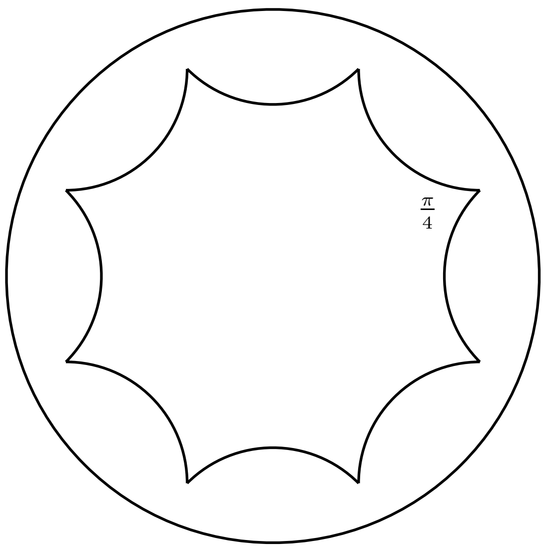
Foundations of Fluid Mechanics

Giovanni Gallavotti



Roma 2000

18/*agosto*/2015; 8:06

Giovanni Gallavotti
Dipartimento di Fisica I^a Università di Roma, *La Sapienza*
P.le A. Moro 2, 00185 Roma, Italia

Tanslation of an Italian original "Ipotesi per una introduzione alla Meccanica dei Fluidi": a paperback copy of the italian edition (free of charge) can be obtained by request (use FAX) to:
CNR-GNFM, via S. Marta 13a, 50139 Firenze, Italia, *tel* +39-055-496912, *fax* +39-055-475915
or its revised versions can be freely downloaded from:
hhtp://ipparco.roma1.infn.it

#######################################

Cover: **Order and Chaos**, (free Author's reinterpretation of a well known geometrical object).

A Daniela

Preface

The imagination is stricken by the substantial conceptual identity between the problems met in the theoretical study of physical phenomena. It is absolutely unexpected and surprising, whether one studies equilibrium statistical mechanics, or quantum field theory, or solid state physics, or celestial mechanics, harmonic analysis, elasticity, general relativity or fluid mechanics and chaos in turbulence.

So that when in 1988 I was made chair of Fluid Mechanics at the Università *La Sapienza*, not to recognize work I did on the subject (there was none) but, rather, to avoid my teaching mechanics, from which I could have a strong cultural influence on mathematical physics in Roma, I was not excessively worried, although I was clearly in the wrong place. The subject is wide, hence in the last decade I could do nothing else but go through books and libraries looking for something that was within the range of the methods and experiences of my past work.

The first great surprise was to realize that the mathematical theory of fluids is in a state even more primitive than I was conscious of. Nevertheless it still seems to me that a detailed analysis of the mathematical problems is essential for any one who wishes research into fluids. Therefore I dedicated (Chap.3) all the space necessary to a complete exposition of the theories of Leray, of Scheffer and of Caffarelli, Kohn and Nirenberg, taken directly from the original works.

The analysis is preceded by a long discussion of the phenomenological aspects concerning the fluid equations and their properties, with particular attention to the meaning of the various approximations. One should not forget that the fluid equations *do not have fundamental nature*, *i.e.* they ultimately are phenomenological equations and for this reason one "cannot ask from them too much". In order to pose appropriate questions it is necessary to dominate the heuristic and phenomenological aspects of the theory. I could not do better than follow the Landau–Lifshitz volume, selecting from it a small, coherent set of properties without (obviously) being able nor wishing to reproduce it (which, in any event, would have been useless), leaving aside most of the themes covered by that rich, agile and modern treatise, which the reader will not set aside in his introductory studies.

In the introductory material (Chaps.1,2) I inserted several modern remarks taken from works that I have come to know either from colleagues or from participating in conferences (or reading the literature). Here and there, rarely, there are a few original comments and ideas (in the sense that I did not find them in the accessed literature).

The second part of the book is dedicated to the qualitative and phenomenological theory of the incompressible Navier–Stokes equation: the lack of existence and uniqueness theorems (in three space dimensions) did not have practical consequences on research, or most of it. Fearless engineers write gigantic codes that are supposed to produce solutions to the equations: they do not care the least (when they are conscious of the problem, which unfortunately seems to be seldom the case) that what they study are *not* the Navier–Stokes equations, but just the informatic code they produced. *No one* is, to date, capable of writing an algorithm that, in an *a priori* known time and within a prefixed approximation, will produce the calculation of any property of the equations solution following an initial datum and forces which are not "very small" or "very special". Statements to the contrary are not rare, and they may appear even on the news: but they are wrong.

It should *not* be concluded from this that engineers or physicists that work out impressive amounts of papers (or build airliners or reentry vehicles) on the "solutions" of the Navier–Stokes equations are dedicating themselves to a useless, or risible, job. On the contrary their work is necessary, difficult and highly qualified. It is, however, important try and understand in which sense their work can be situated in the Galilean vision that wishes that the book of Nature be written in geometrical and mathematical characters. To this question I have dedicated a substantial part of the book(Chap. 4,5): where I expose *descriptive* or *kinematical* methods that are employed in the current research (or, better, in that part of the current research that I manage to have some familiarity with). These are ideas born in the seminal works of Lorenz and Ruelle–Takens, and in part based on stability and bifurcation theory and aim at a much broader and ambitious scope.

Chaotic phenomena are "very fashionable": a lot of ink flowed about them (and many computer chips burnt out) because they attract the attention even of those who like scientific divulgation and philosophy. But their perception is distorted because to make the text interesting for the nonspecialized public, often statements are made which are strong and ambiguous. Like "determinism is over", which is a statement that, if it has some basis of truth, certainly does not underline that nothing changes for those who cherish a deterministic conception of physical reality (a category to which all my colleagues and myself belong) or for those who did cherish it (like Laplace) when the "theory of chaos" was not, yet.

Hence in discussing chaotic properties of the simplest fluid motions I do not investigate at all philosophical themes, nor the semantic interpretation of the words illustrating objective properties. This is so in spite of the "light and non technical" appearance of this part of the book, which is in fact not light at all and it is *very* technical and collects a long sequence of steps, each of which is so simple not to require technical details.

I find it important that anyone is interested in science–related philosophical matters (in Greek times this encompassed all of philosophy but things have changed since; *c.f.r.* [BS98], [Me97]) should necessarily dedicate the time needed for a full understanding of the technical instruments (such as geometry, infinitesimal calculus and Newtonian physics) as already indicated by Galileo. It would be illusory to think one cpuld appreciate modern science without such instruments (*i.e.* "science", which is situated out of the elapsing time and which is called "modern" referring only to some of its "accidents"); divulgence is often terribly close to mystification.

The analysis is set with the aim of studying the initial development of turbulence, following the ideas of Ruelle–Takens, and, mainly, for the introduction and discussion of *Ruelle's principle*. This is a principle that, in my view, has not been appreciated as much as it could, perhaps for its "abstruse nature" or, as I prefer to think, for its originality. I became aware of it at a talk by Ruelle in 1973; I still recall how I was struck by the audacity and novelty of the idea. Since then I started to meditate on how it could lead to concrete applications; a difficult task. In the conclusive Chap. 7 I expose a few recent proposals of applications of the principle.

Section §6.1 is dedicated to the problem of the construction of invariant (*i.e.* stationary) distributions for the Navier–Stokes, equations: collecting from the literature heuristic ideas which seem to me quite interesting, even when far from physical or mathematical applications (or from the solution of the problem).

Kolmogorov "K41" theory cannot be absent in a modern text, no matter how introductory, and it is succinctly discussed in §6.2; while in §6.3 I describe some recent simulations which, in my view, have brought new ideas into the theory of fluids (multifractality): a selection of whose partiality I am aware and which is only partly due to space needs. Partly it is, however, a choice made because it concerns research done in he area of Roma and therefore is more familiar to me.

The last chapter contains several ideas developed precisely while I was teaching the fluid mechanics courses. Often I deal with very recent works which might have no interest at all in a few years from now. Nevertheless I am confident that the reader will pardon my temerity and consider it as a justifiable weakness at the end of a work in which I have limited myself only to classical and well–established results.

I tried to keep the book self-contained, not to avoid references to the literature (that is always present, apart from unavoidable involuntary omissions) but rather to present a unitary and complete viewpoint. Therefore I have inserted, in the form of problems with detailed hints for their solution, a notable amount of results that make the problems perhaps even more interesting than the text itself. I tried use problems with a guided solution to present results that could well have been part of the main text: they are taken from other works or summarize their contents. Students who will consider using the book as an introductory textbook on fluid mechanics should try to solve all the problems in detail, without having recourse to the quoted literature; I think that this is essential in order to dominate a subject that is only apparently easy.[1]

[1] Among the problems one shall find a few classical results (like elementary tides theory) but also (1) phenomenology of nonhomogeneous chemically active continua, (2) Stokes's formula, (3) waves at a free boundary, (4) elliptic equations in regular domains (and the Stokes equation theory), (5) smoke–ring motions, (6) Wolibner–Kato theory for the 2–dimensional Euler equation theory, (7) potential theory needed for Leray's theory,

I wish to thank colleagues and students for the help they provided in correcting my notes. In particular I wish to thank Dr. Federico Bonetto and Dr. Guido Gentile. Some ideas that emerged during endless discussions have influenced the text particularly in the last few sections, and sometimes have avoided errors or unprecise statements.

Roma, 2001.

(8) Sobolev inequalities needed for the CKN theory, (9) several questions on numerical simulations, (10) some details on bifurcation theory, (11) a few comments on continued fractions and on the geodesic flows on surfaces of constant negative curvature, (12) the ergodic theorems of Birkhoff and Oseledec, (13) Lyapunov exponents for hyperbolic dynamical systems, (14) some information theory questions (for entropy). I think that until the last chapter, dedicated to more advanced themes, the only theorem used but not proved (not even with a hint to a proof) is the center manifold theorem (because I did not succeed developing a reasonably short self–contained proof, in spite of its rather elementary nature). Several theorems are hinted at by using a heuristic approach. This is because I find often missing in the literature the heuristic illustration of the ideas, which is generally very simple at least in the simplest nontrivial cases in which they usually were generated.

Indice

CHAPTER I

Continua and generalities about their equations

§1.1 Continua.

A *homogeneous continuum*, chemically inert, in d dimensions is described by

(a) A region Ω in ambient space ($\Omega \subset R^d$), which is the occupied volume.
(b) A function $P \to \rho(P) > 0$, defined on Ω, giving the *mass density*.
(c) A function $P \to T(P)$ defining the *temperature*.
(d) A function $P \to s(P)$ defining the *entropy density* (per unit mass).
(e) A function $P \to \underline{\delta}(P)$ defining the *displacement* with respect to a reference configuration.
(f) A function $P \to \underline{u}(P)$ defining the *velocity field*.
(g) An *equation of state* relating $T(P), s(P), \rho(P)$.
(h) A *stress tensor* $\underset{\approx}{\tau}$, also denoted (τ_{ij}), giving the force per unit surface that the part of the continuum in contact with an ideal surface element $d\sigma$, with normal vector $\underline{n}$, on the side of $\underline{n}$ exercises on the part of continuum in contact with $d\sigma$ on the side opposite to $\underline{n}$, via the formula

$$d\underline{f} = \underset{\approx}{\tau}\,\underline{n}\,d\sigma \qquad (\underset{\approx}{\tau}\,\underline{n})_i = \sum_{j=1}^{d} \tau_{ij} n_j \tag{1.1.1}$$

(i) A *thermal conductivity tensor* $\underset{\approx}{\kappa}$, giving the quantity of heat traversing the surface element $d\sigma$ in the direction of $\underline{n}$ per unit time via the formula

$$dQ = -\underset{\approx}{\kappa}\,\underset{\sim}{n}\,\underline{\partial} T\,d\sigma \tag{1.1.2}$$

(l) A *volume force density* $P \to \underline{g}(P)$.
(m) A relation expressing the stress and conductivity tensors as functions of the observables $\underline{\delta}, \underline{u}, \rho, T, s$.

Relations in (g), (m) are called the continuum *constitutive relations*: in a microscopic theory of continua they must be deducible, in principle, from the atomic model. However in the context in which we shall usually be the constitutive relations have a purely macroscopic character, hence they are phenomenological relations and they must be thought of as essential parts of the considered model of the continuum.
More generally one can consider non homogeneous continua, with more than one chemical components among which chemical reactions may occur: here I shall not deal with such systems, but the foundations of their theory are discussed in some detail in the problems at the end of §1 (*c.f.r.* problems [1.1.7]–[1.1.17]).
We can distinguish between solid and liquid (or fluid) continua. Liquids have a constitutive relation that allows us to express τ in terms of the thermodynamic observables and, furthermore, of the velocity field $\underline{u}$: in other words τ does not depend on the displacement field $\underline{\delta}$.
We always suppose the *validity of the principles of dynamics and thermodynamics*: *i.e.* we assume the validity of a certain number of relations among the observables (listed above) which describe a continuum.
A notation widely used below will be $\underset{\approx}{\tau}$ to denote a tensor τ_{ij}, $i, j = 1, \ldots d$; and $\underset{\approx}{\tau}\,\underline{u}$ to denote the result of the action of the tensor $\underset{\approx}{\tau}$ on the vector $\underline{u}$, *i.e.* the vector whose i-th component is $\sum_j \tau_{ij} u_j$. We shall often adopt the *summation convention over repeated indices*: this means that. for instance, $\sum_{j=1}^3 \tau_{ij} n_j$ will be denoted (unless ambiguous) simply $\tau_{ij} n_j$.
In this way the relations imposed upon the observables describing the continuum by the laws of thermodynamics and mechanics are the following.

(I) Mass conservation.

If Δ is a volume element which in time t evolves into Δ_t it must be

$$\int_\Delta \rho(P,0)dP \equiv \int_{\Delta_t} \rho(P,t)dP \tag{1.1.3}$$

Choosing t infinitesimal one sees that the region Δ_t consists of the points that can be expressed as

$$P' = P + \underline{u}(P)t, \qquad P \in \Delta \tag{1.1.4}$$

and this relation can be thought of as a coordinate transformation $P \to P'$ with Jacobian determinant

$$\det \frac{\partial P'_i}{\partial P_j} = \det\big(1 + \frac{\partial u_i}{\partial P_j} t\big) = 1 + t \sum_{i=1}^3 \frac{\partial u_i}{\partial P_i} + 0(t^2) \tag{1.1.5}$$

so that, neglecting $O(t^2)$:

$$\begin{aligned}\int_{\Delta_t} \rho(P',t)\,dP' = \int_\Delta \rho(P+\underline{u}\,t,t)(1+t\,\underline{\partial}\cdot\underline{u})dP = \\ = \int_\Delta \rho(P)dP + t\int_\Delta (\underline{\partial}\rho\cdot\underline{u}+\partial_t\rho+\rho\underline{\partial}\cdot\underline{u})dP\end{aligned} \tag{1.1.6}$$

hence we find, from (1.1.3):

$$\partial_t\rho + \underline{\partial}\cdot(\rho\,\underline{u}) = 0 \tag{1.1.7}$$

which is the *continuity equation.*

(II) Momentum conservation (I cardinal equation)

$$\frac{d}{dt}\int_\Delta \rho\,\underline{u}\,dP = \int_\Delta \rho\,\underline{g}\,dP + \int_{\partial\Delta} \underset{\approx}{\tau}\,\underline{n}\,d\sigma \tag{1.1.8}$$

To evaluate the derivative one remarks that at time ϑ

$$\begin{aligned}&\int_{\Delta_\vartheta} \underline{u}(P',\vartheta)\,\rho(P',\vartheta)\,dP' = \\ &= \int_\Delta \rho(P+\underline{u}(P)\vartheta,\vartheta)\,(1+\vartheta\underline{\partial}\cdot\underline{u})\,\underline{u}(P+\underline{u}(P)\vartheta,\vartheta)\,dP \\ &\int_{\partial\Delta}(\underset{\approx}{\tau}\,\underline{n})_i\,d\sigma = \int_\Delta\sum_j(\partial_j\tau_{ij})\,dP\end{aligned} \tag{1.1.9}$$

and, therefore (1.1.8) becomes

$$\partial_t(\rho u_i) + \sum_j \partial_j\big(u_j(\rho u_i)\big) = \rho g_i + \sum_{j=1}^3 \partial_j\tau_{ij} \tag{1.1.10}$$

i.e. , by (1.1.7) and the summation convention, we find

$$\partial_t u_j + \underline{u}\cdot\underline{\partial}\,u_j = g_j + \frac{1}{\rho}\partial_k\tau_{jk} \tag{1.1.11}$$

(III) Angular momentum conservation.

This is a property that is automatically satisfied, as a consequence of the definition of stress tensor, (1.1.1): if one allowed a more general stress law $\underline{\tau}_i(\underline{n})d\sigma$, rather than $\tau_{ij}n_jd\sigma$ with a symmetric $\underset{\approx}{\tau}$, one would derive that it imposes that $\tau(\underline{n})_id\sigma$ must have the form $\tau_{ij}n_jd\sigma$, and that $\tau_{ij}=\tau_{ji}$.

Let, indeed, Δ be a set with the form of a tetrahedron with three sides on the coordinate axes and a face with normal vector $\underline{n}$.

Let $\underline{\tau}_1, \underline{\tau}_2, \underline{\tau}_3$ and $\underline{\tau}_n$ be the stresses that act on the four faces, with normal vectors the unit vectors $\underline{i}, \underline{j}, \underline{k}$ of the coordinate axes and $\underline{n}$, respectively.

The angular momentum of Δ with respect to a point $P_0 \in \Delta$ is $\underline{K} = \int_\Delta (P - P_0) \wedge \rho\, \underline{u}\, dP \le 0(\ell^4)$, if ℓ is the diameter of Δ; also the momentum of the volume forces has size $0(\ell^4)$. On the other hand the momentum of the stresses is *a priori* of size $0(\ell^3)$ unless the total force due to the stresses vanishes:[1] hence in order that it be of size of order $0(\ell^4)$ (as it must by consistence to avoid infinite angular acceleration of Δ) it is necessary that a suitable relation between $\underline{\tau}_1, \underline{\tau}_2, \underline{\tau}_3$ and $\underline{\tau}_n$ be verified. To find it note that if the total stress force did not vanish to leading order as $\ell \to 0$, *i.e.* if

$$\underline{0} \ne \underline{\tau}(\underline{n})d\sigma - \big(\underline{\tau}_1 d\sigma_1 + \underline{\tau}_2 d\sigma_2 + \underline{\tau}_3 d\sigma_3\big) \equiv \big(\underline{\tau}(\underline{n}) - \underline{\tau}_1 n_1 - \underline{\tau}_2 n_2 - \underline{\tau}_3 n_3\big) d\sigma \tag{1.1.12}$$

then it would follow that the total force would be $\underline{c}\, d\sigma, \underline{c} \ne 0$, hence the total angular momentum would be of size $0(\ell^3)$ with respect to some point P_0 of Δ. Therefore

$$\tau(\underline{n})_j = \tau_{ji} n_i \tag{1.1.13}$$

(with the summation convention).[2]

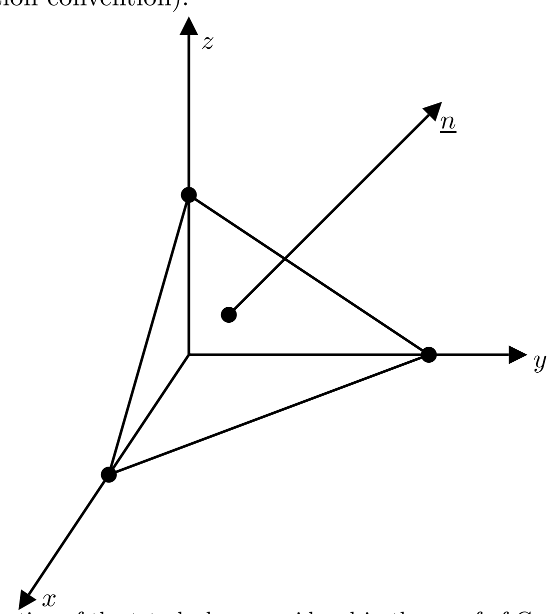


Fig. (1.1.1): Illustration of the tetrahedron considered in the proof of Cauchy's theorem. The unit vectors, $\underline{i}, \underline{j}, \underline{k}$, are not drawn.

Furthermore it is $\tau_{ji} = \tau_{ij}$ as one sees by noting that, at leading order in ℓ, the angular momentum of the stresses on the faces of an *infinitesimal cube*

[1] If the total force does not vanish it has size of order ℓ^2, *i.e.* proportional to the surface area of Δ and then by changing P_0 inside Δ one can find (many) points P_0 with respect to which the momentum of the stresses is of $P(\ell^3)$.

[2] One can also say that this relation follows from the first cardinal equation: indeed if $\underline{c}\, d\sigma \ne \underline{0}$ the total force would have size $0(\ell^2)$ while being equal to the time derivative of the linear momentum which has size $0(\ell^3)$.

with side ℓ, with respect to the cube center, should be of $O(\ell^4)$, as noted above, but it is

$$\ell^3\big[\big(\underline{j}\wedge(\underset{\sim}{\underline{\tau}}\,\underset{\sim}{\underline{j}})\big)+\underline{i}\wedge(\underset{\sim}{\underline{\tau}}\,\underset{\sim}{\underline{i}})+\underline{k}\wedge(\underset{\sim}{\underline{\tau}}\,\underset{\sim}{\underline{k}})\big] \tag{1.1.14}$$

Note that $(\underset{\sim}{\underline{\tau}}\,\underline{j})_i=(\tau_2)_i,(\underset{\sim}{\underline{\tau}}\,\underline{i})_i=\tau_{1i},(\underset{\sim}{\underline{\tau}}\,\underline{u})_i=\tau_{3i}$ are the components of the stresses on the faces having as normals the coordinate unit vectors $\underline{j},\underline{i},\underline{k}$ respectively, and if one imposes that (1.1.14) vanishes, one gets:

$$\begin{aligned}0=&\det\begin{pmatrix}\underline{i}&\underline{j}&\underline{k}\\0&1&0\\\tau_{21}&\tau_{22}&\tau_{23}\end{pmatrix}+\det\begin{pmatrix}\underline{i}&\underline{j}&\underline{k}\\1&0&0\\\tau_{11}&\tau_{12}&\tau_{13}\end{pmatrix}+\det\begin{pmatrix}\underline{i}&\underline{j}&\underline{k}\\0&0&1\\\tau_{31}&\tau_{32}&\tau_{33}\end{pmatrix}\\=&\,\underline{i}\big(\tau_{23}-\tau_{32}\big)+\underline{j}\big(\tau_{13}-\tau_{31}\big)+\underline{k}\big(\tau_{21}-\tau_{12}\big)\end{aligned} \tag{1.1.15}$$

so that $\tau_{ij}=\tau_{ji}$ often called *Cauchy's theorem.*

(IV) Energy conservation.

This is a more delicate conservation law as it involves also the thermodynamic properties of the continuum.

We imagine that every infinitesimal fluid element Δ is a (ideally infinite) system in thermodynamic equilibrium and, hence, with a well defined value of the observables like internal energy, entropy and temperature.., *etc.*

The equation of state, characteristic of the continuum considered, will be a relation expressing the internal energy ε per unit mass in terms of the mass density ρ and of the entropy per unit mass s: $(\rho,s)\to\varepsilon(\rho,s)$.

Then the energy balance in a volume element Δ will be obtained by expressing the variation (per unit time) of its energy (kinetic plus internal)

$$\frac{d}{dt}\int_\Delta(\rho\,\frac{\underline{u}^2}{2}+\rho\,\varepsilon)\,dP \tag{1.1.16}$$

as the work performed by the volume forces, plus the work of the stresses on the boundary of the volume element and, also, plus the heat that penetrates by conduction from the boundary of Δ. This is the sum of the following addends

$$\begin{aligned}&\int_\Delta\rho\,\underline{g}\cdot\underline{u}\,dP+\int_{\partial\Delta}\underset{\sim}{\underline{\tau}}\,\underset{\sim}{\underline{n}\underline{u}}\,d\sigma+\int_{\partial\Delta}\kappa_{ij}(\partial_iT)\cdot n_j\,d\sigma=\\&=\int_\Delta\rho\,\underline{g}\,\cdot\underline{u}\,dP+\int_\Delta\partial_i\left(\tau'_{ij}u_j\right)dP-\int_\Delta\underline{\partial}\cdot(p\,\underline{u})\,dP+\int_\Delta\partial_i(\kappa_{ij}\partial_jT)\,dP\end{aligned} \tag{1.1.17}$$

where we wrote (defining $\underset{\sim}{\underline{\tau}}\,'$) $\tau_{ij}=-p\delta_{ij}+\tau'_{ij}$ with p being the pressure and where we assumed the validity of Fourier's law (1.1.2) for the heat transmission.

Equating (1.1.16),(1.1.17) and using (1.1.11) (multiplied by $\underline{u}$ and integrated over Δ) to eliminate the term with kinetic energy one gets

$$\frac{d}{dt}\int_\Delta \rho\varepsilon dP = \int_\Delta \left[\underset{\approx}{\tau}' \cdot \underline{\partial}\underset{\sim}{u} + \underline{\partial}\cdot(\underset{\approx}{\kappa}\,\underset{\sim}{\partial}T) - p\,\underline{\partial}\cdot\underline{u}\right] dP \tag{1.1.18}$$

and in this relation we recognize that the last term in the r.h.s. is $-p\,d|\Delta|/dt$, *i.e.* it is the work done per unit time by the pressure forces, while the term before the last yields the quantity of heat that enters by conduction in the volume element. The l.h.s. is the variation per unit time of the internal energy. Therefore from the first principle of thermodynamics $dE = dQ - pdV$ we see that the first term in the r.h.s. *must* represent an amount of heat entering the volume element. It is naturally interpreted as the quantity of heat generated by friction forces described by the tensor $\underset{\approx}{\tau}'$.

Note that $\underset{\approx}{\tau}'$ not only contributes to the energy balance through the heat generated per unit volume by friction, *i.e.* $\underset{\approx}{\tau}'\underline{\partial}\underset{\sim}{u}$, but also through the mechanical work $(\underline{\partial}\,\underset{\approx}{\tau}')\underset{\sim}{u}$ per unit volume: such contributions appear, in fact, summed together in (1.1.17) (in the form $\underline{\partial}(\underset{\approx}{\tau}'\underset{\sim}{u}) = \underset{\approx}{\tau}'\underline{\partial}\underset{\sim}{u} + \underset{\sim}{u}\,\underline{\partial}\underset{\approx}{\tau}'$).

This leads, therefore, to interpret $\underset{\approx}{\tau}'$ as an observable associated with the friction forces inside the fluid as well as with the non normal internal stresses.

In order that this interpretation be possible it is necessary, of course, that τ'_{ij} depends solely on the local thermodynamic quantities (s, T) and on the gradient (and, possibly, on the higher order derivatives) of the velocity field and, furthermore, it should vanish if the derivatives of the velocity vanish. Hence in what follows $\tau'_{ij} = 0$ if $\underline{\partial}\,\underset{\sim}{u} = \underset{\approx}{0}$.

The differential form of the (1.1.18) is

$$\partial_t(\rho\varepsilon) + \underline{\partial}\cdot(\rho\varepsilon\underline{u}) \equiv \rho(\partial_t\varepsilon + \underline{u}\cdot\underline{\partial}\varepsilon) = \underset{\approx}{\tau}'\cdot\underline{\partial}\,\underset{\sim}{u} + \underline{\partial}\cdot(\underset{\approx}{\kappa}\,\underset{\sim}{\partial}T) - p\,\underline{\partial}\cdot\underline{u} \tag{1.1.19}$$

having used (in the first identity), the continuity equation.

(V) II^o law of thermodynamics and entropy balance

Eq. (1.1.19) can be combined with the second principle of thermodynamics $T\,dS = dE + p\,d|\Delta|$, and $S = \rho|\Delta|s, E = \rho|\Delta|\varepsilon$, $d|\Delta|/dt = \underline{\partial}\cdot\underline{u}\,|\Delta|$ which gives $T\,d\int_\Delta \rho\,s\,dP = d\int_\Delta \rho\,\varepsilon\,dP - \int_\Delta dP$. One obtains (using also the continuity equation)

$$T\,\rho\left(\partial_t s + \underline{u}\cdot\underline{\partial}\,s\right) = \underset{\approx}{\tau}'\cdot\underline{\partial}\,\underset{\sim}{u} + \underline{\partial}(\underset{\approx}{\kappa}\,\underset{\sim}{\partial}T) \tag{1.1.20}$$

which is the form in which, in applications, energy conservation is often used.

Introducing the *heat current* $\underline{J}_q \stackrel{def}{=} -\underset{\approx}{\kappa}\,\underset{\sim}{\partial}T$ then (1.1.20) can be put in the form

$$\rho\frac{ds}{dt} = -\underline{\partial}\cdot\frac{\underline{J}_q}{T} - \underline{J}_q\cdot\frac{\underline{\partial}T}{T^2} + \underset{\approx}{\tau}'\cdot\frac{1}{T}\underline{\partial}\,\underset{\sim}{u} = -\underline{\partial}\cdot\frac{\underline{J}_q}{T} + \sigma \tag{1.1.21}$$

where σ is interpreted as *entropy density generated per unit volume and unit time.* The σ is also written as

$$\sigma = \sum_j J_j X_j \tag{1.1.22}$$

where X_j is a vector with twelve components consisting in the $-T^{-2}\partial_i T$, $T^{-1}\partial_i u_j$ and J_j consists in the $J_{q\,j}, \underset{\approx}{\tau}'_{ij}$.

Remark: In general thermodynamic forces X_j are identified with parameters measuring how far from thermodynamic macroscopic equilibrium the local state of the fluid is (*e.g.* with a temperature or velocity gradient); and the corresponding currents are identified with the coefficients J_j that allow us to express the entropy creation rate as a linear combination of the forces X_j, *c.f.r.* (1.1.22). It is clear that the identification ("duality") of the forces and the thermodynamic currents is not free of ambiguities because in each problem, given the entropy creation rate σ, one can in general represent σ in several ways in the form (1.1.22). This is an ambiguity analogous to that present in the identification, in mechanics, of the canonical coordinates from a given Lagrangian: one gets different coordinates depending on which are the variables from which one wants to think that the Lagrangian depends.

An unavoidable further problem lies in the fact that a really precise and purely macroscopic definition of entropy creation rate is *not*, to date, well established in systems out of equilibrium unless the systems are *very* close to equilibrium. But this is not the place to enter into a discussion of the foundations of nonequilibrium thermodynamics, see §7.1÷§7.4.

A basic assumption often made in the dynamics of continua is that the relationship between the *thermodynamic forces* X_j and the *currents* or *fluxes* is *linear*, at fixed values of the state observables ρ, p, T

$$J_j = \sum_k L_{jk} X_k \tag{1.1.23}$$

at least if the "thermodynamic forces" X_k are "small". This relation can be combined with the invariance properties for the Galilean transformations, with other possible symmetries (present in fluids with more components and/or chemically active) and with two principles of the thermodynamics of irreversible processes, namely the *Onsager reciprocity* and the *Curie principle.* One obtains, in this way, important restrictions on the tensors $\underset{\approx}{\kappa}, \underset{\approx}{\tau}'$.

Onsager's relations imply that the matrix L is symmetric

$$L_{jk} = L_{kj} \tag{1.1.24}$$

while the Curie's principle says that some among the coefficients L_{jk} vanish. If $L_{jk} = 0$ one says that the current J_j "*does not directly depend on*" (or "*does not directly couple with*") the thermodynamic force X_k. Curie's principle states precisely that the currents J_j that have a vectorial character, *i.e.* are the components of an observable that transforms, under Galilean transformations, as a vector (such as the $\underline{J}_q$), do not couple nor depend directly on thermodynamic forces with different transformation (or "covariance") properties (such as the derivatives $\partial_i u_j$, that have a tensorial character). More generally there is no coupling between thermodynamic forces and currents with different transformation properties with respect to the symmetry groups of the continuum considered.

For instance the matrix κ_{ij} must be symmetric and τ'_{ij} must be (as it already follows from the II^d cardinal equation of dynamics, see above) symmetric and expressible in terms of the derivatives $\partial_i u_j$ via a linear combination of $\partial_i u_j + \partial_j u_i$ e $\delta_{ij}\,\partial_k u_k$ because these are the only tensors that one can form with a linear dependence on the derivatives $\partial_i u_j$

$$\kappa_{ij} = \kappa_{ji}, \qquad \tau'_{ij} = \eta\,(\partial_i u_j + \partial_j u_i) + \eta'\,\partial_k u_k\,\delta_{ij} \tag{1.1.25}$$

see (1.2.6), (1.5.2).

Onsager's relations are a macroscopic consequence of the microscopic reversibility of dynamics, *c.f.r.* [DGM84] and the Curie principle also is rooted on microscopic symmetries, [DGM84].

The second law of thermodynamics is imposed (not without some conceptual difficulties, see problem [1.1.17] below) by requiring that $\sigma \ge 0$: which is obtained by demanding that the tensor $\underset{\approx}{\kappa}$ be positive definite and that $\underset{\approx}{\tau}{}' \cdot \underline{\partial}\,\underset{\sim}{u} \ge 0$, *i.e.* in the case (1.1.25) $\eta, \eta' + 2\eta \ge 0$.

Problems.

[1.1.1]: Let $X(\underline{x},t)$ be a generic observable and define a *current line* as a solution $t \to \underline{x}(t)$ of the equation

$$\dot{\underline{x}} = \underline{u}(\underline{x},t)$$

where $\underline{u}(\underline{x},t)$ is a given velocity field. The *substantial derivative* $\frac{dX}{dt}$, of X, is then defined by the t–derivative of $X(\underline{x}(t),t)$ and it is written as

$$\frac{d}{dt}X(\underline{x}(t),t) = \partial_t X + \underline{u}\cdot\underline{\partial}\,X$$

Show that

$$\rho\frac{d}{dt}X = \partial_t(\rho X) + \underline{\partial}\cdot(\rho X\underline{u})$$

(*Idea:* Use the continuity equation for ρ).

[1.1.2]: Check that the continuity equation can be read as: "the substantial derivative of ρ is $-\rho\underline{\partial}\cdot\underline{u}$".

[1.1.3]: (*A kinetic theory problem*) Consider a monoatomic rarefied gas, whose atoms have mass m and radius σ and occupy the semi space $z>0$. Imagine that the fluid is undergoing isothermal stratified motion with a small shearing velocity field of size $v(z)=z\,v'$ parallel to the x–direction. Let ρ_n be the numerical density (number of atoms per unit volume), λ be the mean free path and $\bar v=(3k_BT/m)^{1/2}$ be the average thermal agitation velocity (k_B is Boltzmann's constant and T is the absolute temperature). Find a heuristic justification, neglecting the horizontal velocity components, for the statement that the number of particles crossing an ideal surface at height $z_0\gg\lambda$ coming from quotae $z>z_0$ and without suffering collisions is, approximately,

$$\int_{z_0}^{z_0+\lambda} dz\int_{-w_z\tau>\lambda}\rho_n\,dxdy\,f(\underline{w})d\underline{w}$$

where $\tau=\lambda/\overline{v}$ is the free flight time, if $f(\underline{w})=e^{\frac{-m\underline{w}^2}{2k_BT}}(\frac{m}{2\pi k_BT})^{3/2}$ is Maxwell's distribution (with k_B representing the Boltzmann's constant.

[1.1.4]: (*kinetic theory for viscosity and heat conductivity)* In the context of problem [1.1.3] deduce that the variations of momentum and thermal kinetic energy (*i.e.* average of $\frac12$ the square of the velocity minus the average velocity) contained in the gas layer at height $z\le z_0$, per unit time and surface, are respectively (we denote by v' the derivative $\frac{dv(z)}{dz}$, by $\overline{v}$ the mean velocity $\overline{v}=(3k_BT/m)^{1/2}$ and the free flight time by $\tau=\lambda/\overline{v}$)

$$\frac{1}{\tau}\int_0^\lambda dh\int_{\lambda/\tau}^\infty\rho_n dw\frac{e^{\frac{-mw^2}{2k_BT}}}{(2\pi k_BT/m)^{1/2}}(2mhv'),$$
$$\frac{1}{\tau}\int_0^\lambda dh\int_{\lambda/\tau}^\infty\rho_n dw\frac{e^{\frac{-mw^2}{2k_BT}}}{(2\pi k_BT/m)^{1/2}}(2\frac32 k_Bh\frac{dT}{dz})$$

Deduce from this that the force per unit surface exerted by the fluid above the height z_0 on the part of the fluid below z_0 is $F=\eta v'$ with:

$$\eta=m\bar v\rho_n\lambda\gamma,\qquad \gamma=\int_{\sqrt3}^\infty e^{-p^2/2}\frac{dp}{\sqrt{2\pi}}$$

Deduce also that the amount of heat crossing per unit time and unit surface the height z_0 is $Q=\kappa\frac{dT}{dz}$ with

$$\kappa=\frac32 k_B\bar v\rho_n\lambda\gamma$$

so that, if the collision cross section is denoted σ^2, it is $\eta=m\frac{\gamma}{\pi\sigma^2}\sqrt{\frac{3kT}{m}}$. (*Idea:* Use the formula (in fact a definition) $\lambda\pi\sigma^2\rho_n=1$ for the mean free path in terms of the atomic diameter σ). See table at the end of the section.

[1.1.5]: (*Clausius–Maxwell relation between specific heat, viscosity and thermal conductivity*) In the context of problems [1.1.3],[1.1.4] assume that the stress tensor of the gas is $\tau'_{ij}=\eta(\partial_iu_j+\partial_ju_i)$ with η constant: compute the force per unit surface that the part of the gas above height z_0 exerts on the part of gas below it and deduce that η is the quantity studied in problem [1.1.4]; and that, therefore, the relation of *Clausius–Maxwell* holds between viscosity, heat conductivity and specific heat at constant volume $c_v\equiv\frac32 RM_A^{-1}$ (if R is the gas constants and M_A is the atomic mass, *e.g.* $4g$ for helium)

$$\kappa=c_v\eta$$

and derive the independence of the viscosity η and of heat conductivity from the density and their proportionality to $\sqrt{T}$. Check that, by refining the calculations, (*e.g.* not neglecting the horizontal components of the velocities) the results only change by numerical factors of $O(1)$ independent on the physical quantities m, ρ, T: in particular the Clausius–Maxwell relation does not change (in rarefied gases).

[1.1.6]: (*energy flux in a perfect fluid*) Show that in a "perfect fluid" (*i.e.* with $\tau', \kappa, g = 0$, *c.f.r.* (1.1.18)) it is

$$\partial_t \int_V \rho\Big(\frac{v^2}{2}+\varepsilon\Big)\,dP = -\int_V \underline{\partial}\cdot\rho\underline{v}\,\Big(\frac{v^2}{2}+w\Big)\,dP = \int_{\partial V}\rho\Big(\frac{v^2}{2}+w\Big)\underline{v}\cdot\underline{n}d\sigma$$

where $w=\varepsilon+p/\rho$. Hence $\rho\Big(\frac{v^2}{2}+w\Big)\underline{v}$ can be interpreted as *energy flux.* (*Idea:* See (1.1.18). Alternatively: if $d\sigma$ is a surface element with external normal $\underline{n}$ the amount of energy crossing $d\sigma$ in the direction $\underline{n}$ is $-\rho(\underline{v}^2/2+\varepsilon)\underline{v}\cdot\underline{n}d\sigma - p\underline{v}\cdot\underline{n}d\sigma$ because the first is the quantity of energy that "exits" through $d\sigma$ per unit time and the second is the work performed through $d\sigma$ by the part of fluid adjacent (but external) to it, per unit time).

[1.1.7] (*mass conservation in mixtures*) Suppose that a fluid consists of a mixture of n different fluids. Let $\rho_1,\ldots,\rho_n$ be the densities and $\underline{u}_1,\ldots,\underline{u}_n$ the respective velocity fields. Then $\rho=\sum_j\rho_j$ will be called the "total density" and $\underline{u}=\rho^{-1}\sum_j\rho_j\underline{u}_j$ the "velocity field" of the fluid. Show that the continuity equations can be written in the form

$$\partial_t\,\rho_k = -\underline{\partial}\cdot(\rho_k\underline{u}_k),\qquad \partial_t\rho=-\underline{\partial}\cdot(\rho\underline{u})$$

We shall set $\underline{J}=\rho\,\underline{u}$ and $\underline{J}_k=\rho_k\,(\underline{u}_k-\underline{u})$: check that if $\frac{d}{dt}\stackrel{def}{=}\partial_t+\underline{u}\cdot\underline{\partial}$ then

$$\frac{d\rho_k}{dt} = -\rho_k\,\underline{\partial}\cdot\underline{u}-\underline{\partial}\cdot\underline{J}_k$$

[1.1.8] (*mass conservation in chemically active mixtures*) In the context of problem [1.1.7] suppose that r chemical reactions are possible between the n species of fluid and that, otherwise, the particles interactions are modeled by hard cores so that the internal energy is entirely kinetic.
If the chemical equation for the j-th reaction is $\sum_{k=1}^n n_{jk}\,[k]=0$, where n_{jk} are stechiometric integers (*e.g.* $2[H_2]+[O_2]-2[H_2O]=0$ involves three species H_2, O_2, and H_2O of molecular mass $2,16,18$ respectively), one defines the *stechiometric coefficients* of the j-th reaction the quantities $\nu_{jk}=m_kn_{jk}$ where m_k is the molecular mass of the k-th species. Then: $\sum_{k=1}^n\nu_{jk}=0$, by mass conservation(*Lavoisier law*), $\sum_k\nu_{jk}\underline{u}_k=\underline{0}$ by momentum conservation and $\sum_k\nu_{jk}\frac12\underline{u}_k^2=\eta_j$ by energy conservation if η_j is the energy yield in the j–th reaction.
Let R_j be the number of chemical reactions of the j-th type that take place per unit volume and unit time (a number which can have either sign: $R_j>0$ means that the reaction proceeds in the direction of transforming molecules with negative stechiometric coefficients into molecules with positive coefficients and viceversa for $R_j<0$). Show that the equations of continuity are modified as

$$\partial_t\,\rho_k=-\underline{\partial}\cdot(\rho_k\underline{u}_k)+\sum_{j=1}^r R_j\,\nu_{jk},\qquad \partial_t\rho=-\underline{\partial}\cdot(\rho\underline{u})$$

Furthermore with the notations of [1.1.7]

$$\frac{d\rho_k}{dt}=-\rho_k\,\underline{\partial}\cdot\underline{u}-\underline{\partial}\cdot\underline{J}_k+\sum_{j=1}^r R_j\,\nu_{jk},\qquad \frac{d\rho}{dt}=-\rho\,\underline{\partial}\cdot\underline{u}$$

Finally setting $c_k = \rho_k/\rho$ it is

$$\rho\frac{d\,c_k}{dt} = -\underline{\partial}\cdot\underline{J}_k + \sum_{j=1}^{r} R_j\,\nu_{jk}$$

[1.1.9] (*momentum conservation in chemically active mixtures*) Check that in fluids with several chemically active components the equation corresponding to the I-th cardinal equation (*i.e.* to momentum conservation) is

$$\rho\,\frac{d\underline{u}}{dt} = -\underline{\partial}\,p + \underset{\sim}{\partial}\,\underset{\approx}{\tau}\,' + \sum_{k=1}^{n}\rho_k\,\underline{g}_k$$

where p is the sum of the partial pressures p_k of each species, $\underset{\approx}{\tau}\,'$ is the sum of the stresses $\underset{\approx}{\tau}\,'_k$ on each species *plus* the tensor $\sum_k \underline{u}_k\,\underset{\sim}{J}_k$, and $\underline{g}_k = -\underline{\partial}V_k$ is the force, with potential energy function V_k, per unit mass acting on the k-th species (which might be species dependent: think, for instance, to a ionized solution in an electric field) *provided* the total potential energy does not change in the chemical reactions (*i.e.* $\sum_k \nu_{jk}\,V_k = 0$). (*Idea:* Write the I-th cardinal equation for each species k:

$$\partial_t(\rho_k\underline{u}_k) + \underset{\sim}{\partial}\cdot(\rho_k\underline{u}_k\underset{\sim}{u}_k) = -\underline{\partial}p_k + \underset{\sim}{\partial}\underset{\approx}{\tau}\,'_k + \rho_k\underline{g}_k + \sum_j R_j\nu_{jk}\underline{u}_k$$

and sum over k taking into account the momentum conservation in [1.1.8].)

[1.1.10] (*energy conservation in chemically active mixtures*)In the context of [1.1.8], [1.1.9] call $\underline{J}_k = (\underline{u}_k - \underline{u})\,\rho_k$ the *diffusion current* of the k-th species, see [1.1.7], and check that the energy conservation equation is

$$\rho\frac{d\varepsilon}{dt} = -p\,\underline{\partial}\cdot\underline{u} + \underset{\approx}{\tau}\,'\,\underset{\sim}{\partial}\underline{u} + \sum_k \underline{g}_k\cdot\underline{J}_k - \underline{\partial}\cdot\underline{J}_q$$

where $\rho\varepsilon \overset{def}{=} \sum_k \rho_k\,(\varepsilon_k + \frac{1}{2}(\underline{u}_k - \underline{u})^2)$ and $\underline{J}_q$ is suitably defined. (*Idea:* The energy in a volume element Δ due to the k–th species is $\int_\Delta \rho_k\,(\varepsilon_k + \frac{1}{2}\underline{u}_k^2)\,d\underline{x}$ if ε_k is the internal energy per particle. We suppose that no interaction takes place between the species other than that giving rise to chemical reactions and other than the hard core pair interaction between the molecules. Then, setting $\vartheta_k \overset{def}{=} (\varepsilon_k + \frac{1}{2}\underline{u}_k^2)$, the energy balance for the k–th species yields

$$\begin{aligned}&\partial_t(\rho_k\vartheta_k) + \underline{\partial}\cdot(\rho_k\vartheta_k\underline{u}_k) = -\underline{\partial}\cdot(p_k\underline{u}_k) + \underset{\sim}{\partial}(\underset{\approx}{\tau}\,'_k\underline{u}_k) +\\ &+ \rho_k\underline{g}_k\cdot\underline{u}_k + \underline{\partial}\cdot(\underset{\approx}{\kappa}\,\underline{u}_k\cdot\underset{\sim}{\partial}T) + \delta_k\end{aligned}$$

where $\underset{\approx}{\kappa}\,\underline{u}_k\cdot\underset{\sim}{\partial}T$ is the heat flux into the species k and δ_k is the total energy variation of the species k due to the chemical reactions, $\delta_k = \sum_j \nu_{jk}(\eta_{jk} + \frac{1}{2}\underline{u}_k^2)$, with η_{jk} being the dissociation energy of the k–th species into the components involved in the j–th reaction so that $\sum_k \delta_k = 0$. Adding and subtracting $\underline{u}$ where appropriate and summing over k one finds

$$\partial_t(\rho\varepsilon+\rho\frac{1}{2}\underline{u}^2)+\underline{\partial}\cdot(\sum_k\rho_k(\varepsilon_k+\frac{1}{2}\underline{u}_k^2)\,(\underline{u}_k-\underline{u})+\rho\varepsilon+\rho\frac{1}{2}\underline{u}^2)=$$

$$=-\underline{\partial}\cdot(p\,\underline{u})-\underline{\partial}\cdot(\sum_k p_k\,(\underline{u}_k-\underline{u}))+\underline{\partial}\cdot(\sum_k \underset{\approx}{\tau}{}'_k\,(\underline{u}_k-\underline{u}))+\underline{\partial}(\underset{\approx}{\tau}{}'\underset{\sim}{u})+$$

$$+\sum_k\rho_k\underline{g}_k\cdot(\underline{u}_k-\underline{u})+(\sum_k\rho_k\underline{g}_k)\cdot\underline{u}+\underline{\partial}\cdot(\underset{\approx}{\kappa}\cdot\underset{\sim}{\partial}T)$$

hence, from [1.1.9], [1.1.8], one gets the above result with

$$\underline{J}_q\stackrel{def}{=}\sum_k\Big((\varepsilon_k+\frac{1}{2}\underline{u}_k^2)\,\underline{J}_k+\underset{\approx}{\tau}{}'_k(\underset{\sim}{u}_k-\underset{\sim}{u})-p_k\,(\underline{u}_k-\underline{u})\Big)-\underset{\approx}{\kappa}\underset{\sim}{\partial}T$$

see also [DGM84].)

[1.1.11] (*heat transport in chemically active mixtures*)The second law of thermodynamics, in the case of chemically active systems, takes the form $TdS=dU+pdV-\sum_k\mu_k\,d(\rho_kV)$ where μ_k is the chemical potential per unit mass of the k-th species. Proceeding as in *(V)* show that [1.1.8],[1.1.9],[1.1.10] imply

$$\rho\,(\partial_ts+\underline{u}\cdot\underline{\partial}s)=-\frac{1}{T}\underline{\partial}\cdot\underline{J}_q+\frac{1}{T}\underset{\approx}{\tau}{}'\cdot\underset{\sim}{\partial}\underset{\sim}{u}+$$
$$+\frac{1}{T}\sum_k\Big(\mu_k\underline{\partial}\cdot\underline{J}_k-\sum_j\nu_{jk}R_j\,\mu_k+\underline{g}_k\cdot\underline{J}_k\Big)$$

(*Idea:* Combine [1.1.10] with the second of [1.1.8] or, using the invariance of ρV (mass conservation) with the third of [1.1.8]).

[1.1.12] (*entropy flow in chemically active mixtures*) Set, see also [1.1.10]

$$A_j=\sum_k\mu_k\nu_{jk},\qquad \underline{J}_s=\frac{1}{T}\,(\underline{J}_q-\sum_{k=1}^n\mu_k\underline{J}_k)$$
$$\sigma=-\,\underline{J}_q\cdot\frac{\underline{\partial}T}{T^2}-\sum_{k=1}^n\underline{J}_k\cdot(\underline{\partial}\frac{\mu_k}{T}-\frac{1}{T}\underline{g}_k)+\underset{\approx}{\tau}{}'\cdot\frac{1}{T}\underset{\sim}{\partial}\underset{\sim}{u}-\sum_{j=1}^rR_j\frac{A_j}{T}$$

and check that the entropy balance equation in [1.1.11], generalizing (1.1.21), can be written

$$\rho\frac{ds}{dt}=-\underline{\partial}\cdot\underline{J}_s+\sigma$$

and (therefore, brushing aside conceptual problems on the identification of the various terms in the balance equation) the quantity $\underline{J}_s$ can be interpreted as the "*entropy current*" transported by the velocity fields, while σ can be interpreted as the quantity of entropy generated per unit volume (by the irreversible processes that develop during the fluids motions). If this interpretation is accepted then the second law of irreversible thermodynamics requires that $\sigma\le0$. If this looks too strict an interpretation one should at least have that $\Gamma=\int\sigma\,dP\le0$.

[1.1.13] (*entropy creation and thermodynamic forces and fluxes*) Check that, defining the "*thermodynamic forces*" $\underline{X}$ and the "*thermodynamic currents*" or "*fluxes*" as

$$\underline{X}=(-\frac{\partial_iT}{T^2},-\partial_i\frac{\mu_k}{T}+g_{ki},\frac{\partial_iu_j}{T},-\frac{A_j}{T}),\quad \underline{J}=(J_{qi},J_{ki},\tau'_{ij},R_j)$$

the entropy generated per unit volume σ, defined in [1.1.12], can be written

$$\sigma = \sum_j J_j X_j$$

thus extending to chemically active multicomponent fluids the results of the theory of homogeneous fluids. Formulate the Curie principle and Onsager reciprocity relations for such fluids. (*Idea:* They are "the same").

[1.1.14] (*entropy creation and constant transport coefficients)* Suppose that in a n components fluid the relation between thermodynamic forces and currents is linear: $J_i = \sum_k L_{ik} X_k$, and that L_{ik} are *constants* and satisfy Onsager relations. Then the *entropy production* per unit time $\Gamma \stackrel{def}{=} \int_\Omega \sigma\, dP$ has time derivative

$$\dot\Gamma = 2\int_\Omega \sum_j J_j \cdot \partial_t X_j\, dP = 2\int_\Omega \sum_j \partial_t J_j \;\cdot\; X_j\, dP$$

(*Idea:* $\Gamma = \int \sum_{jk} L_{jk} X_j X_k\, dP$ and differentiate).

[1.1.15] (*completeness of the equations for mixtures*) Consider the system in [1.1.8] and check that the number of equations equals the number of unknowns, listing the variables and the equations chosen. (*Idea:* For instance we can describe the system by the densities ρ_k, the velocity fields $\underline{u}_k$, the internal energies ε_k; then the first of [1.1.8] are equations for ρ_k, the [1.1.9] gives an equation for the $\underline{u}_k$ and [1.1.10] gives the equation for the ε_k. The equations of state $s_k = s_k(\varepsilon_k, \rho_k)$ and $\mu_k = \mu_k(\varepsilon_k, \rho_k)$ (which are not independent) give the entropy and the chemical potentials, hence the temperature $T = \frac{\partial s_k}{\partial \varepsilon_k}$ (which we have assumed to be the same for all species) and the partial pressures; one also needs the constitutive equations expressing the stresses and, more generally, the fluxes in terms of the forces (*i.e.* the matrix L) so that for instance $R_j = \sum_{j'} L_{jj'} \sum_k \nu_{j'k}\mu_k$ ("*law of mass action*").)

[1.1.16] (*Prigogine's principle*) Consider n fluids in mechanical equilibrium ($p = cost$ and $\underline{u} = \underline{0}$), with boundary conditions in which T is constant in time (at every point of the boundary) and the diffusive current of the k-th species vanishes ($\underline{J}_k = \underline{0}$ at every boundary point). Assume (a strong assumption) that L_{jk} are constants and that there are no volume forces ($\underline{g}_k = \underline{0}$). Taking into account that $\sum_k \underline{J}_k = \underline{0}$, check that the entropy produced per unit time and volume is

$$\sigma = \underline{J}_q \cdot \underline{\partial}\frac{1}{T} - \sum_{j=1}^{r} R_j \frac{A_j}{T} - \sum_{k=1}^{n-1} \underline{J}_k \cdot \underline{\partial}\frac{\mu_k - \mu_n}{T}\,.$$

Check then that the states that make the entropy production Γ, cf. problem [1.1.13], stationary (with respect to the variations of the forces $\underline{X}$) are time independent states (Prigogine). (*Idea:* Lagrange equations for the minimum are $L\underline{X} = \underline{0}$, *i.e.* $\underline{J} = \underline{0}$: hence $R_j = 0, \underline{J}_q = \underline{0}, \underline{J}_k = \underline{0}$ and, therefore, $\frac{d\rho_k}{dt} = 0$, $\frac{d\varepsilon}{dt} = 0$, $\frac{ds}{dt} = 0$ by [1.1.8],[1.1.9],[1.1.10].)

[1.1.17] (*Prigogine's minimal entropy production*) Show that the time independent states of the n fluids in [1.1.16] which minimize (in a strict sense) the entropy production and that are states of mechanical and thermal equilibrium (*i.e.* with p, T constants as a function of time, with $\underline{u} = \underline{0}$ and with a Gibbs function per unit mass $g = \varepsilon - Ts + p\rho^{-1} \equiv \sum_k \mu_k c_k$ which is a strict minimum at every fluid point) are states in stable equilibrium among the thermal and mechanical equilibrium states if the fluids can be regarded as perfect gases (Prigogine). (*Idea:* Imagine perturbing the state by slightly varying T and c_k keeping $\underline{\partial} p, \underline{u} = \underline{0}$; then the system evolves and one has, by [1.1.14]

$$\begin{aligned}\dot\Gamma =&2\int_\Omega \Big(\underline{J}_q\cdot\partial_t\underline{\partial}\frac1T-\sum_{k=1}^{n-1}\underline{J}_k\cdot\partial_t\underline{\partial}\cdot\frac{\mu_k-\mu_n}{T}-\sum_{j=1}^r R_j\partial_t\frac{A_j}{T}\Big)\,dP=\\ =&2\int_\Omega\Big(-\underline{\partial}\cdot\underline{J}_q\partial_t\frac1T+\sum_{k=1}^{n-1}\big(\underline{\partial}\cdot\underline{J}_k-\sum_{j=1}^r R_j\nu_{jk}\big)\,\partial_t\frac{\mu_k-\mu_n}{T}\Big)\,dP\end{aligned}$$

And recalling that $\sum_k \nu_{jk} = 0$ and the continuity equations for the concentrations c_k (in [1.1.8] and in the hint to [1.1.11])

$$\begin{aligned}\dot\Gamma =&2\int_\Omega\Big(-\underline{\partial}\cdot\underline{J}_q\partial_t\frac1T+\sum_{k=1}^{n-1}\Big(\underline{\partial}\cdot\underline{J}_k-\sum_{j=1}^r\nu_{jk}R_j\Big)\partial_t\frac{\mu_k-\mu_n}{T}\Big)\,dP=\\ =&2\int_\Omega(-\underline{\partial}\cdot\underline{J}_q\partial_t\frac1T-\sum_{k=1}^{n-1}\rho\,\frac{dc_k}{dt}\,\partial_t\frac{\mu_k-\mu_n}{T}\Big)\,dP=2\int_\Omega\Big(\Big(-\underline{\partial}\cdot\underline{J}_q-\\ &-\sum_{k=1}^{n-1}\rho\,\frac{dc_k}{dt}\,(\mu_k-\mu_n)\Big)\,\partial_t\frac1T-\frac{\rho}{T}\sum_{h,k=1}^{n-1}\partial_{c_h}(\mu_k-\mu_n)\,\frac{dc_h}{dt}\frac{dc_k}{dt}\Big)\,dP\end{aligned}$$

and by our time independence assumption it is $\frac{dc_h}{dt}=\partial_t c_h$; note that $T^{-1}(\mu_k-\mu_n)$ depends only on c_k, c_n by the perfect gas assumption (in fact in a perfect gas of mass m_k and temperature T the chemical potential is $\mu_k = k_B T\,(\log\rho_k - \frac32\log(k_BT)^{-1} - \frac32\log m_k)$).
However in thermodynamics it is $TdS = dU + pdV - \sum_k \mu_k d(c_k\rho V)$ (note that it is convenient to introduce c_k because $\rho|\Delta|$ is constant in time, while $\rho_k|\Delta|$ is not such, by the second equation in [1.1.8]).
Or: $Tds = d\varepsilon + pd\rho^{-1} - \sum_k \mu_k dc_k$, so that (recalling that $\underline{u} = \underline{0}$) one gets that $(-\underline{\partial}\cdot\underline{J}_q - \sum_{k=1}^{n-1}\rho\,\partial_t c_k\,(\mu_k-\mu_n))\partial_t\frac1T$ is equal, by the equation in [1.1.10], which in the present case becomes $\rho\partial_t\varepsilon = -\underline{\partial}\cdot\underline{J}_q$, to $\rho\partial_t\varepsilon - \rho\sum_k(\mu_k-\mu_n)\partial_t c_k \equiv \rho C\partial_t T$, to the heat generated per unit volume and C is the heat capacity during the transformation.
The thermal equilibrium condition is that in a volume Δ the function $G = U + p|\Delta| - TS = \sum_k \mu_k\rho_k V$ (hence, see [1.1.11], $dG = -SdT + |\Delta|dp + \sum_k \mu_k d\rho_k|\Delta|$) be a minimum at fixed T, p. Therefore in this case $dg = \sum_{k=1}^{n-1}(\mu_k-\mu_n)dc_k$ (where $g = \varepsilon - Ts + p\rho^{-1}$), and one sees that the quadratic form $(M\delta c,\delta c) = \sum_{h,k=1}^{n-1}\partial_{c_h}(\mu_k-\mu_n)\,\delta c_h\delta c_k$ must be positive definite. Hence

$$\dot\Gamma = 2\int\Big(-\frac{\rho C}{T^2}(\partial_t T)^2-\frac{\rho}{T}(M\partial_t c,\partial_t c)\Big)\,dP\le 0$$

in the states that are obtained by a small perturbation of an equilibrium state with minimal Γ.)

[1.1.18] The results on the minimality properties of Γ extend to cases more general than those treated in [1.1.16] but one cannot avoid the condition that the coefficients L_{ik} are constants, and therefore one cannot, on the basis of the above discussion, formulate a universal principle stating that the time independent states in a multicomponent fluid are obtained by *minimizing the entropy production* compatibly with the boundary condition and with the acting forces (because the constancy of L_{jk} is a rather restrictive assumption which is often not satisfied, not even approximately, *c.f.r.* [1.1.5]). One can then infer that there must be some conceptual problem in the interpretation of the thermodynamics of fluids? for instance the definition of entropy produced per unit mass in [1.1.12] and (1.1.21) is not completely free of ambiguities and it has a phenomenological nature. Hence a possible refoundation of nonequilibrium thermodynamics will have to be based on the principles of mechanics rather than on an extension, in some sense arbitrary, of the macroscopic equilibrium thermodynamics. It is conceivable that a generalization of classical thermodynamics to nonequilibrium phenomena (even if statistically stationary) may simply not be possible, at least not without a deep revision of the basic concepts. Nothing, in fact, allows us to believe that a so simple and deep theory, such as equilibrium thermodynamics, is really susceptible of extensions to other nonequilibrium phenomena with the exception of a few cases which are very special (even though very important).

Bibliography: [LL71], [DGM84]. Problems [1.1.7] ÷ [1.1.17] provide a concise exposition of the first 82 pages of [DGM84].

From [LL71]	*kinematic viscosity*	*thermal conductivity*	*Prandtl number*
	$\nu\ cm^2/sec$	$\chi\ cm^2/sec$	ν/χ
Air	$1.50\,10^{-1}$	$2.05\,10^{-1}$	$7.33\,10^{-1}$
Water	$1.00\,10^{-2}$	$1.48\,10^{-3}$	6.75
Alcohol	$2.20\,10^{-2}$	$1.33\,10^{-3}$	$1.66\,10^{+1}$
Glycerine	6.8	$9.38\,10^{-4}$	$7.25\,10^{+3}$
Mercury	$1.20\,10^{-3}$	$2.73\,10^{-2}$	$4.40\,10^{-2}$

§1.2 Equations of motion of a fluid in general. Ideal and incompressible cases. Incompressible Euler, Navier–Stokes and Navier–Stokes-Fourier equations.

At an *internal point* P in the region Ω the equations of motion of a fluid described by the fields $\underline{u}$, T, p are therefore (see §1.2)

$$\begin{aligned}
&(1)\qquad \partial_t\rho+\underline{\partial}\cdot(\rho\underline{u})=0\\
&(2)\qquad \rho\big(\partial_t\underline{u}+\underset{\sim}{\underline{u}}\cdot\underset{\sim}{\underline{\partial}}\underline{u}\big)=-\underline{\partial}p+\rho\underline{g}+\underset{\sim}{\underline{\partial}}\,\underset{\approx}{\underline{\tau}}{}'\\
&(3)\qquad T\rho\big(\partial_t s+\underline{u}\cdot\underline{\partial}s\big)=\underset{\approx}{\underline{\tau}}{}'\underset{\sim}{\underline{\partial}}\underline{u}+\underline{\partial}(\underset{\approx}{\underline{\kappa}}\,\underset{\sim}{\underline{\partial}}T)\\
&(4)\qquad s=s(\rho,\varepsilon),\qquad T^{-1}=\partial_\varepsilon s(\rho,\varepsilon),\qquad p=-T\rho^2\partial_\rho s(\rho,\varepsilon)\\
&(5)\qquad \tau'_{ij}=\theta_{ij}(\underset{\sim}{\underline{\partial}}\underline{u},\rho,T),\qquad \theta_{ij}(0,\rho,T)\equiv 0\\
&(6)\qquad \kappa_{ij}=\xi_{ij}(\rho,T)
\end{aligned}\tag{1.2.1}$$

and the equationa (4) (equation of state), (5), (6) (constitutive equation) imply that (1), (2), (3) are a system of five equations for the five unknowns $\underline{u},T,\rho$.

Equations (5), (6) could be more general, if one allowed the stress tensor to depend also on the higher order derivatives of the velocity field $\underline{u}$, or if one allowed the thermal conductivity tensor, too, to depend on the derivatives of $\underline{u}$. As a rule we shall not consider so general models (see, however, §7.4).

One should note that changing frame of reference the equations (1.2.1) remain invariant: in fact if $t \to \rho(\underline{x},t)$, $T(\underline{x},t)$, $\underline{u}(\underline{x},t)$ is a solution of (1), (2), (3) then $t \to \rho'(\underline{x}',t)$, $T'(\underline{x}',t)$, $\underline{u}'(\underline{x}',t)$, with

$$\rho'(\underline{x}',t) = \rho(\underline{x}'+\underline{v}t,t), \quad T'(\underline{x}',t) = T(\underline{x}'+\underline{v}t,t), \quad \underline{u}'(\underline{x}',t) = \underline{u}(\underline{x}'+\underline{v}t,t)-\underline{v} \tag{1.2.2}$$

gives the motion as seen from an inertial reference frame moving with velocity $\underline{v}$ and coinciding with the preceding frame at time $t = 0$. One checks immediately that (1.2.2) solves (1.2.1): the main point is, naturally, that velocity appears only via its derivatives in the constitutive equations.

This invariance property could not possibly hold if the dependence of the constitutive equations on velocity did not manifest itself through the derivatives of $\underline{u}$: it is (also) for this reason that one does not (usually) consider constitutive equations in which there is an explicit dependence on $\underline{u}$ (and not just on its derivatives).

The functions $s(\rho,\varepsilon)$ are not arbitrary but they must satisfy conditions imposed by the laws of statistical thermodynamics and of statistical mechanics: for instance $\rho^{-1}s(\rho,\varepsilon)$ must be a convex function of its arguments, monotonically increasing in ε and decreasing in ρ, see [Ga99a].

Let us examine the class of particular cases of the (1.2.1) in which ρ is constrained to stay constant: these are the incompressible fluids.

(A) Incompressible non viscous fluid (Euler equations).

The simplest such fluids are the non viscous ($\tau' = 0$) non conducting ($\kappa = 0$) ones:

$$\tau'_{ij} = 0\ , \qquad \kappa_{ij} = 0\ . \tag{1.2.3}$$

In these cases, since ρ is everywhere constant, the (1.2.1) become:

$$\begin{aligned}
&(1) \qquad \underline{\partial}\cdot\underline{u} = 0\\
&(2) \qquad \partial_t\underline{u} + \underset{\sim}{u}\cdot\underset{\sim}{\partial}\underline{u} = -\frac{1}{\rho}\underline{\partial}p + \underline{g}\\
&(3) \qquad \partial_t s + \underline{u}\cdot\underline{\partial}s = 0\\
&(4) \qquad s = \sigma(T)
\end{aligned} \tag{1.2.4}$$

where we chose to think the entropy as a function of T since σ is not singular when $(\partial p/\partial\rho) = \infty$ (a relation expressing incompressibility).

Property (3) gives $(ds/dt) = 0$: *i.e.* s is constant along the lines of current of an incompressible fluid. Hence the case of a fluid which at the initial time is "*isoentropic*", *i.e.* the case $s(\underline{x},0) = s_0 =$ constant, is particularly interesting. This is in fact a property that remains true as time evolves and the temperature will be given at every point by a constant $T = f(s_0)$ and, therefore, it disappears from the equations of motion

$$\begin{aligned}
&\underline{\partial}\cdot\underline{u} = 0\\
&\partial_t\underline{u} + \underset{\sim}{u}\cdot\underset{\sim}{\partial}\underline{u} = -\frac{1}{\rho}\underline{\partial}p + \underline{g}
\end{aligned} \tag{1.2.5}$$

and we get four equations for the four unknowns $\underline{u}, p$, that are called *Euler equations*.

One should, however, ask the question of how could eq. (1.2.4) be possibly related with a real fluid. Density, in such a fluid, would be fixed and entropy would be a function of the temperature alone: but the equation of state would then determine the pressure and, therefore, (1.2.4) would be over determined. For instance if at the initial time s, hence T, were constant over the whole volume, they they would remain constant, hence p would be constant as well (being a function of them) so that $\underline{\partial}\, p = \underline{0}$ and we would have four equations for the three unknowns $\underline{u}$.

This means that (in the incompressible case) the quantity p that appears in (1.2.4) cannot be naively identified with the pressure in the physical sense of the word and, consequently, the interpretation of (1.2.4) is more delicate than it looks at first, *c.f.r.* remarks following (1.3.10).

(B) Incompressible, non heat conducting, viscous fluid (Navier–Stokes equations).

The next simplest case is that of a viscous incompressible non conducting fluid: in this case $\underset{\approx}{\kappa} = \underset{\approx}{0}$ but $\tau'_{ij} \neq 0$. Since τ'_{ij} must vanish for $\partial_i u_j = 0$ the simplest model is the one corresponding to the constitutive equation:

$$\tau'_{ij} = \eta\,(\partial_i u_j + \partial_j u_i) + \eta'\,\underline{\partial}\cdot\,\underline{u}\,\delta_{ij} \qquad\qquad \kappa = 0 \tag{1.2.6}$$

with scalar η, η' depending only on ρ and s (and not on the derivatives of $\underline{u}$), which can be intended as a first order term of a series of τ'_{ij} in powers of $\underset{\sim}{\underline{\partial}\,\underline{u}}$ in which the higher order terms are neglected. The coefficient η is a function of ρ and s which is usually called *dynamic viscosity* while one calls $\nu = \eta/\rho$ *kinematic viscosity*. Incompressibility is expressed by $\underline{\partial}\cdot\underline{u} = 0$ and hence, in incompressible cases, the second term can be omitted.

An incompressible fluid with constitutive equation given by (1.2.6) and $\nu=$ constant is called an *incompressible Navier–Stokes fluid*, or "*NS-fluid*", and it is described by the equations

$$\begin{aligned}
&(1) \qquad \underline{\partial}\cdot\underline{u} = 0\\
&(2) \qquad \partial_t\underline{u} + \underset{\sim}{\underline{u}}\cdot\underset{\sim}{\underline{\partial}}\underline{u} = -\frac{1}{\rho}\underline{\partial}p + \nu\Delta\underline{u} + \underline{g}\\
&(3) \qquad \rho T(\partial_t s + \underline{u}\cdot\underline{\partial}s) = \frac{\eta}{2}\sum_{ij}(\partial_i u_j + \partial_j u_i)^2\\
&(4) \qquad s = \sigma(T)
\end{aligned} \tag{1.2.7}$$

The first two equations should determine p and $\underline{u}$ while the fourth establishes a suitable relation σ between s and T which allows us to compute, via (3) and by integration along the current lines, the entropy density s starting form its initial value. The problem decouples and the "real" equations are the first two, called the *Navier–Stokes equations* or *NS–equations.*

The interpretation problem mentioned immediately after (1.2.5) evidently remains in the present case.

However one should remark that if the fluid is enclosed in a container Ω the above equations might lead to results that might be physically unacceptable: for onstance if the NS equations are subject to a boundary condition $\underline{u} = \underline{0}$ then the third equation shows that it will not be in general possible to fix the temperature at the boundary because the equation would imply that $\partial_t s > 0$ on the boundary and s (hence T) will increase with time and could not be held fixed. Hence in presence of thermoconduction the above equations will be an acceptable model only in special cases; see Sec. 5 for a treatment of the problem in presence of boundaries. The incompressible NS equations will therefore be acceptable only if the internal generation of heat can be completely neglected either because "small" or because it is assumed to be removed by some mechanism which is not described by the equations themselves or which is implemented through special boundary conditions, see Sec. 5.

(C) Incompressible thermoconducting viscous fluid

More difficult is the description of an incompressible fluid which, besides being viscous is also thermoconductor. If the fluid density did not depend from the temperature, then the equations for $\underline{u}$ would be identical to the Navier–Stokes equations with constant density (incompressibility means that the constant density is also pressure independent). In this case entropy should depend only on the temperature, $\sigma(T) = \int c\, dT/T$ for some function $c = c(T)$, and (1.2.1) should become the equations:

$$\begin{aligned}
&(1) \qquad \underline{\partial} \cdot \underline{u} = 0 \\
&(2) \qquad \partial_t \underline{u} + \underset{\sim}{\underline{u}} \cdot \underset{\sim}{\underline{\partial}} \underline{u} = -\frac{1}{\rho} \underline{\partial} p + \nu \Delta \underline{u} + \underline{g} \\
&(3) \qquad \rho\, c\, (\partial_t T + \underline{u} \cdot \underline{\partial} T) = \frac{\eta}{2} \sum_{ij} (\partial_i u_j + \partial_j u_i)^2 + \kappa \Delta T
\end{aligned} \tag{1.2.8}$$

with $\rho = constant$. Hence the problem *seems* again to decouple into the temperature independent one of solving (1) and (2) and then into the one of solving (3) which is the Fourier equation in presence of transport of matter. The interpretation problems mentioned for the Euler and NS equations are still present.

In conclusion physical conditions under which one can assume with good approximation a constant density with a varying temperature are quite rare in applications to fluids.

For instance in convection problems variability of density as a function of temperature is essential: see the analysis in the following §1.3, §1.5.

The (1.2.8), which have therefore a rather limited interest, will be called "Navier–Stokes–Fourier equations".

(D) On the physical meaning of an incompressibility condition.

Since in real fluids it is $(\partial p/\partial\rho)_s < \infty$ we must ask in which cases a real fluid can be considered as incompressible.

To evaluate qualitatively the meaning of an incompressibility hypothesis and its possible validity one can have recourse to *dimensional considerations*

The idea behind the analysis is: imagine that the fluid had a motion which is regular and which is characterized by a "typical velocity variation" δv in the sense that velocity varies of the order of magnitude δv with respect to its average over space and time. Likewise imagine that δT is a "typical temperature variation" with respect to the average temperature and δp is a "typical variation of pressure", *etc.* Furthermore imagine that the above variations show up on a length scale of size l or on a time scale τ.

In this situation quantities like $\underline{\partial}\underset{\sim}{\underline{u}}$, $\partial_t\underline{u}$, $\Delta\underline{u}$, $\underline{\partial}T$, $\partial_t T$, ΔT, $\underline{\partial}p$ can be estimated to have "typically" size of order of magnitude

$$|\underline{\partial}\underset{\sim}{\underline{u}}| \sim \frac{\delta v}{l}\,,\ |\partial_t\underline{u}| \sim \frac{\delta v}{\tau}\,,\ |\Delta\underset{\sim}{\underline{u}}| \sim \frac{\delta v}{l^2}\,,\ |\underline{\partial}T| \sim \frac{\delta T}{l}\,,\ |\Delta T| \sim \frac{\delta T}{l^2},\ \ |\partial p| \sim \frac{\delta p}{l} \tag{1.2.9}$$

which are interpreted as a maximal order of magnitude for such quantities.

Since we suppose that $\underline{u}$, ρ, T are related via the equations of motion it follows that certain relations must hold among the various quantities in (1.2.9). More precisely there must exist instants in which a given term of the equations has the same order of magnitude of any other (otherwise if, for instance, a term was always (much) smaller than another *we could neglect it* and the equation would become simpler).

Hence, for instance, since $\partial_t\underline{u}$ + other terms $= -\rho^{-1}\underline{\partial}p$ (*c.f.r.* (1.2.1), eq. (2)), the remark is that *in some instant and in some point it must be* $\tau^{-1}\delta v \sim \rho^{-1}\delta p/l$; *and since* $\ldots + \underset{\sim}{u}\cdot\underset{\sim}{\underline{\partial}}\underline{u} + \ldots = -\rho^{-1}\underline{\partial}p$ *there will be an instant and a point where* $(\delta v)^2/l \sim \delta p/\rho l$.

In the isoentropic case, *i.e.* when $(\partial\rho/\partial s)_p = -\rho^2(\partial T/\partial p)_s \equiv \rho^2\chi_s$ can be neglected in evaluating density variations,[1], one finds

$$\frac{\Delta\rho}{\rho} \cong \left(\frac{\partial\rho}{\partial p}\right)_s \frac{\delta p}{\rho} \tag{1.2.10}$$

and we realize that the condition of validity of the incompresssibility assumption, *i.e.* $\Delta\rho/\rho \ll 1$, can be obtained by estimating the largest values that $\delta p/\rho$ can take.

(E) The case of incompressible Euler equations

In the case of Euler equations (1.2.5), with $\underline{g} = 0$ for simplicity, the above remark tells us that $\delta p/l\rho$ can reach the following two sizes

$$\frac{\delta p}{l\rho} \sim \frac{\delta v}{\tau} \quad \text{or} \quad \frac{\delta p}{l\rho} \sim \frac{(\delta v)^2}{l} \tag{1.2.11}$$

[1] Where χ_s is the coefficient of heating in adiabatic compressions and the identity is derived from $\delta w = T\,\delta s + \rho^{-1}\delta p$ if w is the *enthalpy* per unit mass.

hence the condition $\Delta\rho/\rho \ll 1$ becomes $\left(\frac{\delta v}{v_s}\right)\frac{l}{\tau v_s} \ll 1$ and $\left(\frac{\delta v}{v_s}\right)^2 \ll 1$ where $v_s^{-2} = (\partial\rho/\partial p)_s$ and v_s has, as it is well known from elasticity theory, the meaning of speed of sound $v_s \equiv v_{sound}$ in the fluid. Hence the condition (1.2.10) becomes

$$\frac{\delta v}{v_{sound}} \ll 1, \quad \frac{l}{\tau v_{sound}} \ll 1 \tag{1.2.12}$$

This relation is interpreted as follows; the Euler fluid can be considered "incompressible" *if the velocity variations are small with respect to the sound speed* and, furthermore, *if the variations manifest themselves over a length scale small with respect to the length run at sound speed over a the time scale over which velocity variations are appreciable.*

(F) The case of the Navier–Stokes equations.

In the case of the NS equations, still with $\underline{g} = \underline{0}$, the new term $\nu\,\Delta\underline{u}$ in the second of (1.2.7) adds to (1.2.11) a new comparison term $\nu\,\delta v\,l^{-2}$: hence the (1.2.10) adds $(\delta v)\nu/v_s^2 l \ll 1$ to the conditions (1.2.12) the $(\delta v)\nu/v_s^2 l \ll 1$; or

$$\frac{\nu}{v_{sound} l} \ll 1 \tag{1.2.13}$$

Since the second of the (1.2.8) coincides with the second of the (1.2.7) and it is the only equation containing p, we see that also (1.2.12), (1.2.13) are all the conditions of validity of the incompressibility assumption, under the hypothesis that the coefficient of heating by compression is neglegible.

If, finally, one supposes that in the considered equations it is also $\underline{g} \ne \underline{0}$ we add a new term in the (2) and we see that δp can also become such that $\delta p/\rho l \sim g$ thus leading to the further condition

$$\frac{lg}{v_{sound}^2} \ll 1 \tag{1.2.14}$$

Which means that, in presence of gravity, *the speed acquired in a free fall from a height equal to the characteristic length over which velocity changes must be small compared to the sound speed in order that the fluid could be considered as incompressible*; hence (1.2.12), (1.2.13) and (1.2.14) express incompressibility conditions in the various cases envisaged, always *if* the heating coefficient for an adiabatic compression can be considered zero.

(G) The case in which heating in adiabatic compressions is not negligible.

If $\left(\frac{\partial\rho}{\partial s}\right)_p \equiv \rho^2\chi_s$, *c.f.r.* (1.2.10), is not zero, *i.e.* if the coefficient of heating under adiabatic compression is not zero, we must add to (1.2.10) the term

$$\left(\frac{\partial\rho}{\partial s}\right)_p \frac{\delta s}{\rho} \equiv \rho\chi_s\delta s \tag{1.2.15}$$

and δs is evaluated, in the Navier–Stokes case, via the third of (1.2.7) yielding the two estimates

$$T\frac{\delta s}{\tau}\simeq\nu\frac{(\delta v)^2}{l^2}\quad\text{and}\quad T\delta s\frac{\delta v}{l}\simeq\nu\frac{(\delta v)^2}{l^2}\tag{1.2.16}$$

And we thus see that incompressibility is justified, in the NS equations case, from the validity of (1.2.12), (1.2.13), (1.2.14) *and in addition*

$$\rho\chi_s\nu\frac{\delta v^2}{l^2}\frac{\tau}{T}\ll 1\quad\text{and}\quad\rho\chi_s\nu\frac{(\delta v)}{lT}\ll 1\tag{1.2.17}$$

Finally, in the case of the (1.2.8) we must add to the last of (1.2.16) the $\rho T\frac{\delta s}{\tau}\approx\kappa\delta T l^{-2}$ e $\rho T\frac{\delta v}{l}\delta s\approx\kappa\frac{\delta T}{l^2}$; and therefore (via (1.2.15)) one finds the further conditions:

$$\kappa\frac{\delta T}{T}\frac{\tau\rho\chi_s}{l^2}\ll 1\quad\text{e}\quad\kappa\frac{\delta T}{T}\frac{\rho\chi_s}{l\delta v}\ll 1\tag{1.2.18}$$

which complete the list of the incompressibility conditions, apart from the problems with boundary conditions, that we have only mentioned in Sect. 1.2.2 and 1.2.3, and which will be studied in some detail in a simple case in Sect. 1.5.

Note to §1.2: dimensional arguments.

One can ask whether the notion of "dimensional argument" can be rendered more precise from a mathematical viewpoint.

It is useful to recall, for this purpose, that analytic functions enjoy a remarkable property: namely if $x\to f(x)$ is a function of the variable x defined for $x\in D$, one says that f is analytic if for every $x_0\in D$ the Taylor series

$$f(x)=\sum_{k=0}^{\infty}f^{(k)}(x_0)(x-x_0)^k/k!\tag{1.2.19}$$

converges for $|x-x_0|$ small enough. Or, equivalently, f is analytic if it is the sum of its own Taylor series around every point.

Then it follows that if f is analytic on D and we suppose that D is the closure of a bounded open set, then it is possible to find a value $\rho>0$ such that the Taylor series of f around *any* $x_0\in D$ has convergence radius at least $\rho>0$. Hence we shall be able to define $f(x)$ for complex values of x: if $|z-x_0|<\rho$ one sets

$$f(z)=\sum_{k=0}^{\infty}f^{(k)}(x_0)(z-x_0)^k/k!\tag{1.2.20}$$

and if the same point z is closer than ρ to two points x_0 and x_0' the two formulae for $f(z)$ obtained by choosing in (1.2.20) once x_0 and another time x_0' must coincide, because the two functions of z so defined must agree for the real z's common to the two intervals of radius ρ and centers x_0,x_0' (hence for infinitely many points and, hence, their identity follows from the identity principle for holomorphic functions).

Thus to say that a function $x\to f(x)$ is analytic on a closed and bounded real domain D is equivalent to saying that it is holomorphic in a complex domain $D_\rho\equiv\{z|\exists x\in D,\ |z-x|\le\rho\}\equiv\{|z-D|<\rho\}$ for some $\rho>0$.

We shall then say that a function f defined on a closed and bounded domain D is "regular on scale ρ" if it is analytic and the convergence radius of its Taylor series around any point is at least ρ, or equivalently if it is holomorphic in D_ρ.

The above notion of regularity is particularly relevant for *dimensional estimates*: indeed if f is *regular on scale* ρ *in* D we shall say that it has a *typical size* $|f|_\rho = \max_{z\in D_\rho} |f(z)|$ and we shall be able to estimate its derivatives as

$$|\partial_x^n f(x)| \le n!|f|_\rho \rho^{-n} \qquad\qquad \forall\, x \in D \tag{1.2.21}$$

i.e. the n-th derivative is simply estimated by dividing the *size of* f by the typical length ρ raised to the n-th power. Just as in the dimensional estimates that are introduced in various arguments in theoretical physics.

Hence the *regularity* on scale ρ and the typical size of a physical quantity that depends on a parameter x have a clear meaning when f is analytic and holomorphically extendible over a distance ρ in the complex and, in the extended domain, it is bounded by a constant $|f|_\rho$ which is identified as the "*typical size*" of f.

In the previous analysis a regular velocity field $\underline{u}(\underline{x}, t)$ must be interpreted as an analytic function in each of the variables x_j and t continuable in the complex, in each variable, by l in the x_j and by τ in the t, remaining bounded therein by δv, and likewise the $s = s(\underline{x}, t)$, $p = p(\underline{x}, t)$ must be analytic and continuable by l and τ, in x_j and t, respectively, staying bounded by δs and δp.

Thus we see that "accepting dimensional estimates" corresponds mathematically to admitting precise regularity properties on the functions under investigation.

Whenever such properties do not hold it becomes necessary to reexamine the dimensional argument: and sometimes it can turn out to be grossly incorrect. This happens when in the problem appear "several scales" very different from each other.

For instance sometimes the function $f(x)$ can be written as a sum of infinitely many functions $f_1(x) + f_2(x) + \ldots$ with f_i regular on scale ρ_i and of order of magnitude δ_i and, furthermore, $\rho_i \underset{i\to\infty}{\longrightarrow} 0$. It is clear that in such cases one shall be very cautious in formulating dimensional arguments. For an explicit example consider a sequence $f_i(x) \equiv c_i f(x/\rho_i)$ with $\rho_i = 2^{-i}$, $c_i = 2^{-i^2}$ or $c_i = 2^{-ki}$ with k integer and fix $f(x)$ to be a rapidly decreasing function (*e.g.* if $D = [0, +\infty)$ we can take $f(x) = e^{-x}$).

Finally we mention that (1.2.21) is a simple consequence of Cauchy's theorem

$$f^{(n)}(x) = \frac{n!}{(2\pi i)} \oint \frac{f(z)}{(z-x)^{n+1}}\, dz \tag{1.2.22}$$

where the contour can be chosen as a circle around x contained in D_ρ: by choosing exactly the radius of the cirdle to be ρ and bounding above the right hand side by the absolute values one immediately gets (1.2.21).

The problems in which there are many length or time scales are called *multiscale problems*: dimensional arguments are in such cases called *scaling arguments*. In recent times new methods for their analysis have been developed, like the "*renormalization group method*". But since a long time they attract the interests of physicists and mathematicians and many beautiful phenomena in mathematics and physics have been understood, I just mention here the almost everywhere convergence of Fourier series for square integrable functions in mathematics and the ultraviolet stability of some quantum fields in three space time dimensions in relativistic physics and, as we shall see in the last sections of this book, in the developed turbulence theory in fluid mechanics, see [Ca66],[Fe72],[BG95].

Problems: *Stokes formula.*

[1.2.1]: Consider a viscous fluid occupying the entire space outside a sphere of radius R and moving with a velocity $\underline{v}_0$ at ∞. Suppose the motion time independent and the velocity so small that one can neglect the transport term $\underset{\sim}{u} \cdot \underset{\sim}{\partial} \underline{u}$ in the NS equation $\underline{0} = -\rho^{-1}\underline{\partial} p + \eta\rho^{-1}\Delta\underline{u}$ and $\underline{\partial} \cdot \underline{u} = 0$. This is the *Stokes equation* which can be written

$$\Delta \operatorname{rot} \underline{u} = \underline{0}, \qquad \underline{\partial} \cdot \underline{u} = 0, \qquad \underline{u} = \underline{0} \quad \text{if } |\underline{x}| = R$$

Show that there is at most one smooth solution $\underline{u}$ tending to $\underline{v}_0$ as $r \to \infty$ and such that $r^2|\underline{\partial}\underline{u}|$ is bounded as $r \to \infty$.

The solution, if existent, must have the form

$$\underline{u} = \underline{v}_0 + f_1(r)\underline{v}_0 + f_2(r)\underline{n} \cdot \underline{v}_0\, \underline{n} + f_3(r)\underline{n} \wedge \underline{v}_0$$

with $r \equiv |\underline{x}|, \underline{n} \equiv \underline{x}/r$ and $f_j \xrightarrow[r\to\infty]{} 0$. Furthermore $f_3 \equiv 0$ by parity symmetry. (*Idea:* Uniqueness follows because the difference $\underline{\delta}$ between two solutions must be such that $\Delta\underline{\delta} = \underline{\partial}\pi$ for some π; hence multiplying both sides by $\underline{\delta}$ and integrating by parts one gets that $\underset{\sim}{\partial}\underline{\delta} = 0$. The equation is linear and the only vectors linearly depending on $\underline{v}_0$ which can be made with $\underline{v}_0$ and $\underline{x}$ are $\underline{v}_0, \underline{x}, \underline{x} \wedge \underline{v}_0$. Uniqueness implies that the solution must be parity invariant).

[1.2.2]: Chose $\underline{v}_0$ along the x–axis and check that if $\underline{u}$ has the form in [1.2.1] then

$$\underline{\partial} \cdot \underline{u} = v_0\, x\, (\frac{f_1'}{r} + 4(\frac{f_2}{r^2}) + r(\frac{f_2}{r^2})')$$
$$\operatorname{rot} \underline{u} = v_0\, \Big(f_1' - \frac{1}{r} f_2\Big)\, (0, \frac{z}{r}, -\frac{y}{r})$$

Check also that the choices $f_1 = r^{-1}, f_2 = r^{-1}$ and $f_1 = -r^{-3}, f_2 = 3r^{-3}$ generate fields $\underline{u}$

$$\frac{1}{r}\Big(\underline{v}_0 + \underline{n}\,\underline{v}_0 \cdot \underline{n}\Big), \text{ or } \frac{1}{r^3}\Big(- \underline{v}_0 + 3\underline{n}\,\underline{v}_0 \cdot \underline{n}\Big)$$

which have 0 divergence and a harmonic rotation (*i.e.* a rotation with 0 laplacian).

[1.2.3]: Hence one can look for a solution like

$$\underline{u} = \underline{v}_0 - \frac{a}{r}\Big(\underline{v}_0 + \underline{n}\,\underline{v}_0 \cdot \underline{n}\Big) + \frac{b}{r^3}\Big(- \underline{v}_0 + 3\underline{n}\,\underline{v}_0 \cdot \underline{n}\Big)$$

Show that the coefficients a and b are uniquely determined by the condition $\underline{u} = \underline{0}$ for $|\underline{x}| = R$ and have the value

$$a = \frac{3}{4}R, \qquad b = \frac{1}{4}R^3$$

(*Idea:* Note, for instance by [1.2.2], that the combinations in front of a and b have 0 divergence and rotation with 0 laplacian).

[1.2.4]: Compute the pressure field associated with the velocity field determined in [1.2.2] showing that $p(\underline{x}) = -3\,\eta\, R\, \underline{v}_0 \cdot \underline{n}/2r^2$. (*Idea:* $-\underline{\partial} p + \eta\Delta\underline{u} = \underline{0}$.)

[1.2.5]: The force exerted on the sphere has, if we choose the z axis parallel to the force, a z–component

$$F = R^2 \int \Big(- p\cos\vartheta + (\tau'_{zz}\cos\vartheta + \tau'_{zx}\sin\vartheta\cos\varphi + \tau'_{zy}\sin\vartheta\sin\varphi)\Big) \sin\vartheta\, d\vartheta d\varphi$$

where $\tau'_{zj} = \eta(\partial_j u_x + \partial_z u_j)$, and show that this implies that $F = \eta R v_0 S$ where S is a constant. Compute S ($S = 6\pi$, *Stokes formula*).

[1.2.6]: (*meaning of approximations*) Discuss under which assumptions the approximation in [1.2.1] can be acceptable. Show that the conditions imposed are realizable around the sphere because they are: $v_0^2/R \ll \nu v_0/R^2$, *i.e.* $v_0\nu^{-1}R \ll 1$ which is read, see the coming sections, by saying that the "Reynolds' number" is small. However such conditions are *not valid* far away from the sphere because there they become $v_0^2\, R/r^2 \ll \nu v_0\, R/r^3$, *i.e.* $v_0\nu^{-1}r \ll 1$. Hence at large distances the velocity field determined via the Stokes' approximations [1.2.1], [1.2.3], cannot be taken as correct. (*Idea:* Compare the size of the transport term $\underset{\sim}{u} \cdot \underset{\sim}{\partial}\, \underline{u}$ to the size of the viscous term $\nu\, \Delta \underline{u}$ by using the solution in [1.2.3].)

Bibliography: The discussion reported in (D,E,F) follows the ideas in [LL71].

§1.3 The rescaling method and estimates of the approximations.

The procedure illustrated in §1.2 to evaluate the orders of magnitude involved in the incompressibility approximations is simple but, in a way, not very systematic.

In fact the claim that (adimensional) quantities $\varepsilon \ll 1$ can be neglected is a satisfactory statement *only* if one is able to evaluate the error made and to show that corrections really have size ε with respect to the terms that are not neglected, as implicitly supposed in the analysis.

This can only be an asymptotic statement and what one really means, or what one should mean, is that it is possible to write the solution of the equations, that we want to approximate, as a series in the parameter ε. But from the discussion we see that ε appears in various forms and it is by no means clear what it does really mean that "we neglect terms of the order ε", in particular, when ε appears both as an order of magnitude of certain quantities and as an argument of relevant functions (as it happens when we say that a function varies on the scale l and $l/\tau v_{sound} = \varepsilon$ is small (*c.f.r.* (1.2.12)).

To make more precise the above intuitive idea we shall translate into a more mathematical form some of the arguments discussed in §1.2, trying to construct, at least in principle, an algorithm that allows us to write the equations necessary for the evaluation of the error as a series in a parameter (on in several parameters) $\ll 1$.

(1) Incompressible Euler equation.

For instance consider the case (A), §1.2, of the incompressible Euler equation, with $\underline{g} = \underline{0}$ for simplicity. Assume that the system is a *perfect gas* with constitutive equations.

$$s = c_V \,\log T - c\,\log\rho, \qquad \tau'_{ij} = 0, \qquad \kappa_{ij} = 0 \tag{1.3.1}$$

which, via the thermodynamic relation $p = -T\rho^2\,(\frac{\partial s}{d\rho})_T$, implies $p = c\,\rho T$.

Furthermore let $v^2_{sound} = (\frac{\partial p}{\partial\rho})_s = cT\,(1 + c/c_V)$ be the velocity sound.[1]

Let $\underline{\bar u}, \bar\rho, \bar s$ be an initial datum with the property of satisfying the first of the (1.2.12). This state can be assigned in terms of three functions $\underline{\bar w}(\underline{\xi}), \bar r(\underline{\xi}), \bar\sigma(\underline{\xi})$ very regular in their arguments $\underline{\xi} \in R^3$ as

$$\underline{\bar u}(\underline{x}) = \varepsilon\, v_{sound}\, \underline{\bar w}(\frac{\underline{x}}{l}), \quad \bar\rho(\underline{x}) = \bar r(\frac{\underline{x}}{l}), \quad \bar s(\underline{x}) = \bar\sigma(\frac{\underline{x}}{l}) \tag{1.3.2}$$

where ε is a very small parameter, so that the initial data in (1.3.2) satisfy *a priori* the condition that the initial velocity $\underline{\bar u}$ be small compared to the

[1] In the case of a perfect monoatomic gas $c_V = \frac{3}{2}RM_0^{-1}, c = RM_0^{-1}$ with R the gases constant, and M_0 the atomic mass.

sound speed; and they vary on a length scale l, which is a parameter with dimension of a length. The velocity v_{sound} depends on T (which depends on $\underline{x}$) and here we define it as equal to the value corresponding to the average value of T (computed from the equation of state in the initial configuration). We shall imagine the system in infinite space and that the functions in (1.3.2) are constants outside a bounded set and that the initial $\underline{\bar{w}}$ vanishes outside this set.

We now ask if there exists a solution to (1.2.1), (1.3.1) satisfying (1.2.12) also at positive times, and if this solution is *well approximated* by the solution of (1.2.4) with the same initial data, and *the better the smaller* ε is.

We shall limit ourselves to the analysis of the case $\bar{r} = constant, \bar{\sigma} = constant$, even though it will be instructive to write a few more general equations.

To pose correctly the question we ask whether (1.2.1), (1.3.1) with the initial datum (1.3.2), has a solution depending *regularly* on t through the "rescaled time" $\vartheta = \varepsilon t l^{-1} v_{sound}$: so that the second'of the (1.2.12) is automatically satisfied (because the time scale τ will be such that $\varepsilon\,\tau\, v_{sound} l^{-1} \simeq 1$ and, hence, $l/(\tau\, v_{sound}) \simeq \varepsilon \ll 1$).

More formally we ask the question whether a solution of (1.2.1) with equation of state (1.3.1) exists such that

$$\begin{aligned} \underline{u}(\underline{x},t) &= \varepsilon\, v_{sound}\, \underline{w}(\underline{x}l^{-1}, \varepsilon\, t\, v_{sound} l^{-1}) \\ \rho(\underline{x},t) &= r(\underline{x}l^{-1}, \varepsilon\, t\, v_{sound} l^{-1}) \\ s(\underline{x},t) &= \sigma(\underline{x}l^{-1}, \varepsilon\, t\, v_{sound} l^{-1}) \end{aligned} \tag{1.3.3}$$

with $\underline{w}(\underline{\xi},\vartheta), r(\underline{\xi},\vartheta), \sigma(\underline{\xi},\vartheta)$ *regular functions of their arguments* and depending on ε so that they can be written as

$$\begin{aligned} \underline{w} &= \underline{w}_0 + \varepsilon \underline{w}_1 + \varepsilon^2 \underline{w}_2 + \ldots \\ r &= r_0 + \varepsilon r_1 + \varepsilon^2 r_2 + \ldots \\ s &= \sigma_0 + \varepsilon \sigma_1 + \varepsilon^2 \sigma_2 + \ldots \end{aligned} \tag{1.3.4}$$

The regularity of $\underline{w}, r, \sigma$ implies that in the case in (1.3.3) the conditions (1.2.12) will continue to be satisfied at positive times and therefore *we expect*, if the discussion in §1.2 is correct, that the (1.3.3) verify the Euler incompressible equation, at a first approximation.

The latter property has now a precise mathematical meaning. In fact inserting the (1.3.4),(1.3.3) in (1.2.1), (1.3.1) and imposing that equations (1.2.1),(1.3.1) are verified at all orders in ε, we obtain equations for the $\underline{w}_j, r_j, \sigma_j$ which, solved with the natural initial data

$$\begin{aligned} &\underline{w}_0\big|_{\vartheta=0} = \underline{\bar{w}}, \quad r_0\big|_{\vartheta=0} = \bar{r} = constant, \quad \sigma_0\big|_{\vartheta=0} = \bar{\sigma} \\ &w_j, r_j, \sigma_j\big|_{\vartheta=0} = 0 \end{aligned} \tag{1.3.5}$$

privide us with "solution to lowest order in ε", given by (1.3.3) with $\underline{w}, r, \sigma$ replaced by $\underline{w}_0, r_0, \sigma_0$ and the higher order corrections.

Then the question that we asked above is whether the functions

$$\varepsilon v_{sound}\, \underline{w}_0(\underline{x}l^{-1}, \varepsilon t v_{sound} l^{-1}), \quad r_0 = constant, \quad \sigma_0(\underline{x}l^{-1}, \varepsilon\, t v_{sound} l^{-1}) \tag{1.3.6}$$

verify incompressible Euler equations (1.2.4). Or

$$\begin{aligned} &r_0 = constant, \qquad \underline{\partial}_{\underline{\xi}} \cdot \underline{w}_0 = 0 \\ &\partial_\vartheta \underline{w}_0 + \underset{\sim}{w}_0 \cdot \underset{\sim}{\partial}_{\underline{\xi}} \underline{w}_0 = -\underline{\partial}_{\underline{\xi}} p', \qquad \partial_\vartheta \sigma_0 + \underset{\sim}{w}_0 \cdot \underset{\sim}{\partial}_{\underline{\xi}} \sigma_0 = 0 \end{aligned} \tag{1.3.7}$$

if $p'(\underline{\xi}, \vartheta)$ is a suitable function.

The equations for the successive orders should determine recursively $\underline{w}_j, r_j, \sigma_j$ and therefore all the corrections, *systematically*.

We would verify in this way, in a precise sense, that the slow velocity motions of the perfect gas under analysis is well approximated by the incompressible Euler equations. And, if we could devise an algorithm to compute the corrections $w_j, r_j, \sigma_j,\ j \ge 1$, it would make sense also to estimate the errors of the approximation

$$\begin{aligned} \underline{u}(\underline{x}, t) =& \varepsilon v_{sound}\, \underline{w}_0(\frac{\underline{x}}{l}, \varepsilon \frac{v_{sound} t}{l}), \quad \rho(\underline{x}, t) = r_0(\frac{\underline{x}}{l}, \varepsilon \frac{v_{sound} t}{l}), \\ s(\underline{x}, t) =& \sigma_0(\frac{\underline{x}}{l}, \varepsilon \frac{v_{sound} t}{l}) \end{aligned} \tag{1.3.8}$$

of the solutions of (1.2.1),(1.3.1) via the solutions of the incompressible Euler equation (1.2.4), (*i.e.* (1.3.6) and (1.3.7)).

It is useful to underline, again, that in our situation, the second of (1.2.12) follows from the first because from (1.3.8) we see that the scale of time evolution is $l v_{sound}^{-1} \varepsilon^{-1}$ (hence the second equation of (1.2.12) becomes $\varepsilon^2 \ll 1$ which coincides with the first); and if the property of approximation of (1.3.3) via the (1.3.4),(1.3.5) holds then the validity of (1.2.12) at the initial time (guaranteed for $\varepsilon \ll 1$ from (1.3.2)) implies its validity at the successive instants, at least up to the time $t = \tau_0 l \varepsilon^{-1} v_{sound}^{-1}$ if τ_0 is the instant until which the Euler equation (1.3.6), in the adimensional variables $\underline{\xi}, \vartheta$, with initial data (1.3.5) admits a solution that stays regular in $\underline{\xi}, \vartheta$.

Therefore we shall proceed to checking that the assumption that (1.2.1) admit a solution that can be developed in powers of ε is a *consistent* assumption and that it really leads to (1.3.7) at the lowest order in ε at least. Once we shall have succeeded, at least formally, we shall have obtained a precise qualitative check of the incompressibility assumptions. A quantitative check will require, then, in principle also an analysis of the series (1.3.4) or at least the analysis of the terms neglected and the possible proof that they tend to 0 for $\varepsilon \to 0$ in a way that can be estimated explicitly.

Note that the assumption $\bar r$ constant and $\bar\sigma=$ constant for $\vartheta=0$ (*i.e.* the case to which we confine our attention) implies that s is constant for all times because the right hand side of the third of the (1.2.1) vanishes. Then it follows that the pressure is a function $p=p(\rho)$ of the density (and $p(\rho)$ is the "adiabatics equation": $p(\rho)=C\rho^{1+c/c_V}$, with C suitable and independent of $\underline x,t$, determined from the initial conditions).

The (1.3.3),(1.3.4),(1.3.5), can be inserted into (1.2.1); taking into account the assumptions on the constitutive equations made when considering (1.3.1) and supposing $\underline g=\underline 0$, one finds (writing only the lowest orders in ε)

$$\begin{aligned}
&\frac{\varepsilon v_{sound}}{l}\big(\partial_\vartheta r_0+\underline\partial_{\underline\xi}\cdot(r_0\underline w_0)\big)=0 \qquad (1.3.9)\\
&r_0\frac{\varepsilon^2 v^2_{sound}}{l}\big(\partial_\vartheta\underline w_0+\underset{\sim}{w}_0\cdot\underset{\sim}{\underline\partial}_{\underline\xi}\underline w_0\big)+\ldots=-\frac{v^2(r)}{l}\big(\underline\partial r_0+\varepsilon\underline\partial r_1+\varepsilon^2\underline\partial r_2\ldots\big)\\
&\frac{\varepsilon v_{sound}}{l}\big(\partial_\vartheta\sigma_0+\underset{\sim}{w}_0\cdot\underset{\sim}{\underline\partial}_{\underline\xi}\sigma_0\big)=0
\end{aligned}$$

where $v^2(r)=\frac{\partial p(\rho)}{\partial\rho}\big|_{\rho=r}$ is essentially still the square of the sound speed (at density r so that we denote it differently from the average quantity v^2_{sound}).

We realize that, in order that the second equations be consistent, it must be $\underline\partial r_0=\underline 0$, *i.e.* the assumptions are consistent only if r_0 is *constant* as a function of $\underline\xi$. And the first of the (1.3.9) will say that $\underline\partial\cdot\underline w_0$ is constant in $\underline\xi$ (being equal to $\partial_\vartheta r_0$ with r_0 constant in $\underline\xi$) and, hence, vanishing if we suppose that $\underline w_0$ tends to zero at infinity for all times ϑ; likewise r_1 must be constant in $\underline\xi$. Hence also $\partial_\vartheta r_0=0$ and $r_0=\bar r$ stays constant and $\underline\partial\cdot\underline w_0=0$. In such case the (1.3.9) become the dimensionless equations:

$$r_0=\bar r,\qquad \underline\partial\cdot\underline w_0=0,\qquad \sigma_0=\bar\sigma,\qquad \partial_\vartheta\underline w_0+\underset{\sim}{w}_0\cdot\underset{\sim}{\underline\partial}_{\underline\xi}\,\underline w_0=-\frac{1}{r_0}\underline\partial r_2 \qquad (1.3.10)$$

and, by the equation of state, $T=$ *constant*.

Thus we have obtained in the rescaled variables, (1.3.7), and in the adiabatic case the incompressible Euler equations.

We see another interesting property: namely what we call "pressure" in the incompressible Euler equations really is, up to a constant, the deviation from the average density to second order in ε.

In principle we should derive (infinitely many) other differential equations which should allow us to evaluate the corrections at the various orders in ε. But such equations would certainly be involved (if possible at all) and of little interest since we shall not have a grip on a theory for them (because we are unable, to this date, to really build a satisfactory theory for the lowest order, *i.e.* for the incompressible Euler equations, as we shall realize in the coming sections). Hence there is a serious risk that what said so far will remain for a long time at a formal level.

The above remarks help understanding the importance of the following theorem that considers the (1.2.1), (1.3.1) with initial data having the form

(1.3.5) with $\underline{\overline{w}} \in C^\infty$, and with $\underline{\overline{w}}_0$ vanishing outside a bounded set. And it allows us to say that the solution of the (1.2.1) tends, as $\varepsilon \to 0$, to the solution of the incompressible Euler equation in the following sense. We consider the solution of the equations (1.3.10), $w_0(\vartheta, \underline{x}), \sigma_0(\vartheta, \underline{x}), r_0$, then the following theorem holds, [Eb77], [EM94]:

1 Theorem *(incompressible Euler limit): The Euler equation (1.3.10) with the initial data (at $\vartheta = 0$) $\underline{w}_0(\underline{x}, 0) = \underline{\overline{w}}(\underline{x})$, $r_0 = \overline{r}$, $\sigma_0 = \overline{\sigma}$ admits a solution of class C^∞, that rapidly vanishes at infinity together with its derivatives, for times $\vartheta \le \tau_0$, if τ_0 is small enough (but depending on the initial data).*
The existence time τ_0, dimensionless by definition, can be chosen so that, for the times $t < \tau_0\, l/\varepsilon v_{sound}$, also Eq. (1.2.1) and Eq. (1.3.1) with initial data of the form (1.3.3) with ε a positive parameter, admit a solution of class C^∞, $\underline{u}_\varepsilon(\underline{x}, t), \rho_\varepsilon(\underline{x}, t), s_\varepsilon(\underline{x}, t) \equiv \bar{\sigma}$. And one has

$$\begin{aligned} &\varepsilon^{-1} |\underline{u}_\varepsilon(\underline{x}, t) - \varepsilon v_{sound}\, \underline{w}_0(\underline{x} l^{-1}, \varepsilon\, v_{sound}\, t\, l^{-1})| \xrightarrow[\varepsilon \to 0]{} 0 \\ &|\rho_\varepsilon(\underline{x}, t) - \bar{r}| \xrightarrow[\varepsilon \to 0]{} 0 \end{aligned} \tag{1.3.11}$$

uniformly for $t < \tau_0\, l/\varepsilon v_{sound}$.

Remarks:

(1) Note that the theorem is formulated without involving at all the higher order terms of the series (1.3.9). Hence, independently of their existence, it is rigorously established that, at least for a small time t, $t < \tau_0 l/\varepsilon\, v_{sound}$ (but of the order of ε^{-1}, *i.e.* "*the smaller ε is the better the incompressibility property is satisfied*"), the incompressible Euler equation provides an effective approximation to the solution of the (1.2.1),(1.3.1).
(2) The role of the previous statement (1) is to insure that the theorem be not empty it is obviously necessary to show that the incompressible Euler equation, with the initial data considered in the theorem, admits a solution up to a time $\tau_0 > 0$, which maintains the necessary properties of regularity. Such a theorem is indeed possible and it will be discussed in §3.1.
(3) It would be interesting to show that the time τ_0 of the theorem is of the order of the maximum time for which the incompressible Euler equation admits a regular solution (with the initial datum considered). This would be particularly interesting in the case of a fluid in a space with dimension $d = 2$: in this case, as we shall see, the incompressible Euler equation admits a global solution (*i.e.* a solution for all times) without losing regularity (*i.e.* data initially of class C^∞ remain such). Unfortunately the proof of the theorem *does not* allow us to conclude this much and the time τ_0 is an estimate that turns out to be shorter than the maximum time for which one can show, see §3.1, existence of regular solutions for the Euler equations.

(2) The incompressible Navier–Stokes equation.

In this case one must add in the right hand side of the second of (1.3.9) the term

$$\varepsilon\frac{\rho\nu v_{sound}}{l^2}\Delta_{\underline{\xi}}\,\underline{w}_0 \tag{1.3.12}$$

plus the corresponding higher orders. In §1.2, we saw that, to derive the conditions of validity of the approximations, the (1.2.13) had to be added to the (1.2.12) if also $\underline{g}=\underline{0}$ is assumed, for simplicity. Proceeding exactly in parallel to the preceding case of the Euler equations, and using the notations of (1.3.9), we see that incompressibility with initial data $\bar r$ and $\bar s$ constant (see (1.3.2)) is consistent if

$$\frac{\nu}{\varepsilon l v_{sound}}\equiv\nu_0\ll 1\qquad\text{independently of }\varepsilon \tag{1.3.13}$$

which is again (1.2.13). This means that (1.2.13) now demands that the length scale over which the fields change be of the order of magnitude of $l=\nu\varepsilon^{-1}v_{sound}^{-1}\nu_0^{-1}$. As in the Euler case one can prove the following theorem. Consider the adiabatic Navier–Stokes equations:

$$\begin{aligned}&r_0=constant,\qquad s_0=constant\\&\partial_\vartheta\underline{w}_0+\underset{\sim}{w}_0\cdot\underset{\sim}{\partial}_{\underline{\xi}}\,\underline{w}_0=-\frac{1}{r_0}\underline{\partial}p_0+\nu_0\Delta\underline{w}_0\end{aligned} \tag{1.3.14}$$

with ν_0 a positive constant, initial data $\underline{w}_0\in C^\infty$ and vanishing outside a bounded region, $r_0=\bar r=constant,\sigma_0=\bar\sigma=constant$; let $\underline{w}_0(\vartheta,\underline{\xi})$ be a solution of class C^∞ (in $\vartheta,\underline{\xi}$). Then, [KM81], [EM94]:

2 Theorem *(incompressibility; the NS case):*
(i) The Navier–Stokes equation (1.3.14) admits a C^∞–solution for times $\vartheta<\tau_0$ for some $\tau_0>0$,
(2) Let $\varepsilon>0$ be a positive parameter. Assume that the constitutive equations are $\kappa_{ij}=0$, (perfect non heat conducting gas) and $\tau'_{ij}=\rho\,\nu(\partial_j u_i+\partial_i u_j)$ with ν verifying (1.3.13) (NS stress). Given $l_0>0$ the time $\tau_0>0$ can be chosen so that (1.2.1) with equation of state of a perfect gas (see (1.3.1)) and initial data

$$\underline{u}(\underline{x})=\varepsilon v_{sound}\underline{\bar w}_0(\frac{\underline{x}}{l}),\quad\rho(\underline{x})=\bar r,\quad s(\underline{x})=\bar\sigma,\qquad\text{with } l=\frac{l_0}{\varepsilon} \tag{1.3.15}$$

admits a C^∞–solution, which we shall denote $\underline{u}_\varepsilon(\underline{x},t),\rho_\varepsilon(\underline{x},t)$, and $s_\varepsilon(\underline{x},t)$, defined for times $t<\tau_0\,l_0/\varepsilon^2 v_{sound}$. Furthermore

$$\begin{aligned}&\varepsilon^{-1}|\underline{u}_\varepsilon(\underline{x},t)-\varepsilon\,v_{sound}\,\underline{w}_0(\varepsilon\,\underline{x}l_0^{-1},\varepsilon^2\,v_{sound}\,tl_0^{-1})|\xrightarrow[\varepsilon\to0]{}0\\&|\rho_\varepsilon(\underline{x},t)-\bar r|\xrightarrow[\varepsilon\to0]{}0,\qquad|s_\varepsilon(\underline{x},t)-\bar\sigma|\xrightarrow[\varepsilon\to0]{}0\end{aligned} \tag{1.3.16}$$

Remarks:

(1) One can make comments identical to the ones that follow theorem 1 above. One sees that in the limit $\varepsilon \to 0$ entropy is conserved: which is no longer obvious since the right hand side of the third of the (1.2.1) no longer vanishes. Nevertheless friction influences the equation for the velocity. This is, at first, looks strange but it is understood if one takes into account that $\underline{w}_0$ is a rescaled velocity and a variation of $O(1)$ due to friction (*i.e.* due to the term $\nu_0\Delta\underline{w}_0$ in (1.3.14)) generates a variation of energy of the order $O(\varepsilon^2)$ and, therefore, a quantity of heat and an increase of entropy (and temperature) $O(\varepsilon^2)$ which is not contradictory to the third of the (1.3.16).

(2) Interpreting theorems 1,2 above one can say: *on time and space scales $O(\varepsilon^{-1})$ the system follows the incompressible Euler equation*; while *on time scales $O(\varepsilon^{-2})$ and space scales $O(\varepsilon^{-1})$ the system follows the Navier–Stokes equations.*

(3) What can one then say if the initial datum is given without any free parameter ε? *i.e.* can the just stated theorems be concretely applicable as approximation theorems when ε is fixed? The risk being that they are just conceptual theorems illustrating the asymptotic nature of the incompressibility assumption.

(4) A proposal is the following. Wishing to apply such theorems in a given particular case one should check that the initial datum can be written in the form (1.3.4). Then if ε is small one shall be able to say that the incompressible Euler equation holds (for times up to $O(\varepsilon^{-1})$), and one should also be able to give the approximation error by using the estimates of the differences in (1.3.11),(1.3.16): indeed such estimates are constructive, *i.e.* computable, in the proofs (not described here) of the theorem. As the time increases, beyond $O(\varepsilon^{-1})$, we expect that the velocity field becomes more uniform in space and that it will, after a time very long with respect to $O(\varepsilon^{-1})$, be described by a regular function of $\varepsilon^{-1}\underline{x}l_0^{-1}$ for some l_0 which should depend on the initial data.

(5) In this situation we shall be in the assumptions of theorem 2 and the fluid will now evolve following the incompressible Navier–Stokes equation, with an approximation controlled up to times of order $O(\varepsilon^{-2})$, and it will proceed towards equilibrium (which is simply the state in which the velocity field vanishes, because we are supposing that there are no external forces) keeping a variability on scales of length of order $\varepsilon^{-1}l_0$ and of time of order $\varepsilon^{-2}\,\nu\,l_0^{-1}\,v_{sound}$.

(6) The above is a scheme of interpretation of an incompressible evolution: however it is just a "proposal" because there are *no other known theorems* that support such proposal and it is not so clear, in the above proposal, the cross–over between the two regimes can be described and how. From what said above not only l_0 is *not* calculated but there is no hint nor any idea on how to calculate it, nor there is any idea on which physical properties a calculation of l_0 could be based.

Bibliography: This section is based on the ideas and results of the paper [EM94]; the original theorems 1, 2 are in [Eb77],[Eb82], [KM81], [KM82]. I have preferred the approach in [EM94] because it is closer in spirit to the analysis by Landau and Lifshitz reported in (D,E,F) of §1.2.

§1.4 Elements of hydrostatics.

Hydrostatics deals with solutions of the Euler, Navier–Stokes or more general continua, with vanishing velocity fields and with time independent thermodynamic functions.

These solutions are very rare, as we shall see by considering a few model cases.

(1) *Hydrostatics in absence of thermoconduction.*

(A) Isoentropic case

Equations (1.2.1) become simply

$$-\frac{1}{\rho}\underline{\partial}\, p + \underline{g} = \underline{0}, \qquad \varepsilon = \varepsilon(\rho, s) \tag{1.4.1}$$

As an example we shall treat the case of a perfect monoatomic gas

$$\varepsilon = \varepsilon_0 \left(\frac{\rho}{\rho_0}\right)^{2/3} e^{(s-s_0)/c_v} \Rightarrow \varepsilon = c_v T, \quad p = \frac{2}{3}\rho c_v T = \rho \frac{RT}{M_A} = \frac{nRT}{v} \tag{1.4.2}$$

where ε_0, ρ_0, s_0 are values of ε, ρ, s in a reference thermodynamic state; $n = M/M_A$ with M the total fluid mass and M_A is the atomic mass (n is called "molar number"); $v = M/\rho$ is the specific volume of the fluid; R is the gas constant $R = 8.31\ 10^7$ erg $^\circ$K^{-1}; c_v is the specific heat at constant volume (per unit mass), *i.e.* $c_v = \frac{3}{2}R/M_A$. If the gas was diatomic the factor 3/2 would become, everywhere, 5/2.

Suppose that the force density $\underline{g}$ is conservative, $\underline{g} = -\underline{\partial} G$. In the isoentropic case the relation between ρ and p is simply the adiabatic equation of state $\rho = R(p)$: $\rho = p^{1/\gamma}\, const$ with $\gamma = 5/3$ for a monoatomic perfect gas. Therefore it is convenient to define the "*pressure potential*"

$$\Phi(p) = \int^p \frac{dp'}{R(p')} \tag{1.4.3}$$

so that (1.4.1) is solved by

$$\Phi(p(\underline{x})) + G(\underline{x}) = \text{constant}, \qquad \rho = R(p(\underline{x})) \tag{1.4.4}$$

that permits us to determine $p(\underline{x})$ in terms of G and consequently to determine $\rho(\underline{x}), \varepsilon(\underline{x})$ *etc* (note that $\Phi(p)$ is strictly increasing in p and, hence, invertible). This holds at least if the values of $G(\underline{x})$ are among those of Φ; *i.e.* they are in $\Phi([0,+\infty))$, up to an additive constant.[1]

One should also remark that if, in the isoentropic case, $\underline{g}$ was not conservative no hydrostatic solution could exist: a non conservative force will necessarily set the fluid in motion. This is physically obvious and (as we shall see) *remains essentially true also in the case of non isoentropic fluids.*

(B) Non isoentropic case.

The non isoentropic hydrostatics is analogous. Now $s = s_0(\underline{x})$ hence $\rho = r(p, s_0)$ so that the equation in the cases when $\underline{g}$ is conservative:

$$-\underline{\partial}\, p = r(p, s_0(\underline{x}))\, \underline{\partial} G \tag{1.4.5}$$

implies that $r(p, s_0(\underline{x}))$ must be a function of the form $R(G(\underline{x}))$, hence $p(\underline{x})$ must also be a function of the form $\pi(G(\underline{x}))$: then $s_0(\underline{x})$ must have the form $\Sigma(G(\underline{x}))$. The interesting consequence is, therefore, that in this case *hydrostatic solutions in which entropy cannot be expressed as a function of the potential of the volume forces is not possible.*

For general $\underline{g}$ one finds that *hydrostatic solutions can exist only if the volume force is proportional to a conservative force.* See problems.

We shall discuss here the latter question in a specific case in which there are no problems on the possible existence of solutions. Since (1.4.1) is very restrictive there are not many such cases and only the particularly symmetric ones are easy to treat.

Consider for instance a fluid occupying the half space $z > 0$ and subject to a gravity force with potential $G = gz$, and look for *stratified hydrostatic solutions*, *i.e.* solutions in which the thermodynamic functions depend only on z. We shall denote them $s = s_0(z)$, $T = T(z)$, $p = p(z)$, $\varepsilon = \varepsilon(z)$ and $\rho = \rho(z)$, $(\rho = r(p, s))$. Then (1.4.5) simplifies and we find

$$-\frac{1}{\rho(z)}\frac{dp}{dz} = \frac{dG}{dz}, \qquad \rho(z) = r(p(z), s_0(z)) \tag{1.4.6}$$

which is an ordinary differential equation for $p(z)$ determining it once the data $p(0) = p_0$ and the function $s = s_0(z)$ are known.

More specifically consider a perfect gas in a gravity field; the (1.4.6) become, if one imagines that $T = T_0(z)$ is *a priori* assigned (instead of the entropy)

$$-\frac{1}{\rho(z)}\frac{dp}{dz} = g, \quad T = T(z), \quad p = \frac{2}{3}\rho c_v T \tag{1.4.7}$$

so that if we take $T = (1 + \gamma z)\, T_0$, $\gamma \ge 0$ we find

$$\frac{2c_vT}{3p}\frac{dp}{dz} = -g \;\Rightarrow\; \frac{dp}{p} = -\frac{3}{2}\frac{1}{c_vT_0}\frac{gdz}{1+\gamma z} \tag{1.4.8}$$

[1] Or, in other words if $G(\underline{x})$ is bounded below, otherwise the equation does not admit hydrostatic solutions.

whose solution is

$$p = (1 + \gamma z)^{-3g/(2c_v T_0 \gamma)}\, p_0 \tag{1.4.9}$$

In the isothermal case, $\gamma = 0$, this becomes the well known

$$p = p_0 e^{-3gz/(2c_v T_0)} \tag{1.4.10}$$

while in the incompressible case (1.4.7) has the equally well known solution

$$p = p_0 - \rho g z \tag{1.4.11}$$

Hence it is possible that a gas in which temperature is not constant stays in a "stratified equilibrium".

(C) Stability of equilibria

Temperature and density gradients can generate instabilities of the equilibrium of a fluid because it could become energetically convenient to displace a volume element of the fluid by exchanging its position with another and by taking advantage of the external field or of the density differences due to temperature differences.

We consider the two following questions about the stratified equilibria in (B) above: (1) under which conditions are they stable, (2) under which conditions it is possible to suppose $\rho =$ constant and therefore use (1.4.11).

The result will be a remarkable stability criterion about the development of convective motions; *in the case of an adiabatic perfect gas*, *i.e.* if heat conduction is negligible, and in a gravity field g there will be stability in a gravity field g if

$$\frac{\partial T}{\partial z} \ge -\frac{g}{c_p} \tag{1.4.12}$$

This means that temperature can decrease with height, but not too much. If the variation ΔT between two horizontal planes at distance h is such that $\Delta T > gh/c_p$, and if the higher plane is colder, convection phenomena start "spontaneously", *i.e.* they are generated by the smallest perturbations.

To obtain the criterion (1.4.12) let $z \to T(z), s(z), p(z), \rho(z)$ be the thermodynamic functions expressed in terms of the height z.

Let Δ be an infinitesimal cube at height z containing gas with specific volume $v = v(p,z)$. Imagine to displace the mass in Δ and to transfer it in a volume Δ' at height $z' = z + \delta z > z$ adiabatically (because we suppose $\underset{\approx}{\kappa} = 0$ and no heat exchange is possible by conduction).

The new volume occupied by the mass $M = \Delta v(p,s)^{-1}$ will be of size $\Delta' = \Delta\, v(p',s)/v(p,s)$ because the gas will have to acquire pressure p' keeping entropy s, (as in absence of heat conduction the transformations are adiabatic).

At the same time the mass originally in Δ' will have to be moved in Δ. Since this mass is $M' = \Delta'/v(p', s')$ it will occupy, at the new pressure p, a volume

$$\begin{cases} \Delta'' = M'v(p, s') = \Delta' \;\; v(p, s')/v(p', s') = \\ = \Delta \;\; v(p', s)v(p, s')/v(p, s)v(p', s') = (1 + O(\delta z^2))\,\Delta \end{cases} \tag{1.4.13}$$

as one sees by a Taylor expansion of $\log v(p, s)$ using that $s - s'$ and $p - p'$ have order of magnitude δz.

We interpret this by saying that the mass to be moved away from Δ' to make space for the one coming from Δ "does indeed fit" in the volume Δ left free (up to a negligible higher order correction). Therefore the proposed transformation will be energetically favored (in a gravity field) if $M = \Delta/v(p, s) < M' = \Delta' \,/v(p', s') = \Delta \; v(p', s)/v(p, s)v(p', s')$, *i.e.* if

$$\frac{v(p', s)}{v(p', s')} > 1 \tag{1.4.14}$$

If (1.4.14) holds then the equilibrium is unstable and small perturbation will induce the permutation of the two volumes of gas generating a nonzero velocity field $\underline{u} \neq 0$ and raise "convective currents"

To see the "usual" meaning of (1.4.14), *i.e.* of

$$-\left(\frac{\partial v}{\partial s}\right)_p \frac{ds}{dz} > 0 \tag{1.4.15}$$

one can use the relation $(\partial v/\partial s)_p \equiv T/c_p(\partial v/\partial T)_p$. Since, in most substances, it is $(\partial v/\partial T)_p > 0$ the (1.4.15) becomes $-ds/dz > 0$, so that:[2]

$$\begin{aligned} \frac{\delta s}{\delta z} &= \left(\frac{\partial s}{\partial T}\right)_p \frac{\delta T}{\delta z} + \left(\frac{\partial s}{\partial p}\right)_T \frac{\delta p}{\delta z} = \frac{c_p}{T}\frac{\delta T}{\delta z} - \left(\frac{\partial v}{\partial T}\right)_p \frac{\delta p}{\delta z} = \\ &= \frac{c_p}{T}\frac{\delta T}{\delta z} + \left(\frac{\partial v}{\partial T}\right)_p \frac{g}{v} < 0 \Leftrightarrow \frac{\delta T}{\delta z} < -\frac{T}{c_p}\frac{g}{v}\left(\frac{\partial v}{\partial T}\right)_p \end{aligned} \tag{1.4.16}$$

is the general instability condition. In the perfect gas case on has stability if (1.4.12) holds.

Finally, on the basis of the analysis in §1.2, we see that the assumption $\rho =$ constant is licit if we limit our interest to a portion of fluid spanning a height H such that $gH \ll v_{sound}^2$, and a variation of temperature δT such that $(\partial\rho/\partial s)_p\,\delta s/\rho \approx \rho\chi_s\,c_p\delta T/T \approx (\rho\chi_s v_{sound}^2/T)\,(\delta T/T) \ll 1$.

(2) *Hydrostatics in presence of thermoconduction.*

In this case too one finds that hydrostatic solutions are rare and special. For the purposes of an example, and to avoid repetitions, we pose a slightly

[2] From $G = U + PV - TS \Rightarrow dG = -SdT + VdP$.

different problem compared to the ones already discussed and we treat only a simple example.

We ask whether a fluid in a container Ω *and in a conservative force field* $\underline{g}$, $\underline{g} = -\underline{\partial} G$, *can, at least in particular circumstances, conduct heat without developing motion* (*i.e.* if it can "look" like a solid conductor). We shall assume therefore that the temperature on the walls is a preassigned function $\underline{\xi} \to \vartheta(\underline{\xi})$.

For a fluid verifying (1.2.8) one has

$$\begin{aligned} &\partial_t \rho = 0 \\ &\underline{\partial} p = -\rho \underline{\partial} G, \qquad T = \tau(v, p) \\ &T \rho \partial_t s = \kappa \Delta T \\ &T(\underline{\xi}) = \vartheta(\underline{\xi}), \quad \underline{\xi} \in \partial\Omega \ . \end{aligned} \tag{1.4.17}$$

The first equation says that $\rho = \rho(\underline{\xi})$ and the second that ρ must be a function of $\underline{\xi}$ through G so that p can be expressed easily in terms of ρ, namely

$$\text{if} \qquad \rho(\underline{\xi}) = V(G(\underline{\xi}))^{-1}, \quad \text{then} \quad p(\underline{\xi}) = \pi(t) + W(G(\underline{\xi})) \tag{1.4.18}$$

where $V(G)$ is a suitable function and $W(G) = -\int_{G_0}^{G} V(G')^{-1}\, dG'$. Since we must have $T = \tau(v, p)$ it is

$$T(\underline{\xi}) = \tau\big(V(G(\underline{\xi})), \pi(t) + W(G(\underline{\xi}))\big) \tag{1.4.19}$$

for each $\xi \in \Omega$, and, hence, also for $\underline{\xi} \in \partial\Omega$. Then it will also be $\pi(t) = \pi_0$ = constant. Thus hydrostatic solutions are possible only if the temperature assigned on the boundary depends on $\underline{\xi}$ via $G(\underline{\xi})$. In this case also $T(\underline{\xi})$ is a function of $G(\underline{\xi})$ and therefore $s(\underline{\xi})$ has the same property.

Furthermore assuming that $\vartheta(\underline{\xi})$ depends on $\underline{\xi}$ via $G(\underline{\xi})$ it is not clear that there is a solution of

$$\partial_t \rho = 0, \quad p = \pi_0 + W(G), \quad \Delta T = 0, \quad T(\underline{\xi}) = \vartheta(\underline{\xi}) \ \text{on} \ \partial\Omega \tag{1.4.20}$$

In fact the last two conditions on T determine T uniquely (as the solution of a "Dirichlet problem" $\Delta T = 0$ in Ω, $T = \vartheta$ on $\partial\Omega$); and it is not necessarily true that T will be a function of $\underline{\xi}$ via $G(\underline{\xi})$: the latter is a very restrictive condition.

To understand how strong the latter restriction is consider the case of a gravity field

$$G(\underline{\xi}) = gz, \quad \vartheta(\underline{\xi}) = T_0(z), \quad \underline{\xi} \in \partial\Omega \tag{1.4.21}$$

In this case we see that T, ρ, s must be functions of z alone and therefore the equation $\Delta T = 0$ becomes $d^2T/dz^2 = 0$, *i.e.* for a suitable γ

$$T(z) \equiv \vartheta(z) \equiv (1 + \gamma g z)\, T_0. \tag{1.4.22}$$

We thus see that to have hydrostatic solutions not only $\vartheta(\underline{\xi})$ *must be a function of* z *alone but it must be a linear function.*

Finally, if $G = gz$ and $T = (1 + \gamma gz)\, T_0$ we see that the (1.4.20) can be satisfied if $W(G)$ is chosen as solution of the equation obtained by imposing the equation of state $T = \tau(v, p)$:

$$(1 + \gamma G)\, T_0 = \tau \left(\frac{1}{W'(G)}, p_0 + W(G) \right) \tag{1.4.23}$$

which is a differential equation for W which, once solved, gives W, V and therefore ρ and p in terms of $G = gz$.

Obviously the conclusion is that convective motions are necessarily generated inside a fluid in a conservative force field and not in thermal equilibrium, apart from very special cases.

The only case in which, under rather general assumptions, one can have static thermoconduction is an incompressible fluid, see (1.2.8); in this case the equations are

$$\begin{aligned} &\rho = \rho_0\,, \qquad \underline{\partial} p = -\rho_0 \underline{\partial} G, \quad \rho_0 c_p \partial_t T = \kappa \Delta T \\ &s = \sigma(T) \equiv \int^T c_p(T) \frac{dT'}{T'} \end{aligned} \tag{1.4.24}$$

where now s depends only on T and $dS/dT = c_p$ because

$$\left(\frac{\partial s}{\partial p} \right)_T = - \left(\frac{\partial v}{\partial T} \right)_p = 0 \tag{1.4.25}$$

while p has to be thought of as no longer related to s or T because

$$\left(\frac{\partial p}{\partial T} \right)_v = - \left(\frac{\partial s}{\partial v} \right)_T = 0 \tag{1.4.26}$$

One can ask how to reconcile the possibility of a solution of (1.4.24), in which T depends on time, with the impossibility of such a solution that we have just shown in the case of a compressible fluid. In fact the incompressible fluid is in a suitable sense a limit case of the compressible fluid.

In reality a compressible fluid close to an incompressible one (in the sense discussed in §1.3) cannot be, for the above discussion, a static thermoconductor and it will start "flowing". However the motion will be the slower the closer we are to a situation in which the fluid can be regarded as incompressible.

Therefore the question of the connection between (1.4.24) and (1.4.20) implies a study of a nonstatic problem and it will be analyzed later (*c.f.r.* §1.5).

(3) *Current lines and the Bernoulli theorem.*

A fluid motion is called *static* if the velocity and thermodynamic fields describing it are time independent.[2]

For such motions it makes sense to define the "*current lines*" as geometric, time independent, curves; they are just the solutions of the differential equations

$$\dot{\underline{\xi}} = \underline{u}(\underline{\xi}) \ . \tag{1.4.27}$$

Current lines play an important role particularly in the case of isoentropic Euler flows. A simple but important property associated with them is "*Bernoulli's theorem*".

Let $\rho = \rho(p)$ be the adiabatic equation of state of the fluid; then we define, as above, the pressure potential $\Phi(p) = \int^p dp'/\rho(p')$ and therefore the Euler equations are

$$\underline{\partial} \cdot (\rho \underline{u}) = 0 \qquad \underset{\sim}{u} \cdot \underset{\sim}{\partial}\, \underline{u} = -\rho^{-1} \underline{\partial}\, p - \underline{\partial}\, G \tag{1.4.28}$$

Multiplying the second equation by $\underline{u}$ we recognize that it becomes

$$\underset{\sim}{u} \cdot \underset{\sim}{\partial} \left[\frac{\underline{u}^2}{2} + \Phi(p) + G \right] = 0 \tag{1.4.29}$$

If $t \to \underline{\xi}(t)$ is a point that moves on a current line according to (1.4.27) and if $X(\underline{\xi}, t)$ is a function then $\partial_t X + \underline{u} \cdot \underline{\partial}\, X$ is the t–derivative dX/dt of $X(\underline{\xi}(t), t)$ evaluated in $(\underline{\xi}(t), t)$. Hence we see that,setting $X(\underline{\xi}, t) = \underline{u}^2(\underline{\xi})/2 + \Phi(p(\underline{\xi})) + G(\underline{\xi})$, the (1.4.29) says that X is constant along the current lines of the fluid:

$$\frac{\underline{u}^2}{2} + \Phi(p) + G = \text{ constant} \tag{1.4.30}$$

This is an equation expressing the *vis vivae* theorem, as the following classical alternative derivation shows.

Let S' be a surface element through $\underline{\xi}'$ with normal $\underline{n}'$ parallel to the fluid velocity $\underline{u}'$ in $\underline{\xi}'$: draw the current line through every point of S', forming in this way a "current tube" which we shall cut, at a point $\underline{\xi}''$, with an element of surface S'' orthogonal to the velocity $\underline{u}''$ in $\underline{\xi}''$.

Consider the fluid enclosed in the current tube at time $t = 0$. At time $t + \delta t$ the surface S' will be displaced forward by $\underline{u}'\delta t$ while the other surface S'' will be displaced forward by $\underline{u}''\delta t$.

The kinetic energy variation of the considered part of the fluid will be, by the static state assumption, simply

$$\frac{1}{2}\underline{u}''^2 \rho'' \underline{u}'' \cdot \underline{n}'' \delta t S'' - \frac{1}{2}\,\underline{u}'^2 \rho' \underline{u}' \cdot \underline{n}' \delta t S' \tag{1.4.31}$$

[2] Often one calls such flows "stationary": here this appellation is avoided because we shall reserve the name "stationary" for states of the fluid that have well defined statistical properties: see Chaps. 5,6,7.

which must equal the work of the applied forces. The external forces perform a work given by the variation of the potential energy (changed in sign)

$$G'\rho'S'\underline{u}'\cdot\underline{n}'\delta t - G''\rho''S''\underline{u}''\cdot\underline{n}''\delta t \tag{1.4.32}$$

while the calculation of the pressure forces is more delicate because we must take into account that such forces not only work on the external faces and on the bases of the tube but also inside it. To compute the work done by the pressure forces we divide the tube into sections $S' = S_1, S_2, \ldots, S_n = S''$ normal to the velocity and spaced so that the center of S_{i+1} follows the center of S_i by an amount much smaller than the quantity $\underline{n}_i\cdot\underline{u}_i\delta t$, if $\underline{u}_i$ and $\underline{n}_i$ are velocity and, respectively, normal vector to S_i.

Under such conditions the fluid element can be regarded as rigid and subject to a force equal to the difference between the pressures on its two bases times their area. Then the work can be computed as

$$\sum_{i=1}^{n-1}(p_i - p_{i+1})\, S_i\, \underline{u}_i\cdot\underline{n}_i\,\delta t \tag{1.4.33}$$

because the pressure forces do not perform work on the lateral face of the current tube (since they are orthogonal to it: recall that the stress tensor is $-p\,\delta_{ij}$).

Mass conservation imposes that $\rho_i S_i \underline{u}_i\cdot\underline{n}_i\,\delta t = Q$ for all i. Hence (1.4.33) becomes

$$Q\sum_{i=1}^{n-1}\frac{p_i - p_{i+1}}{\rho_i} = -Q\int_{p'}^{p"}\frac{dp}{\rho(p)} = -(\Phi(p'') - \Phi(p'))Q\ . \tag{1.4.34}$$

And summing (1.4.31), (1.4.32), (1.4.34) we find

$$\frac{\underline{u}'^2}{2} + \Phi(p') + G' = \frac{\underline{u}''^2}{2} + \Phi(p'') + G'' \tag{1.4.35}$$

In the case of incompressible motions (1.4.35) becomes simpler becauseqq

$$\Phi(p) = \frac{p}{\rho}\quad\text{(incompressible case)} \tag{1.4.36}$$

where ρ is the (constant) fluid density.

From (1.4.35), (1.4.36) we read that increasing the velocity implies that the pressure diminishes (in the incompressible case) or (in the more general isoentropic case) the potential of pressure diminishes. In the incompressible case to a shrinking of the tube section corresponds an increase of the velocity and therefore a decrease of the pressure. It is a property on which several pumps rely.

Problems

[1.4.1] (*integrability of a vector field*) The (1.4.1) shows that only force fields for which there is an integrating factor $\mu(\underline{x})$, *i.e.* such that $\underline{g} = \mu(\underline{x})\underline{\partial} G$ for some G, can generate hydrostatic solutions; show that, in such solutions, the pressure depends on $\underline{x}$ via $G(\underline{x})$ and that also the product $\rho\mu$ is a function of G. Show also that in the 2–dimensional cases every force field admits, at least locally in the vicinity of a point where it does not vanish, an integrating factor (but in general this is only a local property). (*Idea:* Let $\rho = r(p, s)$ be the equation of state and let $\underline{g} = \mu(\underline{x})\underline{\partial} G$; since two scalar functions with proportional gradients have the same level surfaces the (1.4.1) implies that: p is a function of $\underline{x}$ via G: $p(\underline{x}) = \pi(G(\underline{x}))$ and, again by (1.4.1), $\mu\,\rho = r(\pi(G(\underline{x})), s_0(\underline{x}))\mu(\underline{x}) = F(G(\underline{x}))$ for a suitable F).

[1.4.2] In the context of [1.4.1] show that if the entropy density $s_0(\underline{x})$ is known then one can compute the pressure. Note that, however, in general one needs to check compatibility relations between $s_0(\underline{x}), \mu(\underline{x})$ and the equation of state $\rho = r(p, s)$ in order that the equation be soluble. (*Idea:* The pressure must be a function $\pi(G)$. Then $\partial p/\partial G \equiv r(\pi(G(\underline{x})), s_0(\underline{x}))\,\mu(\underline{x}) = \pi'(G(\underline{x}))$ and from this differential equation one deduces π by fixing its value at a point $\underline{x}_0$ and by integrating the equation along a curve which leads from $\underline{x}_0$ to $\underline{x}$, after having expressed $s_0(\underline{x})$ and $\underline{x}$ in terms of G along the curve. The procedure depends upon the curve and therefore compatibility conditions are necessary.)

[1.4.3] In the context of the above two problems assume that the volume force $\underline{g}$ is conservative with potential G, and assume that the entropy $s_0(\underline{x})$ is given and it is a function of the potential, $s_0(\underline{x}) = S(G(\underline{x}))$, show that the compatibility conditions in [1.4.2] are satisfied and that a hydrostatic solution of the second of the (1.2.1) is possible. (*Idea:* The ρ can be expressed in terms of the equation of state $\rho = r(p, s)$ and of the solution $p = \pi(G)$ of the differential equation

$$\frac{\partial}{\partial G}\pi = r(\pi, S(G)) \quad \pi(G_0) = P_0$$

and the hydrostatic solution will then be $p(\underline{x}) = \pi(G(\underline{x}))$.

[1.4.4] (*temperature and in hydrostatic states*)Check that the hydrostatic solutions in [1.4.3] will, in general, correspond to states of the fluid in which temperature changes from point to point and they will, therefore, be really possible only for very special temperature distributions because in general the temperature will be incompatible with the hydrostatic solution of the third of the (1.2.1). (*Idea:* Note that the equation of state allows us to express T as a function of s, p and it will not be, in general, true that the third of the (1.2.1), will hold, *c.f.r.* [1.4.3], unless $\underset{\sim}{\partial}\underset{\approx}{\kappa}\,\underline{\partial} T = 0$ of course).

[1.4.5] (*calm air condition*) Imagine air as a perfect diatomic gas with molecular mass $m_A = 28.8\, m_H$, $m_H =$ hydrogen mass $= 1.67 \cdot 10^{-24}$ g and take $k_B = 1.38 \cdot 10^{-16}$ erg $^\circ K^{-1}$, $g = 9.8 \cdot 10^2$ cm/sec^2, $c_p = \frac{7}{2}\frac{k_B}{m_A} = \frac{7}{2}\frac{R}{M_A}$, $R = k_B N_0$ and $N_0 =$ Avogadro number, $R = 8.31 \cdot 10^7$ erg/$^\circ K$, $N_0 = 6.022\,10^{23}$). Compute, if the ground temperature is $\overline{T} = 20^\circ$C, which is the value of T_0 such that if the temperature at height $z = 10^3$ m is $T \ge T_0$ then convective currents will not develop. (*Idea:* $T_0(z) = \overline{T} - gzc_p^{-1}$ is the limit case as given by (1.4.12); thus one finds $T_0 > \overline{T} - 9.6^\circ K$ (*i.e.* a gradient of $0.96 \cdot 10^{-2}\,^\circ K/m$). If $T < T_0$ air cannot be observed in a hydrostatic stratified equilibrium.)

[1.4.6] In "real" and calm atmosphere in equilibrium the temperature gradient that is observed is $\sim 0.6 \cdot 10^{-2}$ $^\circ$K/m and therefore the calm atmosphere in normal conditions is in stratified equilibrium. Check this statement by finding and consulting some geophysical data.

[1.4.7] (*incompressibility estimate for air*) Express the condition under which a perfect gas in mechanical equilibrium in the gravity field and at constant temperature can be considered as incompressible. (*Idea:* from the discussion in §1.2 one sees that density variations on the scale l over which sensible variations of pressure occur are such that: $\frac{\Delta\rho}{\rho}\simeq\frac{gl}{c^2}$, where c is the sound velocity. Take, in the case of air, $c\simeq 10^3$ km/h. Check that the characteristic scale over which density variation take place is $\simeq c^2/g$, *i.e.* $\simeq 10^4$m. Hence one can consider that in normal conditions air is incompressible (for what concerns the hydrostatic state) over length scales of the order of a kilometer er and therefore one can use (1.4.11) to evaluate the height from a measurement of pressure. For larger heights ρ cannot any more be considered as constant and to compute the height z in terms of p it becomes necessary to know also how temperature changes with height. At least for quota differences not too large it is possible to evaluate the height from pressure measurements, independently of the temperature distribution: it is the principle on which altimeters work. Using a "naive" altimeter, based on the formula $p=-\rho g z+\gamma_0$ (*i.e.* on an empirical gauge performed under ideal atmospheric conditions) can lead to important errors if the atmospheric conditions are not "ideal".) (*Idea:* It is

$$\frac{\Delta\rho}{\rho}\frac{1}{l}=\frac{9.8}{(10^6\cdot 10^{-3}/3\cdot 6)^2}\mathrm{m}^{-1}=1.27\cdot 10^{-4}\mathrm{m}^{-1}.)$$

[1.4.8] (*gravity and calm planetary athmospheres*) Consider a perfect gas in equilibrium in a gravitational field generated by a sphere of given mass and radius and, defining "stratified equilibria" states in which the thermodynamic quantities depend only upon the distance from the center of the sphere, repeat the analysis performed in this section in the case of the half space. Apply the results to the Earth's atmosphere and to that of some other planet (*e.g.* Mars and Venus), computing which could be the maximum temperature gradient compatible with a stratified equilibrium. Compare the results with the average gradients at the surface of the planets as deduced from known astrophysical data. (*Idea:* Part of the problem is to look for, and find, the necessary astrophysical data.)

[1.4.9] (*a case of impossibility of hydrostatic states*) Consider a perfect gas with equation of state (1.3.1) (*i.e.* $s=c_v\log T-c\log\rho$ and therefore $p=\rho Tc$ and $\varepsilon=\frac{3}{2}cT$, where $c=R/M_0, c_v=\frac{3}{2}c$ if R is the gas constant and M_0 is the mass of a mole). Suppose that viscosity and thermal conductivity are given by the Clausius–Maxwell relations ($\eta=c_1T^{1/2}$, $\kappa=c_2T^{1/2}$, with suitable c_1,c_2: *c.f.r.* [1.1.5]). Suppose that the stress tensor is expressed in terms of the viscosity as $\tau_{ij}=\eta(\partial_iu_j+\partial_ju_i)$. Assume also that the gas is enclosed in a cubic container Ω with walls temperature fixed $T_0(P)$, $P\in\partial\Omega$. Show that in general the gas cannot stay in equilibrium (*i.e.* keep $\underline{u}=\underline{0}$, and $T,p=$ constant) and find a distribution of temperature on the walls T_0 which does not permit configurations of (mechanical and thermal) equilibrium in presence of a gravity force. (*Idea:* Show that the equations are

$$\begin{aligned}
&\partial_t\rho+\underline{\partial}(\rho\underline{u})=0\\
&\rho\, c_v\,(\partial_tT+\underline{u}\cdot\underline{\partial}T)=-p\underline{\partial}\cdot\underline{u}+\underline{\partial}(\kappa\underline{\partial}T)\\
&\partial_t\underline{u}+\underset{\sim}{u}\cdot\underset{\sim}{\partial}\underline{u}=-\frac{1}{\rho}\underline{\partial}p++\underset{\approx}{\tau}{}'\cdot\underset{\sim}{\partial}\underline{u}+\frac{\eta}{\rho}\Delta\underline{u}+\underline{g}\\
&p=\rho Tc,\qquad \varepsilon=\frac{3}{2}cT
\end{aligned}$$

and check that $\underline{u}\equiv\underline{0}$ (mechanical equilibrium) implies that $\Delta T+\frac{1}{2}(\underline{\partial}T)^2/T=0$ (*i.e.* $\Delta T^{3/2}=0$) and $\underline{\partial}\log p=\underline{g}/(cT)$), hence that $\underline{g}$ must be parallel to $\underline{\partial}T$ (considering the rotation of the last expression and using that $\underline{g}$ is conservative): this is in general false. For instance if $T_0(x,y,z)=\vartheta x^{2/3}$ for (x,y,z) on $\partial\Omega$ then $T(x,y,z)=\vartheta x^{2/3}$ is solution of the equation for T but its rotation is not parallel to $\underline{g}$.)

[1.4.10] (*elementary tide theory*) Consider a homogeneous spherical planet T of radius R coated by an ocean of depth $h>0$, large enough and of density negligible with respect to that of the planet. Let L be its small, lonely, satellite (also spherical and homogeneous). Denote by M_T and M_L the respective masses and assume that the motion of the two heavenly bodies about their center of mass be circular uniform. Let ρ be the distance TL of the two heavenly bodies: $\rho \gg R \gg h$. Assuming, for simplicity, the satellite on the equator plane and the planet rotation axis orthogonal to it, compute the equilibrium configuration of the fluid surface and evince Newton's formula according to which the *tidal excursion* (*i.e.* the maximal height variation between successive high and low tide) is $\mu = \frac{3}{2}R(\frac{R}{\rho})^3\frac{M_L}{M_T}$. (*Idea:* If G is the center of mass, its distance from the center T is $\rho_B = \frac{M_L}{M_T+M_L}\rho$ and the angular velocity of revolution of the two heavenly bodies is ω, such that $\omega^2\rho = k(M_L + M_T)\rho^{-2}$, if k is the gravitational constant. Let $\underline{n}$ be a unit vector out of T and note that, imagining the observer standing on the frame of reference rotating around G with angular velocity ω (so that the axis TL has a fixed unit vector $\underline{i}$), the potential energy (gravitational plus centrifugal) in the point $r\underline{n}$ has density proportional to

$$-k\frac{M_T}{r} - k\frac{M_L}{(\rho^2 + r^2 - 2\alpha\rho r)^{1/2}} - \frac{1}{2}\omega^2\,(\rho_B^2 + r^2 - 2\rho_B r\alpha)$$

if $\alpha \equiv \underline{i}\cdot\underline{n} \equiv \cos\vartheta$. Develop this in powers of r/ρ to find

$$\begin{aligned}&-k\frac{M_T}{r} - k\frac{M_L}{\rho}\frac{3}{2}\,\alpha^2(\frac{r}{\rho})^2 - \frac{1}{2}k\frac{M_T}{\rho}(\frac{r}{\rho})^2 + const \equiv\\ &\equiv -k\frac{M_T}{\rho}\big(\frac{\rho}{r} + (\frac{r}{\rho})^2(\frac{M_L}{M_T}\frac{3}{2}\alpha^2 + \frac{1}{2})\big) + cost\end{aligned}$$

because the linear terms cancel in virtue of Kepler's law ($\omega^2\rho^3 = k(M_T + M_L)$); and therefore the equation of the equipotential surface is

$$\frac{\rho}{r} + (\frac{r}{\rho})^2\big(\frac{M_L}{M_T}\frac{3}{2}\cos^2\vartheta + \frac{1}{2}\big) = cost$$

Hence, setting $r = (1+\varepsilon)R$, we find: $\varepsilon \simeq (\frac{R}{\rho})^3\frac{M_L}{M_T}\frac{3}{2}\cos^2\vartheta + cost$; the constant is determined by imposing that the solid of equation $r = (1+\varepsilon(\vartheta))R$ has the same volume as the ball $r = R$ and, of course, h has to be large compared to μ (otherwise ...).)

[1.4.11] (*tides and Moon mass*) Knowing that on the open Atlantic (*e.g.* St. Helena island) the tide excursion is of about $90\,cm$, [EH69], and supposing that this would be the tidal excursion on a Earth uniformly covered by a layer of water in a time independent state and subject to the only action of the Moon, estimate the ratio between the mass of the Moon and that of the Earth. Suppose $R = 6378\,Km$, $\rho = 363.3\,10^3\,Km$ equal to the minimum distance Earth–Moon.

[1.4.12] (*ratio of Moon and Sun tides*) Estimate the ratio between the Moon tide and the Sun tide. (*Idea:* $\frac{\varepsilon_L}{\varepsilon_S} = \frac{M_L}{M_S}(\frac{\rho_S}{\rho_L})^3 \simeq 2$, supposing that the Sun mass is $M_S = 10^6 M_T$, and that the Moon mass is (approximately) the one deduced from the problem [1.4.11]).

[1.4.13] Taking into account the result of [1.4.12] compute again the Moon mass and the ratio between the Sun tide and the Moon tide. (*Idea:* The Moon tide will then be of about $50cm$ rather than the $80\,cm$ of [1.4.11].)

[1.4.14] (*tidal slowing of a planet rotation*) Let ω_D the daily rotation velocity of the planet T above, and suppose that the daily rotation takes place on the same plane of the satellite L orbit. Assume that the planet is uniformly coated by a viscous fluid which adheres to the bottom of the ocean while at the surface it is in equilibrium with the satellite (*i.e.* the tide is in phase with the satellite and therefore rotates with an angular velocity $\omega_D - \omega$ with respect to the planet surface). Let $\omega_D \gg \omega$ and let the depth of the ocean be h; suppose as well that the friction force be η times the gradient of velocity: then the momentum of the friction forces with respect to the rotation axis will be

$$A=\int_0^\pi [\eta\frac{(\omega_D-\omega)R\sin\vartheta}{h}]\cdot[R\sin\vartheta]\cdot[2\pi R^2\sin\vartheta\,d\vartheta]=\frac{8\pi}{3}\frac{\eta R^4(\omega_D-\omega)}{h}$$

Estimate the daily and annual deceleration of the planet assuming that the annual revolution velocity is $\omega_D/365$, and that $\eta=0.10\,gs^{-1}\,cm^{-1}$, $R=6.3\,10^3\,Km$, $h=1.Km$, $\omega=2\pi d^{-1}$, $M_T=5.98\times10^{27}\,$g, estimating the number of years necessary in order that the planet rotation velocity (around its axis) be reduced by a factor e^{-1}. (*Idea:* The inertia moment of the planet is $I=\frac{2}{5}M_TR^2$ and therefore $\dot\omega_D=-A/I$, *i.e.* $\omega_D(t)=\omega_D(0)e^{-t/T_0}$, $T_0=3M_Th/20\eta R^2$. The result is $T_0=1.5\cdot10^7$ years which means that (in this friction model) the day would, at the moment, be longer by about .55 sec every century, *i.e.* by about $0.65\cdot10^{-5}$% a century, obviously showing the inadequacy of the model, see [MD00] for a theory of tides and despinning.)

[1.4.15] (*tides on Mars*) Had Mars an ocean uniformly covering it, how wide would the Sun tide be there? (*Idea:* $\sim$ 5.23 cm because the radius of Mars is $3.394\cdot10^8$ cm, its distance to the Sun is $227.94\cdot10^{11}$ cm and its mass is $0.64\cdot10^{27}$ g.)

[1.4.16] (*tides on Europe and Moon*) Assuming that Europe (satellite of Jupiter) had a deep enough uniform ocean estimate the height of the tide generated by Jupiter. Same for the pair Moon–Earth (*Idea:* write a small computer program to solve the general problem of the static tide generated on a satellite by its planet to solve all problems of this kind and play with various cases like Titan–Saturn *etc.*)

Bibliography: [LL71], [BKM74].

§1.5 The convection problem. Rayleigh's equations.

We now investigate in more detail the (1.2.1), and the (1.2.8), to find a "simple model" (*i.e.* simpler than (1.2.1) themselves) for some incompressible motions in which nontrivial thermal phenomena take place. Essentially we search for some concrete case in which (1.2.8) is derived as a "consequence exact in some asymptotic sense" of (1.2.1). We shall find a physically interesting situation, known as the "*Rayleigh regime*", describing a simple incompressible heat conducting and viscous fluid flowing between two surfaces at constant temperature.

(A) General considerations on convection.

The problem we shall address here is to deduce equations, simpler than the general ones in (1.2.1), valid under physically significant situations and that can still describe at least a few of the phenomena of interest, *i.e.* motions generated by density differences due to temperature differences. We look for a system of equations that could play the role plaid by the incompressible NS equations in the study of purely mechanical fluid motions (*i.e.* motions in which temperature variations and the heat and matter transport generated by them can be neglected).

Incompressibility in the simple form of the assumption that $\rho = const$ is obviously not interesting as, by definition, one has convection when the density variations due to temperature variations *are not* negligible.

Therefore we ask whether a physically compressible fluid, a perfect gas to be specific, admits motions that preserve the volume, *i.e.* such that $\underline{\partial}\cdot\underline{u} = 0$ with a good approximation, without having constant density.

In general, however, the divergence $\underline{\partial}\cdot\underline{u}$ *is not* a constant of motion and, therefore, one can doubt that the above question is a well posed one. And in fact we shall find that solutions with $\underline{\partial}\cdot\underline{u} = 0$ can only exist in an approximate sense, giving up the requirement that the equation of state be exactly verified and replacing it, in some sense, by the $\underline{\partial}\cdot\underline{u} = 0$. More precisely we shall find approximations that transform the general equations (1.2.1) into equations that are approximate but which include among them the $\underline{\partial}\cdot\underline{u} = 0$, assuming that it is verified at the initial time, even though it is not necessarily $\rho = const$.

(B) The physical assumptions of the Rayleigh's convection model.

Consider, for definiteness, a perfect gas

$$s = c_V \log e - c \log \rho \Rightarrow s = c_V \log T - c \log \rho + cost \tag{1.5.1}$$

i.e. $p = c\rho T$, $e = c_V T$, where c_V is the specific heat at constant volume, $c = R/M_{mol}$ is the ratio between the gas constant R and the molar mass, and the internal energy is denoted e to avoid confusion with the adimensional small quantity ε introduced in the following. However the assumption of perfect gas is not necessary and the only change in considering a general fluid is that some quantities will have values that cannot be computed unless one specifies the substance under study.

We suppose the fluid to be enclosed between two horizontal planes at height $z = 0$ and $z = H$, subject to a gravity force $\underline{g} = (0, 0, -g)$.

Boundary conditions are fixed by assigning the "ground temperature and the "temperature in quota", $T = T_0$, if $z = 0$, and $T = T_0 - \delta T$, if $z = H$. Furthermore we shall assume that the velocity field is tangent to the planes $z = 0$ and $z = H$ and that the total horizontal momentum $\int u_j\, d\underline{x}$, $j = 1, 2$, vanishes (in the following we shall take the notations $\underline{u} = (u_x, u_y, u_z)$ and $\underline{u} = (u_1, u_2, u_3)$ as equivalent).

We shall study motions in which the pressure is close to the static barometric pressure, $p = p_0 - \rho_0 g z$, and in which density is close to a given ρ_0 and p_0 is large with respect to $\rho_0 g H$.

Supposing that the viscosity coefficients η, η', *c.f.r.* (1.2.6), are constant (the problems in §1.1 show that this is an assumption that can be reasonable if the temperature variations are sufficiently small, see also below) it follows that the "exact" equations are, if $\nu = \eta/\rho, \nu' = \eta'/\rho$, *c.f.r.* §1.1 and (1.2.1):

$$\begin{aligned}&\rho T\big(\partial_t s+\underset{\sim}{u}\cdot\underset{\sim}{\partial}s\big)=\kappa\Delta T+\frac{\eta}{2}(\underset{\sim}{\partial}\underline{u}+\underset{\sim}{\partial}\underline{u})^2+(\eta+\eta')(\underline{\partial}\cdot\underline{u})^2\\&\partial_t\underline{u}+\underset{\sim}{u}\cdot\underset{\sim}{\partial}\underline{u}=\nu\Delta\underline{u}+(\nu+\nu')\underline{\partial}(\underset{\sim}{\partial}\cdot\underset{\sim}{u})-\frac{1}{\rho}\underline{\partial}p+\underline{g}\\&\partial_t\rho+\underline{\partial}\cdot(\rho\underline{u})=0\end{aligned}\tag{1.5.2}$$

to which one adds the equation of state $s = s(e, \rho)$ in (1.5.1), or (equivalently) the two relations $s = c_V \log T - c \log \rho$, $p = c\rho T$, and also the mentioned boundary conditions. We shall suppose κ, ν constant, for simplicity.

It is interesting to make the side–remark that the condition that friction generates entropy at a positive rate is expressed by $\eta(\underset{\sim}{\partial}\underline{u} + \underset{\sim}{\partial}\underline{u})^2 + \eta'(\underline{\partial} \cdot \underline{u})^2/2 \ge 0$, *i.e.* $\eta' \ge -2\eta$. The case $\eta' + \eta = 0$ is therefore possible, from this point of view, hence it is theoretically relevant.

Note that, given $s_0, \underline{u}_0, \rho_0$ at time $t = 0$, we can compute T_0, p_0 at the same instant via the equation of state and, hence, via (1.5.2), $\dot{\underline{u}}, \dot{s}, \dot{\rho}$ (at $t = 0$): so we can compute $s, \underline{u}, \rho$ at time $dt > 0$. Motion is therefore formally determined by the equations (1.5.1),(1.5.2), aside from problems that might be generated by boundary conditions..

If, furthermore, at $t = 0$ one has $\underline{\partial} \cdot \underline{u} = 0$ it is not necessarily $\underline{\partial} \cdot \dot{\underline{u}} = 0$: hence it is not necessarily $\underline{\partial} \cdot \underline{u} = 0$ at time $dt > 0$.

It follows that the condition $\underline{\partial} \cdot \underline{u} = 0$ can be added only provided we eliminate one of the scalar relations, *e.g.* the continuity equation. And this can only be consistent if the incompressibility conditions seen in §1.2 are realized and if, also, the temperature variations do not cause important density variations.

If α is the thermal expansion coefficient at constant pressure ($\alpha \sim T_0^{-1}$ if δT is small, in our perfect gas case), the latter condition simply means that $\varepsilon \equiv \alpha\delta T \ll 1$. And the incompressibility condition seen in §1.2 is formulated in the same way by requiring that a typical variation v of the velocity has to be small with respect to the sound velocity v_{sound}.

An estimate of v can be obtained by remarking that motions that develop starting from a state close to rest are essentially due to the density variations due to temperature variations, which naturally generate a small archimedean force with acceleration $\alpha\,\delta T\,g$.

Thus a typical velocity in a motion close to rest, at least initially, is the one acquired by a weight that falls from a height H with acceleration $g\alpha\delta T$:

$$v = \sqrt{Hg\alpha\delta T} \tag{1.5.3}$$

and the time scale of such motions will be the time of fall, of the order $\tau_c = H/v$. We shall make some simplifying assumptions, namely

(h1) *We shall only consider motions in which the space scale and the time scale over which the velocity varies are of the order of H and, respectively, of $\nu^{-1}H^2$ or $\tau_c = H/v$ assuming that the latter two times have the same order of magnitude.*

Furthermore

(h2) *We shall suppose that all velocities have the same order of magnitude, otherwise (c.f.r. §1.2) the discussion on incompressibility would be more involved; in this way there will be only one "small" parameter ε:*

$$\varepsilon \sim \alpha\delta T \sim \frac{v}{v_{sound}} \sim \frac{\nu H^{-1}}{v_{sound}} \sim \frac{gH}{v_{sound}^2} \tag{1.5.4}$$

Here $\sim$ means that the ratio of the various quantities stays fixed as $\varepsilon \to 0$: in other words the ratios of the various quantities should be regarded as further parameters; the notations are $\alpha = -\rho^{-1}\big(\frac{\partial\rho}{\partial T}\big)_p \approx T^{-1}$ and $v_{sound}^{-2} = \big(\frac{\partial\rho}{\partial p}\big)_T \approx (cT)^{-1}$.[1] Note that the convective instability condition (1.4.12) ($\frac{dT}{dz} > -\frac{g}{c_p}$) becomes, since $\frac{dT}{dz} \approx \frac{\delta T}{H}$, $\frac{\alpha\delta T}{gH}v_{sound}^2 > 1$ and hence the conditions (1.5.4) correspond to unstable situations (although "marginally" so because this parameter is $O(1)$) with respect to the birth of convective motions, at least in absence of thermoconduction, *c.f.r.* (C) in §1.4.

Convective motions in turn can be more or less stable with respect to perturbations: the latter is a different, more delicate, matter that we shall analyze later. Their instability will thus be possible (and even be strong) depending on other characteristic parameters: we shall see that, for instance, convective motions arise even though the adiabatic stability condition (1.5.4) holds but the quantity $R_{Pr} = \nu\rho c_p/\kappa$ is large.

Writing $\rho^{-1}\delta\rho = \rho^{-1}\big(\partial\rho/\partial T\big)_p\delta T + \rho^{-1}\big(\partial\rho/\partial p\big)_T\delta p$ we see that the fluid can be considered incompressible if, estimating $\rho^{-1}\delta p$ as $|\dot{\underline{u}}| \sim v^2H^{-1} \sim |\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{u}|$ and using (1.5.3) (see 1.2)

$$\alpha\delta T \ll 1, \qquad \frac{Hg\alpha\delta T}{v_{sound}^2} \ll 1 \tag{1.5.5}$$

Hence under the assumptions (1.5.4) the incompressibility condition (1.5.5) will simply be $\varepsilon \ll 1$.

Under the hypothesis (1.5.5), and supposing that the velocity and temperature variations in the motions that we consider take place over typical scales of length of the order H and of time of the order $\nu^{-1}H^2$, the equations of

---

[1] In fact the sound velocity is defined as $(\partial\rho/\partial p)_s^{-1/2} = \sqrt{c(1+c_V/c)T} = \sqrt{c_pT}$, rather than by $\sqrt{cT}$, because usually one considers adiabatic motions; but the two definitions give the same order of magnitude in simple ideal gases because $c_V = 3c/2$.

motion will be written, setting $T = T_0 + \vartheta - \delta T/Hz$, as

$$\begin{aligned}
&\underline{\partial} \cdot \underline{u} = 0\\
&\dot{\underline{u}} + \underset{\sim}{u} \cdot \underset{\sim}{\partial} \underline{u} = \nu \Delta \underline{u} + \underline{g} - \frac{1}{\rho} \underline{\partial} p \qquad (1.5.6)\\
&\dot{\vartheta} + \underset{\sim}{u} \cdot \underset{\sim}{\partial} \vartheta - \frac{\delta T}{H} u_z = \chi \Delta \vartheta + \frac{\nu}{2c_p} (\underset{\sim}{\partial} \underline{u} + \underline{\partial} \underset{\sim}{u})^2
\end{aligned}$$

where we set $\chi = \kappa \rho^{-1} c_p^{-1}$; the continuity equation has been eliminated (and its violations will be "small" if (1.5.5) holds)[2] and the last equation is obtained from the first of (1.5.2) by noting that, within our approximations, the thermodynamic transformation undergone by the generic volume element must be thought as a transformation at constant pressure, so that[3] $T ds = c_p dT$.

We shall suppose that ν, χ, c_p in (1.5.6) are constants (again for simplicity). And we shall always imagine, without mention, that the boundary conditions are the ones specified before (1.5.2).

(C) The Rayleigh model.

The (1.5.6), valid under the hypotheses (1.5.5),(1.5.4), are still very involved and it is worth noting that the conditions (1.5.4) allow us to perform further simplifications because there are regimes in which the equations can contain terms of different orders of magnitude.

For instance we can consider the case in which the (1.5.4) hold and one supposes that the external force g tends to 0 and the height H tends to ∞,

---

2 Indeed the terms $\partial_t \rho$ e $\underline{u} \cdot \underline{\partial} \rho$ of the continuity equation have (both by (1.5.4)) order of magnitude $O(\rho \alpha \delta T \, \nu / H^2)$, while the third term $\rho \underline{\partial} \cdot \underline{u}$ has order $O(\rho v / H)$ and, by (1.5.4), $vH \sim \nu$ so that the ratio of the orders of magnitude is $O(\varepsilon)$. To lowest order the continuity equation is thus $\underline{\partial} \cdot \underline{u} = 0$.

3 A more formal discussion is the following. Imagine s as a function of p, T (in a perfect gas it would be $s = c_p \log T + c \log p$); then

$$T ds = c_p dT + \Big(\frac{\partial s}{\partial p}\Big)_T dT \qquad (1.5.7)$$

and we can estimate the ratio $\dot{p}/\dot{T}$ by estimating $\dot{T}$ as $\delta T H^2/\nu$ and $\dot{p}$ by remarking that $O(|\underline{\partial} p|) \sim O(\rho \dot{\underline{u}}) \sim \rho v \nu / H^2$ and hence the variations δp of p have size $O(\delta p) \sim O(\rho v \nu / H)$ and, therefore, $O(\dot{p}) = O(\rho v \nu^2 / H^3)$. Hence if we compare $\dot{p}/p$ to $\dot{T}/T$ we get $\rho v \nu T / p \delta T H$ which has size $O(\varepsilon)$ by (1.5.4): for instance in the free gas case this is $(v/v_{sound}) \cdot (\nu / H v_{sound}) \cdot (1/\alpha \delta T) = O(\varepsilon)$ (we consider a fixed fluid so that the parameters v_{sound}, ρ, ν are regarded as constants). Hence in the variation of entropy we can suppose $p = constant$ to leading order, so that $T ds = c_p dT$ (and in the free gas $c_p = c_v + c$). Note that the perfect gas assumption is not necessary for the above argument: it is only made to perform an explicit computation of $\big(\partial s / \partial p\big)_T$ which is a constant in this argument as it is a property of the fluid.
Note also, as it will be used in the following, that if $\chi \simeq \nu$ (*c.f.r.* [1.1.5]), the term $\eta (\underset{\sim}{\partial} \underline{u} + \underline{\partial} \underset{\sim}{u})^2 / 2$ has size $O(\eta v^2 H^{-2}) = O(\rho \nu v^2 H^{-2}) = O((v/v_{sound})^2 (\nu / H v_{sound})^2) = O(\varepsilon^4)$ so that it can be eliminated from (1.5.6). See [EM94].

as $\varepsilon \to 0$. This facilitates the estimate of the various orders of magnitude in terms of $\varepsilon = \alpha\,\delta T$. It will be possible, in fact, to fix $g = g_0\varepsilon^2$ and $H = h_0\varepsilon^{-1}$ (keeping fixed ν, v_{sound}, p_0, T_0) which, for small ε, is a regime that we shall call the *Rayleigh regime.* In this regime the typical velocity will be $v = \sqrt{\alpha\delta T g H} = O(\varepsilon)$.

In this situation we can see further simplifications, as $\varepsilon \to 0$, because several terms in the last two equations (1.5.6) have order of magnitude in ε which is $O(\varepsilon^3)$; hence all terms of order $O(\varepsilon^4)$ (or smaller) can be neglected in the limit in which $\varepsilon \to 0$. Indeed

(I) the term $\nu c_p^{-1}(\underset{\sim}{\partial}\underline{u})^2$ in of order $O(\varepsilon^4)$; *i.e.* one can neglect the heat generation, by friction, inside the fluid.

(II) $\underline{g} - \rho^{-1}\underline{\partial}p$ differs from $-\alpha\vartheta\underline{g} + \underline{\partial}p'$, for some suitable p', by $O(e^4)$.

Before discussing the validity of the above (I) and (II) note that the typical velocity variations, *c.f.r.* (1.5.3),(1.5.4), will have order $v = \sqrt{gH\alpha\,\delta T} = O(\varepsilon)v_{sound}$ while the typical deviations of temperature and pressure from the hydrostatic equilibrium values will have order $\alpha\,\delta T = \varepsilon$ or v/v_{sound} (*i.e.* again $\le O(\varepsilon)$) respectively, if measured in adimensional form).

Hence neglecting terms of order $(\alpha\,\delta T)^4$ allows us to keep in a significant way the nonlinear terms in (1.5.6), which have order $O(\varepsilon^3)$.

We first discuss (II); we remark that by the definition of ϑ, see (1.5.6), it is

$$\begin{aligned}\mathrm{rot}\,(-\frac{1}{\rho}\underline{\partial}p) =& \frac{1}{\rho^2}(\underline{\partial}\rho \wedge \underline{\partial}p) = \frac{1}{\rho^2}\big(\frac{\partial\rho}{\partial T}\big)_p(\underline{\partial}\vartheta + \frac{\delta T}{H}\frac{\underline{g}}{g}) \wedge \underline{\partial}p = \\ =& \frac{\alpha\rho c}{\rho^2}\underline{\partial}\vartheta \wedge \rho\underline{g} + O(\varepsilon^4) \equiv -\alpha\, c\,\underline{\partial} \wedge (\vartheta\underline{g}) + O(\varepsilon^4)\end{aligned} \tag{1.5.8}$$

because if, as we are assuming, we think that $H = h_0\varepsilon^{-1}, g = g_0\varepsilon^2$, with $\varepsilon = \alpha\,\delta T$ then

(a) in the first line, noting that $dT = d\vartheta - \frac{\delta T}{H}dz$, $\underline{\partial}z = -\underline{g}/g$, we use $(\frac{\partial\rho}{\partial T})_p = \rho/T = \alpha\rho$ and the part of $\underline{\partial}\rho$ proportional to $(\frac{\partial\rho}{\partial p})_T\underline{\partial}p$ does not contribute; furthermore,
(b) $\underline{\partial}p - \rho\underline{g}$ has order $O(\varepsilon^3)$ (*i.e.* the order of $\dot{\underline{u}}$ and hence of the product of νH^{-2} times the typical velocity $v = O(\varepsilon)$, see above). Thus we can replace $\underline{\partial}p$ in the last term in the first line of (1.5.8) with $\rho\underline{g}$ up to $O(\varepsilon^4)$ and

Therefore (II) implies $\mathrm{rot}\,(-\rho^{-1}\underline{\partial}p + \alpha\underline{g}\vartheta) = O(\varepsilon^4)$, *i.e.* for some p' it is

$$-\frac{1}{\rho}\underline{\partial}p = -\alpha\vartheta\underline{g} + \underline{\partial}p' + O((\alpha\,\delta T)^4) \tag{1.5.9}$$

or, since $\underline{g}$ is conservative, also $\underline{g} - \rho^{-1}\underline{\partial}p = -\alpha\vartheta\underline{g} + \underline{\partial}p' + O(\varepsilon^4)$.

We now turn to (I), analyzing with the method of §1.2 its physical significance. In the Rayleigh regime the term that we want to neglect has order of magnitude

$$\frac{\nu}{c_p}(\underset{\sim}{\partial}\,\underline{u})^2 \sim \frac{\nu}{c_p}\frac{Hg\alpha\,\delta T}{H^2} = O(\varepsilon^4) \tag{1.5.10}$$

and it has, therefore, to have order of magnitude small compared to the order of magnitude of the other terms of the equation, *i.e.* $\dot\vartheta, \underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\vartheta$, $\frac{\delta T}{H}u_z, \chi\Delta\vartheta$. And one has

$$\begin{aligned}\dot\vartheta \sim& O(\frac{\nu\delta T}{H^2}), \qquad \frac{\delta T}{H}u_z \sim O(\frac{\delta T}{H}\sqrt{g\alpha\,\delta T\,H})\\ \underset{\sim}{u}\cdot\underset{\sim}{\partial}\vartheta \sim& O(\sqrt{gH\alpha\,\delta T}\frac{1}{H}\delta T), \qquad \chi\Delta\vartheta \sim O(\chi\frac{\delta T}{H^2})\end{aligned} \tag{1.5.11}$$

and comparing (1.5.10) with (1.5.11) one finds that the incompressibility conditions (1.5.5) and the conditions of validity of (I) and (II) can be summarized into

$$\begin{aligned}\sqrt{gH\alpha\,\delta T} \ll v_{sound}, \qquad& \varepsilon \equiv \alpha\,\delta T \ll 1,\\ c_p\,\delta T \gg Hg\alpha\,\delta T, \qquad& \frac{\nu}{c_p}\frac{\sqrt{gH\alpha\,\delta T}}{H\delta T} \ll 1\end{aligned} \tag{1.5.12}$$

supposing, as said above, that motions take place over length and time scales given by H and by $H^2\nu^{-1}$ respectively and the conditions can be simultaneously satisfied by choosing

$$\varepsilon = \alpha\,\delta T, \quad g = g_0\varepsilon^2, \quad H = h_0\varepsilon^{-1}, \quad \frac{g_0 h_0}{c_p T_0} \approx 1, \quad \frac{\nu}{\sqrt{c_p T_0}h_0} \approx 1 \tag{1.5.13}$$

where g_0, h_0, T_0 are fixed and we used $\alpha \approx T^{-1}$, $v_{sound}^2 \sim c_p T$, provided it is $\varepsilon \ll 1$, (note that these relations are just the (1.5.4)).

In such conditions the (1.5.11) have all size $O(\varepsilon^3)$ and the equations, including the boundary conditions specified before (1.5.2), become

$$\begin{aligned}&\underline{\partial}\cdot\underline{u} = 0\\ &\underline{\dot u} + \underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{u} = \nu\Delta\underline{u} - \alpha\vartheta\underline{g} - \underline{\partial}p'\\ &\dot\vartheta + \underset{\sim}{u}\cdot\underset{\sim}{\partial}\vartheta = \chi\Delta\vartheta + \frac{\delta T}{H}u_z\\ &\vartheta(0) = 0 = \vartheta(H), \quad u_z(0) = 0 = u_z(H), \quad \int u_x d\underline{x} = \int u_y d\underline{x} = 0\end{aligned} \tag{1.5.14}$$

and we do not write the equation of state nor the continuity equation because s and ρ no longer appear in (1.5.14) (and, in any event, the equation of state will not hold other than up to a quantity of order $O((\alpha\,\delta T)^2)$. The function p' is related, *but not equal*, to the pressure p: within the approximations it is $p = p_0 - \rho_0 g z + p'$.

We must expect, for consistency, that $\underline{u} = O((\alpha\,\delta T)) = O(\varepsilon)$ and $\vartheta = O(\varepsilon)$, and that the equations make sense up to $O((\alpha\,\delta T)^4)$, *as they now consist entirely of terms of order* $O(\varepsilon^3)$.

In fluidodynamics one defines various *numbers* by forming dimensionless quantities with the parameters that one considers relevant for the stability of the flows studied. In the present case the nonlinear terms in (1.5.14) make sense, and one can define a number measuring the strength of the flow, namely $R = v/v_c$ with $v_c = \nu/H$ (*i.e.* $R = \sqrt{\alpha\,\delta T\,gH}/(\nu/H)$): instabilities can arise for large R, *i.e.* for large velocity variations.

One should stress that, in the considered regime, the Reynolds number is a "free" parameter, ε–independent (by (1.5.13)) in the sense that it is possible to keep R constant while $\varepsilon \to 0$.

In a general flow the "*Reynolds number*" R of a velocity field is defined as the ratio between a typical velocity and the "geometric speed", *i.e.* a velocity formed by the viscosity and a typical length scale. Sometimes there are several numbers that one can imagine to define because there are various different length or time scales. In the present situation the number R, or better R^2, formed by using the "geometric speed" scale ν/H is called *Grashof number*, see [LL71].

In fact there is a second 'natural "geometric speed scale": namely $H^{-1}\chi$. Often $\chi \sim \nu$ (as, in perfect gases, the Clausius–Maxwell relation implies, *c.f.r.* problem [1.1.5]: $\chi = \nu\, c_p/c_V$): but there are materials for which $\nu\chi^{-1} \equiv R_{Pr}$, called the *Prandtl number*, is very large and, therefore, the speed $H^{-1}\chi$ is very different from $H^{-1}\nu$ and instability phenomena can arise at lower velocity gradients. The following table gives an idea of the orders of magnitude (*c.f.r.* [LL71], p. 254) of the experimental values of R_{Pr}

Mercury	0.044
Air	0.733
Water	6.75
Alcohol	16.8
Glycerine	7250.

The convective instability problem, *i.e.* the determination of the values of the parameters R and R_{Pr} in correspondence of which the trivial solution $\underline{u} = \underline{0}$, $\vartheta = 0$ of (1.5.14) loses stability (in the sense of linear stability), was investigated by Rayleigh who did show, as it will be seen in §4.1, that the convective instability is controlled by the size of the product $R^2 R_{Pr}$, sometimes called the *Rayleigh number*:

$$R_{Ray} = \frac{g\alpha\,\delta T\,H^3}{\chi\nu} = R^2\,R_{Pr} \tag{1.5.15}$$

or, sometimes, the *Péclet number*, [LL71].

Remarks:
(1) Note that if the (1.5.4) hold then $R = O(1)$, *i.e.* it stays fixed as $\varepsilon \to 0$. Hence, physically, R_{Ray} large is in general related to large R_{Pr} as well as to large R. However in perfect gases $R_{Pr} = 1$, *c.f.r.* [1.1.5].
(2) Obviously, since the problem allows us to define two independent dimensionless numbers (except in the free gas case, as remarked) we must expect that there is a two–parameters family of phenomena described by (1.5.14) and one should not be surprised that for each of them one could define a characteristic number having the form $R^a R^b_{Pr}$: considering the "large" quantity of possible pairs of real numbers (a, b) one realizes that there is the possibility to make famous not only one's own name, but also that of friends (and enemies), by associating it to a "convective number".

To organize rationally the convective numbers it is useful to define the following adimensional quantities

$$\begin{aligned}\tau =&t\nu H^{-2},\ \xi = xH^{-1},\ \eta = yH^{-1},\ \zeta = zH^{-1},\\ \vartheta^0 =&\frac{\alpha\vartheta}{\alpha\,\delta T},\ \underline{u}^0 = (\sqrt{gH\alpha\,\delta T})^{-1}\,\underline{u}\end{aligned}\tag{1.5.16}$$

where the functions $\underline{u}^0, \vartheta^0$ are regarded as functions of the arguments (τ, ξ, η, ζ).
One checks easily that the Rayleigh equations in the new variables take the form

$$\begin{aligned}&\dot{\underline{u}} + R\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{u} = \Delta\underline{u} - R\vartheta\underline{e} - \underline{\partial}p, \qquad R^2 = \frac{gH^3\alpha\,\delta T}{\nu^2}\\ &\dot{\vartheta} + R\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\vartheta = R_{Pr}^{-1}\Delta\vartheta + Ru_z, \qquad R_{Pr} = \frac{\nu}{\kappa}\\ &\underline{\partial}\cdot\underline{u} = 0 \\ &u_z(0) = u_z(1) = 0, \qquad \vartheta(0) = \vartheta(1) = 0, \qquad \int u_x d\underline{x} = \int u_y d\underline{x} = 0\end{aligned}\tag{1.5.17}$$

where after the change of variables we eliminated the labels 0 and recalled t, x, y, z the adimensional coordinates τ, ξ, η, ζ in (1.5.16) redefining p suitably; furthermore we have set $\underline{e} = (0, 0, -1)$.

The equations (1.5.17) hold under the hypothesis that (1.5.13) hold: note again that in such case $R = g_0 h_0^3 \nu^{-2}$ and $R_{Pr} = \nu\kappa^{-1}$ can be fixed independently of ε. This is important because it shows that various regimes depending on two parameters (the parameters R, R_{Pr}) exist in which the equations are admissible, if ε is small.

Remark: one can also note that if δT was < 0, *i.e.* if the temperature increased with height, the equations (1.5.17) and (1.5.14) would "only" change because of the sign of ϑ in the first of the (1.5.17) or of $\vartheta\underline{g}$ in the second of (1.5.14). This can be seen by looking back at the derivation, or by remarking that changing the sign of δT is equivalent to exchange the role of $z = 0$ and $z = H$. The heat transport equations between two horizontal planes with the one above warmer that the lower one (where "up" and "down" are defined by the direction of gravity) are therefore, if $\delta T > 0$

$$\begin{aligned}
&\dot{\underline{u}} + R\underset{\sim}{u} \cdot \underset{\sim}{\partial}\, \underline{u} = \Delta \underline{u} + R\vartheta \underline{e} - \underline{\partial} p, \qquad R^2 = \frac{gH^3\alpha\,\delta T}{\nu^2}\\
&\dot{\vartheta} + R\underset{\sim}{u} \cdot \underset{\sim}{\partial}\, \vartheta = R_{Pr}^{-1}\Delta\vartheta + Ru_z, \qquad R_{Pr} = \frac{\nu}{\chi}\\
&\underline{\partial} \cdot \underline{u} = 0\\
&u_z(0) = u_z(1) = 0, \qquad \vartheta(0) = \vartheta(1) = 0, \qquad \int u_x d\underline{x} = \int u_y d\underline{x} = 0
\end{aligned} \tag{1.5.18}$$

which are, however, less interesting because they do not imply any (linear) instability of the thermostatic solution, see §4.1.

(D) Rescalings: a systematic analysis.

One can ask for a more systematic way to derive (1.5.17). One can again use the method, actually very general, employed in §1.3.
Suppose that $\varepsilon \equiv \alpha\,\delta T$, $\alpha = T^{-1}$, $v_{sound}^2 = c_p T$ and

$$\begin{aligned}
\varepsilon =&\alpha\delta T, \qquad \varepsilon' = \frac{gH}{cT} \approx \varepsilon, \qquad \varepsilon'' = \frac{\nu}{Hv_{sound}} \approx \varepsilon, \qquad \varepsilon \to 0\\
R =&\frac{gH^3\alpha\,\delta T}{\nu^2}, \qquad \text{and} \quad R_{Pr} = \frac{\nu}{\kappa} \qquad \text{fixed}
\end{aligned} \tag{1.5.19}$$

which is a "regime" that we shall (as above) call the *Rayleigh convective regime* with parameter ε. *We are, automatically, in this regime if the parameters are chosen as in* (1.5.13). We now look for a solution of (1.5.2) which can be written as

$$\begin{aligned}
\underline{u}(\underline{x}, t) =&(gH\alpha\,\delta T)^{1/2}\, \underline{u}^0(\underline{x}H^{-1}, t\nu H^{-2})\\
T(\underline{x}, t) =&T_0 - \frac{\delta T}{H} z + \delta T \vartheta^0(\underline{x}H^{-1}, t\nu H^{-2})\\
\rho(\underline{x}, t) =&\rho_0(zH^{-1}) + \varepsilon r^0(\underline{x}H^{-1}, t\nu H^{-2})\\
p =&\varepsilon p_0(zH^{-1}) + \varepsilon^2 p^0(\underline{x}H^{-1}, t\nu H^{-2}), \qquad g = \varepsilon^2 g_0
\end{aligned} \tag{1.5.20}$$

where $\underline{u}^0(\underline{\xi}, \tau), \vartheta^0(\underline{\xi}, \tau), r^0(\underline{\xi}, \tau)$ *can be thought of as power series* in ε with coefficients *regular* in $\underline{\xi}, \tau$: note that under the assumptions (1.5.19) the three parameters $\varepsilon, \varepsilon', \varepsilon''$ are estimated by ε, which therefore is our only small parameter. Furthermore the functions $T_0 - \delta T z H^{-1}$ and $\rho_0(zH^{-1}), p_0(zH^{-1})$ are solutions of the "time independent problem", *i.e.* of (1.4.6), with boundary conditions $\rho_0(0) = \bar{\rho}$ and T_0; namely

$$\rho_0(\zeta) = \overline{\rho}, \qquad p_0(\zeta) = p_0(0) + \overline{\rho} g_0 h_0 \zeta \tag{1.5.21}$$

obtained from (1.4.10) by applying the equation of state to express ρ_0 in term of p (given by (1.4.10)) and of the temperature $T_0 - \delta T\zeta$.

One can now check, by direct substitution of (1.5.20) into (1.5.2), with $\underline{u}^0 = \underline{u}^1 + \varepsilon\underline{u}^2 + \ldots$, $\rho = \rho_0 + \varepsilon r^1 + \varepsilon r^2 + \ldots$, $p^0 = p^2 + \varepsilon p^3 + \ldots$ and $\vartheta^0 = \vartheta^1 + \varepsilon\vartheta^2 + \ldots$, that the lowest orders $\underline{u}^1, \vartheta^1, \overline{\rho}, p_0$ verify (1.5.17).

As in footnote [3] one has to note that $\dot{p}$ turns out to be of $O(\varepsilon^4)$ in an equation in which all terms have the same size of order $O(\varepsilon^3)$ so that $\dot{p} = 0$ up to order $O(\varepsilon^4)$ and (since $p = c\rho T$) we can replace $T\dot{\rho}$ by $-\rho\dot{T}$ in the entropy equation. Also here the assumption of perfect gas is not essential.

Hence the lowest order of $\underline{u}^0, \vartheta^0, \rho$ describes, at least heuristically, the asymptotic regime in which the (1.5.17) are exact: we shall say that the *convective Rayleigh equation (1.5.17) is expected to be is exact in the limit "*$\varepsilon \to 0$ *with the relations (1.5.13) fixed"*.

This is a mathematically transparent method, apt to clarify the meaning of the approximations and it is the version for the (1.5.2) (*i.e.* for the problem of the gas between two planes at given temperatures) of the analysis of the incompressibility assumption of §1.3. Hence we can hope in the validity of theorems of the type of the ones in §1.3: however such theorems have not (yet) been proved in the present case.

This viewpoint is clearly more systematic because it allows us, in principle, also to find the higher order (in ε) corrections which, in a similar way, should verify suitable equations.

To determine that, in a suitable limit, a certain regime (*i.e.* certain simplified equations) are "asymptotically exact" it is usually necessary to proceed empirically as above (or as in §1.2) and only *a posteriori*, once the structure of the equations has been understood and the relevant adimensional parameters have been identified, it becomes possible to "guess" the right rescaling and the limit in which the equations "become exact".

It is convenient to note here that it is not impossible that for the same equation one can find several distinct "regimes" in which the solutions are described by "rescaled" equations (simpler than the original ones, but usually different and depending on the regime). Although we do not attempt to discuss an example for the model (1.5.2) considered here, there are many other examples: we already met in §1.3 an instance in which a fluid can be found in a "Euler regime" or in the (different) "Navier–Stokes regime".

Problems

[1.5.1]: Examine some consequence of a violation of the (1.5.4).

Bibliography: [LL71]: §50,§53,§56; and [EM93]: from the latter work, in particular, I have drawn many of the basic ideas and the methods of the present section.

§1.6 Kinematics: incompressible fields, vector potentials, decom-

positions of a general field.

It is important to keep in mind various representations of velocity fields in terms of other vector fields with special properties. Much as in electromagnetism it can be important to represent electric or magnetic fields in terms of potentials (like the Coulomb potential or the vector potential). Indeed, sometimes, the basic equations expressed in terms of such auxiliary fields take more transparent or simpler forms. In this section and in the problems at the end of it we shall discuss some among the simplest representation theorems of vector fields of relevance in fluidodynamics (and electromagnetism).

(A) Incompressible fields in the whole space as rotations of a vector potential.

Consider incompressible fluids: the continuity equation will require that

$$\underline{\partial} \cdot \underline{u} = 0 \quad \text{in } \Omega \tag{1.6.1}$$

Suppose $\underline{u}$ of class C^∞ on $\Omega = R^3$ and rapidly decreasing at infinity: *i.e.* for each $p, q \geq 0$ let $|\xi|^q \underset{\sim}{\partial}^p \underline{u}(\xi) \xrightarrow[|\xi|\to\infty]{} 0$.[1] Then there is a vector field $\underline{A}$ such that

$$\text{rot } \underline{A} = \underline{u}, \qquad \underline{\partial} \cdot \underline{A} = 0 \tag{1.6.2}$$

Constructing $\underline{A}$ is elementary if one starts from the Fourier transform representation of $\underline{u}$

$$\underline{u}(\underline{\xi}) = \int \hat{\underline{u}}(\underline{k}) e^{i\underline{k}\cdot\underline{\xi}}\, d\underline{k} \tag{1.6.3}$$

Indeed (1.6.1) means $\underline{k} \cdot \hat{\underline{u}}(\underline{k}) = 0$ *i.e.* $\hat{\underline{u}}(\underline{k}) = i\underline{k} \wedge \underline{a}(\underline{k})$ where $\underline{a}(\underline{k})$ is a suitable vector orthogonal to $\underline{k}$ and unique for $\underline{k} \neq \underline{0}$. Hence

$$\underline{u}(\underline{\xi}) = \int i\underline{k} \wedge \underline{a}(\underline{k}) e^{i\underline{k}\cdot\underline{\xi}} d\underline{k} = \text{rot} \int \underline{a}(\underline{k}) e^{i\underline{k}\cdot\underline{\xi}} d\underline{k} = \text{rot}\, \underline{A}(\xi) \tag{1.6.4}$$

However the vector field $\underline{A}$ in general, although being of class C^∞, will not have a rapid decrease at infinity.

This can be evinced by remarking that for $\underline{k} \to \underline{0}$ the expression for $\underline{a}(\underline{k})$ in terms of $\hat{\underline{u}}(\underline{k})$ will not, in general, be differentiable in $\underline{k}$. In fact from $\underline{\partial} \cdot \underline{u} = 0$ it follows that $\hat{\underline{u}}(\underline{0}) = \underline{0}$.[2] This implies only that $\underline{u}(\underline{k})$ has order $\underline{k}$ for $\underline{k} \to \underline{0}$. However to have that $\underline{a}(\underline{k}) = i\underline{k} \wedge \hat{\underline{u}}(\underline{k})/\underline{k}^2$ be regular in $\underline{k} = \underline{0}$ one should have that $\underline{k} \wedge \hat{\underline{u}}(\underline{k})$ has the form of a product of $\underline{k}^2$ times a regular function of $\underline{k}$: but the vanishing of $\hat{\underline{u}}(\underline{0})$ implies only that $\underline{a}(\underline{k})$ is bounded as

[1] Here and in the following we shall denote with symbols like $\underline{\partial}^p$ a generic derivative of order p with respect to the coordinates $\underline{x}$.

[2] Because $\hat{\underline{u}}(\underline{k})$ is of class C^∞ in $\underline{k}$ by the decrease of $\underline{u}$ and of its derivatives as $\xi \to \infty$ hence $\hat{\underline{u}}(\underline{k}) = \hat{\underline{u}}(\underline{0}) + O(\underline{k})$ and $0 \equiv \underline{k} \cdot \hat{\underline{u}}(\underline{k}) = \underline{k} \cdot \hat{\underline{u}}(\underline{0}) + O(\underline{k}^2)$: since $\underline{k}$ is arbitrary it must be $\hat{\underline{u}}(\underline{0}) = \underline{0}$.

$\underline{k} \to 0$ with, in general, a limit depending on the direction along which one lets $\underline{k}$ tend to $\underline{0}$ (and therefore it is not differentiable in $\underline{k}$ at $\underline{0}$). Hence $\underline{a}(\underline{k})$ is bounded and approaches rapidly zero as $\underline{k} \to \infty$, but it is not everywhere differentiable in $\underline{k}$ and, as a consequence, its Fourier transform $\underline{A}(\xi)$ will be C^∞ but it will not tend to zero rapidly for $\xi \to \infty$.

The above can also be derived directly from the formula, which does not involve Fourier transforms,

$$\underline{A}(\xi) = \frac{1}{4\pi} \int_{R^3} \frac{d\eta}{|\xi - \eta|} \operatorname{rot} \underline{u}(\eta) \tag{1.6.5}$$

i.e. $\underline{A} = -\Delta^{-1} \operatorname{rot} \underline{u}$, well known in electromagnetism ($\underline{A}$ can be interpreted as the magnetic field generated by a current $\underline{u}$ according to the "*Biot–Savart law*").

(B) An incompressible field in a finite convex volume Ω *as a rotation: some sufficient conditions.*

Often one needs to consider divergenceless vector fields representing the velocity $\underline{u}$ of a fluid in a finite container Ω which we always suppose with a very regular boundary of class C^∞. We ask whether given $\underline{u} \in C^\infty(\Omega)$ it is possible to find $\underline{A} \in C^\infty(\Omega)$ so that

$$\underline{u} = \text{ rot } \underline{A} \qquad \text{div } \underline{A} = 0 \tag{1.6.6}$$

A field $\underline{A}$ verifying (1.6.6) and in $C^\infty(\Omega)$ will be called a *vector potential* for $\underline{u}$.

Evidently to show that a field $\underline{A}$ exists it will suffice to show, see Sec. 1.6.1, that $\underline{u}$ can be extended outside Ω to a function $C^\infty(R^3)$ vanishing outside a bounded domain $\tilde{\Omega} \supset \Omega$ and, everywhere, with vanishing divergence.

Referring to the problem in (A) the difficulty is that *the existence of an extension is not evident*. And if Ω contains "holes" it is not true in general (see the problem [1.6.12] for an example). In what follows we shall exhibit one such field by considering convex domains Ω for which the geometric construction is possible. The theorem can be extended to far more general domains (regular convex domains whose boundary points can be connected to ∞ by a curve that has no other points in Ω, *i.e.* domains "with no holes" see problems [1.6.11]%[1.6.14]).

To show the existence of an extension of $\underline{u}$ outside Ω we need to understand better the structure of the divergence free vector fields. We consider the case in which Ω is strictly convex and with analytic boundary (see problem [1.6.8]) for the general case).

Consider the components u_2 and u_3 of $\underline{u}$: we shall extend them to functions, that we still denote with the same name, defined on the whole R^3 and there of class C^∞ and vanishing outside a sphere $\tilde{\Omega}$ of large enough radius containing Ω in its interior. For this purpose we consider the cylinder of all the lines parallel to the axis 1 that cut Ω: the ones tangent to $\partial\Omega$ will define a closed line λ on $\partial\Omega$ and smooth.

Let Σ be a surface of class C^∞ containing λ and intersecting transversally (*i.e.* with a non zero angle) every line parallel to the axis 1. We consider the function u_1 on $\Sigma\cap\Omega$ and extend it to a function defined on the whole Σ, of class C^∞ there, and vanishing outside the sphere $\tilde\Omega$.

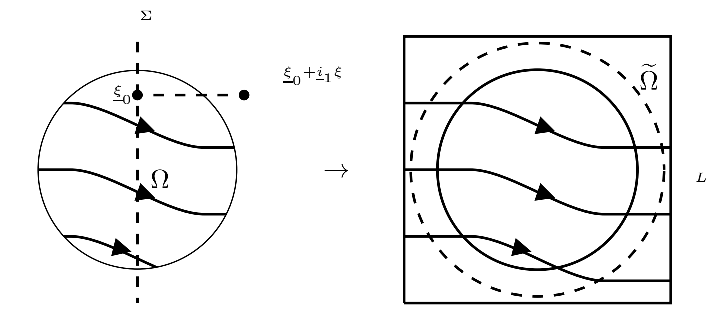


Fig. (1.6.1): The directed lines reprent the field $\underline{u}$, the circle represents Ω and the vertical dotted line the surface Σ while the horizontal dotted line represents a segment parallel to the axis 1. The dotted circle represents $\tilde\Omega$.

Every point of R^3 can be represented as $\underline{\xi}_0+\underline{i}_1\xi$ with $\underline{\xi}_0\in\Sigma$ and $\xi\in R$ and we shall denote it $(\underline{\xi}_0,\xi)$, see Fig. (1.6.1). We then define

$$u_1(\underline{\xi}_0+\underline{i}_1\xi)=u_1(\underline{\xi}_0)+\int_0^\xi-\sum_{j=2}^3\partial_j u_j(\underline{\xi}_0+\underline{i}_1\xi')\,d\xi' \tag{1.6.7}$$

and we thus obtain a velocity field $\underline{u}$ with zero divergence defined in R^3 and extending the field given in Ω. By construction this field is identically zero outside a cylinder parallel to the axis $\underline{i}_1$ containing Ω. Furthermore if one moves along the axis $\underline{i}_1$ and by a distance large enough away from Ω it becomes constant and parallel to the axis 1 itself

$$\underline{u}(\underline{\xi}_0+\xi\underline{i}_1)=\begin{cases}V_+(\underline{\xi}_0)\,\underline{i}_1 & \xi>0 \text{ large}\\ V_-(\underline{\xi}_0)\,\underline{i}_1 & -\xi>0 \text{ large}\end{cases} \tag{1.6.8}$$

where $V_\pm(\xi_0)=u_1(\xi_0)+\int_0^{\pm\infty}-\sum_{j=2}^3\partial_j u_j\,d\xi'$, *c.f.r.* (1.6.7). Here "large" means $|\xi|>L_0$ for some L_0: let $L>L_0$. This is illustrated in the left square in Fig. (1.6.2).

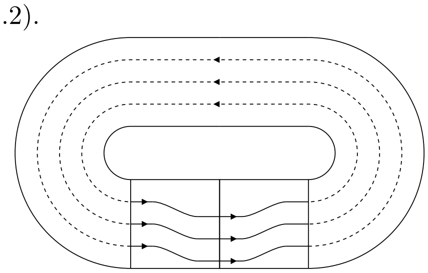


Fig. (1.6.2): The left square contains the extension of the field current lines until they become parallel to the 1–axis. The right square is the mirror image of the left square and the curved dotted lines match the lines exiting from the right square with the corresponding ones of the left square.

We now set

$$\begin{aligned} u'_j(\underline{\xi}_0 + (L+\xi)\underline{i}_1) &= -u_j(\underline{\xi}_0 + (L-\xi)\underline{i}_1) \qquad && j = 2,3 \\ u'_1(\underline{\xi}_0 + (L+\xi)\underline{i}_1) &= +u_1(\underline{\xi}_0 + (L-\xi)\underline{i}_1) \qquad && 0 < \xi < 2L \end{aligned} \tag{1.6.9}$$

It is clear that $\underline{u}'$ extends $\underline{u}$ from $\tilde{\Omega}$ to R^3 and that for $\xi \to \pm\infty$ it has for every "transversal coordinate" $\underline{\xi}_0 \in \Sigma$, the same limit $V_-(\underline{\xi}_0)\underline{i}_1$; furthermore $\underline{\partial} \cdot \underline{u}' = 0$.

Consider now a cylinder Γ parallel to $\underline{i}_1$, such that $\tilde{\Omega} \subset \Gamma$, with a bounded circular base orthogonal to $\underline{i}_1$. Continue the lines of the cylinder which are parallel to the axis into a smooth bundle of curves parallel to the axis $\underline{i}_1$ so that each closes onto itself as symbolically drawn in Fig. (1.6.2).

In the closed tube $\tilde{\Gamma}$ consider a vector field $\underline{\tilde{u}}$ defined so that it remains tangent to the just constructed curves (which define $\tilde{\Gamma}$). The field $\underline{\tilde{u}}$ will be equal to $V_-(\underline{\xi}_0)\underline{i}_1$ at the point where $\tilde{\Gamma}/\Gamma$ joins the curve that has transversal coordinate $\underline{\xi}_0$.

One then continues the function $\underline{\tilde{u}}$ in $\tilde{\Gamma}$, outside Γ, so that the flow of $\underline{\tilde{u}}$ through equal surface elements normal to each curve, among the ones considered, remains constant.

The field $\underline{u}$ is of class C^∞, vanishes out of $\tilde{\Gamma}$ and everywhere it as zero divergence.

We can then think of $\underline{\tilde{u}}$ as a field on R^3 vanishing outside the tube $\tilde{\Gamma}$ and write it as

$$\underline{\tilde{u}} = \text{ rot } \underline{A} \ , \quad \underline{\partial} \cdot \underline{A} = 0 \ . \tag{1.6.10}$$

where $\underline{A}$ is a suitable vector field of class C^∞ (*c.f.r. (A)*). This shows that the restriction of $\underline{A}$ to Ω has the properties that we wish for a vector potential.

(C) Ambiguities for vector potentials of incompressible fields.

It is clear that, once one has found a field $\underline{A}$ that is a vector potential for an incompressible field $\underline{u}$, one can find infinitely many others: it suffices to alter $\underline{A}$ by a gradient field. Then we ask the question: given $\underline{u} \in C^\infty(\Omega)$, with $\underline{\partial} \cdot \underline{u} = 0$ and given a vector potential $\underline{A}_0$ for $\underline{u}$ which ambiguity is left for $\underline{A}$? We still suppose that the domain Ω is simply connected.

If $\underline{A}$ and $\underline{A}_0$ are two vector potentials for $\underline{u}$ then $\underline{A} - \underline{A}_0$ is such that $\underline{\partial} \cdot (\underline{A} - \underline{A}_0) = 0$ and $\text{rot}\,(\underline{A} - \underline{A}_0) = \underline{0}$: *i.e.* there is $\varphi \in C^\infty(\Omega)$ such that

$$\underline{A} = \underline{A}_0 + \underline{\partial}\varphi \ , \quad \Delta\varphi = 0 \ . \tag{1.6.11}$$

Therefore we see that $\underline{A}$ *is determined up to the gradient of a harmonic function* φ.

If we did also require that $\underline{A}\cdot\underline{n}=0$ on $\partial\Omega$ then given a vector potential $\underline{A}_0$ for $\underline{u}$ and calling φ a solution of the *Neumann problem*

$$\Delta\varphi=0,\qquad \partial_{\underline{n}}\varphi=-\underline{A}_0\cdot\underline{n} \tag{1.6.12}$$

we realize that there exists a unique vector potential $\underline{A}$ such that

$$\underline{u}=\ \mathrm{rot}\ \underline{A}\ ,\quad \underline{\partial}\cdot\underline{A}=0\qquad \text{in }\Omega,\quad \underline{A}\cdot\underline{n}=0\qquad \text{su }\partial\Omega \tag{1.6.13}$$

and more generally one can imagine other properties (like boundary conditions or other) apt to single out a vector potential for an incompressible field among the many possible ones.

(D) A regular vector field in Ω can be represented as the sum of a rotation field and a gradient field.

Let $\underline{\xi}\to\underline{w}(\underline{\xi})$ be a vector field defined on R^3 and there of class C^∞ and rapidly decreasing at ∞. If $\underline{\hat w}(\underline{k})$ is its Fourier transform, it will be possible, uniquely, to write, for $\underline{k}\ne 0$,

$$\underline{\hat w}(\underline{k})=i\underline{k}\wedge\underline{a}(\underline{k})+\underline{k}\,f(\underline{k})\quad \text{with}\quad \underline{a}(\underline{k})\cdot\underline{k}=0 \tag{1.6.14}$$

with $\underline{a}(\underline{k})$, $f(\underline{k})$ bounded by $|\underline{k}|^{-1}$ near $\underline{k}=\underline{0}$, rapidly decreasing and of class C^∞ for $\underline{k}\ne\underline{0}$. Hence

$$\underline{w}=\ \mathrm{rot}\ \underline{A}+\underline{\partial}\varphi \tag{1.6.15}$$

Or *every vector field can be represented as a sum of a solenoidal field and a gradient field*, because the fields that have zero divergence are also called solenoidal. Note that, as in the case *(A)*, the potentials will have in general a slow decay (normally proportional to $\underline{\hat w}(\underline{0})$ and decaying almost as $O(|\underline{\xi}|^{-2})$ as $\underline{\xi}\to\infty$.

The same result also holds, therefore, to represent a field $\underline{w}\in C^\infty(\Omega)$ when Ω is a finite region. It suffices to extend $\underline{w}$ to $C^\infty(R^3)$ and apply (1.6.14), (1.6.15).

(E) The space $X_{\mathrm{rot}}(\Omega)$ and its complement in $L_2(\Omega)$.

The space $L_2(\Omega)$ admits a remarkable decomposition into a direct sum of two orthogonal subspaces, "*the rotations and the gradients*". This is not an extension of the decompositions discussed so far: indeed we shall see that also a purely solenoidal field $\underline{u}=\mathrm{rot}\,\underline{X}$ will admit a non trivial decomposition (*i.e.* a representation with a nonzero gradient component), *unless it is also tangent to the boundary of* Ω. In the last case the decomposition that we are about to describe will coincide with the previous ones.

Consider the space $X_{\mathrm{rot}}\subset L_2(\Omega)$ defined as

$$\begin{aligned}X_{\mathrm{rot}}=\{&\text{closure in }L_2(\Omega)\text{ of the fields }\underline{u}\text{ with zero divergence, }C^\infty(\Omega)\\ &\text{and zero in the vicinity of the boundary}\}\ \equiv\ \overline{X^0_{\mathrm{rot}}}\end{aligned} \tag{1.6.16}$$

where $X^0_{\rm rot}$ is the set of the C^∞ fields, with zero divergence and *vanishing* in the vicinity of the boundary $\partial\Omega$.

The space $X_{\rm rot}$ should be thought of, in a sense to be made precise below, as the set of vector fields $\underline{u}$ with zero divergence and with some component vanishing on the boundary $\partial\Omega$. One difficulty, for instance, is the fact that if $\underline{u}\in X_{\rm rot}$ then $\underline{u}$ is in $L_2(\Omega)$ but it is not necessarily differentiable[3]so that the divergence of $\underline{u}$, $\underline{\partial}\cdot\underline{u}$, can only vanish (in general) in a "weak sense"; precisely

$$\int_\Omega \underline{u}\cdot\underline{\partial} f\,d\xi\equiv 0,\qquad \forall\, f\in C^\infty(\Omega) \tag{1.6.17}$$

(as it can be seen by approximating $\underline{u}$ with $\underline{u}_n\in X^0_{\rm rot}$ in the norm of $L_2(\Omega)$, so that $\int_\Omega \underline{u}\cdot\underline{\partial} f\,d\xi=\lim_n\int_\Omega \underline{u}_n\cdot\underline{\partial} f\,d\xi=-\lim_n\int_\Omega\underline{\partial}\cdot\underline{u}_n f\;d\xi=0$, because $\underline{u}_n\cdot\underline{n}\equiv 0$ on $\partial\Omega$). And (1.6.17) shows that if $\underline{u}\in X_{\rm rot}\cap C^\infty(\Omega)$ then

$$\begin{aligned}&-\int_\Omega \underline{\partial}\cdot\underline{u}\,f d\underline{x}+\int_{\partial\Omega}\underline{u}\cdot\underline{n}\,f\,d\sigma=0\Rightarrow\\ &\Rightarrow\underline{\partial}\cdot\underline{u}=0 \text{ in }\Omega,\qquad \underline{u}\cdot\underline{n}=0\ \text{ in }\partial\Omega\end{aligned} \tag{1.6.18}$$

by the arbitrariness of$f$.

It is now convenient to introduce, to simplify notations, the following notion

Definition *(generalized derivatives and distributions): given a function $f\in L_2(\Omega)$ we say that f has a "generalized derivative" F_j, in $L_2(\Omega)$, with respect to x_j, or a derivative $\partial_j f=F_j$ "in the sense of the theory of distributions", if there is a function $F_j\in L_2(\Omega)$ for which we can write*

$$\int_\Omega f(\underline{x})\,\partial_j\varphi(\underline{x})\,d\underline{x}=-\int_\Omega F_j(\underline{x})\,\varphi(\underline{x})\,d\underline{x},\qquad \text{for all }\varphi\in C_0^\infty(\Omega) \tag{1.6.19}$$

where $C_0^\infty(\Omega)$ is the space of the functions φ of class C^∞ and which vanish in the vicinity of the boundary $\partial\Omega$ of Ω.

Analogously we can define the *divergence* in "the sense of distributions" of a fixed $\underline{u}\in L_2(\Omega)$, its *rotation*, and the higher order derivatives.

Thus the regular functions in $X_{\rm rot}$ vanish on $\partial\Omega$ only in the sense that $\underline{u}\cdot\underline{n}=0$. The others verify the (1.6.17) and therefore have $\underline{\partial}\cdot\underline{u}=0$ in the sense of distributions and $\underline{u}\cdot\underline{n}=0$ on $\partial\Omega$ in a weak sense (exactly expressed by the first of (1.6.17)).

The space of the fields $\underline{f}\in L_2(\Omega)$ in the orthogonal complement of $X_{\rm rot}$ also admits an interesting description; if $\underline{f}\in X^\perp_{\rm rot}$, it will be for every C^∞ field $\underline{A}$ vanishing near $\partial\Omega$

$$0\equiv\int\underline{f}\cdot{\rm rot}\,\underline{A}\,d\xi\ \Rightarrow\ {\rm rot}\,\underline{f}=\underline{0}\quad\text{ in the sense of distributions} \tag{1.6.20}$$

[3] Because the operation of closure in L_2 can generate functions that are not regular.

and it is possible to show also (*c.f.r.* problems [1.6.16]%[1.6.19]) that $\underline{f}$ can be written as $\underline{f} = \underline{\partial}\varphi$ with φ function in $L_2(\Omega)$ and gradient (in the sense of distributions) in $L_2(\Omega)$.

Concluding: *the most general vector field in $L_2(\Omega)$ can be written as*

$$\begin{aligned}&\underline{w} = \underline{u} + \underline{\partial}\varphi \qquad \textit{with}\\ &\underline{u} \in X_{\text{rot}}\ ,\ \varphi \in L_2(\Omega)\ \textit{and with gradient in}\ L_2(\Omega)\end{aligned} \tag{1.6.21}$$

and such decomposition is unique (because X_{rot} and $X_{\text{rot}}^{\perp}$ are a decomposition of a Hilbert space in two complementary orthogonal spaces). *Furthermore the functions in X_{rot} must be thought of as vector fields with zero divergence (in the sense of distributions) and tangent to $\partial\Omega$ (in the sense of (1.6.17)). If $\underline{u}$ is continuous in Ω up to and including the boundary then $\underline{u}$ is really tangent on $\partial\Omega$.*

One usually says: *A vector field in $L_2(\Omega)$ can be uniquely written as sum of a solenoidal vector field tangent to the boundary and of a gradient vector field.*

(F) The "gradient–solenoid" decomposition of a regular vector field.

We note that a vector field $\underline{w} \in C^\infty(\Omega)$ can be naturally written in the form

$$\underline{w} = \underline{u} + \underline{\partial}\varphi \quad \text{with} \quad \underline{u}\cdot\underline{n} = 0 \quad \text{on} \quad \partial\Omega \ \textit{and} \quad \underline{\partial}\cdot\underline{u} = 0 \quad \text{on} \quad \Omega \tag{1.6.22}$$

and the decomposition is "regular", *i.e.* one can find $\underline{u} \in C^\infty(\Omega)$ and $\varphi \in C^\infty(\Omega)$, unique in this regularity class (up to an arbitrary additive constant in φ). In fact we shall have that φ satisfies

$$\begin{aligned}\partial_n\varphi &= \underline{n}\cdot\underline{w} \qquad \text{su } \partial\Omega\\ \Delta\varphi &= \underline{\partial}\cdot\underline{w} \qquad \text{su } \Omega\end{aligned} \tag{1.6.23}$$

and this equation determines $\varphi \in C^\infty(\Omega)$ up to an additive constant. Note that (1.6.23) is a *Neumann problem* hence, in order that it be soluble, it is necessary that the compatibility condition expressed by electrostatics Gauss' theorem be satisfied: $\int_{\partial\Omega} \underline{w}\cdot\underline{n}\, d\sigma = \int_\Omega \underline{\partial}\cdot\underline{w}\, d\underline{x}$. This condition is automatically verified because of the integration theorem of Stokes. The regularity in class $C^\infty(\Omega)$ of the solution of the Neumann problem with boundary data of class $C^\infty(\partial\Omega)$ is one of the fundamental properties of the theory of the elliptic equations and one finds it discussed in many treatises (*c.f.r.* for instance [So63]) and it is summarized in the problems of §2.2.

Clearly $\underline{u} = \underline{w} - \underline{\partial}\varphi \in C^\infty(\Omega)$ satisfies (1.6.22). Hence the decomposition (1.6.22) exists and is unique (up to an additive constant in φ), hence the sense in which $\underline{u}\cdot\underline{n} = 0$ is, in such cases, literal.

Problems and complements.

[1.6.1]: Show the impossibility, in general, of a solution of (1.6.22) with $\underline{u} = 0$ on $\partial\Omega$. (*Idea:* consider a field $\underline{w}$ like

$$\begin{aligned} &w_1(x,y,z) = -\partial_y f(x,y)\chi(z) \\ &w_2(x,y,z) = \partial_x f(x,y)\chi(z), \qquad && x,y \in R^2 \\ &w_3(x,y,z) = g(x,y)\chi(z), && z \in [0,+\infty) \end{aligned}$$

where $\chi(z) \equiv 1$ for $|z| \le z_0$ and $\chi(z) \equiv 0$ for $|z| > z_1 > z_0$, and $\chi \in C^\infty(R^3)$. Then the equations (1.6.22) become

$$\begin{cases} \Delta\varphi = g(xy)\chi'(z), & \Omega = R^2 \times [0,+\infty) \\ \partial_z\varphi = g(xy), & z = 0 \end{cases}$$
$$\begin{cases} \partial_x\varphi = -\partial_y f, & z = 0 \\ \partial_y\varphi = \partial_x f \end{cases}$$

but he first pair of equations determines φ up to a constant and, therefore, the other two equations will be false for suitable choices of the *arbitrary* function f.)

[1.6.2]: Show the incompatibility of (1.6.22), with $\underline{u} = \underline{0}$ on $\partial\Omega$ in an example with bounded Ω. (*Idea:* Consider Ω = sphere of radius R, use polar coordinates and imitate the example in [1.6.1] by letting ρ play the role of z (with $\rho \ge 0$, $\rho \le R$) and to (θ,φ) the role of (x,y).)

[1.6.3]: (*extension of a function, brutal method*) If $\vartheta(x)$ is a C^∞ function which is 1 for $x \in (-\infty,+1)$ and 0 for $x \in [+2,+\infty)$, consider the function

$$x \Rightarrow f(x) = \vartheta(x) \sum_{k=0}^{\infty} \frac{c_k}{k!} x^k e^{-xc_k^2} \tag{$*$}$$

and show that it is of class C^∞ in $x \in [0,+\infty)$ *for any sequence* c_k. Furthermore

$$f^{(0)}(0) = c_0, \qquad f^{(p)}(0) = c_p + \text{ polynomial in } c_0,\ldots,c_{p-1}, \quad \forall\, p > 0$$

Show also that $e^{-xc_h^2}$ could be replaced by $e^{-x|c_h|^\alpha}$ with $\alpha > 0$ arbitrary.

[1.6.4]: Using the result in [1.6.3] show that, given an arbitrary sequence $f^{(j)}$, there is a function $f \in C^\infty([0,+\infty)]$ such that $f^{(j)}(0) = f^{(j)}$, $\forall\, j$. Furthermore, setting $F_j = \sum_{k=0}^{j} |f^{(k)}|$ there is a polynomial $L_j(x)$, explicitly computable, for which

$$||f||_{C^{(j)}([0,+\infty))} \le L_j(F_j) \tag{$**$}$$

where $||\cdot||_{C^{(j)}(a,b)} = \sup_{\substack{x\in[a,b] \\ 0\le k\le j}} |\frac{d^k f(x)}{dx^k}|$ is the "C^j–metric" on the functions on $[a,b]$.

[1.6.5]: Check that (*) and (**) imply existence of an extension f^a of f from $[-\infty,0]$ to $[-\infty,a]$, with $a > 0$, which is continuous in the "C^∞–sense (*i.e.* such that if f_n is a sequence for which $||f_n||_{C^j(-\infty,0])} \xrightarrow[n\to\infty]{} 0$ *for every* j then also $||f_n^a||_{C^j(-\infty,a])} \xrightarrow[n\to\infty]{} 0$ *for every* j).

[1.6.6] (*extensions in more dimensions*) Show that the preceding construction can be generalized to prove the extendibility of a C^∞–function, rapidly decreasing in the semi space $z \le 0$, to a C^∞–function defined on the whole space and vanishing for $z > \varepsilon$ with a prefixed ε. The extension can be made so that it is continuous in the sense C^∞ (*c.f.r.* [1.6.5]). (*Idea:* It suffices to consider the case in which $f(\underline{x},z)$ is periodic in $\underline{x}$ with arbitrary period L. In this case f can be written as: $f(\underline{x},z) = \sum_{\underline{\omega}} e^{i\underline{\omega}\cdot\underline{x}} f_{\underline{\omega}}(z)$ where

$\underline{\omega} = 2\pi L^{-1}(n_1, n_2)$ with $\underline{n}$ a bidimensional integer components vector; and we can use the theory of the one–dimensional extensions to extend *each* Fourier coefficient as

$$f_{\underline{\omega}}(z) = \sum_{k=0}^{\infty} \frac{c_k(\underline{\omega})}{k!} z^k \, e^{-z c_k(\underline{\omega})^2 e^{|\underline{\omega}|}}, \qquad \text{per} \quad z \geq 0$$

where one should note the insertion of $e^{|\underline{\omega}|}$ in the exponent, for the purpose of controlling the $\underline{x}$–derivatives, which "only" introduce powers of components of $\underline{\omega}$ in the series.)

[1.6.7]: Show that there is a family of points $\underline{x}_j$ and of C^∞–functions $\chi_j^r(\underline{x})$ vanishing outside a sphere of radius r centered at $\underline{x}_j$ such that: $\sum_j \chi_j^r(\underline{x}) \equiv 1$ and, furthermore, such functions can be chosen to be identical up to translations, *i.e.* to be such that $\chi_j^r(\underline{x}) = \chi^r(\underline{x} - \underline{x}_j)$ for a suitable χ^r. (*Idea:* Consider a pavement of R^3 by semi open cubes with side $r/4$ and centers $\underline{x}_j$; let $f_j(\underline{x})$ be the characteristic function of the cube centered at $\underline{x}_j$. Clearly $\sum_j f_j(\underline{x}) \equiv 1$; let $\gamma(\underline{x}) \geq 0$ be a C^∞–function vanishing outside a small sphere of radius $r/4$ and with integral $\int \gamma(\underline{x}) d\underline{x} = 1$. Let: $\chi_j^r(\underline{x}) = \int \gamma(\underline{x} - \underline{y}) f_j(\underline{y}) d\underline{y}$ and check that such functions have all the properties that we ask for.)

[1.6.8]: (*continuity of extensions of smooth functions*) Use the preceding problems to show that a C^∞–function f in a bounded domain Ω with a C^∞–boundary can be extended to a C^∞–function in the whole space, vanishing after a distance $\varepsilon > 0$ from Ω, with arbitrarily prefixed ε, and so that the extension is continuous in the metric of C^p (for all $p \geq 0$). (*Idea:* Using the partition of the identity of the preceding problem reduce the problem to that of the extension of a function on Ω vanishing outside a small neighborhood of a boundary point; then reduce the problem of extending the latter function to that of extending a C^∞–function defined on a semi space, discussed above.)

[1.6.9] Show that there are sequences $\{a_k\}, \{b_k\}$, $k = 0, 1, \ldots$, such that for each $n = 0, 1, \ldots$:

$$\sum_{k=0}^{\infty} |a_k||b_k|^n < \infty, \quad \sum_{k=0}^{\infty} a_k b_k^n = 1, \quad -b_k \xrightarrow[k\to\infty]{} \infty$$

(taken from [Se64]) (*Idea:* fix $b_k = -2^k$ and determine the numbers X_k^N by imposing

$$\sum_{k=0}^{N-1} X_k^N b_k^n = 1 \qquad n = 0, 1, \ldots, N-1$$

which can be solved via the Cramer rule and the well known properties of the Vandermonde matrices (and determinants) M with $M_{nk} = b_k^n$. In this way one finds $X_k^N = A_k B_{k,N}$ where

$$A_k = \prod_{j=0}^{k-1} \frac{2^j + 1}{2^j - 2^k}, \qquad B_{k,N} = \prod_{j=k+1}^{N-1} \frac{2^j + 1}{2^j - 2^k}$$

where $A_0 = 1$, $B_{N-1,N} = 1$; and then check that: $|A_k| \leq 2^{-(k^2-2k)/2}$, $B_{k,N} \leq e^4$. Furthermore $B_{k,N}$ increases and thus there is the limit $\lim_{N\to\infty} B_{k,N} = B_k$, $1 \leq B_k \leq e^4$. And setting $a_k = B_k A_k$ we see that the sequences a, b so built cheerfully enjoy the wanted properties, [Se64].)

[1.6.10]: (*linear and continuous extension of a smooth function*) Let f be a C^∞–function defined in a semi space $x \in R^d \times [-\infty, 0]$; show that if a_k, b_k denote the sequences of problem [1.6.9], setting

$$f^0(x,t) = \sum_{k=0}^{\infty} a_k \Phi(b_k t) f(x, b_k t) \qquad t > 0 \quad \text{e} \quad f^0(x,t) = f(x,t) \qquad t \leq 0$$

with $\Phi \in C^\infty(R)$, $\Phi(t) \equiv 1$ for $|t| \le 1$, and $\Phi(t) \equiv 0$ for $|t| \ge 2$, then f^0 extends f to the whole R^{d+1} *linearly and continuously* in the C^p–metric (for all $p \ge 0$).

[1.6.11] Consider a regular domain $\Omega \subset R^3$ and assume that on a sphere S containing Ω it is possible to define a C^∞ vector field $\underline{w} \ne \underline{0}$ such that: (1) $\underline{w}$ is orthogonal to ∂S, (2) every flux line of $\underline{w}$ crosses Ω in a connected piece (possibly reduced to a point or to the empty set), (3) there is a regular surface Σ such that the flux lines of $\underline{w}$ are transversal to $\Sigma \cap \Omega$ and every point of Ω is on a line that intersects $\Sigma \cap \Omega$. We shall, briefly, say that Ω *has the section property.* Show that if $\underline{u} \in C^\infty(\Omega)$ and $\underline{\partial} \cdot \underline{u} = 0$ then there is $\underline{A} \in C^\infty(\Omega)$ such that $\underline{u} = \operatorname{rot} \underline{A}$. (*Idea:* The section property is what suffices to repeat the argument in *(B)*.)

[1.6.12] Show that if $\Omega \subset R^3$ is not *simply connected by surfaces* in he sense that not every regular closed surface contained in Ω can be continuously deformed shrinking it to a point, still staying inside Ω, then *it is not in general true that a solenoidal field* $\underline{u} \in C^\infty(\Omega)$ *is the rotation of a field* $\underline{A} \in C^\infty(\Omega)$. (*Idea:* Let Ω = a convex region between two concentric spheres of radii $0 < R_1 < R_2$ and consider the field $\underline{u} = \frac{\underline{r}}{r^3}$ generated by the Coulomb potential. If it could be $\underline{u} = \operatorname{rot} \underline{A}$ for some $\underline{A} \in C^\infty(\Omega)$ one could extend $\underline{A}$ to the whole space in class C^∞ and, in particular, it would be extended to the interior of the smaller sphere. Then it would ensue (by Gauss' theorem) $0 \equiv \int_{|\underline{r}|<R_1} \underline{\partial} \cdot \underline{u} \equiv \int_{|\underline{r}|=R_1} \underline{u} \cdot \underline{n} = 4\pi\,!$ Note the generality of the above example.)

[1.6.13]: (*local extension of an incompressible field*) Let $\Omega \subset R^3$ be regular and connected. Consider the layer of width ε around $\partial\Omega$. It is possible to set up, in a regular set $\Omega_\varepsilon \supset \Omega$ with boundary at distance $> \varepsilon/2$ from that of Ω, an atlas of orthogonal coordinates (x_1, x_2, x_3) such that $\partial\Omega$ has equation $x_3 = 0$ in each chart and such that the direction 3 coincides everywhere on $\partial\Omega$, with the external normal to $\partial\Omega$ and coincides in every point of $\partial\Omega_\varepsilon$, with the external normal to $\partial\Omega_\varepsilon$. Using this coordinate system show that a $C^\infty(\Omega)$ solenoidal field can be extended to a solenoidal field in $C^\infty(\Omega_\varepsilon)$ parallel, in a neighborhood of $\partial\Omega_\varepsilon$, to the direction 3 of the coordinate system in a (arbitrary) chart. The extension can be obtained by extending u_1, u_2 arbitrarily to functions of $C^\infty(\Omega_\varepsilon)$ vanishing near $\partial\Omega_\varepsilon$ and setting

$$u_3(x_1, x_2, x_3) = u_3(x_1, x_2, 0) + \frac{1}{h_1 h_2} \int_0^{x_3} d\xi \Big(-\partial_1(h_2 h_3 u_1) - \partial_2(h_1 h_3 u_2) \Big)$$

if $h_1 dx_1^2 + h_2 dx_2^2 + h_3 dx_3^2$ is the metric of the considered system of coordinates and if the integrand functions are evaluated at the point (x_1, x_2, ξ), while those outside are evaluated at (x_1, x_2, x_3). (*Idea:* Note that the divergence of a vector field with components u_1, u_2, u_3 in an orthogonal system of coordinates is expressible as

$$\frac{1}{h_1 h_2 h_3} \Big(\partial_1(h_2 h_3 u_1) + \partial_2(h_1 h_3 u_2) + \partial_3(h_1 h_2 u_3) \Big)$$

in terms of the orthogonal metric.)

[1.6.14]: A regular simply connected domain $\Omega \subset R^3$ contained in a sphere S has the "property of the normal" if it is possible to define a $C^\infty(\overline{S/\Omega})$–field of unit vectors that extends the external normal vectors to $\partial\Omega$ to the entire domain $\overline{S/\Omega}$ ending, on the boundary ∂S, in the external normal vectors to ∂S. Note that a necessary and sufficient condition so that Ω has this property is that each point can be connected to ∞ by a continuous curve that has no other points in Ω. And note that a necessary and sufficient property for the latter property is that Ω be simply connected by surfaces, see [1.6.12].

[1.6.15] If the regular connected domain $\Omega \subset R^3$ has the property of the normal, see [1.6.14], (or is simply connected by surfaces) then every $C^\infty(\Omega)$ and solenoidal field $\underline{u}$ is the rotation of a field $\underline{A} \in C^\infty(\Omega)$. (*Idea:* The problem [1.6.13] allows us to reduce to the case in which $\underline{u}$ is orthogonal to $\partial\Omega$ and it is parallel to the normal to $\partial\Omega$ in the vicinity of $\partial\Omega$; on the other hand [1.6.14] allows us to extend $\underline{u}$ from Ω to S: it suffices to define $\underline{u}$ as parallel to the vector field connecting the external normals to $\partial\Omega$ to those of ∂S and fix the size of $\underline{u}$ such as to conserve the volume (*c.f.r.* the analogous extension discussed in *(B)* above). Hence the problem is reduced to the "standard" case in which $\underline{u}$ is in $C^\infty(S)$ and ends on ∂S normally to ∂S. The latter is a particular case of the one treated in *(B)*.)

[1.6.16]: (*representing square integrable functions as weak derivatives*) Let $\Omega \subset R^3$ be a simply connected regular region in the interior of a cube Q_1. Let $Q = \cup_{j=1}^8 Q_j$ be the cube with side L twice the previous one and union of 8 copies of Q_1; suppose that the center of Q is the origin. Let Ω_j, $j = 1, \ldots, 8$, be the "copies" of Ω in the eight cubes ($\Omega_1 \equiv \Omega$). We associate with each Q_j a sign $\sigma_j = \pm 1$ so that the sign of Q_1 is $+$ and the sign of every cube is opposite to the one of the adjacent cubes (*i.e.* with a common face). If $f \in L_2(\Omega)$ consider the extension f^e of f to $L_2(Q)$ defined by setting $f = 0$ in Q_1/Ω and then $f^e(x') = \sigma_j f(x)$ if $x' \in Q_j$ is a copy of $x \in Q_1$ (*i.e.* it has the same position relative to the sides of Q_j as x relative to the sides of Q_1). Consider Q as a torus (*i.e.* we identify its opposite sides) so that it makes sense to define the Fourier transform of a function in $L_2(Q)$. Show that the Fourier transform of f^e is a function $\hat f^e(\underline{k})$ (with $\underline{k} = \frac{2\pi}{L}\underline{n}$ for $\underline{n}$ an integer components vector) such that $\hat f^e(\underline{k}) = 0$ if $\underline{k} = (k_1, k_2, k_3)$ has a zero component (*i.e.* if $k_1k_2k_3 = 0$). Deduce that given $p \ge 0$ and $j = 1, 2, 3$ there is a $F^{(p,j)} \in L_2(\Omega)$ such that

$$f = \partial^p_{x_j} F^{(p,j)}, \qquad ||F^{(p,j)}||_{L_2(\Omega)} \le 8\Big(\frac{L}{2\pi}\Big)^p ||f||_{L_2(\Omega)} \tag{!}$$

where the derivatives are intended in the sense of distributions in $L_2(\Omega)$. If $f \in C_0^\infty(\Omega)$ then $F^{(p,j)}$ can be chosen as the restriction of a $C^\infty(Q)$–function. (*Idea:* Define $\bar F^{(p,j)} \in L_2(Q)$ via the Fourier transform of f^e: $\hat{\bar F}^{(p,j)}(\underline{k}) = (ik_j)^{-p}\hat f^e(\underline{k})$ if $k_1k_2k_3 \ne 0$ and $\hat{\bar F}^{(p,j)}(\underline{k}) = 0$ otherwise. Then restrict $\bar F^{(p,j)}(\underline{x})$ to Ω and check the wished properties.)

[1.6.17]: (*Approximating an irrotational field with a gradient*)
Let $\gamma_n(\underline{x}) = e^{(\underline{x}^2 - n^{-2})^{-1}} c_n$ for $|\underline{x}| < 1/n$ and let $\gamma_n = 0$ otherwise. The constant c_n is such that $\int \gamma_n \equiv 1$, ($c_n \propto n^3$). Let $\underline{f} \in L_2(\Omega)$ and $\operatorname{rot} \underline{f} = \underline{0}$ in the sense that for each $\underline{A} \in C_0^\infty(\Omega)$ it is $\int_\Omega \underline{f} \cdot \operatorname{rot} \underline{A} = \underline{0}$. Show, (with the notations of the problem [1.6.16] and if $*$ denotes the convolution product in the domain Q considered as a torus):

(i) $\int_Q \underline{f}^e \cdot \operatorname{rot} \underline{A} = \underline{0}$ for $\underline{A} \in C_0^\infty(\cup_{k=1}^8 \Omega_k)$
(ii) if n is large enough: $\int_Q \gamma_n * \underline{f}^e \cdot \operatorname{rot} \underline{A}^e \equiv 8 \int_\Omega \gamma_n * \underline{f}^e \cdot \operatorname{rot} \underline{A}^e \equiv \underline{0}$
(large enough means that $\frac{1}{n}$ is smaller than the distance between $\partial\Omega$ and the support of $\underline{A}$);
(iii) $\operatorname{rot}(\gamma_n * \underline{f}^e) = \underline{0}$, in $\Omega_k^{(n)} = \big\{\underline{x} | \underline{x} \in \Omega_k,\, d(\underline{x}, \partial\Omega_k) > \frac{1}{n}\big\}$
(iv) $\lim_{n\to\infty} \gamma_n * \underline{f} = \underline{f}$, in $L_2(\Omega)$.

[1.6.18]: Within the context of problems [1.6.16],[1.6.17] suppose Ω convex and let $\varphi_n(\underline{x})$ be defined for $\underline{x} \in \Omega_k^n$ in the interior of $\cup_{k=1}^8 \Omega_k^{(n)}$ by

$$\varphi_n(\underline{x}) = \int_{\Omega_k^{(n)}} \frac{d\underline{y}}{|\Omega_k^{(n)}|} \int_0^1 ds\, (\underline{x} - \underline{y}) \cdot (\gamma_n * \underline{f}^e)(\underline{y} + s(\underline{x} - \underline{y}))$$

Check that for $\underline{x} \in \cup_{k=1}^8 \Omega_k^{(n)}$

$$\partial_j \varphi_n(\underline{x}) = \gamma_n * f_j^e(\underline{x}), \qquad \int_{\Omega_k^{(n)}} \frac{d\underline{y}}{|\Omega_k^{(n)}|} |\varphi_n(\underline{x})| \le C\, ||f||_{L_2(\Omega)}$$

(*Idea:* The first relation is simply due to $\mathrm{rot}\,\gamma_n * f^e = \underline{0}$ in $\cup_{k=1}^8 \Omega_k^{(n)}$, *c.f.r.* problem [1.6.17]. The integral in the second can be bounded by

$$
\begin{aligned}
& L\sqrt{3}\int_0^1 ds \int_{\Omega_k^{(n)}} \frac{d\underline{x}d\underline{y}}{|\Omega_k^{(n)}|} |(\gamma_n * \underline{f}^e)(\underline{y} + s(\underline{x}-\underline{y}))| \le \\
& \le \frac{L\sqrt{3}}{|\Omega_k^{(n)}|}| \int_0^1 \frac{ds}{(1-s)^3} \int_{|\underline{x}-\underline{z}|<(1-s)L\sqrt{3}} d\underline{x}d\underline{z} |(\gamma_n * \underline{f}^e)(\underline{z})| \le \\
& \le C_1 L \int_{\Omega_k^{(n)}} d\underline{z} |(\gamma_n * \underline{f}^e)(\underline{z})| \le \\
& \le C_2 L^{5/2} ||\gamma_n * \underline{f}^e||_{L_2(Q)} \le C L^{5/2} ||\underline{f}||_{L_2(\Omega)}
\end{aligned}
$$

(after changing variables $\underline{y} \to \underline{z} = \underline{y} + (\underline{x} - \underline{y})s$) for suitable constants C_1, C_2, C.)

[1.6.19] (*Representing an irrotational field defined in a convex domain as a gradient*) Let $F_j^{(n)} \in L_2(Q)$ and $\varphi_n \in \cup_{k=1}^8 \Omega_k^{(n)}$ be chosen, in the context of problems [1.3.17], [1.3.18], so that

$$
\partial_j F_j^{(n)} = \gamma_n * f_j^e \text{ in } L_2(Q), \qquad ||\varphi_n||_{L_2(\Omega_k^{(n)})} \le C||f||_{L_2(Q)}
$$

Check first that the bounds obtained in problem [1.6.18] imply that in $\Omega_k^{(n)}$ it is $F_j^{(n)}(\underline{x}) = \partial_j \varphi_n(\underline{x}) + c_k^n$ with the constant c_k^n bounded by $C||f||_{L_2(\Omega)}$.
Then extend φ_n to Q by setting it equal to 0 outside $\cup_{k=1}^8 \Omega_k^{(n)}$ and let $\varphi = \lim \varphi_{n_q}$ be a weak limit in $L_2(\Omega)$ of φ_n on a subsequence $n_q \to \infty$. Show that φ admits generalized first derivatives and $\underline{\partial}\varphi \equiv \underline{f}$. (*Idea:* One remarks that if $b \in C_0^\infty(\Omega)$ it is $\int_\Omega \partial_j b \varphi = \lim_{n\to\infty} \int_\Omega \partial_j b\, \varphi_n = -\int_\Omega b \partial_j \varphi_n = -\int_\Omega b f_j$ so that $|\int \varphi \partial_j b| \le ||f||_{L_2(\Omega)}\, ||b||_{L_2(\Omega)}$.)

[1.6.20]: (*a confined solenoidal field with non closed flux lines*) Consider the vector field defined in the cylinder $0 \le x \le L$, $y^2 + z^2 \le R^2$ by

$$
u_1 = v, \qquad u_2 = -\omega z, \qquad u_3 = \omega y
$$

fixing L such that $\omega v^{-1} L/(2\pi)$ is irrational. Extend the field $\underline{u}$ to $y^2 + z^2 \le 4R^2$ making v, ω C^∞–functions of $y^2 + z^2 \equiv r^2$ vanishing in the vicinity of the external boundary of the cylinder (where $y^2 + z^2 = 4R^2$). Show that the field has zero divergence in the cylinder. Continue the field in such a way that its flux lines exit the right face of the cylinder (where $x = L$ and the other two coordinates are y, z, say) and reenter from the left side (where $x = 0$) at the point corresponding to the same coordinates y, z: this is done by defining $\underline{u}$ outside the cylinder so that the flux thus generated is incompressible (using the continuation technique discussed in *(B)*). Show that the flux lines of $\underline{u}$ exiting in L at the points of coordinates y, z such that $y^2 + z^2 < R^2$ "at every turn around" have new coordinates y', z' still inside the disk of radius R but they *never* close on themselves. Therefore *the flux lines of a confined solenoidal field are not necessarily closed*.

Bibliography: [So63], Sect. 5,6; [CF88a], Sec. 1; [Se64].

§1.7 Vorticity conservation in Euler equation. Clebsch potentials and Hamiltonian form of Euler equations. Bidimensional fluids.

(A) Thomson theorem:

A very important property for Euler fluids is the vorticity conservation law. This law is basic for the understanding (very imperfect to date) of the evolution of structures like "*smoke rings*".

Consider a not necessarily incompressible isoentropic Euler fluid; hence with a relation between density ρ and pressure p given by $\rho = R(p)$, see Sec. 1.4.3, so that the pressure potential $\Phi(p)$ is defined by $\underline{\partial}\Phi = \rho^{-1}\underline{\partial}p$. Suppose also that the external force $\underline{g}$ be conservative: $\underline{g} = -\underline{\partial}G$.

Let γ be a contour that we follow in time, $t \to \gamma(t)$, $\gamma(0) = \gamma$, where $\gamma(t)$ is the contour into which γ evolves if its points follow the current that passes through them at the initial time.

In formulae the contour $\gamma(t)$ has parametric equations $s \to \underline{\ell}(s,t)$ that can be expressed in terms of the equations $s \to \underline{\ell}(s)$ of $\gamma = \gamma(0)$ by saying that $\underline{\ell}(s,t)$ is the value of the solution at time t of

$$\begin{cases} \dot{\xi} = \underline{u}(\xi, t) \\ \xi(0) = \underline{\ell}(s) \end{cases} \tag{1.7.1}$$

Then the *Thomson theorem* holds

$$\frac{d}{dt}\int_{\gamma(t)} \underline{u}(\underline{\ell}, t) \cdot d\underline{\ell} = 0 \tag{1.7.2}$$

i.e. "vorticity" of a contour is conserved along flow lines. One says also that "*vorticity is advected by the fluid flow*".

Checking (1.7.2) is simple; the equation for $\gamma(t)$ is, up to $O(t^2)$,

$$s \to \underline{\ell}(s) + \underline{u}(\underline{\ell}(s), 0)\, t = \underline{\ell}' \tag{1.7.3}$$

so that, up to $O(t^2)$

$$d\underline{\ell}' = d\underline{\ell} + t\underset{\sim}{\partial}\underline{u}(\underline{\ell}, 0)d\underset{\sim}{\ell} + O(t^2) \tag{1.7.4}$$

and, up to $O(t^2)$

$$\begin{aligned} &\int_{\gamma(t)} \underline{u}(\underline{\ell}', t) \cdot d\underline{\ell}' = \int_\gamma (\underline{u}(\underline{\ell}, 0) + t\, \underset{\sim}{u}(\underline{\ell}, 0) \cdot \underset{\sim}{\partial}\, \underline{u}(\underline{\ell}, 0) + t\, \partial_t \underline{u}(\underline{\ell}, 0))\cdot \\ &\cdot (d\underline{\ell} + t\, \underset{\sim}{\partial}\underline{u}(\underline{\ell}, 0))\, \cdot d\underset{\sim}{\ell} = \int_\gamma \underline{u} \cdot d\underline{\ell} + t\int_\gamma (\underset{\sim}{u} \cdot \underset{\sim}{\partial}\, \underline{u}\, d\underline{\ell} + \partial_t \underline{u} d\underline{\ell} + \underline{u} \cdot \underset{\sim}{\partial}\, \underline{u}\, d\underset{\sim}{\ell}) = \\ &= \int_\gamma \underline{u} \cdot d\underline{\ell} + t\int_\gamma (-\underset{\sim}{\partial}(\Phi + G) + \underset{\sim}{\partial}\frac{\underline{u}^2}{2})\, d\underset{\sim}{\ell} = \int_\gamma \underline{u} \cdot d\underline{\ell} \end{aligned} \tag{1.7.5}$$

which gives (1.7.2). If $\underline{g}$ is not conservative the right hand side of (1.7.2) becomes $\int_\gamma \underline{g}\cdot d\underline{\ell}$.

(B) Irrotational isoentropic flows:

An immediate consequence is that, if for $t=0$ $\operatorname{rot}\underline{u}=0$ and if the external force is conservative, then $\operatorname{rot}\underline{u}$ must remain zero at all successive times: indeed if γ is an infinitesimal contour which is the boundary of a surface $d\sigma$ with normal $\underline{n}$ it will be

$$\int_\gamma \underline{u}\cdot d\,\underline{\ell} \equiv \ \operatorname{rot}\,\underline{u}\cdot\underline{n}\,d\sigma \tag{1.7.6}$$

and then the arbitrariness of $d\sigma$ and $\underline{n}$ and the invariance of vorticity just say that $\operatorname{rot}\underline{u}\equiv 0$ at all times $t>0$.

For this reason, in the case of flows which are isoentropic, or more generally such that density is a function of the pressure, "irrotational" or "potential" flows can exist and they are in fact very interesting for their simplicity that allows us to describe them (in simply connected domains) in terms of a scalar function $\varphi(\xi,t)$, called "*velocity potential*", as $\underline{u}=\underline{\partial}\varphi$.

A simple property of such flows is that, possibly changing the potential function φ by a suitable function of time alone, *it must be*

$$\frac{\partial\varphi}{\partial t}+\frac{\underline{u}^2}{2}+\Phi(p)+G=0 \tag{1.7.7}$$

where G is the potential energy of the field of the volume forces (assumed conservative) and $\Phi(p)$ is the pressure potential. For instance in the case of an incompressible fluid in a gravity field it will be $\Phi(p)+G=\frac{1}{\rho}p+gz$.

In fact from the momentum equation (*i.e.* (1.1.11)) we see that

$$\underline{\partial}\partial_t\varphi+\underset{\sim}{u}\cdot\underset{\sim}{\partial}\underline{u}+\underline{\partial}(\Phi+G)=\underline{0} \tag{1.7.8}$$

But $\underset{\sim}{u}\cdot\underset{\sim}{\partial}\underline{u}\equiv\underline{\partial}u^2/2-\underline{u}\wedge(\operatorname{rot}\,\underline{u})$ is in this case $\underline{\partial}\underline{u}^2/2$ hence (1.7.8) expresses that the gradient of (1.7.7) vanishes, *i.e.* the left hand side can only be a function of time which, by changing φ by an additive function of time alone, can be set equal to zero.

Note the relation between (1.7.8) and Bernoulli's theorem, *c.f.r.* (1.4.30): if the motion is "*static* (*i.e.* $\underline{u}$ is time independent) then φ can be chosen time independent and (1.7.8) gives

$$\frac{\underline{u}^2}{2}+\Phi(p)+G=\ \text{constant} \tag{1.7.9}$$

Hence the quantity on the left hand side of (1.7.9) which in static flows is *constant along the current lines* is, in the static irrotational cases, *constant in the whole fluid*.

It is, however, important to keep in mind that *regular* irrotational flows (*i.e.* of class C^∞) are not possible in ideal incompressible fluids in finite, even if simply connected, containers. Simply because in such cases the velocity field would be a gradient field $\underline{u} = \underline{\partial}\varphi$, with φ a suitable scalar *harmonic* function. At the boundary of the container the velocity will have to be tangent and, therefore, the normal derivative of φ will vanish. From the uniqueness of the solutions of the Neumann problem we see that the only harmonic function φ whose gradient enjoys this property is the constant function corresponding to the velocity field $\underline{u} = \underline{0}$.

Hence irrotational isoentropic (or incompressible) motions can exist, possibly, in unbounded domains or they will be singular in some point. They can be also possible in fluids with a free surface, *i.e.* which occupy a region depending on time where the boundary condition is *not* a Neumann boundary condition, *e.g.* in the case of water waves in an open container, see (F).

(C) Eulerian vorticity equation in general and in bidimensional flows:

What said until now is, strictly speaking, valid only if $d = 3$: if $d = 2$ the rotation $\mathrm{rot}\,\underline{u}$ has to be redefined.

One can, however, always imagine that a bidimensional fluid is a tridimensional fluid observed on the plane $z = 0$. It will suffice to define the velocity $\underline{u}(x, y, z) \equiv (\underline{u}(x, y), 0)$ for each z and think of the thermodynamic fields simply as defined everywhere and independent of the z coordinate. We can call this fluid the "3–dimensional extension" of the bidimensional fluid or a 3–dimensional "stratified flow", see also §1.4.

One checks easily that the three dimensional extensions of solutions of the bidimensional equations of Euler and of NS satisfy the corresponding three dimensional equations. Hence the above considerations about the vorticity hold for them.

The rotation of the bidimensional field $\underline{u} = (u_1, u_2)$ regarded as extended to three dimensions is the rotation of $(u_1, u_2, 0)$ hence it is the vector

$$(0, 0, \partial_1 u_2 - \partial_2 u_1) \tag{1.7.10}$$

Therefore we often say that *in two dimensional fluids the rotation of the velocity field* $\underline{u} = (u_1, u_2)$ *is a scalar* ζ:

$$\zeta = \partial_1 u_2 - \partial_2 u_1 \tag{1.7.11}$$

Note that a bidimensional field $\underline{u}$ with zero divergence, that we represent in $d = 3$ as $\underline{u} = \mathrm{rot}\,\underline{A}$, in $d = 2$ is represented instead as

$$\underline{u} = \underline{\partial}^\perp A, \quad A \text{ scalar }, \quad \underline{\partial}^\perp = (\partial_2, -\partial_1) \tag{1.7.12}$$

This can be seen, for instance, by going through the arguments used in the preceding section §1.6 on the representations of velocity fields.

A more analytic form of the just discussed properties can be obtained by remarking that vorticity evolves according to an equation obtained by applying the rotation operator to the momentum equation: $\partial_t \underline{u} + \underset{\sim}{u} \cdot \underset{\sim}{\partial} \underline{u} = -\underline{\partial}(\Phi + G)$. One finds, by using $\frac{1}{2}\underline{\partial}\underset{\sim}{u}^2 = \underline{u} \wedge \text{ rot } \underline{u} + \underset{\sim}{u} \cdot \underset{\sim}{\partial}\underline{u}$ and by setting $\underline{\omega} = \text{rot}\,\underline{u}$, the *Eulerian vorticity equation* for isoentropic fluids:

$$\begin{aligned} &\partial_t \text{ rot } \underline{u} - \text{ rot } (\underline{u} \wedge \text{ rot } \underline{u}) = \underline{0}, \qquad \text{or :} \\ &\partial_t \underline{\omega} + \underset{\sim}{u} \cdot \underset{\sim}{\partial}\, \underline{\omega} = \underset{\sim}{\omega} \cdot \underset{\sim}{\partial}\, \underline{u} \end{aligned} \tag{1.7.13}$$

where we used the identity $\text{rot}\,(\underline{a} \wedge \underline{b}) = \underset{\sim}{b} \cdot \underset{\sim}{\partial}\, \underline{a} - \underset{\sim}{a} \cdot \underset{\sim}{\partial}\, \underline{b}$ if $\underline{\partial} \cdot \underline{a} = \underline{\partial} \cdot \underline{b} = 0$. The (1.7.13) shows again how $\text{rot}\,\underline{u} = \underline{0}$ is possible, if true at time $t = 0$, *i.e.* shows the possibility of irrotational flows.

In the $d = 2$ incompressible case (1.7.13) simplifies considerably, the vorticity being now a scalar, and it becomes, *c.f.r.* (1.7.10), (1.7.11)

$$\partial_t \zeta + \underline{u} \cdot \underline{\partial}\, \zeta = 0 \tag{1.7.14}$$

which says that the *vorticity is conserved along the current lines*.

This gives rise, in the case $d = 2$, and for an isoentropic Euler fluid in a time independent container Ω to the existence of a first integral

$$E(\zeta) = \int_\Omega \zeta(\xi)^2 d\xi \tag{1.7.15}$$

that is called *enstrophy*. And any function f generates via

$$F(\zeta) = \int_\Omega f(\zeta(\xi))\, d\xi \tag{1.7.16}$$

a first integral for the isoentropic Euler equations.

Rather than isoentropic equations one can, of course, consider equations in which the relation between p and ρ is fixed: the only property really used has been that $\rho^{-1}\underline{\partial}\, p = \underline{\partial}\Phi$ for some Φ.

In particular we can apply the above considerations to incompressible fluids.

(D): Hamiltonian form of the incompressible Euler equations and Clebsch potentials. Normal velocity fields.

To exhibit the Hamiltonian structure of the Euler equations one can introduce the "*Clebsch potentials*", denoted (p, q), *c.f.r.* [La32], §167, p. 248, as $C^\infty(R^3)$–functions. The potentials can be used to construct the incompressible velocity field

$$\underline{u} \stackrel{def}{=} q\, \underline{\partial}\, p - \underline{\partial}\, \gamma, \qquad \gamma \stackrel{def}{=} \Delta^{-1} \underset{\sim}{\partial} \cdot (q\, \underset{\sim}{\partial}\, p) \tag{1.7.17}$$

Then we define

$$H(p,q)=\frac{1}{2}\int \underline{u}^2\,d\underline{x}\equiv\frac{1}{2}\int\big(q\underline{\partial}\,p-\underline{\partial}\Delta^{-1}\underset{\sim}{\partial}(q\underset{\sim}{\partial}\,p)\big)^2\,d\underline{x} \tag{1.7.18}$$

and, denoting by $\delta F/\partial f(x)$ the *functional derivative* of a generic functional F of f, compute the Hamiltonian equations of motion

$$\dot p=-\frac{\delta H}{\delta q(x)}=-\underline{u}\cdot\underline{\partial}p,\qquad \dot q=\frac{\delta H}{\delta p(x)}=-\underline{u}\cdot\underline{\partial}q \tag{1.7.19}$$

which are immediately checked if, in the evaluation of the functional derivatives, one uses that $\underline{u}$ has zero divergence and hence γ "does not contribute at all" to the functional equation.[1] We now compute $\dot{\underline{u}}$:

$$\begin{aligned}\dot{\underline{u}}=&\dot q\,\underline{\partial}p+q\,\underline{\partial}\,\dot p-\underline{\partial}\,\dot\gamma=\\
=&-\underset{\sim}{u}\cdot(\underset{\sim}{\partial}\,q)(\underline{\partial}p)-q\underline{\partial}(\underset{\sim}{u}\cdot\underset{\sim}{\partial}p)-\underline{\partial}\dot\gamma= \qquad (1.7.20)\\
=&-\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,(q\,\underline{\partial}\,p)+\underset{\sim}{u}\cdot(q\underset{\sim}{\partial}\,\underline{\partial}\,p)-q(\underline{\partial}\,\underset{\sim}{u})\cdot\underset{\sim}{\partial}\,p-q\underset{\sim}{u}\cdot\underline{\partial}\,\underset{\sim}{\partial}\,p-\underline{\partial}\dot\gamma\\
=&-\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\big(q\,\underline{\partial}p-\underline{\partial}\,\gamma\big)-\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{\partial}\,\gamma-\frac{1}{2}\underline{\partial}\underset{\sim}{u}^2-(\underline{\partial}\,\underset{\sim}{u})\cdot(\underset{\sim}{\partial}\,\gamma)-\partial\,\dot\gamma=\\
=&-\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{u}-\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{\partial}\,\gamma-\frac{1}{2}\underline{\partial}\underset{\sim}{u}^2-\underline{\partial}(\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\gamma)+\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{\partial}\gamma-\underline{\partial}\,\dot\gamma\\
=&-\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{u}-\underline{\partial}\big(\frac{1}{2}\underset{\sim}{u}^2+\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\gamma+\dot\gamma\big)\end{aligned}$$

Hence we see that $\underline{u}$, defined in]equ(7.17), satisfies the incompressible Euler equation and the pressure π is written

$$\begin{aligned}\frac{1}{\rho}\pi=&\frac{1}{2}\underset{\sim}{u}^2\;+\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\gamma+\dot\gamma,\qquad\text{with}\quad\gamma=\Delta^{-1}\underset{\sim}{\partial}\cdot(q\underset{\sim}{\partial}p)\quad\text{e}\\
\dot\gamma=&\Delta^{-1}\underset{\sim}{\partial}\big(-(\underline{u}\cdot\underline{\partial}q)\underset{\sim}{\partial}p-q\underset{\sim}{\partial}\big(\underline{u}\cdot\underline{\partial}p\big)\end{aligned} \tag{1.7.21}$$

Conclusion: *the incompressible Euler equations can be represented in a Hamiltonian form in which the Clebsch potentials are canonically conjugated coordinates.*

We now want to see which is the generality of the representation of an incompressible velocity field as in (1.7.17). The key remark is that, if $\underline{u}$ is given by (1.7.17) then $\underline{\omega}=\mathrm{rot}\,\underline{u}$ is

$$\underline{\omega}\equiv\mathrm{rot}\,\underline{u}=\underline{\partial}q\wedge\underline{\partial}p \tag{1.7.22}$$

[1] Indeed $\dot p(\underline{x})=-\int\underline{u}(\underline{y})\cdot(\delta(\underline{x}-\underline{y})\underline{\partial}_{\underline{y}}p(\underline{y})-\underline{\partial}_{\underline{y}}\frac{\delta\gamma}{\delta q(\underline{x})})\,d\underline{y}\equiv-\underline{u}(\underline{x})\cdot\underline{\partial}p(\underline{x})$ and, likewise, $\dot q(\underline{x})=\int\underline{u}(\underline{y})\cdot q(\underline{y})\underline{\partial}_{\underline{y}}(\delta(\underline{x}-\underline{y})-\frac{\delta\gamma}{\delta p(\underline{x})})\equiv-\underline{u}(\underline{x})\cdot\underline{\partial}q(\underline{x})$.

which shows that the flux lines of $\underline{\omega}$ are orthogonal to both $\underline{\partial} p$ and $\underline{\partial} q$ so that they are the intersections between the level surfaces of the functions q and p. This suggests the following proposition

Proposition: *A smooth incompressible velocity field with vorticity $\underline{\omega}$ never vanishing in a simply connected region Ω admits a representation in terms of Clebsch potentials if and only if there are two regular functions α, β such that the flux lines of the vorticity, inside Ω, can be written in the form: $\alpha(x,y,z) = a$, $\beta(x,y,z) = b$ as the constants a e b vary in some domain D.*

Remarks:
(1) a "domain" is here a set in which every point is an accumulation point of interior points.
2) the proposition holds for any Ω: in particular for very small regions Ω in which $\underline{\omega} \neq \underline{0}$: in such regions the vorticity lines are essentially parallel and, therefore, they satisfy the assumptions of the proposition. It follows that *locally* it is always possible to represent a *smooth* velocity field via Clebsch potentials p, q, γ (and the representation is far from unique, as we shall see).

proof: the remark in (1.7.22) shows the necessity. Viceversa if (α, β) are two functions whose level surfaces intersect along flux lines of $\underline{\omega}$ it will be

$$\underline{\omega} = \lambda\, \underline{\partial}\alpha \wedge \underline{\partial}\beta \tag{1.7.23}$$

for a suitable λ, *function of* α, β. The (1.7.23), in fact, simply expresses that the lines with tangent $\underline{\omega}$ are parallel to the intersections of the level surfaces; furthermore λ is *constant* along the curve determined by fixed values of α and β because $\underline{\omega}$ has zero divergence and hence $\underline{\partial} \cdot \underline{\omega} = \underline{\partial}\lambda \cdot \underline{\partial}\alpha \wedge \underline{\partial}\beta = 0$ *i.e.* λ has zero derivative in the direction of $\underline{\omega}$.[2]
Hence if (q, p) are defined as functions of α, β so that

$$\frac{\partial(q,p)}{\partial(\alpha,\beta)} \equiv \partial_\alpha\, q \partial_\beta\, p - \partial_\beta\, q \partial_\alpha\, p = \lambda(\alpha, \beta) \tag{1.7.24}$$

we get two potentials (q, p) which, thought of as functions of (x, y, z) (via the dependence of (α, β) from (x, y, z)), are by definition such that $\text{rot}\,(q\,\underline{\partial} p - \underline{u}) \equiv \text{rot}\,(q\,\underline{\partial}\, p) - \underline{\omega} = \underline{0}$: *i.e.* there is a γ such that $\underline{u} = q\,\underline{\partial}\, p - \underline{\partial}\,\gamma$, because Ω is simply connected.
The (1.7.24) is always soluble, and in several ways; for instance one can fix $p = \beta$ and deduce

$$p = \beta, \qquad q = \int_{\alpha_0(\beta)}^{\alpha} \lambda(\alpha', \beta) d\alpha' \tag{1.7.25}$$

[2] This can also be seen by noting that the flux of $\underline{\omega}$ through the base of the cylinder of flux lines of $\underline{\omega}$ between surfaces over which α varies by $d\alpha$ and β by $d\beta$ of is $\lambda d\alpha d\beta$ hence λ must be constant in this cylinder because $\underline{\omega}$ has zero divergence.

where $\alpha_0(\beta)$ is an arbitrary function (but obviously such that there is at least one point $(a,b)\in D$ with $a=\alpha_0(b)$): and in this way (1.7.25) give us p,q as functions of α,β and, through them, as functions of (x,y,z).

It is not difficult to realize that there are velocity fields whose vorticity *cannot be* globally described in terms of potentials q,p. A simple example is provided by the field constructed in the problem [1.6.20], regarded as a vorticity field of a suitable velocity field. Since the flux lines fill *densely* a bidimensional surface (the cylindrical surface $0\le x\le L$ and $y^2+z^2=r^2$, if $r<R$), it is clear that distinct flux lines on this surface cannot be intersections of two regular surfaces. Furthermore the example of [1.6.20] can be modified so that the flux lines fill even a three dimensional region (*c.f.r.* problems below).

We then set the following definition

Definition *(normal solenoidal field): A normal solenoidal field in Ω is an incompressible C^∞–field in Ω that admits a Clebsch representation, (1.7.17), in terms of two potentials q,p, with the function γ determined from p,q by solving a Neumann problem in Ω.*

Problem [1.6.20] shows that it is "reasonable" that the normal velocity fields are dense among all solenoidal fields, where dense has to be given a suitable meaning.

One should think, intuitively, the class of normal fields as that of the fields whose vorticity lines are either closed or begin and end at the boundary of Ω (*i.e.* at ∞ when Ω is the whole space), or that are always closed (as it is possible in the case of a parallelepiped with periodic boundary conditions).

The interest of the notion of normal field is not so much related to their density (*i.e.* to the approximability of any solenoidal field by one with flux lines closed or ending on the boundary) rather it is related on the invariance property of such fields. If the initial field is normal then it stays normal if it evolves under the Euler equations; furthermore *the evolution equations for such fields have a Hamiltonian form* provided, of course, the solution of the equation (1.7.19) is regular (of class C^∞ or, somewhat enlarging the notions above, at least of class C^2 so that all the differentiation operations considered on $\underline{\omega}=\mathrm{rot}\,\underline{u}$ be meaningful) and $\underline{\omega}$ does not vanish: this is the content of the above informal analysis (due to Clebsch and Stuart, [La32]).

Obviously it will be interesting to try writing in Hamiltonian form the evolution equations for a velocity field more general than a normal one. This is possible: however it has the "defect" that it poses the equations into Hamiltonian form in a phase space much more general than that the incompressible fields. The incompressible Eulerian motions appear immersed into a phase space in which a rather strange Hamiltonian evolution takes place *but which reduces to the incompressible Euler equations* only on the small subspace representing incompressible velocity fields.

The interest of the Clebsch-Stuart Hamiltonian representation lies in the fact that, although describing only motions of a "dense" class of velocity fields, it describes them with independent *globally defined coordinates* p, q. The alternative Hamiltonian representations of the general incompressible motion the canonical coordinates are not independent from each other: if one wants to describe an incompressible Eulerian field some constraints among them must be imposed (just for the purpose of imposing the possibility of interpreting them as coordinates describing an incompressible velocity field). See the problems for an analysis of this representation.

(E): Lagrangian form of the incompressible Euler equations, Hamiltonian form on the group of diffeomorphisms.

Clebsch's potentials are interesting because they show, for instance, that the Euler equations in 3–dimensions also admit infinitely many constants of motion: they are functions of the form (1.7.16) with ζ replaced by p or q by the (1.7.19). They involve, however, quantities p, q with a mechanical meaning that is not crystal clear and which makes their use quite difficult.
Are there more readily interpretable representations? In fact a fluid is naturally described in terms of the variables $\underline{\delta}(\underline{x})$ giving the "displacement" of the fluid points with respect to a reference position $\underline{x}$ that they occupy in a fluid configuration with constant density ρ_0.
It becomes natural to define a perfectly *incoherent* fluid as the mechanical system with Lagrangian:

$$\mathcal{L}(\dot{\underline{\delta}}, \underline{\delta}) = \int \frac{\rho_0}{2}\,\dot{\underline{\delta}}^2(\underline{x})\, d^3\underline{x} \tag{1.7.26}$$

with equations of motion determined by the action principle, *i.e.* $\rho_0 \frac{d}{dt}\dot{\underline{\delta}} = \underline{0}$, which are the equations of motion of a totally incoherent fluid.

The ideal incompressible fluid should be defined, formally and *alternatively* to the viewpoint of §1.1 as an incoherent fluid on which an ideal incompressibility constraint is imposed.
Denoting as above by $\delta F/\delta f(x)$ the functional derivative of a generic functional F of f, we note that the determinant of the Jacobian matrix $J = \partial\underline{\delta}/\partial\underline{x}$ of the transformation $\underline{x} \to \underline{\delta}(\underline{x})$ is $\det J = \underline{\partial}\delta_1 \wedge \underline{\partial}\delta_2 \cdot \underline{\partial}\delta_3$. Then the action principle applied to the Lagrangian (1.7.26) with the constraint that $\det J = 1$ leads immediately to

$$\rho_0 \frac{d}{dt}\dot{\delta}_k = \frac{\delta}{\delta\,\delta_k(\underline{x})} \int Q(\underline{x}')\,\det J(\underline{\delta}(\underline{x}'))\,d\underline{x}' \equiv -\varepsilon_{kij}\left(\underline{\partial}\delta_i \wedge \underline{\partial}\delta_j\right)\cdot \underline{\partial}Q \tag{1.7.27}$$

where Q is the Lagrange multiplier necessary to impose the constraint: it has to be determined by requiring that $\det J = 1$.
By algebra the (1.7.27) becomes

$$\rho_0 \frac{d}{dt}\dot{\underline{\delta}} = -(\det J)\, J^{-1}\underline{\partial}Q \equiv -\det J\, \frac{\partial \underset{\sim}{x}}{\partial \underline{\delta}}\, \underset{\sim}{\partial} Q \tag{1.7.28}$$

If $\underline{\delta} = \underline{\delta}(\underline{x})$, setting $\underline{u}(\underline{\delta}) = \dot{\underline{\delta}}(\underline{x})$, $\rho = \rho_0/\det J$ and $p(\underline{\delta}) = Q(\underline{x})$ and changing variables from $\underline{x}$ to $\underline{\delta}$ in (1.7.28) we get

$$\frac{d}{dt}\underline{u} = -\frac{1}{\rho}\underline{\partial} p \tag{1.7.29}$$

where now the independent spatial variable is $\underline{\delta}$ that, being dummy, can be renamed $\underline{x}$ thus leading, really, to the incompressible Euler equations because p will have to be determined by imposing incompressibility which in the new variables is $\underline{\partial} \cdot \underline{u} = 0$ (provided the initial datum $\underline{u}$ has zero divergence $\underline{\partial} \cdot \underline{u} = 0$). Hence $p = -\Delta^{-1}\underline{\partial} \cdot (\underset{\sim}{u} \cdot \underset{\sim}{\partial} \underline{u})$.

Naturally $\dot{\underline{\delta}}(\underline{x}) = \underline{u}(\underline{\delta}(\underline{x}))$ is the evolution equation for $\underline{\delta}$ which, however, *decouples* from (1.7.29) in the sense that in (1.7.29) $\underline{\delta}$ no longer appears because it is a dummy variable which can equally well be called $\underline{x}$.

From the Lagrangian form of the equations of motion one can readily derive a Hamiltonian form. It will suffice to consider the Lagrangian

$$\mathcal{L}_i(\dot{\underline{\delta}}, \underline{\delta}) = \int \frac{\rho_0}{2}\, \dot{\underline{\delta}}(\underline{x})^2\, d^3\underline{x} - \int [\Delta^{-1}\underline{\partial}(\underset{\sim}{u} \cdot \underset{\sim}{\partial}\underline{u})]\,(\det J(\underline{\delta}) - 1)\, d\underline{x} \tag{1.7.30}$$

where $J(\underline{\delta}) = \partial\underline{\delta}/\partial\underline{x}$ and, furthermore, the expression in square brackets has to be computed *by thinking of* $\underline{u}$ *as a function of* $\underline{\delta}$ *(i.e.* $\underline{u}(\underline{\delta}) = \dot{\underline{\delta}}(\underline{x}(\underline{\delta}))$ with $\underline{x}(\underline{\delta})$ inverse of $\underline{\delta}(\underline{x})$*) and imagining that the differentiation operators and* Δ *act on the variables* $\underline{\delta}$ *as independent variables which eventually have to be set equal to* $\underline{\delta}(\underline{x})$*.*

In other words we recognize in the expression in square brackets the value of the multiplier introduced in (1.7.27), which turns the just made statements into an immediate formal consequence of the above Lagrangian form of the Euler equation.

The resulting equations for $\underline{p}(\underline{x}) = \delta\mathcal{L}_i/\partial\dot{\underline{\delta}}(\underline{x})$ e $\underline{q}(\underline{x}) = \underline{\delta}(\underline{x})$ are Hamiltonian equations for which, as a consequence of the preceding analysis, "incompressible" initial data, *i.e.* initial data with a zero divergence $\underline{u}$ and $\det J = 1$, evolve remaining incompressible. Hence the Euler equations describe a particular class of motions of a Hamiltonian system, which generates an evolution on the family of maps $\underline{x} \to \underline{\delta}(\underline{x})$ of R^3 into itself: precisely the class of the ideal incompressible motions.

Equation (1.7.30) implies that *for "incompressible data"* we shall have $\underline{p}(\underline{x}) = \rho_0\dot{\underline{\delta}}(\underline{x}) = \rho_0\underline{u}(\underline{\delta}(\underline{x}))$ and, obviously, $\underline{q}(\underline{x}) = \underline{\delta}(\underline{x})$, while if the datum $\underline{u}$ is not incompressible the evolution generated by (1.7.30) *is not* the eulerian evolution: *i.e.* the equations of motion of (1.7.30) are correct (namely are equivalent to Euler's) only for the incompressible motions.

Proceeding formally (along a path that it is an open problem to make mathematically rigorous) one can finally check that the "surface" $\mathcal{S}$ of the incompressible transformations of R^3 is such that the restriction to the phase space $\mathcal{S}\times\mathcal{T}(\mathcal{S})$ (*i.e.* the space of the points in $\mathcal{S}$ and the cotangent vectors $\dot{\underline{\delta}}$ to it) of such Hamiltonian motions is itself a family of Hamiltonian motions on $\mathcal{S}\times\mathcal{T}(\mathcal{S})$. From (1.7.30), we see that the Hamiltonian will be quadratic in $\underline{p}$ (because the (1.7.30) is quadratic in $\dot{\underline{\delta}}$ and, *on the surface* $\mathcal{S}$, $\underline{p}=\rho_0\dot{\underline{\delta}}(\underline{\delta})$) *i.e.* $H(\underline{p},\underline{q})=\frac{1}{2}(\underline{p},G(\underline{q})\underline{p})$ with G a suitable operator on the space of the incompressible transformations.

The surface $\mathcal{S}$ can, finally, be seen as a *group of transformations* (the group of diffeomorphisms of R^3 into itself which preserve the volume) and Arnold has shown that G can then be interpreted as a metric on the group $\mathcal{S}$ which is (left)–invariant and, therefore, the Eulerian motions can be thought of as *geodesic motions* on $\mathcal{S}$, [Ar79].

(F): Hamiltonian form of the Eulerian potential flows; example in 2*–dimensions.*

It is difficult to find concrete applications of the Clebsch potentials or of the Lagrangian and Hamiltonian interpretations of Euler equations given in (E). Nevertheless in special cases the Hamiltonian nature of the Euler equations emerges naturally in other forms and it can become quite useful for the applications.

Furthermore there is an important and vast class of problems that evidently cannot be studied via Clebsch potentials: namely the irrotational incompressible flows in "free boundary" domains representing motions of a fluid in contact with a gas under a gravity field.

As an example we describe a very remarkable application to 2*–dimensional fluid motions. It is the theory of vorticity-less waves, i.e. of potential flows, in* $d=2$.

Consider the plane x,z and on it a fluid whose free surface has equation $z=\zeta(x)$. This means also that we consider only "*non breaking waves*". Let $\underline{u}$ be a velocity field describing the fluid, which we suppose ideal and incompressible. The system evolves via Euler equations and in a (conservative) force field of intensity g oriented towards $z<0$: *gravity.*

The first relation that we shall look for is the *boundary condition* on the free surface of the fluid. It is clear that if $\underline{u}(x,\zeta(x,t),t)=\underline{u}$ is the velocity of a point that at time t is in $(x,\zeta(x,t))=(x,\zeta)$ then at the instant $t+dt$ it will be in $(x+u_xdt,\zeta+u_zdt)$ hence in order that the fluid remains connected (*i.e.* no holes develop in it) it is necessary that this point coincides with $\zeta(x+u_xdt,t+dt)$, up to $O(dt^2)$.

Hence, denoting $\dot{\zeta}=\partial_t\zeta,\zeta'=\partial_x\zeta$, we obtain the boundary condition $-\partial_x\zeta\,u_x+u_z=\partial_t\zeta$ in $(x,\zeta(x,t))$; the latter relation has to be coupled with the condition that the pressure be constant, *e.g.* $p=0$, at the free boundary. If $\underline{n}(x)\stackrel{def}{=}(-\zeta'(x),1)/(1+\zeta'(x)^2)^{1/2}$ is the external normal to the set of the

(x,z) with $z<\zeta(x)$, the two conditions can be written as

$$\begin{aligned}&(\underline{n}\cdot\underline{\partial}\,\varphi)(x,\zeta(x))=\frac{\partial_t\zeta(x)}{(1+\zeta'(x)^2)^{1/2}}\\&\partial_t\varphi+g\zeta+\frac12(\underline{\partial}\varphi)^2=0\end{aligned}\tag{1.7.31}$$

respectively, *c.f.r.* (1.7.7). The relations show that the two functions $(\zeta(x),\dot\zeta(x))$ uniquely determine the velocity field as well as their values at time $t+dt$.

In fact given $\zeta,\dot\zeta$ we can compute the velocity field $\underline{u}=\underline{\partial}\varphi$ by determining φ as the harmonic function satisfying $\Delta\varphi=0$ in $D=\{(x,z)|\,x\in[0,L],z\in(-\infty,\zeta(x))\}$ with boundary condition $\partial_n\varphi$ given by the first of (1.7.31): a Neumann problem; furthermore from the second relation we can compute $\partial_t\varphi(x,z)$ by solving a Dirichlet problem in D with boundary condition $(\partial_t\varphi)(x,\zeta(x))$. The latter field can be used to compute $\ddot\zeta(x)$ by differentiating the boundary condition rewritten in the form $\partial_t\zeta=-\partial_x\zeta\,u_x+u_z$.

We can now put the equations into Hamiltonian form. We expect that, the motion being "ideal", the equations of motion must follow from a naive aplication of the action principle. We shall restrict the analysis to motions which are horizontally periodic over a length L, for simplicity. Since we have seen that the motions are naturally described in the coordinates $\zeta(x),\dot\zeta(x)$ we consider the action functional of $\zeta,\dot\zeta$ as the difference betwen the kinetic energy of the fluid and its potential energy in the gravity field.

Hence the domain D in which we shall consider the fluid is $D=\{(x,z)|\,x\in[0,L],z\in(-\infty,\zeta(x))\}$. In D the velocity field will be $\underline{u}=\underline{\partial}\varphi$ and its time derivative will be $\underline{\partial}\psi$ with φ harmonic, solutions of

$$\Delta\varphi=0\qquad\text{in } D,\qquad \underline{n}\cdot\underline{\partial}\varphi=\frac{\dot\zeta}{(1+\zeta'(x))^{1/2}}\qquad\text{in }\partial D\tag{1.7.32}$$

The potential energy will obviously be infinite, being $\int_0^L\int_{-\infty}^{\zeta(x)}\rho\,g\,z\,dxdz$. However what counts are the energy variations which are formally the same as those of $\rho\int_m^L dx\frac12 g\zeta(x)^2$ where m is any quantity less than the minimum of $\zeta(x)$.

The kinetic energy will be $\frac{\rho}{2}\int_0^L dx\int_{-\infty}^{\zeta(x)}dz\,|\underline{\partial}\varphi(x,z)|^2$. Thus the action of the system, *i.e.* the difference between kinetic and potential energy is, if $ds\overset{def}{=}(1+\zeta'(x)^2)^{1/2}dx$ and if we introduce the auxiliary function $\psi(x)\overset{def}{=}\varphi(x,\zeta(x),$

$$\begin{aligned}\mathcal{L}=&\frac{\rho}{2}\int_0^L\int_{-\infty}^{\zeta(x)}(\underline{\partial}\varphi)^2\,dxdz-\frac{\rho}{2}\int_0^L g\zeta(x)^2dx=\\=&\frac{\rho}{2}\int_0^L(\psi(x)\frac{\dot\zeta(x)}{(1+\zeta'(x))^{1/2}}\,ds-g\zeta(x)^2\,dx)\end{aligned}\tag{1.7.33}$$

after integrating by parts the first integral, defining if $\psi(x)\overset{def}{=}\varphi(x,\zeta(x)$ and making use of the first relation (1.7.31), (1.7.32).

The linearity of (1.7.32) shows that ψ is linear in $\dot\zeta(x)/(1+\zeta'(x))^{1/2}$: *i.e.* $\psi = M\dot\zeta/(1+(\zeta')^2)^{1/2}$ for some linear operator M (which depends on $\zeta(x)$, *i.e.* on the shape of D).

In general if φ_1,φ_2 are two harmonic functions in D with respective "Dirichlet value" φ_1,φ_2 on the boundary, ∂D, and respective "Neumann values" on the same boundary $\sigma_1 = \underline{n}\cdot\underline{\partial}\varphi_1, \sigma_2 = \underline{n}\cdot\underline{\partial}\varphi_2$ then, (on the boundary) by definition, $\varphi_i = M\sigma_i$ for a suitable linear operator M and $\int_{\partial D} ds\,\big((M\sigma_1)\,\sigma_2 - \sigma_1\,(M\sigma_2)\big) = \int_D (\varphi_1\Delta\varphi_2 - \varphi_2\Delta\varphi_1)\,dxdz = 0$ so that the operator M is *symmetric* on $L_2(ds,[0,L])$. Note that the operator M transforms, by definition and by (1.7.34), a Neumann boundary datum for a harmonic function into the corresponding Dirichlet datum. Hence $\mathcal{L}$ becomes

$$\mathcal{L} = \frac{\rho}{2}\int_0^L \big(M(\frac{\dot\zeta}{(1+\zeta')^{1/2}})\frac{\dot\zeta}{(1+\zeta')^{1/2}}\,ds - g\zeta^2\,dx\big) \tag{1.7.34}$$

Therefore we can write the equations of motion in terms of the Hamiltonian corresponding to the Lagrangian $\mathcal{L}$. Note that $\delta\mathcal{L}/\delta\dot\zeta(x) = M(\dot\zeta(x)/(1+(\zeta')^2)^{1/2})$, where we denote by $\delta/\delta\varphi(x)$ a generic functional derivative and we recall that $(1+\zeta'(x)^2)^{-1/2}ds \equiv dx$: therefore we can define the variable $\pi(x)$ conjugate to $\zeta(x)$ as $\pi(x) \stackrel{def}{=} \rho M\dot\zeta/(1+(\zeta')^2)^{1/2} = \rho\psi(x)$ and the Hamiltonian in the conjugate variables π,ζ is

$$\mathcal{H} = \frac{1}{2}\int_0^L \Big(\frac{1}{\rho}\pi(x)G\pi(x) + \rho g\zeta(x)^2\Big)\,dx \tag{1.7.35}$$

where if Γ^D is the operator that solves the Dirichlet problem in D we set $G(\zeta)\pi(x) = \rho\,\big(-\partial_x\zeta(x)\partial_x\big(\Gamma^D\pi\big)(x,z) + \partial_z\big(\Gamma^D\pi\big)(x,z)\big)_{z=\zeta(x)}\big)$.

The check that the equations of motion generated by (1.7.35) are correct, *i.e.* coincide with (1.7.31) proceeds as follows. The first Hamilton equation is just $\dot\zeta = \frac{1}{\rho}G\pi$ which coincides with the first of (1.7.31). More interesting is the computation of the functional derivative $-\delta\mathcal{H}/\delta\zeta(x)$. It is

$$\begin{aligned}
&-g\rho\zeta(x) - \rho\frac{\delta}{\delta\zeta(x)}\int_0^L dy\int_{-\infty}^{\zeta(y)} dz\,\frac{1}{2}(\underline{\partial}\varphi)^2(y,z) =\\
&= -g\rho\zeta(x) - \frac{\rho}{2}(\underline{\partial}\varphi)^2(x,\zeta(x)) -\\
&-\int_D \rho\underline{n}\cdot\underline{\partial}\varphi(y,\zeta(y))\frac{\delta}{\delta\zeta(x)}(\Gamma^D\frac{\pi}{\rho})(y,z)\big|_{z=\zeta(y)}ds
\end{aligned} \tag{1.7.36}$$

where $ds = (1+\zeta'(y)^2)^{1/2}\,dy$. The last functional derivative can be studied by remarking that by its definition $(\Gamma_D\pi)(x,z)|_{z=\zeta(x)} \equiv \pi(x)$ hence near $z=0$ we get

$$\frac{1}{\rho}\Gamma^D\pi(y,z) = \frac{1}{\rho}\,\varphi(y,\zeta(y)) + (z-\zeta(y))\,\frac{1}{\rho}\,\partial_z(\Gamma^D\pi)(y,\zeta(x)) +$$

$$+ O((z-\zeta(x))^2) = \frac{1}{\rho}\pi(y) + (z-\zeta(y))u_z(y) + O((z-\zeta(x))^2) \Rightarrow$$
$$\Rightarrow \frac{1}{\rho}\frac{\delta(\Gamma^D\pi)(y)}{\delta\zeta(x)} = -\delta(x-y)\,u_z(y) \tag{1.7.37}$$

Since $\rho\,\underline{n}\cdot\underline{\partial}\varphi(y,\zeta(y)ds = \rho\dot\zeta(y)dy$ this means that

$$-\frac{\delta\mathcal{H}}{\delta\zeta(x)} = -\rho g\zeta(x) - \frac{\rho}{2}(\underline{\partial}\varphi)^2(x) + \rho\dot\zeta(x)u_z(x) \tag{1.7.38}$$

Furthermore $\dot\pi = \rho\,\frac{d}{dt}(\varphi(x,\zeta(x)) = \rho\,(\partial_t\varphi(x,\zeta(x)) + u_z(x)\dot\zeta(x))$ (because $u_z = \partial_z\varphi$). The latter relation combined with (1.7.38) for $-\delta\mathcal{H}/\delta\zeta(x)$ gives the pressure condition on the free surface.

The above analysis is taken from [DZ94],[DLZ95],[CW95].

(G) Small waves.

The Hamilton equations in (F) considerably simplify in the case of "small waves". Indeed non linear terms will be neglected and the domain D occupied by the fluid will be considered to be the half–space $z<0$. In this approximation M,G take the value that they have for $\zeta\equiv 0$ and admit a simple and well known exact expression. See the problems for a more quantitative analysis of this approximation.

It will then be possible to develop π,ζ in Fourier modes

$$\begin{aligned}\pi(x) =& \frac{1}{\sqrt{2L}}\sum_k p_k e^{ikx}, \qquad \zeta(x) = \frac{1}{\sqrt{2L}}\sum_k q_k e^{ikx}\\ \varphi(x,z) =& \frac{1}{\rho}\frac{1}{\sqrt{2L}}\sum_k p_k e^{ikx}e^{|k|z}\end{aligned} \tag{1.7.39}$$

where k is an integer multiple of $2\pi/L$, and the Hamiltonian can be written in terms of the Fourier coefficients

$$H(p,q) = \frac{1}{2}\sum_k(\frac{|k|}{2\rho}|p_k|^2 + \frac{\rho g}{2}|q_k|^2) \tag{1.7.40}$$

In (1.7.40) there are redundant variables because $p_k = \overline{p_{-k}}$ and $q_k = \overline{q_{-k}}$. Therefore if we decompose p_k, q_k into real and imaginary parts, writing $p_k = p_{1k} + ip_{2k}$ and $q_k = q_{1k} + iq_{2k}$, we find

$$H(p,q) = \frac{\rho g}{4}q_0^2 + \sum_{k>0}\sum_{j=1,2}\frac{1}{2}(\frac{k}{\rho}p_{jk}^2 + \rho g q_{jk}^2) \tag{1.7.41}$$

Recalling that the Fourier transform is a canonical transformation this shows that the problem of the small waves, spatially periodic horizontally with period L in a deep bidimensional fluid is equivalent to studying infinitely many harmonic oscillators: *hence it is exactly soluble.*

The normal modes, *i.e.* the action–angle coordinates, of the problem are naturally given by the Fourier transform and the mode with $k = 0$ is immobile: *i.e.* $q_0 = constant$ $p_0 = constant$. The second relation is evidently trivial because φ, hence also π, are defined up to an additive constant; the first, instead, has the simple interpretation that the average level (*i.e.* $q_0 = \int_0^L \zeta(x)\,dx$) remains constant: *i.e.* the constancy of q_0 expresses mass conservation.

It is interesting to study the velocity fields corresponding to the normal modes of oscillation that we have just derived: see the following problems that also illustrate a more classical approach to linear waves, [LL71].

Problems: *Sound and surface waves. Radiated energy.*

[1.7.0] Think of a 2–dimensional fluid as a stratified 3–dimensional one (see (C) above) and define vorticity as the field perpendicular to the fluid motion and with intensity $\omega = \partial_z u_x - \partial_x u_z$ (see (C) above). Find an expression for the velocity field in terms of Clebsch potentials. (*Idea:* Vorticity lines are straight lines perpendicular to the plane x, z. Hence proceeding as in (D) above we can choose as functions α, β, see (1.7.23), the functions $\alpha = z, \beta = x$, for instance. Then the function $\lambda(z, x)$ in (1.7.24) is precisely: $\lambda(z, x) = \omega(x, z)$ and, by (1.7.25)

$$p = x, \quad q(x, z) = \int_{-\infty}^{z} dz' \omega(x, z')$$

thus the function γ will be $\gamma = \Delta^{-1}(x\Delta q + \partial_x q)$, where Δ^{-1} is the inverse of the Laplace operator in the domain occupied by the fluid.)

[1.7.1]: (*sound waves in a Euler fluid*) Consider a compressible adiabatic Euler fluid, with zero velocity, pressure p_0 and temperature T_0. Imagine to perturb the initial state by a small irrotational perturbation $\underline{u}' = \underline{\partial}\varphi$, vanishing at ∞. Suppose that the motion is adiabatic and that $\underline{u}'$ as well as the variations p' of the pressure and ρ' of the density are small, neglecting their squares. Show that sound waves are generated in the fluid and compute their propagation speed. (*Idea:* Write $\rho = \rho_0 + \rho', p = p_0 + p'$ and check that

$$\partial_t \rho' + \rho_0 \underline{\partial} \cdot \underline{u}' = 0, \qquad \partial_t \underline{u}' + \rho_0^{-1} \underline{\partial} p' = 0, \qquad p' = (\frac{\partial p}{\partial \rho})_s \rho'$$

and, furthermore, $\underline{u}' = \underline{\partial}\varphi$ for a suitable φ; and the second relation implies $\underline{\partial}(\partial_t\varphi + \rho_0^{-1}p') = 0$. Hence $\partial_t\varphi = -\rho_0^{-1}p'$ because at ∞ one has $p' = 0$ and $\underline{u}' = \underline{0}$. So that

$$\partial_t^2 \varphi = -\frac{1}{\rho_0}\partial_t p' = -\frac{1}{\rho_0}\big(\frac{\partial p}{\partial \rho}\big)_s \partial_t \rho' = \big(\frac{\partial p}{\partial \rho}\big)_s \Delta\varphi$$

therefore φ and, by linearity, $\partial_t\varphi$ evolve according to the wave equation with speed $c = \big(\frac{\partial p}{\partial \rho}\big)_s^{1/2}$.)

[1.7.2]: (*free surface boundary condition*) Consider a non viscous incompressible bidimensional fluid, filling the region of the plane (x, z) defined by $z \le \zeta(x, t)$, where $\zeta(x, t)$ is the *free surface* at time t. Show that the boundary condition relating $\underline{u}(\underline{x}, t)$ to $\zeta(x, t)$ is

$$\Big(1 + (\partial_x\zeta)^2\Big)^{1/2} \underline{n} \cdot \underline{u} = \partial_t \zeta$$

if $\underline{n} = \Big(1 + (\partial_x\zeta)^2\Big)^{-1/2}(-\partial_x\zeta, 1)$ is the external normal to the free surface. Show also that the boundary condition has the meaning that the fluid elements that are on the surface *remain there* forever: a necessary consequence of the hypothesis that the displacement of the fluid elements due to the fluid motion is a regular transformation of the domain occupied by the fluid. (*Idea:* In a regular transformation the interior points remain interior.)

[1.7.3]: (*linearity conditions for motions of a free surface*) In the context of [1.7.2] suppose that the fluid evolves with a regular motion in which the surface waves have amplitude a which evolves sensibly over a time scale τ, and suppose that the variations of $\underline{u}$ at fixed time are sensible over a length scale $\lambda \equiv k^{-1}$. Show that the condition under which the transport term (also called inertial term) $\underset{\sim}{u} \cdot \underset{\sim}{\partial} \underline{u}$ in Euler equation can be neglected is that $a \ll \lambda$, *i.e.* $ka \ll 1$. (*Idea:* Proceed as in the analysis of §1.2. The order of magnitude of $\partial_t \underline{u}$ is $a\tau^{-2}$, that of $\underline{u}$ is $a\tau^{-1}$ and that of $\underset{\sim}{u} \cdot \underset{\sim}{\partial} \underline{u}$ is $a^2\tau^{-2}\lambda^{-1}$.)

[1.7.4]: (*linearized Euler equations for motions of a free surface*) In the context of problem [1.7.2] and assuming the inertial term, discussed in [1.7.3], to be negligible consider an irrotational motion: $\underline{u} = \underline{\partial}\varphi$ with pressure at the free surface $p_0 = 0$ constant. Show that Euler equations become

$$\partial_t \varphi + \rho^{-1} p + gz = 0, \quad \Delta\varphi = 0, \quad u_z(x, \zeta(x,t)) = \partial_t \zeta(x,t), \quad p(x,\zeta) = 0$$

(*Idea:* The first is (1.7.7), with $\frac{1}{2}\underline{u}^2$ neglected, *c.f.r.* [1.7.3]; the second is the zero divergence condition for $\underline{u} = \underline{\partial}\varphi$; the third is the boundary condition in [1.7.2] if one neglects the terms of order $O((a/\lambda)^2)$; the fourth expresses that the pressure at the free surface is constant.)

[1.7.5]: (*wave solutions for motions of a free surface*) In the context of [1.7.4] find a solution with $\zeta(x,t) = a\cos(kx - \omega t)$ showing that its existence follows immediately from the analysis in (G) assuming $\omega^2 = kg$; hence surface waves with phase velocity $\sqrt{gk}/k$ and group velocity $\frac{1}{2}\sqrt{g/k}$ are possible. (*Idea:* At $z = \zeta(x,t)$ one must have , *c.f.r.* [1.7.4], $\partial_t \varphi = -ag\cos(kx - \omega t)$; hence $\varphi(x,z) = \frac{ag}{\omega} e^{kz} \sin(kx - \omega t)$ solves the first two equations of [1.7.4], and the $u_z = \partial_t \zeta$ yields $\frac{gk}{\omega} = \omega$.)

[1.7.6]: (*current lines equations in free surface waves*) Write the equations of the current lines of the motion found in [1.7.5]. (*Idea:*

$$\dot{x} = agk\omega^{-1} e^{kz} \cos(kx - \omega t), \quad \dot{z} = agk\omega^{-1} e^{kz} \sin(kx - \omega t).)$$

[1.7.7]: (*circles and current lines in free surface waves*)From the equations deduced in [1.7.6] deduce that if $|z|$ is large then, *approximately*, the current lines with an average position (x_0, z_0) are circles of radius $agk\omega^{-2} e^{kz_0}$ around (x_0, z_0).

[1.7.8]: (*shallow "water" waves*) Show that if the fluid, rather than occupying the region $z \le \zeta(x,t)$, occupies a region $-h \le z \le \zeta(x,t)$ and if $h \gg \lambda \gg a$ then the equations derived in [1.7.4] are simply modified by adding the further boundary condition $\partial_z \varphi|_{z=-h} = 0$. Evaluate the speed of phase propagation analogous to the one in [1.7.5], showing that $\omega^2 = gk \tanh kh$. Show that also this problem can be put into a Hamiltonian form and leads to an integrable Hamiltonian system (of harmonic oscillators); calculate the normal modes.

[1.7.9]: (*velocity in shallow "water" waves*) Show, in the context of [1.7.8], that if $\lambda \gg h$ the phase and group velocity coincide and have value $\sqrt{gh}$.

[1.7.10]: What would change if in the preceding problems we considered the fluid 3–dimensional? (*Idea:* "nothing".)

[1.7.11]: Consider in the context of problem [1.7.1] an irrotational 2–dimensional infinite fluid and consider a plane wave with speed c. Show that the velocity v at a given point and at a given time is related to the density and pressure variations ρ' and p' by $p' = \rho v c$ e $p' = c^2\rho'$. (*Idea:* The second relation follows as explained in the hint for [1.7.2]; and if $\varphi = f(x/c - t)$ is the velocity potential of a plane wave in the direction x one has $v = \partial_x\varphi = c^{-1}f'(x/c - t)$, while $p' = -\rho\partial_t\varphi$, always from the hint for [1.7.1]: hence $p' = -\rho f'(x/c - t)$.)

[1.7.12]: (*energy flux through a surface orthogonal to a plane wave in a fluid*) Note that in a plane wave the average energy flux through a surface with normal $\underline{n}$ parallel to the direction of the wave velocity $\underline{v}$ is $\rho_0 c\,\overline{v^2}$, where the bar denotes time average. (*Idea:* The energy flux is, *c.f.r.* [1.1.6], $\rho\underline{v}(\varepsilon + \frac{\underline{v}^2}{2} + \frac{p}{\rho})$; neglecting $\underline{v}v^2$ one has $\rho\underline{v}(\varepsilon + \frac{p}{\rho}) = w_0\rho\underline{v} + \rho w'\underline{v}$ where $w_0 = \varepsilon_0 + p_0/\rho_0$, and $w' = \left(\frac{\partial w}{\partial p}\right)_s p'$. But w is the enthalpy per unit mass, so that $\left(\frac{\partial w}{\partial p}\right)_s = 1/\rho_0$; hence $\overline{\rho\underline{v}(\varepsilon + p/\rho)} = w_0\,\overline{\rho\underline{v}} + \overline{p'\underline{v}}$. The first average vanishes because $\overline{\rho\underline{v}}$ is the average variation of fluid mass to the left of the surface, while [1.7.11] implies $p' = \rho v c$ hence $\overline{p'v} = \rho c\,\overline{v^2}$.)

[1.7.13]: (*energy emitted by small oscillations or deformations of a large body in an infinite fluid*) In the context [1.7.1] suppose that a body Γ of arbitrary form and linear dimension l is immersed in the fluid. Imagine that the body oscillates either by varying the volume (homotetically) or by displacing the center of mass along the z axis. Let $\omega/2\pi$ be the oscillations and define $\lambda = 2\pi c/\omega$, with c equal to the sound speed.
Supposing that $\lambda \ll l$ and also that such oscillations have a small enough amplitude compared to l we can assume that every surface element of the body emits a plane wave and that the velocity of the point $\underline{x} \in \partial\Gamma$, where the external normal is $\underline{n}$, is $\underline{u}$. Show that the energy emitted per unit time is $\rho_0 c\int_{\partial\Gamma}(\underline{u}\cdot\underline{n})^2\,d\sigma$. (*Idea:* If $\lambda \ll l$ every wave goes its own way and the result follows immediately from [1.7.12].)

[1.7.14]: (*energy emitted by small oscillations or deformations of a small body in an infinite fluid*) In the context of [1.7.13] suppose that $\lambda \gg l$ and that the volume $t \to V(t)$ of the body varies as a given function of time; show that the energy I emitted per unit time is

$$I = \frac{\rho}{4\pi c}\,\overline{(\partial_t^2 V)^2}$$

(*Idea:* In the region $l \ll r \ll \lambda$ one can suppose that $\Delta\varphi = 0$ (because $c^{-2}\partial_t^2\varphi \sim O(\omega^2 c^{-2}\varphi) \sim O(\lambda^{-2}\varphi)$ while the order of magnitude of $\Delta\varphi$ should be much larger being $O(r^{-2}\varphi)$); then for a suitable function a of t we have $\varphi = \frac{a}{r} + O(\frac{1}{r^2})$ and incompressibility demands that the flux through the surface of a (large compared to l) sphere is such that $\dot V = 4\pi a$. Hence $\varphi = -\frac{\dot V(t)}{4\pi r}$. At distance $r \gg \lambda$ it will be, instead, a solution of the wave equation *i.e.* (in the dominant spherical term) $\varphi = -\frac{f(t - r/c)}{4\pi r}$ (for some f because φ must be a solution of the wave equation with speed c): it follows that the behavior for r in both regions can be interpolated by the single expression $\underline{u} = \underline{\partial}\varphi = \frac{(\partial_t^2 v)(t - r/c)\underline{n}}{4\pi c r}$. And integrating by parts on a surface of radius r the result follows.)

[1.7.15]: (*energy emitted by a small rigid body vibrating in a fluid*) Suppose that the body of volume V is rigid, hence $V = const$, and that it moves along the z–axis with an oscillatory motion and pulsation ω. Then the velocity potential will not have ar^{-1} as leading term for r large (although small relative to λ, *i.e.* in the region where the potential can be considered harmonic) because (as in [1.7.14]) this would mean that $\dot V \neq 0$) and its leading behavior will rather be the next (dipole) term $\varphi = \underline{A}(t)\cdot\underline{\partial}\frac{1}{r}$, where $\underline{A}$ is (by symmetry) parallel to the axis z. Noting that $\underline{\partial}\left(\frac{\underline{A}(t - r/c)}{r}\right)$ solves the wave equation, the potential φ can be interpolated for all r ($r \gg l$) by $\varphi \simeq -\frac{\dot{\underline{A}}(t - r/c)\cdot\underline{n}}{cr}$. Show that the velocity at large distance and the irradiated energy per unit time are

$$\underline{v} \simeq \underline{n}\,\underline{n}\cdot\partial_t^2\underline{A}/c^2 r, \quad I = 4\pi\rho\,\overline{(\partial_t^2\underline{A})^2}/3c^3$$

[1.7.16]: (*energy emitted by a small rigid body slowly and harmonically vibrating in a fluid*) Suppose that the body is a sphere of radius R moving as $z(t) = L\cos\omega t$, $L \ll c/\omega$. Show that

$$I \simeq 2\pi\rho\omega^2\Big(\frac{R^3}{2}\Big)^2\frac{2}{3}\frac{(\partial_t^2 z)^2}{c^3}$$

(*Idea:* If a sphere moves at constant velocity $\underline{v}_0$ then the field φ that generates the potential flow is $\varphi = ar^{-2}\underline{v}_0\cdot\underline{n} \equiv -a\underline{v}_0\cdot\underline{\partial}\frac{1}{r}$ (because $\underline{v}_0\cdot\underline{\partial}\frac{1}{r}$ is the only scalar field which is linear in $\underline{v}_0$, vanishing at ∞ and harmonic): and A is determined by imposing $\underline{n}\cdot\underline{\partial}\varphi|_{\underline{r}=R\underline{n}} = \underline{v}_0\cdot\underline{n}$: *i.e.* $a = \frac{1}{2}R^3$. Hence $\underline{A} = \underline{v}_0\frac{R^3}{2}$ and $\dot{\underline{A}} = \frac{R^3}{2}\underline{k}\partial_t^2 z$, if the motion is slow, $\omega L \ll c$, *c.f.r.* [LL71], §68, 69.)

[1.7.17]: (*comparison between energy emitted by a small rigid body and the electro-magnetic energy emitted by a charge when both have a harmonic motion*) Compare the electro-magnetic energy emitted by a charge e oscillating as $z(t) = L\cos\omega t$, $L \ll c/\omega$, with c = speed of light and the energy irradiated by a ball oscillating in the in the same way in a fluid. (*Idea:* The expressions coincide if one identifies the electro-magnetic e^2/c^3 with $2\pi\rho\omega^2\Big(2^{-1}R^3\Big)^2/c^3$: recall that the energy emitted per unit time by an oscillating charge is $\frac{2}{3}e^2c^{-3}\,\overline{\ddot{z}}$, *c.f.r.* [Be64], (11.19) or (13.5).)

[1.7.18]: (*non integrability of nonlinear surface waves in deep fluids*) Recall that a Hamiltonian system obtained by perturbing a system of harmonic oscillators with a perturbing function containing non quadratic terms in the canonical variables of the oscillators can be formally reduced to an integrable system provided suitable non resonance conditions are satisfied, *c.f.r.* [Ga83], §10. Then one can put the problem of trying to show that such conditions are verified in the theory of waves in a deep fluid studied in (F). Attempt an analysis of which kind of problems one has to tackle in order to check the non resonance conditions (and identify some of them). *Comment*: The problem can be exactly solved and one can show that the system can be (remarkably enough) written as an integrable one provided one neglects the terms of order above the third (included), see [DZ94],[DLZ95], or even above the fourth (included), see [CW95]. But if the fifth order terms are not neglected then the system *cannot* be integrated by quadratures, [CW95].

Bibliography: [LL71], §12,13,68,69, [La32] §167, p. 248; [Ar79] Appendix 2, D; [Be64] vol. II, §11, (11.19); [DZ94], [DLZ95], [CW95].

CHAPTER II

Empirical algorithms.

§2.1 Incompressible Euler and Navier–Stokes fluidodynamics. First empirical solutions algorithms. Auxiliary friction and heat equation comparison methods.

(A) Euler equation

Imagine an incompressible Euler fluid in a fixed volume Ω with a (C^∞)–regular boundary. The equations describing it are

$$\begin{array}{lll} (1) & \underline{\partial}\cdot\underline{u}=0, & \text{in } \Omega \\ (2) & \partial_t\underline{u}+\underset{\sim}{u}\cdot\underset{\sim}{\partial}\underline{u}=-\rho^{-1}\underline{\partial}p-\underline{g}, & \text{in } \Omega \\ (3) & \underline{u}\cdot\underline{n}=0, & \text{in } \partial\Omega \\ (4) & \underline{u}(\xi,0)\equiv\underline{u}_0(\xi), & t=0 \end{array} \qquad (2.1.1)$$

where $\underline{n}$ denotes the external normal to $\partial\Omega$ and the "*boundary condition*" (3) expresses the condition that the fluid "glides" (without friction) on the boundary of Ω.

Given a t–independent external field $\underline{g}(\xi)\in C^\infty(\Omega)$ the problem of fluidodynamics with fixed walls is:

(1) Given $\underline{u}_0\in C^\infty(\Omega)$ with $\underline{\partial}\cdot\underline{u}_0=0$, $\underline{u}_0\cdot\underline{n}=0$ on $\partial\Omega$, is there a solution $t\to\underline{u}(\xi,t)$, $p(\xi,t)$ of (2.1.1) valid for small enough times and with $\underline{u}$ and p of class C^∞ or, at least, with continuous derivatives?
(2) Are there solutions global in time?
(3) Are such solutions unique?
(4) Under which assumptions on $\underline{g}$ and $\underline{u}_0$ can one find uniform estimates as a function of time on the derivatives of $\underline{u}$ and p?

Obviously it would be desirable to have a positive answer to (2), (3) while (4) would have importance in view of a consistency check with the physics supposedly described by the equations, which we *stress that have been deduced under the hypothesis that velocity gradients stay small.*

In this section we shall look for a heuristic and constructive algorithm for the existence of the solutions. Given $\underline{u}_0(\xi)$ we can imagine computing $\underline{u}(\xi,t)$, for t very small, as

$$\underline{u}(\xi,t)=\underline{u}_0(\xi)+\dot{\underline{u}}(\xi,0)\,t \tag{2.1.2}$$

However to compute $\dot{\underline{u}}(\xi,0)$ we need to know p. The pressure p can be computed from (2) in (2.1.1); indeed the divergence of (2), and equation (3) in (2.1.1), give

$$\begin{aligned}&-\rho\,\underline{\partial}\cdot(\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{u})+\rho\,\underline{\partial}\cdot\underline{g}=\Delta p &&\text{in }\Omega\\ &\partial_n p=-\rho\,[(\underset{\sim}{u}\cdot\underset{\sim}{\partial})\,\underline{u}]\cdot\underline{n}+\rho\,\underline{g}\cdot\underline{n} &&\text{in }\partial\Omega\end{aligned} \tag{2.1.3}$$

which shows that p_0 is determined up to a constant; note that the inhomogeneous Neumann problem in (2.1.3) satisfies automatically the well known compatibility condition imposing that the integral on the boundary of $\partial_n p$ be equal to the volume integral of the datum of the problem (*i.e.* the l.h.s. of the first of (2.1.3)), because of the integration theorem of Stokes.

Inserting the function p_0 so computed into (2) in (2.1.1) at $t=0$, we see that we can compute $\dot{\underline{u}}(\xi,0)$ and that, by construction, $\underline{\partial}\cdot\dot{\underline{u}}(\xi,0)=0$. Therefore it makes sense to define (2.1.2) and, in fact, we just found an *approximation algorithm* which could even be of interest in numerical simulations. We see also which is the mechanism permitting us to eliminate the pressure by expressing it as a function of the velocity field.

We set, given $t_0>0$, for $k\ge 1$:

$$\begin{aligned}&\underline{u}(\xi,kt_0)=\underline{u}(\xi,(k-1)t_0)+t_0\,\dot{\underline{u}}(\xi,(k-1)t_0)=\underline{u}_k(\xi)\\ &\underline{u}(\xi,0)\equiv\underline{u}_0\end{aligned} \tag{2.1.4}$$

where

$$\begin{aligned}&\dot{\underline{u}}(\xi,(k-1)t_0)=-\frac{1}{\rho}\underline{\partial}p_{k-1}(\xi,(k-1)t_0)-\\ &-\underset{\sim}{u}\,(\xi,(k-1)t_0)\cdot\underset{\sim}{\partial}\,\underline{u}(\xi,(k-1)t_0)+\underline{g}\end{aligned} \tag{2.1.5}$$

$$p_{k-1}=\ \text{solution of}\ \begin{cases}\Delta p_{k-1}=-\rho\underline{\partial}\cdot(\underset{\sim}{u}_{k-1}\cdot\underset{\sim}{\partial}\underline{u}_{k-1})+\underline{\partial}\cdot\underline{g} & \Omega\\ \partial_n p_{k-1}=-\rho(\underset{\sim}{u}_{k-1}\cdot\underset{\sim}{\partial}\underline{u}_{k-1})\underline{n}+\underline{g}\cdot\underline{n} & \partial\Omega\end{cases}$$

and the question is whether the limit

$$\underline{u}(\xi,t)=\lim_{\substack{k\to\infty\\ t_0\to0,\,kt_0=t}}\underline{u}_k(\xi) \tag{2.1.6}$$

exists and gives a solution of (2.1.1).

For the time being we shall consider the problem of the existence of the limits a "technical" one and, therefore, we can consider that the equation (2.1.1) is "formally solved".

The (2.1.6) with k finite and large gives, at least in principle (because the calculation of $\underline{u}_k$ is obviously very difficult also from the point of view of numerical solutions) an approximation algorithm for an incompressible motion that generates evolutions that, on one hand, can be interesting by themselves even not considering applications to fluids, and that, on the other hand, can be considered as models for a real fluid evolution.

The approximations should become better as the "discretization time" t_0 approaches 0.

Remark: the algorithm defined by (2.1.4),(2.1.5),(2.1.6) is, at times, called the *Euler algorithm*: more generally this algorithm provides solutions of a differential equation $\underline{\dot{x}} = \underline{f}(\underline{x})$, with $\underline{x}(0) = \underline{x}_0$ that, at time kt_0, is given by the recursive relation $\underline{x}_k = \underline{x}_{k-1} + t_0 \underline{f}(\underline{x}_{k-1})$, $k \ge 1$.

(B) Navier–Stokes–Euler equation.

We look if, at least in a heuristic sense like the one analyzed in (A), a similar treatment of the incompressible Navier–Stokes equations is possible. The equations are

$$\begin{array}{lll} (1) & \underline{\partial}\cdot\underline{u}=0 & \text{in } \Omega \\ (2) & \partial_t\underline{u}+\underset{\sim}{\underline{u}}\cdot\underset{\sim}{\underline{\partial}}\underline{u}=-\frac{1}{\rho}\underline{\partial}p+\underline{g}+\nu\Delta\underline{u} & \text{in } \Omega \\ (3) & \underline{u}\cdot\underline{n}=0, & \text{in } \partial\Omega \\ (4) & \underline{u}(\xi,0)=\underline{u}_0(\xi) & t=0 \end{array} \qquad (2.1.7)$$

provided one can assume that friction between fluid and boundaries is negligible and, therefore, *the fluid glides without friction along the walls* of the container.

In this case the equations are discussed exactly as in the Euler case (with obvious modifications) and we find, therefore, an approximation algorithm similar to (2.1.4), (2.1.5): in (2.1.5) one has to add on the r.h.s. of the first equation the term $\nu\,\Delta\underline{u}_{k-1}$ while the equation in (2.1.5) for $\Delta\, p_{k-1}$ is unchanged (because $\nu\,\Delta\underline{u}_{k-1}$ has zero divergence) and the equation for $\partial_n p_{k-1}$ is modified by adding $\nu\,\underline{n}\cdot\Delta\underline{u}_{k-1}$.

Therefore there are no really new problems.

(C) Navier–Stokes equations and algorithmic difficulties.

In the applications friction against the walls *is by no means negligible*, to the extent that the physically significant boundary condition is $\underline{u} = 0$ rather than $\underline{u}\cdot\underline{n} = 0$.

We then call the (2.1.7) NSE–equations for the fluid in Ω while we shall call NS–equations for a fluid in Ω the equation

$$\begin{aligned}
&(1)\quad \underline{\partial}\cdot\underline{u}=0 && \text{in } \Omega\\
&(2)\quad \partial_t\,\underline{u}+\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{u}=-\frac{1}{\rho}\underline{\partial}\,p+\underline{g}+\nu\,\Delta\underline{u} && \text{in } \Omega\\
&(3)\quad \underline{u}=0 && \text{in } \partial\Omega\\
&(4)\quad \underline{u}(\xi,0)=\underline{u}_0(\xi) && t=0
\end{aligned} \tag{2.1.8}$$

The analysis of (2.1.8) is, however, radically different: indeed if we attempt at determining p at time $t=0$ we find

$$\begin{aligned}
&\Delta p=-\rho\,\underline{\partial}\cdot(\underset{\sim}{u}_0\cdot\underset{\sim}{\partial}\underline{u}_0)+\rho\underline{\partial}\cdot\underline{g} && \text{in } \Omega\\
&\underline{\partial}\,p=-\rho\underline{g}+\rho\nu\Delta\underline{u}_0 && \text{in } \partial\Omega
\end{aligned} \tag{2.1.9}$$

which in general *will not admit a solution* because it is not necessarily true that the tangential derivatives of p, which obviously are already determined by just the normal derivative (via the solution of the corresponding inhomogeneous Neumann problem), are compatible on $\partial\Omega$ with (2.1.9).

One could, for a moment, hope that the fact that $\underline{u}_0$ is not arbitrary, being with zero divergence, implies *ipso facto* compatibility: but it is easy to convince oneself that this is not the case, see (E) and problems [2.1.8], [2.1.9] below.

This is a serious difficulty showing that we must necessarily expect that on the boundary of Ω interesting and difficult phenomena must take place.

The first effect of the difficulty is that, on the basis of what said until now, it does not yet allow us to give a prescription for a numerical solution of the NS–equation with *viscous adherence* to the boundary.

And one can legitimately suspect that (2.1.8) is not a well posed problem: it will certainly be necessary to interpret suitably (2.1.8) since if we interpret it in a strict sense, in which all functions involved are $C^\infty(\Omega)$, it simply looks inconsistent because it does not allow us to compute $\dot{\underline{u}}$, not even at $t=0$ (because it is defined by an insoluble equation).

A way to proceed to develop an algorithm which, at least on a heuristic basis, permits us to compute a solution to (2.1.8), giving at the same time a suitable interpretation to it and bypassing the problem just met, is the following.

Along the normal direction to $\partial\Omega$, we imagine extending the volume Ω by a length ε and we denote Ω_ε this extended volume; we suppose that the fluid occupies it and there it verifies there the equation

$$\begin{aligned}
&\underline{\partial}\cdot\underline{u}=0 && \text{in } \Omega_\varepsilon\\
&\partial_t\underline{u}+\underset{\sim}{u}\cdot\underset{\sim}{\partial}\underline{u}=-\frac{1}{\rho}\underline{\partial}p+\underline{g}-\nu\Delta\underline{u}-\sigma_\varepsilon(\xi)\underline{u} && \text{in } \Omega_\varepsilon\\
&\underline{u}\cdot\underline{n}=0 && \text{in } \partial\Omega_\varepsilon\\
&\underline{u}(\xi,0)=\underline{u}_0 && \text{in } \Omega_\varepsilon
\end{aligned} \tag{2.1.10}$$

where $\underline{u}_0$ is extended arbitrarily between Ω and Ω_ε, assuming that it is extended, together with a prefixed number p of derivatives ($p \geq 2$), continuously and assuming also that the extension vanishes near the new boundary $\partial\Omega_\varepsilon$. The function $\sigma_\varepsilon(\xi)$ vanishes inside Ω and increases rapidly from 0 to a value $\overline{\sigma}(\varepsilon) \stackrel{def}{=} \overline{\sigma}_\varepsilon$ with (large) average slope $\overline{\sigma}(\varepsilon)/\varepsilon$. This additive term has the interpretation of "*friction*" that slows down the fluid in the corridor between $\partial\Omega$ and $\partial\Omega_\varepsilon$.

The level lines of $\sigma_\varepsilon(\xi)$ will be, by their definition, parallel to $\partial\Omega$ and $\partial\Omega_\varepsilon$, the gradient of σ_ε, denoted $\sigma'_\varepsilon \underline{n}$ with σ'_ε such that $|\underline{n}| = 1$, is a vector field extending, to the layer between the two boundaries, the fields formed by their normals (draw a picture).

One can expect that in the limit $\varepsilon \to 0$ the solution to (2.1.10), for which we can apply the constructive algorithm of the preceding case (B) because on $\partial\Omega_\varepsilon$ the condition $\underline{u} \cdot \underline{n} = 0$ holds, will be such that the limit

$$\lim_{\varepsilon \to 0} \underline{u}_\varepsilon(\xi, t) = \underline{u}(\xi, t) \qquad \xi \in \Omega \tag{2.1.11}$$

exists and $\underline{u}(\xi, t)$ solves (2.1.8), *under the further condition that* $\overline{\sigma}(\varepsilon) \to \infty$, *possibly fast enough, for* $\varepsilon \to 0$.

Remark: If Ω has special forms, *e.g.* it is a cube, then it might be convenient to take Ω_ε to be a torus. We shall do so below in the case of a similar simpler, one dimensional, problem.

The algorithm (2.1.10), that we shall call "*auxiliary friction algorithm*" or "*friction method*", will encounter considerable numerical difficulties because of the term $\sigma_\varepsilon \underline{u}$ and of the divergence towards $+\infty$ (for $\varepsilon \to 0$) of $\bar{\sigma}_\varepsilon$; in fact, for $\varepsilon \to 0$, $\dot{u}_t$ computed from (2.1.10) would tend to ∞, unless one could guarantee *a priori* that $u_\varepsilon \to 0$ when σ_ε becomes large: and it does so quickly enough to control the product $\sigma_\varepsilon \underline{u}_\varepsilon$ (which appears quite difficult a task from a mathematical viewpoint).

(D) Heat equation.

To understand better what is happening let us examine a simpler model case. Consider the heat equation

$$\begin{aligned} \partial_t T &= c\, \partial_x^2 T, \qquad\qquad x \in [-\pi, \pi] \\ T(-\pi) &= T(\pi) = 0 \\ T(x, 0) &= T_0(x) \end{aligned} \tag{2.1.12}$$

If we attempted to find a solution to this equation with the preceding method we should set

$$T_k = T_{k-1} + t_0 \dot{T}_{k-1} \equiv T_{k-1} + t_0 c T''_{k-1} \quad k \geq 1 \tag{2.1.13}$$

but it is clear that, unless $T''(\pm\pi) = 0$, this will already be impossible because T_1 will not verify the boundary conditions.

And even if $T_0''(\pm\pi) = 0$ we see that $T_1''(\pm\pi) \neq 0$ unless $T_0''''(\pm\pi) = 0$. Hence if $T_0(x)$ does not have *all* even derivatives vanishing in the neighborhood of $\pm\pi$ the algorithm will not work.

The "*auxiliary friction method*" of (C) above, setting $c = 1$, would lead to equations

$$\begin{aligned} &\dot T = T'' - \sigma_\varepsilon T \\ &T(-\pi-\varepsilon) = T(\pi+\varepsilon),\, T'(-\pi-\varepsilon) = T'(\pi+\varepsilon) \\ &T(x,0) = T_0(x) \end{aligned} \tag{2.1.14}$$

having extended T_0 arbitrarily out of $[-\pi,\pi]$ to a periodic function (see the comment following (2.1.11)) in $[-\pi-\varepsilon,\pi+\varepsilon]$, and imagining to identify the points $\pi+\varepsilon$ and $-\pi-\varepsilon$, see Fig. (2.1.1) below.

If we choose σ_ε equal to a constant $\bar\sigma_\varepsilon$ for $x \in [\pi + \frac{2}{3}\varepsilon, \pi+\varepsilon]$ and to 0 for $x \in [\pi, \pi + \frac{\varepsilon}{3}]$ and $\sigma_\varepsilon(x) = \sigma_\varepsilon(-x)$, then we can show that the iterative method works, in principle, at least if the initial datum T_0 is regular enough (and if $\bar\sigma_\varepsilon \to \infty$ fast enough as $\varepsilon \to 0$).

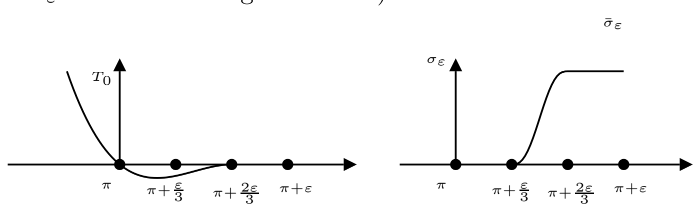


Fig. (2.1.1): Extension of the initial datum $T_0(x)$ to the right of π; and graph of the auxiliary function σ_ε between π and $\pi+\varepsilon$.

This time the problem is so simple that one can show that, *if* $T_0(x)$ *is* C^∞, *and analytic in the interior of* $[-\pi,\pi]$, then the algorithm (2.1.13) converges to the solution of the heat equation with the correct boundary conditions (2.1.12). But the convergence of the function $T_{\varepsilon,k}(\xi,t)$, provided by the algorithm at the k–th step, to a limit $T(\xi,t)$

$$T(\xi,t) = \lim_{\varepsilon\to0} \lim_{\substack{k\to\infty \\ kt_0=t}} T_{\varepsilon,k}(\xi) \tag{2.1.15}$$

is delicate and, as a numerical algorithm, it is not very good because $\dot T$ risks to become very large as $\varepsilon \to 0$, because $\sigma_\varepsilon \to \infty$ badly affects the term $\sigma_\varepsilon T$ in (2.1.14).

Furthermore if T_0 is not analytic but, for instance, "only" C^∞, the algorithm may converge but, in general, it *will not converge* to the usual solution of the heat equation. See below and problems [2.1.1]%[2.1.7].

An apparently better algorithm consists in transforming the equation into

$$T_\varepsilon(\xi,t) = e^{-\sigma_\varepsilon t} T_0(\xi) + \int_0^t e^{-\sigma_\varepsilon(\xi)(t-\tau)} T_\varepsilon''(\xi,\tau)\, d\tau \tag{2.1.16}$$

which we can try to solve iteratively by setting

$$T_{\varepsilon k}(\xi) = e^{-\sigma_\varepsilon t_0 k} T_0(\xi) + t_0 \sum_{h=0}^{k-1} e^{-\sigma_\varepsilon(\xi)(k-h)t_0} T''_{\varepsilon\, h}(\xi), \qquad k \ge 1 \tag{2.1.17}$$

This yields a numerical algorithm in which σ_ε only appears in $e^{-\sigma_\varepsilon t}$, which can be computed without involving large quantities (even when $\sigma_\varepsilon t$ is large). Provided obviously $T''_\varepsilon(\xi)$ does not become too large, which we do not expect, not at least if compared with the corresponding $e^{-\sigma_\varepsilon(\xi)t_0}$ (the reason is that in the well known solution of the heat equation T'' has, usually, a discontinuity at the boundary but it is not infinite). See problems.

(E) An empirical algorithm for NS:

The above discussion suggests an analogous approach for (2.1.8) and (2.1.9) and leads to the following algorithm for (2.1.10). One writes (2.1.10) as

$$\begin{aligned} &\underline{u}\cdot\underline{n} = 0 \text{ on } \partial\Omega, \qquad \underline{\partial}\cdot\underline{u} = 0, \text{ in } \Omega, \qquad \underline{u}(\xi,t) = \qquad (2.1.18)\\ &= \underline{u}_0(\xi)e^{-\sigma_\varepsilon(\xi)t} + \int_0^t e^{-\sigma_\varepsilon(\xi)(t-\tau)}(-\underset{\sim}{u}_\tau\cdot\underset{\sim}{\partial}\underline{u}_\tau - \frac{1}{\rho}\underline{\partial}p_\tau - \underline{g} - \nu\Delta\underline{u}_r)\,d\tau \end{aligned}$$

where u_τ, p_τ denote $u(\xi,\tau), p(\xi.\tau)$ and one sets up the approximation algorithm

$$\begin{aligned} &\underline{u}_k(\xi) = \underline{u}_0(\xi)e^{-\sigma_\varepsilon(\xi)t} + \sum_{h=0}^{k-1} t_0 e^{-\sigma_\varepsilon(\xi)(k-h)t_0}(-\underset{\sim}{u}_h\cdot\underset{\sim}{\partial}\underline{u}_h -\\ &-\frac{1}{\rho}\underline{\partial}p_h + \underline{g} + \nu\Delta\underline{u}_h) \quad k\ge 1 \qquad (2.1.19)\\ &\Delta p_k = -\rho\,\underline{\partial}(\underset{\sim}{u}_k\cdot\underset{\sim}{\partial}\underline{u}_k) + \rho\,\underline{\partial}\cdot\underline{g} - \underline{u}_k\cdot\underline{\partial}\sigma_\varepsilon \qquad \text{in } \Omega_\varepsilon\\ &\partial_n p_k = -\rho(\underset{\sim}{u}\cdot\underset{\sim}{\partial}\underline{u}_k)\cdot\underline{n} + \rho\,\underline{g}\cdot\underline{n} + \rho\,\nu\Delta\underline{u}_k\cdot\underline{n} \qquad \text{in } \partial\Omega_\varepsilon \end{aligned}$$

where $\underline{\partial}\sigma_\varepsilon = \sigma'_\varepsilon\underline{n}$ and σ'_ε has support near $\partial\Omega_\varepsilon$. This algorithm does not encounter obvious problems as $\sigma_\varepsilon \to \infty$, provided $\underline{u}_k\cdot\underline{n}$ is so small that $\sigma'_\varepsilon\underline{u}_k\cdot\underline{n}$ remains bounded as $\varepsilon\to 0$ (see the last term in the second equation). Since we expect that $\underline{u}_k\cdot\underline{n}$ be of order $O(\varepsilon)$, this seems reasonable and we should have

$$\underline{u}(\xi,t) = \lim_{\varepsilon\to 0}\lim_{\substack{k\to\infty\\ kt_0=t}} \underline{u}_{k,\varepsilon}(\xi) \tag{2.1.20}$$

However, unlike the heat equation case, the algorithm (2.1.19) does not eliminate completely the problem, certainly relevant for numerical calculations, that is due to the presence of quantities like $\underline{u}_k\cdot\underline{\partial}\,\sigma_\varepsilon$ which are products of large quantities ($\underline{\partial}\sigma_\varepsilon$) times quantities ($\underline{u}_k$) that we expect to be small (but which we do not know *a priori* that they are really such).

Hence this algorithm has only a theoretical character which, rather than solving a problem, is well suited to illustrate some of its difficulties. We succeed in giving a meaning to a heuristic method of construction of solutions, but the method remains only a matter of principle, conditioned to the solution of an elliptic equation which may lead to a numerical stability problems (at least).

We could ask where did go the conceptual compatibility problem, at the origin of our worries. Obviously it is still around: indeed one should expect that the "exact" solutions of (2.1.18) or of (2.1.10) for $t > 0$ small do not verify any more the boundary conditions $\underline{u} = \underline{0}$ on $\partial\Omega$ (that now is an internal surface in the domain Ω_ε where the approximating solution is studied). But we can hope that the violation of the property of vanishing on the boundary $\partial\Omega$ should rapidly become, as time increases, very small for $t > 0$ and tend to vanish for every prefixed value of $t > 0$ as ε tends to zero, at least if σ_ε becomes vertical enough near $\partial\Omega$ and large enough (for $\varepsilon \to 0$).

The impossibility of satisfying the boundary condition at $t = 0$ (on $\partial\Omega$) implies that $\dot{\underline{u}}_\varepsilon$ will be not zero, and not even small, at $t = 0$ in some point on the boundary $\partial\Omega$ but it will rapidly become very small, the earlier the closest to 0 will ε be: and in the limit $\varepsilon \to 0$ one should attain a limit $\underline{u}$ which at time $t = 0$ *does not* satisfy the NS–equation, *but that verifies it at all later times* $t > 0$.

In this way the friction model for the boundary condition clarifies how it could be that the equation cannot be solved at $t = 0$ but it is solved at all $t > 0$ and which is a physical mechanism that produces such a result.

Should one worry that the solution found (*if* found) had nothing to do with the initial datum $\underline{u}_0$ then it should be noted that, although the initial datum cannot be obtained as a limit from $t > 0$, in the sense that not all derivatives of $\underline{u}$, for $t > 0$ (time derivatives included) tend to the corresponding ones of the initial datum[1] nevertheless motion is generated from the initial datum hence we can expect that the $\underline{u}$ tends to $\underline{u}_0$ in some sense, *e.g.* in $L_2(\Omega)$ or in some other sense (for instance pointwise in the internal points of Ω).

The important point to retain is that the physical meaning of the singularity at $t = 0$ has been understood, at least as a proposal to be checked, through a concrete model together with a description of the mechanism of transition from $t = 0$, where the equation cannot be solved in general, to $t > 0$ where, instead, its solubility does not meet any *a priori* difficulty of principle.

The following questions become natural at this point

[1] We have already seen that this cannot happen for the time derivative which, for the initial datum, is not even defined.

Question (Q1): Do the friction equations (2.1.10), once exactly solved for $\varepsilon > 0$, really have the property of converging for $\varepsilon \to 0$ to a solution of the (2.1.8) (boundary conditions included) at least for $t > 0$.

Question (Q2): Does the algorithm (2.1.19) really converge for $k \to \infty$, $t_0 \to 0$, $kt_0 \to t$, to a solution of (2.1.18).

and we attempt at understanding them by referring to simpler problems in which they already occurr.

(F) Precision of the same algorithms applied to the heat equation, (Q1).

Questions (Q1) and (Q2) will not be analyzed for the NS–equation because of the difficulties that are involved (which, to date, are not yet solved in the case $d = 3$). We can, however, analyze them in the far easier case of the heat equation and the study is very instructive as an illustration of mechanisms and difficulties that one can expect to find again in the case of more ambitious theories like that of the NS–equation. Let us, therefore, look at the problem (Q1) posed in *(E)* in the case of the equations (2.1.12), (2.1.14).

We shall however suppose, for simplicity, that σ_ε in (2.1.14), (2.1.16) is given as a discontinuous function represented in Fig. (2.1.2) by the solid lines rather than by the dashed lines.

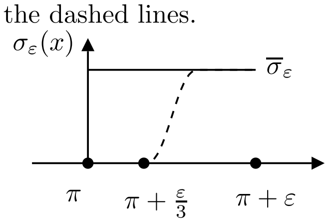


Fig. (2.1.2) The function $\sigma_\varepsilon(x)$ in $[\pi, \pi + \varepsilon]$, while in $[-\pi - \varepsilon, -\pi]$ it is defined as the mirror image: the graph is the solid line jumping from 0 at π to $\overline{\sigma}_\varepsilon$ at points $> \pi$. The dashed line refers to the smooth graph of the function previously used.

Furthermore it is convenient to use the fact that (2.1.14) satisfies periodic boundary conditions and to translate the "potential" σ_ε to the center of the interval (this is done via the change of coordinates $x' = x - \pi - \varepsilon$). In this new representation the boundary layer of width ε is now at the center of the interval $[-\pi - \varepsilon, \pi + \varepsilon]$ as in the Fig. (2.1.3): half of it, namely $[-\varepsilon, 0]$, corresponds to the previous left boundary layer and the other half, namely $[0, \varepsilon]$, corresponds to the previous right boundary layer.

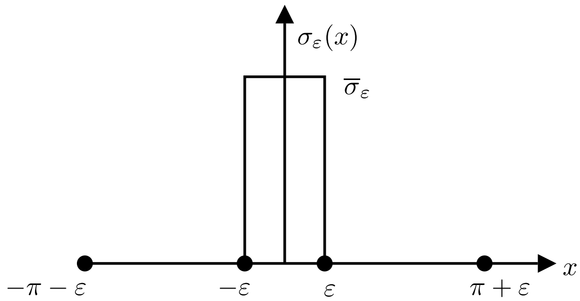


Fig. (2.1.3) The function σ_ε after the shift of th origin. The points $-\pi-\varepsilon$ and $\pi+\varepsilon$ are identified by the periodic boundary conditions.

We shall therefore consider the equation

$$\begin{array}{ll}
(1) & \dot T = T'' - \sigma_\varepsilon T, \qquad \text{in } [-\pi-\varepsilon, \pi+\varepsilon]/\{\pm\varepsilon\} \\
(2) & \text{periodic boundary conditions and matching of the} \\
 & \text{values and of the first derivatives of } T \text{ in } \pm\varepsilon \\
(3) & \sigma_\varepsilon = 0 \text{ if } |x| > \varepsilon \text{ and } \sigma_\varepsilon = \overline{\sigma}_\varepsilon \text{ if } |x| < \varepsilon \\
(4) & \text{initial datum, if } T_0 \text{ is the datum in (2.1.14)}: \\
 & \vartheta_0(x) = T_0(x-\pi-\varepsilon) \text{ if } x > \varepsilon, \text{ and} \\
 & T_0(x+\pi+\varepsilon) \text{ if } x < -\varepsilon \text{ and zero in } |x| < \varepsilon
\end{array} \tag{2.1.21}$$

The exclusion of the points $\pm\varepsilon$ in the first of (2.1.21) is replaced by the matching condition which is now posed because σ_ε has been chosen discontinuous, see Fig. (2.1.3), and the equation (2.1.12) must, as a consequence, be somehow interpreted at the discontinuity points.[2]

The following proposition solves the question analogous to the (Q1), in the case of the heat equation

1 Proposition *(heat equation approximation algorithm): The heat equation, in the form (2.1.21), with ϑ_0 generated (c.f.r. (4) in (2.1.21)) by a smooth datum $T_0(x)$, say $T_0 \in C^\infty([-\pi,\pi])$, admits a solution $t \to T^\varepsilon(x,t)$ which for $\varepsilon \to 0$ and $\overline{\sigma}_\varepsilon \to +\infty$ fast enough converges for all $x \in [-\pi,\pi]$ to a smooth solution of the heat equation (2.1.12), (described in problem* [2.1.6]*).*

The equation can be solved exactly: the proof can therefore be reduced to a simple but instructive check, *c.f.r.* problems [2.1.1]%[2.1.7].

(G) Analysis of the precision of the algorithms in the heat equation case, (Q2).

We now study the question (Q2) following the (2.1.20), *i.e.* we study whether the algorithm (2.1.17) of solution of (2.1.16) does really gener-

---

[2] The matching condition yields a possible interpretation: obviously it is only one of the possible interpretations and as such it has a physical significance, that however we shall not discuss here, given its auxiliary nature in the present context.

ate the wanted smooth solution, *i.e.* the solution guaranteed by the last proposition.

It is easy to convince oneself that *this is not the case*, in general, because this is already not true in the analogous case of the problem (2.1.12) in which the Dirichlet boundary condition is replaced by the even easier periodic boundary condition ($T(-\pi) = T(\pi),\ T'(\pi) = T'(-\pi)$). In such case there is even no need of introducing σ_ε nor of enlarging the domain: because the boundary conditions are respected by the algorithm (2.1.13) and the conceptual problems discussed in (D) (following (2.1.13)) do not even arise.

Note also that the algorithms (2.1.13) and (2.1.17) coincide in this periodic boundary conditions case.

Nevertheless the question whether the algorithm (2.1.13) does really produce the correct solution can be posed in this simpler problem (*i.e.* (2.1.12) with periodic boundary conditions and C^∞ periodic initial datum) and it is clearly a simpler question than the one in (Q2) about the algorithm (2.1.19) for the case of the NS equations with Dirichlet boundary conditions. In fact the following proposition holds

2 Proposition: *(anomaly of approximations convergence for the heat equation) Consider the equation (2.1.12) but with periodic boundary conditions (as above) and C^∞, periodic, initial datum. Let, for $k \ge 1$*

$$T_k = T_0 + ct_0 \sum_{h=1}^{k} T''_{k-h}, \qquad \longleftrightarrow \qquad T_k - T_{k-1} = c\, t_0\, T''_{k-1} \tag{2.1.22}$$

Let $\omega = 2\pi k$ with k integer and let $\hat T_0(\omega)$ be the Fourier transform of $T_0(x)$ then

(i) if, for $\tau, b, \eta > 0$, it is $|\hat T_0(\omega)| < \tau e^{-b|\omega|^{2+\eta}}$ (implying, among other things, that T_0 is analytic entire in x) the sequence (2.1.22) converges to the usual, well known, solution $T(x,t)$ of (2.1.12) for $t_0 \to 0$,

(ii) in general $\hat T_k(\omega)$ has a lmit $\hat T(\omega, t)$ as $k \to \infty$ with $kt_0 \to t$, for every ω

(iii) unless the condition in (i) on the Fourier transform holds it is not in general true that $T_k(x) \xrightarrow[kt_o\to t,\, t_0\to 0]{} T(x,t)$.

proof: The Fourier transform of T_k can be computed as

$$\hat T_k(\omega) = \hat T_{k-1}(\omega)(1 - c\omega^2 t_0) = \hat T_0(\omega)(1 - c\omega^2 t_0)^k \tag{2.1.23}$$

where $\omega = 2\pi n$ with n integer > 0.

Hence (writing $t_0 = \frac{t}{k}$) we see that

$$\hat T_k(\omega) \xrightarrow[\substack{kt_0 = t\\ k\to\infty, t\to 0}]{} e^{-c\,\omega^2 t}\hat T_0(\omega) \tag{2.1.24}$$

But this does not mean that $T_k(\xi) \to T(\xi, t)$, where T is the "usual" solution of (2.1.12) with Fourier transform given by the r.h.s. of (2.1.24). One sees this immediately if T_0 has support $[-a, a]$ strictly contained in $[-\pi, \pi]$, $(a < \pi)$. In this case it is clear (from (2.1.22)) that T_k will remain identically zero in the set where it initially was zero with all its derivatives (*i.e.* outside of $[-a, a]$), hence it cannot be analytic for $t > 0$ (while $T(x, t)$ is analytic).

Convergence is, however, guaranteed if

$$|(1 - c\omega^2 \frac{t}{k})^k \hat{T}_0(\omega)| \le \gamma_t(\omega) \quad \text{and} \qquad \sum_\omega |\gamma_t(\omega)| < \infty \tag{2.1.25}$$

with the series of the $|\gamma_t(\omega)|$ uniformly converging in t for t in an arbitrarily prefixed bounded interval. It is easy to see that this happens if, and essentially only if (see problems [2.1.10], [2.1.11]), $\hat{T}_0(\omega) \to 0$ as fast as $\tau e^{-b|\omega|^2}$, or faster, for $\tau, b > 0$: at least if t is small enough and even for all times if the decay in ω is faster (*e.g.* as assumed in hypothesis in (i) of the proposition).

Therefore to make the method work *strong regularity properties must be imposed* on $T_0(x)$: it must be more regular than an entire analytic function of x (simple analyticity would "just" demand that the Fourier transform decays exponentially, *i.e.* as $\tau e^{-b|\omega|}$ for some $\tau, b > 0$ but the above proof would not work).

(H) Comments:

(1) *A fortiori*, we must expect that very strong regularity conditions have to be imposed upon $\underline{u}_0$, besides imposing suitable properties on the values of $\underline{u}$ and of its derivatives on the boundary $\partial\Omega$ (*c.f.r.* the remark following (2.1.13)), so that the algorithm (2.1.19) could converge to the correct solution of the NS–equation (which, unlike the heat equation, has not yet been shown to even admit a solution).

(2) However the regularity conditions could, in the end, simply reduce to analyticity properties of $\underline{u}_0$ and of the boundary $\partial\Omega$ of Ω (whose regularity also influences that of p_0 hence of $\dot{\underline{u}}_0$ and $\underline{u}$, via the Neumann problem, (2.1.3)).

Unfortunately the problem is open, at least if $d = 3$, and it is not even known whether under conditions of this type the NS equation admits a solution which is well defined and keeps, in general, a regularity comparable to that of the initial data for all times $t > 0$.

(3) Note that the proposition is in apparent contradiction with the theory of the heat equation. Usually one says that the heat transport equation in a conducting rod, with fixed temperature at the extremes (equal to 0 or to each other in the above examples) and with regular initial datum "*admits a unique solution*". The solutions that we construct as limits of the approximations with time step t_0 starting from an initial datum which, for instance, is of class C^∞ and vanishes outside an interval $[-a, a]$ with $a < \pi$, and is analytic inside $(-a, a)$ are, in some sense, solutions of the

heat equation for a rod $[-a, a]$[3] Such solutions, regarded as functions on $[-\pi, \pi]$ vanish outside $[-a, a]$ hence they are not C^∞ in $[-\pi, \pi]$ (in general they will have discontinuous first derivatives at $\pm a$, to say the least).

(4) In other words the key to the uniqueness theorem lies in the requirement that T is really at least once differentiable with continuous derivatives on the *whole* interval $[-\pi, \pi]$ and for all times $t \geq 0$. Obviously one could take the alternative view of calling *solutions of the heat equation* the limits of sequences of discretized equations: in this case we could possibly have an existence and uniqueness theorem *once the discretized approximation is fixed* (possibly under additional conditions on the initial datum) but the solutions would be (in general) different from the "classical" solutions.[4] Furthermore the solution provided by the limit of a sequence of discretized equations may depend on which discretized equation is chosen (*i.e.* on which algorithm of approximation is used).

(5) One thus sees that rather deep interpretation problems arise here, which cannot be solved on a purely mathematical ground: to understand which is the correct meaning to give to a "solution" in physically interesting cases or applications it is necessary to go back to the physical properties that the equation translates into a mathematical model.

It is possible that the physical interpretation requires one or another solution depending on the physical origin of the problem. *It is clear that, if such considerations already apply to a simple equation like the* 1*–dimensional heat equation, with greater reason they will be relevant for equations like the Euler or Navier–Stokes equations* and lead to exposing major existence problems. In general uniqueness problems for partial differential equations are delicate both physically and mathematically: even in cases in which one usually says that there are "no problems" (as one says, for instance, for the heat equation). *This shows, once again, that dogmatic attitudes on notions like existence of solutions, or uniqueness, only lead to failure to see the existence of interesting problems.*

(6) The real importance of the above analysis, already for the heat equation, is shown by the fact that it makes at least less convincing the paradox that claims the incompatibility of the heat equation with special relativity: heat "waves" can apparently travel with infinite speed and "an initial datum with support in a finite region will evolve, by the heat equation, into a datum which does not vanish at an arbitrary distance".

This is, to say the least, a hasty conclusion because, if one defined "solution" what is obtained as limit of the above described Euler algorithm, one would instead find a solution that not only does not have infinite velocity but which in fact does not propagate at all (because as we have seen it will remain nonzero only where it was so initially). This is obviously only one more argument beyond the well known one that remarks that the heat

[3] I do not know a reference for detailed analysis of this property.

[4] Which, when the initial data are at least C^1–functions, *by definition* are those of class C^1 in $[-\pi, \pi]$ and $t \geq 0$.

equation is a phenomenological macroscopic equation derived from assumptions that involve non relativistic ideas and notions (like the "radius of the molecules" which appears in the conductivity coefficient c (see [1.1.5]) and which, alone, would invalidate the "paradoxical" infinite speed propagation of heat "predicted" by the heat equation.

(7) Conclusion: the algorithms discussed in this section are interesting for the conceptual difficulties that they illustrate and for the multiplicity of aspects that they bring up on the theory of fluids. They have limited usefulness for applications, if not none at all.

Problems: *Well posedness of the heat equation and other remarks.*

[2.1.1]: Consider the eigenvalue problem associated with (2.1.14):

$$T'' - \sigma_\varepsilon T = -\lambda^2 T, \qquad \text{in } [-\pi-\varepsilon, \pi+\varepsilon] \tag{$*$}$$

and show that for $|x| > \varepsilon$ the eigenvector T has the form $Ae^{i\lambda x} + Be^{-i\lambda x}$ while for $|x| < \varepsilon$ it has the form $\alpha \cosh \sqrt{\sigma_\varepsilon - \lambda^2}\, x + \beta \sinh \sqrt{\sigma_\varepsilon - \lambda^2}\, x$.

[2.1.2] Find the matching conditions determining λ for the eigenvalue problem in [2.1.1]. (*Idea:* If A_+, B_+ and A_-, B_- are the coefficients of the same solution $x > \varepsilon$ or $x < -\varepsilon$ then the periodicity condition imposes, for each η

$$A_+ e^{i\lambda(\pi+\varepsilon+\eta)} + B_+ e^{-i\lambda(\pi+\varepsilon+\eta)} = A_- e^{i\lambda(-(\pi+\varepsilon)+\eta)} + B_- e^{-i\lambda(-(\pi+\varepsilon)+\eta)}$$

Hence: $A_- = A_+ e^{2i\lambda(\pi+\varepsilon)}$, $B_- = B_+ e^{-2i\lambda(\pi+\varepsilon)}$.
To simplify we study only eigenfunctions which are even in x (which suffices to study the (2.1.14) with an even initial datum: $T_0(x) = T_0(-x)$). The parity condition means: $A_+ e^{2i\lambda(\pi+\varepsilon)} e^{-i\lambda x} + B_+ e^{-2i\lambda(\pi+\varepsilon)} e^{i\lambda x} = A_+ e^{i\lambda x} + B_+ e^{-i\lambda x}$ *i.e.* $B_+ e^{-2i\lambda(\pi+\varepsilon)} = A_+$. Thus for $x \ge 0$ the even eigensolution with eigenvalue λ will have the form

$$\begin{aligned} T(x) &= A\,(e^{i\lambda x} + e^{2i\lambda(\pi+\varepsilon)} e^{-i\lambda x}) && x > \varepsilon \\ T(x) &= \alpha \cosh \sqrt{\sigma_\varepsilon - \lambda^2}\, x && |x| < \varepsilon \end{aligned} \tag{$**$}$$

with suitable A, α. Hence the matching condition will be

$$\begin{aligned} A\,(e^{i\lambda\varepsilon} + e^{2i\lambda(\pi+\varepsilon) - i\lambda\varepsilon}) &= \alpha \cosh \sqrt{\sigma_\varepsilon - \lambda^2}\, \varepsilon \\ i\lambda A\,(e^{i\lambda\varepsilon} - e^{2i\lambda(\pi+\varepsilon) - i\lambda\varepsilon}) &= \alpha \sqrt{\sigma_\varepsilon - \lambda^2} \sinh \sqrt{\sigma_\varepsilon - \lambda^2}\, \varepsilon \end{aligned}$$

so that the eigenvalue λ is determined by

$$\frac{1 + e^{2\pi i\lambda}}{1 - e^{2\pi i\lambda}} \equiv i \operatorname{cotg} \pi\lambda = \frac{\lambda i}{\sqrt{\sigma_\varepsilon - \lambda^2}} \coth \varepsilon \sqrt{\sigma_\varepsilon - \lambda^2} \tag{$***$}$$

The odd eigenfunctions are treated similarly.)

[2.1.3]: Show that if the eigenvalues associated with the even eigenfunctions of (*) are labeled, as λ increases, by $0, 1, 2, \ldots$ it is

$$\lambda_n = \left(n + \frac{1}{2}\right) + O\,(\varepsilon n)\,, \qquad n \le \lambda_n \le n+1 \qquad \text{with } n \text{ integer}$$

if $\sigma_\varepsilon \varepsilon^2 \equiv E \xrightarrow[\varepsilon\to 0]{} \infty$. (*Idea:* It suffices to draw the graph of both sides of (***), paying attention to distinguish the cases $\sigma_\varepsilon > \lambda^2$ and $\sigma_\varepsilon < \lambda^2$).

[2.1.4]: Show that the square of the L_2–norm of $T_n(x)$, $\int_{-\pi-\varepsilon}^{\pi+\varepsilon} |T_n(x)^2 dx$, is

$$|A|^2\, 4\pi\, (1 + \frac{\sin 2\pi\lambda_n}{2\pi\lambda_n}) + \varepsilon|\alpha|^2 (1 + \frac{\sinh 2\sqrt{E-\varepsilon^2\lambda_n^2}}{2\sqrt{E-\varepsilon^2\lambda_n^2}})$$

so that if T_n is normalized to 1 in L_2 it is $|A^2| < 1/8\pi$ for $n \ge 1$ (noting that for $n \ge 1$ it is $\lambda_n \ge 1$). Show also that $|T_n(x)| \le C$ for some n–independent C. (*Idea:* The matching conditions in [2.1.3] imply

$$\big|\frac{\alpha}{A}\big| \le \begin{cases} \min\Big(2/|\cos\varepsilon\sqrt{\lambda_n^2-\sigma_\varepsilon}|, 2/\sqrt{\lambda_n^2-\sigma_\varepsilon}|\sin\varepsilon\sqrt{\lambda_n^2-\sigma_\varepsilon}|\Big) & \text{if } \sigma_\varepsilon < \lambda_n^2 \\ \min\Big(2/\cosh\varepsilon\sqrt{\sigma_\varepsilon-\lambda_n^2}, 2/\sqrt{\sigma_\varepsilon-\lambda_n^2}\sinh\varepsilon\sqrt{\sigma_\varepsilon-\lambda_n^2}\Big) & \text{if } \sigma_\varepsilon \ge \lambda_n^2 \end{cases}$$

so that for $n \ge 1$ (*i.e.* for $\lambda_n \ge 1$) we see that $|\alpha/A| \le 4$.)

[2.1.5]: Check that [2.1.3], [2.1.4] imply that $T_n(x)$, normalized in L_2, can be written as

$$T_n(x) = \frac{(e^{i\lambda_n x} - e^{-i\lambda_n x} + O(\varepsilon n))}{\text{normalization}} = \frac{1}{\sqrt{\pi}} \sin(n+\frac{1}{2})x + O(\varepsilon n), \qquad x > \varepsilon$$

Show also that the results in [2.1.4] agree with the corresponding results for the equation $T'' = -\lambda^2 T$ in $[0, 2\pi]$ with boundary conditions $T(0) = T(2\pi) = 0$, which yield eigenvalues $\lambda_n^2 = -\frac{n^2}{4}$ with eigenvectors $\pi^{-1/2}\sin\frac{nx}{2}$. (*Idea:* Note that the two problems are equivalent if one thinks of $[-\pi,\pi]$ and $[0,2\pi]$ as of two identical circles and if the point $x = \pi$ in the first case is identified with the point $x = 0 = 2\pi$ of the second case: with this change the odd order eigenfunctions for the problem in $[0, 2\pi]$ become the even ones for the problem in $[-\pi,\pi]$ (and *viceversa*)).

[2.1.6]: Show that the solution of (2.1.21) with initial data ϑ_0 even in x and as in (2.1.21) can be written, for $t > 0$, in terms of the eigenfunctions analyzed in problems [2.1.1]%[2.1.4] as

$$T_\varepsilon(\xi, t) = \sum_{k=1}^\infty e^{-\lambda_k^2 t} \langle T_k, \vartheta_0\rangle T_k(x)$$

where $\langle T, T'\rangle \equiv \int_{-\pi-\varepsilon}^{\pi+\varepsilon} T(x)T'(x)\, dx$: see (2.1.21) for the definition of ϑ_0.

[2.1.7]: Suppose that ϑ_0 is even in x. Show that the results of [2.1.4] imply $\langle T_k, \vartheta_0\rangle \xrightarrow[\varepsilon\to 0]{} \langle \overline{T}_k, \tilde{T}_0\rangle$, where $\tilde{T}_0$ is the limit $\lim_{\varepsilon\to 0}\vartheta_0(x)$ and $\overline{T}_k$ are the eigenfunctions of the problem

$$\begin{aligned} &T'' = -\lambda^2 T, \quad \text{in} \quad [-\pi,\pi]/\{0\} \quad \text{with periodicity } 2\pi \\ &T(0) = 0 \end{aligned} \tag{a}$$

i.e. for $x > 0$

$$\overline{T}_k(x) = \frac{1}{\sqrt{\pi}} \sin\Big(k + \frac{1}{2}\Big)\, x, \qquad \bar{\lambda}_k = k + \frac{1}{2} \tag{b}$$

Furthermore check that: $\lambda_k \xrightarrow[\varepsilon\to 0]{} \overline{\lambda}_k$ and $\lambda_k^2 \ge C\,\overline{\lambda}_k^2$ for each $k, |\varepsilon| < 1$, if C is suitably chosen. Infer from this that

$$T_\varepsilon(\xi,t) \xrightarrow[\varepsilon\to 0]{} T(\xi,t) = \sum_{k=1}^{\infty} e^{-\,\overline{\lambda}_k^2 t} \langle \overline{T}_k, \tilde{T}_0\rangle\, \overline{T}_k(x) \qquad (c)$$

and that the limit $T(\xi,t)$ solves the heat equation on $[-\pi,\pi]/\{0\}$ with boundary condition $T(0)=0$ at $x=0$. Check that this easily leads to a solution of the problem (2.1.12) because (2.1.12) can be interpreted as the above "translated" by π. (*Idea:* By [2.1.4] the terms of the series for $T_\varepsilon(\xi,t)$ can be estimated by a constant times $e^{-\lambda_n^2 t}$ and obviously $\langle T_k, T_0\rangle T_k(x)$ converges to the corresponding value in the expression for $T(\xi,t)$.)

[2.1.8]: Consider, in the case $d=2$, a solenoidal nonconservative force $\underline{g}\in C^\infty(\Omega)$ and let $\gamma=\partial\Omega$ be such that $I=\int_\gamma \underline{g}\cdot d\underline{x}\ne 0$. Show that in this case (2.1.9) is not, in general, soluble if the initial datum is $\underline{u}_0=\underline{0}$. Find an example. (*Idea:* Let Ω be a disk of radius r and let $\underline{g}(\underline{x}) \stackrel{def}{=} \underline{\omega}\wedge\underline{x}$ with $\underline{\omega}=\omega\underline{e}$ orthogonal to the disk. Then the normal component of $\underline{g}$ on γ is 0 so that the uniqueness of the solutions for the Neumann problem implies that $p=const$ hence $\underline{\partial}p=\underline{g}=\underline{0}$ but $I=2\pi r^2\omega$).

[2.1.9]: Consider a velocity field in the half space $z\ge 0$ with components

$$u_1=-y\,\chi(z)\,f(x^2+y^2),\qquad u_2=x\,\chi(z)\,f(x^2+y^2),\qquad u_3=0$$

where $\chi(z)=z$ per $z\le h$, $h>0$ prefixed, and $\chi(z)\equiv 0$ for $z>2h$, while $f(r^2)=0$ for $r>R$, with $R>0$ prefixed. Show that $\underline{u}$ has zero divergence but (2.1.9) is no soluble, for χ, f generic. (*Idea:* Note that $\Delta\underline{u}=\chi(z)\,(8f'+4r^2f'')(-y,x,0)$, if $r^2\equiv x^2+y^2$ and $z<h$. Hence $\underline{u}$ and $\Delta\underline{u}$ both vanish on the boundary $z=0$. Furthermore

$$-\underline{\partial}\cdot(\underline{u}\cdot\underline{\partial}\,\underline{u}) \equiv \sigma(z^2,r^2) \equiv\, (f(r^2)^2+2f(r^2)f'(r^2)\,r^2)\chi(z)^2$$

Hence (assuming $\underline{g}=\underline{0}$) the gradient $\underline{\partial}p$ of p vanishes on $z=0$ and the corresponding Neumann problem can be solved by the method of images. The potential p is then the electrostatic potential generated by a charge distribution σ with cylindrical symmetry around the z–axis and with center at the origin and with reflection symmetry across the plane $z=0$.
The at large distance the electric field can be computed, to leading order, in R^{-1} and one sees that its component tangential to the plane $z=0$ does not vanish and it has order R^{-4} if $\int r^2\sigma(r^2,z^2)r\,dr\,dz\ne 0$ (in electrostatic terms one can say that the electric field is dominated at large distance by the lowest nonzero dipole moment which is in the present case the quadrupole, yielding therefore a field proportional to R^{-4}). Check that the dipole moment is identically zero, for any f and χ, beginning with the remark that $\sigma\equiv\chi^2\,\frac{\partial r^2f^2}{\partial\,r^2}$.)

[2.1.10]: Consider the heat equation in $[-\pi,\pi]$, $\dot T=T''$, with periodic boundary conditions and analytic initial datum $T_0(x)$ with Fourier transform $\hat T_0(\omega)$, $\omega=0,\pm1,\ldots$; hence there are constants $\tau,b>0$ such that $|\hat T_0(\omega)|\le\tau e^{-b|\omega|}$. Show that the L_2–norm of $(1-\omega^2t/k)^k\hat T_0(\omega)$ may diverge as $k\to\infty$ although for each ω one has $(1-\omega^2t/k)^k\hat T_0(\omega)\xrightarrow[k\to\infty]{} e^{-\omega^2t}\hat T_0(\omega)$, *c.f.r.* (2.1.25). This implies that in general even if T_0 is analytic the method of approximation in (2.1.22) *does not* converge to the solution neither in the sense of L_2 nor pointwise and staying uniformly bounded. (*Idea:* Take $\hat T_0(k)\equiv\tau e^{-b|\omega|}$ and estimate the sum $\sum_\omega(1-\omega^2t/k)^{2k}\tau^2e^{-2b|\omega|}$ by the single term with $\omega^2t=3k$.)

[2.1.11]: In the context of [2.1.10] show that even if $\hat T_0(\omega)=\tau e^{-b\omega^2}$ still one cannot have L_2 convergence of $T_k(x)$ for all times $t>0$. (*Idea:* Same as previous.)

Bibliography: [Bo79].

§2.2 Another class of empirical algorithms. Spectral method. Stokes problem. Gyroscopic analogy.

A method substantially different from the one discussed in §2.1 is the "*cut–off*" or "*spectral*" method. The name originates from the use of the representation of $\underline{u}$ on the basis generated by the Laplace operator on $X_{\rm rot}(\Omega)$, *c.f.r.* (1.6.16): it is, therefore, a method associated with the spectrum of this operator.

(A) Periodic boundary conditions: spectral algorithm and "reduction" to an ordinary differential equation.

We shall first examine a fluid occupying the d–dimensional torus T^d, *i.e.* an incompressible fluid enclosed in a cubic container with periodic boundary conditions ("opposite sides identified").

In this case the velocity and pressure fields, assumed regular, will admit a Fourier representation which can be regarded, obviously, as the expansion of the fields on the plane waves basis or, equivalently, on the basis generated by the eigenvectors[1] of the Laplace operator on T^d, *i.e.*

$$\underline{u}(\underline{\xi},t)=\sum_{\underline{k}}\hat{\underline{u}}_{\underline{k}}e^{i\underline{k}\cdot\underline{\xi}},\qquad p(\underline{\xi},t)=\sum_{\underline{k}}p_{\underline{k}}e^{i\underline{k}\cdot\underline{\xi}}\tag{2.2.1}$$

where $\underline{k}=2\pi L^{-1}\underline{n}$ with $L=$ side of the container and $\underline{n}$ is a vector with integer components. We adopt the following convention for the Fourier transforms

$$\begin{aligned}&\underline{u}(\underline{x})=\sum_{\underline{k}}e^{i\underline{k}\cdot\underline{x}}\hat{\underline{u}}_{\underline{k}},\qquad \hat{\underline{u}}(\underline{k})=L^{-d}\int_{T^d}e^{-i\underline{k}\cdot\underline{x}}\underline{u}(\underline{x})\,d\underline{x}\\ &||\underline{u}||_2^2\equiv|\underline{u}|_2^2\equiv||\underline{u}||_{L_2(T^d)}=L^d\sum_{\underline{k}}|\hat{\underline{u}}_{\underline{k}}|^2=\int_{T^d}|\underline{u}(\underline{x})|^2\,d\underline{x}\end{aligned}\tag{2.2.2}$$

The incompressibility condition (*i.e.* zero divergence), in the case $d=3$, requires that for $\underline{k}\neq\underline{0}$

$$\hat{\underline{u}}_{\underline{k}}=\gamma^1_{\underline{k}}\,\underline{e}^1_{\underline{k}}+\gamma^2_{\underline{k}}\,\underline{e}^2_{\underline{k}}\equiv\underline{\gamma}_{\underline{k}}\tag{2.2.3}$$

where $\underline{e}^1_{\underline{k}}$, $\underline{e}^2_{\underline{k}}$ are two unit vectors orthogonal to $\underline{k}$. In the case $d=2$ it must be $\hat{\underline{u}}=\gamma_{\underline{k}}\underline{k}^\perp/|\underline{k}|$, if $\underline{k}=(k_1,k_2)$ and $\underline{k}^\perp=(k_2,-k_1)$.

[1] We consider in this section only real vector fields: nevertheless it is occasionally convenient to express them in terms of complex plane waves rather than using the sines and cosines waves. We shall not discuss further this matter of notation.

We consider only incompressible Euler and NS equations in which the applied external force $\underline{g}(\underline{\xi})$ has zero average $L^{-d}\int \underline{g}(\underline{\xi})\,d\underline{\xi} = \underline{0}$: this is to exclude that the center of mass of the fluid accelerates uniformly (note that with periodic boundary conditions the center of mass will move as a body of mass equal to that of the fluid subject to the sum of the volume forces); hence

$$\partial_t \int \rho \underline{u}\, d\underline{\xi} = \rho \int \underline{g}\, d\underline{\xi} = \underline{0} \tag{2.2.4}$$

and, possibly changing reference frame, it is not restrictive to suppose $\int \underline{u}\, d\underline{\xi} = \underline{0}$. Likewise we can fix the arbitrary additive constant in the pressure so that $\int p\, d\underline{\xi} \equiv 0$.

With such conventions and hypotheses we can rewrite the (2.2.1) as

$$\underline{u}(\underline{\xi},t) = \sum_{\underline{k}\neq\underline{0}} \underline{\gamma}_{\underline{k}}(t) e^{i\underline{k}\cdot\underline{\xi}}, \qquad p(\underline{\xi},t) = \sum_{\underline{k}\neq\underline{0}} p_{\underline{k}} e^{i\underline{k}\cdot\underline{\xi}} \tag{2.2.5}$$

and the Euler or NS equations become ordinary equations for the components $\underline{\gamma}_{\underline{k}}$ of the field $\underline{u}$. To write them explicitly remark that

$$\underset{\sim}{u}(\underline{\xi}) \cdot \underset{\sim}{\partial}\, \underline{u}(\underline{\xi}) = \sum_{\underline{h},\underline{k}} e^{i(\underline{h}+\underline{k})\underline{\xi}} (\underset{\sim}{\gamma}_{\underline{h}} \cdot i\underset{\sim}{k})\, \underline{\gamma}_{\underline{k}} \tag{2.2.6}$$

Furthermore define, for $\underline{k} \neq \underline{0}$, the operator $\Pi_{\underline{k}}$ of orthogonal projection of R^3 on the plane orthogonal to $\underline{k}$ by

$$(\prod\nolimits_{\underline{k}} \underline{w})_i = w_i - \frac{\underline{w}\cdot\underline{k}}{\underline{k}^2} k_i \tag{2.2.7}$$

and note the following obvious identity

$$\underset{\sim}{\gamma}_{\underline{k}_1} \cdot \underset{\sim}{k}_2 \underline{\gamma}_{\underline{k}_2} \equiv (\underset{\sim}{\gamma}_{\underline{k}_1} \cdot \underset{\sim}{k}_2) \prod\nolimits_{\underline{k}_1+\underline{k}_2} \underline{\gamma}_{\underline{k}_2} + (\underset{\sim}{\gamma}_{\underline{k}_1} \cdot \underset{\sim}{k}_2)(1 - \prod\nolimits_{\underline{k}_1+\underline{k}_2}) \underline{\gamma}_{\underline{k}_2} \tag{2.2.8}$$

Consequently we see that the partial differential equations

$$\underline{\partial} \cdot \underline{u} = 0\,, \quad \dot{\underline{u}} + \underset{\sim}{u} \cdot \underset{\sim}{\partial} \underline{u} = -\rho^{-1} \underline{\partial} p + \underline{g} + \nu \Delta \underline{u} \tag{2.2.9}$$

can be written as the ordinary differential equations

$$\begin{aligned} \dot{\underline{\gamma}}_{\underline{k}} &= -\nu \underline{k}^2 \gamma_{\underline{k}} - i \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} (\underset{\sim}{\gamma}_{\underline{k}_1} \cdot \underset{\sim}{k}_2) \prod\nolimits_{\underline{k}} \underline{\gamma}_{\underline{k}_2} + \prod\nolimits_{\underline{k}} \hat{\underline{g}}_{\underline{k}} \\ p_{\underline{k}} &= -\rho \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \frac{1}{\underline{k}^2} (\underset{\sim}{\gamma}_{\underline{k}_1} \cdot \underset{\sim}{k}_2)(\underline{k} \cdot \underline{\gamma}_{\underline{k}_2}) - \frac{i\rho}{\underline{k}^2} \hat{\underline{g}}_{\underline{k}} \cdot \underline{k} \end{aligned} \tag{2.2.10}$$

Therefore the pressure "disappears" and the equations for the "essential components" of the fields describing our system become

$$\dot{\underline{\gamma}}_{\underline{k}} = -\nu \underline{k}^2 \gamma_{\underline{k}} - i \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} (\underset{\sim}{\gamma}_{\underline{k}_1} \cdot \underset{\sim}{k}_2) \prod\nolimits_{\underline{k}} \underline{\gamma}_{\underline{k}_2} + \hat{\underline{g}}_{\underline{k}}, \qquad \underline{\gamma}_{\underline{k}}(0) = \underline{\gamma}^0_{\underline{k}} \tag{2.2.11}$$

having assumed that $\Pi_{\underline{k}}\hat{g}_{\underline{k}} \equiv \hat{g}_{\underline{k}}$, since the gradient part of $\hat{g}_{\underline{k}}$, (*i.e.* the component of $\hat{g}_{\underline{k}}$ parallel to $\underline{k}$), can be included in the pressure, as we see from the second of the (2.2.10). To the (2.2.11) *we must always add the reality condition* for $\underline{u}$, *i.e.* $\underline{\gamma}_{\underline{k}} = \overline{\underline{\gamma}}_{-\underline{k}}$: we shall always assume such relation, *c.f.r.* footnote [1] above.

The Euler equations are simply obtained by setting $\nu = 0$.

If $\nu > 0$ the friction term gives rise to very large coefficients as $\underline{k}^2 \to \infty$ and therefore it will possibly generate problems in solution algorithms. It is therefore convenient to rewrite (2.2.11) as

$$\underline{\gamma}_{\underline{k}}(t) = e^{-\nu\underline{k}^2 t}\underline{\gamma}^0_{\underline{k}} + \int_0^t e^{-\nu\underline{k}^2(t-\tau)}\Big(\hat{g}_{\underline{k}} - i\sum_{\underline{k}_1+\underline{k}_2=\underline{k}}(\underline{\gamma}_{\underline{k}_1}(\tau)\cdot \underset{\sim}{k}_2)\prod_{\underline{k}}\underline{\gamma}_{\underline{k}_2}(\tau)\Big)d\tau \tag{2.2.12}$$

in which we see that the friction term is, in fact, a term that can help constructing solution algorithms, because it tends to "eliminate" the components with $|\underline{k}| \gg 1/\sqrt{\nu}$ *i.e.* the "*short wave*" components, also called the "*ultraviolet*" components, also called the "*short wave*" components of the velocity field.

Note that (2.2.12) suggests naturally a solution algorithm

$$\underline{\gamma}^{(n)}_{\underline{k}} = e^{-\nu\underline{k}^2 t}\underline{\gamma}^0_{\underline{k}} + t_0\sum_{m=0}^{n-1} e^{-\nu k^2 t_0(n-m)}\Big(\hat{\underline{g}}_{\underline{k}} - i\sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \underline{\gamma}^{(m)}_{\underline{k}_1}\cdot \underset{\sim}{k}_2\,\Pi_{\underline{k}}\gamma^{(m)}_{\underline{k}_2}\Big) \tag{2.2.13}$$

where the $\gamma^{(m)}$ are computed in mt_0; and therefore

$$\lim_{\substack{n\to\infty\\ nt_0=t}} \underline{\gamma}^{(n)}_{\underline{k}} = \underline{\gamma}_{\underline{k}}(t) \tag{2.2.14}$$

should be a solution of the equation.

This proposal can be subject to criticism of the type analyzed in the preceding §2.1 and we can expect that it might be correct only under further regularity hypotheses on $\underline{u}^0$, *i.e.* on $\underline{\gamma}^0$.

The (2.2.13) requires summing infinitely many terms. For concrete applications it is therefore necessary to find an approximation involving only sums of finitely many terms. One of the most followed methods it to introduce a *ultraviolet cut–off*: this means introducing a parameter N (the *cut–off*) and to constrain $\underline{k}, \underline{k}_1, \underline{k}_2$ in (2.2.12), or in (2.2.11), to be $\le N$.

Thus one obtains a system of finitely many ordinary equations and their solution $\underline{\gamma}^N_{\underline{k}}(t)$ *should* tend to a solution of the NS equations in the limit $N \to \infty$.

In the case of equation (2.2.13) with ultraviolet cut–off N we denote the approximate solution $\underline{\gamma}^{(n)N}_{\underline{k}}(t)$ and in the limit $N \to \infty, n \to \infty, nt_0 = t$ one shoud get a solution of the NS equations. The order in which the above limits have to be taken should have no consequences on the result, or it should be prescribed by the theory of the convergence: but it is not *a priori* clear that the limit really exists, nor that the solution to the equation is can be actually built in this way.

The simplicity of this algorithm, compared to those of §2.1 should be ascribed mainly to the boundary conditions that we are using. The algorithm name of *spectral method* will become more justified when we shall generalize it to the case of non periodic boundary conditions.

The algorithm has a great conceptual and practical advantage which makes it one of the most used algorithms in the numerical solutions of the Euler or NS equations. Unlike the method in §2.1 this algorithm makes manifest that viscosity appears explicitly as a damping factor on the velocity components with large wave number and rather than appearing as a "large" factor ($\sim \nu\underline{k}^2$) it appears as a "small" factor ($\sim e^{-\tau\underline{k}^2\nu}$).

(B) Spectral method in a domain Ω with boundary and the boundary conditions problem.

We shall now build a cut–off algorithm also in the case of a bounded domain Ω with a (smooth) boundary.

We note that the real reason why we succeed at "exponentiating" terms containing viscosity is that the velocity field has been developed in eigenfunctions of the Laplace operator *which is the operator associated with the linear viscous terms of the NS–equation*.

The case of periodic boundary conditions has been very simple, because in absence of boundary it is possible to find a basis for the divergenceless fields $\underline{u}$ which is at the same time a basis of eigenvectors for the Laplace operator Δ appearing in the friction term. In presence of a boundary it will not in general be possible to find eigenvalues of the Laplace operator which have zero divergence and which *at the same time* also vanish on the boundary (*i.e.* there are no eigenvectors, in general, of the Laplace operator with Dirichlet boundary conditions and zero divergence).

However if we define the "divergenceless Laplace operator" as the operator on $X_{\rm rot}$ defined by the quadratic form on $X^0_{\rm rot}$ (*c.f.r.* §1.6, (1.6.16))

$$D(\underline{u}) = \int_\Omega (\underline{\partial}\, \underset{\sim}{u})^2\, dx \tag{2.2.15}$$

one can show the following theorem (*c.f.r.* the problems at the end of the section where the proof is described)

Theorem *(spectral theory of the Laplace operator for divergenceless fields): In the space $X_{\rm rot}$, (c.f.r. (1.6.16)), there is an orthonormal basis of vectors satisfying:*
(1) $\underline{u}_j \in C^\infty(\Omega)\,, \quad \int_\Omega \underline{u}_i \cdot \underline{u}_j = \delta_{ij}$

(2) $\underline{\partial}\cdot\underline{u}_j = 0 \quad in\ \Omega$
(3) there is $\mu_j \in C^\infty(\Omega)$ *and* $\lambda_j > 0$ *such that:* (2.2.16)

$$\Delta\underline{u}_j - \underline{\partial}\,\mu_j = -\lambda_j^2\underline{u}_j \quad in\ \Omega$$

(4) $\underline{u}_j = 0 \quad in\ \partial\Omega$
(5) there are constants $\alpha, c, c', c_k > 0$ *such that*

$$c\, j^{2/d} \le |\lambda_j| \le c'\, j^{2/d}, \qquad |\partial^k \underline{u}_j(\underline{x})| \le c_k j^{\alpha+k/d} \tag{2.2.17}$$

for all $\underline{\underline{x}} \in \Omega$*, if* $d = 2, 3$ *is the space dimension.*

Then each divergenceless datum $\underline{u} \in X_{\rm rot}(\Omega)$ will be written as

$$\underline{u}(\underline{\xi}, t) = \sum_{j=1}^\infty \gamma_j(t)\,\underline{u}_j(\underline{\xi}) \tag{2.2.18}$$

and therefore we can express in terms of the γ_j the results of the actions on $\underline{u}$ of the operators appearing in the Euler and Navier–Stokes equations.

If $\Pi_{\rm grad}$ and $\Pi_{\rm rot} = 1 - \Pi_{\rm grad}$ are the projection operators on the spaces $X_{\rm rot}^\perp = X_{\rm grad}$ and on $X_{\rm rot}$, *c.f.r.* §1.6, the actions of the Laplace operator and of the nonlinear *transport operator* are respectively

$$\begin{aligned}
\Delta\underline{u} &= \sum_{j=1}^\infty -\lambda_j^2\,\gamma_j\underline{u}_j(\underline{\xi}) - \underline{\partial}\,\Big(\sum_{j=1}^\infty \mu_j(\underline{\xi})\gamma_j\Big)\\
\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{u} &= \sum_{j_1,j_2=1}^\infty \gamma_{j_1}\gamma_{j_2}\,\underset{\sim}{u}_{j_1}\cdot\underset{\sim}{\partial}\,\underline{u}_{j_2} = \\
&= \sum_{j_1,j_2=1}^\infty \gamma_{j_1}\gamma_{j_2}\Pi_{\rm rot}\,\big(\underset{\sim}{u}_{j_1}\cdot\underset{\sim}{\partial}\,\underline{u}_{j_2}\big) + \sum_{j_1,j_2=1}^\infty \gamma_{j_1}\gamma_{j_2}\Pi_{\rm grad}\big(\underset{\sim}{u}_{j_1}\cdot\underset{\sim}{\partial}\,\underline{u}_{j_2}\big)
\end{aligned} \tag{2.2.19}$$

and the NS–equation becomes, if we set $\prod_{\rm grad}\big(\underset{\sim}{u}_{j_1}\cdot\underset{\sim}{\partial}\big)\,\underline{u}_{j_2} \equiv \underline{\partial}\,\pi_{j_1 j_2}$,

$$\begin{aligned}
\dot\gamma_j &= -\,\nu\lambda_j^2\gamma_j - \sum_{j_1,j_2=1}^\infty \gamma_{j_1}\gamma_{j_2}\big\langle(\underset{\sim}{u}_{j_1}\cdot\underset{\sim}{\partial}\,\underline{u}_{j_2}),\,\underline{u}_j\big\rangle + g_j\\
\rho^{-1}p &= -\,\nu\sum_{j=1}^\infty \mu_j(\underline{\xi})\,\gamma_j - \sum_{j_1 j_2}\gamma_{j_1}\gamma_{j_2}\,\pi_{j_1 j_2}(\underline{\xi})
\end{aligned} \tag{2.2.20}$$

where $\big\langle\cdot,\cdot\big\rangle$ denotes the scalar product $\langle f, g\rangle = \int_\Omega f(\underline{x})g(\underline{x})\,d\underline{x}$. The (2.2.20) can be written

$$\begin{aligned}
\gamma_j(t) &= e^{-\nu\lambda_j^2 t}\gamma_j^0 + \int_0^t e^{-\nu\lambda_j^2(t-\tau)}\Big(g_j - \sum_{j_1,j_2=1}^\infty \gamma_{j_1}(\tau)\gamma_{j_2}(\tau)C_j^{j_1 j_2}\Big)\,d\tau\\
C_j^{j_1\,j_2} &= \int_\Omega(\underset{\sim}{u}_{j_1}\cdot\underset{\sim}{\partial}\underline{u}_{j_2})\cdot\underline{u}_j\,d\underline{\xi} \equiv \big\langle(\underset{\sim}{u}_{j_1}\cdot\underset{\sim}{\partial}\,\underline{u}_{j_2}),\underline{u}_j\big\rangle
\end{aligned} \tag{2.2.21}$$

This shows that also in the case in which Ω has a boundary it is still possible to write the NS–equations so that viscosity appears as a "smallness" factor rather than as a "large" additive term.

One can also derive a discretization of the Euler equations by setting $\nu = 0$ in (2.2.21). However the boundary condition is now $\underline{u} \cdot \underline{n} = 0$ rather than $\underline{u} = 0$ and therefore we cannot expect that the series for $\underline{u} = \sum_{j=1}^{\infty} \gamma_j \underline{u}_j$ is "well" convergent because, otherwise, this would imply $\underline{u} = 0$ on the boundary, since $\underline{u}_j = 0$ on the boundary. *It would therefore not be a good idea to use the basis above to represent solutions of the Euler equations in the same domain.*

This is a general problem because the Euler and NS–equations are studied, usually, by imposing different boundary conditions. The limit $\nu \to 0$ in which, naively, the NS–equations should "reduce" to the Euler equations *must be* a singular limit and the convergence of the solutions of the NS–equations to solutions of the Euler equation when $\nu \to 0$ must be quite *improper* near the boundary $\partial\Omega$, where interesting surface phenomena will necessarily take place. Of course the case of periodic boundary conditions is *the* remarkable exception.

The (2.2.20),(2.2.21) can be treated as the analogous (2.2.10) and (2.2.12) and reduced, with an ultraviolet cut–off to a finite number of equations, generating a general spectral algorithm.

(C) The Stokes problem.

One calls "*Stokes problem*" the NS–equation linearized around $\underline{u} = \underline{0}$, *c.f.r.* the problems of §1.2

$$\begin{aligned} \dot{\underline{u}} =& -\rho^{-1}\underline{\partial} p + \nu\Delta\underline{u} + \underline{g}, \qquad \underline{\partial}\cdot\underline{u} = 0 \qquad && \text{in } \Omega \\ \underline{u} =& \,\underline{0}, && \text{in } \partial\Omega \\ \underline{u}|_{t=0} =& \,\underline{u}^0 \in X_{\text{rot}}(\Omega) \end{aligned} \tag{2.2.22}$$

and we look for $C^\infty(\Omega\times(0,+\infty))$–solutions that for $t \to 0$ enjoy the property $\underline{u} \to \underline{u}^0$ at least in the sense of $L_2(\Omega)$ (see footnote [1] in (E) of §2.1), *i.e.* in the sense that the mean square deviation of $\underline{u}$ from $\underline{u}^0$, *i.e.* $\int |\underline{u} - \underline{u}^0|^2 \, d\underline{\underline{x}}$, tends to zero with t). We shall take, to simplify, $\underline{g} = \underline{0}$.

The theorem in (B) above allows us to obtain a complete solution of the problem. Indeed we develop $\underline{u}^0$ on the basis $\underline{u}_1, \underline{u}_2, \ldots,$ *c.f.r.* (2.2.16)

$$\underline{u}^0(\underline{x}) = \sum_{j=1}^{\infty} \gamma_j^0 \, \underline{u}_j(\underline{x}) \tag{2.2.23}$$

and we immediately check that the solution is

$$\underline{u}(\underline{x},t) = \sum_{j=1}^{\infty} \gamma_j^0 e^{-\nu\lambda_j^2 t} \underline{u}_j(\underline{x}), \qquad p(\underline{x},t) = -\nu\rho \sum_{j=1}^{\infty} \gamma_j^0 e^{-\nu\lambda_j^2 t} \mu_j(\underline{x}) \tag{2.2.24}$$

From the properties (1)%(5) of the theorem in (B) above it follows that $\underline{u} \in C^\infty(\Omega \times (0, +\infty))$ and that $\underline{\partial} \cdot \underline{u} = 0$ for $t > 0$.

It is easy to see that this solution is unique. One also realizes the strict analogy with the heat equation of which the Stokes equation can be regarded as a "vectorial" version.

In particular it can happen that, if $\underline{u}^0 \in X_{\text{rot}}(\Omega)$ but $\underline{u}^0$ does not have a series representation like (2.2.23) with coefficients γ_j^0 rapidly vanishing as $j \to \infty$ (for instance because $\underline{u}_0$ "does not match well" the boundary conditions so that all we can say is that $\sum_j |\gamma_j^0|^2 < \infty$), then $\nu\Delta\underline{u}^0$ could differ from a field vanishing on the boundary by more than we might expect, *i.e.* by an amount that is not "just" the gradient of a scalar field p: in fact the Neumann problem that should determine p is over determined and it can turn out to be impossible to solve it, as in the cases discussed in §1.6, §2.1 in connection with examples of the same pathology for the heat equation.

The pathology manifests itself *only* at $t = 0$ and it can be explained as in the heat equation case (via a physical model for the boundary condition as given, for instance, by the auxiliary friction method in (C) of §2.1). Obviously in (2.2.22) this problem shows up only at $t = 0$: if $t > 0$ in fact the (2.2.24) show that the boundary condition is strictly satisfied: and the over determined Neumann problem for p *becomes necessarily compatible* and has the second of the (2.2.24) as a solution, which at $t = 0$ might no longer make sense because of the possibly poor convergence of the series.

(D) Comments:

(1) Note that the spectral method for the NS–equations induces us into believing that at least for $t > 0$ the boundary condition is satisfied by the solutions (if existent): one could expect that friction implies that the coefficients $\gamma_j(t)$ tend to zero for $j \to \infty$ much faster than they do at $t = 0$, thanks to the coefficients $e^{-\lambda_j^2\nu(t-\tau)}$. Hence the series (2.2.18) *should be* well convergent and, therefore, its sum should respect the boundary conditions which are automatically satisfied term by term in the series.

(2) *However* we shall see that the argument just given, which is essentially correct in the case of the heat equation (discussed in §2.1) and in the Stokes equation case, becomes now much more delicate and, mainly, if $d = 3$, it is no longer correct, basically because of the non linear terms in the transport equations: *c.f.r.* the analysis in §3.2.

(E) Gyroscopic analogy in $d = 2$.

The NS equations in $d = 2$–dimensions can be put in a form that closely reminds us of the rigid body equations of motion. The NS–equations, (2.2.10), with $\underline{g} = \underline{0}$ for simplicity, can be written in terms of the *scalar* observables $\gamma_{\underline{k}_1}$, $\gamma_{\underline{k}_2}$, $\gamma_{\underline{k}_3}$, related to the vector observable $\underline{\gamma}_{\underline{k}}$ via the $\underline{\gamma}_{\underline{k}} = \gamma_{\underline{k}}\,\underline{k}^\perp/|\underline{k}|$ because, if $d = 2$, the zero divergence property allows us to express $\underline{\gamma}_{\underline{k}}$ in terms of scalar quantities $\gamma_{\underline{k}}$:

$$\underline{\gamma}_{\underline{k}} = \gamma_{\underline{k}} \frac{\underline{k}^\perp}{|\underline{k}|}, \qquad \gamma_{\underline{k}} = -\,\overline{\gamma}_{-\underline{k}}, \qquad \text{if} \quad \underline{k}^\perp = (k_2, -k_1), \qquad \underline{k} = (k_1, k_2) \tag{2.2.25}$$

If $\underline{k}_1 + \underline{k}_2 + \underline{k}_3 = \underline{0}$ we then note that the equations (2.2.10) can be written

$$\begin{aligned}
\dot{\overline{\gamma}}_{\underline{k}_1} &= -\nu \underline{k}_1^2 \overline{\gamma}_{\underline{k}_1} - i\Big\{\gamma_{\underline{k}_2}\gamma_{\underline{k}_3} \frac{(\underline{k}_2^\perp \cdot \underline{k}_3)(\underline{k}_3^\perp \cdot \underline{k}_1^\perp)}{|k_1||k_2|k_3|} + (2 \leftrightarrow 3)\Big\} + ..\\
\dot{\overline{\gamma}}_{\underline{k}_2} &= -\nu \underline{k}_2^2 \overline{\gamma}_{\underline{k}_2} - i\Big\{\gamma_{\underline{k}_3}\gamma_{\underline{k}_1} \frac{(\underline{k}_3^\perp \cdot \underline{k}_1)(\underline{k}_1^\perp \cdot \underline{k}_2^\perp)}{|\underline{k}_1||\underline{k}_2||\underline{k}_3|} + (1 \leftrightarrow 3)\Big\} + .\\
\dot{\overline{\gamma}}_{\underline{k}_3} &= -\nu \underline{k}_3^2 \overline{\gamma}_{\underline{k}_3} - i\Big\{\gamma_{\underline{k}_1}\gamma_{\underline{k}_2} \frac{(\underline{k}_1^\perp \cdot \underline{k}_2)(\underline{k}_2^\perp \cdot \underline{k}_3^\perp)}{|\underline{k}_1|\;|\underline{k}_2|\;|\underline{k}_3|} + (1 \leftrightarrow 2)\Big\} + ..
\end{aligned} \tag{2.2.26}$$

Note the symmetry properties

$$\underline{k}_1^\perp \cdot \underline{k}_2 = \underline{k}_2^\perp \cdot \underline{k}_3 = \underline{k}_3^\perp \cdot \underline{k}_1 \stackrel{def}{=} a(\underline{k}_1, \underline{k}_2, \underline{k}_3) \tag{2.2.27}$$

with $a(\underline{k}_1, \underline{k}_2, \underline{k}_3) \equiv -a$ where a is $\pm$ twice the area of the triangle formed by the vectors $\underline{k}_1, \underline{k}_2, \underline{k}_3$ (it is a symmetric function under permutations of $\underline{k}_1, \underline{k}_2, \underline{k}_3$). The sign is $+$ if the triangle $\underline{k}_1\underline{k}_2\underline{k}_3$ is circled clockwise and $-$ otherwise.

Keeping this symmetry into account together with the relations

$$\begin{aligned}
&\underline{k}_1^\perp \cdot \underline{k}_2 = -\underline{k}_2^\perp \cdot \underline{k}_1, \qquad \underline{k}_3 = -\underline{k}_1 - \underline{k}_2\\
&\underline{k}_1^\perp \cdot \underline{k}_2^\perp \equiv \underline{k}_1 \dot{\underline{k}}_2, \qquad \text{hence, for instance,}\\
&\underline{k}_2^\perp \cdot \underline{k}_3^\perp - \underline{k}_1^\perp \cdot \underline{k}_3^\perp = \underline{k}_1^2 - \underline{k}_2^2
\end{aligned} \tag{2.2.28}$$

one finds (patience is required)

$$\begin{aligned}
\dot{\overline{\gamma}}_{\underline{k}_1} &= -\nu \underline{k}_1^2 \overline{\gamma}_{\underline{k}_1} - i(\underline{k}_3^2 - \underline{k}_2^2)\, \tilde{a}\, \gamma_{\underline{k}_2}\gamma_{\underline{k}_3} + \ldots\\
\dot{\overline{\gamma}}_{\underline{k}_2} &= -\nu \underline{k}_2^2 \overline{\gamma}_{\underline{k}_2} - i(\underline{k}_1^2 - \underline{k}_3^2)\, \tilde{a}\, \gamma_{\underline{k}_1}\gamma_{\underline{k}_3} + \ldots\\
\dot{\overline{\gamma}}_{\underline{k}_3} &= -\nu \underline{k}_3^2 \overline{\gamma}_{\underline{k}_3} - i(\underline{k}_2^2 - \underline{k}_1^2)\, \tilde{a}\, \gamma_{\underline{k}_1}\gamma_{\underline{k}_2} + \ldots
\end{aligned} \tag{2.2.29}$$

where $\tilde{a} = a(\underline{k}_1, \underline{k}_2, \underline{k}_3)/|\underline{k}_1|\,|\underline{k}_2|\,|\underline{k}_3|$.

These equations are analogous to those for the angular velocity of a solid with a fixed point. The analogy becomes even more clear in the variables $\omega_{\underline{k}} = \gamma_{\underline{k}}/|\underline{k}|$ which obey the equations

$$\underline{k}_1^2\, \dot{\overline{\omega}}_{\underline{k}_1} = -\underline{k}_1^4 \nu\, \overline{\omega}_{\underline{k}_1} + (\underline{k}_2^2 - \underline{k}_3^2)\, a\, i\, \omega_{\underline{k}_2}\omega_{\underline{k}_3} + \ldots \text{ etc} \tag{2.2.30}$$

We see also an interesting property: namely every triple or "*triad*" $\underline{k}_1, \underline{k}_2, \underline{k}_3$ such that $\underline{k}_1 + \underline{k}_2 + \underline{k}_3 = \underline{0}$ contributes to the equations (2.2.30) in such a manner that, if the $\underline{\gamma}_{\underline{k}}$ relative to the other values of $\underline{k}$ (different from

$\pm\underline{k}_1, \pm\underline{k}_2, \pm\underline{k}_3$) were zero, the equations would describe the motion (*with friction*) of a "complex" (because the $\omega_{\underline{k}}$ are complex quantities) gyroscope.

Hence Euler equations can be interpreted as describing infinitely many coupled gyroscopes, each associated with a *triad* such that $\underline{k}_1+\underline{k}_2+\underline{k}_3=\underline{0}$: they are not independent and their coupling is described by the constraint that, if a vector $\underline{k}$ is common to two triads, then the $\omega_{\underline{k}}$'s, thought of as components of one or of the other gyroscope, *must be equal* (because $\omega_{\underline{k}}$ depends only upon $\underline{k}$ and not on which of the (infinitely many) triads the vector $\underline{k}$ it is regarded to belong to).

The motions of a complex gyroscope are not as simple as those of the ordinary gyroscopes, not even in absence of friction ($\nu\equiv 0$), and we understand also from this viewpoint the difficulty that we shall meet in the qualitative analysis of the properties of the solutions of the equations.

Note, finally, that the *single* complex gyroscope (*i.e.* described by the equations relative to a single triad) may admit motions that can be interpreted as motions of a system of "real" gyroscopes, even though writing $\omega_{\underline{k}}=\rho_{\underline{k}}e^{i\vartheta_{\underline{k}}}$ with $\rho_{\underline{k}}, \vartheta_{\underline{k}}$ real one finds that in general the phases $\vartheta_{\underline{k}}$ are not constant.

It is indeed easy to see that if the phases of the initial datum have special values then the phases remain constant and the $\rho_{\underline{k}}$ obey equations that are exactly like those obeyed by the three components of the three angular velocities of an ordinary gyroscope (hence in absence of friction they can be integrated by "*quadratures*"). For instance this happens if $\vartheta_j\equiv -3\pi/2$, see also (4.1.27) in §4.1.

(F) Gyroscopic analogy in $d=3$.

A gyroscopic analogy is possible, *c.f.r.* [Wa90], also in the $d=3$ case and it is based on the same identities introduced between (2.2.26) and (2.2.29) and on the new notion of *elicity*. We sketch it quickly here, leaving the details to the interested reader. In the case $d=3$, with $\underline{g}=\underline{0}$, given $\underline{k}$ we introduce, [Wa90], two *complex* mutually orthogonal unit vectors $\underline{h}_s(\underline{k})$, $s=\pm1$, also orthogonal to $\underline{k}$

$$\underline{h}_{s,\underline{k}}=\underline{v}_0(\underline{k})+is\,\underline{v}_1(\underline{k}),\qquad s=\pm1 \tag{2.2.31}$$

where $\underline{v}_0, \underline{v}_1$ are two mutually orthogonal *real* unit vectors orthogonal to $\underline{k}$ and, furthermore, such that $\underline{v}_0(-\underline{k})=\underline{v}_0(\underline{k})$ and $\underline{v}_1(-\underline{k})=-\underline{v}_1(\underline{k})$. In this way $\overline{\underline{h}}_{s,\underline{k}}=\underline{h}_{-s,\underline{k}}=\underline{h}_{s,-\underline{k}}$. Suppose, moreover, that the three vectors $\underline{v}_0, \underline{v}_1, \underline{k}$ form a counterclockwise triple.

The basis $\underline{h}_{-1,\underline{k}}, \underline{h}_{+1,\underline{k}}$ in R^3 will be called the *elicity base* and we shall say that the vector $\underline{h}_{s,\underline{k}}$ has elicity s. Then the Fourier components $\underline{\gamma}_{\kappa}$ of an arbitrary divergenceless velocity field $\underline{v}$ can be written as

$$\underline{\gamma}_{\underline{k}}=\sum_{s=\pm1}\gamma_{\underline{k},s}\,\underline{h}_{s,\underline{k}} \tag{2.2.32}$$

where $\gamma_{\underline{k},s}$ are scalar quantities such that $\gamma_{\underline{k},s}=\overline{\gamma}_{-\underline{k},s}$.

The NS–equations (2.2.10) become

$$\begin{aligned}\dot{\overline{\gamma}}_{\underline{k}_3,s_3} &= -\nu \underline{k}_3^2\,\overline{\gamma}_{\underline{k}_3,s_3} -\\ &- i \sum_{\underline{k}_1+\underline{k}_2+\underline{k}_3=\underline{0}} \gamma_{\underline{k}_1,s_1}\gamma_{\underline{k}_2,s_2}\,[\underline{h}_{s_1,\underline{k}_1}\cdot\underline{k}_2]\,[\underline{h}_{s_2,\underline{k}_2}\cdot\underline{h}_{s_3,\underline{k}_3}]\end{aligned} \tag{2.2.33}$$

The expression $[\underline{h}_{s_1,\underline{k}_1}\cdot\underline{k}_2]\,[\underline{h}_{s_2,\underline{k}_2}\cdot\underline{h}_{s_3,\underline{k}_3}]$ can be studied by noting that the vector $e^{-is\mu}\underline{h}_s$ can be obtained by rotating clockwise by μ the basis $v_0(\underline{k}), s\underline{v}_1(\underline{k})$: this remark allows us to reduce the calculation of this product to the same calculation in the planar case ($d = 2$).

Given the triangle $\underline{k}_1, \underline{k}_2, \underline{k}_3$ and the elicities s_1, s_2, s_3 and having established a Cartesian reference system on its plane, so that the triangle $\underline{k}_1\underline{k}_2\underline{k}_3$ is circled clockwise, we can find three angles μ_1, μ_2, μ_3 such that the clockwise rotation by $s_j\mu_j$ of the basis $\underline{v}_0(\underline{k}_j), s_j\underline{v}_1(\underline{k}_j)$ brings it into a basis $\underline{\tilde{v}}_0(\underline{k}_j), \underline{\tilde{v}}_1(\underline{k}_j)$ with $\underline{\tilde{v}}_0(\underline{k}_j)$ directed as the axis $\underline{k}_j^\perp$ orthogonal to $\underline{k}_j$ and lying in the plane of the triangle and with components (on this plane) $(-k_{j2}, k_{j1})$ if $\underline{k}_j = (k_{j1}, k_{j2})$. Then

$$[\underline{h}_{s_1,\underline{k}_1}\cdot\underline{k}_2]\,[\underline{h}_{s_2,\underline{k}_2}\cdot\underline{h}_{s_3,\underline{k}_3}] = e^{-i\tilde{\mu}}\frac{\underline{k}_1^\perp\cdot\underline{k}_2}{|\underline{k}_1|}\,\Big(\frac{\underline{k}_2^\perp\cdot\underline{k}_3^\perp}{|\underline{k}_2|\,|\underline{k}_3|} - s_2 s_3\Big) \tag{2.2.34}$$

where $\tilde{\mu} = s_1\mu_1 + s_2\mu_2 + s_3\mu_3$; and we see that we can use the expressions already obtained in the case $d = 2$. If a is defined by setting $a = -\underline{k}_1^\perp\cdot\underline{k}_2$ (twice the area of the triangle formed by the vectors $\underline{k}_j$) and if $\mu = -\frac{\pi}{2} + s_1\mu_1 + s_2\mu_2 + s_3\mu_3$ and $\omega_{s,\underline{k}} = |\underline{k}|^{-1}\gamma_{\underline{k},s}$ then by the identities noted in (2.2.27) and (2.2.28) one gets

$$\begin{aligned}&\underline{k}_1^2\,\dot{\overline{\omega}}_{\underline{k}_1,s_1} = -\underline{k}_1^4\nu\,\overline{\omega}_{\underline{k}_1,s_1} -\\ &-(\underline{k}_3^2 - \underline{k}_2^2 + (\underline{s}\wedge\underline{\kappa})_1\kappa_1\,\sigma)\,a\,e^{-i\mu}\,\omega_{\underline{k}_2,s_2}\omega_{\underline{k}_3,s_3} + \ldots \text{ etc}\end{aligned} \tag{2.2.35}$$

where $\sigma = s_1 s_2 s_3$ and $\underline{\kappa}, \underline{s}$ are defined by $\underline{\kappa} = (|\underline{k}_1|, |\underline{k}_2|, |\underline{k}_3|)$ and $\underline{s} = (s_1, s_2, s_3)$: which shows that, once more, the equations can be written in terms of *triads* as in the case $d = 2$. It is therefore still possible to give a "gyroscopic" interpretation to the Euler and NS equations.

The 2–dimensional equations can be obtained from the (2.2.35) simply by eliminating the labels s_j from the ω and setting $\sigma = 0$, because in this case the vectors $\underline{v}_1$ have to be replaced by $\underline{0}$.

We see, furthermore, that if Δ denotes the triad $(\underline{k}_1, s_1), (\underline{k}_2, s_2), (\underline{k}_3, s_3)$ then the quantities

$$E = \frac{1}{6}\sum_\Delta \sum_{\underline{k},s\in\Delta} |\underline{k}|^2|\omega_{\underline{k},s}|^2, \qquad \Omega = \frac{1}{3}\sum_\Delta \sum_{\underline{k},s\in\Delta} |\underline{k}|^4|\omega_{\underline{k},s}|^2 \tag{2.2.36}$$

are constants of motion in the case $\nu = 0, \underline{g} = \underline{0}$ and $d = 2$; in this case the index s should not be present, but we have used the three dimensional; notation for homogeneity purposes.

The first quantity is proportional to the kinetic energy and the second is proportional to the *enstrophy*[2] (*i.e.* to the total vorticity). The quantities E, Ω are sums of positive quantities and we shall see how this property will make them particularly useful in obtaining *a priori* estimates on the solutions of the Euler and NS equations.

Remark the mechanism by which the 2–dimensional fluids conserve energy and enstrophy: it is the same by which a solid with a fixed point conserves energy and angular momentum: here $\underline{k}_1^2$, $\underline{k}_2^2$, $\underline{k}_3^2$ play the roles of principal inertia moments.
In the corresponding case $d = 3$ the energy E is still conserved (because $\underline{s} \wedge \underline{\kappa} \cdot \underline{\kappa} = 0$) while, since in general $\sum_i \kappa_i^2 (\underline{s} \wedge \underline{\kappa})_i \kappa_i \neq 0$, the Ω is no longer a constant of motion.
Nevertheless in the case $d = 3$, always with $\nu = 0, \underline{g} = \underline{0}$, there is another constant of motion because the identity

$$\begin{aligned}&\big(s_3|\underline{k}_3|(\underline{k}_3^2 - \underline{k}_1^2) + \underline{k}_3^2(s_2|\underline{k}_2| - s_1|\underline{k}_1|)\big) + \\ &+ \big(s_2|\underline{k}_2|(\underline{k}_1^2 - \underline{k}_3^2) + \underline{k}_2^2(s_1|\underline{k}_1| - s_3|\underline{k}_3|)\big) + \\ &+ \big(s_1|\underline{k}_1|(\underline{k}_3^2 - \underline{k}_2^2) + \underline{k}_1^2(s_3|\underline{k}_3| - s_2|\underline{k}_2|)\big) \equiv 0\end{aligned} \tag{2.2.37}$$

together with (2.2.35) implies that

$$\tilde{\Omega} = \frac{1}{3} \sum_{\Delta} \sum_{\underline{k}, s} s\, |\underline{k}|\, |\gamma_{\underline{k}, s}|^2 \tag{2.2.38}$$

is a constant of motion. (as it can also be directly seen from the Euler equations by remarking that such quantity is proportional to $\int \underline{u} \cdot \underline{\partial} \wedge \underline{u}^\perp \, dx$, if $\underline{u}^\perp$ is the field with Fourier transform obtained by 90^o-rotation of $\underline{u}_{\underline{k}}$ around $\underline{k}$).

However $\tilde{\Omega}$ cannot be directly used in a priori estimates because it is the sum of quantities with sign not defined.

Remarks:

(1) Note that there can be velocity fields in which all the components have elicity $s = 1$ (or all $s = -1$); it then follows from the (2.2.35) that, given $K > 0$, there exist solutions of the Euler equations having the form $\underline{u}(\underline{x}) = \sum_{\alpha, |\underline{k}_\alpha| = K} e^{i \underline{k} \cdot \underline{x}} c_{\underline{k}} \underline{h}_+(\underline{k})$ where k_α denotes here the component α, $\alpha = 1, 2, 3$, of the vector $\underline{k}$ (which, therefore, has all the components with modulus equal to K). Note in fact that in this case it is $\underline{s} \wedge \underline{\kappa} \equiv 0$ besides $\underline{k}_i^2 = 3K^2$ hence $\underline{k}_i^2 - \underline{k}_j^2 = 0$.
(2) If we consider the NS equation (*i.e.* $\nu \neq 0$) in absence of external field the solutions in (1) are either identically zero or vanish exponentially.

---

[2] From $\varepsilon\nu$ ("inside") and $\sigma\tau\rho\varepsilon\varphi\omega$ ("turn around").

(3) In presence of an external field which also has Fourier components $\underline{g}_{\underline{k}}$ which do not vanish only for $\underline{k}_\alpha \equiv K$ one can find an exact, non zero, time independent solution.

Problems: interior and boundary regularity of solutions of elliptic equations and for Stokes equation.

Here we mainly present the theory of the Laplace operator on divergenceless field in a bounded convex region Ω with smooth boundary. The key idea that we follow is to reduce the problem to the case of in which Ω is instead a torus, where the problem is easy. This is an intuitive and alternative approach with respect to the classical ones, see for instance [Mi70], [LM72], [Ga82].

[2.2.1]: (*weak solutions*) Define a function $x, t \to T(x,t)$ to be a *weak solution* with periodic boundary conditions on $[a,b]$, of the heat equation, (2.1.12), "in the sense of the periodic $C^\infty(a,b)$–functions belonging to a set $\mathcal{P}$ of such functions dense in $L_2([a,b])$" if the function T is in $L_2([a,b])$ and if furthermore

$$\partial_t \int_a^b \varphi(x)\, T(x,t)\, dx - \int_a^b \varphi''(x)\, T(x,t)\, dx = 0, \qquad \text{for all } \varphi \in \mathcal{P}$$

We say that the initial datum of a weak solution is ϑ_0 if it is: $\int_a^b \varphi(x)\, \vartheta_0(x)\, dx = \lim_{t\to 0} \int_a^b \varphi(x)\, T(x,t)\, dx$ for each $\varphi \in \mathcal{P}$.
For instance $\mathcal{P}$ can be the set of *all* $C^\infty([a,b])$ and periodic functions: we say, in such case, that T is a solution "*in the sense of distributions*" in the variable x on the circle $[a,b]$. If $\mathcal{P}$ is the space of the trigonometric polynomials periodic in $[a,b]$ we say that T is the solution in the sense of trigonometric polynomials on $[a,b]$.
We say that a sequence $f_n \in L_2([a,b])$ tends *weakly* to $f \in L_2([a,b])$ in the sense $\mathcal{P}$ if $\lim_{n\to\infty} \int_a^b \varphi(x) f_n(x) dx = \int_a^b \varphi(x) f(x) dx$ for each $\varphi \in \mathcal{P}$.
Show that the heat equation on $[a,b]$ admits a unique weak solution in the sense of trigonometric polynomials (hence in the sense of distributions) for a given initial datum $\vartheta_0 \in C^\infty(a,b)$.
Extend the above notions (when possible) to the case of functions on a d–dimensional bounded domain Ω. (*Idea:* Choose $a = -\pi, b = \pi$ (for simplicity) and write the condition that T is a solution by choosing $\varphi(x) = e^{i\omega x}$, with ω integer.)

[2.2.2]: (*weak solutions and heat equation*) Let ϑ_0 be a C^∞–function with support in $[-a,a]$, $a < L/2$. Show that the algorithm (2.1.23) for the heat equation on a circle of length L (identified with the segment $[-L/2, L/2]$) produces a solution that converges weakly in the sense of trigonometric polynomials to the solution of the heat equation with periodic boundary conditions on $[-L/2, L/2]$. Compare this result with [2.1.10], [2.1.11]. (*Idea:* Let $\hat\vartheta_0(\omega)$, with ω integer multiple of $2\pi L^{-1}$, be the Fourier transform of the initial datum; and note that the Fourier transform of the approximation at time $t = kt_0$, with $t_0 > 0$ and k integer, is $\hat\vartheta_k(\omega) = \hat\vartheta_0(\omega)(1 - \frac{c\omega^2 t}{k})^k$, *c.f.r.* §2.1. The weak convergence becomes equivalent to the statement that $\hat\vartheta_k(\omega)$ tends, for each fixed ω and and for $k \to \infty, t_0 \to 0$ (with $kt_0 = t$), to $\hat\vartheta_0(\omega) e^{-c\omega^2 t}$.)

[2.2.3]: (*Weak solutions ambiguities*) In the context of [2.2.1] we see that the algorithm of [2.2.2], *c.f.r.* (2.1.23), produces a sequence of functions $\hat\vartheta_k(\omega)$ with Fourier transform $\vartheta_{t_0}(x,t)$, $t \equiv kt_0$, periodic on $[-L/2, L/2]$ which, thought of as an element of $L_2([-L/2, L/2])$ converges weakly in the sense of the trigonometric polynomials to a solution $T(x,t)$ of the heat equation on $[-L/2, L/2]$. Note that this happens *for any* $L > a$: *i.e.* in a sense the algorithm produces different weak solutions depending on which is the length of the periodic bar that we imagine to contain the initial heat. Convince oneself that this is not a contradiction, and that on the contrary it is a useful example to meditate on the caution that has to be used when considering the notion of weak solution

[2.2.4]: (*extension of a function to a periodic function with control of its L_2 norm*) Given $\underline{u} \in C_0^\infty(\Omega)$ we can extend it to a C^∞–periodic function on a cube T_Ω with side $L > 2\,\mathrm{diam}\,\Omega$ containing Ω and, as well, a translate Ω' of Ω such that $\Omega' \cap \Omega = \emptyset$. The extension can be done so that the extension vanishes outside Ω and Ω' and, furthermore, on the points of Ω' has a value *opposite* to the one that it has in the corresponding points of Ω. Then the extension, which will be denoted by $\underline{\tilde{u}}$, has a vanishing integral on the whole T_Ω. Show that, (by the definition of D, *c.f.r.* (2.2.15))

$$\int_\Omega |\underline{u}(\underline{x})|^2\, d\underline{x} \equiv \frac{1}{2}|\underline{\tilde{u}}|_2^2 \le \frac{L^2}{8\pi^2} D(\underline{\tilde{u}}) = \frac{L^2}{4\pi^2} D(\underline{u}) \qquad \text{per} \quad \underline{u} \in X^0_{\rm rot}$$

(*Idea:* Write the "norm", *i.e.* the square root of the integral of the square, of the extension of $\underline{u}$ to $L_2(T_\Omega)$ and the value of $D(\underline{\tilde{u}})$ by using the Fourier transform and remarking that $2||\underline{u}||^2_{L_2(\Omega)} \equiv ||\underline{u}||^2_{L_2(T_\Omega)}$, and treat in a similar way $D(\underline{\tilde{u}})$.)

[2.2.5]: (*a lower bound for the Dirichlet quadratic form of a solenoidal vector field*) Given a convex region Ω with analytic boundary $\partial\Omega$ consider the space $X_{\rm rot} \equiv \overline{X^0_{\rm rot}}$ closure in $L_2(\Omega)$ of the space $X^0_{\rm rot}$ of the divergenceless fields vanishing in a neighborhood of the boundary. Consider the quadratic form $(\underline{u},\underline{v})_D$ (called the "*Dirichlet form*") associated with the Laplace operator:

$$(\underline{u},\underline{v})_D = \int_\Omega \underset{\sim}{\partial u} \cdot \underline{\partial} \cdot \underset{\sim}{v}\, d\underline{x}, \qquad \text{and set} \qquad D(\underline{u}) = \int_\Omega (\underset{\sim}{\partial u})^2\, d\underline{x}$$

defined on $X^0_{\rm rot}$. Show that the greatest lower bound of $D(\underline{u})/|\underline{u}|_2^2$ on $X^0_{\rm rot}$ is strictly positive. Show that it is in fact $\ge (2\pi L^{-1})^2$ if L is twice the side of the smallest square containing Ω. (*Idea:* Note that the infimum is greater or equal to the infimum of $\int_Q (\underset{\sim}{\partial}\,\underline{u})^2 / \int_Q \underline{u}^2$ over all C^∞ periodic fields $\underline{u}$ defined and with zero average on a square domain Q containing Ω. Indeed every function in $C_0^\infty(\Omega)$ can be extended trivially, see [2.2.4], to a periodic function $\underline{\tilde{u}} \in C^\infty(Q)$ and if Q has side L one can obviously also request that it has zero average (by defining it as opposite to $\underline{u}$ in the points of the "copy" Ω' of Ω that we can imagine contained in Q and without intersection with Ω): write then $D(\underline{\tilde{u}})$ by using the Fourier transform.)

[2.2.6]: (*bounding in L_2 a solenoidal field with the L_2 norm of its rotation*) Show that in the context of [2.2.4] it is

$$|\underline{u}|_2^2 \le \frac{L^2}{4\pi^2} \int_\Omega (\mathrm{rot}\ \underline{u})^2\, d\underline{\xi}, \qquad \text{in } X^0_{\rm rot}$$

(*Idea:* Make use again of the Fourier transform as in [2.2.4], [2.2.5] and note that $\underline{\partial} \cdot \underline{\tilde{u}} = 0$ implies that $|\underline{k}|^2 |\hat{\underline{\tilde{u}}}(\underline{k})|^2 = |\underline{k} \wedge \hat{\underline{\tilde{u}}}(\underline{k})|^2$, and furthermore $|\underline{u}|_2^2 = \frac{1}{2}|\underline{\tilde{u}}|_2^2 \le \frac{1}{2}\frac{L^2}{4\pi} D(\underline{\tilde{u}}) = \frac{L^2}{4\pi} D(\underline{u})$.)

[2.2.7]: (*"compactness" of Dirichlet forms*) Let $\underline{u}_n \in X^0_{\rm rot}$ be a sequence such that $|\underline{u}_n|_2 = 1$ and $D(\underline{u}_n) \le C^2$, for some $C > 0$, and show the existence of a subsequence of $\underline{u}_n$ converging in L_2 to a limit. (*Idea:* Imagine $\underline{u}_n$ continued to a function $\underline{\tilde{u}}$ defined on T_Ω and changed in sign in Ω', as in [2.2.4], [2.2.5], [2.2.6]. Then the hint of [2.2.6] implies $|\hat{\underline{\tilde{u}}}_n(\underline{k})| \le \frac{C}{|\underline{k}|}\sqrt{\frac{2}{L^d}}$, with the convention (2.2.2) on Fourier transform. Let $\{n_i\}$ be a subsequence such that $\hat{\underline{\tilde{u}}}_{n_i}(\underline{k}) \xrightarrow[i\to\infty]{} \hat{\underline{\tilde{u}}}_\infty(\underline{k})$, $\forall\, \underline{k}$ (which exists because $\underline{k}$ takes countably many values); we see that given an arbitrary $N > 0$:

$$\begin{aligned}
&4|\underline{u}_{n_i} - \underline{u}_{n_j}|_2^2 \equiv |\underline{\tilde{u}}_{n_i} - \underline{\tilde{u}}_{n_j}|^2_{L_2(T_\Omega)} = L^d \sum_{\underline{k}} |\hat{\underline{\tilde{u}}}_{n_i}(\underline{k}) - \hat{\underline{\tilde{u}}}_{n_j}(\underline{k})|^2 \le \\
&\le L^d \sum_{|\underline{k}|\le N} |\hat{\underline{\tilde{u}}}_{n_i}(\underline{k}) - \hat{\underline{\tilde{u}}}_{n_j}(\underline{k})|^2 + \frac{L^d}{N^2} \sum_{|\underline{k}|>N} |\underline{k}|^2 |\hat{\underline{\tilde{u}}}_{n_i}(\underline{k}) - \hat{\underline{\tilde{u}}}_{n_j}(\underline{k})|^2 \qquad (2.2.39)\\
&\xrightarrow[i,j\to\infty]{} \le \frac{2}{N^2} \sup D(\underline{\tilde{u}}_{n_i}) \le \frac{4\,C^2}{N^2}
\end{aligned}$$

hence the arbitrariness of N implies that $\underline{u}_{n_i}$ is a Cauchy sequence in $L_2(T_\Omega)$, hence in $L_2(\Omega)$, which therefore converges to a limit $\underline{u}_\infty \in X_{\rm rot}$.)

[2.2.8]: (*existence of a minimizer for the Dirichlet form on solenoidal vector fields*) Let $\underline{u}_n \in X^0_{\rm rot}$, $n \ge 1$, be a sequence such that there is a $\underline{u}_0$ for which

$$|\underline{u}_n|_2 \equiv 1\,, \quad D(\underline{u}_n) \to \inf_{|\underline{u}|_2=1,\ \underline{u}\in X^0_{\rm rot}} D(\underline{u}) \equiv \lambda_0^2, \qquad |\underline{u}_n - \underline{u}_0|_2 \to 0 \qquad (2.2.40)$$

Show that $D(\underline{u}_n - \underline{u}_m) \xrightarrow[n,m\to\infty]{} 0$ hence, since $\sqrt{D(\underline{u})}$ is a metric, $\underline{u}_0$ is in the "*domain of the closure of the quadratic form*",[3] *i.e.* there is also the limit $D(\underline{u}_0) = \lim_{n\to\infty} D(\underline{u}_n)$ and $D(\underline{u}_0) = \lambda_0^2$. In other words the quadratic form (extended to the functions of its domain reaches its minimum value in $\underline{u}_0$).
Note that if the "surface" $D(\underline{u}) = 1$ is interpreted as an "ellipsoid" in $L_2(\Omega)$ then the search of the above infimum is equivalent to the search of the largest $\underline{u}$ (in the sense of the L_2–norm) on the surface such that $D(\underline{u}) = 1$: this is $\lambda_0^{-1}\underline{u}_0$ so that $\underline{u}_0$ has the interpretation of direction of the largest axis of the ellipsoid and λ_0^{-1} that of its length.
Finally note that strictly analogous results can be derived for the quadratic form $D_1(\underline{u}) = D(u) + \int_\Omega |\underline{u}|^2\, d\underline{x}$. (*Idea:* Note the remarkable *quadrangular equality*:

$$D\left(\frac{\underline{u}_n + \underline{u}_m}{2}\right) + D\left(\frac{\underline{u}_n - \underline{u}_m}{2}\right) = \frac{D(\underline{u}_n)}{2} + \frac{D(\underline{u}_m)}{2} \xrightarrow[n,m\to\infty]{} \lambda_0^2$$

and note that $\frac12|\underline{u}_n + \underline{u}_m|_2 \xrightarrow[m,n\to\infty]{} 1$ and deduce that $\liminf_{n,m\to\infty} D\left(\frac{\underline{u}_n+\underline{u}_m}{2}\right) \ge \lambda_0^2$ and, *hence*, $\frac12 D(\underline{u}_n - \underline{u}_m) \to 0$.)

[2.2.9]: (*recursive construction of the eigenvalues of the Dirichlet form on solenoidal fields*) In the context of [2.2.8], define $\lambda_1 > 0$ as

$$\lambda_1^2 = \inf_{\substack{|\underline{u}|_2=1 \\ \underline{u}\in X^0_{rot},\ \text{and}\ \langle \underline{u},\underline{u}_0\rangle_{L_2}=0}} D(\underline{u})$$

and show that there is a vector field $\underline{u}_1 \in L_2(\Omega)$ such that $D(\underline{u}_1) = \lambda_1^2$, then define λ_2^2 *etc.* In the geometric interpretation of [2.2.8] $\underline{u}_1$ is the direction of the next largest axis

[3] The domain of a quadratic form defined on a linear subspace $\mathcal{D}$ of a (real) Hilbert space H consists in the vectors u for which one can find a sequence $u_n \in \mathcal{D}$ with $||u_n - u_m|| \xrightarrow[n,m\to\infty]{} 0$ and $D(u_n - u_m) \xrightarrow[n,m\to\infty]{} 0$: in such case the sequence $D(u_n)$ converges to a limit ℓ and we set $D(u) \stackrel{def}{=} \ell$, and the set of such vectors u is called the *domain of the closure of the form* or simply the "domain of the form". If u, v are in the domain of the form D one can also extend the "scalar product" $(u,v)_D = \Big(D((u+v)) - D(u) - D(v)\Big)/2$.

of the ellipsoid and λ_1^{-1} is its length. Show also that $(\underline{u}_0, \underline{u}_1)_D = 0$. (*Idea:* Repeat the construction in [2.2.8]. Then remark that if $\underline{w} = x\underline{u}_0 + y\underline{u}_1$ then $D(\underline{w}) \equiv x^2\lambda_0^2 + y^2\lambda_1^2 + 2(\underline{u}_0, \underline{u}_1)_D xy = 1$ is an ellipse in the plane x, y with principal axes coinciding with the x and y axes.)

[2.2.10]: (*minimax principle for solenoidal fields*) Consider the nondecreasing sequence $\lambda_j, j = 0, 1, 2, \dots$ constructed in [2.2.9] and show the validity of the following "*minimax principle*":

$$\lambda_j^2 = \min\max D(\underline{u})$$

where the maximum is taken over the normalized vectors that are in a subspace W_j, with dimension $j+1$, of the domain of the closure (see footnote [3]) of $D(\underline{u})$, while the minimum is over the choices of the subspaces W_j. Find the simple geometric interpretation of this principle in terms of the ellipsoid of [2.2.8], [2.2.9]. Note that in the minimax principle we can replace the $j+1$–dimensional subspaces of the domain of the form with the $j+1$–dimensional subspaces of X^0_{rot}, provided we replace the minimum with an infimum. (*Idea:* The principle is obvious by [2.2.8] in the case $j = 0$ and it is an interpretation of [2.2.9] in the other cases.)

[2.2.11]: (*minimax and bounds on eigenvalues of the solenoidal Dirichlet form*) Consider the sequence $\lambda_j(\Omega)$, $j = 0, 1, \dots$ constructed in [2.2.9],[2.2.10] and consider the analogous sequence associated with the quadratic form $D(\underline{u})$ defined on the fields $\underline{u}$ of class $C^\infty(T_\Omega)$ periodic and with zero divergence on the cube T_Ω, defined in [2.2.4]. Denoting $\lambda_j(T_\Omega)$ the latter, show that $\lambda_j(\Omega) \ge \lambda_j(T_\Omega)$ and $\lim_{j\to\infty}\lambda_j(\Omega) = +\infty$. Likewise we can get upper bounds on the eigenvalues $\lambda_j^2(\Omega)$ by comparing the quadratic form D with the corresponding one on T'_Ω, a cube contained in the interior of Ω with periodic boundary conditions. (*Idea:* Make use of the minimax principle of [2.2.10] and note that every function in $X^0_{\rm rot}(\Omega)$ can be extended to a function on T_Ω. For the limit note that $\lambda_j(T_\Omega)$ are explicitly computable via a Fourier transform. To obtain the lower bound we extend linearly and continuously in the C^p topology for each p the periodic functions on T'_Ω to functions $X^0_{rot}(\Omega')$ defined on a slightly larger domain Ω' with smooth boundary containing the cube T'_Ω and contained in Ω keeping control of the L_2 norms of $\underset{\sim}{\partial u}$: see problems [2.2.33], [2.2.34] and [2.2.35] for more details.)

[2.2.12]: Show that if $\mathcal{H}$ is the closed subspace spanned (in $L_2(\Omega)$) by the vectors $\underline{u}_j$ then in the domain of the closure of the form D orthogonal to $\mathcal{H}$ there cannot exist $\underline{w} \ne \underline{0}$. (*Idea:* If $\underline{w}$ belonged to the domain of the closure of the form one would find, proceeding as in [2.2.8], a vector $\underline{u}$ such that $D(\underline{u}) = \inf D(\underline{w}) = \overline{\lambda}^2 < \infty$, with the infimum taken over all vectors in the domain of the form and orthogonal to $\mathcal{H}$: this contradicts that $\overline{\lambda}_j \to \infty$ as $j \to \infty$; in other words we would have forgotten one element of the sequence $\underline{u}_j$.)

[2.2.13]: (*heuristic equation for the eigenfunctions of the solenoidal Dirichlet form*) Write the condition that $D(\underline{u})$ is minimal in the space of the divergenceless C^∞–fields which vanish on the boundary of Ω and are normalized to 1 in $L_2(\Omega)$, assuming that the vector field $\underline{u}$ which realizes the minimum exists and is a C^∞–function. Show that it verifies

$$\Delta\underline{u} = -\lambda^2\underline{u} - \underline{\partial}\mu$$

where μ is a suitable function. (*Idea:* One can use the Lagrange multipliers method to impose the constraint $\underline{\partial} \cdot \underline{u} = 0$).

[2.2.14]: Show that the function μ in [2.2.13] can be determined via the projection $P_{X^\perp_{\rm rot}}(\Delta\underline{u})$, where $P_{X^\perp_{\rm rot}}$ is the projection operator discussed in (F) of §1.6 (*i.e.* it is the function whose gradient is the gradient part of the gradient–solenoid decomposition of the field $\Delta\underline{u}$).

[2.2.15]: Show that [2.2.13],[2.2.14] imply that we should expect that the vectors of the basis $\underline{u}_n$ satisfy the equations

$$\begin{aligned}\Delta\underline{u}_n =& -\lambda_n^2\underline{u}_n - \underline{\partial}\mu_n, \qquad \underline{\partial}\cdot\underline{u}=0 \qquad \text{in } \Omega\\ \underline{u}_n =& \underline{0}, \qquad \text{in } \partial\Omega\end{aligned}$$

for a suitable sequence of potentials μ_n.

[2.2.16]: Let $f\in C^\infty([0,H])$, show that

$$|f(0)|^2\le 2(H^{-1}||f||_2^2+H||f'||_2^2)$$

(*Idea:* Write $f(0)=f(x)+\int_x^0 f'(\underline{\xi})\,d\underline{\xi}$, and average the square of this relation over $x\in[0,H]$, then apply Schwartz' inequality).

[2.2.17]: Let $f\in C_0^\infty(\Omega\times[0,H])$ be the space of the $f(\underline{x},z)$ of class C^∞ and vanishing if $\underline{x}$ is close to the boundary (assumed regular) $\partial\Omega$ of Ω. Show that

$$\int_\Omega |f(\underline{x},0)|^2\,d\underline{x}\le 2(H\int_\Omega\int_0^H |\underline{\partial}_{\underline{x}} f|^2\,d\underline{x}\,dz+H^{-1}\int_\Omega\int_0^H |f|^2\,d\underline{x}\,dz)$$

(*Idea:* Apply the result of [2.2.16]).

[2.2.18]: (*a "boundary trace" theorem*) Let $f\in C^\infty(\Omega)$, show that there is $C>0$ such that, if L is the side of the smallest cube containing Ω:

$$\int_{\partial\Omega}|f(x)|^2\,d\sigma\le C(L^{-1}||f||_2^2+L||\underline{\partial}f||_2^2)$$

This is an interesting "*trace theorem*" on the boundary of Ω (a "Sobolev inequality"), [So63]. The constant C can also be chose so that it i invariant under homotety (in the sense that dilating the region Ω by a factor $\rho>1$ the constant C does not change and, therefore, C depends only on the geometric form of Ω and not on its size). (*Idea:* Make use of the method of partition of the identity in [1.5.7],[1.5.8] to reduce the present problem to the previous ones).

[2.2.19] (*Traces*) Let Ω be a domain with a bounded smooth manifold as boundary $\partial\Omega$: *i.e.* such that $\partial\Omega$ is covered by a finite number N of small surface elements each of which can be regarded as a graph over a disk δ_i tangent to $\partial\Omega$ at its center $\xi_i\in\partial\Omega$, so that the parametric equations of σ_i can be written $z=z_i(\xi)$, $\xi\in\delta_i$ and the points $\underline{x}$ of Ω close enough to σ_i can be parameterized by $\underline{x}=(\xi,z_i(\xi)+z)$ with $\xi\in\delta_i, z\ge 0$. Note that if Ω has the above properties also the homothetic domains $\rho\Omega$ with $\rho\ge 1$ have the same properties and the σ_i can be so chosen that $diam(\sigma_i)=cL$ if L is the diameter of Ω and N,c are the same for all domains $\rho\Omega$ with $\rho\ge 1$.
Let $f\in\mathcal{C}^\infty(\Omega)$ and define $\partial^{\underline{\alpha}}\equiv\frac{\partial^{|\underline{\alpha}|}}{\partial_1^{\alpha_1}\ldots\partial_d^{\alpha_d}}$ and

$$\begin{aligned}||f||^2_{W^n(\Omega)} =& \sum_{j=0}^n L^{2j-d}\sum_{|\underline{\alpha}|=j}\int_\Omega |\partial^{\underline{\alpha}} f(\underline{x})|^2\,d\underline{x}\\ ||f||^2_{W^n(\sigma_i)} =& ||f_{\delta_i}||_{W^n(\delta_i)}, \qquad ||f||^2_{W^n(\partial\Omega)}=\max_i ||f_{\delta_i}||_{W^n(\delta_i)}\end{aligned}$$

where $f_{\delta_i}(\xi)=f\big(\underline{x}(\xi,z_i(\xi))\big)$. Check that problem [2.2.18] implies

$$||f||_{W^{n-1}(\partial(\Omega))}\le\Gamma\,||f||_{W^n(\Omega)}, \qquad n\ge 1$$

and the constant Γ can be taken to be the same for all domains of the form $\rho\Omega$, $\rho\ge 1$

[2.2.20] (*A "trace" theorem*) Given a scalar function $f\in L_2(\Omega)$ suppose that it admits generalized derivative up to the order n included, with n even. Hence $(-\Delta)^{n/2}$ exists in a generalized sense, because $|\langle f,(-\Delta)^{n/2}g\rangle|\le C||g||_2$ for $g\in C_0^\infty(\Omega)$ and for a suitable C, *c.f.r.* §1.6, (1.6.19) and problem [2.2.1] above. Suppose $n>d/2$.
Show that f is continuous in every point in the interior of Ω, together with its first j derivatives if $n-d/2>j\ge0$. Find an analogous property for n odd.
(*Idea:* Let Q_ε be a cube entirely contained in Ω and let $\chi\in C_0^\infty(Q_\varepsilon)$. The function $\chi f\equiv f_\chi$ thought of as an element of $L_2(Q_\varepsilon)$ admits generalized derivatives of order $\le n$ and $(-\Delta)^{n/2}f_\chi$ exists in a generalized sense (this is clear if $n/2$ is an integer because $\chi\partial^p g=\sum_{j=0}^p\partial^j(\chi^{(j)}\,g)$ where $\chi^{(j)}$ are suitable functions in $C_0^\infty(Q_\varepsilon)$).
Hence thinking of $g\in C_0^\infty(Q_\varepsilon)$ as a periodic function in Q_ε we see that there is a constant C_{Q_ε} such that the relation $|\langle f_\chi,(-\Delta)^{n/2}g\rangle|\le C_{Q_\varepsilon}||g||_{L_2(Q_\varepsilon)}$ holds for each $g\in C_0^\infty(Q_\varepsilon)$, and therefore it must hold for each periodic $C^\infty(Q_\varepsilon)$–function g. Then, if $\hat f_\chi(\underline k)$ is the Fourier transform of f_χ as element of $L_2(Q_\varepsilon)$ (so that $\underline k=2\pi\varepsilon^{-1}\underline m$ with $\underline m$ an integer components vector), we get $\varepsilon^d\sum_{\underline k}|\hat f_\chi(\underline k)|^2|\underline k|^{2n}\le C^2_{Q_\varepsilon}$. Hence, setting $n=\frac d2+j+\eta$ with $1>\eta>0$, we see that the Fourier series of $\partial^j f_\chi$ is bounded above by the series

$$\begin{aligned}&\sum_{\underline k}|\underline k|^j|\hat f_\chi(\underline k)|\equiv\sum_{\underline k}|\underline k|^{j+\eta+d/2}|\hat f_\chi(\underline k)||\underline k|^{-\eta-d/2}\le\\&\le\Big(\sum_{\underline k}|\underline k|^{2n}|\hat f_\chi(\underline k)|^2\Big)^{1/2}\Big(\sum_{\underline k}|\underline k|^{-2\eta-d}\Big)^{1/2}\le\\&\le C_{Q_\varepsilon}\varepsilon^{-d/2}(\frac{\varepsilon}{2\pi})^{\eta+d/2}(\sum_{\underline m}|\underline m|^{-d-2\eta}\Big)^{1/2}\equiv\Gamma_d\varepsilon^{n-j-d/2}C_{Q_\varepsilon}\end{aligned}$$

hence f has j continuous derivatives in Q_ε. If n is odd one can say, *by definition* that $(-\Delta)^{n/2}f$ exists if it is $|<f,\partial^j g>|\le C||g||_2$ for each derivative of order $j\le n$ and for each $g\in C_0^\infty(\Omega)$; then the discussion is entirely parallel to the one in the n even case.)

[2.2.21] (*An auxiliary trace theorem*) Let $W^n(\Omega)$ be the space of the functions $f\in L_2(\Omega)$ with generalized derivatives of order $\le n$ and define.
Show that the method proposed for the solution of [2.2.19] implies that, if $n>j+d/2$ and $d(\underline x,\partial\Omega)$ denote the distance of $\underline x$ from $\partial\Omega$, then

$$L^j|\partial^{\underline\alpha}f(\underline x)|\le(\frac{d(\underline x,\partial\Omega)}{L})^{-j-d/2}\,\Gamma\,||f||_{W^n(\Omega)},\qquad|\underline\alpha|=j$$

and Γ can be chosen to be independent of Ω. (*Idea:* Let Q_1 be the unit cube. Let $\chi_1\in C_0^\infty(Q_1)$ be a function identically equal to 1 in the vicinity of the center of Q_1. Let $\chi_\varepsilon(\underline x)=\chi_1(\underline x\varepsilon^{-1})$ and show that the constant C_{Q_ε} considered in the estimates of problem [2.2.19] can be taken equal to $\gamma L^{d/2}\varepsilon^{-n}||f||_{W^n(\Omega)}$, with γ independent from Ω. Then choose $\varepsilon=d(\underline x,\partial\Omega)$.)

[2.2.22] (*Trace theorem*) Infer from problem [2.2.19], [2.2.21] that if $j<n-1-d/2$ there is a constant Γ such that $\partial^{\underline\alpha}f(\underline x)$, $|\underline\alpha|=j$ with $\underline x\in\partial\Omega$ is continuous and

$$L^j|\partial^{\underline\alpha}f(\underline x)|\le\Gamma\,||f||_{W^n(\Omega)},\qquad|\underline\alpha|=j$$

(*Idea:* a point $\underline x\in\partial\Omega$ will be in some σ_i, *c.f.r.* problem [2.2.19], and at a distance $O(1)$ from its boundary. Then apply problem [2.2.21].) With a little extra effort one can obtain the "same" result under the weaker condition $j<n-d/2$.

[2.2.27] (*a first regularity property of the eigenfunctions of the Dirichlet form: scalar case*) Let $f_0\in L_2(\Omega)$, $f_0=\lim f_n$, $||f_n||_2=1$ and $D(f_n)\to\lambda_0^2$, where λ_0 is associated with the first of the above minimax problems (*c.f.r.* [2.2.10]). Show that the function f_0 admits first generalized derivatives and also the generalized derivative Δ and $-\Delta f_0 = \lambda_0^2 f_0$. (*Idea:* For each $g \in C_0^\infty(\Omega)$ we get, setting $(f,g)_D \equiv \int_\Omega \underline{\partial} f \cdot \underline{\partial} g \, d\underline{\xi}$, that

$$|(f_n, g)_D - \lambda_0^2 \langle f_n, g\rangle| \xrightarrow[n\to\infty]{} 0 \Rightarrow |(f_0, (-\Delta - \lambda_0^2) g|) = 0$$

Hence $|(f_0, -\Delta g)| \le \lambda_0^2 ||g||_2$ and f_0 has a Laplacian in a generalized sense, by definition of generalized derivative, and $-\Delta f_0 = \lambda_0^2 f_0$.)

[2.2.28]: (*smoothess of the lowest eigenfunction of the Dirichlet form*) Check that under the hypotheses of [2.2.27] the f_0 is in $C^\infty(\Omega)$ and $f_0 = 0$ on $\partial\Omega$. (*Idea:* From $(-\Delta f_0) = \lambda_0^2 f_0$ it follows that $(-\Delta)^{n/2} f_0 = \lambda_0^n f_0$, for each $n > 0$ hence, by the results of [2.2.24], [2.2.26], the f_0 verifies the properties wanted. The vanishing on the boundary follows from the trace theorem, [2.2.18], and of the fact that all the functions approximating f_0 have by assumption value zero on $\partial\Omega$).

[2.2.29]: (*pointwise estimates on the derivatives of the lowest eigenfunction of the Dirichlet form*) Show that [2.2.24] and the elliptic estimates in [2.2.26] allow us to estimate the derivatives of the eigenfunctions, normalized in L_2, f_0 as $||f_0||_{C^j(\Omega)} \le \Gamma\, L^{-d/2}\,(1 + (L\lambda_0)^{(j+d)})$ for all values of j.

[2.2.30]: (*pointwise estimates on the derivatives of other eigenfunctions of the Dirichlet form*) Show that, by the minimax principle, and by what has been seen in the above problems, what we have obtained for f_0 applies to the other eigenvectors generated, via the minimax principle, from the Dirichlet quadratic form $D(f)$. In particular: $||f_p||_{C^j(\Omega)} \le \Gamma\, L^{-d/2}\,(1 + (L\lambda_p)^{(j+d)})$ for all values of j, p.

[2.2.31]: (*pointwise upper estimates on the derivatives of other eigenfunctions of the solenoidal Dirichlet form*) Adapt the theory of the quadratic form $D(f)$ on the scalar fields in $f \in C_0^\infty(\Omega)$ to the theory of the form $D(\underline{u})$ in (2.2.15) on the space $X^0_{\rm rot}(\Omega)$ and deduce the theorem that leads to (2.2.16),(2.2.17). (*Idea:* Let $\underline{f}_0 \in L_2(\Omega)$, $\underline{f}_0 = \lim \underline{f}_n$, $\underline{f}_n \in X^0_{rot}(\Omega), ||\underline{f}_n||_2 = 1$ and $D(\underline{f}_n) \to \lambda_0^2$, where λ_0 is associated with the first of the above minimax problems (*c.f.r.* [2.2.10]). Show that the function $\underline{f}_0$ admits a generalized Laplacian Δ and $-\Delta \underline{f}_0 = \lambda_0^2 \underline{f}_0 + \underline{\partial}\mu_0$ with $\underline{\partial}\mu_0 \in L_2(\Omega)$. (*Idea:* For each $\underline{g} \in C_0^\infty(\Omega), \underline{g} \equiv \underline{g}_{rot} + \underline{\partial}\gamma$ we get, setting $(f,g)_D \equiv \int_\Omega \underline{\partial} f \cdot \underline{\partial} g \, d\underline{\xi}$, that

$$|(\underline{f}_n, \underline{g})_D - \lambda_0^2 \langle \underline{f}_n, \underline{g}\rangle| \xrightarrow[n\to\infty]{} 0 \Rightarrow |(\underline{f}_0, (-\Delta - \lambda_0^2) \underline{g}_{rot})| = 0$$

hence $|(\underline{f}_0, -\Delta \underline{g})| = \lambda_0^2 |(\underline{f}_0, \underline{g}_{rot})| \le \lambda_0^2 ||\underline{g}||_2$ (because $||\underline{g}_{rot}||_2 \le ||\underline{g}||_2$ as the solenoid gradient decomposition of $L_2(\Omega)$ is orthogonal, *c.f.r.* Sec. 1.6.5). Hence f_0 has a generalized Laplacian and $D\underline{f}_0 - \lambda_0^2 \underline{f}_0 = \underline{0}$ in the space $\underline{X}_{rot}$: *i.e.* there is μ_0 such that $\underline{\partial}\mu_0 \in X_0^\perp(\Omega)$ and $-\Delta \underline{f}_0 - \lambda_0^2 \underline{f}_0 = \underline{\partial}\mu_0$. Analogously one finds that $\Delta^n f_0 = \lambda_0^n + \underline{\partial}\mu^n$ with $\underline{\partial}\mu^n$ in $X_{rot}^\perp(\Omega)$ for all $n > 0$, *etc.*)

[2.2.32]: (*completeness of the eigenfucnctions of the Dirichlet form*): Show that the sequence $\underline{u}_j$ constructed in [2.2.9] is an orthonormal basis in $X_{rot}(\Omega)$ which, from [2.2.31], is such that $\underline{u}_j$ are $C^\infty(\Omega)$ functions with zero divergence and vanishing on the boundary $\partial\Omega$. (*Idea:* If there existed $\underline{w} \ne \underline{0}$ in $X^0_{rot}(\Omega)$ (*c.f.r.* [2.2.9]) but out of the linear span $\mathcal{H}$ closed in $X_{rot}(\Omega)$ (which here plays the role of L_2 in the scalar problems treated in the preceding problems) one could suppose it orthogonal to $\mathcal{H}$: indeed $\lambda_j^k(\underline{w}, \underline{u}_j) = (\underline{w}, (-\Delta)^k \underline{u}_j) = ((-\Delta)^k \underline{w}, \underline{u}_j)$ hence

$$|(\underline{w}, \underline{u}_j)| \le ||(-\Delta)^k \underline{w}||\, \lambda_j^{-k}, \qquad |\partial^r \underline{u}_j(x)| \le \Gamma L^{-d/2}(1 + (L\lambda_j)^{r+d})$$

where the second inequality follows from [2.2.29], [2.2.30] (adapted to the non scalar case as in [2.2.30]). Furthermore from the first of (2.2.17) we see that $(\underline{w}, \underline{u}_j)$ tend to zero faster than any power in j and, also, that the series $\underline{w}^{||} = \sum_j (\underline{w}, \underline{u}_j)\underline{u}_j(x)$ converges very well, so that its sum is in the domain of the closure of D (one verifies immediately that the sum converges in the sense of the norm $\sqrt{D(\underline{u})}$, so that by the footnote [2] we see that $\underline{w}^{\|}$ is in the domain of D). This means that $\underline{w} - \underline{w}^{||} = \underline{w}^{\perp}$ is a vector in the domain of the form D which does not vanish and which is orthogonal to the space spanned by the vectors $\underline{u}_j$: which is impossible by the remark in [2.2.12].)

[2.2.33]: (*estimates on the large order eigenvalues and eigenfunctions of the scalar Dirichlet form*) Show that $\lambda_j^2 \ge Cj^{2/d}$ for some $C > 0$, *i.e.* find lower bound similar to the upper bound in [2.2.11] to the eigenvalues of the quadratic form D in the scalar case. (*Idea:* Let $Q' \subset \Omega$ be a cube of side size $3L$. It contains 3^d cubes of size L. Let Q be the one among them which is at the center and suppose that Q contains Ω. Let $u \in C^\infty(Q)$ be *periodic* over Q and imagine it extended to the whole Q' by periodicity and to the whole Ω by setting it to 0 outside Q'. Let χ_Q be a $C^\infty(\Omega)$ function that has value 1 on and near Q and vanishes near the boundary of Q' and outside Q'. Let $w = \chi_Q u \in C^\infty(\Omega)$.
Since $\underline{\partial} w = \underline{\partial}\,\chi_Q\, u + \chi_Q\, \underline{\partial}\, u$ we see that

$$\int_\Omega (\underline{\partial} w)^2 dx \le 2 \int_\Omega \Big((\underline{\partial}\chi_Q)^2 |u|^2 + \chi_Q^2 (\underline{\partial} u)^2\Big)\, dx \le \gamma^{-1} \int_Q (u^2 + (\underline{\partial} u)^2)\, dx$$

for γ a suitable constant. Hence if u varies in a $j+1$–dimensional subspace $W \subset C^\infty(T)$ it is

$$\max_{u\in W} \int_Q (u^2 + (\underline{\partial} u)^2)\, dx \ge \gamma \max_{u\in W} \int_\Omega (\underline{\partial} w)^2\, dx$$

Taking the minimum over all $j+1$–dimensional spaces $W \subset C^\infty(T)$ we get $\Lambda_j^2(Q) \ge \lambda_j^2(\Omega)$ by the minimax principle [2.2.10], if $\Lambda_j^2(Q)$ are the eigenvalues associated with the quadratic form $\int_Q (u^2 + (\underline{\partial} u)^2)\, dx$ to which the same arguments and results (including the minimax principle) obtained above for the quadratic form $\int_Q (\underline{\partial} u)^2\, dx$ apply with the obvious changes, see [2.2.8]. The latter eigenvalues have the form $(2\pi L^{-1} \underline{m})^2$ where $\underline{m}$ is an arbitrary integer component vector, with multiplicity 2 for each $\underline{m}$, so that we get the inequality $\lambda_j^2(\Omega) \le Cj^{2/d}$ and the inequality analogous to the first of (2.2.17) in the present scalar case follows from [2.2.29]. [2.2.30] and from the latter inequality; the analogue of the second of (2.2.17) follows from the above inequality and from [2.2.30]. Lower bounds on $\lambda_j(\Omega)$ can be obtained by the minimax principle by enclosing Ω in a cube $Q' \supset Q \supset \Omega$.)

[2.2.34]: (*Estimates on the large order eigenvalues and eigenfunctions of the scalar Dirichlet form*) Let Q, Q', χ_Q be as in [2.2.33]. Let $\underline{u}$ be a C^∞ divergenceless field periodic on Q. Then $\underline{u}$ can be represented on Q as $\underline{u} = \underline{a} + \text{rot}\, \underline{A}$ where $\underline{a}$ is a constant vector and $\underline{A}$ has zero divergence, is $C^\infty(Q)$ and is periodic on Q. We extend $\underline{A}$ to a $C^\infty(Q')$ divergenceless field by periodicity and set it $\underline{0}$ outside Q'. Then $\underline{w} = \text{rot}\,(\chi_Q \frac{1}{2} \underline{x} \wedge \underline{a} + \chi_Q \underline{A})$ extends the field $\underline{u}$ to a field in $X^0_{rot}(\Omega)$. Check that

$$\int_\Omega (\underline{\partial}\, \underline{w})^2 \le \gamma^{-1} \int_Q (\underline{u}^2 + (\underline{\partial}\, \underline{u})^2)\, d\underline{x}$$

for some $\gamma > 0$. This implies the relations in (2.2.17), by the same argument in [2.2.33] and by the fact that the eigenvalues of the quadratic forms $\int_Q (\underline{u}^2 + (\underline{\partial}\, \underline{u})^2)\, d\underline{x}$ and can be explicitly computed and shown to have the form $1 + (2\pi L^{-1} \underline{m})^2$ where $\underline{m}$ is an arbitrary integer component vector, with multiplicity 4 for each $\underline{m}$. (*Idea:* The inequality and the determination of γ can be easily performed by writing the relations in Fourier transform over Q of $\underline{u}, \underline{A}$.)

[2.2.35]: (*bounds on the multipliers μ_j of the eigenfunctions of the solenoidal Dirichlet form*) Show that also the potentials μ_j in [2.2.15] can be bounded by a bound like the second of (2.2.17) $|\partial^k \mu_j| \le c_k j^{\alpha+k/d}$ and estimate α, c_k. (*Idea:* simply use $-\underline{\partial}\mu_j = -\Delta \underline{u}_j + \lambda_j^2 \underline{u}_j$ and then use (2.2.17).)

Bibliography: the gyroscopic analogy in $d = 3$ is taken from [Wa90]; the theory of the elliptic equations and Stokes problem, is based upon [So63], [Mi70]. For a classical approach to the Dirichlet and Neumann problems see also [Ga82].

§2.3 Vorticity algorithms for incompressible Euler and Navier–Stokes fluids. The $d = 2$ case.

So far we tried to set up "*internal approximation algorithms*": by this we mean algorithms in which one avoids (or tries to avoid) approximating the wanted *smooth* solutions with velocity fields that are singular or have high gradients.

The interest of such methods lies in the fact that the fluid motions considered, real or approximate, are always motions in which make the hypotheses underlying the microscopic derivation of the equations can be considered valid.

However one can conceive "*external approximation algorithms*", in which one uses approximations that violate the regularity properties of the macroscopic fields, assumed in deriving the equations of motion: the regularity properties (*necessary for the phyisical consistency of the models*) of the solutions *should* (therefore) be recovered only in the limit in which the approximation converges to the solution.

Certainly such a program can leave us quite perplexed; but it is worth examining because, in spite of what one might fear, it has given positive results in quite a few cases and, in any event, it leads to interesting mathematical problems and to applications in other fields of Physics.

There is essentially only one method and it relies on Thomson's theorem. We shall examine it, as an example, in the case of a periodic container with side size L.

Consider first $d = 2$. The divergence condition is imposed by representing the velocity field $\underline{v}$ as

$$\underline{u} = \underline{\partial}^\perp A \qquad \underline{\partial}^\perp = (\partial_2, -\partial_1) \tag{2.3.1}$$

with A a scalar (smooth, see §1.6), and the vorticity is also a scalar

$$\zeta = \operatorname{rot} \underline{u} = -\Delta A \tag{2.3.2}$$

so that $\underline{u} = -\underline{\partial}^\perp \Delta^{-1} \zeta$, *c.f.r.* (C) in §1.7.

The Euler equations ($\nu = 0$) or the Navier–Stokes ($\nu > 0$) equations can be written in terms of ζ:

$$\begin{cases} \partial_t \zeta + \underline{u} \cdot \underline{\partial} \zeta = \nu \Delta \zeta + \gamma \\ \underline{u} = -\underline{\partial}^\perp \Delta^{-1} \zeta \end{cases} \tag{2.3.3}$$

where $\gamma = \operatorname{rot} \underline{g}$.

We now assume that the initial vorticity field is singular, and precisely it is a linear conbination of Dirac's delta functions

$$\zeta_0(\underline{\xi}) = \sum_{j=1}^n \omega_j \delta(\underline{\xi} - \underline{\xi}_j^0) \tag{2.3.4}$$

i.e. we suppose that the vorticity is concentrated in n points $\underline{\xi}_1^0, \ldots, \underline{\xi}_n^0$ where it is singular and proportional to ω_j: which we take to mean

$$\oint_{\underline{\xi}_j^0} \underline{u}_0(\underline{x}) \cdot d\underline{x} = \omega_j \tag{2.3.5}$$

if the contour turns around point $\underline{\xi}_j^0$ excluding the other $\underline{\xi}$'s.

To find the velocity field corresponding to (2.3.4) we need the inverse of the Laplace operator Δ with periodic boundary conditions. The Green function G, kernel of $-\underline{\Delta}^{-1}$, with periodic boundary conditions has the form

$$G(\underline{\xi}, \underline{\eta}) \equiv \Delta^{-1}_{\underline{\xi}\,\underline{\eta}} = -\frac{1}{2\pi} \log |\underline{\xi} - \underline{\eta}|_L + G_L(\underline{\xi}, \underline{h}) \equiv G_0(|\underline{\xi} - \underline{\eta}|_L) + \Gamma_L(\underline{\xi} - \underline{\eta}) \tag{2.3.6}$$

where $G_0(\underline{\xi} - \underline{\eta}) \equiv -\frac{1}{2\pi} \log |\underline{\xi} - \underline{\eta}|$ is the Green function for the Laplace operator Δ on the whole plane and $|\underline{\xi} - \underline{\eta}|_L$ is the metric on the torus of side L defined by $|\underline{\xi} - \underline{\eta}|_L^2 = \min_{\underline{n}} |\underline{\xi}_i - \underline{\eta}_i - \underline{n}L|^2$; and Γ_L is of class C^∞ for $|\underline{\eta}_i - \underline{\xi}_i| \neq L$ and such that $G(\underline{\xi}, \underline{\eta})$ is L–periodic and C^∞ for $\underline{\xi} \neq \underline{\eta}$. See problems following [2.3.11] for a proof of this interesting property.

The function $\underline{u}^0 = -\partial^\perp \Delta^{-1} \zeta^0$ has singular derivatives (for instance $\operatorname{rot} \underline{u}^0 = \sum_i \omega_i \delta(\underline{\xi} - \underline{\xi}_i)$) and therefore not only we are not in the situation in which it makes physically sense to deduce that the evolution of $\underline{u}^0$ is governed by the Euler equations but, worse, we even have problems at interpreting the equations themselves.

Consider the Euler equation: $\nu = 0$, and suppose that the external force γ vanishes. In reality the interpretation ambiguity is quite trivial, in a sense, because if we suppose that $\zeta(\underline{\xi}, t)$ has the form

$$\zeta(\underline{\xi}, t) = \sum_{i=1}^n \omega_i \delta(\underline{\xi} - \underline{\xi}_j(t)) \tag{2.3.7}$$

where $t \to \underline{\xi}_j(t)$ are suitable functions, then we can find a meaningful equation that has to be verified by $\underline{\xi}_j(t)$ in order to interpret it as a solution of the Euler equations.

Note that (2.3.7) would be consequence of vorticity conservation because it says that the vorticity is transported (see (1.7.14)) by the flow that generates it, *provided the initial value* ζ^0 *of* ζ *is regular*: since, however, ζ_0 is not regular we can think that (2.3.7) is part of the definition of solution of the (2.3.3) which, strictly speaking, does not make mathematical sense when ζ is not regular.

By substitution of (2.3.7) into the Euler equations, (2.3.3) with $\nu = 0$, we find

$$\sum_{i=1}^{n} \omega_i \underline{\partial}\delta(\underline{\xi} - \underline{\xi}_i(r)) \cdot \dot{\underline{\xi}}_i - \sum_{j=1}^{n} \omega_j \underline{\partial}^\perp G(\underline{\xi}, \underline{\xi}_j(t)) \cdot \sum_{p=1}^{n} \omega_p \underline{\partial}\delta(\underline{\xi} - \underline{\xi}_p(t)) = 0 \quad (2.3.8)$$

i.e. setting $\underline{\xi} = \underline{\xi}_i(t)$ we get

$$\dot{\underline{\xi}}_i = \underline{\partial}^\perp_{\underline{\xi}_i} \sum_{j=1}^{n} \omega_j G(\underline{\xi}_i, \underline{\xi}_j) \equiv \underline{\partial}^\perp_{\underline{\xi}_i} \sum_{h \neq i}^{n} \omega_h G(\underline{\xi}_i, \underline{\xi}_h) + \underline{\partial}^\perp_{\underline{\xi}} \omega_i G(\underline{\xi}, \underline{\xi}_i)|_{\underline{\xi} = \underline{\xi}_i} \quad (2.3.9)$$

which has no meaning because the "*autointeraction*" term

$$\underline{\partial}^\perp_{\underline{\xi}} G(\underline{\xi}, \underline{\xi}_i)\Big|_{\underline{\xi} = \underline{\xi}_i} = -\frac{1}{2\pi} \frac{(\underline{\xi} - \underline{\xi}_i)^\perp}{|\underline{\xi} - \underline{\xi}_i|^2}\Big|_{\underline{\xi} = \underline{\xi}_i} \quad (2.3.10)$$

has no meaning.

However one can think that the components of $\partial^\perp_{\underline{\xi}} G(\underline{\xi}, \underline{\xi}_i)|_{\underline{\xi} = \underline{\xi}_i}$ *are numbers having the limit value for* $\underline{\xi} \to \underline{\xi}_i$ *of an odd function of* $\underline{\xi} - \underline{\xi}_i$ *and that they can, therefore, be interpreted as zero.* Obviously this remark can only have a heuristic value and it cannot change the sad fact that (2.3.9) does not have mathematical sense.

Hence we shall *define* a solution of (2.3.3), with $\nu = \gamma = 0$ and initial datum (2.3.4), the (2.3.7) with $\underline{\xi}_j(t)$ given by the solution of the equation

$$\dot{\underline{\xi}}_j = \underline{\partial}^\perp_{\underline{\xi}_j} \sum_{n \neq j} \omega_n G(\underline{\xi}_j, \underline{\xi}_n) \quad (2.3.11)$$

which coincides with (2.3.9) deprived of the meaningless term.

The entire procedure can, rightly, look arbitrary and it is convenient to examine through which mechanisms one can imagine to approximate regular solutions to Euler equation ($\nu = \gamma = 0$ in (2.3.3)) via "solutions" of (2.3.11).

The idea is quite simple. A continuous vorticity field can be thought of as a limit for $\varepsilon \to 0$ of

$$\zeta^0_\varepsilon(\underline{x}) = \sum_i \zeta^0_i \, |\Delta_i| \, \delta(\underline{x} - \underline{x}_i) \quad (2.3.12)$$

where $\zeta_i^0 = \zeta^0(\underline{x}_i)$ and the sum is over small squares Δ_i, and with sides ε and centered at points $\underline{\xi}_i$, of a pavement of T_L.

This means that for the purpose of computing the integrals $\int \zeta^0(\underline{x}) f(\underline{x})\, d\underline{x}$, at least, one can proceed by computing

$$\lim_{\varepsilon\to 0} \int \zeta^0_\varepsilon(\underline{x}) f(\underline{x})\, d\underline{x}, \qquad \text{for all } f \in C^\infty(T_L) \tag{2.3.13}$$

i.e., as it is usual to say, "the functions ζ^0_ε approximates *weakly* ζ^0" as $\varepsilon \to 0$.

Then we can hope that, letting the field (2.3.12) evolve as prescribed by (2.3.11), and defining in this way a singular vorticity field $\zeta_\varepsilon(\underline{x}, t)$, one has that

$$\lim_{\varepsilon\to 0} \zeta_\varepsilon(\underline{x}, t) = \zeta(\underline{x}, t) \tag{2.3.14}$$

exists *in the weak sense*, *i.e.* $\lim_{\varepsilon\to 0} \int \zeta_\varepsilon(\underline{x}, t) f(\underline{x})\, d\underline{x} = \int \zeta(\underline{x}, t) f(\underline{x})\, d\underline{x}$, for all functions $f \in C^\infty(T_L)$, and it is a regular function, if such was ζ^0 to begin with, verifying the (2.3.3) with initial datum ζ_0.

It is therefore very interesting that the latter statement is actually true if $\zeta_0 \in C^\infty(T_L)$, [MP84],[MP92]. This theorem is of the utmost interest because it shows the very "possibility" of external approximations to the solutions of the Euler equation.

The method can be suitably extended to the theory of the Navier–Stokes equations, $\nu \neq 0$, and to the forced fluid case $\gamma \neq 0$. For the moment we supersede such extensions and, instead, we study more in detail the most elementary properties of the equation (2.3.11).

One has to remark first that (2.3.11) can be put in *Hamiltonian form*; set $\underline{x}_i = (x_i, y_i)$

$$\begin{aligned} &p_i = \sqrt{|\omega_i|}\; x_i, \quad q_i = \frac{\omega_i}{\sqrt{|\omega_i|}} y_i, \quad (\underline{p}, \underline{q}) \equiv (p_1, q_1, \ldots, p_n, q_n)\\ &H(\underline{p}, \underline{q}) = -\frac{1}{2} \sum_{h\neq k} \omega_h \omega_k G(\underline{\xi}_h, \underline{\xi}_k) \end{aligned} \tag{2.3.15}$$

where $\underline{\xi}_h, \underline{\xi}_k$ have to be thought as expressed in terms of the (p_h, q_h), (p_k, q_k). Then we can check that (2.3.11) becomes

$$\dot p_j = -\frac{\partial H}{\partial q_j}, \qquad \dot q_j = \frac{\partial H}{\partial p_j} \tag{2.3.16}$$

i.e. the (2.3.11) are equivalent to the system of Hamiltonian equations (2.3.16). If $L = \infty$, *i.e.* for the equations in the whole space, H is

$$H(\underline{p}, \underline{q}) = \frac{1}{8\pi} \sum_{h\neq k} \omega_h \omega_k \log\Big(\Big|\tfrac{p_h}{\sqrt{|\omega_h|}} - \tfrac{p_k}{\sqrt{|\omega_k|}}\Big|^2 + \Big|\tfrac{q_h}{\sigma_h\sqrt{|\omega_h|}} - \tfrac{q_k}{\sigma_k\sqrt{|\omega_k|}}\Big|^2\Big) \tag{2.3.17}$$

where σ_i is the sign of ω_i.

We shall consider the case $L = \infty$ in more detail: but even in the latter case the equations are difficult to solve, except in the trivial case of the "*two vortices problem*" when

$$H = \frac{1}{4\pi}\omega_1\omega_2 \log\Big[\Big(\tfrac{p_1}{\delta_1} - \tfrac{p_2}{\delta_2}\Big)^2 + \Big(\tfrac{q_1}{\delta_1}\sigma_1 - \tfrac{q_2}{\delta_2}\sigma_2\Big)^2\Big] \tag{2.3.18}$$

with $\delta_j = \sqrt{|\omega_j|}$ and the Hamilton equations become, if Δ is the argument of the logarithm,

$$\dot p_1 = -\frac{\omega_1\omega_2}{2\pi}\frac{\frac{\sigma_1}{\delta_1}\Big(q_1\frac{\sigma_1}{\delta_1} - q_2\frac{\sigma_2}{\delta_2}\Big)}{\Delta} \qquad \dot p_2 = +\frac{\omega_1\omega_2}{2\pi}\frac{\frac{\sigma_2}{\delta_2}\Big(q_1\frac{\sigma_1}{\delta_1} - q_2\frac{\sigma_2}{\delta_2}\Big)}{\Delta}$$

$$\dot q_1 = \frac{\omega_1\omega_2}{2\pi}\frac{\frac{1}{\delta_1}\Big(\frac{p_1}{\delta_1} - \frac{p_2}{\delta_2}\Big)}{\Delta} \qquad \dot q_2 = -\frac{\omega_1\omega_2}{2\pi}\frac{\frac{1}{\delta_2}\Big(\frac{p_1}{\delta_1} - \frac{p_2}{\delta_2}\Big)}{\Delta}$$

so that we see that $\sigma_1\delta_1 p_1 + \sigma_2\delta_2 p_2 = \text{const}$, $\delta_1 q_1 + \delta_2 q_2 = \text{const}$, *i.e.* in terms of the original coordinates

$$\omega_1 x_1 + \omega_2 x_2 = \text{cost}, \qquad \omega_1 y_1 + \omega_2 y_2 = \text{cost} \tag{2.3.19}$$

and if $\omega_1 + \omega_2 \neq 0$ we can define the *center of vorticity* x as

$$x = \frac{\omega_1 x_1 + \omega_2 x_2}{\omega_1 + \omega_2}, \qquad y = \frac{\omega_1 y_1 + \omega_2 y_2}{\omega_1 + \omega_2} \tag{2.3.20}$$

If $\omega_1/\omega_2 > 0$ the vorticity center can be interpreted as the "center of mass" of two points with masses equal to $|\omega_i|$: the faster vortex is closer to the vorticity center. If $\omega_1/\omega_2 < 0$ then the center of vorticity leaves both vortices located at points P_1 and P_2 "on the same side". The equations are solved by the motion in which the line P_1P_2 joining the two vortices rotates with angular velocity $(\omega_1 + \omega_2)/(2\pi\Delta)$ counterclockwise, if $\Delta = (x_1 - x_2)^2 + (y_1 - y_2)^2$, around the vorticity center (*c.f.r.* problems). The distance $\sqrt{\Delta}$ has to be > 0 as we shall exclude initial data in which a pair of vortices occupy the same point.

If instead $\omega_1 + \omega_2 = 0$ and $\omega \stackrel{def}{=} \omega_1$, the two vortices proceed along two parallel straight lines perpendicular to the line joining them and with velocity $\omega/(2\pi\sqrt{\Delta})$, going to the right of the vector that joins P_2 to P_1 if $\omega > 0$ (and to the left otherwise), (*c.f.r.* problems).

In general the "problem of n vortices" with intensities $\omega_1, \ldots, \omega_n$, and *vanishing total vorticity* $\omega = \sum_i \omega_i = 0$, admits, if $L = \infty$, *four* first integrals

$$\begin{aligned} I_1 &= \sum_i \sigma_i\sqrt{|\omega_i|}p_i, \qquad I_2 = \sum_i \sqrt{|\omega_i|}q_i,\\ I_3 &= \frac{1}{2}\sum_i \sigma_i(p_i^2 + q_i^2), \qquad I_4 = H(\underline{p}, \underline{q}) \end{aligned} \tag{2.3.21}$$

the I_1, I_2, I_3 can be simply written in the original coordinates as

$$I_1 = \sum_i \omega_i x_i, \quad I_2 = \sum_i \omega_i y_i, \quad I_3 = \frac{1}{2}\sum_i \omega_i |\underline{\xi}_i|^2 \tag{2.3.22}$$

while I_4 is given by (2.3.17). Their constancy in time follows directly from the equations of motion in the coordinates $\underline{x}_i$, (2.3.11) with G given in (2.3.6), by multiplying them with ω_i and summing over i, or multiplying them by $\omega_i \underline{x}_i$ and summing over i. If L is finite only I_1, I_2, I_4 are first integrals (recall that if L is finite we only consider periodic boundary conditions).

In general, however, such integrals are not in involution, in the sense of analytical mechanics, with respect to the Poisson brackets (that we denote with curly brackets as usual). With the exception of a few notable cases.

For instance $\{I_4, I_j\} = 0$ simply expresses that I_1, I_2, I_3 are constants of motion; while $\{I_1, I_2\} = 0$ only if $\sum_i \omega_i \equiv \omega = 0$, (because I_1 only depends on $\underline{p}$ and I_2 only depends on the $\underline{q}$, hence the calculation of the parenthesis is easy and one sees that it yields, in fact, ω); furthermore $\{I_3, I_2\} = I_1$ and $\{I_3, I_1\} = -I_2$.

On the basis of general theorems on integrable systems we must expect that also the three vortices problem *with vanishing total vorticity* be integrable by quadratures. And in fact this is a generally true property. All "*confined motions*" (*i.e.* such that the coordinates of the points stay bounded as $t \to \infty$) will in general be quasi periodic and the others will be reducible to superpositions of uniform rectilinear motions and quasi periodic motions: *c.f.r.* problems. Here the word "superposition" has the meaning of the classical nonlinear superposition that one considers in mechanics in the theory of quadratures and of quasi periodic motions, *c.f.r.* [Ga99b].

The interest of the condition $\omega = 0$ of zero total vorticity is that this condition must automatically hold if one requires that the velocity field generated by the vortices tends to 0 at ∞ quickly (*i.e.* faster than the distance away from the origin): the circulation at ∞ has indeed the value ω.

In reality also the general three vortices problem with $\omega \neq 0$, representing vorticity fields slowly vanishing at ∞, is integrable in general by quadratures, *c.f.r.* problems.

Concerning the four or more vortices problems one can show, by following the same method used by Poincaré to show the non integrability by quadratures of the three body problem in celestial mechanics, that the problem is in general *not integrable* by quadratures: it does not admit enough other analytic constants of motion, [CF88b].

Finally if one considers the Euler equations in domains Ω different from the torus and from R^2 one obtains the (2.3.11) with the Green function $G(\underline{\xi}, \underline{\eta})$ of the Dirichlet problem in Ω. In fact the boundary condition $\underline{u} \cdot \underline{n} = 0$ imposes, by (2.3.1), that the potential A must have *tangential* derivative zero on the boundary of Ω and therefore it must be constant and the constant can be fixed to be 0. Therefore $A = \Delta^{-1}\zeta$ where Δ is the Laplace operator with vanishing boundary condition.

In these cases, in general, only I_4 is a constant of motion: the case in which Ω is a disk is exceptional: because also I_3 is, by symmetry, a constant of motion and, therefore, in this case the two vortices problem is still integrable by quadratures.

Problems: *Few vortices Hamiltonian motions. Periodic Green function.*

[2.3.1]: (*two vortices problem on a plane*) Show that the equations of motion for two vortices of intensity ω_1, ω_2 located at (x_1, y_1) and (x_2, y_2) are respectively

$$\begin{aligned}\dot x_1 =& -\omega_2\,(y_1-y_2)/2\pi\Delta, \qquad \dot x_2 = \omega_1\,(y_1-y_2)/2\pi\Delta\\ \dot y_1 =& \omega_2\,(x_1-x_2)/2\pi\Delta, \qquad \dot y_2 = -\omega_1\,(x_1-x_2)/2\pi\Delta\end{aligned}$$

and deduce that setting $\zeta = (x_1 - x_2) + i(y_1 - y_2)$ it is $\dot\zeta = \frac{i(\omega_1+\omega_2)}{2\pi|\zeta|^2}\zeta$. Derive from this the properties of the motions for the two vortices problem discussed after (2.3.20).

[2.3.2]: Suppose $\omega_1 + \omega_2 \neq 0$, $\omega_1, \omega_2 \neq 0$ and set $c^{-1} = \sqrt{|\omega_1 + \omega_2|}$ and $d^{-1} = \sqrt{|\omega_1^{-1} + \omega_2^{-1}|}$. Let σ_1, σ_2 be, respectively, the signs of ω_1, ω_2 and let ϑ_1, ϑ_2 be the signs, respectively, of $\omega_1 + \omega_2$ and of $\omega_1^{-1} + \omega_2^{-1}$. Show that the transformation $(p_1, p_2, q_1, q_2) \longleftrightarrow (p, p', q, q')$:

$$\begin{aligned} p =& (\sigma_1|\omega_1|^{1/2}p_1 + \sigma_2|\omega_2|^{1/2}p_2)\,c \qquad q = (|\omega_1|^{1/2}q_1 + |\omega_2|^{1/2}q_2)\,c\,\vartheta_1\\ p' =& (p_1|\omega_1|^{-1/2} - p_2|\omega_2|^{-1/2})\,d \qquad q' = (\sigma_1 q_1|\omega_1|^{-1/2} - \sigma_2 q_2|\omega_2|^{-1/2})\,d\,\vartheta_2\end{aligned}$$

is a canonical map. (*Idea:* Poisson brackets between the p, p', q, q' are canonical: check.)

[2.3.3]: (*integrability by quadratures of two vortices planar motions*) Note that in the coordinates (p, p', q, q') the Hamiltonian of the two vortices problem depends only on the coordinates (p', q') and it is integrable by quadratures in the region $(p', q') \neq (0, 0)$, at (p, q) fixed. Show that the action–angle coordinates (A, α) can be identified with the polar coordinates on the plane (p', q'):

$$A = \frac{1}{2}\,(p'^2 + q'^2), \qquad \alpha = \arg(p', q'), \qquad \text{and} \qquad H(p', q') = \frac{\omega_1\omega_2}{4\pi}\log A + cost$$

(*Idea:* The map $(p', q') \longleftrightarrow (A, \alpha)$ is an area preserving map, hence a canonical map.)

[2.3.4]: Show that the results of [2.3.2] and [2.3.3] can be adapted to the case $\omega_1 + \omega_2 = 0$ and study it explicitly. (*Idea:* For instance the map

$$p' = p_1 - p_2, \quad q = (q_1 + q_2), \quad p = (p_1 + p_2)/2, \quad q' = (q_1 - q_2)/2$$

is canonical and transforms H into $\frac{1}{4\pi}\log(p'^2 + q'^2) + const$. However the motions do not have periodic components corresponding, in the preceding cases, to rotations of the vortices around the vorticity center, which is now located at ∞.)

[2.3.5]: Show that, *c.f.r.* (2.3.21), $\{I_3, I_1\} = -I_2$, $\{I_3, I_2\} = I_1$ and $\{I_1, I_2\} = \sum_i \omega_i$.

[2.3.6]: (*the planar three vortices problem*) Given three vortices with intensity ω_1, ω_2, ω_3, with $\omega_j > 0$, consider the transformation $(p_1, p_2, p_3, q_1, q_2, q_3) \longleftrightarrow (p, p', p_3, q, q', q_3)$ of problem [2.3.2] (it is a canonical transformation in which the third canonical coordinates are invariant) and compose it with the transformation of the same type $(p, p', p_3, q, q', q_3) \longleftrightarrow (P, p', p'', Q, q', q'')$ in which (p', q') are invariant while the transformation on (p, p_3, q, q_3) is still built as in [2.3.2] by imagining that in (p, q) there is a vortex with intensity $\omega_1 + \omega_2 = \omega_{12}$ and in (p_3, q_3) there is a vortex with intensity ω_3. Show that in the new coordinates (P, p', p'', Q, q', q'') the Hamiltonian is a function only of (p', p'', q', q'') while I_3 is the sum of a function of (P, Q) only and of a function of (p', p'', q', q'') only.

[2.3.7]: (*integrability by quadratures of the planar three vortices problem*) Show, in the context of [2.3.6], that the P, Q are constants of motion and the surfaces $H = \varepsilon, I_3 = \kappa$ are bounded surfaces in phase space. Hence they are regular 2–dimensional surfaces (*i.e.* they do not have singular points or degenerate into lower dimensional objects) then their connected components are 2–dimensional tori ("Arnold–Liouville theorem"), see [Ar79], and motions on these tori are quasi periodic with two frequences. The three vortices problem, *i.e.* the determination of the motions of the two degrees of freedom system obtained by fixing the values of P, Q, is therefore integrable by quadratures if $\omega_i > 0$. *This does not go on*: the planar four vortices problem is not integrable by quadratures, [CF88b].

[2.3.8]: Show that, by using the result of [2.3.4], the analysis about the three vortices problem and its integrability by quadratures remains true in the case in which vorticities do not have all the same sign, but are such that $\omega_1 + \omega_2 \neq 0$ and $\omega_1 + \omega_2 + \omega_3 \neq 0$. This time, however, the surfaces $I_3 = \kappa, H = \varepsilon$ will not in general be bounded and, therefore, the invariant surfaces will have the form, in suitable coordinates, of a product of a space $R^1 \times T^1$, or R^2, or T^2 and such coordinates can be chosen so that the evolution is linear and motion will be quasi periodic only in the third case.

[2.3.9]: (*integrability by quadratures of two vortices in a disk*) Consider the two vortices problem in a circular region. Show that this is also integrable by quadratures. (*Idea:* This time I_1, I_2 are not constants of motion, but H, I_3 still are (the second because of the circular symmetry of the problem); furthermore all motions are obviously confined.)

[2.3.10]: (*integrability by quadratures of two vortices in a torus*) As in [2.3.9] but assuming that the two vortices are confined on a torus (rather than moving on the plane). Show that the two vortices move rectlinearly because the total vorticity *must* be 0. (*Idea:* Note that on the torus the velocity field $\underline{u}$ must be periodic and of the form $\underline{u} = \underline{\partial}^\perp A$ with a suitable regular A, so that we can only consider vortices with total vorticity zero. This time I_3 is not a constant of motion but I_1, I_2 are, furthermore they are in involution, namely $\{I_1, I_2\} = 0$ if $\{\cdot, \cdot\}$ denotes the *Poisson bracket*.)

[2.3.11]: (*quadratures for three vortices in a torus*) Show that also the three vortex problem on the torus will be integrable outside the level surfaces of H, I_1, I_2 which are not compact. (*Idea:* I_1 and I_2 are in involution, because the total vorticity vanishes on the torus, *c.f.r.* [2.3.10], and they are in involution with H. Then apply Arnold–Liouville theorem, *c.f.r.* [2.3.7].)

[2.3.12]: (*Green's function for periodic boundary conditions*) Consider $G_N(\underline{x} - \underline{y}) = \sum_{|\underline{n}| \le N} G_0(\underline{x} - \underline{y} - \underline{n}L) - \sum_{0 < |\underline{n}| \le N} G_0(\underline{n}L)$, where the sum runs over the integer components vectors $\underline{n} = (n_1, n_2)$ and G_0 is defined after (2.3.6). Check the existence of the limit $\lim_{N\to\infty} G_N(\underline{x} - \underline{y}) = G(\underline{x} - \underline{y})$, which is a periodic function of $\underline{x} - \underline{y}$ with period L in each coordinate, and which differs from $G_0(\underline{x} - \underline{y})$ by a C^∞–function of $\underline{x}, \underline{y}$ for $\underline{x} - \underline{y}$ small with respect to L.
Check that the only singularity of $G(\underline{\xi} - \underline{\eta})$ occurs at $\underline{\xi} = \underline{\eta}$. (*Idea:* Note that $G_0(\underline{x}) = -\frac{1}{2\pi} \log |\underline{x}|$ and $|\underline{\xi} - \underline{n}\, L| = |\underline{n}|\, L\, (1 + (-2\underline{n} \cdot \underline{\xi}\, L + \underline{\xi}^2)/(\underline{n}\, L)^2)^{1/2}$; and setting $\varepsilon = (-2\underline{n} \cdot \underline{\xi}\, L + \underline{\xi}^2)/(\underline{n}\, L)^2$ one has $-G_0(\underline{\xi} - \underline{n}\, L) = \frac{1}{4\pi} \log(1 + \varepsilon) + \frac{1}{2\pi} \log |\underline{n}| L$.
Developing in powers of ε the latter expression becomes $\frac{1}{4\pi}(\varepsilon - \frac{1}{2}\varepsilon^2 + O(\varepsilon^3)) + \frac{1}{2\pi} \log |\underline{n}| L$.

So that $-G_0(\underline{\xi} - \underline{n}L)$ becomes:

$$\frac{1}{4\pi}\Big(-2\frac{\underline{n}\cdot\underline{\xi}\,L}{|\underline{n}|^2L^2}+\frac{\underline{\xi}^2}{|\underline{n}|^2L^2}-2\frac{(\underline{n}\cdot\underline{\xi})^2L^2}{(|\underline{n}|^2L^2)^2}+O(|\underline{n}|^{-3})+\frac{1}{2\pi}\log|\underline{n}|L\Big)$$

Summing over $\underline{n}$ we note that the terms linear in $\underline{\xi}$ add up to 0. Furthermore the terms in $(\underline{n}\cdot\underline{\xi})^2$ have the form $\sum_{i,j} n_i n_j \xi_i\xi_j$ and by symmetry we get the same result if the latter sum is replaced by $\sum_i \xi_i^2 n_i^2$ and by symmetry between the components of $\underline{n}$ we get again the same result if we replace this by $\sum_i \xi_i^2 \underline{n}^2 = 2\underline{\xi}^2\underline{n}^2$. This means that when summing over $\underline{n}$ the contributions from the terms quadratic in $\underline{\xi}$ cancel exactly. Thus summing over $\underline{n}$ gives the same result as summing

$$\sum_{|\underline{n}|<N}\Big[G_0(\underline{\xi}-\underline{n}L)-G_0(\underline{n}L)-\underline{\partial}G_0(\underline{n}L)\cdot\underline{\xi}-\frac{1}{2}\underset{\sim}{\partial}\,G_0(\underline{n}L)\cdot\underline{\xi}\,\underset{\sim}{\xi}\Big]$$

which is a sum of terms of size $O(|\underline{n}|^{-3})$, which converges because the dimension is $d = 2$. And in fact the derivatives of order $\alpha \ge 0$ with respect to $\underline{\xi}$ of the sum above are expressed as sums of quantities which have size in $\underline{n}$ of order $O(|\underline{n}|^{-3-\alpha})$ so that the limit is C^∞ in the sense stated.)

[2.3.13]: Show that if $|\underline{x}-\underline{y}|_L^2$ is defined as $\sum_{i=1}^2(|x_i-y_i| \bmod L)^2$, *i.e.* if $|\underline{x}-\underline{y}|$ is the natural metric on the torus of side L, then the (2.3.6) holds with Γ_L of class C^∞ for $|x_i - y_i| \ne L$ on the torus.

[2.3.14]: (*the images method*) Show that the function $G(\underline{x}-\underline{y})$ is such that $\Delta_{\underline{y}}G(\underline{x}-\underline{y}) = \delta(\underline{x}-\underline{y})$: for this reason the construction in [2.3.12] is called the "*images method*" to construct the Green function of the laplacian with periodic boundary conditions. (*Idea:* It suffices to show this for $\underline{x} = (\frac{L}{2},\frac{L}{2})$, because of the translation invariance of G.)

Bibliography: The theorem of external approximation, following (2.3.14), is taken from [MP84]; for systems integrable by quadratures see [Ar79], [Ga83], [Ga86].

§2.4 Vorticity algorithms for incompressible Euler and Navier–Stokes fluids. The $d = 3$ case.

In the 3–dimensional case the analogue of the point vortex is a closed oriented curve γ, that we shall call *filament*, on which $\operatorname{rot}\underline{u} = \underline{\omega}$ is *concentrated* and is *tangent* to it, so that γ is a flux line for $\underline{\omega}$.

(A) Regular filaments. Divergences and infinities.

To understand the evolution of a vorticity filament consider the Euler equation in the form (1.7.3)

$$\dot{\underline{\omega}}+\underset{\sim}{u}\cdot\underset{\sim}{\partial}\underline{\omega}-\underset{\sim}{\omega}\cdot\underset{\sim}{\partial}\underline{u}=0,\qquad \frac{d\underline{\omega}}{dt}=\underset{\sim}{\omega}\cdot\underset{\sim}{\partial}\underline{u}\tag{2.4.1}$$

It is easy to find the meaning of (2.4.1) as an equation of evolution for a curve γ if we look at a point $\underline{\xi} \in \gamma$ and at an infinitesimal element, or *vorticity element* $\alpha\,\underline{\omega}$ of the filament, with α infinitesimal. The evolution by transport by the fluid of the element between $\underline{\xi}$ and $\underline{\xi}+\alpha\underline{\omega}$ is

$$\begin{aligned}\underline{\xi}\to\underline{\xi}'&=\underline{\xi}+\underline{u}(\underline{\xi})\,dt\\ \underline{\xi}+\alpha\underline{\omega}\to\tilde{\underline{\xi}}'&=\underline{\xi}+\alpha\underline{\omega}+\underline{u}(\underline{\xi}+\alpha\,\underline{\omega})\,dt\end{aligned}\tag{2.4.2}$$

which implies that the arc of γ between $\underline{\xi}$ and $\underline{\xi}+\alpha\,\underline{\omega}$ evolves into the arc between $\underline{\xi}'$ and $\tilde{\underline{\xi}}'$ with

$$\tilde{\underline{\xi}}'-\underline{\xi}'=\alpha\,(\underline{\omega}+\underset{\sim}{\omega}\cdot\underset{\sim}{\partial}\,\underline{u}\,dt)=\alpha\,\underline{\omega}',\qquad\text{if}\quad\underline{\omega}'\stackrel{def}{=}\underline{\omega}+\underset{\sim}{\omega}\cdot\underset{\sim}{\partial}\,\underline{u}\,dt\tag{2.4.3}$$

This shows, by the second relation in (2.4.1), that the line element $\alpha\,\underline{\omega}$ evolves into $\alpha\,\underline{\omega}'$ while the line is transported by the current: hence γ remains always tangent to $\underline{\omega}$ and if the length of a line element of γ is changed, in the evolution of γ, by a factor $(1+\lambda\,dt)$ then $\omega'=\omega(1+\lambda\,dt)$ describes also the corresponding evolution of the modulus $\omega=|\underline{\omega}|$ of $\underline{\omega}$.

Hence the filament shape evolves simply because it is transported and deformed by the fluid. The vorticity, instead, changes *proportionally* to the expansion of the line element corresponding to it: if the line gets longer the vorticity increases.

Since vorticity is a zero divergence field, its flux is constant along its flux lines, in particular along γ; hence if γ is a vorticity filament it must be

$$\underline{\omega}(\underline{\xi})=\Gamma\,\delta_\gamma(\underline{\xi})\,\underline{t}_\gamma(\underline{\xi})\tag{2.4.4}$$

where $\underline{t}_\gamma(\underline{\xi})$ is the unit vector tangent to γ in $\underline{\xi}\in\gamma$ and $\delta_\gamma(\underline{\xi})$ is a uniform distribution concentrated on γ, defined by

$$\int f(\underline{\xi})\delta_\gamma(\underline{\xi})\,d\underline{\xi}\stackrel{def}{=}\int_\gamma f(\underline{\xi})\,dl\tag{2.4.5}$$

for each $f\in C^\infty$, if $d\,l$ is the line element for γ.

To check that Γ *is time independent* imagine the distributions δ_γ realized as (limit of) a function different from 0 in a infinitesimal tubular neighborhood $\mathcal{T}$ with cross–section, in $\underline{\xi}\in\gamma$, given by $s(\underline{\xi})$. Then, denoting $\chi_{\mathcal{T}}(\underline{\xi})$ the characteristic function of $\mathcal{T}$, it must be

$$\underline{\omega}(\underline{\xi})=\Gamma\chi_{\mathcal{T}}(\underline{\xi})\frac{1}{s(\underline{\xi})}\underline{t}_\gamma(\underline{\xi})\tag{2.4.6}$$

if $\underline{t}_\gamma(\underline{\xi})$ is the unit tangent vector to γ in $\underline{\xi}$.

Of course as the time varies the tube $\mathcal{T}$ is transformed into $\mathcal{T}'$ and the section of the tube contracts by $(1+\lambda\,dt)$ while the line element expands by $(1+\lambda\,dt)$ because the tube $\mathcal{T}$ evolution is by an incompressible transport. At the same time we know that vorticity varies by the same factor and, therefore, if $\underline{\xi}$ and $\underline{\xi}'$ are (via the evolution) corresponding points on the curve γ and on its image γ', we note that

$$\underline{\omega}'(\underline{\xi}') = \Gamma\,\chi_{\mathcal{T}'}(\underline{\xi}')(1+\lambda\,dt)\frac{1}{s(\underline{\xi})}\,\underline{t}'_{\mathcal{T}'} \equiv \Gamma\,\chi_{\mathcal{T}'}(\underline{\xi}')\frac{1}{s'(\underline{\xi}')}\,\underline{t}'_{\mathcal{T}'} \tag{2.4.7}$$

which, comparing with (2.4.6), implies that $\Gamma' \equiv \Gamma$.

The velocity field $\underline{u}$ associated with a vorticity filament can be computed via a formula often called *Biot–Savart formula* because it says that the velocity field of the vorticity field is the "*magnetic field*" of an electrric current of intensity Γ circulating on the filament as computed from the Biot-Savart law (units aside, of course)

$$\underline{u}(\underline{x}) = \frac{\Gamma}{4\pi}\oint_\gamma \frac{d\underline{\rho}\wedge(\underline{x}-\underline{\rho})}{|\underline{x}-\underline{\rho}|^3} \tag{2.4.8}$$

where $d\underline{\rho}$ is the line element of γ: *it is in fact the solution of* $\underline{\partial}\wedge\underline{u} = \underline{\omega}$ hence it is the magnetic field generated by the intensity of current $\underline{\omega}$ in (2.4.6).

Then the evolution of a system of several vorticity filaments should (naively) be described by

$$\frac{d\underline{\rho}}{dt} = \sum_{j=1}^{n}\frac{\Gamma_j}{4\pi}\oint_{\gamma_j}\frac{d\underline{l}\wedge(\underline{\rho}-\underline{l})}{|\underline{\rho}-\underline{l}|^3} \qquad \text{if} \quad \underline{\rho}\in\cup_{j=1}^n\gamma_j \tag{2.4.9}$$

because they should be transported by the flow $\underline{u}$.

One can then try to see if a generic vorticity field $\underline{\omega}$ is approximable by a family of many filaments γ with a small vorticity circulation Γ which, as an approximation parameter ε varies, should become denser and denser approximating better and better $\underline{\omega}$ in the sense that, for every fixed $f \in C^\infty(R^3)$

$$\lim_{\varepsilon\to0}\int \underline{f}(\underline{x})\cdot\underline{\omega}_\varepsilon(\underline{x})\,d\underline{x} = \int \underline{f}(\underline{x})\cdot\underline{\omega}(\underline{x})d\underline{x} \tag{2.4.10}$$

Two are the difficulties of this "conjecture", which is suggested by the success of the analogous result in dimension $d = 2$ in §2.3. The most evident is, perhaps, that in this $d = 3$ case it is no longer possible to neglect the autointeraction of the filament. It is already so in the simple case of a circular filament of vorticity $\underline{u}$. Indeed at a point $\xi\in\gamma$, if R is the radius of the circle γ, it will be

$$\underline{\dot\rho} = \frac{\Gamma}{4\pi}\oint_\gamma d\underline{\rho}'\wedge\frac{\underline{\rho}-\underline{\rho}'}{|\underline{\rho}-\underline{\rho}'|^3} \tag{2.4.11}$$

showing that $\underline{\dot\rho}$ is orthogonal to the plane of the filament and it has size v

$$v = \frac{\Gamma}{4\pi}\int_0^{2\pi} 2R^2\,d\alpha\frac{\sin^2\frac{\alpha}{2}}{|(2R\sin\alpha/2)^3|} = \frac{\Gamma}{2^4\pi R}\int_0^{2\pi}\frac{d\alpha}{|\sin\alpha/2|} = \infty \tag{2.4.12}$$

More generally $\underline{u}$ diverges near every point of γ where there is a curvature $R^{-1} > 0$. Hence it has no meaning to consider the evolution of the filament.

A further difficulty, independent of the previous one, is that a generic solenoidal velocity field *does not have* a corresponding vorticity field whose flux lines are closed. Indeed in general the flux lines of the field $\underline{\omega} = \text{rot}\,\underline{u}$, although they cannot "terminate", they will wander around densely filling regions of R^3 without ever closing (*c.f.r.* [1.6.20]). Hence it is not very natural to think of an arbitrary divergenceless velocity field as "well" approximated by fields with closed flux lines.

The latter is an aspect in which the 3–dimensional fluid is deeply different from a 2–dimensional one, in which instead an arbitrary vorticity field is naturally thought of as a limiting case of a field in which vorticity is concentrated in points.

Among the two difficulties the second looks less serious: after all it is a difficulty that can certainly be circumvented by contenting ourselves with approximations of $\underline{\omega}$ with a system of closed vorticity filaments in a sense weak enough and, of course, we are quite free *a priori* to define the meaning of the "approximation" as we wish. The more so as it is an "external" approximation which, therefore, can only be justified *a posteriori*.

The first difficulty is, however, almost "uneliminable".

A way to eliminate it could be to consider filaments so *irregular* to have an undefined tangent and, in fact, such that $d\underline{\rho}' \wedge (\underline{\rho} - \underline{\rho}')$ oscillates so strongly in sign and size to produce a finite result for the integral, (2.4.12), defining the velocity field on the filament points.

Alternatively we could imagine filaments with flux Γ "vanishing" in a sense to define so that the velocity in (2.4.12) is finite.

(B) Thin filament. Smoke ring.

We shall examine the second possibility first, and proceed heuristically to derive the equations of motion of a vorticity filament with "*evanescent*" vorticity, or "*thin filament*".

Given a regular closed curve γ let γ_δ be a tiny tube with radius δ centered around it: imagine that in γ_δ a vorticity field is defined and directed as the tangent $\underline{t}(\underline{x})$ to the curve parallel to γ through $\underline{x} \in \gamma_\delta$. Here it is not important to specify in which sense the tiny tube is filled by curves "parallel" to γ because the result will not depend on such details.

The vorticity $\underline{\omega}$ will therefore be $\Gamma_\delta \sigma_\delta(\underline{x})\underline{t}(\underline{x})$ where $\sigma_\delta(\underline{x})$ is a function that, in the direction perpendicular to γ, decreases in a regular way to 0 near the surface of the tiny tube. Moreover the integral over a section orthogonal to the tiny tube of $\sigma_\delta(\underline{x})$ is fixed to equal 1, so that the tiny tube is a flux tube of the vorticity field with flux Γ_δ.

The velocity field corresponding to the vorticity field $\underline{\omega}$ will be given by the Biot–Savart formula

$$\underline{u}(\underline{x}) = \frac{\Gamma_\delta}{4\pi} \int_{\gamma_\delta} d^3\underline{y}\, \sigma_\delta(\underline{y}) \frac{\underline{t}(\underline{y}) \wedge (\underline{x} - \underline{y})}{|\underline{x} - \underline{y}|^3} \tag{2.4.13}$$

where $\underline{t}$ denotes the unit tangent vector, and the calculation leading to (2.4.12) tells us that if $\Omega = \Gamma_\delta \log \delta$ is kept fixed while $\delta \to 0$, then

$$\lim_{\delta\to 0} \underline{u}_\delta(\underline{x}) = \frac{\Omega}{4\pi} \frac{1}{R(\underline{x})} \underline{b}(\underline{x}) \tag{2.4.14}$$

if $\underline{b}(\underline{x})$ is the unit vector *binormal* (we recall that this is the unit vector orthogonal to the plane of the tangent and the normal) to the curve γ and $R(\underline{x})$ is the radius of curvature of the curve γ in $\underline{x}$.

We then say that the velocity field of a *thin filament* γ with intensity Ω is

$$\underline{u}(\underline{x}) = \frac{\Omega}{4\pi} \frac{1}{R(\underline{x})} \underline{b}(\underline{x}) \qquad \underline{x} \in \gamma \tag{2.4.15}$$

which is sometimes called the "smoke rings equation", because it is a model for the motion of smoke rings, as long as they remain thin and well delimited. A more appropriate name is the equation for the *motion by curvature* of the curve γ.

The simplest case is when γ is a circle of radius R. In such case the (2.4.15) tells us that the circle moves by uniform rectilinear motion orthogonally to its own plane, with velocity $\Omega/4\pi R$ oriented to see the flux on the circle proceed counterclockwise.

The general case can also be "*exactly*" studied: *i.e.* the motion of thin filaments is integrable by quadratures! This is a *very remarkable* result of Hasimoto, *c.f.r.* [Ha72],[DS94].

The key remark is that (2.4.15) implies that the curve moves *without stretching*: the arc length of the curve is invariant. This can be seen form (2.4.3) which shows that the stretching of a vector oriented as the line element is proportional to $\underset{\sim}{\omega} \cdot \underset{\sim}{\partial} \underline{u}$, *i.e.* to $\underline{t} \cdot \partial_s (R^{-1} \underline{b}) = 0$: because the derivative of the binormal unit vector with respect to the curvilinear abscissa is proportional to the normal unit vector, by Frenet formulae, (2.4.16).

The inextensibility of the curve γ during its evolution by curvature allows us to label its points by their curvilinear abscissa with origin on a prefixed point of the curve. During the evolution the points of γ will keep the same abscissa on γ.

It is then important to recall the *Frenet's formulae* that express, on a curve γ, how the three unit vectors $\underline{T} = (\underline{t}, \underline{n}, \underline{b})$, *tangent, normal and binormal* to it, change with the curvilinear abscissa in terms of the *radius of curvature* R and of the *torsion* τ as

$$\partial_s \begin{pmatrix} \underline{t} \\ \underline{n} \\ \underline{b} \end{pmatrix} = \begin{pmatrix} 0 & R^{-1} & 0 \\ -R^{-1} & 0 & -\tau \\ 0 & \tau & 0 \end{pmatrix} \begin{pmatrix} \underline{t} \\ \underline{n} \\ \underline{b} \end{pmatrix} \tag{2.4.16}$$

where torsion and curvature are computed at the point of γ with curvilinear ascissa s, see problem [3.4.1] below.

If we set

$$\psi(s,t) = \frac{1}{R(s,t)} e^{i\sigma(s,t)} \qquad \sigma(s,t) = \kappa(t) + \int_0^s \tau(s',t)\, ds' \tag{2.4.17}$$

where $\kappa(t)$ is a suitable function of t, then *Hasimoto's theorem* can be formulated as

Theorem *(Integrability of the motion by curvature): The function $\psi(s,t)$ satisfies the nonlinear Schrödinger equation*

$$i\frac{4\pi}{\Omega}\partial_t\psi = \partial_s^2\psi + \frac{1}{2}|\psi|^2\psi \tag{2.4.18}$$

which is an equation which is integrable by quadratures.

Therefore, this tells us that R, τ vary, in a sense, quasi–periodically, *c.f.r.* [CD82].

The derivation of (2.4.18) from Frenet's formulae and from the equation of motion by curvature, (2.4.15), is discussed in the problems [2.4.1]%[2.4.3].

Once (2.4.18) is solved the curvature and the torsion at a generic point are known as functions of time on the inextensible curve γ. Therefore the Frenet formulae allow us to compute, always via a quadrature, the unit vectors $\underline{t}(s), \underline{n}(s), \underline{b}(s)$ as functions of time. Hence (2.4.15), with a further quadrature, will also give the actual positions in space of the points of γ as functions of their initial position (which plays the role of a label for the points of γ).

Hence the problem is "completely soluble" by quadratures. However this is not the appropriate place to discuss the qualitative features of the smoke ring motions: it is clear that the motions of thin filaments is not strongly related with the problem of an external algorithm for soving the Euler equations in $d = 3$ which would require considering filaments which are *not thin*, as it already appears from the analysis in (A).

Nevertheless the problem just discussed has some relation with the Euler equation. With the notations introduced above imagine $\delta > 0$ and $\Omega = \Gamma_\delta \log\delta$ fixed and consider the solution of the Euler equation with initial vorticity $\underline{\omega}_\delta(\underline{x})$. Then one can ask whether the following property is valid

Assuming that such solution exists for all times $t > 0$ denote it $\underline{u}_\delta(\underline{x}, t)$. Is then the limit: $\lim_{\delta\to 0} \underline{u}_\delta(\underline{x}, t) = \underline{u}(\underline{x}, t)$ existing? And, if yes, is such limit the solution of the Hasimoto equation of the curvature motion with initial curve γ?

Answering is difficult and, in general, the problem is open. However suppose that the curve γ is a circle, and γ_δ is a tube obtained as the region swept by a disk of radius δ centered at a point of γ and orthogonal to γ by letting the center glide on γ. Suppose also that initial vorticity field

$\underline{\omega}_\delta$ is everywhere perpendicular to this disk. Then the answer to the just posed questions is *affirmative*, see [BCM00]. This is interesting because it clarifies the meaning and the importance of the heuristic considerations at the beginning of this section (B).

(C) Irregular filaments: Brownian filaments.

We now go back to the problem of devising external algorithms for the $d = 3$ Euler equations posed in (A) above and we shall consider the case of an irregular curve. It is then possible that even the velocity given by (2.4.8), suitably interpreted (because the contour integral cannot be, *a priori*, condidered defined over irregular curves) is finite.

The latter possibility can be illustrated through a simple example in which one can see that a very irregular ("fractal") curve can generate a velocity field that is, in some sense, finite on the curve itself.

Consider a curve with parametric equations in cylindrical–polar coordinates (polar on the x, y–plane) is

$$r(\alpha) = R\,(1 + \varepsilon(\alpha)), \qquad z = R\,\vartheta(\alpha) \tag{2.4.19}$$

with $\varepsilon(\alpha), \vartheta(\alpha)$, for the time being, arbitrary.

Setting $\varepsilon(0) = \varepsilon_0, \vartheta(0) = \vartheta_0,\ \varepsilon = \varepsilon(\alpha), \vartheta = \vartheta(\alpha)$ and calling ε', ϑ' the derivatives of ε, ϑ in α and $s \equiv \sin\alpha,\ c \equiv \cos\alpha$ we see that

$$\begin{aligned} d\underline{\rho}' &= (-R(1+\varepsilon)s + R\varepsilon' c\,,\ R(1+\varepsilon)c + R\varepsilon' s\,,\ R\vartheta')\,d\alpha \\ \underline{\rho}' - \underline{\rho} &= (R(1+\varepsilon)c - R(1+\varepsilon_0)\,,\ R(1+\varepsilon)s\,,\ R(\vartheta - \vartheta_0)) \end{aligned} \tag{2.4.20}$$

hence, setting also $\eta = \varepsilon - \varepsilon_0$, $\mu = \vartheta - \vartheta_0$:

$$|\underline{\rho}' - \underline{\rho}|^2 = (2(1-c)(1 + \varepsilon + \varepsilon_0 + \varepsilon\varepsilon_0) + \eta^2 + \mu^2)R^2 \tag{2.4.21}$$

while the components of $d\underline{\rho}' \wedge (\underline{\rho}' - \underline{\rho})$ are immediately computed from (2.4.20) and are

$$\begin{aligned} &\frac{d\underline{\rho}'}{d\alpha} \wedge (\rho' - \rho) = \\ &= \begin{cases} R^2[(1+\varepsilon)c\mu + \varepsilon' s\mu - \vartheta'(1+\varepsilon)s], \\ R^2[\vartheta'(1+\varepsilon)(c-1) + \vartheta'\eta + \mu(1+\varepsilon)s - \mu\varepsilon' c], \\ R^2[-(1+\varepsilon)\eta - (1+\varepsilon)(1+\varepsilon_0)(1-c) + (1+\varepsilon_0)\varepsilon' s] \end{cases} \end{aligned} \tag{2.4.22}$$

Suppose that the curve is chosen as a sample in an ensemble of curves randomly drawn with a probability distribution such that, as α varies near $\alpha_0 = 0$, the quantities ε, ϑ are mutually independent and each is a random function with independent increments. This means in particular that we assume, for $\alpha_1 < \alpha_2 < \alpha_3$, that the quantity $\varepsilon(\alpha_2) - \varepsilon(\alpha_1)$ is, as a random variable, "very little" (see below) dependent from $\varepsilon(\alpha_3) - \varepsilon(\alpha_2)$, and suppose the analogous property on ϑ. Suppose, furthermore, for simplicity, that the fluctuations of ε, η satisfy, for α_1, α_2 near 0, a continuity property like for instance

$$\langle (\varepsilon(\alpha_2) - \varepsilon(\alpha_1))^2 \rangle \le (D|\alpha_2 - \alpha_1|)^{2-2a} \tag{2.4.23}$$

with $D, a > 0$ (the parameter a is a measure of lack of regularity of the random curves); finally suppose that the large fluctuations have small probability (for instance bounded by a Gaussian function). In reality under the above hypotheses one expects that $a \equiv 1/2$ and that the distribution of the increments of $\varepsilon(\alpha)$ and $\vartheta(\alpha)$, for the considered values of α is necessarily Gaussian. Hence it is a distribution of the kind that one encounters in the theory of Brownian motion.

Intuitively we imagine that the random functions ε, ϑ are continuous with probability 1 but have, for α close to $\alpha_0 = 0$, increments between α_1 and α_2, proportional to $(D|\alpha_2 - \alpha_1|)^{1/2}$: hence they are *not differentiable*.

We pose the problem of whether, at least, the *average velocity* of the curve in the point $\underline{\rho}$ corresponding to $\alpha = 0$ is finite. Velocity is given by the integral (2.4.11) which, considering the (2.4.22), in the point $\alpha_0 = 0$ becomes

$$\begin{aligned} \underline{u} = &\frac{\Gamma}{4\pi R} \int \frac{d\alpha}{(\alpha^2 + \eta^2 + \mu^2)^{3/2}} \cdot \\ &\cdot (\mu + \alpha\varepsilon'\mu - \vartheta'\alpha, -\vartheta'\alpha^2/2 + \vartheta'\eta + \alpha\mu - \mu\varepsilon', -\eta - \alpha^2/2 + 2\varepsilon'\alpha) \end{aligned} \tag{2.4.24}$$

where we set $1+\varepsilon \simeq 1, \cos\alpha - 1 = -\alpha^2/2, \sin\alpha = \alpha$ for simplicity, imagining that

(1) ε and ϑ are small perturbations (although random) and

(2) taking into account that the convergence problems in the above analysis are due to what happens for $\alpha \simeq 0$.

The quantities ε', ϑ' suffer from interpretation problems because, by assumption, such derivatives have no meaning: but (2.4.24) and the formal expression $\varepsilon'(\alpha) = (\varepsilon(\alpha + \delta) - \varepsilon(\alpha))/\delta$, in the limit as $\delta \to 0$, shows that by the independence of the increments of ε and ϑ, the terms containing ε' and ϑ' can be considered as contributing zero to the average of (2.4.22), (2.4.24).

Discarding the terms that contain ε', η' we see that the only component of $\underline{u}$ that has nonzero average is the third, and that such component has average

$$\begin{aligned} v = &\frac{\Gamma}{8\pi R} \langle \int d\,\alpha \frac{-\alpha^2}{(\alpha^2 + \eta^2 + \mu^2)^{3/2}} \rangle = \\ = &\frac{-\Gamma}{8\pi R} \int d\alpha d\eta d\mu \, \frac{\alpha^2}{(\alpha^2 + \eta^2 + \mu^2)^{3/2}} f_\alpha(\eta) g_\alpha(\mu) \end{aligned} \tag{2.4.25}$$

where $f_\alpha(\eta), g_\alpha(\mu)$ are the probability distributions of $\eta = \varepsilon(\alpha) - \varepsilon(0), \mu = \vartheta(\alpha) - \vartheta(0)$.

To take advantage, in a simple way, of the assumptions on the distributions of η, ϑ it is convenient setting

$$\eta = (D|\alpha|)^{1/2}\bar{\eta}, \qquad \vartheta = (D|\alpha|)^{1/2}\bar{\vartheta} \tag{2.4.26}$$

and suppose that the variables $\bar{\eta}, \bar{\vartheta}$ have an *α–independent, Gaussian, distribution*: this simplifies some formal aspects of the calculations. We find

$$\begin{aligned} |v| \le &\text{cost} \int d\alpha d\,\overline{\eta} d\,\overline{\mu}\, e^{-\bar{\eta}^2-\bar{\mu}^2} \frac{\alpha^2}{(\alpha^2 + D|\alpha|(\bar{\eta}^2+\bar{\mu}^2))^{3/2}} \simeq \\ \simeq &\text{cost} \int d\alpha\, x dx\, e^{-x^2} \frac{\alpha^2}{(\alpha^2 + D|\alpha|\, x^2)^{3/2}} \simeq \\ \simeq &\text{cost} \int \frac{\alpha^2\, d\alpha}{(\alpha^2)} < \infty \end{aligned} \tag{2.4.27}$$

This remark on the finiteness of $\underline{u}$ admits the following generalization. Imagine a vorticity filament with equations

$$\alpha \to \underline{\rho}(\alpha) + \underline{\xi}(\alpha) \tag{2.4.28}$$

with a C^∞ function $\underline{\rho}(\alpha)$ and with $\underline{\xi}(\alpha)$ sample of a random trajectory that, locally near every one of its points, is "essentially" a Brownian motion. The analysis leading to (2.4.27) can be extended to classes of curves that are periodic and with increments that are "very little" mutually dependent. An example of classes of curves with these properties is illustrated in the problems.[1]

Then the evolution equation for the generic point on the curve, which we shall label by $\underline{\rho}_0 + \underline{\xi}_0$ with $\underline{\rho}_0 = \underline{\rho}(\alpha_0)$, $\underline{\xi}_0 = \underline{\xi}(\alpha_0)$,*c.f.r.* (2.4.28), is written

$$\begin{aligned} \frac{d(\underline{\rho}_0 + \underline{\xi}_0)}{dt} = &- \frac{\Gamma}{4\pi} \int \frac{(d\underline{\rho}' + d\underline{\xi}') \wedge (\underline{\rho}' + \underline{\xi}' - \underline{\rho}_0 - \underline{\xi}_0)}{|\underline{\rho}' - \underline{\rho}_0 + \underline{\xi}' - \underline{\xi}_0|^3} \stackrel{\text{def}}{=} \\ \stackrel{\text{def}}{=} &- \frac{\Gamma}{4\pi} \int \frac{d\underline{\rho}' \wedge (\underline{\rho}' + \underline{\xi}' - \underline{\rho}_0 - \underline{\xi}_0)}{|\underline{\rho}' - \underline{\rho}_0 + \underline{\xi}' - \underline{\xi}_0|^3} \end{aligned} \tag{2.4.29}$$

where having eliminated the terms $d\underline{\xi}'$, that correspond to the terms with the derivatives of ε, ϑ in (2.4.24), is in a sense a step analogous to having eliminated, in $d = 2$, the autointeraction terms of the vortices (see (2.3.9)), and it is hopefully justified by what we have seen in the above particular case in which these autointeraction terms between the vortices gave formally a zero contribution to the average velocity.

[1] Note that we cannot assume that $\underline{\xi}(\alpha)$ it to be exactly a Brownian path with α as time variable because the increments cannot be really independent because the curve must, in the end, be closed. The precise meaning that is given to the motion $\underline{\xi}(\alpha)$ is discussed in detail in the problems following [2.4.4].

The equation (2.4.29) generates then two equations, one for the *average* $\underline{\rho}$ and one for the *fluctuations* $\underline{\xi}$

$$\begin{aligned}\frac{d\underline{\rho}}{dt} &= -\langle\frac{\Gamma}{4\pi}\int\frac{d\underline{\rho}'\wedge(\underline{\rho}'+\underline{\xi}'-\underline{\rho}-\underline{\xi})}{|\underline{\rho}'-\underline{\rho}+\underline{\xi}'-\underline{\xi}|^3}\rangle\\ \frac{d\underline{\xi}}{dt} &= -\frac{\Gamma}{4\pi}\int\frac{d\underline{\rho}'\wedge(\underline{\rho}'+\underline{\xi}'-\underline{\rho}-\underline{\xi})}{|\underline{\rho}'-\underline{\rho}+\underline{\xi}'-\underline{\xi}|^3}-\frac{d\underline{\rho}}{dt}\end{aligned} \tag{2.4.30}$$

To show the correctness of (2.4.30) should mean something similar to what can be shown in $d = 2$. But before proceeding we must stress that now the problem is rather more involved and an answer is not known.

Suppose that $\underline{x} \to \underline{\omega}(\underline{x}) \in C^\infty$ and that the flux lines of the vorticity feld $\underline{\omega}$ are *all closed*.[2] Imagine to cut orthogonally the flux lines by a surface and to pave the suface with small squares of side $\approx \lambda$ and let $\underline{x}_j$ be the center of the j-th square and let Γ_j^λ be the flux of $\underline{\omega}$ through the j-th square. Call γ_j the flux curve of $\underline{\omega}$ passing through $\underline{x}_j$. Then the vorticity field

$$\underline{\omega}^\lambda(\underline{x}) = \sum_j \Gamma_j^\lambda \delta_{\gamma_j}(\underline{x})\,\underline{t}_{\gamma_j}(\underline{x}) \tag{2.4.31}$$

approximates *weakly* the field $\underline{\omega}$ in the limit in which the size λ of the squares tends to 0.

We can now imagine to evolve the curves $\gamma_j + \underline{\xi}_j$, where $\underline{\xi}_j$ is a sample of a random motion "similar" to a Brownian motion but with periodic sample paths and with a suitable mean square dispersion D_j^λ (see problem [2.4.4] and following ones for an example) and compute the vorticity field at time t by using the (2.4.30).

We ask the question whether it is possible to determine D_j^λ so that the vorticity field at time t has a weak limit as $\lambda \to 0$, converging to a *regular* vorticity field solving the Euler equation with initial datum $\underline{\omega}$. It is by no means clear that this or something similar to this could be true.

Heuristically we may expect that by choosing $D_j^\lambda \equiv D^\lambda$, j–independent, and setting $\underline{\omega}^D(\underline{x}, t) = \lim_{\lambda\to 0}\underline{\omega}^\lambda$, then the limit $\lim_{D\to 0}\underline{\omega}^D \equiv \underline{\omega}(\underline{x}, t)$ should satisfy the Euler equations. This should hold *even if* the fluctuations $\underline{\xi}$ are fixed as time independent, thus leading us to consider just the first of the (2.4.30) as a closed system of equations (because now $\underline{\xi}$ has, by assumption, the same distribution at all times).

Hence from the first of the (2.4.30):

$$\frac{d\underline{\rho}}{dt} = -\langle\frac{\Gamma}{4\pi}\int\frac{d\underline{\rho}'\wedge(\underline{\rho}'+\underline{\xi}'-\underline{\rho}-\underline{\xi})}{|\underline{\rho}'-\underline{\rho}+\underline{\xi}'-\underline{\xi}|^3}\rangle \tag{2.4.32}$$

[2] Which is general is not true, even when $\underline{\omega}$ vanishes outside a bounded region, but which constitutes an interesting class of cases.

with $\underline{\xi}(t)$ defined by a time independent distribution (rather than by the second of the (2.4.30)), is an evolution equation that is interesting in itself, even if it turns out to be only indirectly related to the evolution problem for a vorticity filament. In fact to give a meaning to the motion of a filament of vorticity it is necessary to consider both the equations in (2.4.30) and we see that the same "average filament" evolves in a different way depending on the distribution of the initial fluctuations $\underline{\xi}$, *i.e.* depending on the actual structure that is assigned to the filament.

The (2.4.32) and (2.4.30) give a method to give a meaning to the evolution of a filament (and to the notion of filament itself) alternative to the one, also natural, of considering the filament as a tiny vorticity tube initially with a constant section and to follow its evolution. From the viewpoint of numerical simulations the (2.4.32) and (2.4.30) are somewhat simpler than the equations arising from considering the tiny tube model for the vorticity field, because the objects that are described are, respectively one and two–dimensional while the tubes are 3–dimensional.

We see in this way the generation of the idea of making even more "external" the approximation algorithm by using as vorticity filaments, rather than regular closed curves, very irregular curves like the samples of an ensemble of random curves with a probability distribution that assigns essentially independent increments to the coordinates of their points.

But wishing to avoid such radically "external" algorithms of solution of the Euler equations, which present to us obvious conceptual and computational difficulties, it would be necessary to give up using vorticity based algorithms that work so well in 2–dimensional fluids. Therefore abandoning them should be only a "last resrt" because the computational difficulties do not seem overwhelming, as proved by the existence of empirical solution methods for the NS equation (which is an equation of similar complexity), *c.f.r.* [Ch82], [Ch88].

(D) Irregular filaments: quasi periodic filaments.

Another road to pursue for an alternative generalization of the 2-dimensional vorticity algorithms can be obtained by concentrating the vorticity, rather than on closed lines, on lines that are *not* closed and fill densely 2–dimensional or even 3–dimensional surfaces.

Such velocity fields can be observed in real experiments, think for instance to real smoke rings that move in air.

If the filament lines are distributed densely on the surface of a 2–dimensional torus, for instance, or fill its interior, the rotation of $\underline{u}$ for each of them must be infinitesimal and only their density will make sense.

Consider, as an example, the case of a filament filling densely the surface of a torus $\mathcal{T}$ in the simple case in which $\mathcal{T}$ is a 2-dimensional torus and on $\mathcal{T}$ the flux line of $\underline{\omega}$ is a dense quasi periodic trajectory.

We imagine that the torus $\mathcal{T}$ is tangent to the x_1, x_2 plane at the origin and that it has there an external normal parallel to the x_3 axis. The torus will be represented parametrically as

$$\begin{aligned} x_1 =& X_1(\xi_1,\xi_2),\ x_2 = X_2(\xi_1,\xi_2),\ x_3 = X_3(\xi_1,\xi_2) \qquad \text{with}\\ X_1(\xi_1,\xi_2) =& \xi_1 + O(\underline{\xi}^2),\ X_2(\xi_1,\xi_2) = \xi_2 + O(\underline{\xi}^2),\ X_3(\xi_1,\xi_2) = O(\underline{\xi}^2) \end{aligned} \tag{2.4.33}$$

where the $\underline{\xi}$ are angles on a standard torus $\underline{\xi} \in [0,2\pi]^2 \stackrel{def}{=} \mathcal{T}$.

On $\mathcal{T}$ we imagine the curve φ with equations $s \to \underline{\xi}(s) = (s, \eta s)$, $0 \le s \le q_n$, with $\eta =$ irrational and $\eta = \lim_{n\to\infty} p_n/q_n$ where p_n and q_n are the "convergents" of the continued fraction for η(cf. problem [5.1.7] below) p.97). Note that the curve φ will fill densely the torus and that, therefore, if n is large the closed curve φ_n with equations $s \to \underline{\xi}(s) = (s, p_n s/q_n)$ "essentially draws" the torus.

Let $\underline{\tau}$ be the unit vector tangent at the origin to $\mathcal{T}$ in the direction of the curve φ and let $\underline{\nu}$ be the unit vector orthogonal to $\underline{\tau}$ and tangent to the torus at the origin. Consider a surface element $d\sigma = dh\,d\ell$ around the origin where $d\ell$ is the size in the direction of $\underline{\tau}$ and dh is the size in the direction of $\underline{\nu}$.

The sum of the lengths of the segments of the curve φ_n that are contained in $d\sigma$ will be $N = q_n d\sigma/S$, asymptotically in n, where S is a geometric constant (by the "ergodicity" of quasi periodic motions).

We imagine that a vorticity field $\underline{\omega}_n$ is concentrated on the curve φ_n and is parallel to it: so that $\underline{\omega}_n = \gamma q_n^{-1} \underline{\tau}\, \delta_{\varphi_n}(\underline{\xi})$ where δ_{φ_n} is a Dirac's delta distribution uniformly distributed along the curve φ_n. If we now let $n \to \infty$ we see that, for any smooth function $f(\underline{\xi})$, it is

$$\int f(\underline{\xi})\, \underline{\omega}_n(\underline{\xi})\, d\underline{\xi} \xrightarrow[n\to\infty]{} \int \underline{\omega}(\underline{\xi}) f(\underline{\xi})\, d\underline{\xi}, \qquad \underline{\omega}(\underline{\xi}) = \underline{\tau}(\underline{\xi}) \delta_{\mathcal{T}}(\underline{\xi})\, d\underline{\xi} \tag{2.4.34}$$

where $\delta_{\mathcal{T}}$ is a Dirac distribution is concentrated on the surface $\mathcal{T}$ and there proportional to the surface area (the proportionality constant is a function on $\mathcal{T}$ that depends on the actual shape of $\mathcal{T}$ (*i.e.* on the parametric equations $\underline{X}(\underline{\xi})$ in (2.4.32)).

We interpret $\underline{\omega}(\underline{\xi})$ as a vorticity field concentrated on the quasi periodic filament φ on $\mathcal{T}$.

This vorticity distribution induces a velocity $\underline{u}(\xi)$ which is finite. Indeed we can compute it at the origin (to fix the ideas) supposing first that the torus is flat in the vicinity of the origin; setting $\underline{\rho} = (l,h,z)$, $\underline{\tau} = (0,1,0)$, $\underline{\nu}=$ (normal to the torus at the origin), the contribution to the velocity by a neighborhood of size of order ε around the origin is, by the Biot–Savart formula (2.4.8), proportional to the (improper) integral

$$\int_{|l|<\varepsilon} \gamma \frac{\underline{\tau} \wedge \underline{\rho}}{|\underline{\rho}|^3} \delta(z)\, dh\, dz\, dl = \int_{|l|<\varepsilon} \gamma \frac{l\underline{\nu}\, dh\, dl}{(h^2+l^2)^{3/2}} \equiv 0 \tag{2.4.35}$$

for small ε, by parity. In the general case in which the torus has curvature at the origin, or near it, this means that the Biot–Savart integral (which would be logarithmically divergent by "power counting") defining $\underline{u}$ is in fact (improperly) convergent.

If we really concentrate the vorticity on φ_n, then we see that the above interpretation of the limit as $n \to \infty$ simply results from the limit as $n \to \infty$ of the velocity field $\underline{u}_n$ *deprived* of a part that is improperly defined as $n \to \infty$: this is (again) analogous to the prescription in the $d = 2$ case of point vortices in which the deleted part (last term in (2.3.9)) could be interpreted as a rotation of the vortex around itself and with infinite angular speed.

Similarly if $\mathcal{T}$ is a domain bounded by a torus and densely filled by flux lines and if the vorticity on its interior can be considered distributed with a density $\underline{\omega}$ then the integral expressing the value of the field $\underline{u}$ is convergent.

There is, therefore, also the possibility of studying the evolution of a family of flux tubes of dimension 2 or 3 densely filled by one more filament by letting the latter evolve to be transported by the current lines generated by itself. Alternatively one can study cases in which the vorticity is concentrated on surfaces ("vorticity sheets") or in volumes ("tubes") and such as to approximate some smooth vorticity field. In such cases there would be no problem in giving a meaning to the Biot–Savart integral, at least at the initial time.

In practice the algorithm seems simpler in the case of a vorticity filament concentrated on a line dense on a surface, if compared to the case of a vorticity "sheet" concentrated on a surface. But the convergence problems of all the above algorithms are very little studied and only on a empirical (numerical) basis, [Ch82].

Problems.

[2.4.1]: (*time derivative of the principal frame on a curve*) Consider a closed moving curve γ. Show that the three orthogonal vectors $\underline{T} = (\underline{t}, \underline{n}, \underline{b})$ evolve so that there exist three functions A, B, C such that

$$\partial_t \begin{pmatrix} \underline{t} \\ \underline{n} \\ \underline{b} \end{pmatrix} = \begin{pmatrix} 0 & A & B \\ -A & 0 & C \\ -B & -C & 0 \end{pmatrix} \begin{pmatrix} \underline{t} \\ \underline{n} \\ \underline{b} \end{pmatrix} \stackrel{def}{=} M\underline{T} \tag{2.4.36}$$

(*Idea:* Since the three vectors are orthonormal they must evolve as $t \to O(t)\underline{T}(t)$ where $O(t)$ is a rotation matrix. Then the matrix $\dot{O}(t) = MO(t)$ with M an antisymmetric matrix).

[2.4.2]: (*principal frame motion and Frenet relations*) Consider the motion by curvature, *i.e.* according to (2.4.15), of a curve (necessarily inextensible) γ. Then the points $\underline{x} = \underline{r}(s)$ are labeled by their curvilinear abscissa s and, therefore, as time varies their positions can be expressed via a function $\underline{x} = \underline{\rho}(s, t)$. Writing $\partial_s \underline{T} = F\underline{T}$ and $\partial_t \underline{T} = M\underline{T}$, the Frenet formulae (2.4.17) and the relations in [2.4.1], show that the identities $\partial_t\partial_s\underline{T} = \partial_s\partial_t\underline{T}$ imply the relations

$$\begin{aligned} \partial_t R^{-1} =& \partial_s A + B\tau \\ 0 =& \partial_s B - R^{-1}C - \tau A \\ \partial_t \tau =& -\partial_s C - R^{-1}B \end{aligned}$$

and, setting $\Omega' = \Omega/4\pi$, the $\partial_t \underline{\rho} = \underline{u} = \Omega' R^{-1} \underline{b}$ imply $\partial_s(R^{-1}\underline{b}) = \partial_t(\partial_s \underline{r})$, which in turn imply

$$A = \Omega' \frac{\tau}{R}, \quad B = \Omega' \partial_s R^{-1}$$

hence A, B, C are uniquely determined by R, τ. (*Idea:* For the first relations simply differentiate the (2.4.36) with respect to t and (2.4.16) with respect to s using (2.4.16) and, respectively, (2.4.36) to express the ∂_t and, respectively, the ∂_s of the unit vectors: one gets six relations each of the above ones being obtained twice. Proceed similarly for the second relations, by taking also into account that $\partial_s \underline{\rho} = \underline{t}$.)

[2.4.3]: (*Hasimoto's theorem*) Starting from the expressions in [2.4.2] for A, B, C check that the equations for R, τ are

$$\partial_t \begin{pmatrix} R^{-1} \\ \tau \end{pmatrix} = \begin{pmatrix} \partial_s A + B\tau \\ -\partial_s C - R^{-1} B \end{pmatrix} \qquad \partial_s B = R^{-1} C + \tau A$$

and check that, setting $\psi = R^{-1} e^{i\sigma}$, $\sigma = \kappa(t) + \int_0^s \tau(s')ds'$, $\kappa(t) = \int_0^t \Big(2^{-1} R(0,t')^{-2} - C(0,t')\Big)\, dt'$, the "Hasimoto identity" holds, *i.e.* the ψ satisfies the nonlinear Schrödinger equation, (2.4.18).

[2.4.4]: (*A gaussian process*) Consider the periodic functions in $L_2([0,2\pi])$, $\alpha \to \varepsilon_N(\alpha)$ with zero average and with only N harmonics. These are the functions that can be expressed as: $\varepsilon_N(\alpha) = \pi^{-1} \sum_{k=1}^N (c_k \cos k\alpha + s_k \sin k\alpha)$. Define a probability distribution on the set of functions of the type considered, by assigning to the coefficients c_k, s_k the Gaussian distribution

$$\prod_{k=1}^N \frac{e^{-\frac{1}{2}(c_k^2+s_k^2)k^2} dc_k\, ds_k}{\sqrt{2\pi k^{-2}}^2}$$

and show that $\langle(\varepsilon_N(\alpha) - \varepsilon_N(\beta))^2\rangle = 2\,\pi^{-1} \sum_{k=1}^N k^{-2}(1 - \cos k(\alpha-\beta)) \equiv C_N(\alpha-\beta) < |\alpha-\beta|_{2\pi}$ where $|\alpha-\beta|_{2\pi} = \min_n |\alpha-\beta-2\pi n|$ and also that $C_N(x) \xrightarrow[N\to\infty]{} \pi(|\alpha-\beta|_{2\pi} + O((|\alpha-\beta|_{2\pi}^2)^2)$. *The Gaussian process defined in [2.4.4] and discussed in problems following [2.4.4] has periodic sample paths: it differs therefore from the usual brownian motion. However the difference is quite trivial, see [IM65] p.21, problem 3.* (*Idea:* Note that the series limit of $C_N(x)$ as $N \to \infty$ is the Fourier series for the function $|x|_{2\pi} - |x|_{2\pi}^2/2$ in the interval $[-\pi,\pi]$.)

[2.4.5]: In the context of [2.4.4] show that the probability that $|\varepsilon_N(\alpha) - \varepsilon_N(\beta)|$ is larger than $\sqrt{\gamma C_N(\alpha-\beta)}$ is $2\int_\gamma^\infty e^{-\gamma^2} d\gamma/2/\sqrt{2\pi}$. (*Idea:* Note that $\varepsilon_N(\alpha) - \varepsilon_N(\beta)$ must have a Gaussian distribution with dispersion (or "width", or "covariance") $C_N(\alpha-\beta)$,because it is a linear combination of Gaussian random variables (the c_k, s_k).)

[2.4.6]: Show that the probability that, given two "dyadic" points $\alpha, \beta < \pi$ of order p, $\alpha = 2\,\pi\, h 2^{-p}$ and $\beta = \alpha + 2\,\pi\, 2^{-p}$ *adjacent* it is $|\varepsilon_N(\alpha) - \varepsilon_N(\beta)| > \gamma p 2^{-p/2}$ is estimated above by $P_p = c\, 2^p\, \gamma\, p\, \exp -\gamma^2 p^2/2$ for some constant $c > 0$. (*Idea:* The probability of the simultaneous validity of any number of events is bounded by the sum of the respective probabilities: hence the result follows immediately from the problem [2.4.5] because in this case the number of events is 2^p).

[2.4.7]: (*Wiener's theorem for brownian paths*) Given two dyadic points $\alpha = 2\pi h 2^{-p}$ and $\beta = 2\pi k 2^{-q}$, $\alpha, \beta < \pi$, not necessarily of the same order, show that, for instance, if $\alpha < \beta$ and $q < p$ there exists a sequence of n points $\alpha = x_1 \le x_2 \le x_3 \ldots \le x_n = \beta$ such that $x_{i+1} - x_i = 2^{-(p-i)}\sigma_i$ with $\sigma_i = 0, 1$ suitable. This means that x_i and x_{i+1} are either the same or adjacent "on scale" $2^{-(p-i)}$.
Deduce that the probability that $|\varepsilon_N(\alpha) - \varepsilon_N(\beta)| < \gamma\, |\alpha-\beta|^{1/2} \log|\alpha-\beta|^{-1}$ for *any pair* of dyadic points $\alpha, \beta < \pi$ can be estimated by $1 - C\gamma e^{-\gamma^2/2}$, for some constant $C > 0$ (Wiener theorem).

(*Idea:* First note that $(\beta-\alpha)/2\pi = (k2^{p-q}-h)/2^p$ so that expanding in base 2 the numerator $k2^{p-q}-h=\sum_j^{<p} n_j 2^j$ with $n_j=0,1$ we get the representation $(\beta-\alpha)/2\pi = \sum_j \sigma_j 2^{-j}$, with $\sigma_j=0,1$ and trivially related to the n_j.
Write $\varepsilon_N(\alpha)-\varepsilon_N(\beta)=\sum(\varepsilon_N(x_{i+1})-\varepsilon_N(x_i))$ and note that the probability that, *for all* the i, it is $|\varepsilon_N(x_{i+1})-\varepsilon_N(x_i)| < \gamma(p-i)2^{-(p-i)/2}$ is estimated, by the result of [2.4.6], by $1-C\gamma e^{-\gamma^2/2}$.
Moreover $\sum 2^{-(p-i)}\sigma_i \equiv |\beta-\alpha|/2\pi$ hence we get the inequality

$$\sum 2^{-(p-i)/2}(p-i)\sigma_i \le 12\,(|\beta-\alpha|/2\pi)^{1/2}\log_2 4\pi|\beta-\alpha|^{-1}$$

in fact if $\overline{p}$ is the smallest p_j for which $\sigma_j=1$ it is $2^{-\overline{p}} \le |\beta-\alpha|/2\pi < 2^{-\overline{p}+1}$ and $\sum_j p_j\sigma_j 2^{-p_j/2} \le \sum_{m=\overline{p}}^{\infty} m2^{-m/2} \le 12\,\overline{p}2^{-\overline{p}/2}$ from which the latter inequality follows.)

[2.4.8]: Hence the random functions (defined on the dyadics) $\varepsilon_N(\alpha)$ are *uniformly* Hölder continuous, with exponent $\sim 1/2$, and "modulus of continuity"

$$\gamma = \sup_{\alpha,\beta}\frac{|\varepsilon(\alpha)-\varepsilon(\beta)|}{(|\alpha-\beta|_{2\pi}\log|\alpha-\beta|_{2\pi}^{-1})^{1/2}}$$

that is finite with a probability tending to 1 for $\gamma\to\infty$. One can then consider the limit as $N\to\infty$ of the probability distribution on the space $C([0,2\pi])$ of the continuous functions generated by the Gaussian distribution P_N introduced in [2.4.4]. The measurable sets will be defined by the set of functions that in m prefixed angles $\alpha_1,\ldots,\alpha_m$ take values in prefixed intervals $I_1,\ldots,I_m$. Such sets are called "cylinders", for obvious reasons, and they play the role analogous to that of the intervals in the theory of integration of functions of one variables. Furthermore, (in analogy to the ordinary integration theory) all sets approximable will be measurable that can be approximated via a denumerable sequence of operations of union, intersection and complementation on a denumerable collection of cylindrical sets. Check that the probability of each cylindrical set converges to a limit as $N\to\infty$. One can check (using Wiener theorem) that the measure thus constructed is completely additive (*i.e.* if a cylinder can be represented as a countable union of other cylinders then its measure is the sum of the measures of the cylinders that add up to it); and that the set of Hölder continuous functions is measurable and has probability 1. One defines in this way a probability distribution ("periodic Brownian motion") on the space of the continuous functions which are Hölder continuous (even with exponent 1/2).

Bibliography: The tiny filaments theory is mainly taken from [DS94] which discusses some remarkable integrable extensions; see also [Ha72] and the important extension to the theory of motion by curvature in the case of discrete curves, [DS95]. Other pertinent references are [BCM00], [Ne64], [CD82].

CHAPTER III

Analytical theories and mathematical aspects

§3.1 Spectral method and local existence, regularity and uniqueness theorems for Euler and Navier–Stokes equations, $d \ge 2$.

One of the most immediate and elementary applications of the spectral method of §2.2 is the *local* existence, regularity and uniqueness theory of the solutions of the Euler and Navier–Stokes equations in arbitrary dimension.

We shall illustrate the theory only in the case in which Ω is a cube with opposite sides identified (*i.e.* "periodic boundary conditions").
Consider initial data $\underline{u}_0(\underline{x})$ and force density $\underline{g}(\underline{x})$ which are *analytic* in $\underline{x}$, hence such that

$$\begin{aligned} \underline{u}^0(\underline{x}) &= \sum_{\underline{k}\ne\underline{0}} \underline{\gamma}^0_{\underline{k}} e^{i\underline{k}\cdot\underline{x}}, \qquad \underline{\gamma}^0_{\underline{k}} = \overline{\underline{\gamma}}^0_{-\underline{k}}, \quad \underline{k}\cdot\underline{\gamma}^0_{\underline{k}} \equiv 0, \quad |\underline{\gamma}^0_{\underline{k}}| \le V e^{-\xi|\underline{k}|} \\ \underline{g}(\underline{x}) &= \sum_{\underline{k}\ne\underline{0}} \underline{g}_{\underline{k}} e^{i\underline{k}\cdot\underline{x}}, \qquad \underline{g}_{\underline{k}} = \overline{\underline{g}}_{-\underline{k}}, \quad \underline{k}\cdot\underline{g}_{\underline{k}} \equiv 0, \quad |\underline{g}_{\underline{k}}| \le \ G e^{-\xi|\underline{k}|} \end{aligned} \tag{3.1.1}$$

where V, G and ξ are suitable positive constants and $\underline{k}$ is a *nonzero* vector with components integer multiples of $k_0 = 2\pi L^{-1}$ and $|\underline{k}|$ will denote $\sum_i |k_i|$.[1] The averages $\underline{u}_{\underline{0}}$ and $\underline{g}_{\underline{0}}$ are supposed zero, as usual. To simplify calculations we shall assume $\xi \equiv \xi_0/k_0$ with $\xi_0 \le 1$ which, evidently, is not restrictive.
Then the following proposition holds

Proposition *("perturbative local" solution of NS): Consider the NS–equation in d–dimensions (d = 2, 3)*

$$\dot{\underline{u}} = \nu\,\Delta\,\underline{u} - \underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{u} - \rho^{-1}\,\underline{\partial}\,p + \underline{g}, \qquad \underline{\partial}\cdot\underline{u} = 0 \tag{3.1.2}$$

[1] This is a natural definition when $\underline{k}$ are the Fourier transform mode labels for an analytic function, *c.f.r.* [3.1.1].

with initial datum and force satisfying (3.1.1).
There is $B_0 > 0$ *such that if* $\xi_0 \stackrel{def}{=} k_0\xi \le 1$, $V_c \stackrel{def}{=} \nu L^{-1}$, $T_c \stackrel{def}{=} L^2\nu^{-1}$ *and* $V_0 \stackrel{def}{=} B_0\,(V + G\,T_c)\,\xi_0^{-d-1}$ *then (3.1.2), admits a solution* $\underline{u}(\underline{x},t)$ *analytic in* $\underline{x}$ *and* t *with*

$$|\underline{\gamma}_{\underline{k}}(t)| \le V_0 e^{-\xi_0|\underline{k}|/2k_0}, \qquad \text{for} \quad 0 \le t \le t_0 \stackrel{def}{=} T_c\big(\frac{V_c}{V_0}\big)^2 \tag{3.1.3}$$

and the solution is the unique one enjoying the above properties in the time interval $[0, t_0]$.

Remark:

(1) Hence a local existence and regularity theorem holds, and explicit estimates on the duration of the existence and regularity interval are possible: "analytic data evolve remaining analytic at least for a small enough time".
(2) Uniqueness holds even within a much wider class of solutions $t \to \underline{u}(\underline{x},t)$, $p(\underline{x},t)$: for instance it holds for solutions in class $C^1([0,t_0]\times\Omega)$. See problem [3.1.6] below.
(3) Existence regularity and uniqueness can be studied also in other classes of functions. For instance if $\underline{u}^0 \in W^m(\Omega)$ and $m \ge 4$ (so that since $4 > \frac{3}{2}+2$ one has, *c.f.r.* [2.2.20] that $\underline{u}^0 \in C^1(\Omega)$) one can easily show the existence of a time T_m such that the equation (3.1.3) admits a solution in $W^m(\Omega)$ for each $t \in (0, T_m)$: see problems. The analytic case considered here is more involved: but the result is perhaps more interesting, mainly because of the technique that is employed and that illustrates a method often used in other physics problems.

proof: The part concerning uniqueness is absolutely elementary and, as noted, it will show in reality uniqueness among solutions of class $C^1([0,t_0]\times\Omega)$ (hence, *a fortiori*, in the analytic class of the type stated in the proposition). Calling $\underline{u}_1$ and $\underline{u}_2$ two solutions of class $C^1([0,t_0]\times\Omega)$ let $\delta\underline{u} = \underline{u}_1 - \underline{u}_2$ and consider the relation obtained by computing the difference between the two members of the equations that each of the $\underline{u}_i$ verifies, multiplying at the same time both sides by $\delta\underline{u}$ and integrating on Ω. One finds

$$\frac{1}{2}\frac{d}{dt}\int_\Omega(\delta\underline{u})^2\,d\underline{x} = -\nu\int_\Omega(\underset{\sim}{\partial}\,\delta\underline{u})^2\,d\underline{x} + \int_\Omega(-(\underset{\sim}{u}_1\cdot\underset{\sim}{\partial}\,\delta\underline{u} - (\delta\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{u}_2))\cdot\delta\underline{u}\,d\underline{x} \tag{3.1.4}$$

because the pressure term disappears by integration by parts (by using the zero divergence property of the velocity field) and in the same way the first term in the third integral of (3.1.4) also disappears (because it can be written $\int_\Omega -\underset{\sim}{u}_1\cdot\underset{\sim}{\partial}\,(\delta\underline{u})^2\,d\underline{x}$ and it is, therefore, zero for the same reason).

Hence setting $D = \int_\Omega(\delta\underline{u})^2\,d\underline{x}$ one has, for $0 \le t \le t_0$

$$\frac{1}{2}\dot{D} \le D \max_{\underline{x},\,0\le t\le t_0}|\underset{\sim}{\partial}\underline{u}_2(\underline{x},t)| \stackrel{def}{=} D\,M \tag{3.1.5}$$

where M is the indicated maximum ($M < \infty$ because $\underline{u}_2 \in C^1([0,t_0)\times\Omega)$): the (3.1.5) implies $D(t) \le D(0)e^{2Mt}$; hence $D(0) = 0$ yields $D(t) = 0$ for each $0 \le t \le t_0$; *i.e.* one has uniqueness of the solutions of the (3.1.2) verifying (3.1.3).

Note that in (3.1.5) we did not take advantage of the non positivity of the term proportional to the viscosity: hence the just seen uniqueness property remains valid in the case of the Euler equation. However in the remaining discussion we shall make explicitly use of the hypothesis $\nu \ne 0$, *c.f.r.* problem [3.1.6].

We now look at the more interesting question of existence. We shall write (3.1.2) in the spectral form (2.2.10):

$$\dot{\underline{\gamma}}_{\underline{k}} = -\nu\,\underline{k}^2\underline{\gamma}_{\underline{k}} - i\sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \underset{\sim}{\gamma}_{\underline{k}_1}\cdot \underset{\sim}{k}_2\,\Pi_{\underline{k}}\underline{\gamma}_{\underline{k}_2} + \underline{g}_{\underline{k}} \tag{3.1.6}$$

where $\Pi_{\underline{k}}\underline{\hat{g}}_{\underline{k}}$ of (2.2.10) is called here $\underline{g}_{\underline{k}}$. Let us rewrite it as

$$\begin{aligned}\gamma_{\underline{k}}(t) =& \underline{\gamma}^0_{\underline{k}}e^{-\nu\underline{k}^2 t} + \int_0^t e^{-\nu\,\underline{k}^2(t-\tau)}\Big(\underline{g}_{\underline{k}} - \\ &- i\sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \underset{\sim}{\gamma}_{\underline{k}_1}(\tau)\cdot \underset{\sim}{k}_2\,\Pi_{\underline{k}}\underline{\gamma}_{\underline{k}_2}(\tau)\Big)\,d\tau\end{aligned} \tag{3.1.7}$$

We shall set $\overline{\underline{\gamma}}^0_{\underline{k}}(\tau) \equiv \underline{\gamma}^0_{\underline{k}}e^{-\nu\underline{k}^2\tau} + \underline{g}_{\underline{k}}(1-e^{-\nu\,\underline{k}^2\tau})/\nu\,\underline{k}^2$ and we get

$$|\overline{\underline{\gamma}}^0_{\underline{k}}(\tau)| < \overline{V}_0 e^{-\xi_0|\underline{k}|/k_0}, \qquad \overline{V}_0 \equiv V + G\,(\nu k_0)^{-2} \tag{3.1.8}$$

where $k_0 = 2\pi/L$, *c.f.r.* (3.1.1).

Imagine solving (3.1.7) recursively, setting $\underline{\gamma}^0_{\underline{k}}(t) \equiv \overline{\gamma}^0_{\underline{k}}(t)$ and

$$\underline{\gamma}^{n+1}_{\underline{k}}(t) = \overline{\underline{\gamma}}^0_{\underline{k}}(t) - i\int_0^t e^{-\nu\,\underline{k}^2(t-\tau)}\sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \underset{\sim}{\gamma}^n_{\underline{k}_1}(\tau)\cdot \underset{\sim}{k}_2\,\Pi_{\underline{k}}\underline{\gamma}^n_{\underline{k}_2}(\tau)\,d\tau \tag{3.1.9}$$

for $n \ge 0$. Then by "iterating" (3.1.9) once we get

$$\begin{aligned}&\underline{\gamma}^{n+1}_{\underline{k}}(t) = \overline{\underline{\gamma}}^0_{\underline{k}}(t) + \int_0^t e^{-\nu\,\underline{k}^2(t-\tau)}d\tau\,\Big(-i\sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \overline{\underset{\sim}{\gamma}}^0_{\underline{k}_1}(\tau)\cdot \underset{\sim}{k}_2\,\Pi_{\underline{k}}\,\overline{\underline{\gamma}}^0_{\underline{k}_2}(\tau)\Big) + \\ &-i\int_0^t e^{-\nu\,\underline{k}^2(t-\tau)}d\tau\sum_{\underline{k}_1+\underline{k}_2=\underline{k}}\int_0^\tau d\tau'\,\Big(-i\sum_{\underline{k}'_1+\underline{k}'_2=\underline{k}_1} e^{-\nu\,\underline{k}_1^2(\tau-\tau')} \\ &\underset{\sim}{\gamma}^{n-1}_{\underline{k}'_1}(\tau')\cdot \underset{\sim}{k}'_2\,\Pi_{\underline{k}_1}\underline{\gamma}^{(n-1)}_{\underline{k}'_2}(\tau')\Big)\cdot \underline{k}_2\Pi_{\underline{k}}\,\overline{\underline{\gamma}}^{(0)}_{\underline{k}_2}(\tau) + \\ &-i\int_0^t e^{-\nu\,\underline{k}^2(t-\tau)}d\tau\sum_{\underline{k}_1+\underline{k}_2=\underline{k}} -i\int_0^\tau e^{-\nu\,\underline{k}_2^2(\tau-\tau')}\sum_{\underline{k}'_1+\underline{k}'_2=\underline{k}_2}\end{aligned}$$

$$\begin{aligned}&\overline{\underset{\sim}{\gamma}}^{(0)}_{\underline{k}_1}(\tau')\cdot\underset{\sim}{\underline{k}}_2\Pi_{\underline{k}}\Big(\underset{\sim}{\gamma}^{n-1}_{\underline{k}'_1}(\tau')\cdot\underset{\sim}{\underline{k}}'_2\,\Pi_{\underline{k}_2}\overline{\underset{\sim}{\gamma}}^{(n-1)}_{\underline{k}'_2}(\tau')\Big)d\tau' \qquad (3.1.10)\\ &-i\int_0^t e^{-\nu\,\underline{k}^2(t-\tau)}\,d\tau\sum_{\underline{k}_1+\underline{k}_2=\underline{k}}\\ &\Big(-i\sum_{\underline{k}'_1+\underline{k}'_2=\underline{k}_1}\int_0^\tau e^{-\nu\,\underline{k}_1^2(\tau-\tau_2)}\big(\underset{\sim}{\gamma}^{(n-1)}_{\underline{k}'_1}(\tau_2)\cdot\underset{\sim}{\underline{k}}'_2\,\Pi_{\underline{k}_1}\,\underline{\gamma}^{(n-1)}_{\underline{k}'_2}(\tau_2)\big)d\tau_2\Big)\cdot\underline{k}_2\\ &\Pi_{\underline{k}}\Big(-i\sum_{\underline{k}''_1+\underline{k}''_2=\underline{k}}\int_0^\tau e^{-\nu\,\underline{k}_2^2(\tau-\tau_3)}\underset{\sim}{\gamma}^{(n-1)}_{\underline{k}''_1}(\tau_3)\cdot\underset{\sim}{\underline{k}}''_2\,\Pi_{\underline{k}_2}\,\underline{\gamma}^{(n-1)}_{\underline{k}''_2}(\tau_3)d\tau_3\Big)\end{aligned}$$

for $n \ge 1$.

The (3.1.10) can be further iterated: *clearly one needs a better notation* because the (3.1.10) is a straightforward consequence of (3.1.9) and yet it looks very cumbersome and promises to become even more so, unpractically so, upon further iterations.

Therefore we discuss how to find a simpler representation for $\underline{\gamma}^n_{\underline{k}}(t)$: it turns out that it can be represented graphically as

$$\underline{\gamma}^n_{\underline{k}} = \overline{\underline{\gamma}}^0_{\underline{k}}(t) + \sum_{1\le m<2^n}\ \sum_{\Theta\in\,\text{decorated m}-\text{trees}} \text{Val}(\Theta) \qquad (3.1.11)$$

where Θ is a "tree" with m internal vertices ("decorated m-tree"), and Val(Θ) is its "*value*" (that will be defined below); furthermore the tree has a root and every branching happens by doubling, *i.e.* into every internal vertex two branches of the tree enter and one exits.

To understand the notation in detail define a *m–tree* as a connected set of $2m+1$ oriented lines, that we call "branches", with no cycles. Let us denote with (v, v') an oriented line drawn on a plane and going from the point v to the point v'.

(a) Given $2m + 1$ oriented lines of unit length and fixed a point r on the plane, we copy on the plane a first oriented line (v_0, r) that ends in r. We shall say that r is the *root* of the tree.

(b) Then we copy on the plane two other oriented lines that end in v_0, *i.e.* at the starting point of the already drawn line. Let (v_1, v_0) and (v_2, v_0) be these two lines.

(c) Continue the construction by attaching to some of the initial vertices of the last drawn lines pairs of oriented lines with the end vertex in common. Until (after m steps) a figure Θ is obtained that we shall call *tree* with root r and $2m + 1$ branches, or *m–tree* (not decorated).

(d) We shall call *internal vertices* or *nodes* the vertices into which two lines enter (and one exits, necessarily). While the vertex r will be called root, the vertices out of which only a line exits and none enters will be called *final vertices* or *external* and their collection will be denoted by V_f.

(e) Here we shall consider, therefore, only trees with three branches per node, one exiting and two entering: hence the number of nodes is m and it is one unit lower than number of external vertices (root excluded because it *will not be considered a vertex of the tree*). The trees are thought of with the branches entering every node marked, and distinguished, by a label $\delta = 0, 1$; furthermore trees that can be overlapped (labels included) by means of a sequence of rotations of the lines entering the nodes will be considered identical.

In terms of this definition and adding to the trees a certain number of other labels or *decorations* that shall be described in following explanations on how to read the figures, it is possible to reinterpret the (3.1.9) so that it is represented by the following graph

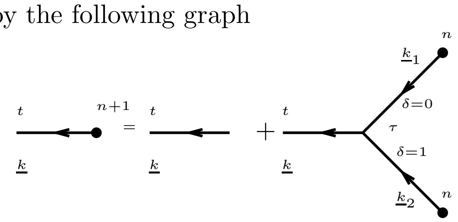


Fig (3.1.1) The representation of (3.1.9).

We read this Fig. (3.1.1) as follows: the l.h.s. represents the l.h.s. of (3.1.9): hence it carries the decorations $t, \underline{k}, n+1, t$ needed to identify it (*i.e.* it is a notation "alternative" to $\underline{\gamma}^{n+1}_{\underline{k}}(t)$). The first graph in the r.h.s. (it is a 0–tree) represents $\overline{\gamma}^{0}_{\underline{k}}(t)$, *i.e.* the first term of the r.h.s. of (3.1.9). Note that it does not carry the label $\bullet$ on the final vertex: a reminder that this branch represents a "known" term.

The second term is read by interpreting the line with labels $\delta = 0$ and $\underline{k}_1$, ending in the node carrying the label τ and starting in a final vertex with label $\bullet^n$, as representing (see (3.1.9)) $\underline{\gamma}^n_{\underline{k}_1}(\tau) \cdot \underline{k}_2$; the other line (with $\delta = 1$) instead is interpreted as $\Pi_{\underline{k}}\, \underline{\gamma}^n_{\underline{k}_2}(\tau)$, where $\underline{k} \equiv \underline{k}_1 + \underline{k}_2$ is the label of the line exiting from the node into which the two lines merge: and this node must be interpreted as the operation $-i \int_0^t d\tau\, e^{-\nu\, \underline{k}^2 (t-\tau)} \cdot$ performed in (3.1.9).

Hence the node represents an integration operation and the label δ distinguishes the two factors with $\underline{\gamma}^n$ in (3.1.9): this is a label made necessary by their not symmetric role. We see that the decorations and the form of the tree identify uniquely the operations to perform: thus they are just an alternative notation to (3.1.9).

Likewise the equation (3.1.10), obtained replacing the functions $\underline{\gamma}^n_{\underline{k}}(\tau)$ in (3.1.9) with the expression that is provided by (3.1.9) itself (with n replaced by $n-1$), is susceptible of a graphical interpretation. An attentive examination of (3.1.10) indeed gives us the following representation, see Fig. (3.1.2), consistent with the preceding one, and *manifestly much simpler than* (3.1.10).

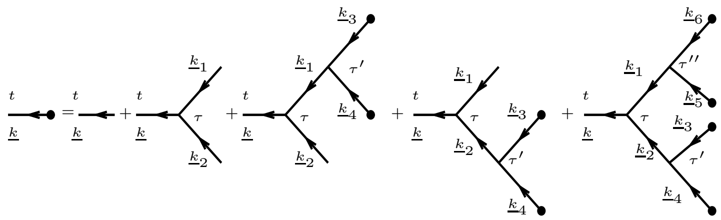


Fig. (3.1.2) The representation of (3.1.10).

where the vertices marked by $\bullet$ should carry the label $n+1$ in the l.h.s. and $n-1$ in the r.h.s., and the lines entering the nodes should carry the label $\delta = 0, 1$ (*not marked in the figure* for simplicity).

We must remark that Fig. (3.1.2) is obtained from Fig. (3.1.1) by simply replacing the final lines (*i.e.* lines with initial vertex in V_f and, therefore, without entering lines) carrying the label $\bullet^n$ by one of the two trees of the r.h.s. of Fig. (3.1.1).

Hence Fig. (3.1.2) can be further "developed" by iterating the construction. At every iteration the trees that represent $\gamma^n_{\underline{k}}(t)$ will have final vertices without $\bullet$ and final vertices still carrying this label that, however, at the k–th step will be $\bullet^{n-k}$. Hence the procedure will stop when $k = n$, provided the vertices with label $\bullet^0$ will be simply drawn *without* this label (as it can be given no other meaning). The graphs that are obtained in this way are trees in the sense defined by (a)%(e) suitably decorated with labels (and with the final branches deprived of the labels $\bullet$).

It is clear that the result of the iteration of the (3.1.9), *i.e.* of the iteration developed until the $\gamma^{(j)}_{\underline{k}}(\tau)$ with $j > 0$ disappear, can be written in the form (3.1.11), as we see already in the case of (3.1.10). For this purpose we shall formally describe the further "decorations" that it is necessary to attach to the trees. The natural decorations, suggested from the above interpretation of (3.1.9) and (3.1.10), are

(1) A label $\underline{k}_\rho$ is attached to every branch ρ of Θ: such vector labels will have to be subjected to the constraint that the sum of the $\underline{k}_\rho$ of the (two) branches ρ entering a node be equal to the label $\underline{k}_\rho$ of the exiting branch. Hence the labels $\underline{k}_\rho$ of the final branches (*i.e.* of those that begin in vertices $v \in V_f$) determine all the labels $\underline{k}_\rho$ of the other branches. We shall set, however, the restriction that the label $\underline{k}_\rho$ of the branch that ends into the root be $\underline{k}$, if we are computing $\gamma_{\underline{k}}(t)$).

We shall call, in analogy with the "*Feynman graphs*" of field theory, the vector $\underline{k}_\rho$ the *momentum flowing* in the line ρ and the momentum of the line that ends in the root will be the *total momentum* of the tree. Note that $\underline{k} = \sum^*_\rho \underline{k}_\rho$ where the sum is restricted only to the final branches (*i.e.* with initial vertex in V_f).

(2) With every node v we associate a time $\tau_v \in [0, t]$ so that the function $v \to \tau_v$ increases as a function of v (consistently with order on the tree,

from the final vertices towards the root). With the root we associate the time t larger than all τ_v.
(3) Finally we recall that, by construction, every branch ρ that precedes a node v carries a label $\delta_\rho = 0, 1$.

If ρ_v is the branch that has v as initial point and v'_ρ, v_ρ denote the final and initial points of the branch ρ, consider the following quantity that we call the *value* Val(Θ) of the decorated tree (Θ)

$$(-i)^m \int (\prod_{v\,\text{nodes}} d\tau_v)(\prod_{\rho\,\text{int}} e^{-\nu\,\underline{k}_\rho^2(\tau_{v'_\rho}-\tau_{v_\rho})}\Pi_{\underline{k}_\rho})(\prod_{v\in V_f} \overline{\underline{\gamma}}^0_{\underline{k}_{\rho_v}}(\tau_v))(\prod_{\rho,\delta_\rho=1} \underline{k}_\rho) \tag{3.1.12}$$

where the first product is over the non trivial vertices, *i.e.* the nodes, of the tree (in number of m); the second product is over the $m-1$ internal branches (*i.e.* over the branches that do not start at a final vertex $v \in V_f$); the third product on $v \in V_f$ concerns the $m+1$ final vertices and the last product is over the m branches with label $\delta_\rho = 1$.
The domain of the integrals on the τ_v is described by the conditions in (2).
The labels of the vectors $\underline{k}$ and $\overline{\underline{\gamma}}^0_{\underline{k}}$ and of the tensors $\Pi_{\underline{k}_\rho}$ that project on the planes orthogonal to $\underline{k}_\rho$ must be contracted between each other in a suitable way (that can be inferred by reading (3.1.9), (3.1.10) and that here it is not interesting to make explicit): *then the sum (3.1.11) of the values (given by (3.1.12)) of the decorated m–trees Θ with $m < 2^n$ will give us the value of $\underline{\gamma}^{(n)}_{\underline{k}}(t)$. The sum over all the decorated trees will be exactly the limit for $n \to \infty$ of $\underline{\gamma}^{(n)}(\tau)$.*
In other words (3.1.12) gives a formula, as a development in a series, for the solution of the NS equation, obviously modulo problems of convergence and provided no double counting errors are made. The latter are simply avoided if we impose to consider only the contribution of decorated trees which are "different", considering as equal two decorated trees that can be overlapped by permuting the branches that enter the same node (with the decorations rigidly attached).

Wishing to perform a bound of the convergence of the series for $t \le t_0$, with $t_0 > 0$, we remark that every factor $\overline{\underline{\gamma}}^0_{\underline{k}}$ can be bounded by $\overline{V}_0 e^{-\xi_0|\underline{k}|/k_0}$ using (3.1.8), with $\overline{V}_0 = V + (\nu k_0)^{-2}\, G$.
Writing $e^{-\xi_0|\underline{k}|/k_0} \equiv e^{-\xi_0|\underline{k}|/2k_0} e^{-\xi_0|\underline{k}|/2k_0}$ and recalling that the sum of the $\underline{k}_\rho$ is conserved at every vertex, we see that $\underline{k} = \sum_{v\in V_f} \underline{k}_{\rho_v}$ and hence the contribution to the sum in (3.1.12) due to the trees with m nodes (hence with $m+1$ final vertices and $m-1$ internal branches) is bounded by

$$\begin{aligned}&\overline{V}_0^{n+1} e^{-\xi_0|\underline{k}|/2k_0} \sum_{\underline{k}_\rho} \int (\prod_{v\in\text{nodes}} d\tau_v)\cdot\\ &\cdot(\prod_{v\ni V_f} e^{-\nu\,\underline{k}^2_{\rho_v}(\tau_{v'}-\tau_v)})(\prod_{v\in V_f} e^{-\xi_0|\underline{k}_v|/2k_0})(\prod_{\rho,\,\delta_\rho=1} |\underline{k}_\rho|)\end{aligned} \tag{3.1.13}$$

where the sum only runs over the $\underline{k}_{\rho_v}$'s with ρ_v being a final branch, (the other $\underline{k}_\rho$'s are fixed by the rule of conservation at the nodes). Recall here that the final vertices $v\in V_f$ are *not* nodes (only the internal vertices are nodes).

The integrals on τ_v can be bounded, enlarging to $[0,t]$ their domains of integration and for *all* trees, by

$$\prod_\rho \frac{1-e^{-\nu\,\underline{k}_\rho^2 t}}{\nu\,\underline{k}_\rho^2}\le\prod_\rho t^\varepsilon\,(\nu\,\underline{k}_\rho^2)^{-(1-\varepsilon)} \tag{3.1.14}$$

for $\varepsilon\in[0,1]$ arbitrary;[2] the product in (3.1.14) runs on the branches ρ which are *not final*, *c.f.r.* above.

Hence by selecting $\varepsilon=1/2$ we see that (3.1.12), relative to a tree with $2m+1$ branches and m nodes, hence, $m+1$ final vertices, is bounded above by

$$e^{-\xi_0|\underline{k}|/2k_0}2^{4n}(t\nu^{-1})^{n/2}\,\overline{V}_0^{m+1}\Big(\sum_{\underline{k}'}\frac{|\underline{k}'|}{k_0}\,e^{-\xi_0|\underline{k}'|/2k_0}\Big)^{m+1} \tag{3.1.15}$$

because the factors $|\underline{k}_\rho|$ corresponding to the *internal* branches with label $\delta_\rho=1$ (in (3.1.13)) are compensated by the corresponding $|\underline{k}_\rho|^{-1}$, (that are generated by (3.1.14) with $\varepsilon=1/2$). The factors $|\underline{k}_\rho|$ relative to the branches with label $\delta_\rho=1$ but that exit from final vertices instead cannot be compensated in this way (as the factors in (3.1.14) are simply not there, see (3.1.12)) hence we left them, just writing them as $k_0\,|\underline{k}_\rho|/k_0$. We put them together with the bound on the factors $\underline{\gamma}^0_{\underline{k}_\rho}$ associated with the $m+1$ final branches (*i.e.* $\overline{V}_0e^{-\xi_0|\underline{k}_\rho|/2k_0}$) generating a certain number $m'<m+1$ of factors that are *some* of the factors in the last term in (3.1.15); each of the other $m+1-m'$ could be bounded by the same sum deprived of the term $|\underline{k}'|/k_0$, which we instead prefer to leave, being ≥1, to get a simpler bound.

The remaining $|\underline{k}_\rho|^{-1}$, corresponding to the branches with label $\delta_\rho=0$ are trivially bounded by k_0^{-1}. And the factor 2^{4m} bounds the number of trees with $2m+1$ branches.[3]

We find, therefore

$$|\underline{\gamma}_{\underline{k}}|\le e^{-\xi_0|\underline{k}|/2k_0}\,\overline{V}_0\sum_{m=0}^\infty\big(\overline{V}_0t^{1/2}\nu^{-1/2}\big)^m\xi_0^{-(d+1)(m+1)}B^{m+1}2^{4m} \tag{3.1.16}$$

[2] Because $(1-e^{-ab})/a$ is bounded by both a^{-1} and by b, for all $a,b>0$.

[3] We see immediately, in fact, that the number of trees with m branches is not larger than the number of paths on the lattice of the positive integers with $2(m-1)$ steps of size 1 (a number bounded by $2^{2(m-1)}$): given a tree think of walking on its branches starting from the root and always choosing the left branch if possible and otherwise coming back until returning to the root (and, by construction, without ever running more than twice on the same branch); we associate with each branch a step forward on the integers lattice, or a step backwards when we could not proceed (choosing the branch on the left) and had to come back. In our case a m–tree contains $2m+1$ branches.

if $\xi_0^{-d-1}B$ is a bound of $\sum_{\underline{k}} \frac{|\underline{k}|}{k_0} e^{-\xi_0|\underline{k}|/2k_0}$ (recall that we are assuming $\xi_0 \le 1$, for simplicity, *c.f.r.* the comment to (3.1.1)).

Hence $t \le \overline{t}_0 = \nu\,(16B\,\overline{V}_0\xi_0^{-d-1})^{-2}$ is the condition of convergence of the series in (3.1.16). The theorem follows by selecting $t_0 = \overline{t}_0/4$, hence such that the sum on n is bounded by 2, and $V_0 = B\,\overline{V}_0\xi_0^{-d-1}$.

At the end one can redefine the numerical constants so that the bound (3.1.3) can be expressed in terms of a single numerical constant B_0.

Remarks: If we define the operator $(-\Delta)^\alpha$, with α real , as the operator that multiplies by $|\underline{k}|^{2\alpha}$ the Fourier transform harmonic of mode $\underline{k}$ of a function in $L_2(\Omega)$, then we note that if the friction term $\nu\,D\underline{u}$ in (3.1.2) is replaced by $-\nu(-\Delta)^\alpha$ with $\alpha > 0$ the proposition remains valid, with some obvious modifications. The main point of the proof is indeed the bound (3.1.14) that can be replaced by

$$\frac{|e^{-\nu|\underline{k}|^{2\alpha}t} - 1|}{\nu|\underline{k}|^{2\alpha}} \le t^\varepsilon(\nu|\underline{k}|^{2\alpha})^{-(1-\varepsilon)} \tag{3.1.17}$$

and the argument can be adapted *provided one can choose* $\varepsilon > 0$ *such that* $2(1-\varepsilon)\alpha > 1$, *i.e. if* $\alpha > 1/2$. This restriction is indeed necessary because only in this way we can to eliminate from the bounds the factors $|\underline{k}_\rho|$ due to the final branches with label $\delta_\rho = 1$, that cannot be otherwise dominated.

Hence in a certain sense the friction term in the NS equations is "*larger than needed*", at least for the purpose of being able to guarantee that analytic initial data generate an analytically regular motion, at least for a short enough time: this is the case not only for the "normal viscosity" (friction term $\nu\Delta\underline{u}$) but also for the *ipoviscous NS* with friction term $-\nu|\Delta|^\alpha$ provided $\alpha > \frac{1}{2}$ and (of course) for the *hyperviscous NS* with $\alpha > 1$.

Concerning the incompressible Euler equation, where $\nu = 0$, the method now illustrated does not lead to conclusions, since we cannot show convergence of the sreis. A local theory of the Euler equation is nevertheless possible (and classical) and we shall illustrate it in the following problems.

Problems. *Classical local theory for the Euler and Navier–Stokes equations with periodic conditions.*

Below we suppose that Ω is a cube with side L and with opposite sides identified: this is done for the sake of simplicity as most of the statements hold also in the case of a boundary condition of the type $\underline{u}\cdot\underline{n} = 0$ at a point where the external normal is $\underline{n}$.

[3.1.1]: (*Cauchy's estimate*) If $\underline{x} \to f(\underline{x})$ is analytic and periodic on the torus $[0,L]^d$, then there is a $\xi > 0$ and a holomorphic function F of d complex variables that extends the function f to a vicinity of $[0,L]^d$ in C^d consisting in the points $\underline{x}$ such that $|\mathrm{Im}\,x_i| < \xi$. Furthermore if $|F| \le \Phi$ for $|\mathrm{Im}\,x_i| < \xi$ the Fourier transform of f verifies $|f_{\underline{k}}| \le \Phi\,e^{-\xi|\underline{k}|}$ where $|\underline{k}|$ is defined as $|\underline{k}| = \sum_i |k_i|$. Find the connection between this result and the theory of the Laurent series. (*Idea:* Study first the case $d = 1$. Write the transform $f_{\underline{k}}$ as an integral on the torus, applying the definition, and "deform the integration path" to the lines $\mathrm{Im}\,x_j = \pm\xi_0$, with $\xi_0 < \xi$, depending on the sign of k_j (*i.e.* use the Cauchy theorem on holomorphic functions). The connection with the theory of the Laurent series is made by thinking of the formula giving the Fourier transform as an integral over the variables $\zeta_j = e^{ix_j}$).

[3.1.2]: (*generalized energy identity*) Show that if $\underline{u}$ is a $C^\infty(\Omega)$ velocity field then

$$\sum_{|\underline{m}|=m}\int_\Omega \vec\partial^{\underline{m}}\underline{u}\cdot\vec\partial^{\underline{m}}(\underset{\sim}{u}\cdot\underset{\sim}{\partial}\underline{u})\,d\underline{x} = \sum_{|\underline{m}|=m}\sum_{\substack{\underline{a}\le\underline{m}\\ |\underline{a}|>0}}\int_\Omega(\vec\partial^{\underline{m}}\underline{u})\cdot(\vec\partial^{\underline{a}}\underset{\sim}{u})\cdot(\underset{\sim}{\partial}\vec\partial^{\underline{m}-\underline{a}}\underline{u})\,d\underline{x}$$

i.e. the term $\underline{a}=\underline{0}$ is missing because of a *cancellation*. (*Idea:* The term with $\underline{a}=\underline{0}$ is a sum of terms like $\int \underline{w}\cdot(\underset{\sim}{u}\cdot\underset{\sim}{\partial})\,\underline{w}\,d\underline{x} \equiv \frac12\int\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{w}^2\,d\underline{x}$ with $\underline{w}$ suitable and this quantity vanishes (integrate by parts, as usual)).

[3.1.3]: Let $\underline{u}\in C^\infty(\Omega)$ and $|\underline{a}|<m-d/2$ (with $|\underline{a}|=\sum|a_i|$ for $\underline{a}=(a_1,\ldots,a_d)$) then, by [2.2.22], $|\underset{\sim}{\partial}^{\underline{a}}\underset{\sim}{u}(\underline{x})|\le\Gamma L^{-|\underline{a}|}||\underline{u}||_{W^m(\Omega)}$. Deduce that, if $m>d+1$

$$\Big|L^{2m-d+1}\sum_{|\underline{m}|=m}\int_\Omega \underset{\sim}{\partial}^{\underline{m}}\underline{u}\cdot\underset{\sim}{\partial}^{\underline{m}}(\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{u})\,d\underline{x}\Big|\le\Gamma_1||\underline{u}||^3_{W_m(\Omega)},\qquad m>d+1$$

with Γ_1 a suitable constant. (*Idea:* Apply [3.1.1] and the Schwartz inequality to each addend, thereby reducing to estimating $\mathcal{N}=||\vec\partial^{\underline{a}}\underset{\sim}{u}\cdot\underset{\sim}{\partial}\vec\partial^{\underline{m}-\underline{a}}\underline{u}||_{L_2(\Omega)}$ with $|\underline{a}|\ge1$. Note that if $|\underline{a}|<m-d/2$ then the estimate of [2.2.22] implies: $L^{|\underline{a}|}|\vec\partial^{\underline{a}}\underline{u}(\underline{x})|<\Gamma_2||\underline{u}||_{W_m(\Omega)}$, allowing us to estimate from above the first of the two factors in the $\mathcal{N}$; if instead $|\underline{a}|\ge m-d/2$ then $|\underline{m}-\underline{a}|+1\equiv m-|\underline{a}|+1\le\frac{d}{2}+1$ hence if $m>d+1\equiv(\frac{d}{2}+1)+\frac{d}{2}$ (hence $|\underline{m}-\underline{a}|+1<m-\frac{d}{2}$) we find, always because of [2.2.22], that

$$L^{(m-|\underline{a}|)+1)}|\underset{\sim}{\partial}\vec\partial^{\underline{m}-\underline{a}}\underline{u}(\underline{x})|\le\Gamma_3||\underline{u}||_{W_m(\Omega)}$$

since $|\underline{a}|\ge1$: which allows us to get an upper bound on the second factor in $\mathcal{N}$).

[3.1.4]: (*bounds uniform in the regularization parameter for Euler flows*) Show that if $\underline{u}^N$ is a solution of the regularized Euler equations, with vanishing or just conservative external volume force, obtained by an "ultraviolet cut–off" (in the sense of §2.2) at $|\underline{k}|<N$ one has, if $m>d+1$

$$\frac12\frac{d}{dt}||\underline{u}^N||^2_{W_m(\Omega)}\le G_m||\underline{u}^N||^3_{W_m(\Omega)}$$

with G_m independent from N. (*Idea:* This follows from [3.1.3] by differentiating $\underline{m}$ times the Euler equation and by multiplying both sides by $\partial^{\underline{m}}\underline{u}$, summing over $\underline{m}$ with $|\underline{m}|\le m$ and integrating over $\underline{x}$: one must remark that the truncation operations do not interfere with the derivation of the conclusions of [3.1.3]. In the sense that the identity in [3.1.2] remains valid if we replace $\underline{u}$ with $\underline{u}^N\equiv P_N\underline{u}$ where P_N is the orthogonal projection, in $L_2(\Omega)$, on the subspace generated by the plane waves of momentum $|\underline{k}|<N$ and if, furthermore, $(\vec\partial^{\underline{a}}\underset{\sim}{u})\cdot(\underset{\sim}{\partial}\vec\partial^{\underline{m}-\underline{a}}\underline{u})$ is replaced by $P_N\big((\vec\partial^{\underline{a}}\underset{\sim}{u}^N)\cdot(\underset{\sim}{\partial}\vec\partial^{\underline{m}-\underline{a}}\underline{u}^N)\big)$: this is immediately checked through the properties of the scalar product in L_2 and of the Fourier transform).

[3.1.5]: (*uniform smoothness of regularized Euler flows*) Show that if $\underline{u}^0\in C^\infty$ then given $m>d+1$ it is, setting $W_m=W_m(\Omega)$

$$||\underline{u}^N(t)||_{W_m}\le\frac{||\underline{u}^0||_{W_m}}{1-G_m||\underline{u}^0||_{W_m}t},\quad 0\le t<T_m$$

with $T_m = (G_m||\underline{u}^0||_{W_m})^{-1}$ and if $\underline{u}^N$ is the solution of the regularized equation with cut–off at $|\underline{k}| < N$ and initial datum $\underline{u}^0$ (truncated with the same cut–off).

[3.1.6]: (*uniqueness for smooth Euler flows*) (Graffi) Show that the Euler equation does not admit more than one solution continuous in $\underline{x}, t$, with first derivatives with respect to $\underline{x}$ and t continuous and bounded in the $\underline{x}$ variables and in every finite time interval $[0,T]$, and verifying the same initial datum. (*Idea:* If $\underline{u}^1$ and $\underline{u}^2$ are two solutions $\underline{\delta} \equiv \underline{u}_1 - \underline{u}_2$ and $D(t)$ is the integral of $\underline{\delta}^2$ on Ω it is

$$\frac{d}{dt}D(t) \equiv -\int_\Omega \underline{\delta}\cdot(\underline{\delta}\cdot\underset{\sim}{\partial}\underline{u}_1)\,d\underline{x} - \int_\Omega \underline{\delta}\cdot(\underline{u}_2\cdot\underset{\sim}{\partial})\,\underline{\delta}\,d\underline{x} \le U\int_\Omega \underline{\delta}^2\,d\underline{x} = U\,D(t)$$

because the second term in the intermediate step vanishes because of the usual reasons, if $U = \max|\underset{\sim}{\partial}\underline{u}_1|$ and if $D(t)$ is defined by the first identity and therefore $D(t) \le D(0)e^{2Ut}$, $t \in [0,T]$, so that *if* $D(0) = 0$ it is $D(t) \equiv 0$).

[3.1.7]: (*local existence and smoothness for Euler flows*) Show, in the context of [3.1.5], that if $\underline{u}^0$ is in $C^\infty(\Omega)$ then $\lim_{N\to\infty}\underline{u}^N(t)$, $t \in [0,T]$ exists and it is a C^m solution of Euler equation for $T < T_{m+1+d/2}$, with T_m defined in [3.1.5], for each $m > 1+d/2$. (*Idea:* Consider a subsequence $\underline{u}^{N_i}$ uniformly convergent together with its first derivatives. Note there is a uniform estimate of the derivatives up to order $k < m - d/2$ *i.e.*, since $m - d/2 > 1$, certainly for $k = 0, 1$; hence the sequences of both $\underline{u}^N$ and $\underset{\sim}{\partial}\underline{u}^N$ are equibounded and equicontinuous so that the Ascoli–Arzelà theorem applies. Note also that the limit of the subsequence is a solution of the Euler equation bounded, and with bounded derivatives, in $\underline{x}$ and t (continuity in t follows from the uniform continuity of the $\underline{u}^N$ in t due to the estimate of the time derivative that one sees from the fact that the $\underline{u}^N$ verify the truncated equation). By [3.1.6] there is only one such solution: hence all subsequences must have the same limit.)

[3.1.8]: (*local continuity with respect to initial data in Euler flows*) Let $\underline{u}_0, \underline{u}'_0 \in C^\infty(\Omega)$ be two initial data for solutions considered in [3.1.7] and suppose that $||\underline{u}_0||_{W_{h+1}}$, $||\underline{u}'_0||_{W_{h+1}} < R$, for a fixed $h > d+1$; then $||\underline{u}||_{W_h} \le 2||\underline{u}_0||_{W_h}$, $\quad ||\underline{u}'||_{W_h} \le 2||\underline{u}'_0||_{W_h}$ and

$$||\underline{u} - \underline{u}'||_{W_h} \le \lambda_{h+1}(R)||\underline{u}_0 - \underline{u}'_0||_{W_h}, \quad 0 \le t \le \frac{1}{2G_{h+1}R} \equiv \tau_h(R)$$

with G_m introduced in [3.1.4] and $\lambda_h(R)$ suitable. (*Idea:* Let $\underline{\delta} = \underline{u} - \underline{u}'$ and note that if $|\underline{m}| \le h$: hence

$$\begin{aligned}&\frac{1}{2}\frac{d}{dt}\int_\Omega (\partial^{\underline{m}}\underline{\delta})^2\,d\underline{x} = -\int_\Omega \partial^{\underline{m}}\underline{\delta}\,\partial^{\underline{m}}(\underset{\sim}{\delta}\cdot\underset{\sim}{\partial}\,\underline{u} + \underset{\sim}{u}'\cdot\underset{\sim}{\partial}\underline{\delta})\,d\underline{x}\\ &\le C_{|\underline{m}|}R||\underline{\delta}||^2_{W_h}, \qquad h > d+1\end{aligned}$$

because if the cancellation in [3.1.2], *i.e.* because $\partial^{|\underline{m}|+1}\underline{\delta}$ never appears, and because of the argument in [3.1.3]: then $\lambda_{h+1}(R) = \exp(R\sum_{|\underline{m}|\le h} L^{2m-d}C_{\underline{m}}\tau_h(R))$, independently of R.)

[3.1.9]: (*local error estimate for regularized approximations Euler flows*) Likewise show that the solution in [3.1.7] enjoys the property

$$||\underline{u} - \underline{u}^N||_{W_h} \le \lambda_{h+1}||\underline{u}_0 - \underline{u}_0^N||_{W_h} + \frac{\tilde{C}_p\lambda_{h+1}}{N^{p-2d}}||\underline{u}_0||^2_{W_{h+p}}$$

with $\tilde{C}_p$ suitable and $p > 2d$. Check that this implies the possibility of a constructive algorithm[4] to approximate the Euler equations solution in the sense of the distance in

[4] By "*constructive*" algorithm we mean an algorithm that produces an approximation of the solution within an *a priori* fixed error ε via a coomputer program that ends in a time that can be *a priori* estimated in terms of ε.

the space W_h valid for times $t \le \tau_{h+p}$, with τ_m defined in [3.1.8] and $p > 2d$. (*Idea:* Write the equations for $\underline{u}$ and $\underline{u}^N$ keeping in mind that $\dot{\underline{u}}^N = -\underset{\sim}{u}^N \cdot \underset{\sim}{\partial}\underline{u}^N + (\underset{\sim}{u}^N \cdot \underset{\sim}{\partial}\underline{u}^N)^{>N}$ where $f^{>N}$ denotes the part of f obtained as sum of the harmonics of f with $|\underline{k}| > N$. Estimate this "ultraviolet part" by using the result of [3.1.4] (subtracting the equations for $\underline{u}$ and $\underline{u}^N$) obtaining an inequality similar to the one in the hint for [3.1.8] with an additional term).

[3.1.10]: (*summary of classical local results for Euler flows*) The results of the above problems can be summarized by saying that, given $h > d+1$ and $\underline{u}_0, \underline{u}'_0 \in C^\infty(\Omega)$, $||\underline{u}_0||_{W_{h+1}}, ||\underline{u}'_0||_{W_{h+1}} \le R$ there is a local solution of the Euler equation, defined for $0 \le t \le \tau_h(R)$ and such that in this time interval

$$\begin{aligned}
&(1) \qquad ||\underline{u}||_{W_h} \le 2R \\
&(2) \qquad ||\underline{u} - \underline{u}'||_{W_h} \le \lambda_{h+1} ||\underline{u}_0 - \underline{u}'_0||_{W_h} \\
&(3) \qquad ||\underline{u} - \underline{u}^N||_{W_h} \le \lambda_{h+1} ||\underline{u}_0 - \underline{u}^N_0||_{W_h} + \frac{C_p \lambda_{h+1}}{N^{p-2d}} ||\underline{u}_0||^2_{W_{h+p}}
\end{aligned}$$

Since in (3) $||\underline{u}_0||^2_{W_{h+p}}$, with $p > 2d$, appears rather than $||\underline{u}_0||^2_{W_h}$ we see that the approximation error on $\underline{u}$ by $\underline{u}^N$ has not been estimated "constructively" on the *whole* time interval $[0, \tau_h(R)]$ along which one can guarantee *a priori* existence of a solution in W_h and bounded in terms of known quantities: *rather it is estimated in the shorter interval* $[0, \tau_{h+p}(R)]$. Note that, on the basis of what said so far, this would not be possible *even* if we knew an *a priori* estimate of the size $||\underline{u}||_{W_k}$, with k arbitrary, in terms of the properties of $\underline{u}_0$ only. It seems that, for a constructive estimate, bounds on $||\underline{u}^N||_{W_k}$ in terms of $\underline{u}_0$ only are also needed. Show that, indeed, if such bounds existed then a constructive algorithm would be possible, valid for all times for which the considered bounds on of $\underline{u}$ and $\underline{u}_0$ hold. The relevance of this comment is that it points out that *although by abstract arguments* we can show existence and even some *a priori* bound on $||\underline{u}||_{W_k}$, in terms of t and $\underline{u}_0$, *nevertheless a constructive estimate* for $\underline{u}^N$ is not known.

The following problems are devoted to deriving a priori estimates on the solution $\underline{u}$ of the Euler equations in $d = 2$ that shows existence, uniqueness and regularity as well as a priori estimates of the size of $||\underline{u}||_{W_h}$, valid at all times and dependent on $\underline{u}_0$ only, provided we accept the sinister axiom of choice (whose use we strongly disrecommend). The constructive part of the theory (as also the rest if the axiom of choice is accepted) is due to Wolibner, Judovic and Kato, c.f.r. [Ka67]. In what follows Ω is chosen $\Omega = [0, L]^2$ with periodic boundary conditions; furthermore, for simplicity, we suppose $\underline{g} = \underline{0}$. Finally we shall consider the Euler equation in an arbitrarily prefixed time interval, $[0, T]$. Given a function $f(\underline{x})$ on Ω or an $f(\underline{x}, t)$ on $\Omega \times [0, T]$ we shall say that it is of of class $C^h(\Omega)$ or $C^{h,k}(\Omega \times [0, T])$ if it has h continuous derivatives in $\underline{x}$ and k continuous derivatives in t; we shall say that it if of class $C^h(\Omega \times [0, T])$ if it has h continuous derivatives in $\underline{x}$ or t. We shall set

$$||f||_h = \sum_{0 \le |\underline{a}| \le h} \max_{\underline{x} \in \Omega} L^{|\underline{a}|}\, |\partial^{\underline{a}}_{\underline{x}} f(\underline{x})|$$

$$||f||_{h,k} = \sum_{\substack{0 \le |\underline{a}| \le h \\ 0 \le \beta \le k}} \max_{(\underline{x}, t) \in \Omega \times [0,T]} L^{|\underline{a}|} T^\beta\, |\partial^{\underline{a}}_{\underline{x}} \partial^\beta_t f(\underline{x}, t)|$$

[3.1.11]: (*Euler flows as vorticity transport in 2–dimensions*) If $\omega(\underline{x}, t) = \operatorname{rot} \underline{u}(\underline{x}, t) \in C^\infty$ then $\omega(\underline{x}, t)$ is a scalar unction with zero average for each t. Show the equivalence between the solutions of the Euler equations in $C^\infty(\Omega\times[0,T])$ in the form of the (3.1.2) with $\nu=0, \underline{g}=\underline{0}$ and the same equations in the form (2.3.3) $\nu,\gamma=0$:

$$\partial_t\omega+\underline{u}\cdot\underline{\partial}\omega=0,\qquad \underline{u}=-\underline{\partial}^\perp\Delta^{-1}\omega$$
$$\int_\Omega \omega(\underline{x},t)\,d\underline{x}=0,\qquad \omega(\underline{x},0)=\omega_0(\underline{x})$$

where $\omega_0=\mathrm{rot}\,\underline{u}_0$. Show that the condition $\int_\Omega \omega d\underline{x}=0$ is necessary in order that $\Delta^{-1}\omega$ be meaningful. (*Idea:* The rotation ω of a C^∞ solution of the Euler equations satisfies the above equations, as noted in §1.7. For the converse the only delicate point is checking existence of the pressure p from the remark that the first equation can be rewritten as $\mathrm{rot}\,(\partial_t\underline{u}+\underset{\sim}{u}\cdot\underset{\sim}{\partial}\underline{u})=0$ as $\underline{\partial}\cdot\underline{u}=\underline{0}$, because the torus Ω is not simply connected. Note that the vanishing of the rotor implies that the circulations of the vector field $\underline{w}\equiv\partial_t\underline{u}+\underset{\sim}{u}\cdot\underset{\sim}{\partial}\underline{u}$ along the torus cycles: $C_1(y)\equiv\{(x,y);\ y=cost\}$ e $C_2(x)=\{(x,y);\ x=cost\}$, defined by $I_1(y)=\oint_{C_1(y)}\underline{w}\cdot d\underline{l}$ and $I_2(x)=\oint_{C_2(x)}\underline{w}\cdot d\underline{l}$, are independent from x and y respectively hence are equal to their averages over these coordinates, *i.e.* $\underline{I}=(I_1,I_2)=L^{-1}\int_\Omega(\partial_t\underline{u}+\underset{\sim}{u}\cdot\underset{\sim}{\partial}\underline{u})\,d\underline{x}$ which we can see to vanish, as a consequence of $\int_\Omega \underline{u}d\underline{x}=\underline{0}$).

[3.1.12]: (*existence of current lines for a smooth voricity field*) Consider the space $\mathcal{M}_0\subset C^\infty$ defined by the

$$(1)\quad \zeta\in C^\infty(\Omega\times[0,T]),\qquad (3)\quad \int_\Omega\zeta(\underline{x},t)\,d\underline{x}=0$$
$$(2)\quad \zeta(\underline{x},0)=\omega_0(\underline{x}),\qquad (4)\quad \max_{(\underline{x},t)\in\Omega\times[0,T]}|\zeta(\underline{x},t)|\le||\omega_0||_0$$

Show that the vector field $\underline{v}^\zeta(\underline{x},t)=-\underline{\partial}^\perp\Delta^{-1}\zeta(\underline{x},t)$ is of class $C^\infty(\Omega\times[0,T])$. And that the equations for the "current" that as the time s varies passes at time $s=t$ through the point $\underline{x}$

$$\partial_s U^\zeta_{s,t}(\underline{x})=\underline{v}^\zeta(U^\zeta_{s,t}(\underline{x}),s),\qquad U^\zeta_{t,t}(\underline{x})\equiv\underline{x}$$

admit a global solution of class C^∞ in $\underline{x}\in\Omega$, $s,t\in[0,T]$. (*Idea:* The solution is global because Ω has no boundary).

[3.1.13]: (*the vorticity map*) Check that the preceding problem implies that the following operator $\mathcal{R}$, "*vorticity map*" is well defined on $\mathcal{M}_0$

$$\mathcal{R}\zeta(\underline{x},t)=\omega_0(U^\zeta_{0,t}(\underline{x}))$$

and that $\mathcal{R}\mathcal{M}_0\subset\mathcal{M}_0$. (*Idea:* One has to check only property (3): note that the field $\underset{\sim}{\zeta}$ has zero divergence, hence the transformation $U^\zeta_{0,t}(\underline{x})=\underline{x}'$ conserves the volume and therefore $\int_\Omega\mathcal{R}\zeta(\underline{x},t)\,d\underline{x}\equiv\int_\Omega\omega_0(\underline{x})\,d\underline{x}$).

[3.1.14]: (*fixed point interpretation of Euler flows*) The existence of a point $\omega\in\mathcal{M}_0$ which is a fixed point for $\mathcal{R}$ implies that ω satisfies the Euler equation. Hence we can write the latter equation as $\omega=\mathcal{R}\omega$. (*Idea:* Note that Euler equation says that vorticity is transported by the current hence at time t it has in $\underline{x}$ the value that at time 0 it had in the point $\underline{x}'$ that is transported by the current to $\underline{x}$ in time t: *i.e.* just $U^\omega_{0.t}(\underline{x})$).

[3.1.15]: (*continuous dependence of the velocity field from the vorticity field*) Show that the field $\underline{v}^\zeta=-\partial^\perp\Delta^{-1}\zeta$, see [3.1.12], verifies the following properties

$$||\underline{v}^\zeta(\underline{x},t)||_0\le C_0L||\zeta||_0,\qquad \frac{|\underline{v}^\zeta(\underline{x},t)-\underline{v}^\zeta(\underline{x}',t)|}{(L^{-1}|\underline{x}-\underline{x}'|)\log_+(L|\underline{x}-\underline{x}'|^{-1})}\le C_0L||\zeta||_0$$

where $\log_+ z \equiv \log(e+z)$ (here $\log_+$ can obviously be replaced by any function of z continuous, positive, increasing and asymptotic to $\log z$ for $z \to \infty$). (*Idea:* Note that $\underline{v}^\zeta = -\underline{\partial}^\perp \Delta^{-1}\zeta$ and the operator $\underline{\partial}\Delta^{-1}$ can be computed as an integral operator starting from the Green function of the Laplacian,, *i.e.* $\frac{1}{2\pi}\log|\underline{x}-\underline{y}|^{-1}$, by the method of images, *c.f.r.* [2.3.12]. One gets formulae containing series, over a label $\underline{n}$ with integer components (*sums over the images of* Ω), that are not absolutely convergent

$$\sum_{\underline{n}} \int_\Omega \frac{d\underline{y}}{2\pi}\,\frac{(\underline{x}-\underline{y}+\underline{n}L)^\perp}{|\underline{x}-\underline{y}+\underline{n}L|^2}\,\zeta(\underline{y}) \equiv \sum_{\underline{n}} \int_\Omega d\underline{y}\,\underline{K}(\underline{x}-\underline{y}+\underline{n}L)\,\zeta(\underline{y})$$

but the hypothesis that ζ has zero average allows us to rewrite them as sums of absolutely convergent series and this reduces the problem to that of showing the validity term by term of the estimates (*i.e.* for each integer vector $\underline{n}$). The sum over all the $\underline{n}$ with $|\underline{n}|>2$ can be trivially bounded and the case that one really has to understand is the case $\underline{n}=\underline{0}$. The latter is also easy for what concerns the first estimate. The $\underline{n}=\underline{0}$ contribution to the second estimate is obtained by writing the difference between the two integrals for $\underline{v}^\zeta(\underline{x},t)$ and $\underline{v}^\zeta(\underline{x}',t)$ as an integral over Ω of the difference of the integrands. The integral over Ω can then be decomposed into the integral over the sphere of radius $2r = 2|\underline{x}-\underline{x}'|$ and center in $\underline{x}$ and in the integral on the complement. The first integral is estimated by introducing the modulus under the integral and bounding separately the two terms: it yields a result proportional to $r||\zeta||_0$ while the second integral will be bounded by Lagrange mean value theorem by estimating the difference $\underline{K}(\underline{x}-\underline{y}) - \underline{K}(\underline{x}'-\underline{y})$ as $r|\underline{y}|^{-2}$, because now $|\underline{y}-\underline{x}|$ and $|\underline{y}-\underline{x}'|$ can be bounded by a constant times $|\underline{y}|$ since $\underline{y}$ is "far" from both $\underline{x}$ and $\underline{x}'$, and the integral over the complement of the small sphere is bounded by $L^{-1}r\,\log L\,r^{-1}$, leading to the second inequality).

[3.1.16]: (*continuous dependence of the flow lines from the vorticity field*) There exists C_0 such that, defining $M_0 = ||\omega||_0$ and $\delta \stackrel{def}{=} e^{-C_0\,M_0\,T}$ for $T>0$, then the currents generated by the two fields $\underline{v}^\zeta, \underline{v}^{\zeta'}$ with $\zeta,\zeta' \in \mathcal{M}_0$ are such that for all $\underline{x}\in\Omega$ and $s,t\in[0,T]$

$$|U^\zeta_{s,t}(\underline{x}) - U^{\zeta'}_{s,t}(\underline{x})| \le C_0\,M_0\,T\,\left(||\zeta-\zeta'||_0 M_0^{-1}\right)^\delta L$$

Hence the $\mathcal{R}$ can be thought of as defined on the closure $\overline{\mathcal{M}}_0$ of $\mathcal{M}_0$ with respect to the metric of the uniform convergence and it can be extended to all continuous vector fields ζ, without any differentiability property, verifying the (2),(3),(4) of problem [3.1.11]. (*Idea:* Setting $\rho_s = L^{-1}(U^\zeta_{s,t}(\underline{x}) - U^{\zeta'}_{s,t}(\underline{x}))$, note that $\rho_t = \underline{0}$; furthermore from problem [3.1.15], by subtracting and adding $\underline{v}^\zeta(U^{\zeta'}_{s,t}(\underline{x}),s)$, there exist constants C_1, C_2, C_0:

$$\begin{aligned}|\partial_s\rho_s| =& |L^{-1}|\underline{v}^\zeta(U^\zeta_{s,t}(\underline{x}),s) - \underline{v}^{\zeta'}(U^{\zeta'}_{s,t}(\underline{x}),s)| \le C_1|\rho_s|\log_+|\rho_s|^{-1} + \\ &+ L^{-1}|\underline{v}^\zeta(U^{\zeta'}_{s,t}(\underline{x}),s) - \underline{v}^{\zeta'}(U^{\zeta'}_{s,t}(\underline{x}),s)| \le \\ &\le C_1|\rho_s|\log_+|\rho_s|^{-1} + C_2||\zeta-\zeta'||_0 \le \\ &\le C_0\big(|\rho_s|\log_+|\rho_s|^{-1} + ||\zeta-\zeta'||_0\big)\end{aligned}$$

Since $0\le s\le T$, by integration it follows that $|\rho_s|\le R$ if R is

$$\int_0^R \frac{d\rho}{\rho\log_+\rho^{-1} + ||\zeta-\zeta'||_0 M_0^{-1}} = C_0 M_0 T$$

which means that R can be taken $R \le K(M_0T)(||\zeta-\zeta'||M_0^{-1})^\delta$ with $K(M_0T)$ a continuous increasing function and $\delta = \exp -M_0C_0T$).

[3.1.17]: (*Hölder continuity of the flow lines*) Show that the current lines verify

$$\frac{|U^\zeta_{s,t}(\underline{x}) - U^\zeta_{s',t'}(\underline{x}')|}{(L^{-1}|\underline{x}-\underline{x}'|)^\delta + (T^{-1}|s-s'|)^\delta + (T^{-1}|t-t'|)^\delta} \le F(M_0T)L$$

where F is an increasing continuous function of its argument; $M_0 \equiv ||\zeta||_0$. (*Idea:* Set $\rho_s = L^{-1}(U^\zeta_{s,t}(\underline{x}) - U^\zeta_{s,t}(\underline{x}'))$, by the inequality in [3.1.15], one finds

$$|\partial_s\rho_s| \le C_0M_0|\rho_s|\log_+|\rho_s|^{-1}, \qquad \rho_t \equiv \underline{x}-\underline{x}'$$

hence $|\rho_s| \le R$, where R is such that

$$\int_{L^{-1}|\underline{x}-\underline{x}'|}^{R} \frac{d\rho}{\rho\log_+\rho^{-1}} = M_0C_0T \quad\Rightarrow\quad R \le (L^{-1}|\underline{x}-\underline{x}'|)^\delta\, F'(M_0T)$$

where F' is a suitable continuous increasing function and δ is as in the preceding problem. Moreover note that, by the first of the inequalities in [3.1.15]: $L^{-1}|U^\zeta_{s,t}(\underline{x}') - U^\zeta_{s',t}(\underline{x}')| \equiv |L^{-1}\int_s^{s'}\underline{v}^\zeta(U^\zeta_{\sigma,t}(\underline{x}'))\,d\sigma| \le C_0M_0|s-s'|$. Finally if $\underline{x}'' = U^\zeta_{t,t'}(\underline{x}')$ then

$$L^{-1}|U^\zeta_{s',t}(\underline{x}') - U^\zeta_{s',t'}(\underline{x}')| \equiv L^{-1}|U^\zeta_{s',t}(\underline{x}') - U^\zeta_{s',t}(\underline{x}'')| \le (L^{-1}|\underline{x}'-\underline{x}''|)^\delta\, F'(M_0T)$$

by what already seen. But $|\underline{x}'-\underline{x}''| \equiv |\int_t^{t'}\underline{v}^\zeta(U^\zeta_{\sigma,t'}(\underline{x}'),\sigma)\,d\sigma| \le LM_0C_0|t-t'|$ and the conclusion follows).

[3.1.18]: (*the vorticity map regularizes*) Show that the $\mathcal{R}$ transforms $\overline{\mathcal{M}}_0$ into the subspace $\overline{\mathcal{M}}_0^\delta$ of the continuous functions verifying (2),(3),(4) of [3.1.12] and the

$$||\zeta||_\delta \equiv \sup_{\Omega\times[0,T]\times\Omega\times[0,T]} \frac{|\zeta(\underline{x},t)-\zeta(\underline{x}',t')|}{(L^{-1}|\underline{x}-\underline{x}'|)^\delta + (T^{-1}|t-t'|)^\delta} \le M_0F''(M_0T) \qquad (!)$$

for a suitable function F''. Hence the continuous transformation $\mathcal{R}$ transforms the convex set $\overline{\mathcal{M}}_0^\delta$ into itself; furthermore such set is, as usually said, "compact" in the uniform convergence topology because it consists in a set of equicontinuous equibounded functions (Ascoli–Arzelá theorem) hence by using the sad axiom of choice (and its consequence expressed by the Schauder fixed point) we infer the existence of a field $\omega \in \overline{\mathcal{M}}_0^\delta$ such that $\omega = \mathcal{R}\omega$. (*Idea:* All follows from the form of $\mathcal{R}$, from the regularity of ω_0 and from the inequality in [3.1.17]. Furthermore the space of the functions that verify the inequality (*) is closed in the uniform convergence topology).

[3.1.19] (*extension of the regularization property of the vorticity map to higher order derivatives*) The result of [3.1.17] can be generalized, with some patience. If we define for $0<\gamma<1$

$$||f||_{k+\gamma} = ||f||_k + \sup_{(\underline{x},t)\in\Omega\times[0,T]} \sum_{|\underline{\alpha}|+\beta=k} \frac{L^{|\underline{\alpha}|}T^\beta|\partial_{\underline{x}}^{\underline{\alpha}}\partial_t^\beta f(\underline{x},t) - \partial_{\underline{x}}^{\underline{\alpha}}\partial_t^\beta f(\underline{x}',t')|}{(L^{-1}|\underline{x}-\underline{x}'|)^\gamma + (T^{-1}|t-t'|)^\gamma}$$

then, if $\gamma+\delta<1$

$$||\mathcal{R}\zeta||_{k+\gamma+\delta} \le F_k||\zeta||_{k+\gamma}$$

and if $\gamma+\delta>1$ then: $||\mathcal{R}\zeta||_{k+1} \le F_k||\zeta||_{k+\gamma}$, where F_k are suitable constants depending only from ω_0 and through its first $k+1$ derivatives. In other words $\mathcal{R}$ "regularizes", transforming C^k into $C^{k+\gamma}$.

[3.1.20] (*global existence and smoothness theorem for Euler flows* (Wolibner, Yudovitch, Kato) Show that the problem [3.1.19] implies that ω, the fixed point of $\mathcal{R}$ in $\overline{\mathcal{M}}_0$, is C^∞; and its derivatives can be bounded to order k in terms of the derivatives of order $\le k-\delta+1$ of the function ω_0. (*Idea:* $\mathcal{R}$ "regularizes" hence $\mathcal{R}\omega$ is more regular than ω; but it is equal to ω so that ω is C^∞).

[3.1.21] Check that, sadly enough, all what has been said above is, sadly, not sufficient to allow us to write a computation program that produces as a result the ω within a prefixed approximation ε in the metric of C^k, for any k ($k = 0$ included). Meditate on the event (or disaster) and if possible find a solution: unlike what is often stated, or unless the estimates of this section are substantially improved, the problem of global existence of solutions of the Euler equation is completely open even in 2 dimensions at least if one demands the "constructivity" of the method employed.

Bibliography: The global existence, smoothness and uniqueness theory for Euler flows in 2 dimensions is taken from [Ka67] and is due to Wolibner, Judovitch and Kato.

§3.2 Weak global existence theorems for NS. Autoregularization, existence, regularity and uniqueness for $d = 2$

We shall consider an incompressible fluid enclosed in a cubic region $\Omega \subset R^d$ with side L, with periodic boundary conditions and subject (for simplicity) to a time independent or quasi periodic volume force $\underline{g}$ of class $C^\infty(\Omega)$ exercising a vanishing total force on the fluid ($\int_\Omega \underline{g}\, d\xi = \underline{0}$).

In this case the center of mass of the fluid moves with rectilinear uniform motion. Calling $\underline{v} = \int \underline{u}\, d\xi / L^d$ the baricenter velocity, one can write the Navier–Stokes equations in the frame which "rotates" on the torus Ω uniformly with velocity $\underline{v}$. The galileian coordinate transformation is:

$$\begin{aligned} \xi' &= \xi - \underline{v}t, & \underline{u}'(\xi', t) &= \underline{u}(\xi' + \underline{v}t, t) - \underline{v}, \\ p'(\xi', t) &= p(\xi' + \underline{v}t, t), & \underline{g}'(\xi', t) &= \underline{g}(\xi + \underline{v}t, t) \end{aligned} \tag{3.2.1}$$

so that one sees that $\underline{u}'$ verifies in Ω the equations

$$\begin{aligned} &\dot{\underline{u}} + (\underset{\sim}{\underline{u}} \cdot \underset{\sim}{\underline{\partial}})\underline{u} = -\underline{\partial}\, p + \nu \Delta \underline{u} + \underline{g} \\ &\underline{\partial} \cdot \underline{u} = 0, \qquad \int_\Omega \underline{u}\, d\xi = \underline{0} \end{aligned} \tag{3.2.2}$$

where units are so chosen that density is $\rho = 1$.

Thus if, *as we shall always suppose*, $\int \underline{g}\, d\xi \equiv \underline{0}$ it is not restrictive to assume that $\underline{v} = \underline{0}$ provided we suppose that $\underline{g}$ is quasi periodic in t; in fact the function $\underline{g}(\xi + \underline{v}t, t)$ is quasi periodic in t even if $\underline{g}$ is time independent, because of its periodicity in ξ.

Condition $\underline{\partial} \cdot \underline{u} = 0$ can be regarded as a constraint and we can eliminate it by choosing, as already seen several times (*e.g. c.f.r.* §2.2), suitable coordinates to represent $\underline{u}$. More precisely we shall take as coordinates the coefficients $\underline{\gamma}_{\underline{k}}$ of its Fourier transform that we define, with the conventions of (2.2.2), as

$$\underline{u}(\xi,t) = \sum_{\underline{k}\neq\underline{0}} \underline{\gamma}_{\underline{k}}(t) e^{i\underline{k}\cdot\xi}, \qquad \underline{\gamma}_{\underline{k}} \equiv \overline{\underline{\gamma}}_{-\underline{k}}, \qquad \underline{\gamma}_{\underline{k}} \cdot \underline{k} \equiv 0 \tag{3.2.3}$$

where $\underline{k} \neq \underline{0}$ expresses the relation $\int \underline{u}\, d\xi \equiv \underline{0}$ (*momentum conservation*), $\underline{\gamma}_{\underline{k}} = \overline{\underline{\gamma}}_{-\underline{k}}$ expresses that $\underline{u}$ is real valued and the condition $\underline{\gamma}_{\underline{k}} \cdot \underline{k} = 0$ is equivalent to $\underline{\partial} \cdot \underline{u} = 0$. Here, by our periodicity assumption on the sides of Ω, the vector $\underline{k}$ is a vector with components that are integer multiples of $k_0 \stackrel{def}{=} 2\pi L^{-1}$ and, therefore, $|\underline{k}| \geq k_0$ because $\underline{k} \neq \underline{0}$; the latter property will be often used in what follows. Consistently with (2.2.2) we shall denote:

$$||f||_2^2 = \int_\Omega |f(\underline{x})|^2\, d\underline{x}, \qquad \text{and} \quad ||\hat f||_2^2 = \sum_{\underline{k}} |\hat f(\underline{k})|^2 \;\longleftrightarrow\; ||f||_2^2 = L^d ||\hat f||_2^2 \tag{3.2.4}$$

so that the $\underline{\gamma}_{\underline{k}}$ are "proper" *Lagrangian* coordinates, which can be freely assigned, without further constraints, as discussed in §2.2.

Assuming $p, \underline{u} \in C^\infty(\Omega)$ and substituting in (2.2.13) one sees that the NS equation becomes the (2.2.10). To recall the notations of §2.2 (*c.f.r.* (2.2.7)%(2.2.11)) let $\underline{\gamma}_{\underline{k}} = \overline{\underline{\gamma}}_{-\underline{k}}$ and $\underline{k} \cdot \underline{\gamma}_{\underline{k}} = 0$ and let $\Pi_{\underline{k}}$ be the orthogonal projection, in R^d, on the plane orthogonal to $\underline{k}$ and let $\Pi^{||}_{\underline{k}}$ be the orthogonal projection, in R^d, on $\underline{k}$:

$$(\Pi_{\underline{k}}\, \underline{w})_i \equiv w_i - k_i \frac{\underline{k}\cdot\underline{w}}{|\underline{k}|^2}\,, \quad \underline{w} \equiv \frac{\underline{w}\cdot\underline{k}}{\underline{k}^2}\underline{k} + \Pi_{\underline{k}}\, \underline{w} \equiv \Pi^{||}_{\underline{k}}\, \underline{w} + \Pi_{\underline{k}}\, \underline{w} \tag{3.2.5}$$

Set $\underline{\varphi}_{\underline{k}} \stackrel{def}{=} \Pi_{\underline{k}}\, \underline{g}_{\underline{k}}$ and define $t_{\underline{k}}$ so that it is $\underline{g}_{\underline{k}} = -i\underline{k}\, t_{\underline{k}} + \underline{\varphi}_{\underline{k}}$; then (2.2.10) can be written:

$$\begin{aligned}
\dot{\underline{\gamma}}_{\underline{k}} =& -\underline{k}^2 \nu\, \underline{\gamma}_{\underline{k}} - i \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \underset{\sim}{\gamma}_{\underline{k}_1} \cdot \underset{\sim}{k}_2\, \Pi_{\underline{k}}\, \underline{\gamma}_{\underline{k}_2} + \underline{\varphi}_{\underline{k}} \stackrel{def}{=} \\
&\stackrel{def}{=} -\underline{k}^2\nu\, \underline{\gamma}_{\underline{k}} + N_{\underline{k}}(\underline{\gamma}) + \underline{\varphi}_{\underline{k}} \\
p_{\underline{k}} =& -\frac{1}{\underline{k}^2} \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} (\underset{\sim}{\gamma}_{\underline{k}_1} \cdot \underset{\sim}{k}_2)\, (\underline{\gamma}_{\underline{k}_2} \cdot \underline{k}) + t_{\underline{k}}
\end{aligned} \tag{3.2.6}$$

It will be very useful to realize that *the nonlinear term $N_{\underline{k}}(\underline{\gamma})$ in (3.2.6) is meaningful as soon as $||\underline{u}||_2^2$ is finite*: for instance if $||\underline{u}||_2^2$ is bounded by a constant E_0, for all t. In fact, one has

$$|N_{\underline{k}}(\underline{\gamma})| \leq |\underline{k}|\, ||\underline{\gamma}||_2^2 \tag{3.2.7}$$

as it can be seen (from the first sum in (3.2.6)) by remarking that $\underline{\gamma}_{\underline{k}_1} \cdot \underline{k}_2 \equiv \underline{\gamma}_{\underline{k}_1} \cdot \underline{k}$ (because $\underline{k}_2 = \underline{k} - \underline{k}_1$ and $\underline{k}_1 \cdot \underline{\gamma}_{\underline{k}_1} = 0$) and, thence, by applying Schwartz' inequality to the sum $\sum_{\underline{k}_1}$, in (3.2.6), keeping $\underline{k}_2 \equiv \underline{k} - \underline{k}_1$.

Remark: An important property of $N_{\underline{k}}(\underline{\gamma})$ has appeared in the proof of (3.2.7): it is related to the orthogonality of $\underline{k}$ to $\underline{\gamma}_{\underline{k}}$, which allowed us to write *two equivalent expressions* for N:

$$N_{\underline{k}}(\underline{\gamma}) = -i \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} (\underset{\sim}{\gamma}_{\underline{k}_1} \cdot \underset{\sim}{k}_2)\, \Pi_{\underline{k}}\, \underline{\gamma}_{\underline{k}_2} \equiv -i \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} (\underset{\sim}{\gamma}_{\underline{k}_1} \cdot \underset{\sim}{k})\, \Pi_{\underline{k}}\, \underline{\gamma}_{\underline{k}_2} \tag{3.2.8}$$

showing that the factors $\underset{\sim}{k}_2$ and $\underset{\sim}{k}$ can be interchanged in this relation: a property that will be used several times in the following.

Therefore the problem consists in solving the first of the equations (3.2.6) subject to the conditions $\underline{\gamma}_{\underline{k}} = \overline{\underline{\gamma}}_{-\underline{k}}$ and to the initial condition $\underline{\gamma}_{\underline{k}}(0) = \underline{\gamma}^0_{\underline{k}}$ (the second equation in (3.2.6) should be regarded just as being the definition of p, as already noted in §2.2, *c.f.r.* (2.2.10)). Such initial condition is automatically imposed if one writes the equations as:

$$\begin{aligned}\underline{\gamma}_{\underline{k}}(t) =& e^{-\nu \underline{k}^2 t} \underline{\gamma}_{\underline{k}}(0) + \int_0^t e^{-\nu \underline{k}^2 (t-\tau)} \cdot \\ & \cdot \Big[\underline{\varphi}_{\underline{k}}(\tau) - i \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \underset{\sim}{\gamma}_{\underline{k}_1}(\tau) \cdot \underset{\sim}{k}_2 \Pi_{\underline{k}}\, \underline{\gamma}_{\underline{k}_2}(\tau) \Big]\, d\tau\end{aligned} \tag{3.2.9}$$

(*c.f.r.* (2.2.11)). Let us, therefore, set

1 Definition: *(weak solutions of NS and Euler equations) We shall say that $t \to \underline{u}(\underline{\xi}, t)$ is a weak solution of the NS equations, (3.2.2), with initial datum $\underline{u}^0(\underline{\xi}) \in L_2(\Omega)$ if there exists $E_0 > 0$ such that:*
(i) for each $t \ge 0$ it is: $||\underline{u}||_2^2 \le E_0$,
(ii) the functions $t \to \underline{\gamma}_{\underline{k}}(t)$ are continuous in t and verify (3.2.9).
(iii) the above makes sense even if $\nu = 0$, i.e. setting $\nu = 0$ and requiring properties (i), (ii) one gets the definition of weak solution for the Euler equations.

Remarks: We shall see that weak solutions with initial datum $\underline{u}^0 \in L_2(\Omega)$ always exist: however, in general, there might be many weak solutions with the same initial data: *the proof of existence is in fact non constructive*.

(1) If $\underline{u}$ is a weak solution then (3.2.9), (3.2.7) imply that $\underline{\gamma}_{\underline{k}}$ is differentiable almost everywhere[1] in t, and that the derivatives verify the first of

[1] This is the "Torricelli–Barrow" theorem: note that continuity of $\underline{\gamma}_{\underline{k}}(t)$ in t for each $\underline{k}$ *does not necessarily imply* continuity in t of $N_{\underline{k}}(\underline{\gamma})$ so that the "almost everywhere" is here necessary and expresses an important aspect of our lack of understanding on NS.

the (3.2.6). Furthermore $\underline{\gamma}_{\underline{k}}(t)$ is the integral of its derivative or, as one says, it is *absolutely continuous*.
(2) Most of what follows will concern weak solutions with initial datum $\underline{u}^0 \in C^\infty(\Omega)$. One might be interested in studying solutions in which the initial datum $\underline{u}^0$ is less regular than C^∞: many results can be easily extended to the case in which the initial datum is such that $\underset{\sim}{\partial}\underline{u}^0 \in L_2(\Omega)$ (*i.e.*, with the notations of problem [2.2.20] of §2.2, $\underline{u}^0 \in W^1(\Omega)$) or even just such that $\underline{u}^0 \in L_2(\Omega)$. It will simply suffice, as we shall point out whenever appropriate, to follow the analysis under this sole assumption. *However* one should note that the very derivation of the fluid mechanics equations (*c.f.r.* §1.1) *assumes regularity* of the velocity field so that non smooth initial data are of little physical interest or, to say the least, require a physical discussion of their meaning.
(3) However we shall see that *only* if $d = 2$ the initial regularity of the solution can be proved to be maintained at positive times. In fact in such case we shall see that even if the initial datum is only in $L_2(\Omega)$ there will be weak solutions that immediately become "regularized": becoming C^∞ at any positive time, and even analytic as a recent theorem shows (see below).
(4) Solutions may *a priori* exist that, starting from a smooth initial datum, evolve at a later time into singular ones, *i.e.* become non smooth (meaning that they do not have a well defined derivative, of some order). Such solutions can have an interesting physical significance. For instance they might signal cases in which the model of a continuum, based on smoothness of its motion, *becomes self contradictory* and therefore it should no longer be considered valid: the motion should be considered in a less phenomenological way (with respect to the theory of §1.1), ultimately possibly reverting to a model based on the microscopic structure of the fluid.
(5) A consequence of the uniform bound assumed on $||\underline{\gamma}||_2$ and of (3.2.7) is that any weak solution will verify, for all $t_0 \geq 0$

$$\begin{aligned}&\underline{\gamma}_{\underline{k}}(t) = e^{-\nu\underline{k}^2(t-t_0)}\underline{\gamma}_{\underline{k}}(t_0)+\\&+\int_{t_0}^t e^{-\nu\underline{k}^2(t-\tau)}\Big[\varphi_{\underline{k}}(\tau) - i\sum_{\substack{\underline{k}_1+\underline{k}_2=\underline{k}\\|\underline{k}_i|\leq\infty}} \underline{\gamma}_{\underline{k}_1}(\tau)\cdot \underset{\sim}{k}_2\,\Pi_{\underline{k}}\,\underline{\gamma}_{\underline{k}_2}(\tau)\Big]\,d\tau\end{aligned} \tag{3.2.10}$$

which we call the "*self-consistence property*" of the above weak solutions: *i.e.* the solution, $t \to \underline{\gamma}(t)$, regarded as a function defined in $[t_0, \infty)$ is still a solution of the NS equation with initial datum $\underline{\gamma}(t_0)$.[2]

To prove existence of weak solutions with initial data in $\underline{u}^0 \in C^\infty$ we replace (3.2.6) with a "*regularized*" equation parameterized by a "cut–off"

[2] Since we do not know, in general, uniqueness this property requires checking (immediate in this case). This follows from (3.2.9) , the additivity of the integrals and the conitnuity of $\underline{\gamma}_{\underline{k}}(t)$ for each $\underline{k}$.

parameter that we shall call R:

$$\begin{aligned}\dot{\underline{\gamma}}^R_{\underline{k}} &= -\,\nu \underline{k}^2 \underline{\gamma}^R_{\underline{k}} - i \sum_{\substack{\underline{k}_1+\underline{k}_2=\underline{k}\\ |\underline{k}_i|\le R}} (\underline{\gamma}^R_{\underline{k}_1}\cdot \underline{k}_2)\,\Pi_{\underline{k}}\,\underline{\gamma}^R_{\underline{k}_2} + \underline{\varphi}_{\underline{k}} \equiv \\ &\equiv -\,\nu \underline{k}^2 \underline{\gamma}^R_{\underline{k}} + N^R_{\underline{k}}(\underline{\gamma}^R) + \underline{\varphi}_{\underline{k}},\ , \qquad |\underline{k}|\le R\end{aligned} \tag{3.2.11}$$

where $\underline{\gamma}^R_{\underline{k}}(t) \stackrel{def}{=} 0$ for $|\underline{k}| > R$ and $\underline{\gamma}^R_{\underline{k}}(0) \stackrel{def}{=} \underline{\gamma}^0_{\underline{k}}$ for $|\underline{k}| \le R$. Note that, as in the case of (3.2.7), it is $|N^R_{\underline{k}}(\underline{\gamma}^R)| \le |\underline{k}|\,||\underline{\gamma}^R||_2^2$ *independently of* R. Then we begin by showing the following proposition about properties of (3.2.11) which are *independent of the value of the regularization parameter* R:

I. Proposition *(a priori bounds on solutions of regularized NS equations): Suppose* $\underline{u}^0 \in C^\infty(\Omega)$,
(i) equation (3.2.11) admits a solution, global for $t \ge 0$*, with initial datum* $\underline{\gamma}^0_{\underline{k}}$, $|\underline{k}| \le R$*, and such solution verifies the a priori estimate:*

$$||\underline{\gamma}^R(t)||_2 \equiv \big(\sum_{\underline{k}} |\underline{\gamma}^R_{\underline{k}}|^2\big)^{1/2} \le \max\big(||\underline{\gamma}^0||_2, \frac{||\underline{\varphi}||_2}{\nu k_0^2}\big) \stackrel{def}{=} \sqrt{E_0 L^{-d}} \tag{3.2.12}$$

for all $t \ge 0$, $R > 0$*, where* E_0 *is defined by (3.2.12); hence if* $\underline{\varphi} \equiv \underline{0}$ *it is* $E_0 \equiv ||\underline{u}^0||_2^2$*, c.f.r. (3.2.4).*
(ii) Furthermore:

$$\int_0^T d\tau \sum_{\underline{k}} \underline{k}^2 |\underline{\gamma}^R_{\underline{k}}(\tau)|^2 \le \frac{1}{2} E_0 L^{-d} \nu^{-1} + T\sqrt{E_0 L^{-d}}\nu^{-1}||\underline{\varphi}||_2 \tag{3.2.13}$$

for all $T > 0$ *and* $R > 0$.

proof: equation (3.2.11) is an ordinary differential equation and it it is enough to multiply (3.2.11) by $\overline{\underline{\gamma}_{\underline{k}}} \equiv \underline{\gamma}_{-\underline{k}}$ and to sum over $|\underline{k}| \le R$ to find, if $(\underline{f}, \underline{h})$ denotes the usual scalar product in $L_2(\Omega)$ of the fields $\underline{f}, \underline{h}$,

$$\begin{aligned}&\frac{1}{2}\frac{d}{dt}||\underline{\gamma}^R||_2^2 = -\nu \sum_{|\underline{k}|\le R} \underline{k}^2 |\underline{\gamma}^R_{\underline{k}}|^2 + (\underline{\varphi}\cdot\underline{\gamma}^R) - i \sum_{\underline{k}_1+\underline{k}_2+\underline{k}_3=\underline{0}} \underline{\gamma}^R_{\underline{k}_1}\cdot \underline{k}_2\,\underline{\gamma}^R_{\underline{k}_2}\cdot \underline{\gamma}^R_{\underline{k}_3} \\ &\equiv -\nu \sum_{|\underline{k}|\le R} \underline{k}^2 |\underline{\gamma}^R_{\underline{k}}|^2 + (\underline{\varphi}\cdot\underline{\gamma}^R) - \frac{i}{2} \sum_{\underline{k}_1+\underline{k}_2+\underline{k}_3=0} \underline{\gamma}^R_{\underline{k}_1}\cdot (\underline{k}_2 + \underline{k}_3)\,\underline{\gamma}^R_{\underline{k}_2}\cdot \underline{\gamma}^R_{\underline{k}_3} \equiv \\ &\equiv -\nu \sum_{|\underline{k}|\le R} \underline{k}^2 |\underline{\gamma}^R_{\underline{k}}|^2 + (\underline{\varphi}\cdot\underline{\gamma}^R)\end{aligned} \tag{3.2.14}$$

having used, in the second step, the symmetry between the summation labels $\underline{k}_2$ and $\underline{k}_3$ and, in the third step, the property that $\underline{k}_2 + \underline{k}_3$ is parallel to $\underline{k}_1$ while $\underline{\gamma}_{\underline{k}_1}$ is, instead, orthogonal to $\underline{k}_1$.

It follows that the right hand side is ≤ 0 if $\sum_{\underline{k}} |\underline{\gamma}^R_{\underline{k}}|^2 > E_0 L^{-d}$: therefore equation (3.2.12) holds. Moreover by integrating (3.2.14) we get:

$$\frac{1}{2}(||\underline{\gamma}^R(t)||_2^2 - ||\underline{\gamma}^R(0)||_2^2) \le -\nu \int_0^t \sum_{|\underline{k}|\le R} \underline{k}^2 |\gamma^R_{\underline{k}}|^2 \, d\tau + \int_0^t ||\underline{\varphi}||_2 ||\underline{\gamma}^R||_2 d\tau \tag{3.2.15}$$

which implies:

$$\int_0^t \sum_{|\underline{k}|\le R} \underline{k}^2 |\gamma^R_{\underline{k}}|^2 \, d\tau \le \frac{1}{2} L^{-d} E_0 \nu^{-1} + t\nu^{-1} ||\underline{\varphi}||_2 \sqrt{E_0 L^{-d}} \tag{3.2.16}$$

As a corollary, always assuming that $\underline{u}^0 \in C^\infty(\Omega)$, we get

II. Corollary *(global existence of weak solutions for the NS equations): Consider the regularized equation (3.2.11); then:*
(i) the functions $\underline{\gamma}^R_{\underline{k}}(t)$ *are bounded by* $\sqrt{E_0 L^{-d}}$ *and have first derivative with respect to t that can be bounded above by* $\nu \underline{k}^2 \sqrt{E_0 L^{-d}} + |\underline{k}| E_0 L^{-d} + ||\underline{\varphi}||_2$. *Hence there is a sequence* $R_j \to \infty$ *such that the limits* $\gamma^{R_j}_{\underline{k}}(t) \to \gamma^\infty_{\underline{k}}(t)$ *exist, for all* $\underline{k}$, *uniformly in every bounded interval inside* $t \ge 0$. *Each such* $\underline{\gamma}^\infty(t)$ *will be called a weak limit, as* $R \to \infty$, *of* $\underline{\gamma}^R(t)$.
(ii) Every weak limit $\underline{\gamma}^\infty$ *verifies, for all* $t \ge 0$

$$\begin{aligned} &||\underline{\gamma}^\infty||_2 \le \sqrt{E_0 L^{-d}} \\ &\int_0^t d\tau \sum_{\underline{k}} \underline{k}^2 |\underline{\gamma}^\infty_{\underline{k}}(\tau)|^2 \le \frac{1}{2} E_0 L^{-d} \nu^{-1} + t\nu^{-1} \sqrt{E_0 L^{-d}} ||\underline{\varphi}||_2 \end{aligned} \tag{3.2.17}$$

It also verifies (3.2.9): therefore it will be a weak solution.

proof: Boundedness of $\dot{\underline{\gamma}}_{\underline{k}}(t)$ immediately follows from (3.2.11) and from the successive remark; hence properties (i) and (ii) follow. Property (iii) is slightly more delicate to check; rewriting (3.2.11) for $|\underline{k}| \le R$ as:

$$\begin{aligned} &\underline{\gamma}^R_{\underline{k}}(t) = e^{-\nu \underline{k}^2 t} \underline{\gamma}^0_{\underline{k}} + \\ &+ \int_0^t e^{-\nu \underline{k}^2 (t-\tau)} \Big[\varphi_{\underline{k}}(\tau) - i \sum_{\substack{\underline{k}_1 + \underline{k}_2 = \underline{k} \\ |\underline{k}_i| \le R}} \underset{\sim}{\gamma}^R_{\underline{k}_1}(\tau) \cdot \underset{\sim}{k}_2 \, \Pi_{\underline{k}} \, \underline{\gamma}^R_{\underline{k}_2}(\tau) \Big] \, d\tau \end{aligned} \tag{3.2.18}$$

the problem is to take the limit under the integral sign in (3.2.18). Given that the first *a priori* bound in (3.2.17) guarantees the absolute convergence of the series obtained by taking the term by term limits, the passage to the limit will be possible if we shall show that the series in (3.2.18) is uniformly convergent for $R \to \infty$ (*i.e.* the remainder of its partial sum of order N approaches 0 uniformly in R as $N \to \infty$).

Fix $N > 0$ and recall (3.2.8) and the remark in (3.2.13): note that if $|\underline{k}_1|$ or $|\underline{k}_2|$ are $\ge N/2$ then $|\underline{k}_1| \ge N/2$ and $|\underline{k}_2| \ge k_0 \equiv 2\pi L^{-1}$ or viceversa, hence

$$\begin{aligned}&\int_0^t \sum_{\substack{\underline{k}_1+\underline{k}_2=\underline{k}\\ |\underline{k}_1|+|\underline{k}_2|>N}} \Big|\underline{\gamma}^R_{\underline{k}_1}(\tau)\cdot\underline{k}_2\Big|\Big|\underline{\gamma}^R_{\underline{k}_2}(\tau)\Big| \le \int_0^t |\underline{k}| \sum_{\substack{\underline{k}_1+\underline{k}_2=\underline{k}\\ |\underline{k}_1|+|\underline{k}_2|\ge N}} \Big|\underline{\gamma}^R_{\underline{k}_1}(\tau)\Big|\,\Big|\underline{\gamma}^R_{\underline{k}_2}(\tau)\Big| \le\\
&\le \frac{2|\underline{k}|}{Nk_0}\int_0^t \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} |\underline{k}_1|\,|\underline{\gamma}^R_{\underline{k}_1}(\tau)|\,|\underline{k}_2|\,|\underline{\gamma}^R_{\underline{k}_2}(\tau)| \le \frac{2|\underline{k}|}{Nk_0}\int_0^t \sum_{\underline{k}} |\underline{k}|^2|\underline{\gamma}_{\underline{k}}(\tau)|^2 \le\\
&\le \frac{2|\underline{k}|}{Nk_0}\Big(\frac{E_0L^{-d}}{2\nu} + \frac{t}{\nu}||\underline{\varphi}||_2\sqrt{E_0L^{-d}}\Big) \xrightarrow[N\to\infty]{} 0 \end{aligned} \tag{3.2.19}$$

where the integrals are performed with respect to $d\tau$ and we must understand that $\underline{\gamma}^R_{\underline{k}_i} \equiv \underline{0}$ if $|\underline{k}_i| \ge R$ or if $|\underline{k}_i| = 0$; in the first step of the second line we multiply and divide by $|\underline{k}_1|\,|\underline{k}_2|$ and bound from below the denominator $|\underline{k}_1|^{-1}|\underline{k}_2|^{-1}$ by $Nk_0/2$; and, finally, the vector $\underline{k}_2$, in the first inequality, is replaced by $\underline{k}$ using (3.2.8) obtaining (iii).

Remarks:
(1) Equations (3.2.9) have therefore a weak solution verifying (3.2.17) and (3.2.10): hence, almost everywhere in t, the first of (3.2.6). Consequently, the NS equations admit a weak solution, independently of the dimension $d \ge 2$ of the space into which the fluid flows, for all initial data $\underline{u}^0 \in C^\infty(\Omega)$ (with $\underline{\partial}\cdot\underline{u}^0 = 0$). Such solution is consistent with itself in the sense (3.2.10).
(2) *(Non smooth cases)*: However the same proof would work if we only assumed that $\underline{u}^0 \in L_2(\Omega)$ (with $\underline{\partial}\cdot\underline{u} = 0$ in the sense of distributions, *i.e.* $\underline{\gamma}^0_{\underline{k}}\cdot\underline{k} = 0$). We note also that the solutions discussed in the corollary have finite total vorticity $S(t) = \sum_{\underline{k}} |\underline{k}|^2|\underline{\gamma}_{\underline{k}}(t)|^2$ for almost all t: by (3.2.17) we do not even have to suppose that $\underline{u}^0 \in W^1(\Omega)$ because (3.2.16) puts a bound on the integral of the square of the $W^1(\Omega)$ norm of the solution which only depends on the L_2 norm of $\underline{u}^0$. Thus the NS equations in dimension $d = 2, 3$ admit a weak solution for all initial data in L_2 and such solution has finite vorticity for almost all times.

The solutions that arise in the above corollary might be not unique and it might be possible to exhibit other weak solutions by other methods. It is therefore convenient to give them a special name to distinguish them from weak solutions developed (later) by following other methods and which might be different. Hence we set the following definition

2 Definition *(C–weak solutions): A weak solution obtained through the limits in (i) of corollary II with initial datum in* $\underline{u}^0 \in L_2(\Omega)$*, i.e. from the solutions of the cut–off regularized equations (3.2.11), will be called a "C-weak soluion" of the NS equation.*

We shall see that there could be weak solutions different from them, even if $\underline{u}^0 \in W^1(\Omega)$, *c.f.r.* below and §3.3.

Equations (3.2.9) have other remarkable consequences. The most notable is certainly the *autoregularization theorem* valid *for all weak solutions*, (hence in particular for the C-weak solutions of definition 2

III. Proposition *(autoregularization): Let $\underline{u}(t)$ be a weak solution with initial datum $\underline{u}^0 \in C^\infty$. Given $T > 0$ suppose that, for some $\alpha \ge 0$, there exists a constant $C_\alpha > 0$ such that for all $0 \le t \le T$ it is $\sup_{\underline{k}} |\underline{k}|^\alpha |\underline{\gamma}_{\underline{k}}(t)| \le C_\alpha$. Then if $\alpha > d-1$ it will be $C_\beta < \infty$ for all $\beta > 0$: hence $\underline{u} \in C^\infty(\Omega)$.*

Remark: In other words $\underline{\gamma}_{\underline{k}}$ is the Fourier transform of an infinitely smooth solution the NS equation, if it is "just" the Fourier transform of a regular enough solution. Note that if $\underline{u} \in C^\infty(\Omega)$ is a weak solution considered in corollary II then $\underline{u}$ is also C^∞ in t because the time derivatives can be expressed in terms of the $\underline{x}$–derivatives by differentiating suitably many times the equations verified by $\underline{\gamma}_{\underline{k}}$ or $\underline{u}$ (*i.e.* the NS equations).

proof: Given (3.2.9) we can bound the nonlinear term:

$$|N_{\underline{k}}(\underline{\gamma})| \le \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} |\underline{\gamma}_{\underline{k}_1}|\,|\underline{\gamma}_{\underline{k}_2}|\,|\underline{k}| \le C_\alpha^2\,|\underline{k}| \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \frac{1}{|\underline{k}_1|^\alpha |\underline{k}_2|^\alpha} \le$$

$$\le \begin{cases} \infty & \alpha \le d/2 \\ |\underline{k}|\,C_\alpha^2\,B_\alpha |\underline{k}|^{-(2\alpha-d)} & d/2 < \alpha < d \\ |\underline{k}|\,C_\alpha^2\,B_\alpha\,|\underline{k}|^{-\alpha} \log|\underline{k}| & \alpha = d \\ |\underline{k}|\,C_\alpha^2\,B_\alpha\,|\underline{k}|^{-\alpha} & \alpha > d \end{cases} \tag{3.2.20}$$

where all vectors $\underline{k}, \underline{k}_i$ are different from $\underline{0}$.

Since $\int_0^t e^{-\nu \underline{k}^2 \tau} d\tau \le 1/\nu \underline{k}^2$ then, if $\alpha > d/2$ and if we take into account (3.2.8), we find an estimate of $\int_0^t N_{\underline{k}}(\underline{\gamma}(\tau))\,d\tau$ as

$$\Big|\int_0^t d\tau \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \underset{\sim}{\gamma}_{\underline{k}_1} \cdot \underset{\sim}{k}_2 \Pi_{\underline{k}}\, \underline{\gamma}_{\underline{k}_2} e^{-\underline{k}^2\nu(t-\tau)}\Big| \le \frac{C_\alpha^2 B_\alpha}{\nu|\underline{k}|} \begin{cases} |\underline{k}|^{-(2\alpha-d)} & \alpha < d \\ |\underline{k}|^{-\alpha} \log|\underline{k}| & \alpha = d \\ |\underline{k}|^{-\alpha} & \alpha > d \end{cases} \tag{3.2.21}$$

Suppose $\underline{u}^0 \in C^\infty(\Omega)$, hence $\sup_{\underline{k}} |\underline{k}|^\beta\,|\underline{\gamma}^0_{\underline{k}}| \stackrel{def}{=} C^0_\beta < \infty$ for all $\beta \ge 0$, and suppose $|\underline{k}|^\alpha\,|\underline{\gamma}_{\underline{k}}(t)| = C_\alpha < \infty$, for some $\alpha > d-1$. Then setting $\eta = \min(\alpha - d + 1, 1) > 0$ we have, therefore, shown by (3.2.21) that there exists $C'_{\alpha+\eta} < \infty$ such that

$$|\underline{\gamma}_{\underline{k}}(t)| \le \frac{C^0_{\alpha+\eta}}{|\underline{k}|^{\alpha+\eta}} + \frac{C'_{\alpha+\eta}}{|\underline{k}|^{\alpha+\eta}} \tag{3.2.22}$$

where the first term arises by bounding the first and second terms in (3.2.9) by $|\underline{\gamma}^0_{\underline{k}}| + \frac{|\underline{\varphi}_{\underline{k}}}{\nu \underline{k}^2} \le C^0_{\alpha+\eta}\,|\underline{k}|^{-\alpha-\eta}$: hence the claim in the theorem follows (by indefinitely repeating the argument gaining each time some extra decay in $|\underline{k}|$).

Remarks:

(1) It is important to note that if C_{α_0} is a constant such that $|\underline{k}|^{\alpha_0}|\underline{\gamma}_{\underline{k}}| \le C_{\alpha_0}$ and $\alpha_0 > d-1$, then not only it follows that $|\underline{k}|^{\alpha}|\underline{\gamma}_{\underline{k}}| \le C_\alpha$ for all $\alpha > \alpha_0$ and suitable C_α, but also that the constant C_α can be explicitly bounded in terms of C_{α_0} and of the quantities C^0_α relative to the initial datum and the forcing (*i.e.* $(|\gamma^0_{\underline{k}}| + \frac{|\varphi_{\underline{k}}|}{\nu \underline{k}^2})\,|\underline{k}|^\alpha, \le C^0_\alpha$) by a function of C_{α_0}, C^0_α that we shall call $B_\alpha(\alpha_0, C_{\alpha_0}, C^0_\alpha)$. An explicit bound on B_α can be easily derived from (3.2.20) but we shall not need it now (*c.f.r.* the proof of proposition IX below).
(2) Furthermore the bounds described above hold also for the cut–off regularized equations with a cut–off parameter $R < \infty$: this means that $|\gamma^R_{\underline{k}}||\underline{k}|^{\alpha_0} \le C_{\alpha_0}$ and $\alpha_0 > d-1$, imply $|\gamma^R_{\underline{k}}(t)||\underline{k}|^{\alpha} \le B_\alpha(\alpha_0, C_{\alpha_0}, C^0_\alpha)$ where the function B *coincides* with that of the non regularized case. We shall see that the last remarks will allow us to bound, in the case $d = 2$, the difference between $\underline{\gamma}$ and $\underline{\gamma}^R$.
(3) *(Non smooth cases):* In the above proof the assumption that $\underline{u}^0 \in C^\infty$ has been used only to insure that $\sup |\underline{k}|^\alpha |\underline{\gamma}^0_{\underline{k}}| \stackrel{def}{=} \overline{C}^0_\alpha < \infty$. We have in fact shown a lot more in the above proof. Indeed the first term in (3.2.9) can be alternatively bounded rather than by $\overline{C}^0_{\alpha+\eta}|\underline{k}|^{-\alpha-\eta}$ (which could be ∞ if we only assume that $\underline{u}^0 \in L_2(\Omega)$) by the quantity $\sqrt{L^{-d}E_0}e^{-\nu\underline{k}^2 t}$ which is certainly a finite bound, and a very good one, for $t > 0$. This bound, under the only assumption that $\underline{u}^0 \in L_2$ and $|\underline{k}|^\alpha\,|\underline{\gamma}_{\underline{k}}(t)| < C_\alpha$ for some $\alpha > d-1$ (which is a non trivial assumption if $d > 2$, see below), implies

$$
\begin{aligned}
|\underline{\gamma}_{\underline{k}}(t)| \le& \sqrt{L^{-d}E_0}\, e^{-\nu \underline{k}^2 t} + \frac{C'_{\alpha+\eta}}{|\underline{k}|^{\alpha+\eta}} + \frac{|\underline{\varphi}_{\underline{k}}|}{\nu \underline{k}^2}\\
|\underline{\gamma}_{\underline{k}}(t)| \le& \frac{k_0^{\alpha+\eta}\sqrt{L^{-d}E_0}\,\max_{x\ge0} x^{\alpha+\eta}e^{-x^2}}{(k_0^2\nu\, t)^{(\alpha+\eta)/2}\,|\underline{k}|^{\alpha+\eta}} + \frac{C'_{\alpha+\eta}}{|\underline{k}|^{\alpha+\eta}} + \frac{|\underline{\varphi}_{\underline{k}}|}{\nu \underline{k}^2} \le\\
\le& \big(1 + \frac{1}{(k_0^2\nu\, t)^{(\alpha+\eta)/2}}\big)\,\frac{C''_{\alpha+\eta}}{|\underline{k}|^{\alpha+\eta}} \qquad (3.2.23)\\
C''_{\alpha+\eta} =& C'_{\alpha+\eta} + k_0^{\alpha+\eta}\sqrt{L^{-d}E_0}\,\max_{x\ge0} x^{\alpha+\eta}e^{-x^2} + \max\frac{|\underline{k}|^{\alpha+\eta}||\underline{\varphi}_{\underline{k}}|}{\nu|\underline{k}|^2}
\end{aligned}
$$

Recursively this means, of course, that the solution will be C^∞ for $t > 0$: *even* if the initial datum $\underline{u}^0$ is just L_2 and *if*, for some, $\alpha > d-1$ one has the further (in general very nontrivial if $d > 2$, see below) information that $\sup_{\underline{k}} |\underline{k}|^\alpha\,|\underline{\gamma}_{\underline{k}}(t)| < \infty$. In this case the quantities $\sup_{\underline{k}} |\underline{k}|^\beta\,|\underline{\gamma}_{\underline{k}}(t)|$ are $< \infty$ for all $\beta \ge 0$ and they can be bounded by a t dependent quantity which diverges as $t \to 0$ as an easily computable inverse power of t, *c.f.r.* (3.2.23).

The above regularization theorems are very useful in the two dimensional case, $d = 2$: in such case in fact they imply the following theorem:

IV. Proposition: *(existence and smoothness for NS in dimension 2) If $d = 2$ the C-weak solutions*[3] *with datum $\underline{u}^0 \in L_2(\Omega)$ are velocity fields of class C^∞ in $\underline{x}$ and t, for $t > 0$, and even for $t \ge 0$ when one supposes that $\underline{u}^0$ is of class C^∞ as well.*
Furthermore if $\underline{u}^0 \in W^1(\Omega)$ (i.e. such that $\underset{\sim}{\partial}\underline{u} \in L_2(\Omega)$, c.f.r. problem [2.2.20]) and if one sets: $F_{-1} = \big(\sum_{\underline{k}} |\varphi_{\underline{k}}|^2 |\underline{k}|^2\big)^{1/2}$, one finds that the solutions $\underline{\gamma}^R_{\underline{k}}(t)$ of (3.2.18) for $R \le \infty$ verify

$$\Big(\sum_{|\underline{k}|\le R} |\underline{k}|^2 |\underline{\gamma}^R_{\underline{k}}(t)|^2\Big)^{1/2} \le \qquad (3.2.24)$$
$$\le \max\Big[\big(F_{-1}\nu^{-1}k_0^{-2}\big),\ \big(\sum_{\underline{k}} \underline{k}^2 |\gamma_{\underline{k}}(0)|^2\big)^{1/2}\Big] \stackrel{def}{=} \sqrt{S_0}$$

where k_0 is the minimum value that $|\underline{k}|$ can take (i.e. $k_0 = 2\pi L^{-1}$).

proof: Suppose first that $\underline{u}^0 \in W^1(\Omega)$ so that $S_0 < \infty$. Note that:

$$\sum_{|\underline{k}_3|\le R} N_{\underline{k}_3}(\underline{\gamma}^R) \cdot \overline{\underline{\gamma}}^R_{k_3} |\underline{k}_3|^2 \equiv 0, \qquad \text{if } d = 2 \qquad (3.2.25)$$

that can be proved by remarking that in this case:

$$\underline{\gamma}_{\underline{k}} = \gamma_{\underline{k}} \frac{\underline{k}^\perp}{|\underline{k}|}, \qquad \gamma_{\underline{k}} = -\overline{\gamma}_{-\underline{k}}, \qquad \text{if} \quad \underline{k}^\perp = (k_2, -k_1), \qquad \underline{k} = (k_1, k_2) \qquad (3.2.26)$$

with $\gamma_{\underline{k}}$ *scalar* and therefore:

$$\underset{\sim}{k}_3^2\, (\underline{\gamma}_{\underline{k}_1} \cdot \underset{\sim}{k}_2)(\underline{\gamma}_{\underline{k}_2} \cdot \underline{\gamma}_{\underline{k}_3}) = \frac{(\underline{k}_1^\perp \cdot \underline{k}_2)\,(\underline{k}_2^\perp \cdot \underline{k}_3^\perp)}{|\underline{k}_1|\ |\underline{k}_2|\ |\underline{k}_3|} \underline{k}_3^2\, \gamma_{\underline{k}_1} \gamma_{\underline{k}_2} \gamma_{\underline{k}_3} \qquad (3.2.27)$$

Summing over the permutations of $\underline{k}_1, \underline{k}_2, \underline{k}_3$ and using $\underline{k}_1^\perp \cdot \underline{k}_2 = \underline{k}_2^\perp \cdot \underline{k}_3 = \underline{k}_3^\perp \cdot \underline{k}_1$ we find zero (see also (2.2.29), that imply the (3.2.25), for a rapid and interesting way to reach this conclusion). Hence the same argument used in connection with the above *a priori* estimate leads to:

$$\frac{d}{dt}\frac{1}{2} \sum_{|\underline{k}|\le R} |\underline{k}|^2 |\underline{\gamma}^R_{\underline{k}}|^2 = -\nu \sum_{|\underline{k}|\le R} |\underline{k}|^4 |\underline{\gamma}^R_{\underline{k}}|^2 + \sum_{\underline{k}} \underline{k}^2 \varphi_{\underline{k}} \underline{\gamma}^R_{\underline{k}} \qquad (3.2.28)$$

Hence one gets: $\dot S/2 \le -\nu k_0^2 S + F_{-1}\sqrt{S}$: *i.e.* $\dot S/2 \le 0$ if $S > S_0$, with S_0 defined in (3.2.24), so that the inequality (3.2.24) follows.
Thus the weak solutions must verify:

$$S(t) = \sum_{\underline{k}} |\underline{k}|^2 |\underline{\gamma}_{\underline{k}}(t)|^2 \le S_0, \qquad \forall\ t \ge 0 \qquad (3.2.29)$$

[3] See definition 2 and the remarks preceding proposition III.

and this implies, by Schwartz' inequality:

$$
\begin{aligned}
|N_{\underline{k}}(\underline{\gamma})| &\le |\underline{k}| \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} |\underline{\gamma}_{\underline{k}_1}|\,|\underline{\gamma}_{\underline{k}_2}| \le \\
&\le |\underline{k}| \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \frac{|\underline{k}_1|\,|\underline{\gamma}_{\underline{k}_1}|\,|\underline{k}_2|\,|\underline{\gamma}_{\underline{k}_2}|}{|\underline{k}_1|\,|\underline{k}_2|} \le \frac{2S_0}{k_0}
\end{aligned}
\tag{3.2.30}
$$

because if $\underline{k}_1 + \underline{k}_2 = \underline{k}$ then at least one among $\underline{k}_1$ and $\underline{k}_2$ has modulus $\ge |\underline{k}|/2$ (and both have modulus $\ge k_0$, being non zero) and therefore we can say that: $|\underline{k}_1||\underline{k}_2| \ge k_0\,|\underline{k}|/2$.

Hence, by (3.2.30),(3.2.9), and if $|\underline{\varphi}_{\underline{k}}| \le G_\alpha |\underline{k}|^{-\alpha}$, we deduce:

$$
|\underline{\gamma}^R_{\underline{k}}(t)| \le e^{-\nu \underline{k}^2 t} |\underline{\gamma}^R_{\underline{k}}(0)| + \Big(\frac{G_0}{\nu \underline{k}^2} + \frac{2S_0}{\nu \underline{k}^2\, k_0}\Big)(1 - e^{-\nu \underline{k}^2 t})
\tag{3.2.31}
$$

and we see as well that $|\underline{\gamma}^\infty_{\underline{k}}(t) - e^{-\underline{k}^2 \nu t} \underline{\gamma}^0_{\underline{k}}| \le (G_0 \nu^{-1} + 2S_0/k_0)/|\underline{k}|^2$, which will be useful later.

Hence there is a constant $C_2(t)$ such that $|\underline{\gamma}_{\underline{k}}(t)| \le C_2(t)/|\underline{k}^2|$ with $C_2(t) = const\,(k_0^2 + (\nu t)^{-1})$ and we *can apply the autoregularization theorem for* $\alpha > d - 1$ *(as* $\alpha = 2$ *and* $d - 1 = 1$*) to the evolution considered for* $t > 0$ (using the self-consistency property of corollary II).

Hence the solution (which we shall show, in proposition VI below, to be unique) $\underline{u}^\infty(\underline{x}, t)$ is C^∞ in $\underline{x}$ for each t, and each of its derivative is uniformly bounded in time (in terms of the constants $C_\alpha(t)$ which as $t \to 0$ may diverge if $\underline{u}^0 \in W^1(\Omega)$ but which stay finite if $\underline{u}^0 \in C^\infty(\Omega)$). It follows that, by differentiating with respect to t both sides of the equation (3.2.9), for instance, that $\underline{u}$ in class C^∞ in $\underline{x}, t$, for $t > 0$.

The case $\underline{u}^0 \in L_2(\Omega)$ (when S_0 could be ∞) is immediately reduced to the one just treated. Because the estimate (3.2.17) remains valid for the weak solutions of corollary II even when the datum is just L_2 and it imples that $\underline{u}(t) \in W^1(\Omega)$ for almost all times $t > 0$: therefore it suffices to consider as initial datum the value of the weak solution $\underline{u}(t)$ at such an instant and take into account the self-consistence property of the weak solutions discussed in corollary II, *c.f.r.* (3.2.10).

Remarks:

(1) Note that the derivation just discussed shows that (3.2.31) holds also for $\underline{\gamma}^R_{\underline{k}}(t)$ if $R < \infty$, a property that will be useful below, *c.f.r.* proposition VII below.

(2) One concludes, from the autoregularization property with $\alpha = 2$ and $d = 2$, that initial data in C^∞ evolve into solutions of the Navier–Stokes equation which are of class C^∞ for $t \ge 0$ and one concludes that also data initially in L_2 evolve into solutions that for $t > 0$ are in C^∞. Hence the

analogy with the heat equation and with the Stokes equation (*c.f.r.* §9, (C)) is rather strong if $d = 2$. [4]

A further autoregularization theorem *valid for* $d \ge 2$ *and for all weak solutions*, becomes evident when one ponders the above proof and it is:

V. Proposition *(finite vorticity implies smoothness in dimension* $d = 2, 3$*)* *Suppose that* $t \to \underline{\gamma}_{\underline{k}}(t)$ *verifies for* $t \in [0, T]$ *the (3.2.9) or (3.2.11) with* $d \le 3$ *and initial datum* $\underline{u}^0$ *of class* C^∞*. Furthermore, given* $T > 0$*, suppose that* $\sum_{\underline{k}} |\underline{k}|^2 |\underline{\gamma}_{\underline{k}}(t)|^2 \le S_1$ *for* $0 \le t \le T$*. Then for each* $\alpha \ge 0$ *there exists a constant* C_α *for which it is:* $|\underline{\gamma}_{\underline{k}}(t)| \le C_\alpha |\underline{k}|^{-\alpha}$*: i.e. the* $\underline{u}$ *is in class* C^∞ *in* $\underline{x}$ *(hence it is* C^∞ *also in* $t \ge 0$*). See the following comment (4) for the case in which* $\underline{u} \in L_2(\Omega)$*.*

proof: clearly, by the assumptions, the result holds for $\alpha \le 1$ and $C_1 \equiv S_1^{1/2}$. From the expression of the inertial term $N_{\underline{k}}(\underline{\gamma})$ in (3.2.9) we see that (by replacing, as done often in the preceding pages, $\underset{\sim}{k}_2$ with $\underset{\sim}{k}$, before bounding the left hand side term) (3.2.30) holds:

$$|N_{\underline{k}}(\underline{\gamma})| \le \frac{2}{k_0} S_1 \tag{3.2.32}$$

hence the integral in (3.2.21) is bounded by $S_1/(k_0\, \nu\, \underline{k}^2)$ and we see, from (3.2.9) or (3.2.11), the existence of a constant C_2' for which $|\underline{\gamma}_{\underline{k}}| \le e^{-\nu \underline{k}^2 t} |\underline{\gamma}^0_{\underline{k}}| + C_2' |\underline{k}|^{-2}$. Since $\underline{u}^0 \in C^\infty$ then it is $|\underline{\gamma}_{\underline{k}}| \le C_2 |\underline{k}|^{-2}$ for a conveniently chosen constant C_2. But then, again

$$\begin{aligned} |N_{\underline{k}}(\underline{\gamma})| \le & \sum_{\underline{k}_1 + \underline{k}_2 = \underline{k}} |\gamma_{\underline{k}_1}|\, |\underline{k}_2|\, |\gamma_{\underline{k}_2}| \le \sum_{\underline{k}_1 + \underline{k}_2 = \underline{k}} \frac{|\underline{k}_1|\, |\gamma_{\underline{k}_1}|\, |\underline{k}_2|^2\, |\gamma_{\underline{k}_2}|}{|\underline{k}_1|\, |\underline{k}_2|} \le \\ & \le C_2 S_1^{1/2} \Big(\sum_{\underline{k}_1 + \underline{k}_2 = \underline{k}} \frac{1}{|\underline{k}_1|^2\, |\underline{k}_2|^2} \Big)^{1/2} \le C_2 \sqrt{S_1} \frac{2^{1/4} B}{k_0^{1/4} |\underline{k}|^{1/4}} \end{aligned} \tag{3.2.33}$$

because the last sum is bounded by $2^{1/4}(k_0|\underline{k}|)^{-1/4}$ times the square root of $\big(\sum (|\underline{k}_1||\underline{k}_2|)^{-(2-1/4)}\big)$; and the latter quantity is bounded via Schwartz's inequality by $B = \sum_{\underline{k}} |\underline{k}|^{-(4-1/2)}$; therefore, proceeding as above, we bound again the integral in (3.2.21) and find $|\underline{\gamma}_{\underline{k}}| < C_{2+1/4} \sqrt{S_1} |\underline{k}|^{-(2+1/4)}$. But $2 + 1/4 > d - 1$ if $d \le 3$, and then the preceding general autoregularization theorem applies.

Remarks: The above result is particularly interesting if $d = 3$ but it also holds if $d = 2$.

---

[4] Although we cannot now prove that data in L_2 become analytic for $t > 0$, but only that they become C^∞, however see propositions VIII and IX below.

(1) Hence every weak solution of the Navier–Stokes equation with smooth initial datum and total vorticity bounded in every finite time interval is in class C^∞ for $t \ge 0$; and the maxima of its derivatives can be explicitly bounded in terms of the initial data and of the vorticity estimate S_1. This happens, for instance, if $d = 2$: *c.f.r.* proposition IV.
(2) The difference between the cases $d = 2$ and $d = 3$ is that if $d = 3$ we do not know how to control *a priori* the quantity $\sum_{\underline{k}} |\underline{k}|^2 |\underline{\gamma}_{\underline{k}}(t)|^2$, *i.e.* we do not know how to prove the existence of a quantity S_1 that bounds it; while if $d = 2$ vorticity conservation guarantees such a property. For $d = 2$ this is the contents of proposition IV.
(3) Note that the theorem holds for (3.2.9) and for (3.2.11) as well and *with the same bounds*, independent of R.
(4) *(non smooth case)*: Looking more closely at the above proof we realize that it provides informations on the cases in which $\underline{\gamma}^0$ is just in W^1 (*i.e.* the quantity denoted S_1 in the proof is initially finite): in such cases (taking advantage of the factor $e^{-\nu \underline{k}^2 t}$ that will multiply $\underline{\gamma}^0$ making it rapidly decreasing in $|\underline{k}|$ for $t > 0$) for each $\alpha \ge 0$ the above proof shows that there exists $C_\alpha(t)$ for which we have: $|\underline{\gamma}_{\underline{k}}(t)| \le C_\alpha(t)|\underline{k}|^{-\alpha}$, and $C_\alpha(t)$ can be chosen continuous for $T \ge t > 0$, *i.e.* excluding $t = 0$. *Hence in $d = 2, 3$ weak solutions with bounded vorticity are C^∞ for $t > 0$.*

Concerning uniqueness we can prove the following proposition, which is strengthened in the remarks following it

VI. Proposition *(uniqueness of smooth solutions of NS): If $d \ge 2$ and if $\underline{\gamma}^1_{\underline{k}}$ and $\underline{\gamma}^2_{\underline{k}}$ are two (weak) solutions rapidly decreasing for $|\underline{k}| \to \infty$, in the sense that both are bounded in a given interval $0 \le t \le T$ by $C_\alpha |\underline{k}|^{-\alpha}$ for every $\underline{k}$ and for some $\alpha > d-1$, then the two solutions coincide if they have the same initial datum $\underline{u}^0 \in C^\infty(\Omega)$.*

proof: By the autoregularization theorem we deduce that then $|\underline{\gamma}^i_{\underline{k}}| \le C_\alpha |\underline{k}|^{-\alpha}$ for every $\alpha > 0$ (and a suitable corresponding constant C_α).
Their difference will verify, if we define $\Delta \equiv ||\underline{\gamma}^1 - \underline{\gamma}^2||_2^2$ and $\Delta_1 \equiv ||\,|(\underline{\gamma}^1 - \underline{\gamma}^2)|\,|\underline{k}|\,||_2^2$:

$$\begin{aligned}
\frac{d}{dt}\frac{1}{2}\Delta &\le -\nu\Delta_1 + \sum_{\underline{k}} |\underline{\gamma}^1_{\underline{k}} - \underline{\gamma}^2_{\underline{k}}| \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} |\underset{\sim}{\gamma}{}^1_{\underline{k}_1} \cdot \underset{\sim}{k}_2\, \Pi_{\underline{k}}\, \underline{\gamma}^1_{\underline{k}_2} - \underset{\sim}{\gamma}{}^2_{\underline{k}_1} \cdot \underset{\sim}{k}_2\, \Pi_{\underline{k}}\, \underline{\gamma}^2_{\underline{k}_2}| \\
&\le -\nu\Delta_1 + \sum_{\underline{k}} |\underline{\gamma}^1_{\underline{k}} - \underline{\gamma}^2_{\underline{k}}| \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} |(\underset{\sim}{\gamma}{}^1_{\underline{k}_1} - \underset{\sim}{\gamma}{}^2_{\underline{k}_1}) \cdot \underset{\sim}{k}_2\, \Pi_{\underline{k}}\, \underline{\gamma}^1_{\underline{k}_2}| + \\
&+ \sum_{\underline{k}} |\underline{\gamma}^1_{\underline{k}} - \underline{\gamma}^2_{\underline{k}}| \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} |(\underset{\sim}{\gamma}{}^2_{\underline{k}_1} \cdot \underset{\sim}{k}_2\, \Pi_{\underline{k}}\, (\underline{\gamma}^1_{\underline{k}_2} - \underline{\gamma}^2_{\underline{k}_2})|
\end{aligned} \tag{3.2.34}$$

and if we replace $\underset{\sim}{k}_2$ with $\underline{k}$ and if we use the regularity property $|\underline{\gamma}^i_{\underline{k}}| \le C_\alpha |\underline{k}|^{-\alpha}$ with, for instance, $\alpha = 4$ (note that under the present hypotheses this inequality holds for all $\alpha > 0$) we find:

$$\frac{1}{2}\dot{\Delta} \le -\nu\Delta_1 + 2\sum_{\underline{k}}\sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \big(|\underline{k}|\,|\underline{\gamma}^1_{\underline{k}} - \underline{\gamma}^2_{\underline{k}}|\big)\,\big(|\underline{\gamma}^1_{\underline{k}_1} - \underline{\gamma}^2_{\underline{k}_1}|\big)\,\frac{C_4}{|\underline{k}_2|^4} \le$$
$$\le -\nu\Delta_1 + 2\sum_{\underline{k}}\sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \big(|\underline{k}|^2|\underline{\gamma}^1_{\underline{k}} - \underline{\gamma}^2_{\underline{k}}|^2\frac{\varepsilon}{2} + \frac{1}{2\varepsilon}|\underline{\gamma}^1_{\underline{k}_1} - \underline{\gamma}^2_{\underline{k}_1}|^2\big)\frac{C_4}{\underline{k}_2^4} \tag{3.2.35}$$

having made use of the inequality $|ab| \le a^2/(2\varepsilon) + \varepsilon b^2/2$ (holding for each $\varepsilon > 0$) and $|\underline{k}| \ge k_0$. Hence, if $B = \sum_{\underline{k}} C_4|\underline{k}|^{-4}$:

$$\frac{1}{2}\dot{\Delta}(t) \le -\nu\Delta_1 + \varepsilon B\Delta_1 + B\varepsilon^{-1}\Delta \qquad \text{for every } \varepsilon > 0 \tag{3.2.36}$$

and, by choosing ε so that $\varepsilon B = \nu$, we deduce:

$$\dot{\Delta} \le 2B^2\nu^{-1}\,\Delta \qquad \Rightarrow \qquad \Delta(t) \le \Delta(0)\,e^{2B^2\nu^{-1}t} \tag{3.2.37}$$

so that $\Delta(0) = 0$ implies $\Delta(t) \equiv 0$, for $t > 0$.

Remarks: The theorem can also be proved directly by noting that $\underline{u}^i(t)$ must be C^∞ by proposition V: hence they must coincide by a uniqueness theorem that can be obtained along the lines followed in the case of the Euler equation in problem [3.1.6]. The above proof is interesting because it leads to the remarks that follow and because it suggests the proof of the following proposition VII.

(1) *(Non smooth cases)*: If we suppose that the two weak solutions correspond to an initial datum $\underline{u}^0 \in L_2(\Omega)$ only and *furthermore* are such that $\lim_{t\to0}||\underline{u}^i(t) - \underline{u}^0||_2 = 0$, $i = 1, 2$ then we conclude from the above proof that the two solutions coincide. In fact we shall take as initial time a time $t_0 > 0$ where the solutions have become C^∞ (by proposition III): of course $\Delta(t_0)$ might be $\ne 0$ and to prove uniqueness we still need to check that $\Delta(t_0) \to 0$ as $t_0 \to 0$. This would be implied by $\lim_{t\to0}||\underline{u}^i(t) - \underline{u}^0||_2 = 0$, $i = 1, 2$.

However if $\underline{u}(t)$ is a weak solution with datum L_2 it is not known whether $\underline{u}(t)$ in general tends to $\underline{u}^0$ in L_2 as $t \to 0$!

If instead $\underline{u}^0 \in W^1(\Omega)$ and $d = 2$, from the relation $|\underline{\gamma}_{\underline{k}}(t) - e^{-\underline{k}^2\nu t}\underline{\gamma}^0_{\underline{k}}| \le (G_0\nu^{-1} + 2S_0/k_0)/|\underline{k}|^2$, *c.f.r.* the comment to (3.2.31), we deduce that $\lim_{t\to0}||\underline{u}(t) - \underline{u}^0||_2^2 = 0$ and therefore $\Delta(t_0) \to 0$ as $t_0 \to 0$. Hence *in* $d = 2$ *the weak solutions with datum in* $W^1(\Omega)$ *are unique*. And the only case, in $d = 2$, in which weak solutions might be not unique is when $\underline{u}^0 \in L_2(\Omega)$ but $\underline{u}^0 \notin W^1(\Omega)$: a problem which to my knowledge is unsolved in this case.

(2) The argument used in the above proof allows us to find an explicit estimate of the error that is made on the solutions by truncating the Navier Stokes equation. In fact the following *theorem of spectral approximability* holds:

VII. Proposition *(constructive approximations errors estimate for NS solutions in 2 dimensions): Suppose $d = 2$ and that $\underline{u}^0$ is in class C^∞; calling $\Delta_R(t)$ the square of the L_2–norm of the difference between the solution of the regularized equations with a regularization parameter R (see (3.2.11)) and the solution of class C^∞ discussed in the previous proposition, it follows that $\Delta_R(t)$ can be bounded for every integer $q > 0$ by:*

$$\Delta_R(t) \le V_q e^{Mt}\, R^{-q} \tag{3.2.38}$$

where V_q, M are suitable constants computable in terms of the initial datum but which do not depend on the time t at which the solutions are considered to be evaluated.

Remark: Equation (3.2.38) shows that the truncation method, often also called *spectral method*, provides us with a *truly constructive algorithm* for the solution of the *bidimensional* NS equations. Note that the error estimate is exponentially increasing with the time t: this is a property that, in general, one cannot expect to improve because, as usual in differential equations, data that initially differ by little keep differing more and more as time increases and diverge exponentially, even when they evolve with the same differential equation (furthermore in our case, the truncated equation describing $\underline{\gamma}^R$ is also somewhat different from the not truncated one). This exponential divergence is well established on heuristic and experimental grounds in the case of the NS equations at even moderately large Reynolds number, *c.f.r.* the following Ch. 4,5,6,7.

proof: Suppose for simplicity that there is no external force: $\underline{g} = \underline{0}$.
Equation (3.2.31) holds, as already noted, also for $\underline{\gamma}^R$. Call then C_2 the constant bounding above $|\underline{k}|^2 |\underline{\gamma}_{\underline{k}}|$, and note that (3.2.31) shows that a possible choice is:

$$C_2 = C_2^0 + \nu^{-1} G_0 + 2 S_0 \nu^{-1} k_0^{-1} \tag{3.2.39}$$

where C_p^0 and G_p are upper bounds of $|\underline{k}|^p |\underline{\gamma}^0_{\underline{k}}|$ and of $|\underline{k}|^p |\underline{g}_{\underline{k}}|$, respectively, for each $\underline{k}$.

Let $C_p = B_p(C_2, C_p^0, G_p)$ be the constant, independent of t, such that:

$$|\underline{\gamma}^R_{\underline{k}}|, |\underline{\gamma}_{\underline{k}}| < C_p |\underline{k}|^{-p} \tag{3.2.40}$$

which exists by the autoregularization theorem and is $< \infty$ for all $p \ge 0$; and consider $p \ge 4$.

Then the equation verified by $\underline{\gamma} - \underline{\gamma}^R$ is, for $|\underline{k}| \le R$:

$$\dot{\underline{\gamma}}_{\underline{k}} - \dot{\underline{\gamma}}^R_{\underline{k}} = -\nu \underline{k}^2 (\underline{\gamma}_{\underline{k}} - \underline{\gamma}^R_{\underline{k}}) - i \sum_{\underline{k}_1 + \underline{k}_2 = \underline{k}} \big(\underset{\sim}{\gamma}_{\underline{k}_1} \cdot \underset{\sim}{k}_2\, \Pi_{\underline{k}}\, \underline{\gamma}_{\underline{k}_2} - \underset{\sim}{\gamma}^R_{\underline{k}_1} \cdot \underset{\sim}{k}_2\, \Pi_{\underline{k}}\, \underline{\gamma}^R_{\underline{k}_2} \big) \tag{3.2.41}$$

where $\gamma^R_{\underline{k}_i} \equiv \underline{0}$ if $|\underline{k}|_i > R$: clearly also for such large values of $|\underline{k}_i|$ the $\underline{\gamma}_{\underline{k}_i}$ is in general not zero.

Multiply scalarly both sides by $\overline{\gamma}_{\underline{k}} - \overline{\gamma}^R_{\underline{k}}$ and summing over $\underline{k}$, we find, with the notations and the procedure followed for the (3.2.36), (3.2.37):

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\Delta_R \le & -\nu\Delta^1_R + \varepsilon B\Delta^1_R + \\ & + B\varepsilon^{-1}\Delta_R + \sum_{\underline{k}} \frac{2C_p}{|\underline{k}|^p} \sum_{\substack{|\underline{k}_1| \text{ or } |\underline{k}_2|>R \\ \underline{k}_1+\underline{k}_2=\underline{k}}} |\underline{k}||\gamma_{\underline{k}_1}|\,|\gamma_{\underline{k}_2}| \le \\ \le & B^2\nu^{-1}\Delta_R + \frac{4C_p^3}{(2\pi L^{-1})^{p-1}}\Big(\sum_{|\underline{k}_1|>R}\frac{1}{|\underline{k}_1|^p}\Big)\Big(\sum_{\underline{k}_2}\frac{1}{|\underline{k}_2|^p}\Big) \le \\ \le & B^2\nu^{-1}\Delta_R + \frac{4C_p^3K_p}{(2\pi L^{-1})^{p-1}R^{p-2}}\end{aligned} \tag{3.2.42}$$

where $\Delta_R(t) = \sum_{\underline{k}} |\gamma_{\underline{k}} - \gamma^R_{\underline{k}}|^2$ and $\Delta^1_R = \sum_{\underline{k}} |\underline{k}|^2\,|\gamma_{\underline{k}} - \gamma^R_{\underline{k}}|^2$; and K_p estimates the quantity $(\sum_{\underline{k}} |\underline{k}|^{-p})^2$. Hence:

$$\begin{aligned}\Delta_R(t) \le & e^{2B^2\nu^{-1}t}\Delta_R(0) + \int_0^t e^{2B^2\nu^{-1}(t-\tau)}\frac{8C_p^3K_p}{(2\pi L^{-1})^{p-1}R^{p-2}}\,d\tau \le \\ \le & e^{2B^2\nu^{-1}t}\Delta_R(0) + \frac{8C_p^3K_p}{2(2\pi L^{-1})^{p-1}B^2\nu^{-1}}\frac{1}{R^{p-2}}\end{aligned} \tag{3.2.43}$$

recalling that $B \equiv C_4 \sum_{\underline{k}} |\underline{k}|^{-4}$ (see (3.2.36).

But $\Delta_R(0)$ is small for large R: in fact the $\Delta_R(0) \equiv \sum_{|\underline{k}|\ge R} |\gamma_{\underline{k}}(0)|^2$ is bounded above by ${C^0_r}^2 \sum_{|\underline{k}|\ge R} |\underline{k}|^{-2r}$, if $|\gamma_{\underline{k}}(0)|\,|\underline{k}|^r \le C^0_r$ for all $r \ge 0$; and therefore $\Delta_R(0) \le {C^0_r}^2 R^{-2r+2}B'_r$ for a suitably chosen B'_r and:

$$\Delta_R(t) \le e^{2B^2\nu^{-1}t}\Big(\frac{B'_r}{R^{2r-2}} + \frac{B"_p}{R^{p-2}}\Big) \tag{3.2.44}$$

with a suitable $B"_p$. This yields, in particular, an explicit estimate of the difference between $\gamma_{\underline{k}}(t)$ and $\gamma^R_{\underline{k}}(t)$ at $\underline{k}$ fixed, $|\underline{k}| \le R$, and the (3.2.38) follows by choosing the arbitrary parameter r so that $2r - 2 = p - 2$, *i.e.* $r = p/2$.

We conclude by showing that if $d = 2$ the weak solutions with mildly regular initial data become analytic at positive time

Proposition VIII *(an analytic regularity result for NS in $d = 2$, (Mattingly, Sinai)): If $d = 2$ the (unique) solution of the NS equations with initial datum $\underline{u}^0$ and forcing $\underline{g}$ such that $|\gamma_{\underline{k}}(0)| < U|\underline{k}|^{-3}, |\underline{g}_{\underline{k}}| < Fe^{-\kappa|\underline{k}|}$ with $U, F, \kappa > 0$, is analytic for $t \in (0, T]$ with T arbitrarily fixed.*

Proof: We follow, and implement in our context, the idea in [MS99]. Consider the regularized equations with cut-off R; by propositions III, IV *for* *all times* it is: $|\gamma^R_{\underline{k}}(t)| \le \frac{C(t)}{|\underline{k}|^3}$ for a non decreasing, finite, function $C(t) > 0$ depending on U but not on R.

Hence, by choosing ε sufficiently small, for $i, j = 1, 2$ (i is the component index, j distinguishes real or imaginary part) we have[5]

$$|\gamma^{(i,j)R}_{\underline{k}}(t)| < 2 \cdot e^{-|\underline{k}|\varepsilon t} \frac{C(T)}{|\underline{k}|^3} \tag{3.2.45}$$

for $|\underline{k}| \le K_0 \equiv 8\nu^{-1} 2^3 C(T)\, 2B$ and $t \in [0, T]$, with $B = \sum_{|\underline{k}|\neq 0} |\underline{k}|^{-3}$. This means that we can take $\varepsilon = (K_0 T)^{-1} \log(U/2C(T))$.

We take, *just for simplicity*, $\underline{g} = \underline{0}$. Then assuming the existence of a time t when the (3.2.45) is violated for some $\overline{k} > K_0$ we define $\overline{t}_R$, with $0 < \overline{t}_R < T$, so that at $t = \overline{t}_R$, "for the first time", $\gamma^{(i,j)R}_{\overline{\underline{k}}}(\overline{t}) = \pm 2 \cdot e^{-|\overline{\underline{k}}|\varepsilon \overline{t}} C(T)\, |\overline{\underline{k}}|^{-3}$ for at least one $\underline{k} = \overline{\underline{k}}$ ($|\underline{k}| > K_0$) and $i, j = 1$ or 2; in the + case, for instance, $\dot\gamma^{(i,j)R}_{\underline{k}}$ will be bounded by:

$$\begin{aligned}\dot\gamma^{(i,j)R}_{\overline{\underline{k}}}(\overline{t}) &\le -\nu |2\overline{\underline{k}}|^2 \gamma^{(i,j)R}_{\overline{\underline{k}}} + |\overline{\underline{k}}| \sum_{\underline{k}_1 + \underline{k}_2 = \overline{\underline{k}}} |\gamma^R_{\underline{k}_1}| \cdot |\gamma^R_{\underline{k}_2}| \le \\ &\le -\nu |2\overline{\underline{k}}|^2 2 \cdot e^{-|\overline{\underline{k}}|\varepsilon \overline{t}} \frac{C(T)}{|\overline{\underline{k}}|^3} + |\overline{\underline{k}}|\, e^{-|\underline{k}_1|\varepsilon \overline{t}} \cdot e^{-|\underline{k}_2|\varepsilon \overline{t}} \cdot \frac{2C(T)}{(|\underline{k}|/2)^3} C(T) \cdot 4\, B\end{aligned} \tag{3.2.46}$$

because if $\underline{k}_1 + \underline{k}_2 = \overline{\underline{k}}$ then either $|\underline{k}_1| > |\overline{\underline{k}}|/2$ or $|\underline{k}_2| > |\overline{\underline{k}}|/2$.

Therefore the first term in the r.h.s., since $|\overline{\underline{k}}| \le |\underline{k}_1| + |\underline{k}_2|$, will be a quantity which for $|\overline{\underline{k}}| > K_0$ is less than $-\frac{1}{2}\, |\overline{\underline{k}}|\, \nu\, |\gamma^{(i,j)R}_{\overline{\underline{k}}}| < 0$, because K_0 was defined just to make this true.

Hence the derivative of $\gamma^{(i,j)R}_{\underline{k}}$ is less than the "speed of contraction of the boundary" of the region where (3.2.45) is satisfied for all $|\underline{k}| < R$. Therefore for every R, $\overline{t}_R = T$. Since the regularized solution converges *for every* $\underline{k}$, as $R \to \infty$, to the solution of NS equations, the latter solution verifies (3.2.45) *for all* $\underline{k}$. Hence it is analytic.

Remark: (Non smooth case) Hence, if $d = 2$, the weak solutions with data in $L_2(\Omega)$ also become analytic at positive time because by proposition IV they become immediately C^∞.

Of course the size ε of the "strip of analyticity" depends on the prefixed T and tends to 0 as $T \to \infty$: this is quite different with respect to the "analyticity improving" nature of the heat equation or of the Stokes equation. Then one can ask if there are weaker regularity properties that the solution acquires for positive time, say for $t > t_0 > 0$, and that do not depend on T.

---

[5] The choice of the factor 2 is arbitrary: a constant > 1 would be equally suitable for our purposes.

A (quite weak) regularization result of this type can be obtained independently of the proposition VIII and it follows immediately from the autoregularization estimates

Proposition IX *(a C^∞–regularity estimate for NS in $d = 2$): Let $\underline{u}^0 \in L_2$, $d = 2$, and let $\underline{u}(t)$ a corresponding C-weak*[6] *solution of the NS equation. There exist two functions $H(t), h(t)$ finite and non increasing for $t > 0$, divergent as $t \to 0$, such that*

$$|\gamma_{\underline{k}}(t)| \le H(t_0) \cdot e^{-(\log|\underline{k}|)^2/h(t_0)}, \qquad \text{for all} \quad t \ge t_0 > 0 \qquad (3.2.47)$$

Proof: (sketch) In fact the constant C_α in proposition III can be estimated (by (3.2.21), and we leave this as a problem) as $C_{\alpha+1} < \alpha! B' + 2^\alpha B'' C_\alpha$ for some constants B', B'' and $\alpha > 0$. This immediately gives: $C_\alpha \le \alpha!\, 2^{\alpha^2} D$ for some $D > 0$, and consequently (as in deriving (3.2.23)) the estimate

$$|\gamma_{\underline{k}}(t)| \le H(t_0) \cdot e^{-(\log|\underline{k}|)^2\, h(t_0)^{-1}}, \qquad \text{for all} \quad t \ge t_0 > 0 \qquad (3.2.48)$$

where $H(t_0), h(t_0) < +\infty$ if $t_0 > 0$ are constants depending on the initial conditions and on $\nu, \underline{g}$, *but not on time $t > t_0$*.

Noting that $\exp -(\log|\underline{k}|)^2/h$ tends to 0 as $|\underline{k}| \to \infty$ faster than any power, this means that the C-weak solution condidered here (unique at least if $\underline{u}^0 \in W^1(\Omega)$ by proposition VI) has a Fourier transofrm that acquires a C^∞ regularity expressed quantitatively by the parameters H, h in (3.2.47) at any time $t_0 > 0$ and keeps it forever *with the same parameters* (at least).

Problems

[3.2.1]: Estimate the function B_α, described in remark (1) following the proof of proposition III, in the cases $d = 2, \alpha_0 = 2$ and $d = 3, \alpha_0 = 2$ for $\alpha = \alpha_0 + 1, \alpha_0 + 2$.

[3.2.2]: Estimate the constants B_α in (3.2.20).

[3.2.3]: Prove proposition VII without assuming the absence of the external force; determine the $\underline{g}$–dependence of the constants V_q, M.

Bibliography: This section is based on my lecture notes (see [Bo79]) on the work [FP67], which discusses the above matters in a domain with boundary

[6] Recall that we have proved uniqueness of the weak solutions only under the slightly stronger condition $\underline{u}^0 \in W^1(\Omega)$.

and no slip boundary conditions. Proposition VIII is, however, a recent result in [MS99].

§3.3 Regularity: partial results for the NS equation in $d = 3$. The theory of Leray.

The theory of §3.2 is very unsatisfactory in the 3–dimensional case, because it is nonconstructive and lacks a uniqueness theorem.

The consequent ambiguity *only* consists in having been found as a limit of an unspecified subsequence extracted from a family of approximate solutions and therefore one can fear that by selecting different subsequences one obtains different solutions. It also consists in the *possibility* that there are other regularizations which, through a limiting procedure, could lead to yet other solutions.

We shall consider a cubic container Ω with side L and periodic boundary conditions, and we shall fix units so that the density is $\rho = 1$. Therefore there are two scales intrinsically associated with the system: a time scale T_c and a velocity scale V_c:

$$V_c \equiv \frac{\nu}{L}, \qquad T_c \equiv \frac{L^2}{\nu} \tag{3.3.1}$$

where ν is the kinematic viscosity. It will be natural to call them together with the length scale L, the scales "characteristic of the geometry of the system"

(A) Leray's regularization.:

An interesting regularization, different from the one by *cut–off* used so far is *Leray's regularization.*

Let $\underline{x} \to \chi(\underline{x}) \ge 0$ be a C^∞ function defined on Ω and not vanishing in a small neighborhood of the origin and with integral $\int \chi(\underline{x})\, d\underline{x} \equiv 1$: the function $\chi(\underline{x})$ can be regarded as a periodic function on Ω or as a function on R^3 with value 0 outside Ω, as we shall imagine that Ω is centered at the origin, to fix the ideas. For $\lambda \ge 1$ also the function $\chi_\lambda(\underline{x}) \stackrel{def}{=} \lambda^3 \chi(\lambda \underline{x})$ can be regarded as a periodic function on Ω or as a function on R^3: it is an "approximate Dirac's δ–function".

It will be relevant to point out that there is a simple relation between the Fourier transforms of the function χ regarded as defined on Ω and its Fourier transform when it is regarded as a function on R^3: namely if $\hat\chi(\underline{k})$, $\underline{k} \in R^3$ is the Fourier transform of χ as defined on R^3, then the Fourier tranform of χ regarded as a function on Ω is simply $\hat\chi(\underline{k})$ evaluated at $\underline{k} = \underline{n}\, k_0$ with $k_0 = 2\pi L^{-1}$ and $\underline{n}$ is an integer components vector (just write the definitions of the Fourier transforms to check this statement). Note that $\hat\chi(\underline{k})$ decreases faster than any power as $|\underline{k}| \to \infty$.

We shall examine the regularization with (dimensionless) parameter λ of the NS equation "*in the sense of Leray*" defined as

$$\dot{\underline{u}} = \nu\Delta\underline{u} - \langle\underline{u}\rangle_\lambda \cdot \underset{\sim}{\underline{\partial}}\, \underline{u} - \underline{\partial}p + \underline{g}, \qquad \underline{\partial}\cdot\underline{u} = 0, \qquad \int_\Omega \underline{u}\, d\underline{x} = \underline{0} \qquad (3.3.2)$$

where $\langle\underline{u}\rangle_\lambda \equiv \int_{R^3} \chi_\lambda(\underline{y})\, \underline{u}(\underline{x}+\underline{y})\, d^3\underline{y}$, and $\underline{u}, \underline{g}$ are divergenceless fields. The volume force $\underline{g}$ will be assumed time indpendent (for simplicity).

In (3.3.2) we see that the approximation corresponding to the regularization consists in having the velocity field at a point $\underline{x}$ no longer "transported", as it should, by the velocity field itself as in the Euler and NS equations, but *rather by the average of the velocity* on a region of size of the order of λ^{-1} around $\underline{x}$.

Rewrite (3.3.2) as an equation for the Fourier components $\underline{\gamma}_{\underline{k}} \equiv \overline{\underline{\gamma}}_{-\underline{k}}$ of the velocity field (*c.f.r.* §2.2)

$$\dot{\underline{\gamma}}^\lambda_{\underline{k}} = -\nu\, \underline{k}^2\, \underline{\gamma}^\lambda_{\underline{k}} - i \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \hat{\chi}(\underline{k}_1\lambda^{-1}) \underset{\sim}{\gamma}^\lambda_{\underline{k}_1} \cdot \underset{\sim}{k}_2\, \Pi_{\underline{k}}\, \underline{\gamma}^\lambda_{\underline{k}_2} + \underline{g}_{\underline{k}} \qquad (3.3.3)$$

where $\hat{\chi}(0) = 1$ e $\underline{k}\cdot\underline{\gamma}^\lambda_{\underline{k}} = 0$, $\underline{\gamma}^\lambda_{\underline{0}} \equiv \underline{g}_{\underline{0}} \equiv \underline{0}$ and the functions $\underline{\gamma}_{\underline{k}}(0) \equiv \underline{\gamma}^0_{\underline{k}}$ e $\underline{g}_{\underline{k}}$ are assigned as Fourier transforms (with the conventions fixed in (2.2.2))

$$\underline{\gamma}^0_{\underline{k}} = L^{-3} \int_\Omega e^{-i\underline{k}\cdot\underline{x}}\, \underline{u}^0(\underline{x}) d\,\underline{x}, \qquad \underline{g}_{\underline{k}} = L^{-3} \int_\Omega e^{-i\underline{k}\cdot\underline{x}}\, \underline{g}(\underline{x}) d\,\underline{x} \qquad (3.3.4)$$

of a datum $\underline{u}^0 \in C^\infty(\Omega)$ and of the intensity of external force $\underline{g} \in C^\infty(\Omega)$. We shall use the convention on the Fourier transform so that $\underline{u}^0 = \sum_{\underline{k}} \underline{\gamma}^0_{\underline{k}} e^{i\underline{k}\cdot\underline{x}}$, hence $\sum_{\underline{k}} |\underline{\gamma}^0_{\underline{k}}|^2 = L^{-3} \int |u^0(\underline{x})|^2\, d\underline{x}$. Both $\underline{u}$ and its Fourier transform $\underline{\gamma}$ have the dimension of a velocity.

Remarks:
(1) The (3.3.3) is similar to the truncation regularizations considered in the previous section, because the large values of $\underline{k}_1$, $|\underline{k}_1| \gg \lambda$, are "*suppressed*" in (3.3.3). Nevertheless (3.3.3) *does not* reduce to the regularization by truncation, not even in the case in which $\hat{\chi}(\underline{k}_1/\lambda)$ is chosen as a characteristic function of the set of $\underline{k}$'s with $|\underline{k}| < \lambda L^{-1}$: indeed with such a choice of $\hat{\chi}$ the vectors $\underline{k}_2$ and $\underline{k}$ in (3.3.3) remain free and only $|\underline{k}_1|$ is forced to be $\le \lambda$. In the equations obtained by truncation (for the NS equation) discussed in the preceding section, instead, also $|\underline{k}_2|$ and $|\underline{k}|$ are constrained to be $\le \lambda$ (λL^{-1} correspond to R of the preceding section).
(2) We have not, however, chosen χ with a $\hat{\chi}$ being a characteristic function of a sphere not just because the corresponding $\chi(\underline{x})$ would have a so slow decrease at ∞ to make improper the integral defining the $\langle\underline{u}\rangle_\lambda(\underline{x})$; but also, and mainly, because the following inequality (fundamental to the theory) would not be generally true

$$|\langle\underline{u}\rangle_\lambda(\underline{x})| \le \max_{\underline{y}\in\Omega} |\underline{u}(\underline{y})| \qquad (3.3.5)$$

With our choice of χ as an approximate δ–function the (3.3.5) is correct and simple to check: *but* it is based upon the positivity of the function $\chi(\underline{y})$ and on the fact that its integral over the whole space is equal to 1.

(B) Properties of the regularized equation and new weak solutions.

The theory of (3.3.3) at λ fixed is very simple. The key is that if $||\underline{\gamma}||_2 < \infty$ then, for each $\lambda > 1$ and each sequence $\{\underline{\gamma}_{\underline{k}}\}$ an identity, that we have already seen to be the root of the energy conservation relation, remains true in the form

$$\sum_{\substack{\underline{k}_1+\underline{k}_2+\underline{k}_3=0\\ |\underline{k}_i|\le N}} \hat{\chi}\big(\underline{k}_1\lambda^{-1}\big)\,\underset{\sim}{\gamma}_{\underline{k}_1}\cdot\underset{\sim}{k}_2\,\underline{\gamma}_{\underline{k}_2}\cdot\underline{\gamma}_{\underline{k}_3}=0 \tag{3.3.6}$$

for all integers N, because it is based only on the symmetry of this expression with respect to $\underline{k}_2$ and $\underline{k}_3$ and to the orthogonality between $\underset{\sim}{k}_2+\underset{\sim}{k}_3\equiv-\underset{\sim}{k}_1$ and $\underset{\sim}{\gamma}_{\underline{k}_1}$.

This means that one can envisage the same method of §3.2 to solve the equation (3.3.3) as a limit of solutions $\gamma^{\lambda,N}_{\underline{k}}(t)$ of the cut–off equations (here N is a cut–off parameter whose role is completely different from that of λ as the latter is necessary to make sense of the equations while N is introduced here as a technical tool to establish properties of the regularize equations and it will soon disappear) with $|\underline{k}|\le N$

$$\begin{aligned}\underline{\gamma}^{\lambda,N}_{\underline{k}}(t)=e^{-\nu\underline{k}^2t}\underline{\gamma}^0_{\underline{k}}+\int_0^t e^{-\nu\underline{k}^2(t-\tau)}\Big(\underline{g}_{\underline{k}}-i\sum_{\substack{\underline{k}_1+\underline{k}_2=\underline{k}\\ |\underline{k}_i|\le N}}\hat{\chi}\big(\lambda^{-1}\underline{k}_1\big)\cdot\\ \cdot\,\underset{\sim}{\gamma}^{\lambda,N}_{\underline{k}_1}(\tau)\cdot\underset{\sim}{k}_2\,\Pi_{\underline{k}}\gamma^{\lambda,N}_{\underline{k}_2}(\tau)\Big)\,d\tau\end{aligned} \tag{3.3.7}$$

In fact one proceeds as in §3.2 and the approach yields a solution verifying the "same" *a priori* estimates, discussed in (3.2.12) and (3.2.13)

$$\begin{aligned}&||\underset{\sim}{\gamma}^{\lambda,N}(t)||_2^2\le E_0L^{-3},\\ &\sum_{\underline{k}}\int_0^t d\tau\underline{k}^2|\gamma^{\lambda,N}_{\underline{k}}(\tau)|^2\le\frac{1}{2}\,E_0L^{-3}\,\nu^{-1}+t\sqrt{E_0L^{-3}}\,\nu^{-1}\ ||\underline{g}||_2\end{aligned} \tag{3.3.8}$$

where $L^3||\underset{\sim}{\gamma}^{\lambda,N}(t)||_2^2=\int|\underline{u}^{\lambda,N}(t)|^2\,d\underline{x}$ and the regularization parameter N plays the same role of R in §3.2 and λ is, for the time being, kept fixed.

By the argument of proposition I of §3.2 the equation (3.3.8), in turn, implies existence of a limit as $N\to\infty$, $\gamma^{\lambda}_{\underline{k}}(t)$, possibly obtained on a subsequence $N_j\to\infty$, verifying the (3.3.7) and (3.3.3).

This time however the autoregolarization is stronger (because the (3.3.3) *is not* the NS equation). Indeed if $|\underline{k}|^\alpha |\underline{\gamma}^{\lambda,N}_{\underline{k}}(t)| \le C_\alpha$ for all $\underline{k}$ and $|t| \le T$, then $\forall\ |t| \le T$:

$$|\underline{\gamma}^{\lambda,N}_{\underline{k}}(t)| \le e^{-\nu \underline{k}^2 t} |\underline{\gamma}^0_{\underline{k}}| + \frac{|\underline{g}_{\underline{k}}|}{\nu \underline{k}^2} + \int_0^t d\tau\, |\underline{k}| e^{-\nu \underline{k}^2 (t-\tau)}. \tag{3.3.9}$$

$$\cdot \sum_{\underline{k}_1 + \underline{k}_2 = \underline{k}} \hat{\chi}(\lambda^{-1} \underline{k}_1) \frac{C_\alpha^2}{|\underline{k}_1|^\alpha |\underline{k}_2|^\alpha} \le e^{-\nu \underline{k}^2 t} |\underline{\gamma}^0_{\underline{k}}| + \frac{|\underline{g}_{\underline{k}}|}{\nu \underline{k}^2} + \frac{C_\alpha^2 B(\lambda)}{\nu |\underline{k}|^{\alpha+1}}$$

where $B(\lambda) = 2^{1+\alpha} k_0^{-\alpha} \sum_{\underline{k}} |\underline{k}|^{2\alpha} \hat{\chi}(\underline{k}\lambda^{-1}) < \infty$, if $k_0 = 2\pi L^{-1}$ is the minimum value of $|\underline{k}|$ (note that, obviously, $B(\lambda) \xrightarrow[\lambda\to\infty]{} \infty$).

Hence if $\underline{u}^0 \in C^\infty$, $\underline{g} \in C^\infty$ we see that there is $C_{\alpha+1} < \infty$ such that $|\underline{\gamma}^{\lambda,N}_{\underline{k}}(t)| \le C_{\alpha+1} |\underline{k}|^{-\alpha-1}$, $\forall\ 0 \le t \le T$. Hence repeating the argument we see that $|\gamma^{\lambda,N}_{\underline{k}}(t)|\ |\underline{k}|^{\alpha'} \le C_{\alpha'}$, $\forall\ \alpha' \ge \alpha$, $t \le T$.

This means that all limits of convergent subsequences as $N \to \infty$ of $\underline{\gamma}^{\lambda,N}$ are C^∞, because the (3.3.8) trivially guarantees the validity of this estimate for $\alpha = 0$. And we find immediately, as in the case $d = 2$, that such solution of (3.3.3) is unique and therefore $\underline{\gamma}^\lambda = \lim_{N\to\infty} \underline{\gamma}^{\lambda,N}$, without need of considering subsequences (*c.f.r.* (3.2.37)).

But in trying to perform also the limit $\lambda \to \infty$ one risks loss of regularity obtaining only a weak solution, in the same sense of §3.2, as a limit on a suitable subsequence $\lambda_j \to \infty$.

The latter weak solution will still verify, by the same argument used in §3.2, the *a priori* estimates (3.3.8) or (3.2.12), with the same right hand side members, and the NS equation

$$\begin{aligned}\underline{\gamma}_{\underline{k}}(t) =& e^{-\nu \underline{k}^2 (t-t_0)} \underline{\gamma}(t_0)_{\underline{k}} + \int_{t_0}^{t_0+t} e^{-\nu \underline{k}^2 (t-\tau)} \Big(\underline{g}_{\underline{k}} - i \sum_{\underline{k}_1 + \underline{k}_2 = \underline{k}} \\ & \underset{\sim}{\gamma}_{\underline{k}_1}(\tau) \cdot \underset{\sim}{k}_2\, \Pi_{\underline{k}}\, \underline{\gamma}_{\underline{k}_2}(\tau) \Big)\, d\tau\end{aligned} \tag{3.3.10}$$

where $t_0 \ge 0$, *c.f.r.* (3.2.10).

Remark: Not having established a uniqueness theorem for weak solutions it is not necessarily true that such solutions coincide with the ones discussed in §3.2. Hence it should not be a surprise that *it will be possible* to prove that such new solutions enjoy properties that we *would not know* how to get for the others.

(C) The local bounds of Leray. Uniformity in the regularization parameter.

Let $t_0 \ge 0$ be an arbitrary time; let

$$J_0(t_0) = \sum_{\underline{k}} |\underline{\gamma}^\lambda_{\underline{k}}(t_0)|^2\ , \quad J_1(t_0) = L^2 \sum_{\underline{k}} |\underline{k}|^2 |\underline{\gamma}^\lambda_{\underline{k}}(t_0)|^2\ , \quad G_0 = \sum_{\underline{k}} |\underline{g}_{\underline{k}}|^2 \tag{3.3.11}$$

and introduce

$$D_m(t) = \sup_{\underline{x},|\underline{\alpha}|=m} |L^m \partial_{\underline{x}}^{\underline{\alpha}} \underline{u}^\lambda(\underline{x},t)|, \qquad J_m(t) = L^{2m} \sum_{\underline{k}} |\underline{k}|^{2m} |\underline{\gamma}_{\underline{k}}^\lambda(t)|^2 \tag{3.3.12}$$

noting that the quantities $D.$ have dimension of a velocity while $J.$ of velocity squares and $\sqrt{G_0}$ is an acceleration.

It is useful to define, for the purpose of a clearer formulation of Leray's theory, the *Reynolds' number*. Given the importance of this notion we give a formal definition

1. Definition *(Reynolds number): If $\underline{u}$ is a velocity field in Ω we define the "Reynolds number" R and the "dimensionless strength" R_g of the external force in terms of "geometric" scales of velocity and time, c.f.r. (3.3.1), and of the velocities $V_1 \equiv \sqrt{J_1}$ and $W_0 = T_c\sqrt{G_0}$ which we shall call, respectively, the "velocity variation scale" of the field $\underline{u}$ and the "external force" or "free fall" velocity scale. The definitions are*

$$R \equiv \frac{V_1}{V_c}, \qquad R_g = \frac{W_0}{V_c} \tag{3.3.13}$$

The quantities V_1, J_1, R are properties of the velocity field $\underline{u}$, while R_g is instead a function of the density of external force.

In particular R depends on time t if $\underline{u}$ does; and $R < \infty$ is tantamount to saying $\underline{u} \in W^1(\Omega)$. We show that

I. Proposition *(regularization independent a priori bounds): Let $\underline{u}^0, \underline{g} \in C^\infty(\Omega)$. There exist λ–independent dimensionless constants $F < 1$, F_m, $m = 0, 1, 2\ldots$ such that if $t \to \underline{u}^\lambda(t)$ is a solution of the regularized Navier Stokes equations (3.3.2) which at a time $t_0 \ge 0$ has Reynolds number $R(t_0) < \infty$ then*

$$\begin{aligned} D_m(t) \le& \Big(\frac{V_c R(t_0)}{\sqrt[4]{L^{-2}\nu(t-t_0)}} + W_0\Big) \frac{F_m}{(L^{-2}\nu\,(t-t_0))^{\frac{m}{2}}}, \\ R(t)^2 \le& 8(R(t_0)^2 + R_g^2) \end{aligned} \tag{3.3.14}$$

for all $t \in [t_0, t_0 + T_0]$, with T_0 given by

$$T_0 = F\,\frac{T_c}{R(t_0)^4 + R_g^2 + 1} \tag{3.3.15}$$

with R_g defined in (3.3.13).

Remark: Note that while the velocities D_m depend on λ their bounds, together with the corresponding time of validity, *do not depend on* λ. They are therefore bounded in terms of the Reynolds number at the instant t_0 only. But in general the latter depends on λ, *except when* $t_0 = 0$.

proof: Rewrite (3.3.7) in "position space", *i.e.* for $\underline{u}^\lambda(\underline{x},t)$. We find

$$\begin{aligned} u^\lambda_j(\underline{x},t) = &\int_\Omega d\underline{y}\,\Gamma(\underline{x}-\underline{y},t-t_0)\,u^\lambda_j(\underline{y},t_0)+\\ &+\int_{t_0}^t d\tau\int_\Omega \Gamma(\underline{x}-\underline{y},\tau-t_0)\,g_j(\underline{y},\tau)\,d\underline{y}+\\ &-\sum_{p,h=1}^3\int_{t_0}^t d\tau\int_\Omega d\underline{y}\,\partial_p T_{jh}(\underline{x}-\underline{y},t-\tau)\,\langle u^\lambda_p(\underline{y},\tau)\rangle_\lambda\,u^\lambda_h(\underline{y},\tau)\end{aligned} \tag{3.3.16}$$

where we have set

$$\begin{aligned}\Gamma(\underline{x},t) =& L^{-3}\sum_{\underline{k}} e^{-\nu\underline{k}^2t}e^{i\underline{k}\cdot\underline{x}},\\ T(\underline{x},t)_{ij} =& L^{-3}\sum_{\underline{k}\ne\underline{0}} e^{-\nu\underline{k}^2t}\big(\delta_{ij}-\frac{k_ik_j}{|\underline{k}|^2}\big)e^{i\underline{k}\cdot\underline{x}}\end{aligned} \tag{3.3.17}$$

which are Green's functions for the heat equation on the torus Ω of side L.
The following properties of Γ, T for $t > 0$ or for $t = 0$ will be used

$$\begin{aligned}(\partial_t-\nu\Delta)\Gamma(\underline{x},t)=0,\qquad &\Gamma(\underline{x},0)=\delta(\underline{x})\\ (\partial_t-\nu\Delta)T(\underline{x},t)=0,\qquad &T_{jh}(\underline{x},0)=\delta(\underline{x})\delta_{jh}-\partial_j\partial_hG(\underline{x})\end{aligned} \tag{3.3.18}$$

here $\delta(\underline{x})$ is Dirac's delta and $G(\underline{x})$ is the Green's function of the Laplace operator on the torus Ω. The latter can be expressed, via the images method (*c.f.r.* problems [2.3.12]÷[2.3.14]) and [3.1.12], as

$$\begin{aligned}G(\underline{x}) = &-\frac{1}{4\pi}\sum_{\underline{n}}\big\{\frac{1}{|\underline{x}+\underline{n}L|}-\frac{1}{|\underline{n}L|}+\frac{\underline{x}\cdot\underline{n}L}{|\underline{n}L|^3}+\\ &+\frac12\frac{1}{|\underline{n}L|^3}\big(\underline{x}^2-3\frac{(\underline{x}\cdot\underline{n}L)^2}{|\underline{n}L|^2}\big)\big\}=\\ &\equiv\lim_{N\to\infty}-\frac{1}{4\pi}\sum_{|\underline{n}|\le N}\big(\frac{1}{|\underline{x}+\underline{n}L|}-\frac{1}{|\underline{n}L|}\big),\end{aligned} \tag{3.3.19}$$

where $\underline{n} = (n_1, n_2, n_3)$ is an integer components vector.[1]
The Γ, T can also be computed by the method of images (from (3.3.18))

$$\begin{aligned}\Gamma(\underline{x},t) = &\sum_{\underline{n}}\frac{e^{-(\underline{x}+\underline{n}L)^2/4\nu t}}{(4\pi\nu t)^{3/2}}\\ T_{jn}(\underline{x},t) = &\Gamma(\underline{x},t)\delta_{jn}+\partial_j\partial_n\int_\Omega G(\underline{x}-\underline{y})\Gamma(\underline{y},t)\,d\underline{y}\end{aligned} \tag{3.3.20}$$

[1] Note that the first formula in (3.3.19) implies the second because the first is absolutely convergent and therefore can be trivially obtained as the limit for $N \to \infty$ of the sum over $|\underline{n}| \le N$; but if the sum over $\underline{n}$ is restricted to the $\underline{n}$ such that $|\underline{n}| \le N$ then the linear terms in $\underline{x}$ and the quadratic ones as well add up to zero (for every N).

Finally Γ, T verify the following basic inequalities, for $\nu t \le L^2$

$$0 \le \Gamma(\underline{x},t) \le \frac{C_0}{(\underline{x}^2+\nu t)^{3/2}} \qquad \int_\Omega \Gamma(\underline{x},t)\,d\underline{x} \equiv 1 \tag{3.3.21}$$

for a suitable C_0, and if $\underline{\alpha} = (\alpha_1,\alpha_2,\alpha_3)$, with $\alpha_j \ge 0$ integer, and $|\underline{\alpha}| \stackrel{def}{=} \alpha_1+\alpha_2+\alpha_3$, then

$$|\partial^{\underline{\alpha}}_{\underline{x}}\Gamma(\underline{x},t)|\ ,\ |\partial^{\underline{\alpha}}_{\underline{x}}T_{ij}(\underline{x},t)| \le \begin{cases} C_{|\underline{\alpha}|}\,(\underline{x}^2+\nu t)^{-(3+|\underline{\alpha}|)/2} & \text{for } \nu t \le L^2\\ C_{|\underline{\alpha}|}L^{-(3+|\underline{\alpha}|)}\,(L^2/\nu t)^{|\underline{\alpha}|/2} & \text{for } \nu t > L^2 \end{cases} \tag{3.3.22}$$

See problems [3.3.6] and [3.3.9] for a check of the inequalities.

If the derivatives $\partial^{\underline{\alpha}}\Gamma$ and $\partial^{\underline{\alpha}}T$ are considered as convolution operators on $L_2(\Omega)$ which are, therefore, defined by

$$\begin{aligned}(\partial^{\underline{\alpha}}\Gamma * f)(\underline{x}) &= \int_\Omega \partial^{\underline{\alpha}}\Gamma(\underline{x}-\underline{y},t)f(\underline{y})\,d\underline{y}\\ (\partial^{\underline{\alpha}}\underset{\sim}{\underline{T}} * \underline{f})(\underline{x}) &= \int_\Omega \partial^{\underline{\alpha}}\underset{\sim}{\underline{T}}(\underline{x}-\underline{y},t)\underline{f}(\underline{y})\,d\underline{y}\end{aligned} \tag{3.3.23}$$

we bound immediately their sizes in $L_2(\Omega)$ by evaluating, via (3.3.17), the "norms" in $L_2(\Omega)$ via the Fourier transforms and we get, if $|\underline{\alpha}| = m$

$$||\partial^{\underline{\alpha}}\Gamma * \cdot||_2\ ,\ ||\partial^{\underline{\alpha}}T * \cdot||_2 \le \sup_{\underline{k}} |\underline{k}|^{|\underline{\alpha}|}e^{-\nu\underline{k}^2 t} \le \frac{B_m}{(\nu t)^{m/2}} \tag{3.3.24}$$

if B_m are suitable dimensionless constants.

Returning to (3.3.17) we recall that, besides the characteristic time $T_c = L^2/\nu$ and the characteristic velocity $V_c = \nu/L$, *c.f.r.* (3.3.13), we associated with the fluid the velocities $V_1 = J_1(t_0)^{1/2}$ and $W_0 = T_c\,G_0^{1/2}$. Then, in terms of these quantities $D_0(t)$, *c.f.r.* (3.3.22), can be estimated starting from (3.3.16) by

$$\begin{aligned} D_0(t) \le & \sum_{\underline{k}} \frac{e^{-\nu\underline{k}^2(t-t_0)}}{|\underline{k}|}|\underline{k}|\ |\underset{\sim}{\gamma}{}^{\lambda}_{\underline{k}}(t_0)| + \int_{t_0}^t \sum_{\underline{k}} e^{-\nu\underline{k}^2(\tau-t_0)}|g_{\underline{k}}|\,d\tau + \\ & + C_1 \int_{t_0}^t d\tau \int_\Omega d\underline{y} \frac{D_0(\tau)^2}{(|\underline{x}-y|^2+\nu(t-\tau))^2} \le \\ & \le F_0'\big(\frac{V_1}{\sqrt[4]{(t-t_0)/T_c}} + W_0\big) + F_0''\int_{t_0}^t \frac{D_0(\tau)^2}{\sqrt{(t-\tau)/T_c}}\frac{d\tau}{T_cV_c}\end{aligned} \tag{3.3.25}$$

where F_0' and F_0'' are suitable dimensionless constants and $\nu(t-t_0) \le L^2$.

Since $D_0(t)$ is the maximum of $\underline{u}^\lambda(\underline{x},t)$ in $\underline{x}$ and since $\underline{u}^\lambda$ is C^∞ in $\underline{x}, t$, *because it solves the regularized equation*, for $t \ge t_0$, then $D_0(t)$ is continuous in t. Therefore we conclude the validity of the inequality

$$D_0(t) < 2\,\big(\frac{V_1}{((t-t_0)/T_c)^{1/4}} + W_0\big)\,F_0' \tag{3.3.26}$$

for $t - t_0 > 0$ small enough. This is simply the statement that $D_0(t)$ is continuous in t: *but* $t - t_0$ could have to be extremely smaal and it depends on λ in a, so far, uncontrolled way.

However, for $t < T_c$ the inequality (3.3.26) *will stay valid as long as the upper bound of $D_0(t)$ in (3.3.25), evaluated by replacing $D_0(t)$ with the bound (3.3.26), stays smaller than the right hand side of (3.3.26) itself*, as a moment of thought reveals.

Remark: the condition $t < T_c$ appears because the inequalities have been derived supposing $t - \tau < T_c$ (*i.e.* $\nu(t-\tau) < L^2$).

Therefore, if $t_0+\tau_0$ is the maximal value of $t > t_0$ up to which the inequality (3.3.26) holds, the (3.3.25) implies that the time τ_0 is not smaller than the maximum $\tau \ge 0$ for which

$$\begin{aligned}&\big((\tau/T_c)^{-1/4}\,V_1 + W_0\big)\,F_0' + \\ &+ 8F"_0 {F_0'}^2 \int_0^\tau \Big(\big(\frac{T_c}{\tau'}\big)^{1/2}\,V_1^2 + W_0^2\Big)\,\big(\frac{T_c}{\tau-\tau'}\big)^{1/2}\,\frac{d\tau'}{T_cV_c} \le \\ &\le 2((\tau/T_c)^{-1/4}\,V_1 + W_0)\,F_0'\end{aligned} \tag{3.3.27}$$

Note in fact that the l.h.s. is an estimate of the r.h.s. of (3.3.25) obtained from it by replacing $D_0(t)$ by (3.2.29).

The (3.3.27) is, taking into account the τ–independence of the integral $\pi = \int_0^\tau [\tau'(\tau-\tau')]^{-1/2}\,d\tau'$, a consequence of

$$8F_0'F_0''\pi\frac{(V_1^2 + (\tau/T_c)^{1/2}W_0^2)V_c^{-1}}{V_1 + (\tau/T_c)^{1/4}W_0} \le \frac{1}{(\tau/T_c)^{1/4}} \tag{3.3.28}$$

so that, setting $F' = 1/(8\pi F_0'F_0'')$ and $x \equiv (\tau/T_c)^{1/4}$, the condition is

$$x < F'\frac{V_c(V_1 + xW_0)}{V_1^2 + x^2W_0^2} \tag{3.3.29}$$

which, by $\frac{a+b}{c+d} \ge \min(\frac{a}{c}, \frac{b}{d}$ is implied by $x < F' \min\big(V_c/V_1,\, V_c/(xW_0)\big)$.

For simplicity we shall impose the validity of the last condition (cf. remark above) and of the $t < T_c$ by assuming the simpler (but more restrictive)

$$\tau \le T_0 \stackrel{def}{=} F\,\frac{T_c}{\big(\frac{V_1}{V_c}\big)^4 + \big(\frac{W_0}{V_c}\big)^2 + 1} \equiv F\,\frac{T_c}{R^4 + R_g^2 + 1} \tag{3.3.30}$$

for F small enough, which has a conveniently simple form (for our purposes) in terms of dimensionless quantities.

Remark: The main feature of the definition of the estimate T_0 for τ_0 is of being a function of the Reynolds number R which is *independent of the* *cut–off parameter* λ. Of course the Reynolds number in general depends on λ unless $t_0 = 0$, as already noted).

A further consequence of the (3.3.16) and (3.3.24), that is obtained in a similar way to the one followed in deriving (3.3.25), is

$$J_1(t)^{1/2} \le J_1(t_0)^{1/2} + T_c G_0^{1/2} + \int_{t_0}^{t} D_0(\tau) \frac{B_1}{\sqrt{\nu(t-\tau)}} J_1(\tau)^{1/2}\, d\tau \quad (3.3.31)$$

This can also be bounded by the same kind of reasoning just presented for the bound on $D_0(t)$ in deriving (3.3.26), (3.3.27) and (3.3.30). One gets

$$J_1(t)^{1/2} \le 2(J_1(t_0)^{1/2} + T_c G_0^{1/2}), \qquad t \le T_0 \quad (3.3.32)$$

under the same condition for τ_0, (3.3.30), with a possibly different constant $\tilde{F}$ in place of F.

Hence we have obtained the existence of suitable dimensionless constants $\tilde{F}, \hat{F}$ such that

$$\begin{aligned} &\text{if} \qquad t \le T_0 = T_c \, \min\big(1, \frac{\tilde{F}}{R^4 + R_g^2}\big) \qquad \text{then :}\\ &D_0(t) \le \hat{F}\Big(V_1\big(\frac{t-t_0}{T_c}\big)^{-1/4} + W_0\Big) \stackrel{def}{=} d(t-t_0) \\ &J_1(t) \le 8\big(J_1(t_0) + T_c^2 G_0\big) \stackrel{def}{=} j(t_0) \end{aligned} \quad (3.3.33)$$

We remark, again, the λ–independence of the right hand side and that the function $d(\varepsilon)$ is decreasing with ε.

Thus the proposition is proved in the case $m = 0$. The $m > 0$ cases are treated analogously, or they follow as special cases from propositions IV,V,VI below.

(D) Local existence and regularity. Leray's local theorem.

Proposition I yields the following corollary

II. Proposition *(local existence, regulatity and uniqueness, (Leray)): There is a dimensionless constant $F > 0$ such that if $\underline{u}^0 \in L_2(\Omega)$ is a velocity field with Reynolds number $R < \infty$ (which is equivalent to $\underline{u}^0 \in W^1(\Omega)$) and if the dimensionless strength of the external force is R_g (in the sense of definition 1, (3.3.13)) then there is a weak solution with initial datum $\underline{u}^0$ at $t = t_0$, of the NS equations which verifies the following properties*

$$\begin{aligned} &(1) \quad \underline{u} \in C^\infty((t_0, t_0 + T_0) \times \Omega), \qquad T_0 = F\, \frac{T_c}{R^4 + R_g^2 + 1} \\ &(2) \quad ||\underline{u}(t) - \underline{u}^0||_2 \xrightarrow[t \to t_0]{} 0 \\ &(3) \quad \int_0^{T_0} D_0(\tau)^2\, d\tau < \infty \end{aligned} \quad (3.3.34)$$

where $D_0(t)$ *is the maximum* $\max_{\underline{x}} |\underline{u}(\underline{x},t)|$. *Two weak solutions enjoying properties 1,2,3 necessarily coincide. Finally the solution can be obtained via a constructive algorithm.*

Remarks: (1) We recall that "weak solution" means (*c.f.r.* definition 1 in §3.2) a function $t \to \underline{u}(\underline{x},t)$ with finite $L_2(\Omega)$–norm (*i.e.* finite kinetic energy) verifying a uniform estimate in every finite time interval and making (3.3.10) an identity, as well as (3.2.6), almost everywhere in $t > t_0$.
(2) If there is no force all Leray's solutions will become eventually smooth, see problem [3.3.4]. These are examples of several results of Leray on global existence (see also problem [3.3.5]). The conclusion is that we miss an existence and uniqueness theorem under *general* initial data.

proof: Setting $t_0 = 0$ property (1) follows immediately from the estimates (3.3.14), from the autoregularization property of proposition V of §3.2 and from the remark (3) to the latter proposition.
Furthermore from (3.3.7), and $\int_0^t e^{-\nu\underline{k}^2\tau}\,d\tau \le t^\varepsilon/(\nu\underline{k}^2)^{1-\varepsilon}$ for each $\varepsilon \in [0,1]$, we deduce for all $\lambda \ge 0$

$$\begin{aligned}&|\underline{\gamma}^\lambda_{\underline{k}}(t) - e^{-\nu\underline{k}^2 t}\underline{\gamma}^0_{\underline{k}}| \le \frac{|\underline{g}_{\underline{k}}|t^\varepsilon}{(\nu\underline{k}^2)^{1-\varepsilon}} +\\ &+\frac{t^\varepsilon}{(\nu\underline{k}^2)^{1-\varepsilon}}\frac{\sqrt{E_0L^{-3}}}{L}\;2\big(J_1(0)^{1/2} + T_cG_0^{1/2}\big)\end{aligned} \tag{3.3.35}$$

where the second term is obtained from (3.3.7) by bounding the sum proportionally to $||\underline{\gamma}||_2 \cdot ||\underline{k}\,\underline{\gamma}||_2$ (the $||\cdot||_2$ is defined in (3.2.4)), via the energy estimate (3.2.12) and the estimate (3.3.33). This shows (if $\varepsilon > 0$ is chosen so that $(1-\varepsilon)\cdot 4 > 3$) that

$$||\underline{\gamma}^\lambda(t) - \underline{\gamma}(0)||_2 \le \big(\sum_{\underline{k}} |\underline{\gamma}^0_{\underline{k}}|^2(1-e^{-\nu\underline{k}^2 t})^2\big)^{1/2} + O(t^\varepsilon) \xrightarrow[t\to 0]{} 0 \tag{3.3.36}$$

hence $||\underline{\gamma}^\infty(t) - \underline{\gamma}(0)||_2 \le \lim_{\lambda_j\to\infty} ||\underline{\gamma}^{\lambda_j}(t) - \underline{\gamma}^0(0)||_2 \xrightarrow[t\to 0]{} 0$ and property (2) of (3.3.34) follows.
Property (3) is implied by the second of (3.3.33), saying that $D_0(\tau)$ diverges at most as $\tau^{-1/4}$ as $\tau \to 0$.
It remains to check uniqueness. Indeed, given two solutions $\underline{u}^1$ e $\underline{u}^2$ verifying (3.3.34), let $\Delta = ||\underline{\gamma}^1 - \underline{\gamma}^2||_2^2$ and $\Delta_1 = |||\underline{k}|\,(\underline{\gamma}^1 - \underline{\gamma}^2)||_2)$. Proceeding as usual, *c.f.r.* proposition VI of §3.2, we then get

$$\begin{aligned}&\frac{d}{dt}\frac{\Delta}{2} = -\nu\Delta_1 - \sum_{\underline{k}_1+\underline{k}_2+\underline{k}_3=0} i(\underset{\sim}{\gamma}^1_{\underline{k}_1}\cdot\underset{\sim}{k}_2\underline{\gamma}^1_{\underline{k}_2} - \underset{\sim}{\gamma}^2_{\underline{k}_1}\cdot\underset{\sim}{k}_2\underline{\gamma}^2_{\underline{k}_2})\cdot(\underline{\gamma}^1_{\underline{k}_3} - \underline{\gamma}^2_{\underline{k}_3}) =\\ &= -\nu\Delta_1 + \sum i(\underset{\sim}{\gamma}^1_{\underline{k}_1} - \underset{\sim}{\gamma}^2_{\underline{k}_1})\cdot\underset{\sim}{k}_3\underline{\gamma}^1_{k_2}\cdot(\underline{\gamma}^1_{\underline{k}_3} - \underline{\gamma}^2_{\underline{k}_3})+\\ &+\sum i\underset{\sim}{\gamma}^2_{\underline{k}_1}\cdot\underset{\sim}{k}_2(\underline{\gamma}^1_{\underline{k}_2} - \underline{\gamma}^2_{\underline{k}_2})\cdot(\underline{\gamma}^1_{\underline{k}_3} - \underline{\gamma}^2_{\underline{k}_3}) =\end{aligned} \tag{3.3.37}$$

$$\begin{aligned}
&= -\nu\Delta_1 + L^{-d}\int (\underline{u}^1-\underline{u}^2)\cdot\big[\underset{\sim}{u}^1\underset{\sim}{\partial}\cdot(\underset{\sim}{u}^1-\underset{\sim}{u}^2)\big]+\\
&+ L^{-d}\int\big[\underset{\sim}{u}^2\cdot\underset{\sim}{\partial}(\underline{u}^1-\underline{u}^2)\big]\cdot(\underline{u}^1-\underline{u}^2)\;\le\;-\nu\Delta_1+\\
&+2D_0(t)\sqrt{\Delta_1}\sqrt{\Delta}\;\;\le \nu\Delta\max_{y>0}\big(-y^2+2\nu^{-1}D_0(t)y\big)\le\Delta\,\nu^{-1}D_0^2(t)
\end{aligned}$$

hence

$$\Delta(t)\le\Delta(0)\,e^{2\nu^{-1}\int_0^t D_0(\tau)^2\,d\tau} \tag{3.3.38}$$

thus, since $\Delta(0)=0$, we see that the third of the (3.3.34) implies $\Delta(t)\equiv 0$, for $t\le T_0$.[2]

(E) Exceptionality of singularities. Global Leray's theorem. Leray's solutions.

2. Definition *(L-weak solutions): Consider solutions $\underline{u}^\lambda(t)$ of the Leray's regularized equation with C^∞ initial datum $\underline{u}(0)=\underline{u}^0$ and for all times t. Let $\lambda_j\to\infty$ be a sequence such that $\underline{\gamma}^{\lambda_j}_{\underline{k}}(t)$ converges, for all t and $\underline{k}$, to a weak solution $\gamma^\infty_{\underline{k}}(t)$ with Fourier transform $\underline{u}^\infty(\underline{x},t)$. Such "Leray's solutions" may be distinct from the C-weak solutions of §3.2: hence we shall call them "L-weak solutions".*

Keeping in mind (3.3.11) and the deinition 1 following it, let

$$R^2(t)V_c^2\overset{def}{=}J_1(t)=\liminf_{j\to\infty}\,L^2\sum_{\underline{k}}|\underline{k}|^2|\underline{\gamma}^{\lambda_j}_{\underline{k}}(t)|^2 \tag{3.3.39}$$

Note that all quantites do depend also on the sequence $\{\lambda_j\}$. Then the second of (3.3.8) gives

$$\begin{aligned}
&\int_0^T R(t)^2\,\frac{dt}{T}\le\liminf_{j\to\infty}\int_0^T\;V_c^{-2}L^2\sum_{\underline{k}}|\underline{k}|^2|\underline{\gamma}^{\lambda_j}_{\underline{k}}(t)|^2\,\frac{dt}{T}\le\\
&\le\Big(\frac{L^2}{T}E_0L^{-3}\nu^{-1}+L^2\,\nu^{-1}\sqrt{E_0L^{-3}}\,||\underline{g}||_2\Big)V_c^{-2}<\infty
\end{aligned} \tag{3.3.40}$$

hence $R(t)^2<\infty$ almost everywhere.
Let $\mathcal{E}_n=\{t|R(t)^2<n\}$ and set,*c.f.r.* (3.3.15)

$$\tau_n=F\,\frac{T_c}{n^2+R_g^2+1} \tag{3.3.41}$$

[2] Strictly speaking, *c.f.r.* definition 2, one gets $\Delta(t)\le\Delta(t_0)\exp 2\nu^{-1}\int_{t_0}^t D_0(\tau)^2\,d\tau$ for $t_0>0$ because $\underline{u}^i\in C^\infty$ for $t>t_0$; then it follows from (2) in (3.3.34) that also $\underline{u}^i(t_0)$ tend to the same limit as $t_0\to 0$, namely to $\underline{u}(0)$, in L_2: so that $\Delta(t_0)$ also tends to $\Delta(0)=0$ as $t_0\to 0$ and (3.3.38) follows.

The set $\cup_n \mathcal{E}_n$ has a zero measure complement. Call $t_1^{(n)}, t_2^{(n)}, \ldots$ a denumerable family of times in $\mathcal{E}_n$ such that the set $\tilde{\mathcal{E}}_n$:

$$\tilde{\mathcal{E}}_n \equiv \cup_{j=1}^{\infty} [t_j^{(n)}, t_j^{(n)} + \tau_n) \quad \text{coincides with} \qquad \cup_{t\in\mathcal{E}_n} [t, t+\tau_n) \tag{3.3.42}$$

and it is not difficult to see, by abstract thinking (*i.e.* no estimates are needed), that such a family exists, and the open set $\mathcal{E}_n^0 \stackrel{def}{=} \tilde{\mathcal{E}}_n \supset \cup_{t\in\mathcal{E}_n}(t, t+\tau_n)$ is contained in $\tilde{\mathcal{E}}_n$.

As a consequence of propositions I and II we see that $u^\infty(\underline{x}, t)$ is C^∞ on the set $\mathcal{E}_n^0$ and there it verifies, together with all its derivatives, estimates depending only on n and upon the distance of t from $\partial\mathcal{E}_n^0$.

Since $\cup_n \mathcal{E}_n^0$ differs by an at most denumerable set [3] from the set $\cup_n \tilde{\mathcal{E}}_n$ and $\cup_n \tilde{\mathcal{E}}_n \supset \cup_n \mathcal{E}_n$ it follows that $\cup_n \mathcal{E}_n^0$ is open and has a zero measure complement.

Therefore

III. Proposition *(L-weak solutions can only have sporadic singularities (Leray)): There exist weak solutions of the 3–dimensional NS equation which are C^∞ on an open set with a zero measure complement. All Leray's solutions enjoy this property.*

This is, in essence, the high point of Leray's theory. A few more substantially stronger properties have been very recently obtained: we shall analyze them in the following sections §3.4,§3.5.

(F) Characterization of the singularities. Leray–Serrin theorem.

A further corollary, where $\nu = 1$ for simplicity, is the following theorem (*Serrin*).

IV. Proposition *(velocity is unbounded near singularities): Let $\underline{u}(\underline{x}, t)$ be a weak solution of the NS equation verifying $||\underline{u}||_2^2 < E_0$ for $t > 0$ and*

$$\begin{aligned}\underline{u}(\underline{x}, t) = &\int_\Omega \Gamma(\underline{x} - \underline{y}, t)\, \underline{u}^0(\underline{y})\, d\underline{y} + \int_0^t d\tau \int_\Omega \Gamma(\underline{x} - \underline{y}, t - \tau)\, \underline{g}(\underline{y})\, d\underline{y} + \\ &- \int_0^t d\tau \int_\Omega \vec{\partial} \underset{\approx}{\underline{T}}\, (\underline{x} - \underline{y}, t - \tau)\, \vec{u}(\underline{y}, \tau)\, \underset{\sim}{u}\, (\underline{y}, \tau)\, d\underline{y}\end{aligned} \tag{3.3.43}$$

Given $t_0 > 0$ suppose that $|\underline{u}(\underline{x}, t)| \le M$, $(\underline{x}, t) \in U_\rho(\underline{x}_0, t_0) \equiv$ sphere of radius $\rho < t_0$ around $(\underline{x}_0, t_0)$, for some $M < \infty$: then $\underline{u} \in C^\infty(U_{\rho/2}(\underline{x}_0, t_0))$.

Remarks:

(1) This means that the only way a singularity can manifest itself, in a weak solution (in the sense (3.3.43)) of the NS equations, is through a divergence

[3] Since $\mathcal{E}_n^0$ is defined in terms of open intervals $(t, t+\tau_n)$ rather than semiclosed $[t, t+\tau_n)$ the set $\mathcal{E}_n^0$ might fail to contain the points $\cup_{n,j} t_j^{(n)}$.

of the velocity field itself. For instance it is impossible to have a singular derivative *without* having the velocity itself unbounded. Hence, if $d \ge 3$ velocity discontinuities are impossible (and even less so shock waves), for instance. Naturally if $\underline{u}(\underline{x}, t)$ is modified on a set of points $(\underline{x}, t)$ with zero measure it remains a weak solution (because the Fourier transform, in terms of which the notion of weak solution is defined, does not change), hence the condition $|\underline{u}(\underline{x}, t)| \le M$ *for each* $(\underline{x}, t) \in U_\lambda(\underline{x}_0, t_0)$ can be replaced by the condition: *for almost all* $(\underline{x}, t) \in U_\rho(\underline{x}_0, t_0)$.
(2) It will become clear that the above result is not strong enough to overcome the difficulties of a local theory of regularity of the L–weak solutions. Therefore one looks for other results of the same type and it would be desirable to have results concerning regularity implied by *a priori* informations on the vorticity. We have already seen that bounded total vorticity implies regularity (*c.f.r.* §3.2 proposition 5): however it is very difficult to go really beyond; hence it is interesting to note that also other properties of the vorticity may imply regularity. A striking result in this direction, although insufficient for concluding regularity (if true at all) of L–weak solutions, is the Constantin–Fefferman theorem that we describe without proof in proposition VII in (G) below.

proof: This is a consequence of a few remarkable properties of the integrals, often called with the generic name of "*heat kernels*",

$$\begin{aligned} V(\underline{x}, t) =& \mathcal{P}' F \equiv \int_\Omega P(\underline{x} - \underline{y}, t - \tau) F(\underline{y}, \tau)\, d\underline{y} \\ V(\underline{x}, t) =& \mathcal{P} F \equiv \int_0^t \int_\Omega \partial_j P(\underline{x} - \underline{y}, t - \tau) F(\underline{y}, \tau)\, d\tau\, d\underline{y} \end{aligned} \tag{3.3.44}$$

where $P = \Gamma$ or $P = T$, see (3.3.20), as operators on a function F which is bounded by a constant M. The properties below can be regarded as establishing the analogue of the autoregularization property directly "in position space". They are expressed by the following propositions

V. Proposition *(regularization induced by heat kernels): There is a suitable function* K_t^0*, decreasing in* t *(hence bounded for* $t > 0$*) such that*

$$|V(\underline{x}, t)| + \frac{|V(\underline{x}, t) - V(\underline{x}', t)|}{(L^{-1}|\underline{x} - \underline{x}'|)^{1/2}} \le M\, K_t^0 \qquad \forall\ \underline{x}, \underline{x}' \tag{3.3.45}$$

In other terms the operators in (3.3.44) transform bounded functions into Hölder continuous functions with exponent $1/2$.

See problems [3.3.10], [3.3.11] below for a proof, and furthermore

VI. Proposition*(stronger regularization induced by heat kernels): Likewise if* F *verifies (3.3.45) (with* F *in place of* V*) then* $\mathcal{P}F$ *and* $\mathcal{P}'F$ *are differentiable in* $\underline{x}$ *and, for a suitable* t*–decreasing (hence finite for* $t > 0$*)* K_t^1*, have derivatives bounded by*

$$|\partial_i V(\underline{x}, t)| \le K_t^1 M \tag{3.3.46}$$

i.e. the operator in (3.3.44) transforms Hölder continuous functions into differentiable ones.

A proof is described in problem [3.3.12] below.

Accepting for the time being (3.3.45), (3.3.46), *c.f.r.* problems [3.3.11], [3.3.12], we can complete the proof of proposition IV.

Let $\chi = \chi_{\underline{x}_0,t_0}(\underline{y},\tau) \le 1$ be a C^∞ function vanishing outside the ball U_ρ and equal to 1 inside the ball $U_{3\rho/4}$. Since Γ, T are C^∞ if $\underline{\xi}^2 + \tau > 0$ the exspression (3.3.43) yields a C^∞ function for $(\underline{x},t) \in U_{\rho/2}$ if $\underline{u}$ is everywhere replaced, in the r.h.s., by $(1-\chi_{\underline{x}_0,t_0})\underline{u}$.

Thus it remains to analyze (3.3.43) with $\underline{u}$ replaced, in the r.h.s., by $\chi_{\underline{x}_0,t_0}\,\underline{u}$. But the function $\chi\,\underline{u}$ is, by assumption, bounded by M so that by propositions V, VI (3.3.43) is differentiable in $\underline{x}$. Hence $\underline{u}$ is differentiable in $\underline{x}$, and still by (3.3.43) we see that $\underset{\sim}{\partial}\,\underline{u}$ is given by expressions similar to (3.3.43) as the new $\underline{x}$ derivatives of T, Γ can be "transferred", by integration by parts, on the functions $\underline{u}$. Hence by the same argument $\underset{\sim}{\partial}\,\underline{u}$ is differentiable in $\underline{x}$. Repeating the argument we see that $\underline{u}$ is infinitely differentiable in $U_{\rho/2}$. Knowing differentiability in $\underline{x}$ and differentiating (3.3.43) with respect to t one can integrate by parts and "transfer" the time derivatives acting on the kernels Γ, T into $\underline{x}$–derivatives of the functions $\underline{u}$ (which exist by the argument above) because $\partial_t\Gamma = \nu\,\Delta\Gamma$ and $\partial_t T = \nu\,\Delta\,T$. Hence $\underline{u}$ is C^∞ in t as well for $(\underline{x},t) \in U_{\rho/2}$.

(G) Vorticity orientation uncertainty at singularities.

It is interesting, see remark (2) in (F) above, to quote the following "vorticity based" regularity proposition (Constantin and Fefferman, *c.f.r.* [CF93]):

Proposition VII *(vorticity orientation is uncertain at singularities):* *Suppose that $\underline{u}(t)$ is a L–weak solution of the NS equations with periodic boundary conditions in a cubic container Ω with C^∞ external force $\underline{g}$ and initial datum $\underline{u}^0 \in C^\infty$. Let $\underline{\omega} = \underline{\partial} \wedge \underline{u}$ be its vorticity field. Let $\underline{u}^{(N)}$ be the sequence of approximating solutions with Leray regularization parameter N; let their vorticity be $\underline{\omega}^{(N)}$ and $\underline{\xi}^{(N)} = \underline{\omega}^{(N)}/|\underline{\omega}^{(N)}|$ be the its "direction". Suppose that there exist constants $X, \rho > 0$ such that if both $|\underline{\omega}^{(N)}(\underline{x},t)| > X$ and $|\underline{\omega}^{(N)}(\underline{x}+\underline{y},t)| > X$ then the projection of $\underline{\xi}^{(N)}(\underline{x}+\underline{y},t)$ on the plane orthogonal to $\underline{\xi}^{(N)}(\underline{x},t)$ is bounded above by $|\underline{y}|/\rho$ for all $0 \le t \le T$, where T is a positive time, and for all pairs $\underline{x}, \underline{y}$, uniformly in the cut–off parameter N.*
Then the solution is of class C^∞ in the time interval $[0,T]$.

Hence the "NS *solution is smooth unless in the (smooth) approximations vorticity at $\underline{x}$ changes wildly direction*" as $\underline{x}$ varies. This result is remarkable because it gives a regularity property under conditions involving the vorticity. The result does not apply to C-weak solutions (*c.f.r.* (3.2.11)).

As it will appear clearly in the following sections, particularly in §3.5, what is really missing in the theory of regularity of NS solutions is a proper way of taking into account that vorticity is transported (one also says "advected") by the fluid flow. Even the strongest regularity result, the CKN theorem, *c.f.r.* §3.5, gives detailed local information (unlike the above theorem that relies on an assumption that involves vorticities at all points at a given instant) but *without* referring to the validity of the Thomson theorem at zero voscosity.

(H) Large containers.

The theory just developed is dimensionally satisfactory only if the initial velocity field $\underline{u}_0$ and the density of force $\underline{g}$ are regular on a length scale L equal to that of the container.

Sometimes, however, one considers situations in which the length scale characteristic of the initial data, be it r, and of the forces is different from that of the container: *e.g.* "a vortex in the sea", see problem [3.3.8] below. In such cases it will be important to keep the roles of the two scales distinct.

Suppose that data and forces are "on scale" r: mathematically this means that data and forces are appreciably nonzero on a small sphere of radius r where their properties are simply described by parameters $\tilde{v}$ and $\tilde{g}$ that fix the order of magnitude. A precise way to formulate the latter property is to assume that $\underline{u}^0$ and $\underline{g}$ are analytic functions of $\underline{x}$ with holomorphy domain $|\mathrm{Im}\, x_i| < r$ and are bounded in this complex polistrip by $\tilde{v}$ and $\tilde{g}$, respectively, and furthermore that they are negligible ("small") for $|\mathrm{Re}\, x_i| > r$, (*c.f.r.* the final note in §1.2).

Then the relevant norms for the formulation of the theorems can be estimated as follows

$$\begin{aligned}&\int_\Omega d\underline{x}\, |\underset{\sim}{\partial}\underline{u}^0(\underline{x})|^2 \sim \tilde{v}^2 r,\\ &\int_0^t d\tau \int_\Omega |\Gamma(\underline{x}-\underline{y}, t-\tau)\underline{g}(\underline{y})|\, d\underline{y} \sim \frac{\tilde{g} r^2}{\nu}\end{aligned} \tag{3.3.47}$$

where the second quantity is what becomes of the W_0 in (3.3.26) in the proof of Leray's's bounds.[4] Hence the relevant quantities in the formulation of the results of Leray's theorem are

$$\begin{aligned}&V_c = \frac{\nu}{L}, \quad T_c = \frac{L^2}{\nu}, \quad V_1 \sim \tilde{v}(\frac{r}{L})^{1/2}, \quad W_0 \sim \tilde{g}\frac{r^2}{\nu}\\ &R_g^2 \sim \tilde{g}\frac{r^2 L}{\nu^2}, \qquad R \sim \frac{\tilde{v}(Lr)^{1/2}}{\nu},\end{aligned} \tag{3.3.48}$$

It is therefore useful to introduce "local" scales of velocity $\tilde{V}_c$ and time $\tilde{T}_c$ as well as the Reynolds number $\tilde{R}$ and the dimensionless strength number,

[4] To estimate (3.3.47) one can bound Γ with (3.3.21) and use the property of "smallness" of $\underline{g}$ outside the ball of radius r.

$\tilde R_g$

$$\tilde V_c = \frac{\nu}{r}, \qquad \tilde T_c = \frac{r^2}{\nu}, \qquad \tilde R = \frac{\tilde v r}{\nu}, \qquad \tilde R_g = \frac{\tilde g r^3}{\nu^2} \tag{3.3.49}$$

It follows that the characteristic time scale T_0 of the local solution of Leray and the corresponding velocity estimate $D_0(t)$ are

$$\begin{aligned} T_0 =&\bar F T_c\,(R^4 + R_g^2)^{-1} = \bar F \tilde T_c\,(\tilde R^4 + \tilde R_g^2)^{-1}, \\ D_0(t) \le& F\tilde v\,\big((r^2\,(\nu(t-t_0))^{-1})^{1/4} + \tilde g r^2\,\nu^{-1}\big) \end{aligned} \tag{3.3.50}$$

which, as we should have expected, are *independent* from L and essentially the same that we would find if the container had side size $\sim r$.

Problems: *Further results in Leray's theory*

[3.3.1]: Check that there would be difficulties in showing that the Leray's solutions with initial data $\underline u^0 \in W^1(\Omega)$ (*i.e.* $J_1(0) < \infty$) coincide with the C-weak solutions of §3.2 with the same initial data. Show that they would coincide if the C-weak solutions verified property (2) in (3.3.34), *i.e.* took the initial value with continuity in $L_2(\Omega)$ as $t \to 0$. (*Idea:* it is not known whether the two notions coincide.)

[3.3.2]: (Leray's) Check that the technique used to obtain (3.3.25) can be adapted to show that if $E(t) = L^3 \sum_{\underline k} |\gamma_{\underline k}(t)|^2$ then there are B_1, B_2 such that

$$D_0(t) \le B_1\Big(D_0(0) \wedge \frac{V_1}{\sqrt[4]{t/T_c}} + W_0\Big) + B_2 \int_0^t \Big(\frac{D_0(\tau)^2}{\sqrt{(t-\tau)/T_c}} \wedge^* \frac{L^{-3}E(\tau)}{((t-\tau)/T_c)^2}\Big)\frac{d\tau}{T_c V_c} \tag{3.3.51}$$

where $a \wedge^* b$ means a if $\nu(t-\tau) \le L^2$ and $\min(a,b)$ otherwise. (*Idea:* Note that $\int_\Omega |\Gamma(\underline x, t)|\, d\underline x \le B_1$ for a suitable B_1 and for each t, by (3.3.21) and, furthermore, if $t - \tau \ge T_c$, by the Schwartz inequality and (3.3.22):

$$\sup_{\underline y} |\partial_{\underline y} \underline T(\underline y, t-\tau)| \int_\Omega d\underline y |\,\langle \underline u(\underline y, \tau)\rangle_\lambda|\, |\underline u(\underline y, \tau)| \le \frac{L^{-4}E(\tau)}{((t-\tau)/T_c)^{\frac12}} C_1 \tag{3.3.52}$$

where C_1 is a suitable constant. Furthermore

$$\begin{aligned} &\sum_{\underline k} e^{-\nu \underline k^2 t} |\underline\gamma_{\underline k}(0)| \le \sum_{\underline k} \frac{e^{-\nu \underline k^2 t}}{|\underline k|} |\underline k| |\underline\gamma_{\underline k}(0)| \le \\ &\le (\frac{J_1(0)}{L^{-1}})^{1/2} (\sum_{\underline k} \frac{e^{-\nu \underline k^2 t}}{\underline k^2})^{1/2} \le B \frac{V_1}{(t/T_c)^{1/4}} \end{aligned} \tag{3.3.53}$$

for a suitable B. Hence (3.3.51) follows.)

[3.3.3]: (*small initial data and global solutions for NS in $d = 3$, (Leray)*) Let $V_0 = D_0(0) + W_0$ and check that if $V_0/V_c < \delta$ is small enough then the solution exists for all times and $D_0(t)$ can be bounded proportionally to V_0. (*Idea:* If $E(t) = L^3 \sum_{\underline k} |\underline\gamma_{\underline k}(t)|^2$ the *a priori* energy estimate holds also in this case, see the comment preceding (3.3.10), and it will be $L^{-3}E(\tau) \le B_3 V_0^2$ for all τ (*c.f.r.* (3.2.12)). Then by the inequality in problem [3.3.2] it will be $D_0(t) \le 2B_1 V_0$ as long as it is

$$B_2 \int_0^t \frac{4B_1^2\,(1 + t/T_c) V_0^2}{\sqrt{(t-\tau)/T_c}} \wedge^* \frac{B_3 V_0^2}{((t-\tau)/T_c)^2}\, \frac{d\tau}{T_c V_c} < B_1 V_0 \sqrt{1 + t/T_c} \tag{3.3.54}$$

where B_2 is the constant of [3.3.2], *i.e.* dividing the integral between 0 and $t-T_c$ and between $t-T_c$ and t, as long as

$$\frac{V_0^2B_2}{V_c}\Big(\int_0^{t-T_c}\frac{B_3}{((t-\tau)/T_c)^{\frac12}}\frac{d\tau}{T_c}+\int_{t-T_c}^t\frac{4B_1^2\sqrt{1+t/T_c}}{\sqrt{(t-\tau)/T_c}}\frac{d\tau}{T_c}\Big)<B_1V_0\sqrt{1+t/T_c} \tag{3.3.55}$$

where the first term should be omitted if $t\le T_c$ while, in this case, the second should be the integral over $[0,t]$.

Therefore there is a $B<\infty$ majorizing the sum of the integrals (*for every t*): and if $V_0<V_cB_1/(BB_2)$ (which implies $\delta\le B_1/(B\,B_2)$) the $D_0(t)\le 2B_1V_0\sqrt{1+t/T_c}$ will hold for all times: apply at this point the Leray's–Serrin theorem).

[3.3.4] (Leray): Supposing $\underline g=\underline 0$ show that if $\underline u^0\in L_2(\Omega)$ is an initial datum for the NS equations then there exists a time $\bar T$ large enough such that (all) the L-weak solutions with $\underline u^0$ as initial datum are of class C^∞ for $t>\bar T$. And $\bar T$ can be chosen

$$\bar T=B\,T_c\log_+\left(E(0)L^{-3}/V_c^2\right)^{-1} \tag{3.3.56}$$

where B is a suitable constant and $\log_+x$ denotes the maximum betwen 1 and $\log x$. (*Idea:* Since $\underline g=\underline 0$ we have, as in (3.2.14) with $\underset{\sim}{\varphi}=\underline 0$, $\dot E\le-\nu k_0^2E$ and $E(t)\le E(0)e^{-\nu k_0^2t}$, $k_0=2\pi L^{-1}$. Furthermore (3.2.13) give for all $T\ge0$ and using the definition (3.3.13)

$$\int_T^{T+T_c}\frac{d\tau}{T_c}\,J_1(\tau)\le\frac12\frac{E(0)L^{-1}}{\nu T_c}\,e^{-\nu k_0^2T}\stackrel{def}{=}V_c^2r_T^2 \tag{3.3.57}$$

Hence if $\tilde T$ is large enough and $T>\tilde T$ it is (*c.f.r.* (3.3.15))

$$T_c\,F\,r_T^{-4}>3T_c \tag{3.3.58}$$

and in every interval $[T-2T_c,T-T_c]$ of length T_c there will be a time t_0 where $J_1(t_0)<V_c^2r_T^2$, *i.e.* where the Reynolds' number $R(t_0)$ will be smaller than r_T, to the right of which the solution is bounded by (3.3.14). In particular, keeping in mind the definitions (3.3.1) and equation (3.3.14), it will be bounded in the interval $[T,T+T_c]$ by $D_0(t)\le V_c\,F_0\,r_{\tilde T}$ and by the arbitrariness of T the bound on $D_0(t)$ holds for all $t\ge\tilde T+2T_c$. Hence the regularized solution is bounded uniformly in the regularization parameter and for $t\ge\tilde T+2T_c$ provided (3.3.58) holds for $T>\tilde T=(4\nu k_0^2)^{-1}\log(\frac{3}{F}\left(E(0)L^{-1}/2\nu T_cV_c^2\right)^{-2})$, *i.e.* recalling the definition (3.3.1) we get (3.3.56) by proposition IV.)

[3.3.5]: (*a second global result*(Leray)) Supposing $\underline g=\underline 0$ show that if, for p large enough,

$$\Big(\frac{(L^{-3}E(0))^{1/2}}{V_c}\Big)^{p-3}\Big(\int_\Omega\Big|\frac{\underline u(\underline x,0)}{V_c}\Big|^p\frac{d\underline x}{L^3}\Big)^{1/p}<\varepsilon$$

and ε is small enough then there exists a unique C^∞ solution defined for all times. (*Idea:* Similar to problem [3.3.2].)

[3.3.6]: Consider an approximate δ–function as in (A) $N^d\chi(N\underline x)$ and define $\langle\underline u\rangle_N(\underline x)$ as $\int_{R^d}N^d\chi(N\underline y)\underline u(\underline x+\underline y)\,d\underline y$, interpreting $\underline u$ as extended to the entire space R^d by defining it a speriodic with period L in all the d coordinate directions. We call *Euler equation with Leray's regularization* the equation

$$\underline{\dot u}+\langle\underline u\rangle_N\cdot\underset{\sim}{\partial}\,\underline u=-\underline\partial p+\underline g,\qquad \underline\partial\cdot\underline u=0$$

Adapt the analysis of the problems of §3.1 to check that if $\underline{u}^N$ is a solution of the regularized (with Leray's regularization) Euler equation then, for $\underline{g} = \underline{0}$

$$\frac{1}{2}\frac{d}{dt}||\underline{u}^N||^2_{W_m} \le G_m||\underline{u}^N||^3_{W_m}$$

with G_m independent from N. (*Idea:* This is the "same" as the corresponding inequality for the Euler equation, see (3.1.2), (3.1 .3). Differentiate $\underline{m}$ times Euler equation and multiply both sides by $\partial^{\underline{m}}\underline{u}$; sum over $\underline{m}$ with $|\underline{m}| \le m$ and integrate over $\underline{x}$. One should note that the conclusions of [3.3.6] (and [3.3.2]) hold also if $\underset{\sim}{u}$ is replaced by $\langle \underset{\sim}{u} \rangle_N$ because the differentiations commute with the averaging operation and the modulus of an average is majorized by the average of the moduli *because* χ is nonnegative).

[3.3.7]: Combine the analysis of the problems in §3.1 with the ideas of this section and with [3.3.6] to derive an *a priori* estimate on the kinetic energy $E(t)$, of $\underline{u}^N$, which is independent of N (so that $E(t)^{1/2} \le E(0)^{1/2} + ||\underline{g}||_{L_2} t$). And to show that the Leray regularized Euler equation of [3.3.6] has a solution C^∞ in $t, \underline{x}$ for each initial datum $\underline{u}^0 \in C^\infty$,

[3.3.8]: Consider a vortex in R^3, regarded as motion of water in normal conditions with velocity

$$\underline{u} = \frac{1}{2}\underline{\omega}(\underline{r}) \wedge \underline{r}, \qquad \underline{\omega}(\underline{r}) = \Gamma e^{-\underline{r}^2/r_0^2}\,\underline{k}$$

where r_0 is a length scale and $\underline{k}$ is the unit vector of the z axis; Γ is an inverse time scale. Suppose that no volume force acts on the fluid. Compute:
(1) The Reynolds number (in the sense of Leray's, (3.3.13), and in the sense discussed for "large containers" in (H) above).
(2) Assuming that the vortex has radius $r = 1.\, m$ with $\Gamma = 1.\, s^{-1}$ estimate the time of existence of the local Leray's solution (refer to the theory of large containers in (H)).
(3) Estimate how large should the time scale Γ^{-1} (*c.f.r.* [3.3.3]) be to be sure of the existence of a global solution, given that the length scale is $r = 1.m$?
(*Idea:* Compute the constants of Leray's theory and apply it, via the extension to large containers. The kinematical viscosity of water is $\nu = 0.01 cm^2/sec$ and its density is $1.g/cm^3$ (normal conditions 4^oC, $1\,atm$).

[3.3.9]: Check that $\Gamma(\underline{x}, t)$ defined in (3.3.17) is:

$$\Gamma(\underline{x}, t) = \frac{e^{-\underline{x}^2/4t\nu}}{\sqrt{4\pi\nu t}^3} + \sum_{\underline{n}\ne 0} \frac{e^{-(\underline{x}+\underline{n}L)^2/4t\nu}}{\sqrt{4\pi\nu t}^3} \tag{3.3.59}$$

and show that, therefore, it suffices to prove the (3.3.22) for $|\underline{x}|$ and $|t|$ small. Note that it is necessary to check only the first term, the others are C^∞ corrections (*c.f.r.* problem [2.3.12])). (*Idea:* By (3.3.18) Γ is the heat equation kernel (on the torus) and the first term in (3.3.59) is the heat equation kernel in R^3; all terms are regular near the origin except the first.)

[3.3.10]: *(properties of heat kernels)* Let, in this and in the following problems, $\nu = 1$ for simplicity. Note that (3.3.22) can be studied for $\nu t < L^2$ by assuming alternatively that $x^2 \le t$ and $x^2 > t$ and showing that in the first case it is: $|\partial^{\underline{\alpha}} P| \le C_{\underline{\alpha}}(\nu t)^{-(3+|\underline{\alpha}|)/2}$, where $P = \Gamma$ or $P = T$; and in the second: $|\partial^{\underline{\alpha}} P| \le C_{\underline{\alpha}}\,(\underline{x}^2)^{-(3+|\underline{\alpha}|)/2}$, if $C_{\underline{\alpha}}$ is a suitable constant. For $\nu t > L^2$ the r.h.s. is replaced by $C_{\underline{\alpha}} L^{-(3+|\underline{\alpha}|)}(L^2/\nu t)^{|\alpha|/2}$.
(*Idea:* In the case of Γ this is a simple direct check. So we can assume that the (3.3.22) holds for Γ. Note that the second term in T (*c.f.r.* (3.3.20)) is not proportional to Γ; it can, however, be thought of as the convolution product between the derivatives $\partial_i\partial_j\, G$ of the Green's function of the Laplace operator and Γ, *c.f.r.* (3.3.22). Then study

$$\partial_i\partial_j G * \Gamma(\underline{x}, t) \equiv \partial_i\partial_j \int_\Omega G(\underline{x} - \underline{y})\Gamma(\underline{y}, t)\, d\underline{y} \tag{3.3.60}$$

with $|\underline{x}|$, $|t|$ small and consider only the term $\underline{n}=0$ in (3.3.59), because the others contribute a correction of class C^∞ in $\underline{x}$. Again we can replace $G(\underline{x})$ with $-1/4\pi|\underline{x}|$ since $G(\underline{x})\equiv-\frac{1}{4\pi|\underline{x}|}+\gamma(\underline{x})$ and $\gamma(\underline{x})$ is of class C^∞ for $|\underline{x}|$ small and with derivatives trivially bounded proportionally to $L^{-3-|\underline{\alpha}|}$, *c.f.r.* problems [2.3.12]÷[2.3.14].

Hence, if $t\ge\underline{x}^2$ and if the constants $C^j, C_{|\underline{\alpha}|+2}$ are suitably chosen we use the bound (3.2.22) for Γ

$$\begin{aligned}
&|\partial^{\underline{\alpha}}\partial_i\partial_j\int\frac{1}{|\underline{x}-\underline{y}|}\frac{e^{-\underline{y}^2/4t}}{\sqrt{4\pi t}^3}\,d\underline{y}|\le C_{|\underline{\alpha}|+2}\int\frac{1}{|\underline{x}-\underline{y}|}\frac{d\underline{y}}{(\underline{y}^2+t)^{(5+|\underline{\alpha}|)/2}}\\
&\le C_{|\underline{\alpha}|+2}\Big(\int_{|\underline{y}|<2|\underline{x}|}\cdot+\int_{|\underline{y}|>2|\underline{x}|}\cdot\Big)\le C^1\frac{1}{t^{(5+|\underline{\alpha}|)/2}}|\underline{x}|^2+\int_{|\underline{y}|>2|\underline{x}|}\cdot\le\\
&\le C^2\Big(\frac{1}{t^{(3+|\underline{\alpha}|)/2}}+\int\frac{d^3\underline{y}}{|\underline{y}|(\underline{y}^2+t)^{(5+|\underline{\alpha}|)/2)}}\Big)\le\frac{C^3}{t^{(3+|\underline{\alpha}|)/2}}
\end{aligned}$$

If instead $t<\underline{x}^2$ divide the integral in the part $|\underline{y}|<|\underline{x}|/8$ and in the part with $|\underline{y}|>|\underline{x}|/8$. The second part can be bounded as needed (change variable $\underline{y}'=\frac{y}{|\underline{x}|}$ to get $\le|\underline{x}|^{-(3+|\underline{\alpha}|)}\int\frac{d\underline{y}'}{|\underline{e}-\underline{y}'|}|\underline{y}'|^{-(5+|\underline{\alpha}|)}$, if $\underline{e}$ is the unit vector $\underline{x}/|\underline{x}|$). Therefore it remains to study: $\int_{|\underline{y}|<|\underline{x}|/8}\frac{d\underline{y}}{|\underline{x}-\underline{y}|}\partial^{\underline{\alpha}+2}\frac{e^{-y^2/4t}}{\sqrt{4\pi t}^3}$. Note that the integral can be performed by parts generating various terms, on the boundary of the sphere $|\underline{y}|=|\underline{x}|/8$, which can be majorized by their maximum (on the sphere). One obtains quantities proportional to expressions like

$$\begin{aligned}
&\int_{|\underline{y}|=|\underline{x}|/8}d\sigma\partial^{p-1}\frac{1}{|\underline{x}-\underline{y}|}\partial^{|\underline{\alpha}|+2-p}\big(\frac{e^{-y^2/4t}}{\sqrt{4\pi t}^3}\big)\le\\
&\le\text{const}\ \frac{|\underline{x}|^2}{|x|^p}\frac{1}{(x^2+t)^{(3+|\underline{\alpha}|+2-p)/2}}\le\text{const}|\underline{x}|^{-(|\underline{\alpha}|+3)}
\end{aligned}$$

where we denote, generically, by $\partial^{|\underline{\alpha}|+2-p}$ and ∂^p a derivative with respect to the components of $\underline{y}$ and of order $|\underline{\alpha}|+2-p$ or of order p).

Furthermore one has to consider the volume integral. This involves a sum of quantities bounded by $\text{const}|\underline{x}-\underline{y}|^{-|\underline{\alpha}|-3}$ *i.e.* by $\text{const}\,|\underline{x}|^{-|\underline{\alpha}|-3}$ (as $|\underline{y}|\ll|\underline{x}|$) multiplied by $e^{-\underline{y}^2/4t}/(4\pi t)^{3/2}$. Hence by using that the integral of the heat kernel, as $\underline{y}$ varies in the whole space, has value 1, we find that the part with $t<\underline{x}^2$ is also bounded by $\text{const}|\underline{x}|^{-|\underline{\alpha}|-3}$. Hence the (3.3.22) follow).

[3.3.11]: (*proof of proposition V*) Referring to the first of (3.3.22) consider the second of (3.3.44) and prove that, if $|F|\le M$, the (3.3.45) holds. An analogous analysis holds for the first of the (3.3.44). (*Idea:* For instance let $P=\Gamma$ and note: $|V(\underline{x},t)-V(\underline{x}',t)|\le\int_0^t\int_\Omega|\partial_j\Gamma(\underline{x}-\underline{y},t-\tau)-\partial_j\Gamma(\underline{x}'-\underline{y},t-\tau)|M\,d\underline{y}\,d\tau$. Decompose the integral into the sum of the integrals extended to the domain $|\underline{y}-(\underline{x}+\underline{x}')/2|<|\underline{x}-\underline{x}'|$ and to its complement. The first integral is bounded by majorizing the modulus of the difference by the sum of the moduli and making use of (3.3.22) to bound each of the two terms (performing first the integral over τ and then that on $\underline{y}$).

If B is a suitable constant we get: $\le 2MC_1\int_0^t\int_{|\underline{y}|<2|\underline{x}-\underline{x}'|}(\underline{y}^2+t-\tau)^{-2}\,d\underline{y}\,d\tau\le 2MC_1B|\underline{x}-\underline{x}'|$. On the other hand the integral over $|\underline{y}-(\underline{x}+\underline{x}')/2|\ge|\underline{x}-\underline{x}'|$ is simply bounded by applying Taylor formula to the integrand majorizing it by $\int_0^1 d\sigma\,|\underline{\partial}\partial_j\Gamma(\underline{x}+\sigma(\underline{x}'-\underline{x})-y,t-\tau)|\,|\underline{x}'-\underline{x}|$ and applying then the (3.3.22) to estimate the expression with the gradient as $\le MC_2|\underline{x}' - \underline{x}| \int_{|\underline{y}|>|\underline{x}'-\underline{x}|/2} \int_0^t \frac{d\underline{y}\, d\tau}{(\underline{y}^2+t-\tau)^{5/2}}$. We get

$$\le MC_2|\underline{x}' - \underline{x}| \cdot B \int_{|\underline{x}'-\underline{x}|/2}^{L} \frac{dy}{\underline{y}^3} \le MC_2B'|\underline{x} - \underline{x}'| \log \frac{2L}{|\underline{x} - \underline{x}'|}$$

if B, B' are suitable constants; and this also shows what we wanted in the case $P = T$ (because the proof depends only on (3.3.22)) and it estimates K^0_t, which we thus see that can even be chosen *independent of* t, furthermore the Hölder exponent $\frac{1}{2}$ in Proposition 5 can be replaced by any η with $\eta < 1$.)

[3.3.12]: (*proof of proposition VI*) suppose that F is Hölder continuous with exponent $\alpha > 0$, *i.e.* suppose that, for a fixed $0 < \alpha \le 1$,

$$|F(\underline{x}, t)| + \frac{|F(\underline{x}, t) - F(\underline{x}', t)|}{(L^{-1}|\underline{x} - \underline{x}'|)^\alpha} \le M_\alpha \tag{3.3.61}$$

Then show that the function V in the second of (3.3.44) is differentiable if $P \equiv \Gamma$ and verifies (3.3.46), (the case $P = T$ will be identical because we shall use only (3.3.22)). Analogously deduce the validity of (3.3.46) for the first of the (3.3.44). (*Idea:* Differentiability can be studied by evaluating the differential ratio in the direction of the unit vector $\underline{e}$

$$\Delta \equiv \frac{V(\underline{x} + \varepsilon\underline{e}, t) - V(\underline{x}, t)}{\varepsilon} = \int_0^t \int_\Omega \frac{1}{\varepsilon}(\partial_j\Gamma(\underline{x} + \varepsilon\underline{e} - \underline{y}, t - \tau) - \partial_j\Gamma(\underline{x} - \underline{y}, t - \tau))F(\underline{y}, \tau)$$

and *noting that if* $F(\underline{y}, \tau)$ *was replaced by* $F(\underline{x}, t)$ *the second member would vanish*, by integration by parts. Hence

$$\Delta \equiv \int_0^t d\tau \int_\Omega d\underline{y}\frac{1}{\varepsilon}(\partial_j\Gamma(\underline{x} + \varepsilon\underline{e} - \underline{y}, t - \tau) - \partial_j\Gamma(\underline{x} - \underline{y}, t - \tau))(F(\underline{y}, \tau) - F(\underline{x}, \tau))$$

Divide the integral into the part over $|\underline{y} - \underline{x}| < \varepsilon^\beta$, $\beta < 1$ and in the remainder choosing β so that $\beta(1 + \alpha) > 1$, if α is the Hölder continuity exponent. The contribution to Δ of the term with $|\underline{y} - \underline{x}| < \varepsilon^\beta$ is then estimated by

$$\Delta \le \frac{2}{\varepsilon} \int_{|\underline{y}-\underline{x}|<2\varepsilon^\beta} \frac{C_2}{(|\underline{x} - \underline{y}|^2 + (t - \tau))^2} M_\alpha|\underline{y} - \underline{x}|^\alpha \, d\underline{y}\, d\tau \le \bar{C}_2 M_\alpha \varepsilon^{-1}\varepsilon^{(1+\alpha)\beta} \xrightarrow[\varepsilon\to 0]{} 0$$

with suitable $\bar{C}_2$, having separately bounded the $\partial_j\Gamma$, via (3.3.22).

While the contribution to the integral coming from $|\underline{y} - \underline{x}| > \varepsilon^\beta$ can be estimated by using the Taylor formula to write the integral as

$$\sum_i \int_0^1 d\sigma\, d\tau \int_{|\underline{y}-\underline{x}|\ge\varepsilon^\beta} d\underline{y}\, \partial_{ij}\Gamma(\underline{x} - \underline{y} + \sigma\, \varepsilon\, \underline{e},\, t - \tau)\, (F(\underline{y}, \tau) - F(\underline{x}, \tau))\, e_i \tag{3.3.62}$$

And note that the integrand is bounded, for all the ε, $\underline{e}$ (since $|\underline{x} - \underline{y}| \ge \varepsilon^\beta \gg \varepsilon$,because $\beta < 1$ and by using again the (3.3.21)), by: $(\frac{1}{2}|\underline{x} - \underline{y}|^2 + t - \tau)^{-5/2} M_\alpha|\underline{y} - \underline{x}|^\alpha$, *i.e.* by an ε–independent function whose integral is

$$\int_0^t d\tau \int_\Omega \frac{M_\alpha|\underline{y} - \underline{x}|^\alpha \, d\underline{y}}{(\frac{1}{2}(\underline{x} - \underline{y})^2 + t - \tau)^{5/2}} \le C \int_\Omega M_\alpha|\underline{x} - \underline{y}|^{-3+\alpha}\, d\underline{y} < \overline{C}_\alpha M_\alpha L^\alpha < \infty$$

hence the integral that expresses the differential ratio is uniformly convergent in ε and, therefore, it is possible to take the limit in (3.3.62) under the integral sign to evaluate the limit as $\varepsilon \to 0$. Thus V is differentiable and its derivative is given by $\partial_i V(\underline{x}, t) = \int d\tau \int_\Omega d\underline{y}\, \partial_{ij}\Gamma(\underline{x} - \underline{y}, t - \tau)(F(\underline{y}, \tau) - F(\underline{x}, \tau))$ and $|\partial_i V(\underline{x}, t)| \le \overline{C}_\alpha M_\alpha L^\alpha$, (note that the integral is convergent and bounded as described by the preceding majorizations).

Bibliography: Leray's theory, exposed in this section, is taken from [Le34]. Proposition VII is taken from [CF93] where it is discussed for bounded vorticity solutions in R^3 rather than in a cubic container with periodic boundary conditions.

§3.4 Fractal dimension of singularities of the Navier–Stokes equation, $d = 3$.

Here we ask which could be the structure of the possible set of the singularity points of the solutions of the Navier–Stokes equation in $d = 3$.

We have already seen in §3.3 that the set of times at which a singularity is possible has zero measure (on the time axis).

Obviously sets of zero measure can be quite structured and even large in other senses. One can think to the example of the Cantor set which is non denumerable and obtained from an interval I by deleting an open concentric subinterval of length $1/3$ that of I and then repeating recursively this operation on each of the remaining intervals (called n–th generation intervals after n steps); or one can think to the set of rational points which is dense.

(A) Dimension and measure of Hausdorff.

An interesting geometric characteristic of the size of a set is given by the Hausdorff dimension and by the Hausdorff measure, *c.f.r.* [DS60], p.174.

1. Definition *(Hausdorff α–measure): The Hausdorff α-measure of a set A contained in a metric space M is defined by considering for each $\delta > 0$ all coverings $\mathcal{C}_\delta$ of A by closed sets F with diameter $0 < d(F) \le \delta$ and setting*

$$\mu_\alpha(A) = \lim_{\delta \to 0} \inf_{\mathcal{C}_\delta} \sum_{F \in \mathcal{C}_\delta} d(F)^\alpha \tag{3.4.1}$$

Remarks:

(1) The limit over δ exists because the quantity $\inf_{\mathcal{C}_\delta} \ldots$ is monotonic non-decreasing.

(2) It is possible to show that the function defined on the sets A of M by $A \to \mu_\alpha(A)$ is completely additive on the smallest family of sets containing all closed sets and invariant with respect to the operations of complementation

and countable union (which is called the *σ-algebra Σ of the Borel sets of M*), *c.f.r.* [DS60].

One checks immediately that given $A \in \Sigma$ there is a α_c such that

$$\begin{aligned} \mu_\alpha(A) =& \infty \qquad \text{if} \quad \alpha < \alpha_c \\ \mu_\alpha(A) =& 0 \qquad \text{if} \quad \alpha > \alpha_c \end{aligned} \tag{3.4.2}$$

and it is therefore natural to set up the following definition

2. Definition *(Hausdorff measure and Hausdorff dimension): Given a Borel set $A \subset R^d$ the quantity α_c, (3.4.2), is called* Hausdorff dimension *of A, while $\mu_{\alpha_c}(A)$ is called* Hausdorff measure *of A.*

It is not difficult to check that

(1) Denumerable sets in $[0, 1]$ have zero Hausdorff dimension and measure.
(2) Hausdorff dimension of n-dimensional regular surfaces in R^d is n and, furthermore, the Hausdorff measure of their Borel subsets defines on the surface a measure μ_{α_c} that is equivalent to the area measure μ: namely there is a $\rho(x)$ such that $\mu_{\alpha_c}(dx) = \rho(x)\,\mu(dx)$.
(3) The Cantor set, defined also as the set of all numbers in $[0, 1]$ which in the representation in base 3 do not contain the digit 1, has Hausdorff dimension

$$\alpha_c = \log_3 2 \tag{3.4.3}$$

Indeed with 2^n disjoint segments with size 3^{-n}, uniquely determined (the n–th generation segments), one covers the whole set C; hence

$$\mu_{\alpha,\delta} \stackrel{def}{=} \inf_{\mathcal{C}_\delta} \sum_{F \in \mathcal{C}_\delta} d(F)^\alpha \le 1 \qquad \text{if} \quad \alpha = \alpha_0 = \log_3 2 \tag{3.4.4}$$

and $\mu_{\alpha_0}(C) \le 1$: *i.e.* $\mu_\alpha(C) = 0$ if $\alpha > \alpha_0$. Furthermore, *c.f.r.* problem [3.4.3] below, if $\alpha < \alpha_0$ one checks that the covering $\mathcal{C}^0$ realizing the smallest value of $\sum_{F \in \mathcal{C}_\delta} d(F)^\alpha$ with $\delta = 3^{-n}$ is precisely the just considered one consisting in the 2^n intervals of length 3^{-n} of the n–th generation and the value of the sum on such covering diverges for $n \to \infty$. Hence $\mu_\alpha(C) = \infty$ if $\alpha < \alpha_0$ so that $\alpha_0 \equiv \alpha_c$ and $\mu_{\alpha_c}(C) = 1$.

(B) Hausdorff dimension of singular times in the Navier–Stokes solutions ($d = 3$).

We now attempt to estimate the Hausdorff dimension of the sets of times $t \le T < \infty$ at which appear singularities of a given weak solution of Leray, *i.e.* a solution of the type discussed in §3.3, definition 2, (E). Here T is an *a priori* arbitrarily prefixed time.

For simplicity we assume that the density of volume force vanishes and we recall that in §3.3 we have shown that if at time t_0 it is $J_1(t_0) =$

$L^{-1}\int(\partial \underline{u})^2 d\underline{x} < \infty$, *i.e.* if the Reynolds number $R(t_0) = J_1(t_0)^{1/2}/V_c \equiv V_1/V_c$, *c.f.r.* (3.3.13), is $< +\infty$, then the solution stays regular in a time interval $(t_0, t_0 + \tau]$ with (see proposition II in §3.3):

$$\tau = F \frac{T_c}{R(t_0)^4 + R_g^2 + 1} \tag{3.4.5}$$

From this it will follow, see below, that there are $A > 0, \gamma > 0$ such that if

$$\liminf_{\sigma \to 0} \Big(\frac{\sigma}{T_c}\Big)^\gamma \int_{t-\sigma}^t \frac{d\vartheta}{\sigma} R^2(\vartheta) < A \tag{3.4.6}$$

then $\tau > \sigma$ and the solution is regular in an interval that contains t so that the instant t is an instant at which the solution is regular. Here, as in the following, we could fix $\gamma = 1/2$: but γ is left arbitrary in order to make clearer why the choice $\gamma = 1/2$ is the "best".

We first show that, indeed, from (3.4.6) we deduce the existence of a sequence $\sigma_i \to 0$ such that

$$\int_{t-\sigma_i}^t \frac{d\vartheta}{\sigma_i} R^2(\vartheta) < A\Big(\frac{\sigma_i}{T_c}\Big)^{-\gamma} \tag{3.4.7}$$

therefore, the l.h.s. being a time average, there must exist $\vartheta_{0i} \in (t - \sigma_i, t)$ such that

$$R^2(\vartheta_{0i}) < A\Big(\frac{\sigma_i}{T_c}\Big)^{-\gamma} \tag{3.4.8}$$

and then the solution is regular in the interval $(\vartheta_{0i}, \vartheta_{0i} + \tau_i)$ with length τ_i at least

$$\tau_i = F T_c \frac{(\sigma_i/T_c)^{2\gamma}}{A^2 + R_g^2(\sigma_i/T_c)^{2\gamma}} > \sigma_i \tag{3.4.9}$$

provided $\gamma \le 1/2$, and σ_i is small enough and if A is small enough: if $\gamma = \frac12$ we take $A^2 = \frac12 F$, for instance. Under these conditions the size of the regularity interval is longer than σ_i and *therefore it contains t itself.*

It follows that, if t is in the set S of the times at which a singularity is present, it must be

$$\liminf_{\sigma \to 0} \Big(\frac{\sigma}{T_c}\Big)^\gamma \int_{t-\sigma}^t \frac{d\vartheta}{\sigma} R^2(\vartheta) \ge A \qquad \text{if } t \in S \tag{3.4.10}$$

i.e. every singularity point is covered by a family of infinitely many intervals F with diameters σ *arbitrarily small* and satisfying

$$\int_{t-\sigma}^t d\vartheta\, R^2(\vartheta) \ge \frac{A}{2}\sigma\Big(\frac{\sigma}{T_c}\Big)^{-\gamma} \tag{3.4.11}$$

From Vitali's covering theorem (*c.f.r.* problem [3.4.1]) it follows that, given $\delta > 0$, one can find a denumerable family of intervals $F_1, F_2, \ldots$, with $F_i =$

$(t_i - \sigma_i, t_i)$, pairwise disjoint and verifying the (3.4.11) and $\sigma_i < \delta/4$, such that the intervals $5F_i \stackrel{def}{=} (t_i - 7\sigma_i/2, t_i + 5\sigma_i/2)$ (obtained by dilating the intervals F_i by a factor 5 about their center) cover S

$$S \subset \cup_i 5F_i \tag{3.4.12}$$

Consider therefore the covering $\mathcal{C}$ generated by the sets $5F_i$ and compute the sum in (3.4.1) with $\alpha = 1 - \gamma$:

$$\begin{aligned} &\sum_i (5\sigma_i)\Big(\frac{5\sigma_i}{T_c}\Big)^{-\gamma} = 5^{1-\gamma} \sum_i \sigma_i \Big(\frac{\sigma_i}{T_c}\Big)^{-\gamma} < \\ &< \frac{2\, 5^{1-\gamma}}{A} \sum_i \int_{F_i} d\vartheta\, R^2(\vartheta) \le \frac{2\, 5^{1-\gamma}}{A} \int_0^T d\vartheta\, R^2(\vartheta) < \infty \end{aligned} \tag{3.4.13}$$

where we have made use of the *a priori* estimates on vorticity derived in (3.3.8), and we must recall that $\gamma \le 1/2$ is a necessary condition in order that what has been derived be valid (*c.f.r.* comment to (3.4.9)).

Hence it is clear that for each $\alpha \ge 1/2$ it is $\mu_\alpha(S) < \infty$ (pick, in fact, $\alpha = 1 - \gamma$, with $\gamma \le 1/2$) hence the Hausdorff dimension of S is $\alpha_c \le 1/2$. Obviously the choice that gives the best regularity result (with the informations that we gathered) is precisely $\gamma = 1/2$.

Moreover one can check that $\mu_{1/2}(S) = 0$: indeed we know that S has zero measure, hence there is an open set $G \supset S$ with measure smaller than a prefixed ε. And we can choose the intervals F_i considered above so that they also verify $F_i \subset G$: hence we can replace the integral in the right hand side of (3.4.13) with the integral over G hence, since the integrand is summable, we shall find that the value of the integral can be supposed as small as wished, so that $\mu_{1/2}(S) = 0$.

(C) Hausdorff dimension in space–time of the solutions of NS, $(d = 3)$.

The problem of which is the Hausdorff dimension of the points $(\underline{x}, t) \in \Omega \times [0, T]$ which are singularity points for the Leray's solutions is quite different.

Indeed, *a priori* , it could even happen that, at one of the times $t \in S$ where the solution has a singularity as a function of time, *all* points $(\underline{x}, t)$, with $\underline{x} \in \Omega$, are singularity points and therefore the set S_0 of the singularity points thought of as a set in space–time could have dimension 3 (and perhaps even 3.5 if we take into account the dimension of the singular times discussed in (B) above).

With some optimism one can think that a version "local in space" of Leray's theorem, see proposition II of §3.3 which is "only local in time", holds. Under the influence of such wishful thought we then examine what we can expect as an estimate of the Hausdorff dimension of S_0.

The notions of characteristic time T_c, of characteristic velocities V_c, V_1, W_0, of characteristic acceleration $\sqrt{G_0}$, of characteristic size of the forcing R_g, introduced in §3.3 and playing a major role in the development of the theory of L–weak solutions of Leray, were "global" notions in the sense that they were associated with properties of the whole fluid in Ω and were not associated with any particular point or subregion of Ω.

These notions can be "localized", *i.e.* given a different meaning at different points of Ω, in a rather naive way: namely given $\underline{x} \in \Omega$ and a length scale r we can imagine that the whole fluid consists of the part that is contained in a ball $S(\underline{x}, r)$ of radius r around $\underline{x}$ and then define the various characteristic scales by replacing the container size L by r and the container Ω with $S(\underline{x}, r)$. Thus regarding a small fluid volume as in some way similar, apart from obvious scaling, to a large one is the natural idea behind the following definitions and the guesses that they inspire: it is an idea that already proved fruitful in the discussion of "large containers" in §3.3, (G).

Consider a time ϑ and define the *local Reynolds number* on scale r at time ϑ as the ratio between a quantity characterizing the velocity variation on scale r, near $\underline{x}$, and a characteristic velocity V_{cr} associated with the geometric dimension r, *c.f.r.* (3.3.1), (3.3.13). It will be the ratio between $V_{1r} \stackrel{def}{=} (r^{-1} \int_{s(\underline{x},r)} (\underline{\partial} \underset{\sim}{\underline{u}})^2 d\underline{\xi})^{1/2}$ and $V_{cr} = \nu r^{-1}$ namely

$$R_r^2(\vartheta) \stackrel{def}{=} (\frac{V_{1r}}{V_{cr}})^2 = \frac{r}{\nu^2} \int_{S(\underline{x},r)} (\underline{\partial} \underset{\sim}{\underline{u}}(\underline{\xi}, \vartheta))^2 d\underline{\xi} < \infty \tag{3.4.14}$$

where $S(\underline{x}, r)$ is the sphere of radius r and center $\underline{x}$.

Likewise, in analogy with (3.3.13), we can define the "local" strength of the forcing by "localizing" the definition in (3.3.13). Let $G_{0r} = r^{-3} \int_{S(\underline{x},r)} |\underline{g}(\underline{x})|^2 \, d\underline{x}$ be a local acceleration scale; let W_{cr} be the corresponding local velocity scale $W_{0r} = T_{cr} \sqrt{G_{0r}}$ with $T_{cr} = r^2 \nu^{-1}$, see (3.3.1), (3.3.13), and let the dimensionless strength R_{gr} of the forcing be

$$R_{gr} \stackrel{def}{=} \frac{W_{0r}}{V_{cr}} = \frac{T_{cr} \sqrt{G_{0r}}}{V_{cr}} = \frac{r^3}{\nu^2} \Big(r^{-3} \int_{S(\underline{x},r)} |\underline{g}(\underline{x})|^2 d\underline{x} \Big)^{1/2} \tag{3.4.15}$$

hence since we take $\underline{g}$ to be smooth it is $R_{gr}^2 \le C \tilde{g}^2 r^6 \nu^{-4}$ where C is a geometric constant and $\tilde{g}$ is an estimate of the maximum of the density of volume force $\underline{g}$.

Guess: *A "local version" of the Leray theorem in question, see Proposition II in Sect. §3.3 and remark (2) following it, "could" say that, under the condition (3.4.14) the solution is regular in the space–time region*

$$\underline{x} \in S(\underline{x}, r) \qquad \text{and} \qquad t \in (\vartheta, \vartheta + T_{cr} \min(1, \frac{F}{R_r^4 + R_{gr}^2})) \tag{3.4.16}$$

Obviously such a statement is very strong, far from proven in the analysis that we performed until now, and somewhat surprising: in fact since the fluid is incompressible sound waves propagate at infinite speed and therefore one can fear that a singularity that at time t is at a position $\underline{\xi}$ far from $\underline{x}$ could arrive, in an arbitrarily short time, near $\underline{x}$, even though no singularity was present near $\underline{x}$ at time t.

On the other hand it does not seem absurd that (3.4.14) or something similar to it implies (by the temporarily assumed validity of a space–local version of Leray theorem) regularity in $\underline{x}$ for a short successive time because it is also difficult that the wave associated with the singularity far away from $\underline{x}$ does not dissolve right away because of the friction. After all the theorem of Leray–Serrin, *c.f.r.* §3.3, excludes the possibility of real shock waves in an incompressible NS fluid.

Assuming, *still temporarily*, that (3.4.14) implies regularity of the solution in the vicinity of the points $(\underline{x}, t)$ with $t \in (\vartheta, \vartheta + FT_{cr}(R_r^4(\vartheta) + R_{gr}^2)^{-1})$ see (3.4.16) above, then we can use the ideas already used to estimate the time-singularities measure.

Indeed if regularity of $\underline{u}$ holds in the vicinity of $\underline{x}, \vartheta$ it follows (just from the regularity of $\underline{u}$) that $\lim_{r\to 0} \frac{\nu}{r^2} \int_{t-r^2/\nu}^t d\vartheta R_r^2(\vartheta) = 0$ (because $R_r^2(\vartheta)$ would have size $O(r^4)$).

Viceversa if the implication of (3.4.14) on regularity in the space–time set specified in (3.4.16) around $\underline{x}$ and in the time intervals $(\vartheta, \vartheta - r^2\nu^{-1})$ is accepted (keep in mind, however, that it is a property that we are considering for the sake of establishing some intuition about what should be attempted in the coming analysis), we could argue as follows.

Suppose *knowing that*

$$\limsup_{r\to 0} \frac{\nu}{r^2} \int_{t-r^2/\nu}^t d\vartheta R_r^2(\vartheta) < \varepsilon \tag{3.4.17}$$

for some $\varepsilon > 0$. Then a sequence $r_i \to 0$ would exist such that

$$\frac{\nu}{r_i^2} \int_{t-r_i^2/\nu}^t d\vartheta\, R_{r_i}^2(\vartheta) \;<\; \varepsilon \tag{3.4.18}$$

Hence, the latter expression being a time average, a time $\vartheta_i \in (t, t - r_i^2\nu^{-1})$ would exist where $R(\vartheta_i)^2 < \varepsilon$ and also, by the regularity of the external force $\underline{g}$, $R_{gr}^2 \le C^2\tilde{g}^2 r_i^6 \nu^{-4} < \varepsilon^2$ if r_i is small enough.

So that regularity would follow in the vicinity of

$$\underline{x} \times (\vartheta_i, \vartheta_i + \frac{r_i^2}{\nu}) \tag{3.4.19}$$

hence in $(\underline{x}, t)$, *provided* $2^{-1}F\varepsilon^{-2} > 1$ as a consequence of the "guess" above.

It would follow that the set S_0 of the space–time singularity points could be covered by sets $C_r = S(\underline{x}, r) \times (t - r^2\nu^{-1}, t]$ with r arbitrarily small[1] and

[1] Note that in order that this be true it suffices to require the validity of (3.4.17) with the lower limit only: here we require the upper limit because, as we shall see in the following, the (3.4.17) is, in the latter more restrictive form, closer to the property that one can really prove.

will be such that

$$\frac{1}{r\nu}\int_{t-r^2\nu^{-1}}^{t} d\vartheta \int_{S(\underline{x},r)} d\underline{x}\,(\underline{\partial}\underset{\sim}{u})^2 > \varepsilon \tag{3.4.20}$$

which is the negation of the property in (3.4.17).

Again by a covering theorem of Vitali (*c.f.r.* problems [3.4.1],[3.4.2]), we could find a family F_i of sets of the form $F_i = S(\underline{x}_i, r_i) \times (t_i - \nu^{-1} r_i^2, t_i]$ pairwise disjoint and such that the sets $5F_i$= set of points $(\underline{x}', t')$ such that $|\underline{x}' - \underline{x}_i| < 5r_i$ and $|t' - t_i - \nu^{-1}\frac{1}{2}r_i^2| < \nu^{-1}(5r_i)^2$ covers the singularity set S_0.[2] One could then estimate the sum in (3.4.1) for such a covering, by using that the sets F_i are pairwise disjoint and that $5F_i$ has diameter, if $\max r_i$ is small enough, not larger than $11r_i$:

$$\sum_i (11r_i) \le \frac{11}{\nu\varepsilon}\sum_i \int_{F_i} (\underline{\partial}\underset{\sim}{u})^2 d\underline{\xi} dt \le \frac{11}{\nu\varepsilon}\int_0^T \int_\Omega (\underline{\partial}\underset{\sim}{u})^2 d\underline{\xi} dt < \infty \tag{3.4.21}$$

i.e. the 1-measure of Hausdorff $\mu_1(S_0)$ would be $< \infty$ hence the Hausdorff dimension of S_0 would be ≤ 1.

Since S_0 has zero measure, being contained in $\Omega \times S$ where S is the set of times at which a singularity occurs somewhere, see (3.4.10), it follows (still from the covering theorems) that in fact it is possible to choose the sets F_i so that their union U is contained into an open set G which differs from S_0 by a set of measure that exceeds by as little as desired that of S_0, (which is zero); one follows the same method used above in the analysis of the time–singularity. Hence we can replace the last integral in (3.4.21) with an integral extended to the union U of the $F_i' s$: the latter integral can be made as small as wished by letting the measure of G to 0. It follows that not only the Hausdorff dimension of S_0 is ≤ 1, but also the $\mu_1(S_0) = 0$.

Remark: One could in this way exclude that the set S_0 of the space–time singularities contains a regular curve: singularities, *if existent*, cannot move along trajectories (like flow lines) otherwise the dimension of S would be $1 > 1/2$) nor they can be distributed, at fixed time, along lines and, hence, in a sense they must appear isolates and immediately disappear (always assuming their real existence).

Going back to the basic assumption behind the above wishful[3] reasoning, we realize that the condition (3.4.16) can be replaced, for the purposes of the argument discussed above and to conclude that the Hausdorff 1–measure of S_0 vanishes, by the *weaker statement* in (3.4.17) with ε small enough, that implies regularity in $(\underline{x}, t)$.

---

[2] Here the constant 5, as well as the other numerical constants that we meet below like $5, 11$ have no importance for our purposes and are just simple constants for which the estimates work.

[3] Being based on the guess above.

Or, with an obvious modification of the argument discussed above, it would suffice that the regularity in $(\underline{x}, t)$ was implied by a relation similar to (3.4.17) in which the cylinder $S(\underline{x}, r) \times (t - r^2\nu^{-1}, t]$ is replaced by a similar cylinder with $(\underline{x}, t)$ in its interior, *i.e.* if the regularity was implied by a relation of the type

$$\begin{aligned} &\limsup_{r\to 0} \frac{\nu}{r^2} \int_{t-r^2/2\nu}^{t+r^2/2\nu} R_r(\vartheta)^2 < \varepsilon, \qquad \text{or} \\ &\limsup_{r\to 0} r^{-1} \int_{t-r^2/2\nu}^{t+r^2/2\nu} \int_{S(\underline{x},r)} \frac{d\vartheta}{\nu}\, d\underline{\xi}\, (\underset{\sim}{\underline{\partial u}})^2 < \varepsilon \end{aligned} \tag{3.4.22}$$

with ε small enough.

The latter property can be actually proved to hold [CKN82]: it will be discussed in detail in §3.5.

Remark: A conjecture (much debated and that I favor) that is behind all our discussions is that *if the initial datum $\underline{u}^0$ is in $C^\infty(\Omega)$ then there exists a solution to the equation of Navier Stokes that is of class C^∞ in $(\underline{x}, t)$*", *i.e.* $S_0 = \emptyset$!

The problem is, still, open: counterexamples to the conjecture are not known (*i.e.* singular weak solutions with initial data and external force of class C^∞) but the matter is much debated and different alternative conjectures are possible (*c.f.r.* [PS87]).

In this respect one should keep in mind that if $d \geq 4$ it is possible to show that *not all* smooth initial data evolve into regular solutions: counterexamples to smoothness can indeed be constructed, *c.f.r.* [Sc77].

Problems.

[3.4.1]: (*covering theorem*, (Vitali)) Let S be an arbitrary set inside a sphere of R^n. Consider a *covering* of S with little open balls with the *Vitali property*: *i.e.* such that every point of S is contained in a family of open balls of the covering whose radii have a zero greatest lower bound. Giben $\eta > 0$ show that if $\lambda > 1$ is large enough it is possible to find a denumerable family $F_1, F_2, \ldots$ of pairwise disjoint balls of the covering with diameter $< \eta$ such that $\cup_i \lambda F_i \supset S$ where λF_i denotes the ball with the same center of F_i and radius λ times longer. Furthermore λ can be chosen independent of S, see also [3.4.2]. (*Idea:* Let $\mathcal{F}$ be the covering and let $a = max_{\mathcal{F}}\, diam(F)$. Define $a_k = a2^{-k}$ and let $\mathcal{F}_1$ be a *maximal* family of *pairwise disjoint* ball of $\mathcal{F}$ with radii $\geq a2^{-1}$ and $< a$. Likewise let $\mathcal{F}_2$ be a maximal set of balls of $\mathcal{F}$ with radii between $a2^{-2}$ and $a2^{-1}$ pairwise disjoint between themselves and with the ones of the family $\mathcal{F}_1$. Inductively we define $\mathcal{F}_1, \ldots, \mathcal{F}_k, \ldots$. It is now important to note that if $x \not\in \cup_k \mathcal{F}_k$ it must be: $distance(x, \mathcal{F}_k) < \lambda a 2^{-k}$ for some k, if λ is large enough. If indeed δ is the radius of a ball S_δ containing x and if $a2^{-k_0} \leq \delta < a2^{-k_0+1}$ then the point of S_δ farthest away from x is at most at distance $\leq 2\delta < 4a2^{-k_0}$; and if, therefore, it was $d(x, \mathcal{F}_{k_0}) \geq 4a2^{-k_0}$ we would find that the set $\mathcal{F}_{k_0}$ could be made larger by adding to it S_δ, against the maximality supposed for $\mathcal{F}_{k_0}$.)

[3.4.2]: Show that if the balls in [3.4.1] are replaced by the *parabolic cylinders* which are Cartesian products of a radius r ball in the first k coordinates and one of radius r^α, with $\alpha \geq 1$ in the $n - k$ remaining ones, then the result of problem [3.4.1] still holds if one interprets λF_i as the parabolic cylinder obtaine by applying to F_i a homothety, with respect to the center of F_i, of scale λ on the first k coordinates and λ^2 on the others. Check that if $\alpha = 2$ the value $\lambda = 5$ is sufficient.

[3.4.3]: Check that the Hausdorff dimension of the Cantor set C is $\log_3 2$, *c.f.r.* (3.4.3). *Idea:* It remains to see, given (3.4.4), that if $\alpha < \alpha_0$ then $\mu_\alpha(C) = \infty$. If $\delta = 3^{-n}$ the covering $\mathcal{C}_n$ of C with the n–th generation intervals is "the best" among those with sets of diameter $\leq 3^{-n}$ because another covering could be refined by deleting from each if its intervals the points that are out of the n–th generation intervals. Furthermore the inequality $1 < 2 \cdot 3^{-\alpha}$ for $\alpha < \log_3 2$ shows that it will not be convenient to further subdivide the intervals of $\mathcal{C}_n$ for the purpose of diminishing the sum $\sum |F_i|^\alpha$. Hence for $\delta = 3^{-n}$ the minimum value of the sum is $2^n 3^{-n\alpha} \xrightarrow[n\to\infty]{} \infty$.

Bibliography: See [DS60], vol. I, p. 174, (Hausdorff dimension and measure); [Ka76] p. 74, (Vitali covering); see also [DS60]; [Sc77],[CKN82] (fractal dimension of the singularities).

§**3.5 Local homogeneity and regularity. CKN theory.**

The theory of space–time singularities, *i.e.* the proofs of the statements that have been heuristically discussed in §3.4, will be partly based upon simple *kinematic inequalities*, which therefore have little to do with the Navier–Stokes equation, and partly they will be based on the local energy conservation which follows as a consequence of the Navier–Stokes equations.
We suppose that the volume Ω is a 3-dimensional torus (*i.e.* we assume "periodic" boundary conditions) and that the initial datum is C^∞.

(A) Energy balance for weak solutions.

Energy conservation for the regularized equations (3.3.2) says that the kinetic energy variation in a given volume element Δ of the fluid, in a time interval $[t_0, t_1]$, plus the energy dissipated therein by friction, equals the sum of the kinetic energy that in the time interval $t \in [t_0, t_1]$ enters in the volume element plus the work performed by the pressure forces (on the element boundary) plus the work of the volume forces (if any), *c.f.r.* (1.1.17). The analytic form of this relation is simply obtained by multiplying both sides of the first of the (3.3.2) by $\underline{u}$ and integrating on the volume element Δ and over the time interval $[t_0, t_1]$.
The relation that one gets can be generalized to the case in which the volume element has a time dependent shape. And an even more general relation can be obtained by multiplying both sides of (3.3.2) by $\varphi(\underline{x}, t)\underline{u}(\underline{x}, t)$ where φ is a $C^\infty(\Omega \times (0, s])$ function with $\varphi(\underline{x}, t)$ zero for t near 0 (here s is a positive parameter).
The preceding cases are obtained as limiting cases of choices of φ in the limit in which it becomes the characteristic function of the space–time volume element $\Delta \times [t_0, t_1]$. Making use of a regular function $\varphi(\underline{x}, t)$ is useful, particularly in the rather "desperate" situation in which we are when using the theory of Leray, in which the "solutions" $\underline{u}$ (obtained by removing, in (3.3.2), the regularization) are only weak solutions and, therefore, the

relations that are obtained can be interpreted as valid only after suitable integrations by parts that allow us to avoid introducing derivatives of $\underline{u}$ (whose existence is not guaranteed by the theory) at the "expense" of differentiating the "test function" φ.

We shall assume the absence of volume forces: this is a simplicity assumption as the extension of the theory to cases with time independent smooth (*e.g.* C^∞) volume forces is trivial.

Performing analytically the computation of the energy balance, described above in words, in the case of the regularized equation (3.3.2) and via a few integrations by parts[1] we get the following relation

$$\begin{aligned}&\frac{1}{2}\int_\Omega d\underline{\xi}|\underline{u}(\underline{\xi},s)|^2\varphi(\underline{\xi},s)+\nu\int_0^s dt\int_\Omega \varphi(\underline{\xi},t)|\underline{\partial}\underset{\sim}{u}(\underline{\xi},t)|^2 d\underline{x}=\\&=\int_0^s\int_\Omega\Big[\frac{1}{2}(\varphi_t+\nu\,\Delta\varphi)|\underline{u}|^2+|\underline{u}|^2\langle\underset{\sim}{u}\rangle_\lambda\cdot\underset{\sim}{\partial}\varphi+p\,\underline{u}\cdot\underline{\partial}\varphi\Big]\,dt\,d\underline{\xi}\end{aligned}\tag{3.5.1}$$

where $\varphi_t\equiv\partial_t\varphi$ and $\underline{u}=\underline{u}^\lambda$ is in fact depending also on the regularization parameter λ; here p is the pressure $p=-\sum_{ij}\Delta^{-1}\partial_i\partial_j(u_iu_j)$.

Suppose that the solution of (3.3.2) with fixed initial datum $\underline{u}_0$ converges, for $\lambda\to\infty$, to a "Leray solution" $\underline{u}$, described in (E) and definition 2 of §3.3, possibly over a subsequence $\lambda_n\to\infty$.

The (3.5.1) implies, see below, that (any, in case of non uniqueness) Leray solution $\underline{u}$ verifies the *energy inequality*:

$$\begin{aligned}&\frac{1}{2}\int_\Omega|\underline{u}(\underline{\xi},s)|^2\varphi(\underline{\xi},s)\,d\underline{\xi}+\nu\int_{t\le s}\int_\Omega\varphi(\underline{\xi},t)|\underline{\partial}\underset{\sim}{u}(\underline{\xi},t)|^2d\underline{\xi}dt\le\\&\le\int_{t\le s}\int_\Omega\Big[\frac{1}{2}(\varphi_t+\nu\,\Delta\varphi)|\underline{u}|^2+|\underline{u}|^2\underset{\sim}{u}\cdot\underset{\sim}{\partial}\varphi+p\,\underset{\sim}{u}\cdot\underset{\sim}{\partial}\varphi\Big]\,d\underline{\xi}\,dt\end{aligned}\tag{3.5.2}$$

where the pressure p is given by $p=-\sum_{ij}\Delta^{-1}\partial_i\partial_j(u_iu_j)\equiv-\Delta^{-1}\underset{\sim}{\partial}\,\underline{\partial}(\underline{u}\,\underset{\sim}{u})$ $\equiv-\Delta^{-1}(\underset{\sim}{\partial}\,\underline{u})^2$.

Remarks:

(1) It is important to note that in this relation one might expect the equal sign =: as we shall see the fact that we cannot do better than just obtaining an inequality means that the limit necessary to reach a Leray solution can introduce a "spurious dissipation" that we are simply unable to understand on the basis of what we know (today) about the Leray solutions.

(2) The above "strange" phenomenon reflects our inability to develop a complete theory of the Navier–Stokes equation, and it is possible to conjecture that no other dissipation can take place and that a (yet to come) complete theory of the equations could show this. Hence we should take

[1] As discussed in §3.3 the solutions of (3.3.2) are $C^\infty(\Omega\times[0,\infty))$ so that there is no need to justify integrating by parts.

the inequality sign in (3.5.2) as one more manifestation of the inadequacy of the Leray's solution.

(B) General Sobolev's inequalities and further a priori bounds.

The proof of (3.5.2) and of the other inequalities that we shall quote and use in this section is elementary and based, *c.f.r.* problem [3.5.15] below, on a few general "kinematic inequalities" that we now list (all of them will be used in this section although to check (3.5.2) only (S) and (CZ) are employed)

(P) *Poincarè inequality:*

$$\int_{B_r} d\underline{x}\,|f-F|^\alpha \le C^P_\alpha\, r^{3-2\alpha}\left(\int_{B_r} d\underline{x}\,|\underline{\partial} f|\right)^\alpha, \qquad 1\le\alpha\le\frac{3}{2} \tag{3.5.3}$$

where F is the average of f on the ball B_r with radius r and C^P_α is a suitable constant. We shall denote (3.5.3) by (P).

(S) *Sobolev inequality:*

$$\begin{aligned}\int_{B_r} |\underline{u}|^q\, d\underline{x} \le C^S_q \Big[\left(\int_{B_r} (\underline{\partial}\,\underset{\sim}{\underline{u}})^2\, d\underline{x}\right)^a \cdot \left(\int_{B_r} |\underline{u}|^2\, d\underline{x}\right)^{q/2-a} + \\ + r^{-2a}\left(\int_{B_r} |\underline{u}|^2\, d\underline{x}\right)^{q/2} \Big]\end{aligned} \tag{3.5.4}$$

if $2\le q\le 6,\quad a=\frac{3}{4}(q-2)$, where B_r is a ball of radius r and the integrals are performed with respect to $d\underline{x}$. The C^S_q is a suitable constant; the second term of the right hand side can be omitted if $\underline{u}$ has zero average over B_r. We shall denote (3.5.4) by (S), [So63].

(CZ) *Calderon–Zygmund inequality:*

$$\int_\Omega |\sum_{i,j}(\Delta^{-1}\partial_i\partial_j)(u_iu_j)|^q d\underline{\xi} \le C^L_q \int_\Omega |\underline{u}|^{2q} d\underline{\xi}\,, \quad 1<q<\infty \tag{3.5.5}$$

which we shall denote (CZ): here Ω is the torus of side L and C^L_q is a suitable constant, [St93].

(H) *Hölder inequality*:

$$\left|\int f_1f_2\ldots f_n\right| \le \prod_{i=1}^n \left(\int |f_i|^{p_i}\right)^{\frac{1}{p_i}}, \qquad \sum_{i=1}^n \frac{1}{p_i}=1 \tag{3.5.6}$$

which we shall denote (H): the integrals are performed over an arbitrary domain with respect to an arbitrary measure.

Remarks:

The (H) are a trivial extension of the Schwartz-Hölder inequalities; while (S) and (P) (mainly in the cases, important in what follows, $q = 6$ and $\alpha = \frac{3}{2}$) and (CZ) are less elementary and we refer to the literature, footnote at p.43 [So63], [St93], and p. 213, 219 of [LL01].

A proof of (3.5.2), given the (3.5.3)÷(3.5.6), is illustrated in problem [3.5.1]. Another important consequence of the inequalities is

I *Proposition: Let $\underline{u}$ be a Leray solution verifying (therefore) the a priori bounds in (3.3.8): $\int_\Omega |\underline{u}(\underline{x},t)|^2 d\underline{x} \le E_0$ and $\int_0^T dt \int_\Omega |\underset{\sim}{\partial}\underline{u}(\underline{x},t)|^2\, d\underline{x} \le E_0\nu^{-1}$ then*

$$\int_0^T dt \int_\Omega d\underline{x}\, |\underline{u}|^{10/3} + \int_0^T dt \int_\Omega d\underline{x}\, |p|^{5/3} \le C\nu^{-1} E_0^{5/3} \tag{3.5.7}$$

where C can be chosen $C^S_{\frac{10}{3}}\,(1 + C^L_{\frac{5}{3}})$.

proof: Apply (S) with $q = \frac{10}{3}$ and $a = 1$:

$$\begin{aligned}\int_\Omega |\underline{u}|^{\frac{10}{3}} d\underline{x} \le C^S_{\frac{10}{3}} \Big(\int_\Omega (\underset{\sim}{\partial}\underline{u})^2\, d\underline{x}\Big)^1 \cdot \Big(\int_\Omega \underline{u}^2\, d\underline{x}\Big)^{\frac{5}{3}-1} \le \\ \le C^S_{\frac{10}{3}} E_0^{\frac{2}{3}} \int_\Omega |\underset{\sim}{\partial}\underline{u}|^2\end{aligned} \tag{3.5.8}$$

hence integrating over t between 0 and T using also the second *a priori* estimate, we find

$$\int_0^T dt \int_\Omega |\underline{u}|^{\frac{10}{3}} d\underline{x} \le C^S_{\frac{10}{3}} E_0^{\frac{2}{3}} \int_0^T dt \int_\Omega d\underline{x}\, (\underset{\sim}{\partial}\underline{u})^2 \le C^S_{\frac{10}{3}} \nu^{-1} E_0^{1+\frac{2}{3}} \tag{3.5.9}$$

while the (CZ) yields: $\int_\Omega d\underline{x}\, |p|^{\frac{5}{3}} \le C^L_{\frac{5}{3}} \int_\Omega d\underline{x}\, |\underline{u}|^{\frac{10}{3}}$ which, integrated over t and combined with (3.5.9), gives the announced result.

(C) Pseudo Navier Stokes velocity–pressure pairs. Scaling operators.

As already mentioned the CKN theory will not fully use that $\underline{u}$ verifies the Navier–Stokes equation: in order to better realize this (unpleasant) property it is convenient to define separately the only properties of the Leray solutions that are really needed to develop the theory, *i.e.* to obtain an estimate of the fractal dimension of the space–time singularities set S_0. This leads to the following notion

1. Definition *(pseudo NS velocity field): Let $t \to (\underline{u}(\cdot,t), p(\cdot,t))$ be a function with values in the space of zero average square integrable "velocity" and "pressure" fields on Ω. Suppose that for each $\varphi \in C^\infty(\Omega \times (0,T])$ con*

$\varphi(\underline{x},t)$ vanishing for t near zero the following properties hold. For each $T<\infty$ and $s\le T$:

$$(a)\quad \int_\Omega \underline{u}\,d\underline{x}=\underline{0},\qquad \underline{\partial}\cdot\underline{u}=\underline{0},\qquad p=-\sum_{i,j}\partial_i\partial_j\Delta^{-1}(u_iu_j)$$

$$b)\quad \int_0^T dt\int_\Omega d\underline{x}\,|\underline{u}|^{10/3}+\int_0^T dt\int_\Omega d\underline{x}\,|p|^{5/3}<\infty \tag{3.5.10}$$

$$(c)\quad \frac{1}{2}\int_\Omega d\underline{x}\,|\underline{u}(\underline{x},s)|^2\varphi(\underline{x},s)+\nu\int_{t\le s}\int_\Omega \varphi(x,t)|\underset{\sim}{\underline{\partial u}}|^2 d\underline{x}dt\le$$
$$\le\int_{t\le s}\int_\Omega\Big[\frac{1}{2}(\varphi_t+\nu\,\Delta\varphi)|\underline{u}|^2+\frac{1}{2}|\underline{u}|^2\underset{\sim}{u}\cdot\underset{\sim}{\partial}\varphi+p\,\underset{\sim}{u}\cdot\underset{\sim}{\partial}\varphi\Big]\,d\underline{x}\,dt$$

Then we shall say that the pair $(\underline{u},p)$ is a pseudo NS velocity and pressure pair. The singularity set of $(\underline{u},p)$ will be defined as the set S_0 of the points $(\underline{x},t)\in\Omega\times[0,T]$ that do not admit a vicinity U where $|\underline{u}|$ is bounded.[2]

The remaining part of this section will concern the general properties of the pseudo NS pairs and their regularity at a given point $(\underline{x},t)$: it will not have more to do with the velocity and pressure fields that solve the Navier–Stokes equations. It is indeed easy to convince oneself that the (3.5.10), in spite of the arbitrariness of φ *are not equivalent*, not even formally, to the Navier–Stokes equations, and they pose on $\underline{u},p$ restrictions far less severe. We should not be surprised, therefore, if it turned out possible to exhibit pseudo NS pairs that really have singularities on "large sets" of space–time. In a way it is already surprising that the pseudo NS fields verify the regularity properties discussed below.

The analysis of the latter properties (of pseudo NS fields) is based on the reciprocal relations between certain quantities that we shall call "*dimensionless operators*" relative to the space–time point $(\underline{x}_0,t_0)$

2 Definition: *(dimensionless "operators" for NS) Let $(\underline{x}_0,t_0)\in\Omega\times(0,\infty)$ and suppose*

$$\begin{aligned}\Delta_r(t_0)=&\{t|\,|t-t_0|<r^2\nu^{-1}\}\\ B_r(\underline{x}_0)=&\{\underline{\xi}\,|\,|\underline{\xi}-\underline{x}_0|<r\}\equiv B_r\\ Q_r(\underline{x}_0,t_0)=&\{(\underline{\xi},\vartheta)\,|\,|\underline{\xi}-\underline{x}_0|<r,\,|\vartheta-t_0|<r^2\nu^{-1}\}=\\ =&\Delta_r(t_0)\times B_r(\underline{x}_0)\equiv Q_r\end{aligned}\tag{3.5.11}$$

define:[3]

[2] Here we mean bounded outside a set of zero measure in U or, as one says, *essentially bounded* because it is clear that, being $\underline{u},p$ in $L_2(\Omega)$, they are defined up to a set of zero measure and it would not make sense to ask that they are bounded everywhere without specifying which realization of the functions we take.

[3] If $r\ge L/2$ this is interpreted as $B_r\equiv\Omega$.

(i) "dimensionless kinetic energy operator" on scale r:

$$A(r) = \frac{1}{\nu^2 r} \sup_{|t-t_0|\le\nu^{-1}r^2} \int_{B_r} |\underline{u}(\underline{\xi},t)|^2\, d\underline{\xi} \tag{3.5.12}$$

and we say that the dimension of A is 1 : this refers to the factor r^{-1} that is used to make the integral dimensionless.
(ii) "local Reynolds number" on scale r:

$$\delta(r) = \frac{1}{\nu r} \int_{Q_r} d\vartheta d\underline{\xi}\, |\underset{\sim}{\underline{\partial}}\, \underline{u}|^2 \tag{3.5.13}$$

and we say that the dimension of δ is 1 : this refers to the factor r^{-1} that is used to make the integral dimensionless.
(iii) "dimensionless energy flux" on scale r:

$$G(r) = \frac{1}{\nu^2 r^2} \int_{Q_r} d\vartheta d\underline{\xi}\, |\underline{u}|^3 \tag{3.5.14}$$

and we say that the dimension of G is 2: this refers to the factor r^{-2} that is used to make the integral dimensionless.
(iv) "dimensionless pressure power" forces on scale r:

$$J(r) = \frac{1}{\nu^2 r^2} \int_{Q_r} d\underline{\xi} d\vartheta\, |\underline{u}|\, |p| \tag{3.5.15}$$

and we say that the dimension of J is 2 : this refers to the factor r^{-2} that is used to make the integral dimensionless.
(v) "dimensionless non locality" on scale r:

$$K(r) = \frac{r^{-13/4}}{\nu^{3/2}} \int_{\Delta_r} d\vartheta \Big(\int_{B_r} |p|\, d\underline{\xi} \Big)^{5/4} \tag{3.5.16}$$

and we say that the dimension of K is $13/4$: this refers to the factor $r^{-13/4}$ that is used to make the integral dimensionless.
(vi) "dimensionless intensity" on scale r:

$$S(r) = \nu^{-7/3} r^{-5/3} \int_{Q_r} (|\underline{u}|^{10/3} + |p|^{5/3}) d\vartheta d\underline{\xi} \tag{3.5.17}$$

where the pressure is always defined in terms of $\underline{u}$ by the expression $p = -\sum_{i,j=1}^3 \partial_i \partial_j \Delta^{-1}(u_i u_j)$. And we say that the dimension of A is $5/3$: this refers to the factor $r^{-5/3}$ that is used to make the integral dimensionless.

Remarks:

(1) The $A(r), \ldots$ *are not* operators in the common sense of functional analysis. Their name is due to their analogy with the quantities that appear in

problems that are studied with the methods of the "renormalization group" (which, also, are not operators in the common sense of the words). Perhaps a more appropriate name could be "dimensionless observables": but we shall call them operators to stress the analogy of what follows with the methods of the renormalization group.

(2) The $A(r), G(r), J(r), K(r), S(r)$ are in fact estimates of the quantities that their name evokes. We omit the qualifier "estimate" when referring to them for brevity.

(3) The interest of (i)÷(iv) becomes manifest if we note that the energy inequality (3.5.10) can be expressed in terms of such quantities if φ is suitably chosen. Indeed let

$$\varphi = \chi(\underline{x}, t)\, \frac{\exp -\big(\frac{(\underline{x}-\underline{x}_0)^2}{4(\nu(t_0-t)+2r^2)}\big)}{(4\pi\nu(t-t_0)+8\pi r^2)^{3/2}} \tag{3.5.18}$$

where $\chi(\underline{x}, t)$ is C^∞ and has value 1 if $(\underline{x}, t) \in Q_{r/2}$ and 0 if $(\underline{x}, t) \notin Q_r$. Then there exists a constant $C > 0$ such that

$$\begin{aligned} &|\varphi| < \frac{C}{r^3}, \quad |\underline{\partial}\varphi| < \frac{C}{r^4}, \qquad |\partial_t\varphi + \nu\Delta\varphi| < \frac{C}{\nu^{-1}r^5}, \qquad \text{everywhere}\\ &|\varphi| > \frac{1}{Cr^3}, \qquad\qquad\qquad\qquad\qquad \text{if} \quad (\underline{x}, t) \ni Q_{r/2} \end{aligned} \tag{3.5.19}$$

Hence (3.5.10) implies

$$\frac{\nu^2}{Cr^2}\big(A(\tfrac{r}{2}) + \delta(\tfrac{r}{2})\big) \le C\,\big(\frac{1}{\nu^{-1}r^5}\int_{Q_r} |\underline{u}|^2 + \frac{1}{r^4}\int_{Q_r} |\underline{u}|^3 + \frac{1}{r^4}\int_{Q_r} |\underline{u}||p|\big) \tag{3.5.20}$$

and, since $\int_{Q_r} |\underline{u}|^2 \le C\,(\int_{Q_r} |\underline{u}|^3)^{2/3}(\nu^{-1}r^5)^{1/3}$ with a suitable C, it follows that for some $\tilde C$

$$A(\tfrac{r}{2}) + \delta(\tfrac{r}{2}) \le \tilde C\, \big(G(r)^{2/3} + G(r) + J(r)\big) \tag{3.5.21}$$

(4) Note that the operator $\delta(r)$ is an average of the "local Reynolds' number" of §3.4, see (3.4.14), (3.4.22), which has therefore dimension 1 in the above sense.

(5) The operator (v) appears if one tries to bound $J(\frac{r}{2})$ in terms of $A(r) + \delta(r)$: such an estimate is indeed possible and it leads to the following *local Scheffer theorem*

(D) The theorems of Scheffer and of Caffarelli–Kohn–Nirenberg.

We can state the strongest results known (in general and to date) about the regularity of the weak solutions of Navier Stokes equations (which however hold also for the pseudo Navier Stokes velocity–pressure pairs).

II Theorem *(upper bound on the dimension of the sporadic set of singular times for NS, (Scheffer)): There are two constants $\varepsilon_s, C > 0$ such that if*

$G(r) + J(r) + K(r) < \varepsilon_s$ for a certain value of r, then $\underline{u}$ is bounded in $Q_{\frac{r}{2}}(\underline{x}_0, t_0)$:

$$|\underline{u}(\underline{x}, t)| \le C \frac{\varepsilon_s^{1/3}}{r}, \quad (\underline{x}, t) \in Q_{\frac{r}{2}}(\underline{x}_0, t_0), \quad \text{almost everywhere} \tag{3.5.22}$$

having set $\nu = 1$.

Remarks: (1) *c.f.r.* problems [3.5.5]÷[3.5.11] for a guide to the proof.
(2) This theorem can be conveniently combined, for the purpose of checking its hypotheses, with the inequality: $J(r) + G(r) + K(r) \le C\left(S(r)^{9/10} + S(r)^{3/4}\right)$, which follows immediately from inequality (H) and from the definitions of the operators, with a suitable C.
(3) In other words *if the operator $S(r)$ is small enough then $(\underline{x}_0, t_0)$ is a regular point.*
(4) This implies, with the *a priori* bound (3.5.7), and by a repetition of the analysis in §3.4, with the $S(r)$ playing the role of(3.4.22), that the fractal dimension of the space–time singularities set is $\le 5/3$. In fact an *a priori* estimate on the global value of an operator with dimension α implies that the Hausdorff' measure of the set of points around which the operator is large does not exceed α: in §3.4 the dimension 1 operator $\delta(r)$, *i.e.* (3.5.13) or (3.4.22), was used together with the *a priori* vorticity estimate (3.3.8) (*c.f.r.* (3.4.21)) obtaining a Hausdorff's dimension bound 1; here the operator $S(r)$ has dimension 5/3 and therefore together with the *a priori* bound (3.5.7) it yields and estimate $\le 5/3$ for the Hausdorff dimension of the singularity set. This also justifies the introduction of the operator $S(r)$.

It is now easy, in terms of the just defined operators, to illustrate the strategy of the proof of the following CKN theorem (due to Caffarelli, Kohn, Nirenberg, [CKN82]) which by the arguments in §3.4 (*c.f.r.* (3.4.22)) implies in turn that the fractal dimension of the space time singularities set S_0 for a pseudo NS field is ≤ 1 and that its 1–measure of Hausdorff $\mu_1(S_0)$ vanishes.

III Theorem: *(upper bound on the Hausdorff dimension of the sporadic singular points in space-time ("CKN theorem")) There is a constant ε_{ckn} such that if $(\underline{u}, p)$ is a pseudo NS pair of velocity and pressure fields and*

$$\limsup_{r\to 0} \frac{1}{\nu r} \int_{Q_r(\underline{x}_0, t_0)} |\underset{\sim}{\partial} \underline{u}(\underline{x}', t')|^2 \, d\underline{x}' dt' \equiv \limsup_{r \to 0} \delta(r) < \varepsilon_{ckn} \tag{3.5.23}$$

then $\underline{u}(\underline{x}', t')$, $p(\underline{x}', t')$ are C^∞ in the vicinity of $(\underline{x}_0, t_0)$.[4]

For fixed $(\underline{x}_0, t_0)$, consider the "sequence of length scales": $r_n \equiv L2^n$, with $n = 0, -1, -2, \ldots$. We shall set $\alpha_n \equiv A(r_n)$, $\kappa_n = K_n^{8/5}$, $j_n = J_n$, $g_n - G_n^{2/3}$,

[4] This means that near $(\underline{x}, t)$ the functions $\underline{u}(\underline{x}', t'), p(\underline{x}', t')$ coincide with C^∞ functions apart form a set of zero measure (recall that the pseudo NS fields are defined as fields in $L_2(\Omega)$).

$\delta_n = \delta(r_n)$ which is a natural definition as it will shortly appear. And define $\underline{X}_n \equiv (\alpha_n, \kappa_n, j_n, g_n) \in R^4_+$. Then the proof of this theorem is based on a bound that allows us to estimate the size of $\underline{X}_n$, defined as the sum of its components, in terms of the size of $\underline{X}_{n+p}$ provided the Reynolds number δ_{n+p} on scale $n+p$ is $\le \delta$.

The inequality will have the form, if $p > 0$ and $0 < \delta < 1$,

$$\underline{X}_n \le \mathcal{B}_p(\underline{X}_{n+p}; \delta) \tag{3.5.24}$$

where $\mathcal{B}_p(\cdot; \delta)$ is a map of the whole R^4_+ into itself and the inequality has to be understood "component wise", *i.e.* in the sense that each component of the l.h.s. is bounded by the corresponding component of the r.h.s. We call $|\underline{X}|$ the sum of the components of $\underline{X} \in R^4_+$.

The map $\mathcal{B}_p(\cdot; \delta)$, which to some readers will appear as strongly related to the "*beta function*" for the "running couplings" of the "renormalization group approaches",[5] will enjoy the following property

IV *Proposition: Suppose that p is large enough; given $\rho > 0$ there exists $\delta_p(\rho) > 0$ such that if $\delta < \delta_p(\rho)$ then the iterates of the map $\mathcal{B}_p(\cdot; \delta)$ contract any given ball in R^4_+, within a finite number of iterations, into the ball of radius ρ: i.e. $|\mathcal{B}^k_p(\underline{X}; \delta)| < \rho$ for all large k's.*

Assuming the above proposition theorem III follows:

proof of theorem III: Let $\rho = \varepsilon_s$, *c.f.r.* theorem II, and let p be so large that proposition IV holds. We set $\varepsilon_{ckn} = \delta_p(\varepsilon_s)$ and it will be, by the assumption (3.5.23), that $\delta_n < \varepsilon_{ckn}$ for all $n \le n_0$ for a suitable n_0 (recall that the scale labels n are negative).

Therefore it follows that $|\mathcal{B}^k_p(\underline{X}_{n_0}; \varepsilon_{ckn})| < \varepsilon_s$ for some k. Therefore by the theorem II we conclude that $(\underline{x}_0, t_0)$ is a regularity point.

(E) Proof that the renormalization map contracts.

Proposition IV follows immediately from the following general "Sobolev inequalities"

(1) *"Kinematic inequalities": i.e.* inequalities depending only on the fact that $\underline{u}$ is a divergence zero, average zero and is in $L_2(\Omega)$ and $p = -\Delta^{-1}(\underset{\sim}{\partial}\, \underline{u})^2$

$$\begin{aligned} J_n &\le C\,(2^{-p/5} A^{1/5}_{n+p} G^{1/5}_n K^{4/5}_{n+p} + 2^{2p} A^{1/2}_{n+p} \delta_{n+p}) \\ K_n &\le C\,(2^{-p/2} K_{n+p} + 2^{5p/4} A^{5/8}_{n+p} \delta^{5/8}_{n+p}) \\ G^{2/3}_n &\le C\,(2^{-2p} A_{n+p} + 2^{2p} A^{1/2}_{n+p} \delta^{1/2}_{n+p}) \end{aligned} \tag{3.5.25}$$

[5] As it relates properties of operators on a scale to those on a different scale. Note, however, that the couplings on scale n, *i.e.* the components of $\underline{X}_n$, provide information on those of X_{n+p} rather than on those of $\underline{X}_{n-p}$ as usual in the renormalization group methods, see [BG95].

where C denotes a suitable constant (*independent on the particular pseudo NS field*). The proof of the inequalities (3.5.25) is not difficult, assuming the (S,H,CZ,P) inequalities above, and it is illustrated in the problems [3.5.1], [3.5.2], [3.5.3].

(2) *"Dynamical inequality: i.e.* an inequality based on the energy inequality (c) in (3.5.10) which implies, quite easily, the following "*dynamic inequality*"[6]

$$A_n \le C\,(2^p G_{n+p}^{2/3} + 2^p A_{n+p}\delta_{n+p} + 2^p J_{n+p}) \tag{3.5.26}$$

whose proof is illustrated in problem [3.5.4].

proof of proposition IV: Assume the above inequalities (3.5.25), (3.5.26) and setting $\alpha_n = A_n, \kappa_n = K_n^{8/5}, j_n = J_n, g_n = G_n^{2/3}$, $\delta_{n+p} = \delta$ and, as above, $\underline{X}_n = (\alpha_n, \kappa_n, j_n, g_n)$. The r.h.s. of the inequalities defines the map $\mathcal{B}_p(\underline{X};\delta)$.

If one stares long enough at them one realizes that the contraction property of the proposition is an immediate consequence of

(1) The exponents to which $\varepsilon = 2^{-p}$ is raised in the various terms are either positive or not; in the latter cases the inverse power of ε is always appearing multiplied by a power of δ_{n+p} which we can take so small to compensate for the size of ε to any negative power, *except in the one case corresponding to the last term in* (3.5.26) where we see ε^{-1} without any compensating δ_{n+p}.

(2) Furthermore the sum of the powers of the components of $\underline{X}_n$ in each term of the inequalities is *always* ≤ 1: this means that the inequalities are "almost linear" and a linear map that "bounds" $\mathcal{B}_p$ exists and it is described by a matrix with small entries *except one off–diagonal element.* The iterates of the matrix therefore contract unless the large matrix element "is ill placed" in the matrix: and one easily sees that it is not.

A formal argument can be deviced in many ways: we present one in which several choices appear that are quite arbitrary and that the reader can replace with alternatives. In a way one should really try to see why a formal argument is not necessary.

The relation (3.5.26) can be "iterated" by using the expressions (3.5.25) for G_{n+p}, J_{n+p} and then the first of (3.5.25) to express $G_{n+p}^{1/5}$ in terms of A_{n+2p} with n replaced by $n+p$:

$$\begin{aligned}\alpha_n \le C\,(&2^{-p}\alpha_{n+2p} + 2^{3p}\delta_{n+2p}^{1/2}\alpha_{n+2p}^{1/2} + \\ &+ 2^{p/5}(\alpha_{n+2p}\kappa_{n+2p})^{1/2} + 2^{7p/5}\delta_{n+2p}\alpha_{n+2p}^{7/20}\kappa_{n+2p}^{1/2} + \\ &+ 2^{3p}\delta_{n+2p}\alpha_{n+2p})\end{aligned} \tag{3.5.27}$$

[6] We call it "dynamic" because it follows from the energy inequality, *i.e.* from the equations of motion.

It is convenient to take advantage of the simple inequalities $(ab)^{\frac{1}{2}} \le za + z^{-1}b$ and $a^x \le 1 + a$ for $a, b, z, x > 0$, $x \le 1$.

The (3.5.27) can be turned into a relation between α_n and $\alpha_{n+p}, \kappa_{n+p}$ by replacing p by $\frac{1}{2}p$. Furthermore, in the relation between α_n and $\alpha_{n+p}, \kappa_{n+p}$ obtained after the latter replacement, we choose $z = 2^{-p/5}$ to disentangle $2^{p/10}(\alpha_{n+p}\kappa_{n+p})^{1/2}$ we obtain recurrent (generous) estimates for α_n, κ_2

$$\begin{aligned}\alpha_n \;&\le C\,(2^{-p/10}\alpha_{n+p} + 2^{3p/10}\kappa_{n+p} + \xi^{\alpha}_{n+p})\\ \kappa_n \;&\le C\,(2^{-4p/5}\kappa_{n+p} + \xi^{\kappa}_{n+p})\\ \xi^{\alpha}_{n+p} &\stackrel{def}{=} 2^{3p}\delta_{n+p}(\alpha_{n+p} + \kappa_{n+p} + 1)\\ \xi^{\kappa}_{n+p} &\stackrel{def}{=} 2^{3p}\delta_{n+p}\alpha_{n+p}\end{aligned} \tag{3.5.28}$$

We fix p once and for all such that $2^{-p/10}C < \frac{1}{3}$.

Then if $C2^{3p}\delta_n$ is small enough, *i.e.* if δ_n is small enough, say for $\delta_n < \overline{\delta}$ *for all* $|n| \ge \overline{n}$, the matrix $M = C\begin{pmatrix} 2^{-p/10} + 2^{3p}\delta_{n+p} & 2^{3p/10} + 2^{3p}\delta_{n+p}\\ 0 & 2^{-4p/5} + 2^{3p}\delta_{n+p}\end{pmatrix}$ will have the two eigenvalues $< \frac{1}{2}$ and iteration of (3.5.27) will contract any ball in the plane α, κ to the ball of radius $2\,\overline{\delta}$.

If α_n, κ_n are bounded by a constant $\overline{\delta}$ for all $|n|$ large enough the (3.5.25) show that also g_n, j_n are going to be eventually bounded proportionally to $\overline{\delta}$.

Hence by imposing that δ is so small that $|\underline{X}_n| = |\alpha_n| + \kappa_n + j_n + g_n < \rho$ we see that proposition IV holds.

Problems. The CKN theory.

In the following problems we shall set $\nu = 1$, with no loss of generality, thus fixing the units so that time is a square length. The symbols $(\underline{u}, p)$ will denote a pseudo NS field, according to definition 1 in (C). Moreover, for notational simplicity, we shall set $A_\rho \equiv A(\rho)$, $G_\rho \equiv G(\rho), \ldots$, and sometimes we shall write $A_{r_n}, G_{r_n} \ldots$ as $A_n, \ldots$ with an abuse that should not generate ambiguities. The validity of the (3.5.10) for Leray's solution is checked in problem [3.5.15], at the end of the problems section, to stress that the theorems of Scheffer and CKN concern pseudo NS velocity–pressure fields: however it is independent of the first 14 problems.

[3.5.1]: Let $\rho = r_{n+p}$ and $r = r_n$, with $r_n = L2^n$, *c.f.r.* lines following (3.5.23), and apply (S),(3.5.4), with $q = 3$ and $a = \frac{3}{4}$, to the field $\underline{u}$, at t fixed in Δ_r and using definition 2 deduce

$$\begin{aligned}\int_{B_r} |\underline{u}|^3 d\underline{x} \le C^S_3[\Big(\int_{B_r} |\underset{\sim}{\partial}\underline{u}|^2\, d\underline{x}\Big)^{\frac{3}{4}}\Big(\int_{B_r} |\underline{u}|^2\, d\underline{x}\Big)^{\frac{3}{4}} + r^{-3/2}\Big(\int_{B_r}|\underline{u}|^2\Big)^{3/2}] \le\\ \le C^S_3[\rho^{3/4}A^{3/4}_\rho\Big(\int_{B_r}|\underset{\sim}{\partial}\underline{u}|^2 d\underline{x}\Big)^{3/4} + r^{-3/2}\Big(\int_{B_r}|\underline{u}|^2\Big)^{3/2}]\end{aligned}$$

Infer from the above the third of (3.5.25). (*Idea:* Let $\overline{|\underline{u}|^2_\rho}$ be the average of $\underline{u}^2$ on the ball B_ρ; apply the inequality (P), with $\alpha = 1$, to show that there is $C > 0$ such that

$$\int_{B_r} d\underline{x}\, |\underline{u}|^2 \le \Big(\int_{B_\rho} d\underline{x}\, \Big||\underline{u}|^2 - \overline{|\underline{u}|^2_\rho}\Big|\Big) + \overline{|\underline{u}|^2_\rho} \int_{B_r} d\underline{x} \le$$
$$\le C\rho \int_{B_\rho} d\underline{x}\, |\underline{u}||\underset{\sim}{\partial}\underline{u}| + C\Big(\frac{r}{\rho}\Big)^3 \int_{B_\rho} d\underline{x}\, |\underline{u}|^2 \le C\rho^{3/2} A_\rho^{1/2} \Big(\int_{B_\rho} d\underline{x}\, |\underset{\sim}{\partial}\underline{u}|^2\Big)^{1/2} +$$
$$+ C\Big(\frac{r}{\rho}\Big)^3 \rho A_\rho$$

where the dependence from $t \in \Delta_r$ is not explicitly indicated; hence

$$\int_{B_r} d\underline{x}\, |\underline{u}|^3 \le C\,(r\rho^{-1})^3 A_\rho^{3/2} + C\,(\rho^{3/4} + \rho^{9/4} r^{-3/2}) A_\rho^{3/4} \Big(\int_{B_\rho} d\underline{x}\, |\underset{\sim}{\partial}\underline{u}|^2\Big)^{3/4}$$

then integrate both sides with respect to $t \in \Delta_r$ and apply (H) and definition 2.)

[3.5.2]: Let $\varphi \le 1$ be a non negative C^∞ function with value 1 if $|\underline{x}| \le 3\rho/4$ and 0 if $|\underline{x}| > 4\rho/5$; we suppose that it has the "scaling" form $\varphi = \varphi_1(\underline{x}/\rho$ with $\varphi_1 \ge 0$ a C^∞ function fixed once and for all. Let B_ρ be the ball centered at $\underline{x}$ with radius ρ; and note that, if $\rho = r_{n+p}$ and $r = r_n$, pressure can be written, at each time (without explicitly exhibiting the time dependence), as $p(\underline{x}) = p'(\underline{x}) + p''(\underline{x})$ with

$$p'(\underline{x}) = \frac{1}{4\pi} \int_{B_\rho} \frac{1}{|\underline{x} - \underline{y}|} p(y) \Delta\varphi(\underline{y})\, d\underline{y} + \frac{1}{2\pi} \int_{B_\rho} \frac{\underline{x} - \underline{y}}{|\underline{x} - \underline{y}|^3} \cdot \underline{\partial}\varphi(\underline{y})\, p(\underline{y})\, d\underline{y}$$
$$p''(\underline{x}) = \frac{1}{4\pi} \int_{B_\rho} \frac{1}{|\underline{x} - \underline{y}|} \varphi(\underline{y})\, (\underline{\partial}\underline{u}(\underline{y})) \cdot (\underline{\partial}\underset{\sim}{u}(\underline{y}))\, d\underline{y}$$

if $|\underline{x}| < 3\rho/4$; and also $|p'(\underline{x})| \le C\rho^{-3} \int_{B_\rho} d\underline{y}\, |p(\underline{y})|$ and all functions are evaluated at a fixed $t \in \Delta_r$. Deduce from this remark the first of the (3.5.25). (*Idea:* First note the identity $p = -(4\pi)^{-1} \int_{B_\rho} |\underline{x} - \underline{y}|^{-1}\, \Delta\,(\varphi\, p)$ for $\underline{x} \in B_r$ because if $\underline{x} \in B_{3\rho/4}$ it is $\varphi\, p \equiv p$. Then note the identity $\Delta\,(\varphi\, p) = p\,\Delta\,\varphi + 2\underline{\partial}\, p \cdot \underline{\partial}\varphi + \varphi\,\Delta\, p$ and since $\Delta\, p = -\underline{\partial} \cdot (\underset{\sim}{u} \cdot \underset{\sim}{\partial}\, \underline{u}) = -(\underline{\partial}\, \underset{\sim}{u}) \cdot (\underset{\sim}{\partial}\, \underline{u})$: the second of the latter relations generates p'' while $p\Delta\,\varphi$ combines with the contribution from $2\underline{\partial}\, p \cdot \underline{\partial}\varphi$, after integrating the latter by parts, and generates the two contributions to p'.
From the expression for p'' we see that

$$\int_{B_r} d\underline{x}\, |p''(\underline{x})|^2 \le \int_{B_\rho \times B_\rho} d\underline{y}\, d\underline{y}'\, |\underset{\sim}{\partial}\underline{u}(\underline{y})|^2 |\underset{\sim}{\partial}\underline{u}(\underline{y}')|^2 \int_{B_r} d\underline{x} \frac{1}{|\underline{x} - \underline{y}||\underline{x} - \underline{y}'|} \le$$
$$\le C\rho (\int_{B_\rho} d\underline{y}\, |\underset{\sim}{\partial}\underline{u}(\underline{y})|^2)^2 \qquad (!)$$

The part with p' is more interesting: since its integral expression above contains inside the integral kernels apparently singular at $\underline{x} = \underline{y}$ like $|\underline{x} - \underline{y}|^{-1}\, \Delta\,\varphi$ and $|\underline{x} - \underline{y}|^{-1}\, \underline{\partial}\,\varphi$ one notes that this is not true because the derivatives of φ vanish if $\underline{y} \in B_{3\rho/4}$ (where $\varphi \equiv 1$) so that $|\underline{x} - \underline{y}|^{-k}$ can be bounded "dimensionally" by ρ^{-k} in the whole region $B_\rho/B_{3\rho/4}$ for all $k \ge 0$ (this remark also shows why one should think p as sum of p' and p'').
Thus replacing the (apparently) singular kernels with their dimensional bounds we get

$$\int_{B_r} d\underline{x}\, |\underline{u}||p'| \le \frac{C}{\rho^3} \Big(\int_{B_r} d\underline{x}\, |\underline{u}|\Big) \cdot \Big(\int_{B_\rho} d\underline{x}\, |p|\Big)$$

which can be bounded by using inequality (H) as

$$\begin{aligned}&\le \frac{C}{\rho^3}\Big(\int_{B_r} d\underline{x}\,|\underline{u}|^{2/5}\cdot|\underline{u}|^{3/5}\cdot 1\Big)\cdot\Big(\int_{B_\rho} d\underline{x}\,|p|\Big)\le\\ &\le \frac{C}{\rho^3}\Big(\int_{B_r} d\underline{x}\,|\underline{u}|^2\Big)^{1/5}\cdot\Big(\int_{B_r} d\underline{x}\,|\underline{u}|^3\Big)^{1/5}(r^3)^{3/5}\cdot\int_{B_\rho} d\underline{x}\,|p|\le\\ &\le \frac{C\,r^{9/5}}{\rho^3}\,(\rho A_\rho)^{1/5}\Big(\int_{B_r} d\underline{x}\,|\underline{u}|^3\Big)^{1/5}\cdot\Big(\int_{B_\rho} d\underline{x}\,|p|\Big)\end{aligned}$$

where all functions depend on $\underline{x}$ (besides t) and then, integrating over $t\in\Delta_r$ and dividing by r^2 one finds, for a suitable $C>0$:

$$\frac{1}{r^2}\int_{Q_r} dtd\underline{x}\,|\underline{u}||p'|\le C\,(\frac{r}{\rho})^{1/5}G_r^{1/5}K_\rho^{4/5}A_\rho^{1/5}$$

that is combined with $\int_{B_r} d\underline{x}|\underline{u}||p''|\le(\int_{B_r} d\underline{x}\,|\underline{u}|^2)^{1/2}(\int_{B_r} d\underline{x}\,|p''|^2)^{1/2}$ which, integrating over time, dividing by ρ^2 and using inequality (!) for $\int_{B_r} d\underline{x}\,|p''|^2$ yields: $r^{-2}\int_{Q_r} dtd\underline{x}\,|\underline{u}||p''|\le C(\rho r^{-1})^2A_\rho^{1/2}\delta_\rho)$.

[3.5.3] In the context of the hint and notations for p of the preceding problem check that $\int_{B_r} d\underline{x}\,|p'|\le C(r\rho^{-1})^3\int_{B_\rho} d\underline{x}|p|$. Integrate over t the power $5/4$ of this inequality, rendered adimensional by dividing it by $r^{13/4}$; one gets: $r^{-13/4}\int_{\Delta_r}(\int|p'|)^{5/4}\le C(r\rho^{-1})^{1/2}K_\rho$, which yields the first term of the second inequality in (3.5.25). Complete the derivation of the second of (3.5.25). (*Idea:* Note that $p''(\underline{x},t)$ can be written, in the interior of B_r, as $p''=\tilde p+\hat p$ con:

$$\tilde p(\underline{x})=-\frac{1}{4\pi}\int_{B_\rho}\frac{\underline{x}-\underline{y}}{|\underline{x}-\underline{y}|^3}\varphi(\underline{y})\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{u}\,d\underline{y},\qquad \hat p(\underline{x})=-\frac{1}{4\pi}\int_{B_\rho}\frac{\underline{\partial}\varphi(\underline{y})\cdot(\underset{\sim}{u}\cdot\underset{\sim}{\partial})\underline{u}}{|\underline{x}-\underline{y}|}\,d\underline{y}$$

(always at fixed t and not declaring explicitly the t–dependence). Hence by using $|\underline{x}-\underline{y}|>\rho/4$, for $\underline{x}\in B_r$ and $\underline{y}\in B_\rho/B_{3\rho/4}$, *i.e.* for $\underline{y}$ in the part of B_ρ where $\underline{\partial}\,\varphi\ne\underline{0}$) we find

$$\begin{aligned}\int_{B_r}|\tilde p|\,d\underline{x}&\le C\int_{B_\rho} d\underline{y}\Big(\int_{B_r}\frac{d\underline{x}}{|\underline{x}-\underline{y}|^2}\,|\underline{u}(\underline{y})|\,|\underset{\sim}{\partial}\,\underset{\sim}{u}\,(\underline{y})|\Big)\le\\ &\le Cr\Big(\int_{B_\rho}|\underline{u}|^2\Big)^{1/2}\Big(\int_{B_\rho}|\underset{\sim}{\partial}\underline{u}|^2\Big)^{1/2}\le Cr\rho^{1/2}A_\rho^{1/2}\Big(\int_{B_\rho}|\underset{\sim}{\partial}\underline{u}|^2\Big)^{1/2}\\ \int_{B_r}|\hat p|\,d\underline{x}&\le C\frac{r^3}{\rho^2}\int_{B_\rho}|\underline{u}||\underset{\sim}{\partial}\underline{u}|\le C\,r\rho^{1/2}A_\rho^{1/2}\Big(\int_{B_\rho}|\underset{\sim}{\partial}\underline{u}|^2\Big)^{1/2}\end{aligned}$$

and $\Big(\int_{B_r}|p''|\Big)^{5/4}$ is bounded by raising the right hand sides of the last inequalities to the power $5/4$ and integrating over t, and finally applying inequality (H) to generate the integral $\Big(\int_{Q_\rho}|\underset{\sim}{\partial}\underline{u}|^2\Big)^{5/8}$).

[3.5.4]: Deduce that (3.5.26) holds for a pseudo–NS field $(\underline{u},p)$, *c.f.r.* definition 1. (*Idea:* Let $\varphi(\underline{x},t)$ be a C^∞ function which is 1 on $Q_{\rho/2}$ and 0 outside Q_ρ; it is: $0\le\varphi(\underline{x},t)\le 1$,

$|\underline{\partial}\varphi| \le \frac{C}{\rho}$, $|\Delta\varphi + \partial_t\varphi| \le \frac{C}{\rho^2}$, if we suppose that φ has the form $\varphi(\underline{x}, t) = \varphi_2(\frac{\underline{x}}{\rho}, \frac{t}{\rho^2}) \ge 0$ for some φ_2 suitably fixed and smooth. Then, by applying the third of (3.5.10) and using the notations of the preceding problems, if $\bar{t} \in \Delta_{\rho/2}(t_0)$, it is

$$\begin{aligned}
&\int_{B_r\times\{\bar t\}} |\underline{u}(\underline{x}, t)|^2 d\underline{x} \le \frac{C}{\rho^2}\int_{Q_\rho} dtd\underline{x}\, |\underline{u}|^2 + \int_{Q_\rho} dtd\underline{x}\, (|\underline{u}|^2 + 2p)\, \underline{u}\cdot\underline{\partial}\varphi \le \\
&\le \frac{C}{\rho^2}\int_{Q_\rho} dtd\underline{x}\, |\underline{u}|^2 + \Big|\int_{Q_\rho} dtd\underline{x}\, (|\underline{u}|^2 - \overline{|\underline{u}|^2_\rho})\underline{u}\cdot\underline{\partial}\varphi\Big| + 2\int_{Q_\rho} dtd\underline{x}\, p\, \underline{u}\cdot\underline{\partial}\varphi \le \\
&\le \frac{C}{\rho^{1/3}}\Big(\int_{Q_\rho} dtd\underline{x}\, |\underline{u}|^3\Big)^{2/3} + \Big|\int_{Q_\rho} dtd\underline{x}\, \Big(|\underline{u}|^2 - \overline{|\underline{u}|^2_\rho}\Big)\, \underline{u}\cdot\underline{\partial}\varphi\Big| + \frac{2C}{\rho}\int_{B_\rho} dtd\underline{x}\, |p||\underline{u}| \le \\
&\le C\rho G_\rho^{2/3} + C\rho J_\rho + \rho\Big|\frac{1}{\rho}\int_{Q_\rho} dtd\underline{x}\, (|\underline{u}|^2 - \overline{|\underline{u}|^2_\rho})\, \underline{u}\cdot\underline{\partial}\varphi\Big| \qquad (*)
\end{aligned}$$

We now use the following inequality, at t fixed and with the integrals over $d\underline{x}$

$$\begin{aligned}
&\frac{1}{\rho}\Big|\int_{B_\rho} d\underline{x}\, (|\underline{u}|^2 - \overline{|\underline{u}|^2_\rho})\underline{u}\cdot\underline{\partial}\varphi\Big| \le \frac{C}{\rho^2}\int_{B_\rho} d\underline{x}\, |\underline{u}|\, \Big||\underline{u}|^2 - \overline{|\underline{u}|^2_\rho}\Big| \le \\
&\le \frac{C}{\rho^2}\Big(\int_{B_\rho} d\underline{x}\, |\underline{u}|^3\Big)^{1/3}\Big(\int_{B_\rho} |\underline{u}^2 - \overline{|\underline{u}|^2_\rho}|^{3/2}\Big)^{2/3}
\end{aligned}$$

and we also take into account inequality (P) with $f = \underline{u}^2$ and $\alpha = 3/2$ which yields (always at t fixed and with integrals over $d\underline{x}$):

$$\Big(\int_{B_\rho} \Big|\underline{u}^2 - \overline{|\underline{u}|^2_\rho}\Big|^{3/2}\Big)^{2/3} \le C\Big(\int_{B_\rho} |\underline{u}||\underline{\partial}\underline{u}|\Big)$$

then we see that

$$\begin{aligned}
&\int_{B_\rho} \Big||\underline{u}|^2 - \overline{|\underline{u}|^2_\rho}\Big|\, |\underline{u}|\, |\underline{\partial}\varphi| \le \frac{C}{\rho}\Big(\int_{B_\rho} |\underline{u}|^3\Big)^{1/3}\Big(\int_{B_\rho} |\underline{u}||\underline{\partial}\underline{u}|\Big) \le \\
&\le \frac{C}{\rho}\Big(\int_{B_\rho} |\underline{u}|^3\Big)^{1/3}\Big(\int_{B_\rho} |\underline{u}|^2\Big)^{1/2}\Big(\int_{B_\rho} |\underline{\partial}\underline{u}|^2\Big)^{1/2} \le \\
&\le \frac{C}{\rho}\rho^{1/2}A_\rho^{1/2}\Big(\int_{B_\rho} |\underline{u}|^3\Big)^{1/3}\cdot\Big(\int_{B_\rho} |\underline{\partial}\underline{u}|^2\Big)^{1/2}\cdot 1
\end{aligned}$$

Integrating over t and applying (H) with exponents $3, 2, 6$, respectively, on the last three factors of the right hand side we get

$$\frac{1}{\rho^2}\int_{Q_\rho} |\underline{u}|\,\Big||\underline{u}|^2 - \overline{|\underline{u}|^2_\rho}\Big| \le C A_\rho^{1/2} G_\rho^{1/3}\delta_\rho^{1/2} \le C\,(G_\rho^{2/3} + A_\rho\delta_\rho)$$

and placing this in the first of the preceding inequalities (*) we obtain the desired result).

The following problems provide a guide to the proof of theorem II. Below we replace, unless explicitly stated the sets B_r, Q_r, Δ_r *introduced in definition 2, in (C) above, and employed in the previous problems with* B_r^0, Δ_r^0, Q_r^0 *with* $B_r^0 = \{\underline{x}|\, |\underline{x} - \underline{x}_0| < r\}$,

$\Delta^0_r = \{t \,|\, t_0 > t > t - r^2\}$, $Q^0_r = \{(\underline{x}, t) | \, |\underline{x} - \underline{x}_0| < r,\, t_0 > t > t - r^2\} = B^0_r \times \Delta^0_r$. *Likewise we shall set* $B^0_{r_n} = B^0_n, \Delta^0_{r_n} = \Delta^0_n, Q^0_{r_n} = Q^0_n$ *and we shall define new operators* A, δ, G, J, K, S *by the same expressions in (3.5.11)%(3.5.17) in (C) above but with the just defined new meaning of the integration domains. However, to avoid confusion, we shall call them* $A^0, \delta^0, \ldots$ *with a superscript* 0 *added.*

[3.5.5]: With the above conventions check the following inequalities

$$A^0_n \le C A^0_{n+1}, \qquad G^0_n \le C {G^0}_{n+1}, \qquad G^0_n \le C\,({A^0_n}^{3/2} + {A^0_n}^{3/4} {\delta^0_n}^{3/4})$$

(*Idea:* The first two are trivial consequences of the fact that the integration domains of the right hand sides are larger than those of the left hand sides, and the radii of the balls differ only by a factor 2 so that C can be chosen 2 in the first inequality and 4 in the second. The third inequality follows from (S) with $a = \frac{3}{4}$, $q = 3$:

$$\int_{B^0_r} |\underline{u}|^3 \le C\left[\Big(\int_{B^0_r} |\underline{\partial}\underline{u}|^2\Big)^{3/4} \Big(\int_{B^0_r} |\underline{u}|^2\Big)^{3/4} + r^{-3/2}\Big(\int_{B^0_r} |\underline{u}|^2\Big)^{3/2}\right] \le$$
$$\le C\left[r^{3/4} {A^0_r}^{3/4} \Big(\int_{B^0_r} |\underline{\partial}\underline{u}|^2\Big)^{3/4} + {A^0_r}^{3/2}\right]$$

where the integrals are over $d\underline{x}$ at t fixed; and integrating over t we estimate G^0_r by applying (H) to the last integral over t.)

[3.5.6]: Let $n_0 = n + p$ and $Q^0_n = \{(\underline{x}, t) | \, |\underline{x} - \underline{x}_0| < r_n,\, t_0 > t > t - r^2_n\} \stackrel{def}{=} B^0_n \times \Delta^0_n$ consider the function:

$$\varphi_n(\underline{x}, t) = \frac{\exp(-(\underline{x} - \underline{x}_0)^2 / 4(r^2_n + t_0 - t))}{(4\pi(r^2_n + t_0 - t))^{3/2}}, \qquad (\underline{x}, t) \in Q^0_{n_0}$$

and a function $\chi_{n_0}(\underline{x}, t) = 1$ on $Q^0_{n_0 - 1}$ and 0 outside $Q^0_{n_0}$, for instance choosing,, a function which has the form $\chi_{n_0}(\underline{x}, t) = \tilde{\varphi}(r^{-1}_{n_0} \underline{x}, r^{-1/2}_{n_0} t) \ge 0$, with $\tilde{\varphi}$ a C^∞ function fixed once and for all. Then write (3.5.10) using $\varphi = \varphi_n \chi_{n_0}$ and deduce the inequality

$$\frac{A^0_n + \delta^0_n}{r^2_n} \le C\Big[r^{-2}_{n+p} {G^0}^{2/3}_{n+p} + \sum_{k=n+1}^{n+p} r^{-2}_k G^0_k + r^{-2}_{n+p} {J^0}_{n+p} + \sum_{k=n+1}^{n+p-1} r^{-2}_k L_k\Big] \qquad (@)$$

where $L_k = r^{-2}_k \int_{Q^0_k} d\underline{x}\, dt\, |\underline{u}| |p - \overline{p^k}|$ with $\overline{p^k}$ equals the average of p on the ball B^0_k; for each $p > 0$. (*Idea:* Consider the function φ and note that $\varphi \ge (C r^3_n)^{-1}$ in Q^0_n, which allows us to estimate *from below* the left hand side term in (3.5.10), with $(C r^2_n)^{-1}(A^0_n + {\delta^0}_n)$. Moreover one checks that

$$|\varphi| \le \frac{C}{r^3_m}, \qquad |\underline{\partial}\varphi| \le \frac{C}{r^4_m}, \quad n \le m \le n + p \equiv n_0, \quad \text{in} \quad Q^0_{m+1}/Q^0_m$$
$$|\partial_t \varphi + \Delta\varphi| \le \frac{C}{r^5_{n_0}} \qquad \text{in} \quad Q^0_{n_0}$$

and the second relation follows noting that $\partial_t \varphi + \Delta\varphi \equiv 0$ in the "dangerous places", *i.e.* $\chi = 1$, because φ is a solution of the heat equation (backward in time). Hence the first term in the right hand side of (3.5.10) can be bounded *from above* by

$$\int_{Q^0_{n_0}} |\underline{u}|^2 |\partial_t \varphi_n + \Delta\varphi_n| \le \frac{C}{r^5_{n_0}} \int_{Q^0_{n_0}} |\underline{u}|^2 \le \frac{C}{r^5_{n_0}} \Big(\int_{Q^0_{n_0}} |\underline{u}|^3\Big)^{2/3} r^{5/3}_{n_0} \le \frac{C}{r^2_{n_0}} {G^0}^{2/3}_{n_0}$$

getting the first term in the r.h.s. of (@).

Using here the scaling properties of the function φ the second term is bounded by

$$\int_{Q^0_{n_0}} |\underline{u}|^3 |\underline{\partial}\varphi_n| \le \frac{C}{r_n^4}\int_{Q^0_{n+1}} |\underline{u}|^3 + \sum_{k=n+2}^{n_0} \frac{C}{r_k^4}\int_{Q^0_k/Q^0_{k-1}} |\underline{u}|^3 \le$$
$$\le \sum_{k=n+1}^{n_0} \frac{C}{r_k^4}\int_{Q^0_k} |\underline{u}|^3 \le C\sum_{k=n+1}^{n_0} \frac{G^0_k}{r_k^2}$$

Calling the third term (*c.f.r.* (3.5.1)) Z we see that it is bounded by

$$Z \le \Big|\int_{Q^0_{n_0}} p\,\underline{u}\cdot\underline{\partial}\,\chi_{n_0}\varphi_n\Big| \le \Big|\int_{Q^0_{n+1}} p\,\underline{u}\cdot\underline{\partial}\,\chi_{n+1}\varphi_n\Big| +$$
$$+\sum_{k=n+2}^{n_0}\Big|\int_{Q^0_k} p\,\underline{u}\cdot\underline{\partial}\,(\chi_k-\chi_{k-1})\varphi_n\Big| \le \Big|\int_{Q^0_{n+1}} (p-\overline{p^{n+1}})\,\underline{u}\cdot\underline{\partial}\,\chi_{n+1}\varphi_n\Big| +$$
$$+\sum_{k=n+2}^{n_0-1}\Big|\int_{Q^0_k} (p-\overline{p^k})\,\underline{u}\cdot\underline{\partial}\,(\chi_k-\chi_{k-1})\varphi_n\Big| + \int_{Q^0_{n_0}} |\underline{u}|\,|p|\,|\underline{\partial}(\chi_{n_0}-\chi_{n_0-1})\varphi_n$$

where $\overline{p^m}$ denotes the average of p over B^0_m (which only depends on t): the possibility of replacing p by $p-\overline{p}$ in the integrals is simply due to the fact that the 0 divergence of $\underline{u}$ allows us to ad to p *any* constant because, by integration by parts, it will contribute 0 to the value of the integral.

From the last inequality it follows

$$Z \le \sum_{k=n+1}^{n_0-1} \frac{C}{r_k^4}\int_{Q^0_k} |p-\overline{p^k}|\,|\underline{u}| + J^0_{n_0} r_{n_0}^{-2} = \sum_{k=n+1}^{n_0-1}\frac{C}{r_k^2}L_k + J^0_{n_0} r_{n_0}^{-2}$$

then sum the above estimates.)

[3.5.7] If $\underline{x}_0$ the center of Ω the function $\chi_{n_0}p$ can be regarded, if $n_0 < -1$, as defined on the whole R^3 and zero outside the torus Ω. Then if Δ is the Laplace operator *on the whole* R^3 note that the expression of p in terms of $\underline{u}$ (*c.f.r.* (a) of (3.5.10)) implies that in $Q^0_{n_0}$:

$$\chi_{n_0}p = \Delta^{-1}\Delta\chi_{n_0}p \equiv \Delta^{-1}\Big(p\Delta\chi_{n_0} + 2(\underline{\partial}\chi_{n_0})\cdot(\underline{\partial}p) - \chi_{n_0}\underset{\sim}{\underline{\partial}\underline{\partial}}\cdot(\underset{\sim}{\underline{u}\,\underline{u}})\Big)$$

Check that this expression can be rewritten, for $n < n_0$, as

$$-\underset{\sim}{\underline{\partial}\underline{\partial}}\Delta^{-1}(\chi_{n_0}\underset{\sim}{\underline{u}\underline{u}}\vartheta_{n+1}) - \underset{\sim}{\underline{\partial}\underline{\partial}}\Delta^{-1}(\chi_{n_0}\underset{\sim}{\underline{u}\underline{u}}(1-\vartheta_{n+1}))-$$
$$-2\underline{\partial}\Delta^{-1}(\underset{\sim}{\underline{\partial}}\chi_{n_0}\underset{\sim}{\underline{u}\underline{u}}) - \Delta^{-1}((\underset{\sim}{\underline{\partial}\underline{\partial}}\chi_{n_0})\underset{\sim}{\underline{u}\underline{u}})-$$
$$-\Delta^{-1}((\Delta\chi_{n_0})p) - 2\underline{\partial}\Delta^{-1}(\underline{\partial}\chi_{n_0}\,p)$$

where ϑ_k is the characteristic function of $B^0{}_k$.

[3.5.8] In the context of the previous problem check that in $Q^0_{n_0}$ it is $p=p_1+p_2+p_3+p_4$ with

$$p_1=-\,(\underset{\sim}{\partial}\underline{\partial}\Delta^{-1})\cdot\big(\chi_{n_0}\vartheta_{n+1}\underline{u}\,\underset{\sim}{u}\big),\qquad p_2=-\frac{1}{4\pi}\int_{B^0{}_{n_0}/B^0{}_{n+1}}\big(\underset{\sim}{\underline{\partial}\partial}\frac{1}{|\underline{x}-\underline{y}|}\big)\cdot\chi_{n_0}\underline{u}\,\underset{\sim}{u}$$

$$p_3=-\,\frac{1}{2\pi}\int_{B^0_{n_0}}\frac{\underline{x}-\underline{y}}{|\underline{x}-\underline{y}|^3}(\underset{\sim}{\partial}\chi_{n_0})\underline{u}\,\underset{\sim}{u}+\frac{1}{4\pi}\int_{B^0_{n_0}}\frac{1}{|\underline{x}-\underline{y}|}(\underset{\sim}{\partial}\underline{\partial}\chi_{n_0})\underset{\sim}{u}\,\underline{u}$$

$$p_4=\frac{1}{4\pi}\int\frac{1}{|\underline{x}-\underline{y}|}p(\underline{y})\Delta\chi_{n_0}+\frac{2}{4\pi}\int p(\underline{y})\frac{\underline{x}-\underline{y}}{|\underline{x}-\underline{y}|^3}\cdot\underline{\partial}\chi_{n_0}$$

where $n<n_0$ and the integrals are over $\underline{y}$ at t fixed, and the functions in the left hand side are evaluated in $\underline{x},t$.

[3.5.9]: Consider the quantity L_n introduced in [3.5.6] and show that, setting $n_0=n+p$, $p>0$, it is

$$L_n\le C\Big[\Big(\frac{r_{n+1}}{r_{n_0}}\Big)^{7/5}A^0_{n+1}{}^{1/5}G^0_{n+1}{}^{1/5}K^0_{n_0}{}^{4/5}+\Big(\frac{r_{n+1}}{r_{n_0}}\Big)^{5/3}G^0_{n+1}{}^{1/3}G^0_{n_0}{}^{2/3}+$$
$$+G^0_{n+1}+r^3_{n+1}G^0_{n+1}{}^{1/3}\sum_{k=n+1}^{n_0}r_k^{-3}A^0_k\Big]$$

(*Idea:* Refer to [3.5.8] to bound L_n by: $\sum_{i=1}^4 r_n^{-2}\int_{Q^0_n}|\underline{u}||p_i-\overline{p_i^n}|$ where $\overline{p_i^n}$ is the average of p_i over B^0_n; and estimate separately the four terms. for the first it is not necessary to subtract the average and the difference $|p_1-\overline{p_1^n}|$ can be divided into the sum of the absolute values each of which contributes equally to the final estimate which is obtained via the (CZ), and the (H)

$$\int_{B^0_{n+1}}|p_1-\overline{p_1}||\underline{u}|\le2\Big(\int_{B^0_{n+1}}|p_1|^{3/2}\Big)^{2/3}\Big(\int_{B^0_{n+1}}|\underline{u}|^3\Big)^{1/3}\le C\int_{B^0_{n+1}}|\underline{u}|^3$$

and the contribution of p_1 at L_n is bounded, therefore, by CG^0_{n+1}: note that this would not be true with p instead of p_1 because in the right hand side there would be $\int_\Omega|p|^{3/2}$ rather than $\int_{B^0_{n+1}}|p|^{3/2}$, because the (CZ) is a "nonlocal" inequality. The term with p_2 is bounded as

$$\int_{\Delta^0_n}\int_{B^0_n}|p_2-\overline{p_2^n}||\underline{u}|\le\int_{\Delta^0_n}\int_{B^0_n}|\underline{u}|\,r_n\max|\underline{\partial}p_2|\le$$
$$\le r_n\Big(\int_{Q^0_n}\frac{|\underline{u}|^3}{r_n^2}\Big)^{1/3}r_n^{2/3}r_n^{10/3}\max_{Q^0_n}|\underline{\partial}p_2|\le$$
$$\le r_n^5G^0_n{}^{1/3}\sum_{m=n+1}^{n_0-1}\max_{t\in\Delta^0_m}\int_{B^0_{m+1}/B^0_m}\frac{|\underline{u}|^2}{r_m^4}=r_n^5G^0_n{}^{1/3}\sum_{m=n+1}^{n_0}\frac{A^0_m}{r_m^3}$$

Analogously the term with p_3 is bounded by using $|\underline{\partial}p_3|\le Cr_{n_0}^{-4}\int_{B^0_{n_0}}|\underline{u}|^2$ which is majorized by $Cr_{n_0}^{-3}(\int_{B^0_{n_0}}|\underline{u}|^3)^{2/3}$ obtaining

$$\frac{1}{r_n^2}\int_{Q^0_n}|\underline{u}||p_3-\overline{p_3^n}|\le\frac{C}{r_n^2}r_{n_0}^{-3}\int_{\Delta^0_n}[(\int_{B^0_n}|\underline{u}|^3)^{2/3}r_n\int_{B^0_n}|\underline{u}|]\le$$
$$\le\frac{C}{r_n^2}r_n^3r_{n_0}^{-3}\int_{\Delta^0_n}(\int_{B^0_n}|\underline{u}|^3)^{2/3}(\int_{B^0_n}|\underline{u}|^3)^{1/3}\le$$
$$\le\frac{C}{r_n^2}(\frac{r_n}{r_{n_0}})^3r_{n_0}^{4/3}r_n^{2/3}G^0_{n_0}{}^{2/3}G^0_n{}^{1/3}=C(\frac{r_n}{r_{n_0}})^{5/3}G^0_{n_0}{}^{2/3}G^0_n{}^{1/3}$$

Finally the term with p_4 is bounded (taking into account that the derivatives $\Delta\chi_n, \underline{\partial}\chi_n$ vanish where the kernels become bigger than what suggested by their dimension) by noting that

$$\int_{B_n^0} |p_4 - \overline{p_4^n}||\underline{u}| \le C r_n \int_{B_n^0} |\underline{u}| \max_{B_n^0} |\underline{\partial} p_4| \le C r_n \Big(\int_{B_n^0} |\underline{u}|\Big)\Big(\int_{B_{n_0}^0} \frac{|p|}{r_{n_0}^4}\Big)$$

Denoting with $\tilde K_{n_0}^0$ the operator $K_{n_0}^0$ without the factor $r_{n_0}^{-13/4}$ which makes it dimensionless, and introducing, similarly, $\tilde A_n^0, \tilde G_n^0$) we obtain the following chain of inequalities, using repeatedly (H)

$$\begin{aligned}
&\frac{1}{r_n^2}\int_{Q_n^0} |p_4 - \overline{p_4^n}||\underline{u}| \le \frac{C}{r_n^2} r_n \Big(\int_{\Delta_n^0}\Big(\int_{B_{n_0}^0} \frac{|p|}{r_{n_0}^4}\Big)^{5/4}\Big)^{4/5}\Big(\int_{\Delta_n^0}\Big(\int_{B_n^0} |\underline{u}|\Big)^5\Big)^{1/5} \le\\
&\le \frac{C}{r_n^2}\frac{r_n}{r_{n_0}^4} \tilde K_{n_0}^{0\;4/5}\Big(\int_{\Delta_n^0}\Big(\int_{B_n^0} |\underline{u}|^{2/5}|\underline{u}|^{3/5}\cdot 1\Big)5\Big)^{1/5} \le\\
&\le \frac{C}{r_n^2}\frac{r_n}{r_{n_0}^4} \tilde K_{n_0}^{0\;4/5}\Big(\int_{B_n^0} |\underline{u}|^2\Big)^{1/5}\Big(\int_{Q_n^0} |\underline{u}|^3\Big)^{1/5} r_n^{9/5} \le\\
&\le \frac{C}{r_n^2}\frac{r_n}{r_{n_0}^4} r_n^{12/5} \tilde K_{n_0}^{0\;4/5} \tilde A_n^{0\;1/5} \tilde G_n^{0\;1/5} \le C\Big(\frac{r_n}{r_{n_0}}\Big)^{7/5} A_n^{0\;1/5} G_n^{0\;1/5} K_{n_0}^{0\,4/5}
\end{aligned}$$

Finally use the inequalities of [3.5.5] and combine the estimates above on the terms p_j, $j = 1,..,4$.)

[3.5.10] Let $T_n = A_n^0 + \delta_n^0$; combine inequalities of [3.5.6] and [3.5.9], and [3.5.5] to deduce

$$\begin{aligned}
T_n \le & 2^{2n}\Big(2^{-2n_0}\varepsilon + \sum_{k=n+1}^{n_0} 2^{-2k} T_k^{3/2} + 2^{-2n_0}\varepsilon + 2^{-7n_0/5}\varepsilon \sum_{k=n+1}^{n_0} 2^{-3k/5} T_k^{1/2} +\\
&+ \varepsilon 2^{-5n_0/3} \sum_{k=n_0+2}^{n_0} 2^{-k/3} T_k^{1/2} + \sum_{k=n+1}^{n_0} 2^{-2k} T_k^{3/2} + \sum_{k=n+1}^{n_0-1} 2^k T_k^{1/2} \sum_{q=k}^{n_0} 2^{-3q} T_q\Big)\\
&\varepsilon \equiv C \max(G_{n_0}^{0\;2/3}, K_{n_0}^{0 4/5}, J^0{}_{n_0})
\end{aligned}$$

and show that, by induction, if ε is small enough then $r_n^{-2} T_n \le \varepsilon^{2/3} r_{n_0}^{-2}$.

[3.5.11]: If $G(r_0) + J(r_0) + K(r_0) < \varepsilon_s$ with ε_s small enough, then given $(\underline{x}', t') \in Q_{r_0/4}(\underline{x}_0, t_0)$, show that if one calls $G_r^0, J_r^0, K_r^0, A_r^0, \delta_r^0$ the operators associated with $Q_r^0(\underline{x}', t')$ then

$$\limsup_{n\to\infty} \frac{1}{r_n^2} A_n^0 \le C \frac{\varepsilon_s^{2/3}}{r_0^2}$$

for a suitable constant C. (*Idea:* Note that $Q_{r_0/4}^0(\underline{x}', t') \subset Q_{r_0}(\underline{x}_0, t_0)$ hence $G_{r_0/4}^0, J_{r_0/4}^0, \ldots$ are bounded by a constant, ($\le 4^2$), times the respective $G(r_0), J(r_0)$. ... Then apply the result of [3.5.10]).

[3.5.12]: Check that the result of [3.5.11] implies theorem II. (*Idea:* Indeed

$$\frac{1}{r_n^2} A_n^0 \ge \frac{1}{r_n^3}\int_{B_n^0} |\underline{u}(\underline{x}, t')|^2 d\underline{x} \xrightarrow[n\to-\infty]{} \frac{4\pi}{3}\,|\underline{u}(\underline{x}', t')|^2$$

where B^0_n is the ball centered at $\underline{x}'$, *for almost all* the points $(\underline{x},', t') \in Q^0_{r_0/4}$; hence $|\underline{u}(\underline{x}', t'))|$ is bounded in $Q^0_{r_0/4}$ and one can apply proposition IV §3.3).

[3.5.13]: Let f be a function with zero average over B^0_r. Since $f(\underline{x}) = f(\underline{y}) + \int_0^1 ds\, \underline{\partial} f(\underline{y} + (\underline{x} - \underline{y})s) \cdot (\underline{x} - \underline{y})$ for each $\underline{y} \in B^0{}_r$, averaging this identity over $\underline{y}$ one gets

$$f(\underline{x}) = \int_{B^0_r} \frac{d\underline{y}}{|B^0_r|} \int_0^1 ds\, \underline{\partial} f(\underline{y} + (\underline{x} - \underline{y})s) \cdot (\underline{x} - \underline{y})$$

Assuming $\alpha = 1$ integer prove (P). (*Idea:* Change variables as $\underline{y} \to \underline{z} = \underline{y} + (\underline{x} - \underline{y})s$ so that for α integer

$$\int_{B^0_r} |f(\underline{x})|^\alpha \frac{d\underline{x}}{|B^0_r|} \equiv \int_{B^0_r} \frac{d\underline{x}}{|B^0_r|} \Big| \int_0^1 \int_{B^0_r} \frac{d\underline{z}}{|B^0_r|} \frac{ds}{(1-s)^3} \underline{\partial} f(\underline{z}) \cdot (\underline{z} - \underline{x}) \Big|^\alpha$$

where the integration domain of $\underline{z}$ depends from $\underline{x}$ and s, and it is contained in the ball with radius $2(1-s)r$ around $\underline{x}$. The integral can then be bounded by

$$\int \frac{d\underline{z}_1}{|B^0_r|} \frac{ds_1}{1 - s_1} \cdots \frac{d\underline{z}_\alpha}{|B^0_r|} \frac{ds_\alpha}{(1-s_\alpha)^3} (2r)^\alpha |\underline{\partial} f(\underline{z}_1)| \ldots |\underline{\partial} f(\underline{z}_\alpha)| \int \frac{d\underline{x}}{|B^0_r|}$$

where $\underline{x}$ varies in a domain with $|\underline{x} - \underline{z}_i| \le 2(1-s_i)r$ for each i. Hence the integral over $\frac{d\underline{x}}{|B^0_r|}$ is bounded by $8(1-s_i)^3$ for each i. Performing a geometric average of such bounds (over α terms)

$$\int_{B^0_r} |f(\underline{x})|^\alpha \frac{d\underline{x}}{|B^0_r|} \le 2^{\alpha+3} r^\alpha \prod_{i=1}^{\alpha} \int \frac{d\underline{z}_i\, ds_i}{|B^0_r|(1-s_i)} ||\underline{\partial} f(\underline{z}_i)|\, (1-s_i)^{3/\alpha} \le$$

$$\le 2^{\alpha+3} r^\alpha \Big(\int_{B^0_r} |\underline{\partial} f(\underline{z})| \frac{d\underline{z}}{|B^0_r|} \Big)^\alpha \cdot \Big(\int_0^1 \frac{ds}{(1-s)^{3-3/\alpha}} \Big)^\alpha$$

getting (P) and an explicit estimate of the constant C^P_α.)

[3.5.14]: Differentiate twice with respect to α^{-1} and check the convexity of $\alpha^{-1} \to ||f||_\alpha \equiv (\int |f(\underline{x})|^\alpha\, d\underline{x}/|B^0_r|)^{1/\alpha}$. Use this to get (P) for each $1 \le \alpha < \alpha_0$ if it holds for $1, \alpha_0$. (*Idea:* Since (P) can be written $||f||_\alpha \le C_\alpha\, (\int |\underline{\partial} f|\, d\underline{x}/r^2)$ then if $\alpha^{-1} = \vartheta \alpha_0^{-1} + (1-\vartheta)$ with α_0 integer it follows that C_α can be taken $C_\alpha = \vartheta C_{\alpha_0} + (1-\vartheta)\, C_{\alpha_0+1}$).

[3.5.15]: Consider a sequence $\underline{u}^\lambda$ of solutions of the Leray regularized equations , *c.f.r.* §3.3, (D), which converges *weakly* (*i.e.* for each Fourier component) to a Leray solution. By construction the $\underline{u}^\lambda, \underline{u}$ verify the *a priori* bounds in (3.3.8) and (hence) (3.5.7). Deduce that $\underline{u}$ verifies the (3.5.10). (*Idea:* Only (c) has to be proved. Note that if $\underline{u}^\lambda \to \underline{u}^0$ weakly, then the left hand side of (3.5.1) is semi continuous hence the value computed with $\underline{u}^0$ is not larger than the limit of the right hand side in (3.5.1). On the other hand the right hand side of (3.5.1) is *continuous* in the limit $l \to \infty$. Indeed given $N > 0$ weak convergence implies

$$\lim_{\lambda\to\infty} \int_0^{T_0} dt \int_\Omega |\underline{u}^\lambda - \underline{u}^0|^2\, d\underline{x} \equiv \int_0^{T_0} dt \sum_{0<|\underline{k}|} |\underline{\gamma}^\lambda_{\underline{k}}(t) - \underline{\gamma}^0_{\underline{k}}(t)|^2 \le$$

$$\le \lim_{\lambda\to\infty} \sum_{0<|\underline{k}|<N} \int_0^{T_0} dt |\underline{\gamma}^\lambda_{\underline{k}}(t) - \underline{\gamma}^0_{\underline{k}}(t)|^2 + \sum_{|\underline{k}|\ge N} \int_0^{T_0} dt \frac{|\underline{k}|^2}{N^2} |\underline{\gamma}^\lambda_{\underline{k}}(t) - \underline{\gamma}^0_{\underline{k}}(t)|^2 \le$$

$$\le \lim_{\lambda\to\infty}\sum_{0<|\underline{k}|<N}\int_0^{T_0} dt|\underline{\gamma}^\lambda_{\underline{k}}(t)-\underline{\gamma}^0_{\underline{k}}(t)|^2+\frac{1}{N^2}\int_0^{T_0} dt\int_\Omega|\underset{\sim}{\partial}(\underline{u}^\lambda-\underline{u}^0)|^2=$$

$$=\lim_{\lambda\to\infty}\frac{1}{N^2}\int_0^{T_0} dt\int_\Omega|\underset{\sim}{\partial}(\underline{u}^\lambda-\underline{u}^0)|^2\le\frac{2E_0\nu^{-1}}{N^2}$$

using the *a priori* bound in (3.3.8) (with zero force) and component wise convergence of $\gamma^\lambda_{\underline{k}}(t)$ at $\gamma^0_{\underline{k}}(t)$. Hence $\int_0^{T_0}\int_\Omega|\underline{u}^\lambda-\underline{u}^0|^2\to 0$ showing the convergence of the first two terms of the right hand side of (3.5.1) to the corresponding terms of (c) in (3.5.10).

Apply, next, the inequality (S), (3.5.4), with $q=3$, $a=\frac{3}{4}$, $\frac{q}{2}-a=\frac{3}{4}$, and again by the *a priori* bounds in (3.3.8) we get

$$\int_0^{T_0} dt\int_\Omega|\underline{u}^\lambda-\underline{u}^0|^3\,d\underline{x}\le C\int_0^{T_0} dt\,||\underset{\sim}{\partial}(\underline{u}^\lambda-\underline{u}^0)||_2^{3/2}\,||\underline{u}^\lambda-\underline{u}^0||_2^{3/2}\le$$
$$\le C\Big(\int_0^{T_0} dt\,||\underset{\sim}{\partial}(\underline{u}^\lambda-\underline{u}^0)||_2^2\Big)^{3/4}\Big(\int_0^{T_0} dt\,||\underline{u}^\lambda-\underline{u}^0||_2^6\Big)^{1/4}\le$$
$$\le C(2E_0\nu^{-1})^{3/4}(2\sqrt{E_0})\int_0^{T_0} dt\,||\underline{u}^\lambda-\underline{u}^0||_2^2\xrightarrow[\lambda\to\infty]{}0$$

showing continuity of the third term in the second member of (3.5.10). Finally, and analogously, if we recall that $p^\lambda=-\Delta^{-1}\sum_{ij}\partial_i\partial_j(u^\lambda_i u^\lambda_j)$ and if we apply the inequalities (CZ) and (H), we get

$$\int_0^{T_0} dt\int_\Omega d\underline{x}|p^\lambda\underline{u}^\lambda-p^0\underline{u}^0|\le\int\int|p^\lambda-p^0|\,|\underline{u}^\lambda|+\int\int|p^0|\,|\underline{u}^\lambda-\underline{u}^0|\le$$
$$\Big(\int\int|p^\lambda-p^0|^{3/2}\Big)^{2/3}\Big(\int\int|\underline{u}^\lambda|^3\Big)^{1/3}+\Big(\int\int|p^0|^{3/2}\Big)^{2/3}\Big(\int\int|\underline{u}^\lambda-\underline{u}^0|^3\Big)^{1/3}$$

where the last integral tends to zero by the previous relation while the first, via (CZ), will be such that $\int_0^{T_0}\int_\Omega|p^\lambda-p^0|^{3/2}\le\Big(\int\int|\underline{u}^\lambda-\underline{u}^0|^3\Big)^{2/3}\xrightarrow[\lambda\to\infty]{}0$ proving the continuity o the fourth term in the right hand side of (c) in (3.5.10). Hence the right hand side is continuous in the considered limit).

Bibliography: [CKN82], [Ga93].

CHAPTER IV

Incipient turbulence and chaos

Lack of periodicity is very common in natural systems, and is one of the distinguishing features of turbulent flow. Because instantaneous flow patterns are so irregular, attention is often confined to the statistics of turbulence, which, in contrast to the details of turbulence, often behave in a regular well–organized manner. The short–range weather forecaster, however, is forced willy–nilly to predict the details of the large scale turbulent eddies –the cyclones and anticyclones– which continually arrange themselves into new patterns. Thus there are occasions when more than the statistics of irregular flow are of very real concern. (E.N. Lorenz, 1962, [Lo63] p. 131.)

§4.1 Fluids theory in absence of existence and uniqueness theorems for the basic fluidodynamics equations. Truncated NS equations. The Rayleigh's and Lorenz' models.

Analysing the fundamental problems of the NS equation has, in particular, brought up clearly the lack of an adequate algorithm, *i.e.* convergent and constructive, for its solution. Furthermore even if we knew that the fluids equations had unique and regular solutions, for regular initial data, (for the NS equation (this is true if $d = 2$ and likely if $d = 3$, false if $d \ge 4$) this would not help much to the understanding of the physical properties of such solutions at large times.

An analogous situation is met in kinetic theory of gases. Assuming that the interaction between atoms is bounded below, one easily finds that Newton's equations admit global solutions. However this gives us no help in the derivation of the properties of gases (equation of state, fluctuations, transport coefficients, *etc*). And as soon as N, number of particles, is of the order of few decades it becomes impossible to make use of the methods for the construction of solutions, not because they are no longer valid (they are) but because not even the largest conceivable computers (not to speak of the existing ones) will ever be able to apply the solution algorithms to perform, accurately, and in a reasonable time the necessary calculations (*i.e.* in a time comparable to that of human life or even to the age of the Universe).

This did not hinder the development of a deep and (at least in various respects) satisfying theory of gases and materials.

Wishing to provide a theoretical frame for the study of the asymptotic behavior in time of 2 and 3–dimensional fluid motions we are *therefore* forced to change attitude.

The question that we shall attempt to answer is whether it is possible to set up a theory without really entering into the mathematical details of the properties of the solutions of the NS or Rayleigh equations (for instance). And we should add the adjective "possible" to the word "solutions" because, as we saw, no known methods exist to construct not even one of the weak solutions of the NS equation (or for that matter of the Rayleigh equations), aside from the trivial cases in which one can construct explicitly a global solution *c.f.r.* problems in §3.3. We shall mostly refer to the incompressible NS equations but identical comments can be made for other fluidodynamical equations.

The first remark is that the NS equation, in spite of the privileged role that it plaid so far in our analysis, *is not a fundamental equation.* We must keep in mind that it is a phenomenological equation, obtained under various assumptions, *c.f.r.* (B) in §1.2. It is possible to modify a little the mathematical interpretation of the assumptions made in its derivation to obtain equations different from the NS equations but which should be physically equivalent to them for what concerns predictions about fluid motions.

For instance considering, always for the sake of simplicity, a fluid in a cubic container with periodic boundary conditions, the hypothesis that the velocity gradient "be small", made in justifying the constitutive equation

$$T_{ij} = -p\delta_{ij} + \eta(\partial_i u_j + \partial_j u_i) \tag{4.1.1}$$

could be interpreted by saying that not only $\partial_i u_j$ must be small, but that there should exist a minimal length scale λ below which there are no variations of $\underline{u}$, *i.e.* by saying that the Fourier transform $\hat{\underline{u}}_{\underline{k}}$ of $\underline{u}$ does not have components with wavelength shorter than λ, *i.e.* with $|\underline{k}| > 2\pi\lambda^{-1}$.

Hence the equations that would be obtained would be the NS equations *truncated* at $|\underline{k}| < 2\pi\lambda^{-1}$: for which we have seen that there are global existence and uniqueness theorems and constructive approximation algorithms. If however the velocity develops, in the course of time, Fourier components which have non negligible amplitudes at $|\underline{k}| \simeq 2\pi\lambda^{-1}$ it will be necessary to give up using the NS equation to describe the motion: and pass to more elaborated equations that could be non fluidodynamic equations, possibly even involving the atomic nature of the fluid.

It is, for instance, clear that if $\lambda \simeq$ mean free path in the fluid and if the harmonics of $\underline{u}$ with wavelength $\simeq \lambda$ have importance in describing the motion, then (4.1.1), and hence the NS equation itself, is inadequate for the representation of the motion.

We can therefore take the point of view that a theory of fluids can be developed by using the NS equations truncated with an ultraviolet cut–off at a wave vector $|\underline{k}| \simeq K \stackrel{def}{=} W/\nu$ where W is a velocity variation characteristic of the initial datum and ν is the dynamical viscosity so that W/ν is a typical quantity with the dimension of inverse length: at least if initially the Fourier modes with non negligible amplitude are those with $|\underline{k}| \ll W\nu^{-1}$. Note that $W/\nu = L^{-1}R$ if R is the Reynolds number.

We shall worry about the validity of the model only if, in the course of time, the motion under investigation will develop Fourier harmonics with $|\underline{k}|$ of the order of $\sim K$. In such case it will be possible to continue using the same equations but with a larger ultraviolet cut–off, and so on until we are forced to use a cut–off so large that the continuum model for the fluid becomes unreasonable (*e.g.* when the cut–off reaches the atomic scale).

It is an empirical fact that by letting smooth initial data (with "few" harmonics) evolve they do not develop, as time goes, harmonics with an ever shorter wave length but the motion evolves asymptotically, if the external force is time idependent, by "confining" the relevant harmonics and their amplitudes. In §6.2, *c.f.r.* equation (6.2.9), we shall see that $K = L^{-1}R$ is usually a "generous" ultraviolet cut–off if one considers the evolution of a velocity field obtained after the system has evolved a long enough time to have reached a stationary state. Indeed one expects an effective cut–off to act at wave numbers smaller than $K' \stackrel{def}{=} L^{-1}R^{3/4}$, in the sense that the relevant harmonics will be related to modes $\le K'$ substantially smaller than the cut–off K, while higher harmonics, even if initially present, will rapidly decay so that the equations with a cut off at $K' = L^{-1}R^{3/4}$ should already faithfully represent the fluid motion.

This attitude allows us to attempt at developing an empirical theory of incompressible fluid motions by describing them with differential equations that are, or that are believed to be, equivalent to the NS equations, unless we realize that the theory itself implies phenomena that force us to change the model to obtain a more correct description.

In this way we can set up a procedure to build mathematical models for fluids that often reveal themselves self consistent (in the sense that they do not evolve interesting initial data towards situations in which the approximations and assumptions made in deriving the model fail).

The interest of this viewpoint is that one is led to models that do not present mathematical difficulties, for what concerns existence, uniqueness and constructive approximability of the solutions, and nevertheless are models that can be used to illustrate physical phenomena.

A classic example, and important as well, is provided us by the *Lorenz' model* for convection. It is a model derived via a hyper simplification of the Rayleigh's convection model, §1.5. The Lorenz' treatment of the model (including its derivation) has a special historical importance because it establishes in a clear way a method of analysis that has been followed in most successive theoretical studies of the onset of turbulence, [Lo63].

To avoid considering the Lorenz' model as a mathematical curiosity it is useful to keep in mind the analysis of §1.5, where we devoted attention to the physical foundations of Rayleigh's model.

We shall also illustrate some truncations of the bidimensional NS equation with periodic boundary conditions. Both the Lorenz' model and the truncations of the NS equation that we shall discuss, will have to be considered of mathematical or of exemplificatory interest, except perhaps in cases of very low Reynolds numbers: they will serve to illustrate in the coming sections, concrete examples in which certain mechanisms become manifest that we believe are at the origin of the development of turbulence and of chaotic motions. "Developed" or "strong" turbulence is a phenomenon important only at very large Reynolds' numbers and it will the object of study after that of the initial turbulence that appears at (relatively) small Reynolds' number, see Ch. VI, VII.

(A) The 2–dimensional Saltzman's equations.

We recall that in the theory of convection, §1.5, we denoted $\underline{u}$ the velocity field and ϑ the variation of the temperature field with respect to the linear temperature profile (*i.e.* the trivial "*thermostatic solution*") that one would have if the fluid remained motionless, just conducting heat from the top (colder) plate of the container to the bottom (warmer) one: see (1.5.17).

The 2–dimensional version of Rayleigh's model is obtained by assuming that $\underline{u},\vartheta$ in (1.5.17) are y–independent. In this case there is a velocity potential $\psi=\psi(x,z)$ such that:

$$\underline{u}=(-\partial_z\psi,0,\partial_x\psi) \tag{4.1.2}$$

and (1.5.17) can be rewritten in terms of ψ rather than of $\underline{u}$. With the notation

$$\underset{\sim}{u}\cdot\underset{\sim}{\partial}\, f\equiv(-\partial_z\psi\,\partial_x\, f+\partial_x\,\psi\,\partial_z\, f)\equiv\frac{\partial(\psi,f)}{\partial(x,z)} \tag{4.1.3}$$

The Rayleigh's equations (1.5.17) take a form called "*Saltzman's equation*"

$$\begin{aligned}\partial_t\vartheta+R\frac{\partial(\psi,\vartheta)}{\partial(x,z)}=&\sigma^{-1}\Delta\vartheta+R\partial_x\psi, \qquad \psi|_{z=0}=\psi|_{z=1}=0\\ \partial_t\Delta\psi+R\frac{\partial(\psi,\Delta\psi)}{\partial(x,z)}=&\Delta^2\psi+R\partial_x\vartheta, \qquad\quad \vartheta|_{z=0}=\vartheta|_{z=1}=0\end{aligned} \tag{4.1.4}$$

which differ from the Saltzman's equation in [Lo63] only because we write them in a dimensionless form; R is the Reynolds number σ is the "Prandtl's number", (*c.f.r.* §1.5).

The equation $\underline{\partial}\cdot\underline{u}=0$ is now an identity, if $\underline{u}$ is given by (4.1.2), while the condition $\psi(x,0)\equiv\psi(x',1)$ translates the condition of vanishing total horizontal momentum and at the same time the condition of 0 horizontal velocity at the boundaries $z=0,1$ ("bottom" and "top" of the fluid).[1] The

[1] If $\psi(x,0)$ and $\psi(x,1)$ are constants it is: $u_z=\partial_x\psi=0$ if $z=0,1$. Furthermore, the two constant values of ψ for $z=0$ and $z=1$ are related because we require that the total horizontal momentum per unit horizontal length vanishes: in fact if the two constant values are equal it is: $\int u_x dxdz=-\int\partial_z\psi dxdz=-\int(\psi(x,1)-\psi(x,0))dx=0$.

fact that the constant value of ψ, common to $z = 0$ and $z = 1$, is set equal to 0 simply reflects the property of ψ of being defined up to an arbitrary additive constant.

(B) A priori estimates.

Consider first a continuously differentiable solution of the general 3–dimensional Rayleigh's equations, (1.5.17)

If $\delta T < 0$, *i.e.* if the temperature $T_0 - \delta T$ of the upper surface is warmer than that, T_0, of the lower surface, then by multiplying the first of (1.5.18) by $\underline{u}$ and the second by ϑ, one finds

$$\frac{1}{2}\frac{d}{dt}E(t) = -S(t) \qquad \text{with} \qquad (4.1.5)$$
$$E(t) = \int \underline{u}^2\, d\underline{x} + \int \vartheta^2\, d\underline{x}, \qquad S(t) = \int (\underset{\sim}{\partial}\underline{u})^2\, d\underline{x} + \sigma^{-1}\int (\underline{\partial}\vartheta)^2\, d\underline{x}$$

hence $E(t) \le E(0)$.

If, instead, $\delta T > 0$ (*i.e.* the upper surface is colder than the lower) we find from (1.5.17):

$$\frac{1}{2}\frac{d}{dt}E(t) = -S(t) + 2R\int \vartheta u_z\, d\underline{x} \le -S(t) + RE(t) \tag{4.1.6}$$

(having exploited $2ab \le a^2 + b^2$), hence

$$E(t) \le E(0)e^{2Rt} \tag{4.1.7}$$

We can obain more interesting bounds if we consider the problem with periodic horizontal boundary conditions, *i.e.* we demand that $\underline{u}, \vartheta$ be periodic functions in x, y at fixed z, with period l^{-1}, where $l > 0$ is a fixed parameter (we are using dimensionless units, so that l is a dimensionless quantity). Then

$$S(t) \ge ((2\pi l)^2 + \pi^2)\, \min(1, \sigma^{-1})\, E(t) \stackrel{def}{=} \pi^2\, \mu\, E(t) \tag{4.1.8}$$

because the expansion of the fields $\underline{u}, \vartheta$ in Fourier series in x and in a sine series in z has minimum momentum $2\pi l$ in the x–direction and π in the z–direction (in our units, the vertical dimension is 1). Hence, in the *uninteresting* case $\delta T < 0$ (*c.f.r.* (1.5.18)), (4.1.5) implies

$$E(t) \le E(0)e^{-2\pi^2\,\mu\, t} \xrightarrow[t\to+\infty]{} 0 \tag{4.1.9}$$

and the "no–motion" state $\vartheta \equiv 0$, $\underline{u} = \underline{0}$ is always a *stable equilibrium.*

While if $\delta T > 0$ the no–motion state is stable (by (4.1.6), (4.1.8)) if $2R < \pi^2\mu$, *i.e.* if R is small enough; and, for general data, we can only say

$$E(t) \le E(0)e^{-(2\pi^2\mu - 2R)t} \tag{4.1.10}$$

although, even for R large, the estimate (4.1.10) could be excessively pessimistic (at least in some cases one can check that it can be improved to $E(t) \le const$).

In the 2–dimensional case one gets from (4.1.4) an estimate of $(\underset{\sim}{\partial}\underline{u})^2$ and $(\Delta\psi)^2$, similar to the (4.1.10) but with $\underline{u}^2 = (\underline{\partial}\psi)^2$ replaced by $(\Delta\psi)^2$:

$$\begin{aligned}&\tfrac{1}{2}\dot{E}_1 \equiv \tfrac{1}{2}\tfrac{d}{dt}\int(\vartheta^2 + (\Delta\psi)^2)\,d\underline{x} \le \\ &\le -\int\big((\Delta\underline{\partial}\psi)^2 + \sigma^{-1}(\underline{\partial}\vartheta)^2\big)\,d\underline{x} + R\int(\vartheta\partial_x\psi + \Delta\psi\,\partial_x\vartheta)\,d\underline{x}\end{aligned} \qquad (4.1.11)$$

This implies that the quantity $E_1 \equiv \int(\vartheta^2 + (\Delta\psi)^2)\,d\underline{x}$ is bounded by

$$E_1(t) \le E_1(0)e^{2Ct} \qquad (4.1.12)$$

and C can be estimated in terms of R, σ, l, *c.f.r.* problems.

The *a priori* bound in (4.1.12) suffices, with (4.1.8), to develop a theory of existence of weak solutions. And, *in the 2-dimensional case*, the (4.1.12) permits us to derive a proof of an existence, uniqueness and regularity theorem (when the initial data are C^∞) analogous to the one discussed in §3.2 for the NS equations.

From the bounds on E_1 (hence on $\int(\underset{\sim}{\partial}\underline{u})^2 d\underline{x}, \int\vartheta^2 d\underline{x}$), and thinking the equations written as in (1.5.17), we can repeat the auto–regularization theory of §3.2 in the present case. The only substantial change is that now the boundary conditions are different, and the velocity and temperature fields must be developed over a basis different from the Fourier basis, *c.f.r.* (4.1.13). One finds that if $|\vartheta_{\underline{k}}| < C_\beta|\underline{k}|^{-\beta}$ and $|\underline{u}_{\underline{k}}| < C_\alpha|\underline{k}|^{-\alpha}$ with $\beta \ge 0$ and $\alpha \ge 2$, and $C_a, C_\beta < \infty$, then $\vartheta, \underline{u}$ are C^∞.

(C) The Lorenz' model.

The model is obtained by considering the (4.1.4) with periodic conditions in the variable $x \in [0, 2a^{-1}]$, where a is a dimensionless parameter (we are using here the dimensionless equations: in the original coordinates the length scale associated with $2a^{-1}$ would be: $2a^{-1}H$).

Supposing that ϑ and ψ are of class C^∞ and choosing a suitable basis adapted to periodic conditions in x and to vanishing conditions at $z = 0, 1$, we can write

$$\begin{aligned}\vartheta(x,z) &= \sum_{k_1=-\infty}^{\infty}\sum_{k_2=1}^{\infty}\vartheta_{k_1k_2}e^{i\pi k_1 lx}\sin\pi k_2 z\\ \psi(x,z) &= \sum_{k_1=-\infty}^{\infty}\sum_{k_2=1}^{\infty}\psi_{k_1k_2}e^{i\pi k_1 lx}\sin\pi k_2 z\end{aligned} \qquad (4.1.13)$$

with $\vartheta_{k_1k_2} = \overline{\vartheta}_{-k_1k_2}$ and $\psi_{k_1k_2} = \overline{\psi}_{-k_1k_2}$. We can then study special solutions that admit a development like (4.1.13) with $\vartheta_{\underline{k}}$ and $\psi_{\underline{k}}$ real: in fact, for symmetry reasons, if the initial data can be written as in (4.1.13) with $\vartheta_{\underline{k}}, \psi_{\underline{k}}$ real then this representation remains valid for all times $t > 0$, for every regular solution of the (4.1.4). The equations (4.1.4) become (after some trigonometry)

$$\begin{aligned}
\dot\vartheta_{r_1r_2} =& -\sigma^{-1}\pi^2(l^2r_1^2+r_2^2)\vartheta_{r_1r_2}+\pi l r_1 R\psi_{r_1r_2}-\frac{\pi^2 l}{4}R\cdot\\
&\cdot\sum_{\substack{\varepsilon=\pm1\\ \eta=\pm1}}\sum_{\substack{|k_1+\eta h_1|=r_1\\ |k_2+\varepsilon h_2|=r_2}}\big(k_1h_2-\eta k_2h_1\big)\psi_{k_1k_2}\vartheta_{h_1h_2}\sigma_2 \qquad (4.1.14)\\
\dot\psi_{r_1r_2} =& -\pi^2(l^2r_1^2+r_2^2)\psi_{r_1r_2}+\frac{\pi l r_1}{\pi^2(r_1^2l^2+r_2^2)}R\,\vartheta_{r_1r_2}-\frac{\pi^2 l}{4}R\cdot\\
&\cdot\sum_{\substack{\varepsilon=\pm1\\ \eta=\pm1}}\sum_{\substack{|k_1+\eta h_1|=r_1\\ |k_2+\varepsilon h_2|=r_2}}\frac{\pi^2(l^2h_1^2+h_2^2)}{\pi^2(l^2r_1^2+r_2^2)}\big(\eta k_1h_2-\varepsilon k_2h_1\big)\psi_{k_1k_2}\psi_{h_1h_2}\sigma_1\sigma_2
\end{aligned}$$

where σ_1, σ_2 are the signs of $k_1+\eta h_1$ and $k_2+\varepsilon h_2$, respectively. And $\underline{r}, \underline{k}, \underline{h}$ are integer, non negative, components vectors with one component at least always positive (*c.f.r.* (4.1.13)).

The (4.1.14) are the "spectral form", *c.f.r.* §2.2, of the (4.1.4) for solutions with initial data that can be written as (4.1.13).

It is believed, and it has been shown by (numerical) experiments of Saltzman's, [Sa62], [Lo63], that at least as R and σ vary in suitable intervals there exist initial data which, following (4.1.14), evolve as $t \to +\infty$ towards a state in which only three spectral components $\vartheta_{\underline{k}}, \psi_{\underline{k}}$ are not negligible; and precisely the components

$$\psi_{11} \stackrel{def}{=} \sqrt{2}X, \quad \vartheta_{11} \stackrel{def}{=} \sqrt{2}Y, \quad \vartheta_{02} \stackrel{def}{=} -Z \qquad (4.1.15)$$

The equations (4.1.14) truncated *over the latter* 3 *components* become

$$\begin{aligned}
\dot X =& -\pi^2(1+a^2)X+\frac{\pi a R}{\pi^2(1+a^2)}Y\\
\dot Y =& -\sigma^{-1}\pi^2(1+a^2)Y-\pi^2 aRXZ+\pi aRX \qquad (4.1.16)\\
\dot Z =& -4\sigma^{-1}\pi^2 Z+\pi^2 aRXY
\end{aligned}$$

To simplify (4.1.16) we rescale the variables with suitable scaling parameters λ, μ, ν, ξ (which are symbols that are not to be confused with physical quantities that we occasionally denoted with the same letters)

$$x \equiv \lambda X, \qquad y = \mu Y, \qquad z \equiv \nu Z, \qquad \tau \equiv \xi t \qquad (4.1.17)$$

so that the equations can be rewritten as

$$\begin{aligned}
\dot x =& \xi^{-1}\Big(-\pi^2(1+a^2)x+\frac{\pi a\lambda R}{\pi^2(1+a^2)\mu}y\Big)\\
\dot y =& \xi^{-1}\Big(-\sigma^{-1}\pi^2(1+a^2)y-\frac{\pi^2 aR\mu}{\lambda\nu}xz+\frac{\pi aR\mu}{\lambda}x\Big) \qquad (4.1.18)\\
\dot z =& \xi^{-1}\Big(-4\sigma^{-1}\pi^2 z+\frac{\pi^2 aR\nu}{\lambda\mu}xy\Big)
\end{aligned}$$

Following Lorenz' normalizations we shall choose

$$\begin{aligned}
&\xi^{-1}\sigma^{-1}\pi^2(1+a^2)=1,\quad \xi^{-1}\frac{\pi^2 aR\nu}{\lambda\mu}=1,\quad \xi^{-1}\frac{\pi^2 aR\mu}{\lambda\nu}=1,\\
&r\stackrel{def}{=}\xi^{-1}\frac{\pi aR\mu}{\lambda},\quad \xi^{-1}\frac{\pi aR\lambda}{\pi^2(1+a^2)\mu}\stackrel{def}{=}\sigma,\qquad b\equiv\frac{4}{(1+a^2)}
\end{aligned}\tag{4.1.19}$$

obtaining the "*Lorenz' model*" equations

$$\begin{aligned}
\dot x=&-\sigma x+\sigma y\\
\dot y=&-y-xz+rx\\
\dot z=&-bz+xy
\end{aligned}\qquad\text{with}\qquad
\begin{aligned}
b=&4(1+a^2)^{-1}\\
\sigma=&R_{Pr}\\
r=&\frac{(\pi a)^2R^2\sigma}{(\pi^2(1+a^2))^3}\equiv\frac{R_{Ray}}{R_a}
\end{aligned}\tag{4.1.20}$$

where $R_a\equiv\pi^4(1+a^2)^3a^{-2}$ and R_{Pr},R_{Ray} are the Prandtl and Rayleigh numbers, defined in §1.5, *c.f.r.* for instance (1.5.15).

Remark: (symmetry) It is important to note that the equations (4.1.20) are invariant under a (simple) symmetry group; namely under the 2–elements group consisting of the transformations:

$$x\to\varepsilon\,x,\;\; y\to\varepsilon\,y,\;\; z\to z\qquad \varepsilon=\pm1\tag{4.1.21}$$

which we call the "symmetry group" of (4.1.20).

The (4.1.20) are such that the thermostatic solution $\vartheta=\psi=0$ loses stability for $R_{Ray}>R_a$, *i.e.* for $r>1$, (as shown by Rayleigh in the more general context of the 3–dimensional Rayleigh's equation, and as it can be checked immediately by studying the linearized equation near a thermostatic solution). The value of a that yields the minimum of R_a is $a^2=1/2$, which shows that the convective instability arises by generating convective motions that are spatially periodic with a period of length $2a^{-1}H\equiv2\sqrt2H$ (in dimensional units).

The equations were studied by Lorenz, [Lo63], choosing

$$a^2=\frac12,\qquad b=\frac83,\qquad \sigma=10.\tag{4.1.22}$$

A choice simply due to the fact that $a^2=1/2$ (which implies the value $b=8/3$, *c.f.r.* (4.1.20)), is the value of a for which the static solution $\vartheta=\psi=0$ becomes unstable at the smallest value of the number R_{Ray}. The value of σ is a value of the order of magnitude of the Prandtl's number of water in normal conditions, *c.f.r.* table in §1.5.

(D) Truncated NS Models.

Likewise we can study systems of ordinary equations obtained by truncation from the NS equations. It is interesting to study them together with the Lorenz' model in order to illustrate other phenomena that, although very frequent, do not show up in its phenomenology.

We shall only consider a few truncations of the NS equations in 2–dimensions with periodic boundary conditions. The latter equations, written in spectral form, see (3.2.6), (3.2.26), are

$$\dot{\gamma}_{\underline{k}} = -\underline{k}^2\gamma_{\underline{k}} - i\sum_{\Delta(\underline{k})}\gamma_{\underline{k}_2}\gamma_{\underline{k}_3}\frac{(\underline{k}_2^\perp\cdot\underline{k}_3)(\underline{k}_3^2-\underline{k}_2^2)}{|\underline{k}_1|\,|\underline{k}_2|\,|\underline{k}_3|} + g_{\underline{k}} \tag{4.1.23}$$

where the sum runs over the set, denoted $\Delta(\underline{k})$, of all unordered pairs of coordinates, $\underline{k}_2,\underline{k}_3$ such that $\underline{k}_2+\underline{k}_3=\underline{k}$. It is the sum over the "triads of interacting modes" considered in §2.2; the velocity and external force fields are, respectively,

$$\underline{u}(\underline{x}) = \sum_{\underline{k}\neq\underline{0}}\gamma_{\underline{k}}\frac{\underline{k}^\perp}{|\underline{k}|}e^{i\underline{k}\cdot\underline{x}}, \qquad \underline{g}(\underline{x}) = \sum_{\underline{k}\neq\underline{0}}g_{\underline{k}}\frac{\underline{k}^\perp}{|\underline{k}|}e^{i\underline{k}\cdot\underline{x}}, \tag{4.1.24}$$

and the viscosity ν is set equal to 1, for simplicity; *c.f.r.* also (2.2.10), (2.2.25), (3.2.26).

The truncations will have $\underline{k}$ and $\Delta(\underline{k})$ chosen among the following seven "modes"

$$\begin{aligned} &\underline{k}_1 =(1,1), \qquad \underline{k}_2=(3,0), \qquad \underline{k}_3=(2,-1), \qquad \underline{k}_4=(1,2),\\ &\underline{k}_5 =(0,1), \qquad \underline{k}_6=(1,0), \qquad \underline{k}_7=(1,-2)\end{aligned} \tag{4.1.25}$$

and their opposites.

One checks that the equations truncated on the modes in (4.1.25) are (setting $\gamma_i\equiv\gamma_{\underline{k}_i}$):

$$\begin{aligned}
&\dot{\gamma}_1 = -2\gamma_1 + \frac{4i}{\sqrt{10}}\gamma_2\,\overline{\gamma}_3 - \frac{4i}{\sqrt{10}}\gamma_4\,\overline{\gamma}_5, \qquad \dot{\gamma}_2 = -9\gamma_2 - \frac{3i}{\sqrt{10}}\gamma_1\gamma_3\\
&\dot{\gamma}_3 = -5\gamma_3 - \frac{7i}{\sqrt{10}}\,\overline{\gamma}_1\gamma_2 - \frac{9i}{5\sqrt{2}}\gamma_1\gamma_7 + r, \qquad \dot{\gamma}_4 = -5\gamma_4 - \frac{i}{\sqrt{10}}\gamma_1\gamma_5\\
&\dot{\gamma}_5 = -\gamma_5 + \frac{3i}{\sqrt{10}}\,\overline{\gamma}_1\gamma_4 + \frac{i}{\sqrt{2}}\gamma_1\,\overline{\gamma}_6, \qquad \dot{\gamma}_6 = -\gamma_6 - \frac{i}{\sqrt{2}}\gamma_1\,\overline{\gamma}_5\\
&\dot{\gamma}_7 = -5\gamma_7 - \frac{9i}{5\sqrt{2}}\gamma_3\,\overline{\gamma}_1
\end{aligned} \tag{4.1.26}$$

where the external force has been supposed to have only one component: with intensity r on the mode $\pm\underline{k}_3$. The (4.1.26) admit special solutions in which

$$\begin{aligned} \gamma_j = x_j = \text{real}, &\qquad j=1,3,5\\ \gamma_j = ix_j = \text{imaginary}, &\qquad j=2,4,6,7\end{aligned} \tag{4.1.27}$$

and such a solutions are generated by initial data of the form (4.1.27) (an easily checked symmetry property) provided r is real, as we shall suppose from now on.

The equations (1.4.26) for complex data become a set of seven real equations for solutions of the form (4.1.27)

$$\begin{aligned}
\dot x_1 =& -2x_1-\frac{4}{\sqrt{10}}x_2x_3+\frac{4}{\sqrt{10}}x_4x_5, \qquad \dot x_2=-9x_2-\frac{3}{\sqrt{10}}x_1x_3\\
\dot x_3 =& -5x_3+\frac{7}{\sqrt{10}}x_1x_2+\frac{9}{5\sqrt2}x_1x_7+r, \qquad \dot x_4=-5x_4-\frac{1}{\sqrt{10}}x_1x_5\\
\dot x_5 =& -x_5-\frac{3}{\sqrt{10}}x_1x_4+\frac{1}{\sqrt2}x_1x_6, \qquad \dot x_6=-x_6-\frac{1}{\sqrt2}x_1x_5\\
\dot x_7 =& -5x_7-\frac{9}{5\sqrt2}x_3x_1
\end{aligned}\tag{4.1.28}$$

which we shall call *NS_7–model* or seven modes truncated model of the 2–dimensional NS equations (studied in [FT79]).[2]

Remark: (symmetry) We note that the equations (4.1.28) are invariant under a (simple) symmetry group; namely under the 4–elements group consisting of the transformations:

$$\begin{aligned}
&\gamma_1\to\varepsilon\,\eta\,\gamma_1, \qquad \gamma_2\to\varepsilon\,\eta\,\gamma_2, \qquad \gamma_3\to\gamma_3,\\
&\gamma_4\to\eta\,\gamma_4, \qquad \gamma_5\to\varepsilon\,\gamma_5, \qquad \gamma_6\to\eta\,\gamma_6, \qquad \gamma_7\to\varepsilon\,\eta\,\gamma_7
\end{aligned}\tag{4.1.29}$$

with $\varepsilon,\eta=\pm1$, which we call the "symmetry group" of (4.1.28).

A simpler model, (considered in [BF79], [FT79])[3] is the following NS_5–model obtained by considering the special truncation

$$\begin{aligned}
\dot x_1 =& -2x_1-\frac{4}{\sqrt{10}}x_2x_3+\frac{4}{\sqrt{10}}x_4x_5, \qquad \dot x_2=-9x_2-\frac{3}{\sqrt{10}}x_1x_3\\
\dot x_3 =& -5x_3+\frac{7}{\sqrt{10}}x_1x_2+r, \qquad \dot x_4=-5x_4-\frac{1}{\sqrt{10}}x_1x_5\\
\dot x_5 =& -x_5-\frac{3}{\sqrt{10}}x_1x_4
\end{aligned}\tag{4.1.30}$$

which is obtained by setting $x_6=x_7\equiv0$ in (4.1.28). The (4.1.30) are invariant under the same 4–elements *symmetry group* in (4.1.29):

$$\begin{aligned}
&\gamma_1\to\varepsilon\,\eta\,\gamma_1, \qquad \gamma_2\to\varepsilon\,\eta\,\gamma_2, \qquad \gamma_3\to\gamma_3,\\
&\gamma_4\to\eta\,\gamma_4, \qquad \gamma_5\to\varepsilon\,\gamma_5,
\end{aligned}\tag{4.1.31}$$

[2] : In the reference the equations are written for variables ξ_i related to those called here x_i by $x_i\equiv\sqrt{50}\xi_i$, x_2 is denoted $-x_2$ and $r=R\sqrt{50}$ if R is the quantity defined in [FT79]; see also [Ga83], Cap. V, §8, (5.8.14).

[3] In the references the equations are written for the variables ξ_i related to the x_i by $x_i\equiv\sqrt{10}\xi_i$ and x_2 is denoted $-x_2$, see also [Ga83], Cap. V, §8, (5.8.9).

with $\varepsilon, \eta = \pm 1$, which we call the "symmetry group" of (4.1.30).

As we see, the equations (4.1.28), (4.1.30) are quite similar to those of the Lorenz' model. All such equations can be interpreted as equations describing systems of coupled gyroscopes, subject to an external torque, to friction and to some dynamical (*i.e.* anholonomic) constraints that force equality of various components of the angular velocities. This is a particular case of the general remark discussed in (E) of §2.2.

Likewise one could consider truncations involving more modes or also truncations on the corresponding equations in $d = 3$ dimensions. But the Lorenz' and NS models with five or seven modes, that we denoted respectively, NS_5 and NS_7, will be sufficient to illustrate many phenomena of the qualitative theory of the onset of chaotic motions and turbulence.

Problems.

[4.1.1]: Show that if $\underline{u}$ and ϑ verify the boundary conditions in (4.1.4) and are periodic in x with period l^{-1} then there exists a constant $C_1 > 0$ such that: $\int (\partial_x \psi)^2 dx dz \le C_1 l^2 \int (\Delta\psi)^2 dx\, dz$, where the integral in x is over the period. Show that the right hand side of (4.1.11) can be bounded above by $C(R, \sigma, l) \int (\vartheta^2 + (\Delta\psi)^2)\, dx\, dz$ and compute an explicit estimate for C_1 and the function $C(R, \sigma, l)$. (*Idea:* Use (4.1.3); and, for the calculation of C use $ab \le \varepsilon^{-1} a^2 + \varepsilon b^2$ with $a = \vartheta$ and $b = \partial_x \psi$, or $a = \Delta\psi$ and $b = \partial_x \vartheta$, fixing suitably ε in the two cases, to compensate the "undesired term" with the first integral on the right hand side of (4.1.11).)

[4.1.2]: Show that if $r < 1$ the origin is a global attracting set for the solutions of (4.1.20): all initial data tend to zero exponentially. (*Idea:* Multiply (4.1.20) by $(\alpha\, x, y, z)$, with $\alpha > 0$ getting, if $E = \alpha\, x^2 + y^2 + z^2$: $\frac{1}{2}\dot{E} = -\alpha\, \sigma x^2 - y^2 - bz^2 + (\alpha\, \sigma + r)xy$ and note that the matrix (often called the "stability matrix") $\begin{pmatrix} \alpha\sigma & -\frac{1}{2}(r + \alpha\,\sigma) \\ -\frac{1}{2}(r + \alpha\,\sigma) & 1 \end{pmatrix}$ has determinant $\alpha\,\sigma - \frac{1}{4}(\alpha\,\sigma + r)^2$. The latter has a maximum value at $\alpha\,\sigma = 2 - r$ where it has value $1 - r$. As long as the determinant is > 0 the matrix is positive definite, hence $\ge \lambda(r) > 0$. Setting $\delta = \min(\lambda(r), b)$ one realizes that $\frac{1}{2}\dot{E} \le -\delta E$ with $\delta > 0$ if $r < 1$.)

[4.1.3]: Show that solutions of (4.1.20) evolve so that (x, y, z) enters into any ball of radius $> br^{-1}(\sigma + r)$ after a finite time (dependingon the initial condition). (*Idea:* Given [4.1.2] it suffices to chek the statement for $r \ge 1$: change variables setting $z = \zeta - z_0$ and rewrite the equations for x, y, ζ; multiply them, scalarly, by (x, y, ζ) obataining, with the choice $z_0 = -(\sigma + r)$: $\frac{1}{2}\dot{E} = -\sigma x^2 - y^2 - bz^2 - b(\sigma + r)\zeta \le -\delta'\, E + b(\sigma + r)\sqrt{E}$, where $E = x^2 + y^2 + \zeta^2$, $\delta' = \min(\sigma, b, 1)$ hence $\dot{E} < 0$ if $E > b^2(\sigma + r)^2$.)

[4.1.4]: Show that if $r = 0$ every initial datum for the Lorenz' model, (4.1.20),(4.1.22), evolves approaching zero esponentially for $t \to \infty$, *i.e.* as e^{-ct} for some time constant c, and estimate the time constant . (*Idea:* Multiply (4.1.20) by x, y, z respectively and use that $\sigma > b > 1$ to deduce that the time constant is ≥ 1, *c.f.r.* [4.1.2] above).

[4.1.5] Show that if $r = 0$ the solutions of (4.1.26) and (4.1.28) approach zero esponentially and show that the time constant (see [4.1.4]) is ≥ 1 in both cases. Check that there are solutions, in both cases, that tend to zero exactly as e^{-t}. (*Idea:* The special solutions correspond to initial data with all components vanishing but a suitable one).

[4.1.6] Show that, if $r \ne 0$, the equations (4.1.26), (4.1.28) have, for small r, a time independent solution $\underline{x}_r$, often also called a "*fixed point*", that attracts exponentially initial data close enough to $\underline{x}_0$. And if $r \ne 0$ is small also the equations (4.1.20), (4.1.22) enjoy the same property. (*Idea:* By continuity the linear stability matrix, *i.e.* the matrix of the coefficients obtained by linearizing the equations near the time independent solution, keeps eigenvalues with real part $\lesssim 0$ by continuity, because for $r = 0$ they are ≤ -1 by the time constant estimates in [4.1.4], [4.1.5]).

[4.1.7]: Show that the first stability loss of (4.1.20), (4.1.22) happens at $r = 1$ with a single eigenvalue λ of the stability matrix becoming zero. Check that this bifurcation generates two new time independent solutions no longer invariant with respect to the symmetry, implicit in (4.1.20), $(x, y) \longleftrightarrow (-x, -y)$. (*Idea:* Compute explicitly the time independent solutions).

[4.1.8]: Identify the symmetries of (4.1.20), (4.1.28) and check that as r increases they are successively "broken" generating, by bifurcation, time independent motions with less and less symmetries. (*Idea:* Just solve explicitly for the time independent solutions).

[4.1.9]: Compute the maximum value r_c of r for which the time independent solutions of (4.1.20), (4.1.22) without symmetry are stable and show that the stability loss takes place because two nonreal conjugate eigenvalues reach the imaginary axis for $r = r_c$. (*Idea:* Compute the nontrivial time independent solutions, *i.e.* $z = r - 1$, $x = y = \pm\sqrt{r-1}$, and linearize the equation around one of them thus obtaining an expression for the stability matrix. Compute its characteristic equation, which will have the form: $\lambda^3 + c_2\lambda^2 + c_1\lambda + c_0 = 0$ with c_j simply expressed in terms of the parameters, r, b, σ. The critical value of r in correspondence of which such time independent solutions lose stability is found by requiring that the above characteristic equation has two purely imaginary solutions $\lambda = \pm i\mu$ (one easily realizes that a third solution is always < 0); hence it is the value of r for which there is μ such that $-\mu^3 + c_1\mu = 0$ *and* $-c_2\mu^2 + c_0 = 0$; *i.e.* it is the value of r for which $c_1 = c_0/c_2$: this turns out to be a *linear equation* for r with solution $r_c = 470/19 = 24.736842$.)

[4.1.10]: Show that (4.1.26) for $r > \bar r_c$, with $\bar r_c$ suitably chosen, does not admit new stable or unstable time independent solutions (*i.e.* the number of time independent solutions is constant for r large enough), and there exists $\bar r_{c,0} > \bar r_c$ such that for $r > \bar r_{c,0}$ no time independent solution is stable. (*Idea:* Solve for all the time independent solutions and compute exactly the time independent solutions and their respective stability matrices; and discuss the sign of the real part of the eigenvalues. Show that that the fourth order equation for the eigenvalues of the time independent solutions with the least symmetry admit a purely imaginary solution $i\mu$ for $r = r_{c,0}$, (which implies a discussion of an equation of second degree only)).

[4.1.11]: The loss of stability, at $r = r_c$, of the time independent solutions of the Lorenz' model is *inverse*: in the sense that no stable periodic motions are generated . Check this empirically via a qualitative computer analysis. In the next sections we shall introduce the notion of vague attractivity which will allow an analytic check for r near r_c (this is a cumbersome, but instructive, check that the loss of stability takes place via the imaginary axis crossing by two conjugate *nonreal* eigenvalues of the linear stability matrix and with a vague attractivity index which is positive, see §4.2, §4.3).

[4.1.12]: Consider a 3–mode truncation of the $d = 2$ NS equation and suppose that the *shortest modes* $\underline{k}_0 = (0, \pm 1)$ are the sole modes with a novanishing component of the force $\underline{g}_{\underline{k}}$: $\underline{g}_{\underline{k}} = r\delta_{\underline{k},\underline{k}_0}\underline{k}_0^\perp$. Check that *laminar motion* $\underline{\gamma}_{\underline{k}} = \delta_{\underline{k},\underline{k}_0}\, r\underline{k}_0^\perp/\nu\underline{k}_0^2$ is linearly stable for *all* values of r. (*Idea:* Diagonalize the stability matrix).

[4.1.13]: As in [4.1.12] but with a 5 modes truncation. Is this true for any finite truncation? (*Idea:* Yes (*Marchioro's theorem*). Write (as suggested by *Falkoff*) the NS equations in dimension 2 and check that if $\Delta \stackrel{def}{=} \sum_{|\underline{k}|>k_0} (\frac{\underline{k}^2}{\underline{k}_0^2} - 1)|\gamma_{\underline{k}}|^2$ then $\frac{1}{2}\dot\Delta = \sum_{|\underline{k}|>k_0} \underline{k}^2 (\frac{\underline{k}^2}{\underline{k}_0^2} - 1)|\gamma_{\underline{k}}|^2$, hence $\frac{1}{2}\dot\Delta \le -\nu k_0^2\Delta$ and $\Delta(t) \le e^{-2\nu k_0^2 t}\Delta(0)$, *etc.*; see also proposition 4 in Sec. 3.2)

Bibliography: The analysis of subsections (A)÷(C) is taken from [Lo63] and that in (D) from [FT79]; the problems are taken from the same papers, see also [Ga83], Cap. V, §9. The last two problems admit a remarkable generalization: the $d = 2$ NS equation, *whether or not* truncated, forced on the *shortest mode* (*i.e.* on the *largest length scale*) is such that the laminar solution is stable for all values of the force intensity (*Marchioro's theorem*), *c.f.r.* [Ma86]. This means that in the $d = 2$ NS equation forced on the shortest mode turbulence is not possible. This result admits some generalizations, see [Ma86],[Ma87].

§4.2: Onset of chaos. Elements of bifurcation theory.

In particular regimes of motion, fluids can be described by *simple* equations with few degrees of freedom: this happens when the velocity, temperature, density fields can be expanded on suitable bases of functions adapted to the boundary conditions of the specific problem and have few appreciably nonzero components. This is the very definition of "simple" or "low excitation" motion.

Examples can be imagined on the basis of the analysis in §4.1. The Lorenz' and truncated NS models have been derived by developing the fields on the Fourier basis, or on similar bases adapted to the considered boundary conditions (*c.f.r.* (4.1.13) for an example): sometimes they can, in spite of their simplicity, be considered as good models for understanding some of the fluids properties. The Lorenz' model was indeed originally suggested by the results of certain numerical experiments;[1] however the examples of §4.1 are already rather complex and it is convenient to look at the matter starting somewhat more systematically.

For the purposes of a more technical discussion it is convenient to confine ourselves, for simplicity, to studying motions of a fluid subject to constant external forces. In general, because of the presence of viscous forces, motions will asymptotically (as $t \to \infty$) dwell in a small part of the space of possible states which we shall call *phase space* (*i.e.* velocity, pressure, density, temperature ... fields): the phase space volume contracts, in a sense to be better understood in the following, and even large sets of initial data rapidly reduce themselves "in size", essentially no matter how we decide to measure their size.[2]

Sometimes, after a short transient time, large regions of phase space are "squeezed" on an "attracting set" consisting of a simple fixed point (*i.e.* of a time independent state of motion of the fluid), converging asymptotically

[1] The experiments (by Saltzman, *c.f.r.* §4.1 and [Lo63]) which indicated that sometimes the motion was well represented in terms of temperature and velocity fields with only three components on a natural basis.

[2] The size can, for instance, be measured by the volume occupied in phase space or in terms of its Hausdorff dimension, see §3.4: two possible notions with a very different meaning, as we shall see in Sect. 5.5.3 and in Ch. VII.

to it. Sometimes it will develop asymptotically on a periodic motion (hence converging to a closed curve in phase space), or to a quasi periodic motion (hence converging to a torus in phase space) or to an irregular motion (converging to a "*strange attracting set*").[3]

To visualize the above remarks think of the velocity, temperature, density, *etc*, by developing them over a suitable basis (for instance in the case of the NS equations with periodic boundary conditions the Fourier basis will be convenient). In this way every state of the fluid is represented by the sequence $x = \{x_i\}$ of the components of the coordinates on the chosen basis. A motion of the system will be a sequence $x(t) = \{x_i(t)\}$ whose entries are functions of time.

We shall say that the system has *apparently* "n degrees of freedom" if we shall be able to describe its state completely, at least for the purposes of our observations and of the precision with which we are planning to perform them, by n components of the sequence x. We shall also say that the system has "n degrees of freedom" or "n apparent modes". Imagining to increase the force we must think that, usually, the motions become more complicated and this can become mathematically manifest via the necessity of increasing the number of nonzero components, *i.e.* of nontrivial coordinates, of the vector $x = \{x_i\}$ describing the state of the system.

The attribute of "*apparent*" is necessary because, evidently, other coordinates could exist in terms of which it is still possible to describe the motions in a satisfactory way and which are *less in number*.

For instance a periodic motion of a fluid is described, in any coordinate system, by a periodic function $\underline{U}_{\underline{x}}(\varphi)$ of period 2π in φ and such that $\underline{u}(\underline{x},t) = \underline{U}_{\underline{x}}(\omega t)$, if $T = 2\pi/\omega$ is its period. It is clear that $\underline{u}(\underline{x},t)$ can also have many Fourier harmonics in $\underline{x}$, say n, of comparable amplitudes $\underline{\hat{u}}_{\underline{k}}(t)$; but, although the number of apparent degrees of freedom of the motion is at least n, the "true number" of degrees of freedom is, in this case, 1 because the "best" coordinate is precisely φ.

A first rough idea of the fluid motion and of its complexity can be obtained by just counting the number of "essential" coordinates necessary to describe it. As just noted this number might depend on the choice of the basis used to define the coordinates: but we should expect that one cannot describe a given motion with less than a certain minimum number of coordinates. We shall see that the attempt to define precisely the *minimum number* of coordinates naturally leads to the notion, or better to various notions, of fractal dimension of the region in phase space visited after a transient time. Such minimum number can also be called the "*dimension of the motion*" or the "*effective number of degrees of freedom*".

Adopting the above terminology (admittedly of phenomenological–empiri-

[3] Where the adjective "strange" refers more to the complex nature of the asymptotic motion than to a strange geometric structure of the attracting set: we shall see that very complex motions are possible even when the attracting set is a very smooth geometric surface.

cal nature that we shall attempt to make more and more precise) we can say

(1) A time independent motion will be described by coordinates x which are constant, or "fixed", in time: *i.e.* by a 'fixed point". An asymptotically time independent motion will appear as a motion in which the coordinates tend to assume a time independent value as $t \to \infty$.
(2) An asymptotically periodic motion will appear in phase space as attracted by a closed curve run periodically, *i.e.* by a curve described by a point $x(t)$ whose coordinates are all asymptotically periodic. For t large the motion will then be, with a good approximation, $t \to \{x_i(t)\} = \{\xi_i(\omega t)\}$ with $\xi_i(\varphi)$ periodic with period 2π, and the curve $\varphi \to \{\xi_i(\varphi)\}$, $\varphi \in [0, 2\pi]$, is its "orbit" or "trajectory".
(3) A quasi periodic motion with q periods $2\pi\omega_1^{-1}, \ldots, 2\pi\omega_q^{-1}$ will appear as a function $t \to x(t) = \{\xi_i(\omega_1 t, \ldots, \omega_q t)\}$ with $\xi_i(\varphi_1, \ldots, \varphi_q)$ periodic of period 2π in their arguments. And the geometric set $\underline{\varphi} \to \{\xi_i(\underline{\varphi})\}$, $\underline{\varphi} \in T^q$, is the *invariant torus* described by the quasi periodic motion. Evidently it will not be reasonably possible to use less than one coordinate to describe a periodic motion, nor less than q coordinates to describe a quasi periodic motion with q rationally independent periods.[4]
(4) A motion that is not of any of the above types, *i.e.* neither time independent, nor periodic, nor quasi periodic, is called "*strange*".

Until recently it was, indeed, believed that the phenomenon of the onset of turbulence was extremely simple, see [LL71], Chap. 3. In the just introduced language it would consist in successive appearances of new independent frequencies;[5] their number growing as the driving force increased, with a consequent generation of new quasi periodic motions. Until the number of "excited" modes and of frequencies (q in the above notations), had become so large to make it difficult identifying them so that a relatively simple quasi periodic motion could look, instead, as erratic and unpredictable.

In the above "Aristotelic" vision according to which an arbitrary motion can conceptually be considered as composed by (possibly very many) circular uniform motions a "different one" which does not take place on a torus and that cannot be regarded as composed by circular motions is therefore "strange".

To an experimentalist who assumes the quasi periodic viewpoint it can be not too interesting to study the onset of turbulence, because it would appear as consisting in a progressive "excitation" of new modes of motion: a dull phenomenon, differing in different systems only by insignificant details.

---

[4] Rational independence of $(\omega_1, \ldots, \omega_q)$ means that no linear combination $\sum_i n_i\omega_i$ with n_i integers vanishes unless $n_i = 0$ for all $i = 1, \ldots, q$.

[5] This means an increasing number of "modes" periodically moving with frequencies among which no linear relation with integer coefficients exists.

It is remarkable that the force of the above aristotelian–ptolemaic conception of motion, [LL71], Chap. 3,[6] has been such that no one had, until the 1970's, the idea of checking experimentally whether it was in agreement with the natural phenomena observable in fluids. This is most remarkable after it had become clear, starting with Copernicus and Kepler, how fertile was to subject to critique the astronomical conceptions on quasi periodicity: a critique which prompted the birth of modern Newtonian science, although at the beginning it did in fact temporarily reinforce the ancient astronomical and kinematical conceptions which attained the maximum splendor with the work of Laplace in which all celestial motions were still quasi periodic, *c.f.r.* [Ga99b]. And it is surprising that the quasi periodic theory of turbulence was abandoned only in the 1970's, 100 years after a definitive critique of the fundamental nature of quasi periodic motions had become available with the work of Poincaré.

Following the analysis of Ruelle–Takens, [RT71], [Ru89], [ER81], formulated at the end of the 1960's and preceded by the independent, and somewhat different in spirit, critique of Lorenz (*c.f.r.* comments at the end of (C) in §4.2, and [Lo63]), experimentalists were induced to perform experiments considered of little interest until then. They could in this way observe the ubiquity of "*strange motions*", governed by attracting sets in phase space that are neither fixed points, nor periodic motions, nor quasi periodic compositions of independent periodic motions.

The observations established that at the onset of turbulence motions have few degrees of freedom: observed after an initial transient time they can be described by few coordinates whose time evolution can be modeled by simple equations. But they also established that *nevertheless* such equations produce soon (as one increases the strength of the forcing) "strange" motions, which are not quasi periodic.

Furthermore from the works [Lo63], [RT71], and others, it emerged that motions can increase in complexity, still staying with few degrees of freedom, by following *few* different "*routes*" or "*scenarios*" which however can show up in rather arbitrary order in different systems giving thus rise to quite rich phenomenologies; much as the complexity of matter grows up from the "simplicity" of atoms.

This relies on an important mathematical fact: namely differential equations with few degrees of freedom often present few types of really different asymptotic motions and all that a given model does is to organize the appearance of the various asymptotic regimes according to a certain order.

---

[6] The reader might have trouble finding even traces of this once widely held point of view. In fact the Ch. 3 of [LL71] has been rewritten in the subsequent editions (postumous to one of the authors). While the new version is certainly closer to the modern views on turbulence it is clear that the complete elimination of the original theory will eventualy produce great problems to historians of science. A more sensible choice would have been, perhaps, to keep the "ancient" viewpoint in small print at the end of the new chapter.

An order which, unfortunately, remains to date *essentially inscrutable*: in the sense that, given a few degrees of freedom system, it is usually impossible to predict the order and the many quantitative aspects of its asymptotic dynamical properties without actual experimentation.

Nevertheless the small number of *a priori* possibilities allows us to use the existing theory also for a certain non negligible number of quantitative predictions and, very important, also for the conception and design of the experiments.

A basic mathematical instrument to analyze the scenarios of the onset of turbulence is *bifurcation theory.* Here we give a general idea of the most well known results: their complete (and general) mathematical analysis is beyond the scopes of this volume: see problems for a few more details.

Consider a system described by a mathematical model consisting in a differential equation on a phase space R^n with n dimensions

$$\underline{\dot{x}} = \underline{f}(\underline{x}; r) \tag{4.2.1}$$

depending on a parameter r, to be thought of as a measure of the strength of the driving force; the vector field $\underline{f}(\underline{x}; r)$ is a C^∞ function, to fix ideas, and such that *a priori* it is $|\underline{x}(t)| \le \overline{C}(\underline{x}(0))$ with $C(\cdot)$ a suitable function. We shall denote $\underline{f}(\underline{x}; r)$ also as $\underline{f}_r(\underline{x})$ for ease of notation.

With the above assumptions, the last of which is an *a priori* estimate which usually translates a consequence of energy conservation, there are no regularity or existence problems for the solutions. These are problems that we have already given up considering in general (see Ch. III and §4.1): accordingly it is clear that (4.2.1) admits regular global solutions, unique for given initial data.

Typically imagine that $\underline{x} = (x_1, \ldots, x_n)$ are the components of a development of the velocity field on a basis $\underline{u}_k(\underline{\xi})$ and of the temperature fields on another suitable basis: let the bases be chosen to represent correctly the boundary conditions, hence likely to require a small (if not minimal) number of "non negligible" components. For instance if the fluid is entirely described by the velocity field (as in the case of incompressible NS fluids) it will then be

$$\underline{u}(\underline{\xi}) = \sum_{k=1}^{n} x_k \underline{u}_k(\underline{\xi}) \tag{4.2.2}$$

where n depends on the complexity of the motion that one studies. In some cases, n can be very small: as an example consider equation (4.1.20) regarded as a description in terms of 3 "not negligible" coordinates of the large time behavior of certain solutions of Rayleigh's equations, (4.1.14).

Remarks.

(1) It is convenient to repeat that we have always in mind studying systems with *a priori* infinitely many degrees of freedom (like a fluid): and, nevertheless, such systems may admit asymptotic motions that can be described

with few coordinates whose *minimum number* n we call the *"effective number of degrees freedom"*. This reduction of the number of degrees of freedom is due to the dissipative nature of the motions that we study.

(2) Of course since the equations that we consider depend on a parameter r the number n will vary with the parameter. But if the parameter is kept in a prefixed (finite) range of values we can imagine n to be large enough to describe all asymptotic motions that develop when r is within the given range. This n might be larger than needed for certain values of r's.

(3) As the driving force increases we expect that the velocity field acquires a structure with more complexity so that the number of coordinates ("components" or "modes") necessary to describe it increases; with the consequent change of the asymptotic states of the system from what we called "laminar" to more interesting ones.

(4) The number of coordinates should not, nevertheless, increase too rapidly with the driving strength r (although, naturally, the number n could increase without limit as $r \to \infty$). In any event one always imagines that n can be taken so large to provide enough coordinates to describe the asymptotic motions in every prefixed interval in which one is interested to let r vary.

(5) At first sight this viewpoint is perhaps not very intuitive and, at the same time, it is very rich: we are in fact saying in a more precise form what stated previously, *i.e.* that the motion of the fluid can show a quite high complexity in spite of being describable in terms of few parameters. Or, in other words, that it is not necessary that the system be described by many parameters to show "strange" behavior.

Obviously it will be more common to observe "strange" motions in systems with many effective degrees of freedom: but motions of too high complexity are difficult to study both from the theoretical viewpoint and the experimental one while those that arise in systems with few (effective) degrees of freedom are easier to study and to classify, hence they provide us with a good basis for the theory and it is interesting that remarkable phenomena of turbulence are observable when n is still small. *A new viewpoint on the study of turbulence is thus born, in which the phenomenology of the onset of turbulence acquires an important theoretical interest.*

We can audaciously hope that in the future the phenomena that appear at the onset of turbulence can become the blocks with which to construct the theory of "strong" turbulence, with many degrees of freedom, in a way similar (perhaps) to the construction of a theory of macroscopic systems from the theory of elementary atomic interactions, *i.e.* in a way similar to Statistical Mechanics.

The possibility of this viewpoint emerged after the works of Ruelle–Takens and it seems supported by a large amount of experimental and theoretical checks.

(A) Laminar motion.

Coming back to (4.2.1) imagine n fixed and suppose that for $r = 0$ ("no driving") the system has $\underline{x} = \underline{0}$ as the unique time independent solution. We shall suppose that this solution is globally attracting in the sense that if $\underline{x}_0 \neq \underline{0}$ is any initial state then the solution of the equations of motion with $\underline{x}_0$ as initial datum is such that

$$\underline{x}(t) \xrightarrow[t\to\infty]{} \underline{0} \tag{4.2.3}$$

This approach to $\underline{0}$ will usually take place exponentially fast (*i.e.* $|\underline{x}(t)| < c_1\, e^{-c_2 t}, c_1, c_2 > 0$ for $t \to \infty$) and with a time constant equal to the largest real part of the eigenvalues of the "stability matrix" $J_{ij} = \partial_{x_i} f_j(\underline{0}; 0)$, also called the "*Jacobian matrix*" of the fixed point.

In fact for $\underline{x}_0 \simeq \underline{0}$ the equation appears to be well approximated by the linear equation $\dot{\underline{x}} = J\underline{x}$: then $|\underline{x}(t)| \simeq |e^{Jt}\underline{x}_0| \leq const\, e^{\mathrm{Re}\,\lambda\, t}|\underline{x}_0|$, if λ is the eigenvalue of J with largest real part.[7]

This property is, for instance, easily verified in the models of §4.1, *c.f.r.* problems [4.1.4],[4.1.5],[4.1.13].

This situation persists for small r, at least for what concerns the solutions that begin close enough to the laminar motion which for $r > 0$ is defined as the phase space point $\underline{x}_r$ such that $\underline{f}(\underline{x}_r; r) = \underline{0}$, where $\underline{x}_r$ is a regular function of r tending to $\underline{0}$ for $r \to 0$ (*c.f.r.* problems [4.1.6], [4.1.7]).

(B) Loss of stability of the laminar motion.

As r increases the picture changes and the "laminar" solution may lose stability when r reaches a certain critical value r_c. This happens when (and *if*) an eigenvalue λ_r, of the stability matrix $J_{ij}(r) = \partial_{x_i} f_j(\underline{x}_r; r)$, with largest real part reaches (as r increases) the imaginary axis of the complex plane on which we imagine drawing the curves followed by the eigenvalues of $J(r)$ as functions of r.

This can happen in two ways: either because $\lambda_{r_c} = 0$ or because $\mathrm{Re}\,\lambda_{r_c} = 0$ but $\mathrm{Im}\,\lambda_{r_c} = \omega_0 \neq 0$. In the second case there will be two eigenvalues ($\lambda_{r_c} = \pm i\omega_0$) reaching the imaginary axis (because the matrix J is real, hence its eigenvalues are in conjugate pairs).

In the first case *we should not expect*, in general, that for $r > r_c$ the time independent solution $\underline{x}_r$, that we are following as r increases, continues to exist. This in fact would mean that the equation $\underline{f}(\underline{x}_{r_c} + \underline{\delta}; r) = \underline{0}$ is soluble with $\underline{\delta}$ tending to zero for $r \to r_c^{\pm}$: an event that should appear "unreasonable" (to reasonable people) for the arguments that follow.

---

[7] In this respect the Grobman–Hartman theorem is relevant, *c.f.r.* problems [4.3.8], [4.3.9]: it, however, rather shows that identifying the nonlinear equation, near the fixed point, with its linearized version is not "always valid" and that it can be understood literally in the sense of its "generic" validity, *c.f.r.* (C) for a definition and an analysis of the notion of genericity).

Let $\underline{v}$ be the eigenvector of $J(r_c)$ with zero eigenvalue; then we can consider the graph of the first component f^1_r of the vector $\underline{f}_r$ in a coordinate system in which the n eigenvectors of $J(r_c)$ as coordinate axes, assuming for simplicity that $J(r_c)$ is diagonalizable and that $\underline{v}$ is the eigenvector with label 1; and the fixed point $\underline{x}_{r_c}$ is on the coordinate axis parallel to $\underline{v}$.

Then the graph, *c.f.r.* Fig. (4.2.1), of f^1_r as function of the component x_1 of $\underline{x}$ along $\underline{v}$, keeping the other components fixed to the value that they have in the point $\underline{x}_{r_c}$ taken as origin, will appear as a curve with a local maximum (or a minimum) near the origin: indeed "*in general*" the second derivative of f^1_r in the $\underline{v}$ direction will be nonzero near $r = r_c$ and one can therefore exclude the possibility of an inflection point (as a "rare" and "not general" event).

As r approaches r_c the maximum (or minimum) point on this curve gets closer to the origin and for $r = r_c$ coincides, by the described construction, with it and the curve appears tangent to the abscissae axis (*i.e.* to $\underline{v}$) and lies, locally, on the lower half plane (or upper if it is a minimum).

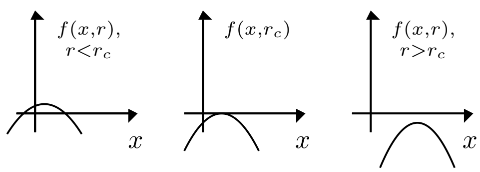


Fig. (4.2.1): *Loss of stability by fixed point through a stability matrix eigenvalue going through* 0. *There a collision between the stable fixed point (right) with an unstable one (left) and the two "annihilate" each other.*

Letting r increase it will happen, in general, that the graph of the curve will continue its "descent" (or "ascent") motion and the maximum (respectively the minimum), *c.f.r.* Fig. (4.2.1), will go below (above) the axis $\underline{v}$ hence it will no longer be possible to solve the equation $f^1(\underline{x}; r) = 0$, at least not with $\underline{x}_r \xrightarrow[r\to r_c]{} \underline{x}_{r_c}$. There will not exist, therefore, ("in general") solutions to $\underline{f}(\underline{x}; r) = \underline{0}$ near $\underline{x}_{r_c}$, for $r > r_c$.[8]

Naturally *there are also other possibilities*: for instance the "velocity of descent" of the maximum of the curve as a function of r could become zero for $r = r_c$ and change sign (in this case the fixed point solution would continue to exist). Or for $r = r_c$ the point $\underline{x}_r$ could become a horizontal inflection.

However the latter alternatives are not generically possible in the sense that, in order for them to happen, special relationships must hold between the coordinates of $\underline{x}_r$ and r for $r = r_c$; for instance the first possibility means

[8] Note that the above argument works in any dimension although it has a 1–dimensional flavour (and it is straighforward in 1 dimension). The possibility of reducing it to $d = 1$ is due to the hypothesis that J is diagonalizable so that the basis which is used exists. In general what this is saying is that the existence of a fixed point for $r > 0$, given that $\underline{f}(\underline{x}_0, 0) = \underline{0}$ cannot be inferred (and often does not exist) unless $\det J|_{r=0} \neq 0$, as well known from the implicit functions theorem.

that, if $\underline{x}_{r,max}$ is the point of a maximum for $f_1(\underline{x})$ as a function of x_1, then it must be $\partial_r f_1(\underline{x}_{r,max}) = 0$ for $r = r_c$ which in general has no reason to be true: *i.e.* there is no reason why *simultaneously* the derivatives of f^1_r with respect to *both* x_1 and r should vanish at $r = r_c, \underline{x} = \underline{r}_{r_c}$.

It is therefore worth pausing to clarify formally the notion of "*genericity*" that, so far, we have been using in an intuitive sense only.

(C) The notion of genericity.

We should first stress that we are considering various equations that are of interest to us because they are models of some physical phenomenon and we have already given up, *a priori*, to describe it *ab initio* by going back to the atomic hypothesis. Hence it should be possible to modify the equations of our models while still producing the same results within reasonable approximations.

For instance if $\underline{f}_r$ depends on certain empirical parameters it should not be possible to obtain qualitatively different results (at least if suitably interpreted) by changing the value of the parameters within the experimental errors. We want, in other terms, that the predictions of the model be insensitive to "reasonable" changes of the model itself.

Certainly, however, some properties of the models cannot be modified. In fact if, *a priori* , we know that fundamental principles imply that $\underline{f}$ should enjoy a certain property then modifications of the model that violate the property should be excluded.

Hence if $\underline{\partial} \cdot I\,\underline{f} = 0$ has, for a suitable matrix I, the meaning of energy conservation (as it is the case for Hamilton's equations) we shall not permit changes in the equation, *i.e.* in $\underline{f}$, that do not keep this property. If some conservation law requires that $\underline{f}(\underline{x})$ be odd in $\underline{x}$, then we shall not permit modifications of $\underline{f}$ violating such property (think for instance to the third law of dynamics that imposes that the force between two particles be an odd function of their relative position).

In every model of a physical phenomenon a few very special properties are imposed on $\underline{f}$: which usually translate physical laws, regarded as fundamental and whose validity should, therefore, not be discussed. After imposing such properties the function $\underline{f}$, *i.e.* the equations of motion, will have still (many) free parameters: and, then, the qualitative and quantitative aspects of the theory should not be sensitive to their small variations.

Suppose that a given problem is modeled by a differential equation $\dot{x} = f(x)$ with f in a space $\mathcal{F}$ of functions verifying all properties that *a priori* the system must verify (like *conservation laws* or *symmetries*). *If* one believes that *all a priori necessary properties are satisfied* one can then take the attitude, which appears the only reasonable one, that the significant physical properties do not change by changing a little f within the space $\mathcal{F}$.

"Little" must be understood in the sense of some metric measuring distances between functions in $\mathcal{F}$ and which should translate into a quantitative form which functions, on the basis of physical considerations, we are willing to consider a "small change". For instance

(1) If f must be a function of one variable of the form $f(x) = r + ax^2$ then $\mathcal{F}$ could be identified with the pairs (r, a) and a natural metric could be the ordinary distance in R^2.
(2) If f must be a function of the form $f(x) = r + \sum_{k=1}^{\infty} a_k x^{2k}$ with $0 \le a_k \le 1$ then $\mathcal{F}$ is the space of such functions and it could be metrized in various ways (which, however, give raise to quite different notions of distance); for instance one could define the distance between f^1, f^2 as $|r^1 - r^2| + \sum_k |a_k^1 - a_k^2| 2^{-k}$, or as $\max_k(|r^1 - r^2| + |a_k^1 - a_k^2|)$, *etc.*
(3) If f must be an analytic function, holomorphic and bounded for $|x| < 1$ and if $\{a_k\}$ are its Taylor coefficients (so that $f(x) = \sum_{k=0}^{\infty} a_k x^k$) one could define the distance between f^1 and f^2 as $d(f^1, f^2) = \sup_{|x|<1} |f^1(x) - f^2(x)|$ or (but not equivalently) $d(f^1, f^2) = \sum_k |a_k^1 - a_k^2| 3^{-k}$.
(4) If f is a function of class $C^1([0, 1])$ the distance could correspond to the metric induced by the norm on C^1, defined by the maximum of the modulus of f and of its derivative, or (not equivalently) by the norm defined by the integral of the absolute value of f plus that of its derivative.

With the above remarks and if $\mathcal{F}$ is a separable metric space,[9] we define

Definition *(stability and genericity): A property $\mathcal{P}$ of the functions $f \in \mathcal{F}$ is "stable" if, given an element $f \in \mathcal{F}$ which enjoys the property $\mathcal{P}$, it holds also for the elements close enough to f. A property is "*stable at $f \in \mathcal{F}$*" if it holds for all the elements close enough to f. A property $\mathcal{P}$ is "generic" in $\mathcal{F}$ if it holds for an open dense set in $\mathcal{F}$.*

Remark: In mathematics one often uses a somewhat different definition, calling generic also a property that holds on a set that, although not necessarily open and dense, is nevertheless a countable intersection of dense open sets:[10] it is not useful to discuss it here as in any event one should *not* attribute excessive importance to the details of the notion of genericity.

Example 1: Consider the space of the functions of the real variable x of the form $f(x) = r - ax^2$ with $(a, r) \in R^2$, and consider the equation $f(x) = 0$. Existence of a solution is a stable property if $r/a > 0$: but it is not generic because the complementary set $r/a < 0$ is open.

[9] Separability means that in the metric space $\mathcal{F}$ there is a denumerable dense set of points: it would be very inappropriate in our context to consider more general spaces, "never" met in applications.

[10] Which is still dense under very general assumptions on the space $\mathcal{F}$ in which the sets are contained, by "Baire's theorem", *c.f.r.* [DS60], chap. I: *e.g.* it is sufficient that $\mathcal{F}$ be a metric space which is separable and complete.

Example 2 (*nongenericity of the bifurcation through* $\lambda = 0$): As a more interesting example consider the functions $f(x;r)$ of class C^2 (*i.e.* with two continuous derivatives) such that
(a) for $r \le r_c$ there is a solution x_r of the equation $f(x;r) = 0$ continuously dependent on r, and
(b) $f'(x_r;r) < 0$ for $r < r_c$ and $f'(x_{r_c};r_c) = 0$ (here $f' \equiv \partial_x f$).
Considering the differential equation $\dot{x} = f(x,r)$ (a), (b) imply that for $r < r_c$ the time independent solution x_r is linearly stable for $r < r_c$ and becomes unstable (or, more accurately, marginally stable) at $r = r_c$.

Let $\mathcal{F}$ be the space of the above considered functions regarded as a metric space with some metric (*e.g.* one can define the distance between two functions as the supremum of $\sum_{j=0}^{2} |\partial_x^j (f-g)|$). Then generically in $\mathcal{F}$ it will be $\partial_x^2 f(x_{r_c}, r_c) \ne 0$ and $\partial_r f(x_{r_c}, r_c) \ne 0$.

Hence *generically* the graph of $f(x, r_c)$ in the vicinity of $x = x_{r_c}$ will have a local maximum (or minimum) either above or below the axis $f = 0$; and the local maximum (or minimum) $y(r)$ of f near x_r will have in $r = r_c$ a derivative $y'(r_c)$ *different from zero.* Hence the graph of f "evolves" with r as described by the graphs of the figure Fig. (4.2.1) above.

The Fig. (4.2.1) shows the nongenericity of the existence of a solution x_r of the equation $f(x;r) = 0$ continuous as a function of r for $r > r_c$, in the cases in which, for $r = r_c$, the derivative of f in x_{r_c} vanishes. This argument can be suitably extended to cases of equations in more dimensions.

Note that a solution of $f(x;r) = 0$ is a time independent solution $x = x_r$ of the equation of motion $\dot{x} = f(x;r)$; hence the stability loss when an eigenvalue of the stability matrix reaches 0 at $r = r_c$, *leads generically to the disappearance of the time independent solution from the vicinity of* x_{r_c}, as illustrated by the Fig. (4.2.1). We say that if a fixed point for a differential equation $\underline{\dot{x}} = \underline{f}(\underline{x};r)$ loses stability because an eigenvalue of the stability matrix reaches 0 then, generically, *there is no bifurcation*: *i.e.* the fixed point cannot be continued for $r > r_c$ and possibly coexist (unstable) with other stable nearby fixed points.

Different is the case in which f has properties that guarantee *a priori* the existence of the solution x_r for r near r_c: see the example 5 below.

Example 3 (genericity of the annihilation phenomenon): A more accurate analysis shows that instead, generically, the existence of a solution x_r continuous in r of $f(x,r) = 0$ for $r \le r_c$, with $f'(x_r, r) < 0$ and $f'(x_{r_c}, r_c) = 0$), is accompanied by the existence of another family, x'_r, of solutions of the same equation verifying however $f'(x'_r, r) > 0$ and such that $x_r - x'_r \xrightarrow[r\to r_c]{} 0$; this is made manifest, in the one–dimensional case, by the preceding Fig. (4.2.1). Generically the loss of stability due to the passage through 0 of a real eigenvalue of the stability matrix comes together with a "collision" (always as the parameter r varies playing the role of "time" in our description) between two solutions, one stable and one not, which reciprocally "annihilate" and disappear from the vicinity of the collision point for $r > r_c$.

Example 4 (symmetric bifurcation through $\lambda = 0$): However consider the space $\mathcal{F}_0$ of the C^2 functions $f(x,r)$ (in the sense of the preceding examples) *which are odd in* x and which for $r < r_c$ have a negative derivative at the origin and for $r = r_c$ have zero derivative. Then *it is no longer generically true* that the solution $x = 0$ disappears for $r > r_c$: it is always false, instead. This is not a contradiction because belonging to $\mathcal{F}_0$ *is not generic* in the space $\mathcal{F}$ of the functions of example 2. But it is, tautologically, *generic* in $\mathcal{F}_0$ being the very property that defines such space.

It is interesting to see that in the latter space the stability loss with an eigenvalue passing through zero is not only accompanied by the "survival" of the solution $x = 0$, generically becoming unstable, but also by the appearance of two stable solutions x^+_r, x^-_r approaching $x_{r_c} = 0$ for $r \to r_c$ and existing for $r < r_c$ *or* for $r > r_c$, depending on whether the sign of $f''(0; r_c)$ is positive or negative.

The proof can be obtained by contemplating the following drawing, Fig. (4.2.2), illustrating the one dimensional case and the result can be extended to more dimensions, *c.f.r.* problem [4.2.9].

This shows how delicate can be the discussion and the interpretation of "genericity"; the same property can be generic in a certain context (*i.e.* in a certain space $\mathcal{F}$) and not generic in another (*i.e.* in another space $\mathcal{F}_0$).

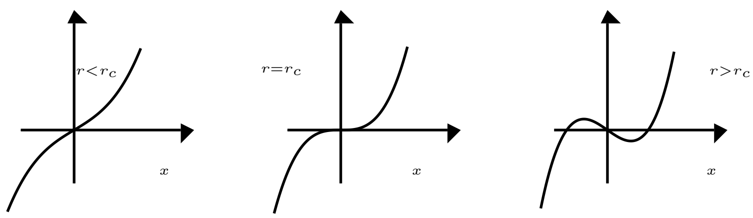


Fig. (4.2.2): *A bifurcation in presence of symmetry: the funcion f is supposed odd in x and the origin remains a fixed point even though it loses stability through* 0.

Interpreting the property $\underline{f}(\underline{x}, r) = -\underline{f}(-\underline{x}, r)$ as a "*symmetry*" of $\underline{f}$ we see that bifurcations "*in presence of a symmetry*" can be quite different and "non generic" when compared with the generic bifurcation behavior in absence of symmetry. This is a general fact.

Example 5: If S is an invertible C^k *map*, $k \geq 1$, of a compact C^k surface, *i.e.* a C^k "diffeomorphism", and if x is a periodic point for S with period τ, we say that x is *hyperbolic* if the stability matrix of x, regarded as fixed point of S^τ, does not have eigenvalues of modulus 1.

The δ–local *stable manifold* (or *unstable manifold*) of O is defined as the surface $W^{\delta,s}_O$ (or $W^{\delta,u}_O$) that

(a) is a graph defined on the ball of radius δ centered at O on the plane spanned by the eigenvectors of the stability matrix of S^τ in O corresponding to eigenvalues with modulus < 1 (or, respectively, with modulus > 1) and

(b) enjoy the property that $d(S^{n\tau}y, O) \xrightarrow[n\to+\infty]{} 0_{s,y}$ for $y \in W^{\delta,s}_O$ (or, respectively, $d(S^{-n\tau}y, O) \xrightarrow[n\to+\infty]{} 0$ for $y \in W^{\delta,u}_O$).

If S is of class C^k then one can prove that, if δ is small enough, $W_O^{\delta,s}$ and $W_O^{\delta,u}$ exist and are of class C^k and, furthermore, they depend regularly (in C^k) on parameters from which S possibly depends (assuming of course that S is of class C^k also with respect to these parameters), *c.f.r.* [Ru89b], p. 28.
An example of a classical genericity theorem is the following theorem (Kupka–Smale theorem, *c.f.r.* [Ru89b], p. 46.)

Theorem *(genericity of hyperbolicity and transversality of periodic orbits in C^∞ maps): Consider the set of all C^∞ diffeomorfisms of a C^∞–regular compact manifold M. Then*
(a) for all $T_0 > 0$ the set $\mathcal{I}_{T_0}$ of diffeomorphisms such that all periodic points with period $\le T_0$ are hyperbolic is open and dense.[11] *Each $f \in \mathcal{I}_{T_0}$ contains finitely many such periodic orbits: hence it is possible to define a $\delta = d_{f,T_0} > 0$ so that each periodic point for f with period $\le T_0$ admits a δ–local stable and unstable manifolds.*
(b) for all $T_1 > 0$ consider the set $\mathcal{I}_{T_0,T_1} \subseteq \mathcal{I}_{T_0}$ of diffeomorphisms such that if O, O' are points of two periodic orbits of period $\le T_0$ then $S^{T_1} W_O^{\delta,u}$ and $S^{-T_1} W_{O'}^{\delta,s}$ have intersections (if any) which are linearly independent, i.e. their tangent planes form an angle different from 0 or π. Then $\mathcal{I}_{T_0,T_1}$ is open and dense,
Hence the above hyperbolicity and transversality properties are two generic properties for each T_0, T_1.

We deduce that the set of C^∞–maps such that all their periodic points are hyperbolic and with global stable and unstable manifolds intersecting transversally is a countable intersection of open dense sets. In particular, by general results in the theory of sets they form a dense set (it is Baire's theorem, *c.f.r.* footnote[10]).

Example 6: (Hamiltonian equations) A Hamiltonian differential equation in R^{2n} (*i.e.* a vector field in R^{2n}) is not generic in the space of differential equations for motions in R^{2n}. One expects, therefore, in such systems to observe as generic phenomena various phenomena which are not generic in the space of all differential equations.

We conclude these remarks by noting another delicate aspect of the notion of genericity: at times, in real or numerical experiments, it may happen that one "forgets" to note an important property of a system, *e.g.* a symmetry property; in this case one might therefore be "surprised" to see occurrence of a nongeneric phenomenon that, being nongeneric, we do not expect to

---

11 If f, g are two C^∞ functions defined on a ball in R^n and $\delta(\xi) \stackrel{def}{=} |\xi|/(1+|\xi|)$ then the distance in C^∞ between f and g can be, for instance, defined by $d(f,g) \stackrel{def}{=} \sum_{k=0}^\infty 2^{-k}\delta(|\partial^k f - \partial^k g|)$ having set $|\partial^k f - \partial^k g| = \max_{x\in M} |\partial^k f(x) - \partial^k g(x)|$. By making use of an atlas of maps for the manifold M this naturally leads to a metric on the diffeomorphisms of M.

see. When this happens it is usually a very interesting event, because it may lead to the discovery of a missed conservation law or of a forgotten symmetry.

The idea that only generic properties may have relevance led, see [RT71], *to the modern viewpoint on turbulence theory and to the reduction and classification (partial but on the whole rather satisfactory) of the a priori possible phenomena occurring in incipient turbulence.*

The generality of the latter idea, its key role and its concrete use combined with the theory of dynamical systems to yield a general theory of turbulence, distinguish the ideas of Ruelle–Takens from those, which were formulated several years earlier, of Lorenz.

The latter did not, perhaps, fully stress the importance of the chaotic phenomena he discovered and he "confined" himself to stress the actual existence of chaotic motions[12] among solutions of simple equations describing for instance atmospheric evolution, and pointing out that in the theory of fluids equations one should expect instability in the form of nonperiodicity and a consequent impossibility of predictions of details of the asymptotic motion (as time t tends to ∞).

Instead Ruelle–Takens's viewpoint was more general (as they tried to relate *and prove* instability by means of the genericity notion) and was therefore better communicated and understood and immediately broadly tested. It is however also important to recall that while the observations of Lorenz were made at the beginning of the 1960's those of Ruelle–Takens were made at the end of the same decade when the electronic computers had become simple and common enough to be used to perform immediate large scale checks of the new ideas, which is indeed what happened, with important feedback on the further development of the very same ideas.

(D) Generic routes to the loss of stability of a laminar motion. Spontaneously broken symmetry.

Going back to the stability of laminar motion, regarded as a solution of the equation $\underline{f}(\underline{x}_r; r) = \underline{0}$, suppose that the stability loss is due to an eigenvalue of the stability matrix $J(r)$ of the fixed point $\underline{x}_r$ which reaches $\lambda_r = 0$ at $r = r_c$. In this case for $r > r_c$,, generically (see examples 2,4 in (C)), the laminar motion disappears and no motions remain confined near $\underline{x}_{r_c}$: even if the initial datum is close to $\underline{x}_{r_c}$ the asymptotic behavior will then be regulated by properties of the equations of motion in some *other region* of phase space, away from the point $\underline{x}_{r_c}$ (in whose vicinity, for $r < r_c$, wander the motions with initial data close to $\underline{x}_r$).

Hence we completely lose control of what happens because, in our generality, we do not know how the function $\underline{f}_r$ behaves far away from the fixed

[12] Simply and acutely called "nonperiodic", putting on the same level periodic and quasi periodic motions by distinguishing them from their "opposites", *i.e.* from the nonperiodic motions.

point that has "ceased to exist" at $r = r_c$.

It is now necessary to say that often systems have some symmetry and the laminar motion that exists for $r = 0$, and then is followed by continuity as r grows has, usually, the same symmetry. *This is the case in all examples in* §4.1, *c.f.r.* (4.1.21), (4.1.29), (4.1.31). Therefore it can happen that the losses of stability are "nongeneric": in the same way as in the example 4 of (C) above. Through them the initial fixed point may be replaced in its role of attracting set, by other fixed points endowed with *less symmetry* which, in turn, can lose further symmetries as r increases. This is a sequene of events that we can call *spontaneous symmetry breakings*,

An interesting class of examples of sequences of bifurcations "non generic" because of symmetry breaking is provided by the models in §4.1; see problems [4.1.7], [4.1.8] and in the successive §4.4).

The bifurcations can continue until, when r attains a certain value, *all symmetries are broken.* The fixed points that remained stable up to this value of r lose their stability, this time in a "generic" way. If it happens "through $\lambda = 0$" one cannot say much, *in general*, as already noted. However if the bifurcation takes place because a pair of conjugate eigenvalues cross the imaginary axis, then we can still examine a few "generic" possibilities, some of which fairly simple.

In fact a stability loss due to a pair of conjugate eigenvalues of the stability matrix $J(r)$ passing through the imaginary axis at $r = r_c$ does not destroy the existence of the fixed point. This is trivially so because in this case the stability matrix $J(r_c)$ *does not have zero determinant* so that, by the implicit functions theorem, one can still find a fixed point that for $r > r_c$ "continues" the fixed point $\underline{x}_r$, that lost stability at $r = r_c$, into a generically unstable one.

The eigenvalues of $J(r)$, for $r > r_c$, will contain, generically, two conjugate eigenvalues $\lambda_\pm(r)$ with positive real part unless $\partial_r \mathrm{Re}\,\lambda_+(r_c) = 0$, a non-generic possibility that we exclude. Possibly changing the definition of r we can suppose that $\mathrm{Re}\,\lambda_\pm = r - r_c$

$$\lambda_\pm(r) = (r - r_c) \pm i\omega(r) \tag{4.2.4}$$

where $\pm\omega(r)$ is the imaginary part, with $\omega(r_c) \equiv \omega_0 > 0$:

When $r > r_c$ the fixed point $\underline{x}_r$, being unstable, "repels" essentially all motions that develop close enough to it. If the *nonlinear terms* of the Taylor series of $\underline{f}$ in $\underline{x}_{r_c}$ force, when $r = r_c$, initial data close to $\underline{x}_{r_c}$ to approach as $t \to \infty$ the fixed point then for $r > r_c$ the linear repulsion of the fixed point is compensated at a suitable distance from the fixed point $\underline{x}_r$, by the nonlinear terms. In this case we say that the vector field $\underline{f}$ is *vaguely attracting* at $\underline{x}_{r_c}$.

In the simple 2–dimensional case (see problem [4.2.4]) one checks the existence of a "*suitable*" system of coordinates $\underline{x} = (x_1, x_2)$ in which

(a) the nonlinear terms $\underline{w}$ of $\underline{f}$ are, for $r = r_c$, at least of third order in $\rho = |\underline{x} - \underline{x}_{r_c}|$ and, at the same time,

(b) if $\underline{\rho} = (\rho_1, \rho_2)$ and $\underline{\rho}^\perp = (\rho_2, -\rho_1)$ the third order in $\rho = |\underline{x} - \underline{x}_{r_c}|$ has the form $\gamma\,\rho^2 \underline{\rho} - \gamma'\,\rho^2 \underline{\rho}^\perp$, up to higher orders in ρ, for some $\gamma(r), \gamma'(r)$ computable in terms of the first three derivatives of $\underline{f}$ at $(\underline{x}_{r_c}, r_c)$; hence

$$\underline{f}(\underline{x}) = (r - r_c)\underline{\rho} - \omega'(r)\underline{\rho}^\perp + \gamma\,\rho^2\,\underline{\rho} + O(\rho^4) \tag{4.2.5}$$

where $\omega' = \omega'(\rho, r) = \omega(r) + \gamma'(r)\,\rho^2$. Finding such a system of coordinates is always possible (and easy, if tedious, to do explicitly) as discussed in problem [4.2.4].

Hence the notion of vague attractivity is simply formulated *in the above coordinates* as a property of the sign of the third order terms in the Taylor expansion of $\underline{f}_r(\underline{x})$ in $\underline{x} - \underline{x}_{r_c}$ and $(r - r_c)$: if $\gamma < 0$ then for r small the fixed point $\underline{x}_{r_c}$ is *vaguely attractive* and the vectors $\underline{w}$ "point towards the origin", while if $\gamma > 0$ they "point away from the origin" which is then *vaguely repulsive* (*c.f.r.* problem [4.2.4] and [Ga83] Chap. 5).

In the cases of "*vague repulsivity*" (*i.e.* $\gamma > 0$) not much can be said for $r > r_c$ since the motion will abandon the vicinity of the fixed point that loses stability and we shall be in a situation similar to the one corresponding to the passage of an eigenvalue through 0. If regions far away from O evolve again towards O (arriving close to it along an attracting direction) one says that there is an "*intermittency phenomenon*": its analysis is very close to the case (H1) in §4.3 and, to avoid quasi verbatim repetitons we do not discuss it here.

However in the case of *vague attractivity* (*i.e.* $\gamma < 0$) and for $r > r_c$, but $r - r_c$ small, the distance ρ (in the special system of coordinates introduced above) is such that the first term in (4.2.5) $(r - r_c)\rho$ can be "*balanced*" by the cubic term $\gamma\rho^3$ which, by assumption, would tend to recall the motion to the fixed point: then a periodic orbit is generated, evidently stable.

In the chosen coordinates it is a circular orbit (up to orders higher than the third in the distance to the unstable fixed point) and with radius $O((-\gamma)^{-1/2}\sqrt{r - r_c})$. Indeed, in the special coordinates (x_1, x_2) that we consider, the transformation can be written, setting $z = x_1 + ix_2$, simply

$$\dot z = \big(r - r_c + i\omega'(|z|, r)\big)\,z - \gamma\,|z|^2\,z \tag{4.2.6}$$

with an approximation of order $O(|z|^4)$: showing the validity of the above statement on the radius and showing also that the period will be $\sim 2\pi/\omega_0$ if $\omega_0 = \omega(r_c)$, see [4.2.4], because if $\rho = \sqrt{-\gamma^{-1}(r - r_c)}$ the function $z(t) = \rho e^{i\omega'(\rho,r)t}$ is an exact solution of the approximate equation.

In the vague attractivity case the asymptotic motion with initial data close enough to the unstable fixed point $\underline{x}_r$, will therefore be of somewhat higher complexity compared to the one taking place for $r < r_c$: no longer evolving towards a fixed point but towards a periodic stable motion. And the system, that for $r \le r_c$ had asymptotic motions without any time scale (*i.e.* it had just fixed points) will now acquire a "*time scale*" equal to the period $\sim 2\pi\omega_0^{-1}$ of this periodic orbit.

What we said so far in the 2–dimensional case can be extended, without essential changes (see the problems), to higher dimensions and it constitutes the "*Hopf's bifurcation theory*". Basically this is so because the motion in the directions transversal to that of the plane of the two conjugate eigenvectors with small real part eigenvalues is a motion which exponentially fast contracts to 0 on a time scale regulated by the maximum of the real part of the remaining eigenvalues (which is negative) so that the motion is "effectively two–dimensional", see problems.

When the evolution (with r) leads, as r increases, to a periodic stable orbit we can proceed by analyzing the stability properties of the latter. Again several possibilities arise and we shall discuss them in §4.3.

Problems. Hopf bifurcation.

[4.2.1]: Let $\dot{\underline{x}} = \underline{f}_r(\underline{x})$, $\underline{x} \in R^n$, be a differential equation (of class C^∞), that for $r < r_c = 0$ has a fixed point that loses stability because two conjugate eigenvalues $\lambda(r) = r + i\omega(r)$ e $\bar\lambda(r) = r - i\omega(r)$, with respective eigenvectors $v(r), \bar v(r)$, as r increases, pass across the real axis with an imaginary part $\omega_0 = \omega(0) > 0$. Note that the hypothesis $\mathrm{Re}\,\lambda(r) = r$ is not very restrictive and it can be replaced by the more general $\frac{d}{dr}\mathrm{Re}\,\lambda(r_c) > 0$. (*Idea:* If the derivative dos not vanish one can change variable setting $r' = \mathrm{Re}\,\lambda(r)$ for r close to $r_c = 0$).

[4.2.2]: In the context of [4.2.1] suppose $n = 2$ and, if $x = (x_1, x_2) \in R^2$ show that the linearized equations can be written, setting $z = x_1 + ix_2$, as $\dot z = \lambda(r)z$. Furthermore the equations can be written to third order (*i.e.* by neglecting fourth order terms) as

$$\dot z = \lambda(r)z + az^2 + b\bar z^2 + cz\bar z + a_1 z^3 + a_2 z^2\bar z + a_3 z\bar z^2 + a_4\bar z^3$$

with a, b, c, a_i suitable complex numbers and $\bar z$ denotes the complex conjugate of z..

[4.2.3]: In the context of [4.2.2] show the existence of a coordinate change $z = \zeta + \alpha\zeta^2 + \beta\zeta\bar\zeta + \gamma\bar\zeta^2$ that allows us to write, near the origin and near $r = 0$, the equation truncated to third order in a form in which no second order terms appear. (*Idea:* This is just a "brute force" check. Check that the coefficients α, β, γ can be determined as wanted).

[4.2.4]: *(normal form vague attractivity theorem)* In the context of [4.2.1],[4.2.2],[4.2.3] assume that the equation truncated to third order does not contain quadratic terms (not restrictive by [4.2.3]).Show that it is possible a further change of coordinates $z = \zeta + \alpha_1\zeta^3 + \alpha_2\zeta^2\bar\zeta + \alpha_3\zeta\bar\zeta^2 + \alpha_4\bar\zeta^3$ such that the equation assumes the *normal form*

$$\dot\zeta = \lambda(r)\zeta + \vartheta|\zeta|^2\zeta + O(|\zeta|^4)$$

with $\vartheta = \vartheta(r)$ a suitable complex function of r. Check also that the value of $\mathrm{Re}\,\vartheta$ can be computed as a function only of the derivatives of order ≤ 3, with respect to $\underline{x}$ at $\underline{x} = \underline{0}$, of the function $\underline{f}_r(\underline{x})$ defining the differential equation. The time independent solution is said "vaguely attractive" if $\gamma \equiv \mathrm{Re}\,\vartheta < 0$ for $r = r_c$. (*Idea:* Same as in the preceding problem, *c.f.r.* [Ga83] §5.6, §5.7).

[4.2.5]: *(Hopf's bifurcation in 2 dimensions)* In the context of problem [4.2.4], and supposing $\gamma = \mathrm{Re}\,\vartheta < 0$ at $r = r_c$, show that the equation $\dot{\underline{x}} = \underline{f}_r(\underline{x})$ admits a periodic attractive solution which, in the coordinates defined in the preceding problem, runs on a curve with equation $|\zeta| = \sqrt{(r-r_c)(-\gamma)^{-1}} + O((r-r_c))$, for $r > r_c$ and $r - r_c$ small enough. The period of the motion is $\sim 2\pi/\omega_0$. (*Idea:* Note that this is obvious for the third order truncated equation. The non truncated case reduces to the truncated one via a suitable application of the implicit functions theorem (*c.f.r.* problem [4.2.10]), as the higher order terms are "negligible").

[4.2.6]: *(Inverse Hopf's bifurcation in* 2 *dimensions)* In the context of problem [4.2.4], and supposing $\gamma = \mathrm{Re}\,\vartheta > 0$, show that the equation $\dot{\underline{x}} = \underline{f}_r(\underline{x})$ admits a periodic repulsive solution which, in the coordinates defined in the preceding problem, runs on a curve with equation $|\zeta| = \sqrt{(r-r_c)\gamma^{-1}} + O((r-r_c))$, for $r < r_c$ and $r - r_c$ small enough. The period of the motion is $\sim 2\pi/\omega_0$. (*Idea:* as in the preceding problem).

[4.2.7]: *(Center manifold theorem)* Under the assumptions of [4.2.1], with $n > 2$, introduce the system of coordinates $\underline{x} = (x, y, \underline{z})$ with x–axis parallel to $\mathrm{Re}\,\underline{v}(0)$, y–axis parallel to $\mathrm{Im}\,\underline{v}(0)$ and $\underline{z} \in R^{n-2}$ in the plane orthogonal to that of $\underline{v}$, $\overline{\underline{v}}$. Suppose that for some $r_0, \nu > 0$ the remaining eigenvalues of the stability matrix of $\underline{x} = \underline{0}$ stay with real part not larger than $-\nu < 0$, for $r \in (-r_0, r_0)$.
Then for all $k \geq 2$ and a suitable $\delta_k > 0$ it is possible to find a 2–dimensional surface $\Sigma_k(r)$ of class C^k and equations

$$\underline{z} = \underline{\zeta}_r(x, y), \qquad |x|, |y| < \delta_k,\ |r| \leq \frac{r_0}{2}$$

such that

(i) if $|x_0|, |y_0| \leq 2\delta_k,\ |z_k| \leq \delta_k$ the solutions $t \to S_t(\underline{x}_0)$ (with initial datum $\underline{x}_0 = (x_0, y_0, \zeta_r(x_0, y_0)))$ of the equation remain on Σ_k if initially $\underline{x} \in \Sigma_k$ (*i.e.* Σ_k is S_t-invariant) and
(ii) there is D such that if $|S_\tau(\underline{x})| < \delta_k$ for $0 \leq \tau \leq t$ then $d(S_t(\underline{x}), \Sigma_k) \leq De^{-\nu t/2}$: *i.e.* the surface Σ_k is attractive for motions that are near the origin, as long as they stay near it; *i.e.* the motion gets close to Σ_k and it can go away from the origin only by "gliding" along Σ_k and
(iii) $|\underline{\zeta}_r(x, y)| \leq C(x^2 + y^2)$, with C a suitable constant.

This is a version of the *center manifold theorem*, *c.f.r.* [Ga83], §5.6, for instance.

[4.2.8] Show that the above theorem, combined with the results of the previous problem allows us to prove the following

Theorem *(Hopf bifurcation theorem) Suppose that for $r = r_c$ the differential equation in problem [4.2.1] has the origin as a fixed point which loses stability as r grows through r_c because of the crossing of the imaginary axis by two eigenvalues of the stability matrix and in the way described in [4.2.1]. Suppose that a suitable polynomial formed with the derivatives of $\underline{f}_r$ of order ≤ 3, evaluated at the time independent point and for $r = r_c$, is negative. Then for $r > r_c$ and $r - r_c$ small enough there is a periodic attractive solution located within a distance $O(\sqrt{r-r_c})$ from the origin; the period of this orbit is $\sim 2\pi/\omega_0$ and it tends to $2\pi/\omega_0$ as $r - r_c \to 0^+$.* (*Idea:* The center manifold theorem with k large (*e.g.* $k = 4$ is certainly enough) reduces the proof to the $n = 2$ case, precedently treated).

[4.2.9] (*Loss of stability through $\lambda = 0$ in higher dimension*) Consider a C^2 function $\underline{f}(\underline{x}, r)$ which for $r \leq r_c$ vanishes at a point $\underline{x}_r$ continuously dependent on r. Suppose that the stability matrix $\partial_{x_j} f_i = J_{ji}$ at the point $\underline{x}_r$ has only one real eigenvalue λ_r which, among the other eigenvalues, also has largest real part. Call the corresponding eigenvector $\underline{v}_r$ and, furthermore, suppose $\lambda_{r_c} = 0$. Using the center manifold theorem in [4.2.7] extend to dimension ≥ 2 the genericity and non genericity properties in examples 2,4 above.

[4.2.10] (*Implicit function to neglect higher orders in the Hopf bifurcation in* [4.2.5]) For simplicity suppose that $\omega(r) = \omega, \vartheta(r) = \gamma < 0$ are real constants. The parametric equation for the invariant curve of the equation in problem [4.2.4] can be written as

$$z(\varphi) = \rho_0\,(1+\delta(\varphi))\,e^{i(\varphi+h(\varphi))}, \qquad \rho_0 \stackrel{def}{=} (-\gamma^{-1}\,r)^{1/2}$$

where $\varphi \in T^1 = [0, 2\pi]$ and δ, h are functions in $C^1(T^1)$, and suppose that the solution to the equation in problem [4.2.4] is, without neglecting the higher order terms, equivalent to $\dot\varphi = \omega + \varepsilon$ for some constant ε. Check that the condition for this to happen has the form $(\delta', h', \varepsilon') = (\delta, h, \varepsilon)$ where $(\delta', h', \varepsilon') = H(\delta, h, \varepsilon)$ with H an operator on the space $C^1(T^1) \times C^1(T^1) \times R$ defined for $||\delta||_1, ||h||_1, |\varepsilon| < 2^{-1}$ by

$$\begin{aligned}
\delta' &= r^{3/2}\left(\partial_\varphi + 2r/(\omega+\varepsilon)\right)^{-1}\Big(-r^{-1/2}\delta^2(1+\delta)(\omega+\varepsilon)^{-1} + R_1(\varphi;\delta,h,\varepsilon)\Big)\\
h' &= r^{3/2}\,(\omega+\varepsilon)^{-1}\,\partial_\varphi^{-1}\Big(R_2(\varphi;\delta,h,\varepsilon) - \int_0^{2\pi} R_2(\psi;\delta,h,\varepsilon)\,d\psi/2\pi\Big)\\
\varepsilon' &= r^{3/2}\int_0^{2\pi} R_2(\psi;\delta,h,\varepsilon)\,d\psi/2\pi
\end{aligned}$$

with R_j, as well as the following A_j, B_j, D_j, C_j, a regular function in $r, \delta, h, \varepsilon, e^{i\varphi}$:

$$R_j(\varphi;\delta,h,\varepsilon) = A_j(\varphi) + B_j(\varphi;\delta,h,\varepsilon)\,\delta\, + C_j(\varphi;\delta,h,\varepsilon)\,h\, + D_j(\varphi;\delta,h,\varepsilon)\,\varepsilon$$

and the operators ∂_φ^{-1} and $(\partial_\varphi + 2r/(\omega+\varepsilon))^{-1}$ are defined as the multiplication by $(in)^{-1}$, $(in + 2r/(\omega+\varepsilon))^{-1}$ of the Fourier transforms of the operands. Note that the action of ∂_φ^{-1} is well defined because the operator acts on a 0–average function. Check that there is $c > 0$ such that H is a contraction in $C^1 \times C^1 \times R$ if $||\delta||_{C^1} < c$, $||h||_{C^1} < c$, $|\varepsilon| < c$ if r is small enough: hence H has a fixed point. For an alternative simpler approach see [Ga83]. (*Idea:* Note that ∂_φ^{-1} and $(\partial_\varphi + 2r/(\omega+\varepsilon))^{-1}$ can be written as integrals over φ once they make sense.)

Bibliography: [RT71],[Ga83]. A complete exposition of the bifurcation theory related to the above problems can be found in [Ru89b].

§4.3 Bifurcation theory. End of the onset of turbulence.

We shall study the stability of the periodic motions continuing the analysis of §4.2.

(E) Stability loss of a periodic motion. Hopf bifurcation.

Stability of periodic motions of the general equation (4.2.1) in R^n can be studied via a *Poincaré's map* defined as follows.

Consider a periodic motion Γ with period T_r as a closed curve in phase space and intersect it in one of its points O with a flat surface Σ transversal to Γ (*i.e. not* tangent to Γ) of dimension $n-1$, if n is the dimension of phase space. Imagine fixing on Σ coordinates $\underline{\eta} = (\eta_1, \ldots, \eta_{n-1})$ with origin O. As the intensity r of the driving force, *c.f.r.* (4.2.1), varies Γ, O, Σ change and Γ may even cease to exist; but if we study the system for r near some value r_c then Σ can be taken as fixed.

If $\underline{\xi}$ is a point of Σ in a vicinity U of the origin O, we imagine taking it as initial datum of a solution of the equation of motion $\underline{\dot x} = \underline{f}(\underline{x}; r)$, *c.f.r.* (4.2.1). It will follow closely the periodic motion on a trajectory that will come back to intersect Σ in a point $\underline{\xi}'$ after a time approximately equal to the period T_r of O if, as we shall suppose, U is chosen small enough.

1 Definition *(Poincaré map): We shall denote $S_\Sigma\underline{\xi}$ the point $\underline{\xi}'$ on Σ reached in this way starting from $\underline{\xi} \in \Sigma$. The map S_Σ, of the section Σ into itself, has O as a fixed point and is called a "Poincaré map". The map can be described in a system of cartesian coordinates $\underline{\eta} = (\eta_1, \ldots, \eta_{n-1})$ on Σ in which O has coordinates that we shall call $\underline{x}_r$. The matrix M of the derivatives in O of S_Σ with respect to the coordinates $\underline{\eta}$, i.e. $M_{ij} = \partial_{\eta_i} S_\Sigma(\underline{\eta})_j \,|_{\underline{\eta}=\underline{x}_r}$ will be called stability matrix of the periodic motion Γ.*

The stability matrix M (which is a $(n-1)\times(n-1)$ square matrix) besides depending on Γ depends also on the choice of O on Γ, on the choice of the section Σ through O, and finally on the system of coordinates defined on Σ. However it is not difficult to realize that the spectrum of the eigenevalues of M *does not depend on any of such choices*, *c.f.r.* problem [4.3.1].

The stability of the orbit is discussed naturally in terms of the map S_Σ. Indeed the orbit will be stable and exponentially attracting nearby points if the stability matrix M of Γ will have all eigenvalues with absolute value < 1. As the parameter r (which parameterizes the original equation and hence S_Σ itself) increases a loss of stability is revealed by one of the eigenvalues of M reaching the unit circle. We assume that this happens at $r = r_c$.

There are three *generic* possibilities

(1) an eigenvalue with maximum modulus of the matrix $M(r)$ reaches the unit circle in a nonreal point $e^{i\delta_c}$, $\delta_c \neq 0$, together with a conjugate eigenvalue (because $M(r)$ is real), while the others stay away of the real axis.
(2) the eigenvalue with maximum modulus is simple and reaches the unit circle in -1; or
(3) the eigenvalue with maximum modulus is simple and reaches the unit circle in 1.

Other possibilities (like more than two eigenvalues reaching simultaneously the unit circle) are nongeneric and will not be considered.

In the first two cases the determinant $\det(M(r_c) - 1)$ does not vanish and, by an implicit functions theorem, one finds that the fixed point $\underline{x}_r$ of the map S_Σ can be *"continued"* for $r > r_c$ as a, generically unstable, fixed point of S_Σ: which corresponds to the existence of an unstable periodic motion. This continuation, in the third case, is generically impossible as in the analogous case, examined in §4.2, of a fixed point for a differential equation that loses stability because an eigenvalue of the stability matrix reaches 0, see §4.2 and problem [4.3.2].

Consider the first case: it can happen that for $r = r_c$, *i.e.* at the moment of stability loss by the periodic motion Γ, the nonlinear terms of the Poincaré's map S_Σ are attractive (analogously to what we saw in §4.2 in discussing the vague attrattivity of fixed points). Then for fixed r slightly larger thn r_c

these terms will balance the linear repulsivity acquired by O and, at a certain distance from O, an invariant curve $\gamma = S_\Sigma \gamma$ is created on Σ: the action of S_Σ, "rotates" γ over itself with a *rotation number* approximately equal to $\delta_c/2\pi$ where δ_c is the argument of the complex eigenvalue $\lambda_{r_c} = e^{i\delta_c}$.[1] In this case one says that the periodic motion is vaguely attractive for $r = r_c$.

This means that motions with initial data on γ generate, in the n dimensional phase space, a surface $\mathcal{T}$ that is stable and is, topologically, an *invariant bidimensional torus* intersecting Σ on the closed curve γ. Furthermore a second time scale is generated, whose ratio to the period of the periodic orbit, still existing but now unstable, is approximately $2\pi/\delta_c$ and the motion on the invariant torus is quasi periodic or periodic.

We must expect, however, that as r increases beyond r_c quasi periodicity and periodicity alternate because the rotation number will continuously change from rational to irrational (in fact the map S_Σ, hence γ, change continuously and the rotation number depends continuously on the map, see problems [5.1.28], [5.1.29]). Below we shall see that rational values will be generically taken over small intervals of r so that the graph of the function $\delta(r)$ will look like a "smoothed staircase".

While the motion "turns" on $\mathcal{T}$ following essentially (at least for small $r - r_c$) the nearby periodic orbit Γ (by now unstable) the trajectory wraps around on the torus $\mathcal{T}$ to "reappear" right on the section Σ on the curve γ but in a point rotated *in the average* by an arc of length δ in units in which the curve Γ has length 2π. Thus the orbit closes approximately every $2\pi/\delta$ "revolutions" and this gives also the physical meaning of the rotations number.

Vague attractivity, as in the case of the fixed points of §4.2 can be discussed more precisely. For instance we consider the case in which the dimension of the surface Σ is 2 and the point O is chosen as origin of the coordinates x_1, x_2 in the plane Σ.

The coordinates can be combined to form a complex number $z = x_1 + ix_2$ and they can be chosen in analogy with the choice described in problem [4.2.4]. *This time, however, a further condition becomes necessary: namely that for $r = r_c$ none of the eigenvalues λ_i of the stability matrix shall verify*

---

[1] Since γ is invariant the Poincaré map S_Σ can be represented on γ as a map of a circle into itself. One simply identifies the point of γ with curvilinear abscissa (counted from an origin fixed on γ) equal to s with the point of the unit circle with angle $2\pi\, s/\ell$, where ℓ is the length of γ; and the map S_Σ becomes a regular map $s \to g(s)$ of the circle into itself. As we shall see, *c.f.r.* (4.3.1), the curve γ is in suitable coordinates very close to a circle and the map $g(s)$ is very close to a rotation, hence $g(s)$ is monotonically increasing and the map is invertible as a map of the circle into itself. Imagining of "developing" the circle into a straight line the map S_Σ becomes a function g defined on $(-\infty, +\infty)$ such that $g(s + 2\pi) = g(s) + 2\pi$. Then, a theorem of Poincaré, *c.f.r.* problem [4.3.3], gives the existence of the limit $\rho = \lim_{n\to\infty} (2\pi\, n)^{-1} g^n(s)$ and its independence on s. It is natural to call this number ρ the "*rotation number*" of the map g: it is indeed the fraction of 2π that in the average the iterates of g impose to the points on the circle at every action. A perfect rotation $s \to s + \delta$ will have rotation numer $\delta/2\pi$.

$\lambda_i^3 = 1$ *or* $\lambda_i^4 = 1$.

Assuming in what follows the latter (generically true) condition a change of coordinates can be devised which will be at least of class C^1 and such that the map S_Σ assumes, in the new coordinates, the form

$$S_\Sigma z = \lambda(r)\, z\, e^{c(r)\,|z|^2 + O(z^4)} \tag{4.3.1}$$

where $\lambda(r_c) = e^{i\delta_c}$, $c(r)$ is a complex number, and $\lambda(r), c(r), O(z^4)$ are functions of z, r of class C^1 at least.

If Re $c(r_c) = \overline{c} < 0$ one has *vague attractivity* (the case in which third order terms neither attract nor repel for $r = r_c$, *i.e.* $\overline{c} = 0$, is not generic). The analysis is entirely analogous, *mutatis mutandis*, to the corresponding one in §4.2.

We suppose that the r–derivative of the absolute value part of $\lambda(r)$ at $r = r_c$ does not vanish so that it is not restrictive to suppose that also

$$\lambda(r) = e^{r - r_c + i\delta(r)} \tag{4.3.1a}$$

Neglecting $O(z^4)$ the circle $r - r_c - \mathrm{Re}\, c(r)\, |z|^2 = 0$ is exactly invariant, by (4.3.1), and the motion on it is a rotation by an angle ϑ about equal to $\delta(r_c)$ (precisely equal to $\delta(r) + \mathrm{Im}\, c(r)\, |z|^2$). This implies that there is an invariant curve which is, approximately, a circle of radius ρ which the map S_Σ, approximately, rotates by an angle ϑ

$$\rho = \big(-\, \overline{c}^{-1}\, (r - r_c) + O((r - r_c)^2)^{1/2}, \qquad \vartheta = \delta(r_c) + O(r - r_c) \tag{4.3.2}$$

because the terms of $O(z^4)$ cannot change too much the picture, as it can be seen by an application of some implicit function theorem, see problem [4.3.10].

Generically the above condition on the powers 3 and 4 of the eigenvalues λ_i will be verified; hence, in the vague attractivity case, an invariant torus is generated which is run by motions with two time scales (*for* $r - r_c$ *small positive and up to infinitesimal corrections in* $r - r_c$) respectively equal to $T_{r_c} = 2\pi\omega_0^{-1}$ (period of the periodic motion that lost stability), and $T_1 = 2\pi T_{r_c} \delta^{-1}$. This is again called a *Hopf bifurcation c.f.r.* problems of §4.2. As r decreases towards r_c the torus becomes confused, one says "collides", with the orbit Γ from which it "inherited" stability for $r > r_c$.

If, instead, the nonlinear terms of the map S_Σ tend, for $r = r_c$, to move trajectories away from O (*i.e.* if in the special coordinates mentioned above we have $\overline{c} > 0$) we say that there is *vague repulsivity* and in this case the nonlinear terms (generically of order of $r - r_c$) cannot balance the linear ones and the motion gets away from the vicinity of the unstable periodic orbit.

We are in a situation similar to the one met in §4.2 when a generic loss of stability took place because the eigenvalue with largest real part reached 0.

As in that case, if $\overline{c} > 0$, one still finds a curve invariant for S_Σ, *unstable* and existing fo r $r < r_c$: this therefore implies that in phase space there is an invariant unstable torus which for $r \to r_c$ merges (one says it "collides") with the periodic orbit which inherits its instability for $r > r_c$.

The nature of the bifurcation in the non generic cases in which $\lambda_i^3 = 1$ or $\lambda_i^4 = 1$ is quite involved and not yet completely understood.

At this point *one could be led in temptation* and think that the motion on the invariant torus, born via a vaguely attractive Hopf bifurcation from a periodic orbit, is run (possibly only generically) by *quasi periodic* motions with two angular velocities approximately equal to $\omega(r), \delta(r)$ and which vary with continuity. Naively we could think that in general the rotation number of the motion on the invariant curve γ will be $\delta^*(r)/2\pi$ for some $\delta^*(r)$ tending, see (4.3.2), to δ_{r_c} as $r \to r_c$, *in a strictly monotonic way*, at least if $\partial_r \mathrm{Re}\,\lambda(r_c) > 0$.

A more attentive exam of this thought would, however, reveal it a rushed conclusion. If it was so, indeed, the rotation number would evolve with continuity assuming, as $r > r_c$ varies near r_c, values at times rational and at times irrational. If for a certain r we have a rational rotation number this implies that the motion on γ is, in reality, asymptotic to a periodic motion. Then we imagine the map S_Σ to describe, as r varies, a curve in the space of the smooth maps of the circle: such maps have the generic property that when their rotation number is rational the motions that they generate are asymptotic to a finite number of periodic motions which are either stable or unstable with a stability matrix which is not equal to 1.[2]

Therefore we *cannot* expect that the torus is covered by periodic orbits as it could be when the rotation number varying continuously and strictly monotonic becomes rational: by doing so we take too fast for granted the same picture as in the cases in which the rotation on the torus is *exactly* linear.[3]

Indeed if we think that as r changes the curve described by S_Σ enters and exits, in the space of circle maps, open regions where the motions are regulated by a few attracting periodic orbits (with a stability eigenvalue which is < 1) then their asymptotic motions will be periodic and will keep a "*fixed*" period as r varies in (possibly very tiny) intervals: because the period, being an integer, cannot vary by a small amount. The torus will be covered, possibly, by periodic orbits only at the extreme values of such intervals of r.

We should expect that as r increases the rotation number will remain fixed

---

[2] This is a very special case, intuitively appealing, of a more general theorem, *Peixoto's theorem*, that says that generically a differential equation on a 2–*dimensional* torus has a finite number of periodic solutions or fixed points, some of which attractive and some repulsive.

[3] Note If the rotation number is rational and the rotation is linear the torus is covered by a continuum of distinct periodic orbits. If the rotation number is irrational the torus is, instead, run densely by the motion on it which, therefore, will not have any non trivial attracting sets.

over segments with positive length > 0 on the axis r, and on such segments it will keep a rational value.

When this happens one says that in such intervals of r a *phase locking phenomenon* or *resonance* takes place between the two rotations (the one along the periodic orbit and the one on the section Σ).

This will not, however, forbid that the rotation number varies with continuity with r (because *natura non facit saltus*, as is well known). But the graph of the rotation number as a function of r will have a characteristic aspect, appearing piecewise *flat* and *hence* not strictly monotonic. Such a graph is often called a *Devil's staircase* (an obscure etymology because it is unclear why the latter Being could have spared time to dedicate himself to building such stairs, see problem [4.3.6]).

The case in which at stability loss of the fixed point is vaguely repulsive implies that the motion wanders away from the unstable orbit and not much can be said in general. If regions far away from O evolve again towards O (arriving close to it along an attracting direction) one says that there is an "*intermittency phenomenon*": its analysis is very close to the case (H1) below and, to avoid quasi verbatim repetitons we do not discuss it here.

(F) Loss of stability of a periodic motion. Period doubling bifurcation.

This is the case when the stability loss of the periodic orbit is due to the crossing through -1, for $r = r_c$, by an eigenvalue of the stability matrix $M(r)$. Again we must distinguish the *vaguely repulsive* case, (in which the nonlinear terms repel away from the fixed point for $r = r_c$), from the opposite case, *vaguely attractive*.

The case in which the third order terms do neither action will not be considered because it is nongeneric.

In the first case, again, for $r > r_c$ the motion simply drifts away from the periodic motion which has become unstable and nothing more can be said in general: an intermittency phenomenon can happen also in this case if regions far from O are mapped back close to O and its features can be discussed as in the case (H1) below. In the second case instead one easily sees that a new stable *periodic point* for S_Σ is generated, with period 2.

Indeed one can consider a point P near O and displaced by a small δ in the direction of the eigenvector $\underline{v}$ corresponding to the eigenvalue of M which is approximately -1. Then P returns close to O roughly on the *opposite side relative to* O (with respect to $\underline{v}$) and at a distance $\delta' > \delta$ a little further away than δ from the origin.

Iterating the point cannot get too far because, by the assumed vague attractivity, the cubic terms will eventually balance the tendency to get away from the linearly unstable O:[4] the motion therefore settles into an "equilibrium state" visiting alternatively at each turn, *i.e.* at every application of

[4] A proof is similar, but much simpler, to the one suggested in (E) in the analogous case of the Hopf bifurcation, *c.f.r.* problem [4.3.10] below.

the map S_Σ, two points located where the linear repulsion and the nonlinear attraction compensate each other. This is a periodic orbit for the map S_Σ with period 2 which is the intersection between the plane Σ and a periodic trajectory in phase space with a period of about twice[5] that $2\pi/\omega_0$ of the orbit that lost stability.

Geometrically this periodic orbit will be very close to the unstable orbit and almost twice as long: this explains the name of "*period doubling bifurcation*" that is given to this phenomenon.

(G) Stability loss of a periodic motion. Bifurcation through $\lambda = 1$.

This case is analogous to the already encountered one of fixed points of the differential equation (4.2.1) losing stability for $\lambda = 0$: the fixed point O_r for S_Σ generically ceases to exist at $r = r_c$; and for $r > r_c$ the motion migrates away from the vicinity of the periodic point that we followed up to $r = r_c$ and, in general, nothing can be said. Obviously in models with symmetries we can repeat what already said for the fixed points of (4.2.1) and it is possible that before the "*jump into the dark of the unknown*" (*i.e.* the migration of the motions towards more stable attracting sets, far away from the site in phase space where the orbit Γ_c that lost stability was located), a few nongeneric bifurcations may develop, accompanied by symmetry breakings. Or it is possible that regions far from Γ_{r_c} are again mapped close to Γ_{r_c} and an intermittency phenomenon arises: we describe it in the very similar case (H1) below.

(H) And then? Chaos! and its "scenarios".

When motions get far away from motions that have lost stability various possibilities arise: the simplest is when motions repelled by fixed points or periodic motions that have become unstable get near attracting sets, possibly far in phase space, that are still simple, *i.e.* of the same type of those considered so far (fixed points or periodic motions). In this case, as r changes further, all what has been said can be proposed again *in the same terms.* And the "real" problem is "postponed" to larger values of r. But things may be different, possibly after one ore more repetitions of the above analysis, and the examples that follow provide an interesting illustration.

(H1) Intermittency scenario.

Consider a periodic orbit Γ_r that loses stability at $r = r_c$ because of the crossing through 1 or -1 of an eigenvalue of the stability matrix. Generically the periodic orbit will disappear if the crossing is through 1 while if the crossing is through -1 it will persist unstable, varying continuously with r.

The Poincaré map S_Σ continues to contract approximately in all *but one* direction for $r - r_c$ small. The direction will be approximately that of the eigenvector $\underline{v}_c$ of the stability matrix $M(r_c)$ of S_Σ at the fixed point O on

[5] Approximately because $r \neq r_c$.

which the map S_Σ is defined for $r \le r_c$. We call O_c the position of O at $r = r_c$.

And motion in that direction while not having small amplitude might still be essentially one dimensional and S_Σ can sometimes be considered as a map of a segment into itself. This will happen if the nonlinear terms of S_Σ at $r = r_c$ tend to recall, again, near O_c the motions that go far enough from $O_c \in \Sigma$.

We then see that the motion appears as having an approximately periodic component (developing near the orbit Γ_{r_c} that has lost stability) and a component transversal to it.

One can also say that it looks as if the orbit Γ_{r_c} lied (approximately if $r > r_c$) on a 2–dimensional invariant surface intersecting Σ along a curve τ approximately tangent to the direction $\underline{v}_c$ of the eigenvector corresponding to the real eigenvalue that has reached the unit circle.

At "each turn" the motion returns essentially to this curve τ, because the map S_Σ does not cause the motion to go far away from it, being unstable only in the direction $\underline{v}_c$ and *having assumed* that S_Σ brings back to near O_c points of Σ that are far from it (at least for r only slightly beyond r_c).

On the curve τ (that, near O_c, is very close to the straight line parallel to the eigenvector $\underline{v}_c$) we define the abscissa s and use it to describe the points of τ; let $s^* = s(r_c)$ be the abscissa of the point O_c of τ (we can imagine that $s^* = 0$).

In this case it can, sometimes, be a good *approximation* to consider the projection of the motion on the line parallel to the direction $\underline{v}_c$. Denoting always with s the abscissa on this line it will be possible to describe the motion by a map $s \to g_r(s)$. We are interested in the cases in which initial data close to the fixed point that has lost stability do not go too far away along this line, but remain indefinitely confined to a segment, say to $[-s_0, s_0]$, moving back and forth on it.

A picture illustrating the phenomenon is the following; represent the curve τ as a segment in the direction of $\underline{v}_c$ and a family of maps of the segment into itself which leads to a loss of stability at $r = r_c$ with eigenvalue 1, remaining attractive away from the point $s^* = 0$.

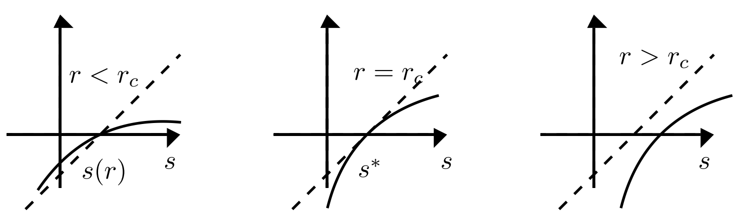


Fig. (4.3.1) *Case with eigenvalue +1. The s axis represents τ, the point $s(r)$ the fixed point; the map tends, for $r > r_c$, to send points away from s^*; therefore if the non linear terms eventually prevail the map should look quite different further away from s^*. Below in Fig. (4.3.2) a more global picture is presented. The case with eigenvalue −1 is simpler and is illustrated in Fig. (4.3.3).*

Reducibility of the motion generated by S_Σ on the section Σ can, however, *only be an approximation* because the map of $[-s_0, s_0]$ in itself will have necessarily (see below) to consist in an expansion of $[-s_0, s_0]$ followed by a *folding* of the interval into itself, *c.f.r.* Fig. (4.3.1). Therefore there will exist points $s = g_r(s_1) = g_r(s_2)$ that are images of distinct points s_1 and s_2. Hence if the motion did really develop on an invariant surface intersecting Σ on the curve τ then the uniqueness of solutions for the differential equation, that regulates the motions that we consider, would be violated. In Fig. (4.3.1) the "folding" is not illustrated, *i.e.* one should imagine that it happens at a distance from s^* not shown in the Fig. (4.3.1): see Fig. (4.3.2) for an example.

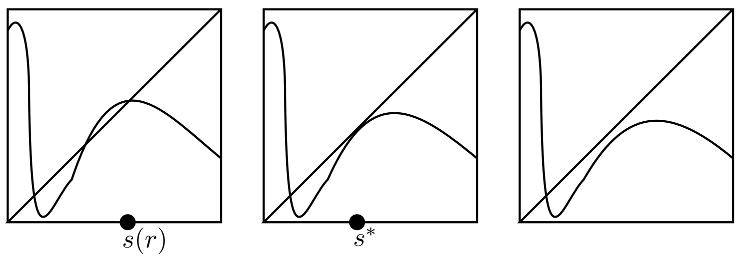


Fig. (4.3.2) *An example of an intermittency phenomenon: the picture represents the graph of $g(s)$ over $[-s_0, s_0]$ abd the square containing the graph is usually drawn as a visual aid to see "by inspection" that the interval $[-s_0, s_0]$ is invariant. The diagonal helps visualizing the fixed points and the iterates of the map.*

Hence the "one dimensional" description cannot be but approximated and in reality one must think that the segment $[-s_0, s_0]$ consists of several (infinitely many) segments which are essentially parallel and extremely close, so that they are not distinguishable by our observations. Nevertheless in this approximation we have the possibility of thinking the motion as 1–dimensional, at least in some cases in which the loss of stability is due to a single eigenvalue crossing in 1 or -1 the unit circle.

In the case of *crossing through* 1 we can imagine that g_r has a graph of the above type: respectively before the stability loss, at the stability loss and afterwards.

Stability loss will then be accompanied by an *intermittency phenomenon*, in fact for $r > r_c$ the map g_r that for $r = r_c$ was tangent to the diagonal will not have any longer points in common with the diagonal near the last tangency point and therefore motions *will spend a long time* near the point $s^*(r_c)$ of "last existence" of the fixed point (*no longer existent* for $r > r_c$). Then they will go away and, possibly, return after having visited other regions of phase space (that under the assumptions used here reduces to an interval).[6]

[6] Unless, of course, the picture differs from that in Fig. (4.3.2) and the points are attracted by *other* stable fixed points or periodic orbits: in such cases, however, we can repeat the whole analysis.

In this situation the motion does not appear periodic nor quasi periodic: there are time intervals, even very long if $r \sim r_c$, in which it looks essentially periodic. This is so because the motion returns close to what for $r = r_c$ was still a fixed point of S_Σ and which will still be such in an approximate sense if $r - r_c > 0$ is small.

Such long time intervals are followed, because of the phenomenon illustrated by the third of Fig. (4.3.2), by erratic motions in which the point gets away from the periodic orbit that has lost stability until it is again "captured" by it because the map is, far from $s^*(r_c) = 0$, still pushing towards s^*. One checks, indeed, that for small $r - r_c > 0$ a number $O((r - r_c)^{-1/2})$ of iterations of S_Σ are necessary to get appreciably far away from the point with abscissa $s^*_{r_c}$ (which for $r > r_c$ no longer is a fixed point).

In this way "chaotic" or "turbulent" motions arise following what is called the "*Pomeau–Manneville scenario*" or "*intermittency scenario*", [PM80], [Ec81].

(H2) Period doubling scenario.

A different scenario develops when the stability loss is due to the crossing by an eigenvalue of $M(r)$ of the unit circle through -1. In this case the one dimensional representation described in (H1) will still be possible, but the graph of g_r, having to cut at 90^o the bisectrix (of the second and third quadrants of the plane (s, g_r)) at $r = r_c$ (so that $g'_{r_c}(0) = -1$), cannot become tangent to the graph of S_Σ, for $r > r_c$ and $r - r_c$ small.

This time the unstable fixed point O with abscissa $s(r)$ on the curve τ continues to exist even for $r - r_c > 0$ (at least for a while) because the stability matrix of O does not have an eigenvalue 1 for $r = r_c$ (hence the fixed point continues to exist as guaranteed by an implicit functions theorem). We imagine that the absolute value of the derivative of g_r at the unstable fixed point increases: and we can then apply a general theory of bifurcations associated with interval maps whose graph has the form in Fig. (4.3.3)

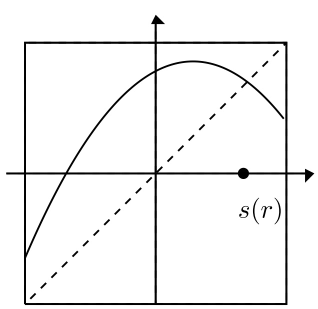


Fig. (4.3.3): *This illustrates a map g with a fixed point $s(r)$ with a stability eigenvalue about -1, and smaller than 1 in modulus. As r varies one should imagine that the slope at $s(r)$ reaches for $r = r_c$ the value -1 and then grows further (in modulus) for $r > r_c$ giving rise to a period doubling bifurcation.*

which is one of the simplest forms that allow us to define a map that stretches and then folds on itself (another has been illustrated in Fig. (4.3.2)).

This theory, due to Feigenbaum, predicts, not "generically" but "quite generally", that the first period doubling bifurcation is followed by an *infinite chain*, appearing within a short interval of r, *of period doubling bifurcations*, in which the initial orbit successively doubles in shape and period. The intervals of r at the end of which the successive bifurcations take place contract exponentially in size with a ratio asymptotically equal to 1/4.68.. which is a universal constant (*i.e.* independent of the system considered within a vast class of possible systems) and is called the *Feigenbaum constant*.

Chaotic, nonperiodic, motions appear (or can appear) at the end of the chain of period doublings: and this transition to chaos is called the *Feigenbaum scenario* or *period doubling scenario*, *c.f.r.* [Fe78],[Fe80], [Ec81],[CE80].

Of course if the loss of stability occurs with O being vaguely repulsive *an intermittency phenomenon* is possible if the non linear terms can still be such to map regions far away from O back to the vicinity of O. Its discussion is completely analogous to the one in (H1).

(H3) The Ruelle–Takens scenario.

It remains to consider what happens when some Hopf's bifurcation generates a torus and, as r increases this torus loses stability. This torus will appear in a Poincaré's section as a closed curve γ (intersection between the torus and the section surface Σ).

Attempting to generalize the notion of Poincaré's map to analyze stability of the quasi periodic motion is doomed to failure because the map should transform a curve γ_0 (*i.e.* a small deformation of the invariant curve γ, representing the intersection between the surface Σ and the torus of which we study the stability) into a new curve γ_0', image of the former "after one turn". The failure is not so much to ascribe to the fact that such a map would act on an infinite dimensional space (acting upon curves in Σ rather than on point on Σ) but, rather, because in general *it will not be well defined* since new motions can develop on the torus.

For instance small periodic orbits or invariant sets appearing on the surface of the torus or nearby make the extension of Poincaré's map, that we would like to define, to depend in an important way on the choice of Σ and can make it ill defined (while one notes that in the preceding case of the stability analysis for a periodic orbit all sections are equivalent). For instance if there is an invariant set E on the torus which has a small diameter and the surface Σ does not intersect it we cannot derive from the behavior of the Poincaré's map, any instability phenomenon due to motions that stay inside E, simply because their trajectories never visit Σ.

Often one finds that the two time scales associated with the motion on the torus become commensurate (as r varies), *i.e.* one has *resonance* or *phase locking*, *c.f.r.* (F), and motion is asymptotically periodic and not

quasi periodic.
Or it may happen that the loss of stability generates attracting sets that are contained in a surface of higher dimension which is not a torus or even if it is a torus the motion on it, generically, will *neither be periodic nor quasi periodic* making it difficult to build a general theory of the motions.

The problem is real because, as shown in [RT71], the stability loss of a bidimensional torus $\mathcal{T}_2$, *even admitting that it generates an invariant three dimensional torus* $\mathcal{T}_3$, is followed, generically, by a motion on $\mathcal{T}_3$ that is not a quasi periodic motion with three frequencies (as in the classic *aristotelic scenario*) but rather by a motion regulated by a *strange attracting set*, *i.e.* by an attracting set which is neither a time independent point nor a periodic orbit nor a quasi periodic motion. Hence the stability of an invariant torus provides us with a third scenario, called the *scenario of Ruelle–Takens*, because they pointed out the above stated genericity of strange motions generated by the loss of stability of a quasi periodic motion with two frequencies, [RT71],[Ec81].
To be precise in [RT71] it is shown that if for some reason an invariant torus of dimension *four*, only later reduced to three, is generated at a certain value of the parameter r, then generically motion on it will *not* be quasi periodic. And the point of the work was to criticize the idea that it could be common that a 2–dimensional torus run quasi periodically bifurcates into a 3–dimensional torus also run quasi periodically which in turn bifurcates into a 4 dimensional torus run quasi periodically and so on.

(I) Conclusions:

The three scenarios just discussed do not exhaust all possibilities but they cover fairly well the instabilities that are generically possible when a periodic motion or a quasi periodic one (with two frequencies) become unstable.
This is so at least in the cases in which the stability loss is due to the passage through the unit circle of either one real eigenvalue or of two conjugate ones of the stability matrix of the Poincaré map, while the others remain well inside the unit circle. Hence it is reasonable to think that the unstable motion generated on the surface on which the Poincaré map is defined is, within a good approximation, one dimensional and developing on a segment in the first case or on a closed curve in the second.
There are also *other possibilities*, like the eventual moving far away from the fixed point (or periodic motion) that has become unstable as it happens in cases of vague repulsivity: this case either give rise to a "*palingenesis*" because the phenomenology reappears anew because the motions near the unstable ones migrate elsewhere in phase space towards a stable fixed point or periodic orbit or invariant 2-dimensional torus, or it can be considered as a *fourth scenario* that we can call the *scenario of direct transition to chaos*, *i.e.* when the motions departing from the vicinity of the ones that have become unstable do not return intermittently close to them and are asymptotically governed by a strange attracting set.

It can also happen that a large number of eigenvalues of the stability matrix dwell near the unit circle (or the imaginary axis); a case in which the motion may "appear" chaotic because of the many time scales that regulate its development. One can hope, and this seems indeed the case, that this rarely happens because of the nongenericity of the crossing of the unit circle by more than one real or two conjugate eigenvalues (or of the imaginary axis by more than one real or two conjugate eigenvalues in the case of stability loss by a fixed point). In any event one can think that this is part of a *fifth scenario* or *scenario of the remaining cases*.

Among the "other cases" there is also the possibility of successive formation of quasi periodic motions with an increasing number of frequencies, *i.e.* there is the possibility of a scenario sometimes called the *scenario of Landau* (or of *Hopf*, who also proposed it, or *quasi periodicity scenario*). They are certainly mathematically possible and examples can be exhibited which, however, seem to have little relevance for fluidodynamics where this scenario has *never* been observed when involving quasi periodic motions with more than three frequencies; and the cases with three frequencies, or interpretable as such, observed are only two: one in a real experiment, *c.f.r.* [LFL83], and one in numerical simulations in a 7 modes truncation of the three dimensional Navier–Stokes equation, *c.f.r.* [GZ93].

Note also that, as discussed in (E) apropos of the Devil's staircases, already what happens in an ordinary Hopf bifurcation is in a way in "conflict" with the quasi periodicity scenario with increasing number of frequencies: in fact phase locking phenomena (with the consequent formation of stable periodic orbits) are generated rather than quasi periodic motions with monotonically varying rotation number.[7]

At this point it must also be said, as suggested by Arnold in support of Landau's scenario, that although it is true that quasi periodic motions cannot generically follow losses of stability by two frequencies motions on 2–dimensional tori, it remains that there are generic sets with *extremely small volume*[8] while there are nongeneric sets with *extremely large volume*.[9]

Examples of the latter phenomenon are, indeed, quasi periodic motions that arise in the theory of conservative systems: they are motions that are important, for instance, for the applications to celestial mechanics. They take place on sets with large positive volume in the space of the control parameters, and are therefore even easy to observe. Not more difficult than

[7] It is unreasonable to think that Landau ignored the phase locking phenomenon in this context: probably he thought that it had negligible consequences in the development of turbulence and that "things" wernt as if it was unexistent. The phase locking intervals would become shorter and shorter as the forcing strength increased. This is a very fruitful way of thinking common in Physics, think for instance (see [Ga99b]) to the geocentric hypothesis of ancient astronomy or to the ergodic hypothesis, see Ch. VII for other examples.

[8] Think of the union of intervals of size $\varepsilon\, 2^{-n}$ centered around the n–th rational point.

[9] Think to the set of points ω in $[0,1]$ such that $|\omega q-p|>Cq^{-2}$ for a suitable C and for all pairs (p,q) with $q>0$: these are the "*diophantine*" numbers with exponent 2, which form a set of measure 1 but with a dense complement, *c.f.r.* problems in §5.1.

the phenomena that take place in correspondence of the complementary values of the control parameters (which may be even stable, *i.e.* take place in open sets).

Hence there is the possibility that by simply changing the notion of small or large, understanding for "large" what has a large volume (rather than what is "generic"), that the quasi periodicity scenario appears as acceptable. Thus, mathematically speaking, the critique that one can really move to the scenario of Landau is that we do not observe it or that it is *extremely rare*.

As r increases further all the possibilities seen above will "overlap" as they can appear in various regions of phase space and thus coexist, making it very difficult even a phenomenology of developed turbulence.

But for what concerns the onset of turbulence the first four scenarios described above provide a rather general list of possibilities and rarely transitions to chaos have been observed that significantly deviate from one of them (I only know of the two quoted above, [LFL83], [GZ93].

Finally one should note that although the growth of r usually favors development of instabilities, it is not necessary that bifurcations should always happen in the direction in which we have discussed them: sometimes as r increases the same phenomena are observed but in reverse order. For instance it is possible, and it is observed, that periodic orbits, instead of doubling, halve in a "reverse" Feigenbaum sequence reaching a final state consisting in a stable periodic orbit which, as r increases further, might become again unstable following any of the possible scenarios. Furthermore what happens in certain phase space regions is, or can be, largely independent of what happens in others.

When for the same value of r of the driving there are several attracting sets (with different attracting basins, of course) one says that the system presents a phenomenon of *hysteresis* also called of *non–uniqueness*, see [Co65] p. 416 and [FSG79] p. 107.

Several direct "transitions to chaos", (fourth scenario), *can be instead regarded as normal inverse evolutions according to one of the first three scenarios.* Indeed what at times happens is that in a "direct transition" to chaos motions get away towards an attracting set which, at the considered value of r, is chaotic but which is the result of an independent evolution (following one of the first three scenarios) of another attractive set located elsewhere in phase space. In conclusion we can say that the scenarios alternative to the four described ones have never been observed with certainty in models of physical interest (although they are mathematically possible, as simple examples show).

Therefore we shall call "normal" an evolution towards chaotic motions that follows one of the four scenarios described above: "anomalous" will be other possibilities including the *aristotelic scenario* in which quasi periodic motions with more and more frequencies appear without ever giving rise to a truly strange motion, the scenario often attributed to Landau who in this case is meant by some to play the unpleasant role of modern Ptolemy, (which to Landau should not sound bad as the critiques to Ptolemy are often not well founded either and in any event hardly diminish his scientific stature, [Ga99b].

(J) The experiments.

I shall not present here how the analysis of the preceding sections really looks from an experimental point of view: *i.e.* which are the actual observations and the related techniques and methodolgy. The reader will find a simple and informative expository introduction in [SG78].

Problems.

[4.3.1]: Consider the spectrum of the stability matrix M of the Poincaré map associated with a periodic orbit Γ. Check that it does not depend on the point O nor on the section plane Σ chosen to define it. (*Idea:* Check that changes of O, or of Σ, change M by a similarity transformation; *i.e.* there is a matrix J such that the new stability matrix M' is $M' = JMJ^{-1}$.)

[4.3.2]: (*Hopf bifucation for maps*) Proceeding in analogy to the analysis of §4.2, formulate the Hopf bifurcation theory for periodic orbits. (*Idea:* The role of the imaginary axis is now taken by the unit circle. The theory is essentially identical with the novelty of the possible stability loss through $\lambda = -1$, which gives rise to the period doubling bifurcations. The normal forms approach, *c.f.r.* problem [4.2.4] is also very similar. If one formulates and accepts a theorem analogous to the center manifold theorem, *c.f.r.* problem [4.2.7], the analogue of the Hopf bifurcation theorem is discussed in the same way taking into account, when necessity arises, the further condition mentioned above (*i.e.* $\lambda_i^a \neq 1$ for $a = 3, 4$) that there should not be eigenvalues of the stability matrix at $r = r_c$ which are third or fourth roots of unity, *c.f.r.* [Ga83]).

[4.3.3]: (*existence of the rotation number for circle maps*) Let $\alpha \to g(\alpha)$ be a C^∞ increasing function defined for $\alpha \in (-\infty, \infty)$ and such that:

$$g(\alpha + 2\pi) = g(\alpha) + 2\pi$$

Note that the map of the circle $[0, 2\pi]$ defined by $\alpha \to g(\alpha) \bmod 2\pi$ is also of class C^∞: show that the limit $\lim_n \frac{1}{n} g^n(\alpha)$, if existing, is independent of α. Show then that the limit exists (*Poincaré's theorem*).
(*Idea:* Monotony of g implies that $g^n(0) \leq g^n(\alpha) \leq g^n(2\pi) = g^{n-1}(g(2\pi)) = g^{n-1}(g(0) + 2\pi) = \ldots = g^n(0) + 2\pi$: this gives α–independence. Likewise $g^n(2\pi\, k) = g^n(0) + 2\pi\, k$: next note that

$$g^{n+m}(0) \leq g^n(0) + g^m(0) + 4\pi$$

because setting $g^m(0) = 2\pi k_m + \delta_m$, $0 \leq \delta_m < 2\pi$ it is $g^{n+m}(0) \leq g^n(2\pi k_m + \delta_m) \leq g^n(2\pi(k_m + 1)) = g^n(0) + 2\pi(k_m + 1) \leq g^n(0) + g^m(0) + 4\pi$. Hence calling $L = \liminf_{m\to\infty} m^{-1} g^m(0)$ and if m_0 is such that $m_0^{-1} g^{m_0}(0) < L + \varepsilon$ with ε positive, we shall write $n = k\, m_0 + r_0$ with $0 \leq r_0 < m_0$. Then $g^n(0) \leq k g^{m_0}(0) + 4\pi(k + 2)$ and dividing by n both sides and taking the limits as $n \to \infty$ we get $\limsup n^{-1} g^n(0) \leq m_0^{-1} g^{m_0}(0) + 8\pi/m_0 \leq L + \varepsilon + 8\pi/m_0$ and by the arbitrariness of ε and the possibility of choosing m_0 as large as wanted we get $\limsup n^{-1} g^n(0) \leq \liminf m^{-1} g^m(0)$, *i.e.* the limit esists. See §5.4 for an alternative proof. See problem [5.1.28], [5.1.29], below for a constructive proof and an algorithm to construct the rotation number and verify its continuity as a function of the map g.)

[4.3.4]: By using only a computer analyze the bifurcations of (4.1.30) and, proceeding by experimentation, check the existence of period doubling bifurcations. Try to find an interval of r, where pairs of stable and unstable periodic motions annihilate as r decreases, via a bifurcation through 0, *c.f.r.* [FT79]. (The results are described also in [Ga83], Chap. V, §8).

[4.3.5]: Study empirically, with a computer, the structure of the bifurcations of the time independent solutions and of the (possible) periodic orbits of (4.1.28), *c.f.r.* [Fr83]. (The results are described also in [Ga83], Chap. V, §8).

[4.3.6]: Build an example of a function on $[0,1]$ whose graph is a "devil's staircase", *i.e.* which is a nondecreasing function such that every rational value is taken in an interval in which the function is constant and, furthermore, which is continuous and strictly increasing in all points in which it has an irrational value. (*Idea:* Consider the n–th rational point x_n, $n=1,2,\ldots$ (enumerating rationals arbitrarily). Let $f(x)=x_1$ for x in the "first triadic interval" I_1 (*i.e.* in the open interval containing the numbers whose first digit in base 3 is not 1: $(\frac{1}{3},\frac{2}{3})$).[10] Consider x_2 and if $x_2<x_1$ let $f(x)=x_2$ in the first triadic interval to the left of I_1, *i.e.* $(1/9,2/9)$, or if $x_2>x_1$ in the first triadic interval to the right of I_1, *i.e.* $(7/9,8/9)$. Proceed iteratively and let $f(x)$ be the function that we reach in this way: extend it by continuity to the other points (that form the Cantor set) obtaining a function that has as graph the stair of the wicked Being.)

[4.3.7]: (*a resonance*) Conjecture that every C^∞ map, S, of R^n in itself that has the origin as a hyperbolic fixed point (*i.e.* with a stability matrix without eigenvalues λ_i of modulo 1) can be transformed by a coordinate change of class C^2 at least into a map S' which is *exactly linear* in the vicinity of the origin. Show that the conjecture is false by showing that the map

$$x_1'=\lambda^2x_1+N_{123}x_2x_3,\quad x_2'=\lambda^{-1}x_2,\quad x_3'=\lambda^3x_3$$

with $\lambda>1$ gives a counterexample because it cannot be transformed into $\xi_1'=\lambda^2\xi_1$, $\xi_2'=\lambda^{-1}\xi_2$, $\xi_3'=\lambda^3\xi_3$ with a C^2 change of coordinates. (*Idea:* Show that $\xi_j'=x_j+\sum_{i,k}m_{j,ik}x_ix_k+o(x^2)$ gives rise to incompatible conditions for the coefficients $m_{j,ik}$ if $N_{123}\ne0$: because they should verify $(\lambda_j-\lambda_i\lambda_k)m_{ijk}=N_{ijk}$.)

[4.3.8]: The result in problem [4.3.7] is not incompatible with the *Grobman–Hartman theorem* which, instead, states that if S is of class C^1 at least then it can be transformed in the vicinity of the origin into a linear map via a coordinate change locally invertible and *continuous*, *but not necessarily differentiable*. One can show that if S is of class C^∞ then there exists, near a fixed point (the origin to fix the ideas), a C^1 coordinate change that linearizes locally the map if $\lambda_j\ne\lambda_i\lambda_k$ for all values of i,j,k. One says in this case that S verifies a "nonresonance condition" of order 2 at the fixed point. If, furthermore, $\lambda_i\ne\lambda_1^{k_1}\ldots\lambda_n^{k_n}$ for all n–ples of nonnegative integers $k_1,\ldots,k_n$ with sum ≥2, one says that S verifies a "nonresonance condition" to all orders at the fixed point: then the coordinate change can be chosen to be of class C^k, for all $k>0$, close to the fixed point (how close may depend on the value prefixed for k), *c.f.r.* [Ru89b], moreover it depends continuously, in class C^1, from any parameters on which S possibly depends, provided S is C^∞ in these parameters). Analogous results hold for the exact linearization of differential equations near a fixed point where the stability matrix has eigenvalues λ_j with $\lambda_j\ne\sum_i k_i\lambda_i$ for $\underline{k}$ as above, *c.f.r.* [Ru89b], p. 25.

[4.3.9]: (*stable and unstable manifolds: normal form in absence of resonances*) Assume valid the *nonresonance* property to arbitrary order and prove that the results quoted in problem [4.3.8] can be used to show existence, regularity and regular dependence (on the (possible) parameters on which S depends) of two manifolds $W_O^{\delta,s}$ e $W_O^{\delta,u}$ contained in a vicinity of O, of small enough radius δ, such that $S^nx\xrightarrow[n\pm\infty]{}O$ if $x\in W_O^{\delta,s}$ or, respectively, $x\in W_O^{\delta,u}$. (*Idea:* This is obvious if S is linear, and it remains true also in the nonresonant nonlinear case, by using the coordinate change quoted in problem [4.3.8]. This is a little involved proof of the existence of the stable and unstable manifolds of an hyperbolic fixed point. Existence can be proved with more elementary means and under more general conditions, in which the equation is *not even linearizable*: the manifolds

[10] In general given an interval I it can be divided in three thirds, I_0,I_1,I_2: called here the 0-th, the 1-st and the 2–d (*i.e.* we use the C-language conventions).

always exist for any hyperbolic fixed point and have *the same regularity* of S provided the latter is at least of class C^1, *c.f.r.* [Ru89b], p.28.)

[4.3.10]: Applying an implicit function theorem prove that even if the terms $O(z^4)$ in eq. (4.3.1) are not neglected there is a curve γ invariant for S_Σ and close to the circle in eq. (4.3.2). (*Idea:* Let $\varepsilon = -2\,\mathrm{Re}\,c(r)$ and let the parametric equations of the curve γ be $z = \rho(1+\rho\,\xi(\alpha))e^{i\alpha}$ with $\alpha \in [0,2\pi]$ with ρ as in eq. (4.3.2). Then the equation for γ becomes $K\xi(\alpha) = \xi(\alpha)$ with

$$\begin{aligned} K\xi(\alpha') =& (1-\varepsilon)\xi(\alpha) + \rho\xi(\alpha)^2\,\overline{\Lambda} + \rho^2\Lambda' \\ \alpha' =& \alpha + \delta + \rho^2\,\mathrm{Im}\,c(r) + \rho^2\xi(\alpha)\tilde{\Lambda} + \rho^4\Lambda \end{aligned}$$

where $\Lambda, \Lambda', \overline{\Lambda}, \tilde{\Lambda}$ are smooth functions of $\xi\alpha$. Then for $r - r_c > 0$ small (*i.e.* for ρ small) one finds, if $||\xi||_{C^1} \le 1$,

$$||K\xi_1 - K\xi_2||_{C^1} \le (1-\varepsilon/2)||\xi_1 - \xi_2||_{C^1}$$

so that the solution can be found simply as $\lim_{n\to\infty} K^n\xi_0$ with $\xi_0 = 0$, for instance.)

Bibliography: [RT71], [Fe78], [Fe80], [Ec81], [CE80], [ER81], [Ru89], [GZ93], [LL71], [SG78] and principally [Ru89b]. The scenario of Landau in its original formulation can be found in the first editions of the treatise [LL71]: in the more recent ones the text of the latter book has been modified and updated to take into account the novelties generated by [RT71].

§4.4: Dynamical tables.

In this section we analyze properties of the states that are reached at large time starting from randomly chosen initial data. Such states depend on the value of a parameter, r in the examples of §4.1, that we conventionally call the *Reynolds number* and which measures the intensity of the driving force applied to the system to keep it in motion.

This parameter will be considered, in order to offer a more cogent visual image, as a "time": so that we shall say that "as r increases" a bifurcation takes place from a certain stationary state (not necessarily time independent, nor quasi periodic but just statistically stationary) which leads to the "creation" of new stable stationary states. Or that "as r increases", somewhere in phase space happens a "creation" of a pair of fixed points or of periodic orbits one stable and one unstable, or an *"annihilation"* of such a pair happens, *etc.*

The following view of the onset of turbulence emerges from the discussion of §4.2, §4.3. The initial laminar motions (which in phase space is a fixed point, *i.e.* a "time independent motion") loses stability through successive bifurcations from which new laminar motions are born; they are less symmetric, at least if the system possesses some symmetries, as the examples in §4.1, see (4.1.21), (4.1.29), (4.1.31). The new motions attract, in turn, random initial data (by random we mean chosen with a probability distribution that has a density with respect to the volume in phase space).

This may go on until (and if) the laminar motion disappears because of a "collision" with a time independent unstable motion giving rise directly or intermittently to a chaotic motion. Or it becomes unstable via a Hopf bifurcation generating a stable periodic motion. In the latter case the periodic stable motion may lose stability as r increases becoming in turn unstable. This can happen in various ways

(1) by a "collision" with an unstable periodic orbit with a possible change of the asymptotic motion to a chaotic one following an intermittent or a direct scenario, or
(2) giving raise to a quasi periodic motion on a 2–dimensional torus following the Ruelle–Takens scenario, or
(3) giving rise to an asymptotic motion which is chaotic after several period doubling bifurcations following the Feigenbaum scenario.

In case (2) the torus will evolve possibly losing stability and generating strange attractors; likewise in case (3) the periodic motions generated by the period doubling bifurcations will continue to double in an ever faster succession (as r grows) eventually generating a chaotic attracting set.

It is also possible that, at certain values of r, stable periodic orbits or stable fixed points (or other types of attracting sets) appear somewhere else in phase space in regions which seemed to have no other distinguishing property and which dispute the role of attracting set to attracting sets existent elsewhere in phase space: this is the *hysteresis* phenomenon, *c.f.r.* (I) in §4.3.

The "dispute" takes place because the attraction basins of the attracting sets will in general have common boundaries which often show rather high geometrical complexity; so that it can be difficult to decide to which basin of attraction a particular initial datum belongs, and which will be the long time behavior of a motion following the given initial datum.

For instance a pair of fixed points or of periodic motions, one stable and one unstable, can be born at a certain value of r, in some region of phase space and evolve in a way similar to the evolution of the "principal series" of attracting sets, which we define loosely as

Definition *(principal series): we shall call "series" a family of invariant sets in phase space parameterized by the forcing strength r and varying continuosly with r (several invariant sets may correspond to each r because of possible bifurcations). The family of fixed points, periodic orbits, invariant tori, other invariant sets that can be continuosly traced back to an attracting set that exists at $r = 0$ will be called a "principal series" of bifurcations, the others (when existent) "secondary series".*

Remark: Usually there will be only one such family because for $r = 0$ there will be a single globally attracting (time independent) point, that we called *laminar motion.*

As soon as there exist two or more distinct attracting sets, phase space divides into regions attracted by different attracting sets; and naturally new phenomena are possible if one looks at data starting on the boundary that separates the two basins of attraction. Such phenomena are "difficult to see" because it is impossible to get initial data that are exactly on the boundary of the basins, unless the system is endowed with some special symmetry that imposes a simple and precise form for the boundary between two basins.

The order in which different types of attracting sets appear in a bifurcation series, principal or secondary, is not necessarily an ordered sequence of fixed points first, then periodic orbits, then invariant tori *etc*: some steps may be missing or the order can be changed, even inverted.

(A) Bifurcations and their graphical representation.

It is important to find a method to represent the above phenomenology to understand by inspection the situation in a given particular case.

A convenient method consists in using the following "*dynamical tables*".

Consider the family of the stable fixed points and attracting sets that can be continuously followed as r grows starting from the fundamental laminar motion of the system, identified with a laminar motion existing for $r = 0$, that we suppose for simplicity existent unique and globally attractive for $r = 0$ and for r very small, *i.e.* we consider the "principal series" of bifurcations according to the definition above.

We shall represent in a plane the axis r as abscissae axis: a point on a line parallel to the axis r will symbolize the laminar motion corresponding to the parameter r.

The stability loss at $r = r_c$ will be marked by a point I, *c.f.r.* Fig. (4.4.1). The segment OI will be denoted F_0 (with F standing for "fixed").

If in r_c there is a bifurcation with symmetry loss we draw two new lines departing from I, *i.e.* as many as the newly created fixed points, drawing as dashed the lines representing the unstable fixed points and as continuous lines the stable ones. The vertical axis does not have *meaning* other than a symbolic one and the plane is only used to have space to draw a picture

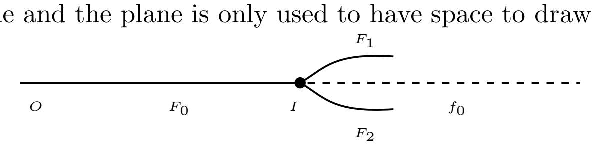


Fig. (4.4.1) *The segment OI represents a family of fixed points which bifurcate at I with a symmetry breaking bifurcation into two steble fixed points and an unstable one.*

Hence from a figure in which all continuous families of fixed points have been represented as lines parallel to the x axis we see immediately how many fixed points coexist in correspondence of a given r and from which "history" they arise as r grows: it is enough to draw a line perpendicular to the axis and count the intersections with the graph.

To continue in a systematic way it is convenient to establish more general rules for the drawings that we shall make. We shall use the following symbols

Fig. (4.4.2) *The drawings represent a family (F) of fixed points continuously evolving as r varies in* (r_1, r_2)*, or a family of periodic orbits (O), or a family of stable invariant tori (T), or of a strange attracting sets (S), respectively.*

to denote that in correspondence of the interval (r_1, r_2) there is a stable fixed point continuously depending on r in the case F, or a stable periodic orbit in the case O, or an invariant torus run quasi periodically in the case T, or of strange attracting set in the case S.

The same symbols, in *dashed form*, will be used to denote the corresponding unstable entities. For brevity we shall call "*eigenvalues*" of a fixed point or of a periodic orbit the eigenvalues of its stability matrix.

A bifurcation will be denoted by a black disk, that will carry a label to distinguish its *type*: for instance

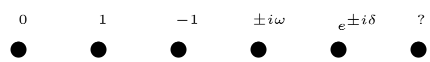


Fig. (4.4.3) *Symbols for types of bifurcations.*

will denote respectively the following cases

(1) Case denoted 0: a stability loss of a fixed point due to the passage through 0 of a real eigenvalue (with loss of some symmetry and persistence of the fixed point as an unstable point, or with a creation of a pair of fixed points (or annihilation: with ensuing direct or intermittent chaos).

(2) Case denoted 1: a stability loss by a periodic orbit due to the passage through 1 of a real eigenvalue with consequent breaking of some symmetry or a creation (annihilation) of a pair of periodic orbits (one stable and one unstable). Hence on the black disk of Fig. (4.4.1) we should put the label 0 over the label I.

(3) Case denoted -1: a stability loss by a periodic orbit due to the passage through -1 of a real eigenvalue. In this case a consequent appearance of a stable periodic orbit of almost double period and persistence of the preceding orbit as an unstable one is possible; but sometimes the orbit is not vaguely attractive and direct or intermittent chaotic transitions appear.

(4) Case denoted $\pm i\omega$: a stability loss by a fixed point due to the crossing of the imaginary axis by a pair of conjugate eigenvalues with either the consequent appearance of a periodic stable orbit and permanence of the fixed point as an unstable one, or with the disappearance of a stable periodic orbit and appearance of chaotic motion via a direct or intermittent transition.

(5) Case denoted $e^{\pm i\delta}$: a stability loss by a periodic orbit due to the passage through the unit circle of a pair of conjugate eigenvalues; also in this case the loss of stabiloity could result in a (direct or intermittent) chaotic transition or in the appearance of an invariant torus (stable or unstable) and persistence of the periodic orbit as an unstable orbit.

(6) There are, sometimes, bifurcations whose type is difficult to analyze or to understand or other: this case will be denoted by affixing on the bullet a label ? .

An *arrow pointer* will indicate that the history of an orbit or of an attracting set proceeds, as the parameter r varies, in the direction marked by the arrow and that it can possibly undergo further bifurcations *that are not indicated*.

(B) An example: the dynamical table for the model NS_5.

The discussion of the sections §4.2, §4.3 combined with the graphic conventions just established allows us to summarize the theoretical results on the equation (4.1.30), which are described in the problems of §4.1, §4.3 and the experimental ones obtained mainly in [FT79], via a *dynamical table* that provides us immediately with a "global view" of the phenomenology.

We recall that the NS_5, *i.e.* (4.1.30), possess a symmetry: they are invariant under a simple symmetry group with four elements, see §4.1.

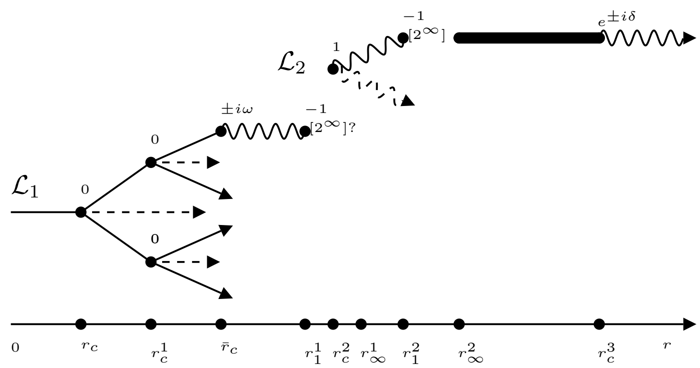


Fig. (4.4.4) *A rather detailed dynamical table representing the phenomenology observed in the model NS_5. The r_c, r_c^1, $\bar r_c$ are symmetry breaking bifurcations. Note the hysteresis in a small interval to the right of $r_c^2 < r_\infty^1$.*

The above Fig. (4.4.4) illustrates, in fact, the real dynamical table of the equation (4.1.30), computed in [FT79] (*c.f.r.* also the problems of §4.1, §4.3). The points r_c, r_c^1 marked on the r axis are the threshold values of the successive symmetry breaking bifurcations. The first marks the breaking of the symmetry (4.1.31) in the sense that the two new stable fixed points generated by the bifurcation are mapped into each other by the transformations with $\varepsilon = \pm 1$ but are still invariant under the subgroup with $\eta = \pm 1$ and $\varepsilon = 1$; at r_c^1 the symmetry is completely broken and the 4 resulting stable fixed points are distinct and mapped into each other by the group of symmetry.

In $\bar r_c$ *each* of the four stable fixed points generated by the preceding bifurcations loses stability undergoing a Hopf bifurcation, and each generates a stable periodic orbit (hence four symmetric orbits are generated): one of them is indicated by the wavy line (the others are related to it by the symmetries of the model and are hinted by the arrow pointers).

This orbit loses stability by period doubling in r_1^1 and a sequence of infinitely many period doubling bifurcations follows, indicated by the symbol $[2^\infty]$. One observes, in fact, *only three* such period doubling bifurcations because at the value $r_c^2 < r_\infty^1 = \{$ value where the sequence of period doubling bifurcations "should" accumulate (according to Feigenbaum's theory) $\}$ a hysteresis phenomenon develops (*c.f.r.* (I) in §4.3).

Namely, in *another region* of phase space (not far, but clearly distinct, from the one where the orbits, that we followed so far, underwent period doubling bifurcations), a pair of periodic orbits (one stable and one unstable) are "created".

The attraction basin of the new stable orbit rapidly extends, as r grows within a very small interval. While r is in this interval a *hysteresis phenomenon occurs* and some initial data are attracted by the new stable orbit and others are attracted by the doubling periodic orbits. Soon the attraction basin of the new orbit appears to swallow the region where the doubling periodic orbits should be located (hence they either disappeared or possess too small a basin of attraction to be observable in the simulations). In this way the "main series" of bifurcations denoted $\mathcal{L}_1$, terminates (without glory) its history.

Note that a system, like the one being considered, which has various symmetries necessarily presents hysteresis phenomena when the attracting sets do not have full symmetry (*i.e.* when symmetry breaking bifurcations have occurred). This is therefore the case for $r > \overline{r}_c$: the dynamic table in reality consists, for $r > \overline{r}_c$, of four equal tables, only one of which is described in Fig. (4.4.4) (the others would emerge, by symmetry,from the other "arrow pointers" associated with the other fixed points). Between these 4 *coexisting* attracting sets, images of each other under the symmetry group, hysteresis phenomena naturally take place. But the hysteresis phenomenon between the lines $\mathcal{L}_1$ and $\mathcal{L}_2$ in Fig. (4.4.4) has a *different nature*, *i.e.* it is not trivially due to a coexistence of symmetric attracting sets, because no element of the first series is a symmetric image of one of the second.

In Fig. (4.4.4) we have not marked the unstable periodic orbits that remain, as a "wake", after each series of period doubling bifurcations: such unstable orbits can, in fact, be followed in the numerical simulations at least for a while (as r increases). Drawing them would make the figure too involved; but we can imagine that the symbol $[2^\infty]$ includes them. The interrogation mark that follows $[2^\infty]$ means that in reality one observes only few period doubling bifurcations because, as discussed above, at a certain point between r_c^2 and the value r_∞^1 (extrapolated from the three or four really observed values of the period doubling bifurcations, on the basis of Feigenbaum's universality theory) a "collision" seems[1] to take place between the stable orbits produced by the period doubling bifurcations and the attraction basin of the stable periodic orbits of the *new series* $\mathcal{L}_2$, with the consequence that the family $\mathcal{L}_1$ seems to end.

[1] Keep in mind that we are discussing experimental result

The new series $\mathcal{L}_2$, as described, is born (as r increases) as a stable periodic orbit which at birth is marginally stable with a stability matrix with a single eigenvalue 1 corresponding to a creation of a pair of periodic orbits, one stable and one unstable: following the stable orbit one observes (experimentally) that it loses stability for some $r = r_1^2$, by a period doubling to which an infinite sequence of period doubling bifurcations follow, in the sense that one is able to observe many of them and the only obstacle to observe more seems to be the numerical precision needed.

At the end of the evolution of this family which is at a a value r_∞^2 where the successive thresholds r_j^2 accumulate as $j \to \infty$ (at a fast rate numerically compatible with Feigenbaum's constant, 4.68...), one observes the birth of a strange attracting set, which is however followed at larger r by a periodic stable orbit, see Fig. (4.4.4).

The latter, if followed *backwards decreasing* r, loses stability at $r = r_c^3$ because of the passage of two complex eigenvalues though the unit circle *without* being followed by the birth of an invariant torus of dimension 2 (because one observes instead a chaotic motion): this means that the bifurcation (at decreasing r) is an "inverse" bifurcation. And, indeed, at $r = r_c^3$ the stable periodic orbit that exists for $r > r_c^3$ has, in fact, been observed to collide with another unstable periodic orbit which also exists for $r > r_c^3$ (at least for $r - r_c^3$ small) and the two annihilate leaving (for $r < r_c^3$) a strange attracting set with chaotic motion on it. The unstable periodic orbit has not beeen represented in the picture.

The set of lines connected with the one representing the laminar motion at small Reynolds number is the *main series* of bifurcations and the others are the *secondary series* (only one of them in the case of Fig. (4.4.4), if one does not count the other 3 existing because of the symmetry (4.1.29)).

The just described phenomenology is, however, as it often happens with experiments (real or numerical) valid only in a first approximation. Refining the measurements one can indeed expect to be able to see further details: for instance the interval $[r_\infty^2, r_c^3]$, which appears at first to correspond to motions regulated by "just" 4 symmetric strange attracting sets as shown in Fig. (4.4.4), reveals itself endowed with much higher complexity.

For instance one can "resolve", via experimentally more delicate and difficult observations and measurements, *c.f.r.* [Fr83], some features of the motions that take place for r in this interval as shown in Fig. (4.4.5) where the "*window*" (describing the birth of the periodic orbit and the successive Feigenbaum bifurcations) for $r > r_F$ has a really small width of the order of 10^{-2}! (not on scale in Fig. (4.4.5)) and terminates in the point r_F^∞ where the period doubling bifurcations accumulate: to it again a chaotic motion follows.

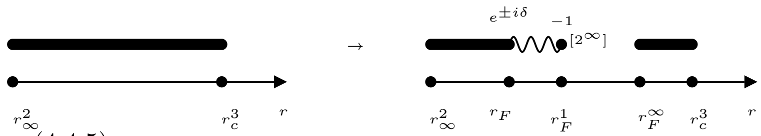


Fig. (4.4.5): *Resolution of an interval of values of r where the motion appears, at first, always chaotic. At a more accurate analysis the interval* $[\overline{r}^2_\infty, r^3_c]$ *splits into regions where it is still chaotic and into tiny regions (not on scale in the drawing) where stable periodic orbits exist and evolve through a rapid history of period doubling bifurcations.*

(C) The Lorenz model table.

The dynamical table for the Lorenz equation, built on the basis of the results discussed in the problems of §4.1, is simpler and shows how the laminar motion evolves, as r increases, by first losing symmetry at $r = 1$ (see (4.1.21)) into two new stable fixed points mapped into each other by the 2–elements symmetry group. Each of them loses stability, for some $r = r_c$, with two conjugate eigenvalues crossing the imaginary axis. However the bifurcation is not "direct", *i.e.* there will *not* be generation of a stable periodic orbit from each of the two fixed points, (*c.f.r.* Fig. (4.4.6)) at $r = r_c$; rather the points will lose stability because they "collide" with an unstable periodic orbit (represented by the dashed wavy line in the figure). After the stability loss the fixed points remain unstable and, as Fig. (4.4.6) indicates, no periodic stable motion is generated but we have directly a chaotic motion (this is a direct transition to chaos, according to the fourth scenario in §4.3).

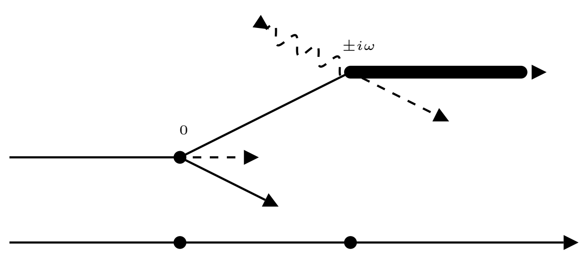


Fig. (4.4.6) *Dynamical table for the Lorenz model representing a symmetry breaking followed by a collision, between the resulting stable time independent point and an unstable periodic orbit, which engenders chaos.*

A more attentive analysis shows that also in this model there are intervals of r in (r_c, ∞) in which there is an attracting set consisting of a periodic (stable) motion which as r varies evolves with a sequence of period doubling bifurcations as in the case illustrated in Fig. (4.4.5), *c.f.r.* [Fr80].

(D) Remarks.

From bifurcation theory we have seen that certain bifurcations are not generic: hence usually we must imagine, when trying to build a dynamical table for a particular system, only tables in which generic bifurcations appear. This is for instance the case of the previously examined realistic dynamic tables.

In this way we can only build a relatively small number of combinations at each bifurcation. Examples of nongeneric (hence forbidden in the tables) bifurcations can be taken from the analysis of §4.2 and, obviously, there are many more, among which the bifurcation in the Fig. (4.4.1) in *absence* of an accompanying breaking of some symmetry. Some examples are provided in Fig. (4.4.7).

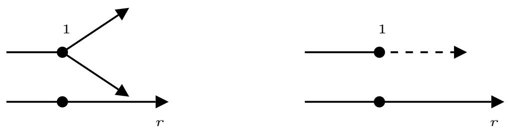


Fig. (4.4.7) *Forbidden non generic bifurcations.*

In principle we should expect that any dynamical table which only contains generic bifurcations could be realized by the phenomenology of some model: and in fact their variety is very large. One can simply look at [Fr83], [Fr80], [FT79] to be convinced. Examining the dynamical tables presented in such papers one sees by inspection the involved phenomenology of the sequences of bifurcations, with various hysteresis phenomena (*i.e.* coexistence of various attracting sets in correspondence to a fixed value of r, as in the case of Fig (4.4.4) in the interval between r_c^2 and r_∞^1).

In several cases *resolutions* of "apparently simple" attracting sets have been observed which, as in the example of Fig. (4.4.5), reveal a far more complex structure than that of Fig (4.4.4). But the greater complexity manifests itself over much smaller intervals of r and it can be easily "missed" if the observations are not very precise or not systematic enough.

Dynamical tables provide us with a method to rapidly visualize and organize *without boring and monotonous descriptions* the general aspects of the phenomenology of a given experiment.[2]

One should refrain from thinking that compiling dynamical tables is simple *routine*: they usually require and describe accurate and patient analysis. Particularly when hysteresis phenomena are relevant and make it difficult to follow a given attracting set as the Reynolds number varies; or when one wishes to follow the history of an attracting set that after a bifurcation survives but becomes unstable. The latter feat, understandably quite difficult, is at times necessary because the same attractive set may lose stability in correspondence of a certain Reynolds number to reacquire it again in correspondence of another value (typically "colliding" with another attracting set and "ceding" to it the instability while "taking" the stability, if usage of a pictorial language is allowed here).

[2] We shall no longer distinguish, unless necessary, between numerical experiments and experiments performed on real fluids: because we take for granted that both are "real" and with equal "dignity".

(E) Another example: the dynamic table of the NS_7 model.

The NS_7 model of §4.1 has been studied in detail in [FT79] where measurements sufficient to establish the following dynamical table have been performed.

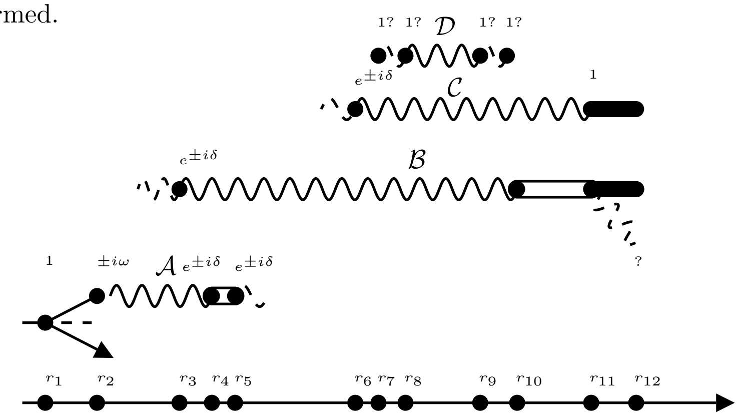


Fig. (4.4.8) *A NS_7 equations dynamical table. For $r < r_1$, "last bifurcation breaking the symmetry" in (4.1.29), the series is the same as that preceding r_c^1 in Fig. (4.4.4). Note the hysteresis phenomena between $\mathcal{B},\mathcal{C},\mathcal{D}$ or $\mathcal{A},\mathcal{B}$ or $\mathcal{B},\mathcal{C}$.*

The strange attracting sets that are born at large r are not different in the sense that *apparently* the bifurcation series $\mathcal{C}$ and $\mathcal{B}$ both end in the same attracting set for $r > r_{11}$ (even though in the table it appears drawn as different) modulo the symmetries of the model.

We see in this table several hysteresis phenomena. In this case too we did not represent the trivial hysteresis: one should always take into account that if a differential equation has certain symmetries then the attracting sets either are symmetric or the sets that are obtained by applying the symmetry transformations to one of them generate others (that are "copies") with the same properties. This gives rise to hysteresis phenomena that are "trivial" (in the sense of the comments to Fig. (4.4.4)) and are not marked because we have not drawn in the table the attracting or repelling sets that can be obtained from the drawn ones by applying the symmetries of the model, *c.f.r.* (4.1.29) (with the exception of the case of the bifurcation at r_1).

Not all bifurcations in the table of Fig. (4.4.8) are clearly understood: in particular those of the line $\mathcal{D}$. The bifurcations on which hangs the type label "?" have not been studied (because there remains always something to do, usually for good reasons).

Since the details that can be seen in dynamic tables depend on the precision with which measurements are performed (or, better, can be performed) it is expected that by performing more accurate measurements one can find some change of the tables in the sense that more details may appear, as already seen in the case of the NS_5 model.

(F) Table of the 2–dimensional Navier–Stokes equation at small Reynolds number. Navier–Stokes)

Finally we conclude by considering the dynamic table of the 2–dimensional Navier–Stokes equation with a large regularization parameter, *c.f.r.* Fig. (4.4.9). This delicate study indicates that the phenomenology that occurs at small values of the Reynolds number *appears to stabilize* as a function of the number of Fourier modes used in the regularization.

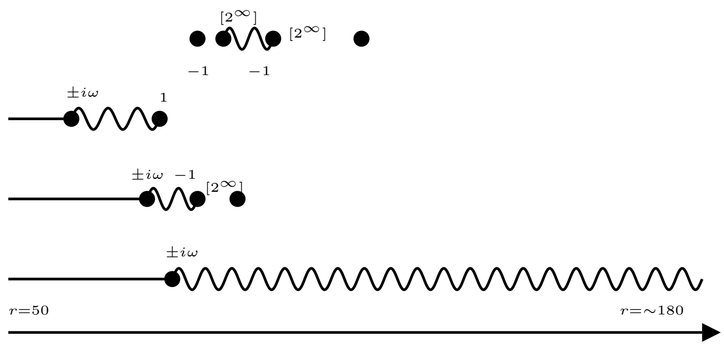


Fig. (4.4.9) *The stabilized phenomenology of the NS equations up to Reynolds number* ~ 200 *and with truncations at order* $M = 26$, 37, 50, 64 *(results, however, do not depend on* M*). Note the hysteresis phenomena.*

Up to Reynolds numbers of the order of 200 it seems that at least the laminar motions become independent from the value of the ultraviolet cut–off M that defines the regularization by selecting the Fourier modes such that $|\underline{k}| = \max(|k_1|, |k_2|) \le M$, *c.f.r.* §3.1, §3.2.

We consider the dynamical tables of these regularizations of the equations, that we shall denote as $NS^{(M)}$, with $M = 26, 37, 50,\ 64$, taken from [FGN88], drawn for r between $r = 50$ to about $r = 180$. The force acts only on one mode (and its opposite): $\underline{k} = (2, -1)$ and it is real. We only consider solutions in which the various Fourier components $\gamma_{\underline{k}}$ are either real or purely imaginary, extending the remark seen in §4.1, *c.f.r.* (4.1.27).

Note that there is a very good reason for the choice of the mode on which the forcing acts: the simplest choice, *i.e.* the mode with minimum length $\underline{k}_0 = (0, \pm 1), (\pm 1, 0)$, is not sufficient to create interesting phenomenology: see problem [4.1.13]. However any other choice of the mode seems sufficient to generate interesting phenomenologies. How much the results depend on which mode is actually chosen for the forcing has not been studied: however it seems that the choice does not affect "substantially" the phenomenology and it would be interesting to clarify this point. Also the restriction that the components $\gamma_{\underline{k}}$ are real or imaginary (rather than complex) is an important issue partially studied in the literature and deserving further investigation.

The tables coincide from $r = 0$ up to about $r = 180$, hence we draw only the table relative to $M = 64$ (which is a model with 5200 equations or so). The above Fig. (4.4.9) illustrates an interval of r going from 50 to ~ 180.

For larger values important variations of the form of the table as function of the cut–off M are still observed, particularly for $r > 200$. In all cases the motions eventually generate a quasi periodic bidimensional motion; however this happens at *cut–off dependent* values of r, between ~ 250 and ~ 600.

By the discussion in §4.3, (F), the "quasi periodic" motion should appear in a sequence of phase locked periodic motions: in the experiments the motion always appears quasi periodic which means that the steps of the "devil staircase" formed by the graph of the rotation number as a function of r are so short that they are not observable within the precision of the experiments.

This shows how delicate can be the process of (believed) "convergence" as $M \to \infty$ of the attracting sets, and of the corresponding phenomenologies, of the equations $NS^{(M)}$ to those of the full Navier–Stokes equation. And strictly speaking we only know that such convergence really takes place through (scanty) empirical evidence: the theory developed in Chap. 3, of the 2–dimensional Navier–Stokes equations (with periodic boundary conditions), although rather complete from the viewpoints of existence, uniqueness and regularity, *does not allow us to obtain asymptotic results as* $t \to \infty$ and it is therefore of little use for the questions which interested us here.

Problems.

[4.4.1]: On the basis of the phenomenology described in [FZ85], [FZ92] draw (from their original data) the dynamic tables describing the asymptotic motions of the systems considered.

[4.4.2]: Draw the dynamic tables describing the asymptotic motions studied in [Ri82] (from his original data).

Bibliography: The experimental results of simulations, exposed above, are taken mainly from the papers: [FT79], [FGN88], [Fr80], [Fr83], [Ri82], [Ga83]. The dynamical tables have been introduced and widely used in these references. For experiments on real fluids see [FSG79], [SG78].

CHAPTER V

Ordering chaos

§5.1 Quantitative description of chaotic motions before developed turbulence. Continuous spectrum.

After the discussions of the previous chapters it becomes imperative to find quantitative methods of study, or even simply of description, of the various phenomena that one expects to observe in experiments on fluids.

This section, as well as the others in this chapter, will be devoted to this kind of questions.

Everywhere in chapter 5, unless explicitly stated, M will be supposed to be at least a closed bounded set in a euclidean space. We begin by setting the following formal definition

1 Definition *(smooth flows, continuous dynamical systems):*
A "continuous dynamical system" (or "flow") is defined by giving a phase space M, that we suppose to be a regular bounded surface in R^n, and by a group of regular (C^∞) transformations S_t, with $t \ge 0$ or $t \in R$ and $S_t S_{t'} = S_{t+t'}$, generated by the solutions of a differential equation on M: the dynamical system is denoted (M, S_t). We shall call observable *any regular ($C^\infty(M)$) function $O : M \to R$. Occasionally we shall also consider functions O that are only piecewise regular (C^∞) calling them* piecewise regular observables.[1]

Let $\underline{u} \in M$ be a point in phase space and let $t \to S_t\underline{u} = \underline{u}(t)$ be a motion that develops on an attracting set, or that does so asymptotically. Let $O(\underline{u})$ be an observable, considered as the time t varies, *i.e.* consider the function

[1] Regular surfaces are *always assumed connected*. In some cases M will be a surface in a infinite dimensional space, like $C^\infty(\Omega)$. If M has infinite dimension the notion of observable must be made precise on a case by case basis. For instance if M is the space of the C^∞ divergenceless vector fields on a domain $\Omega \subset R^2$ or $\Omega \subset R^3$ we shall require the observables to be functions $O(\underline{u})$ that can be expressed as polynomials in the values of $\underline{u}$ and of its derivatives, evaluated in a finite number of points, or their integrals over Ω with regular weight functions.

$t \to F(t) = O(\underline{u}(t))$: we shall call it the *history* of O on the considered motion.

For instance if $\underline{u}(\underline{x})$ is the velocity field of a fluid described by a finite truncation of the NS equations a typical observable will be the first component of the velocity at a given point $\underline{x}_0$ of the container: $O(\underline{u}) = u_1(\underline{x}_0)$; and the *history* of this observable on a given motion of the fluid is the function $t \to F_O(t) = O(\underline{u}(\underline{x}_0, t)) = u_1(\underline{x}_0, t)$.

A first simple qualitative property that it is interesting to associate with a motion is the *power spectrum* of an observable O. It can be defined as the function $p \to A(p)$:

$$\begin{aligned} A(p) = \lim_{T\to\infty} A_T(p) = \lim_{T\to\infty} \frac{1}{2T} \Big| \int_0^T e^{-ipt} F_O(t)\, dt \Big|^2 = \\ = \lim_{T\to\infty} \frac{1}{2T} \int_0^T dt \int_0^T d\tau\, F_O(t) F_O(\tau) \cos p(t-\tau) = \\ = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^T d\vartheta \int_{(-\vartheta)\vee 0}^{T-(\vartheta\vee 0)} d\tau\, F_O(\tau+\vartheta) F_O(\tau)\, \cos p\vartheta \end{aligned} \tag{5.1.1}$$

where A and A_T are here implicitly defined and $a \vee b = \max(a,b)$, $a \wedge b = \min(a,b)$. If the limit

$$\Omega(\vartheta) = \lim_{T\to\infty} \frac{1}{T} \int_{(-\vartheta)\vee 0}^{T-(\vartheta\vee 0)} dt\, F_O(t+\vartheta) F_O(t) \stackrel{def}{=} \langle F_O(\cdot+\vartheta) F_O(\cdot) \rangle \tag{5.1.2}$$

(which defines the right hand side of (5.1.2)) existed, and *if we were allowed* to permute the limit over T and the integral over ϑ in (5.1.1), then we would have

$$A(p) = \frac{1}{2} \int_{-\infty}^{\infty} d\vartheta\, \Omega(\vartheta) \cos p\vartheta \tag{5.1.3}$$

However the exchange of these limits is always difficult to discuss (even though "essentially always possible") and to avoid posing too early subtle mathematical questions the following definition will be adopted

2 Definition *(correlation functions):*
If the average (5.1.2) exists the power spectrum *of the observable O on the motion $t \to u(t)$, is the Fourier transform, (5.1.3), of the "correlation function" $\Omega(\vartheta) = \langle F_O(\cdot+\vartheta) F_O(\cdot) \rangle$ of the history $F_{O,u}(t) = O(u(t))$.*

To understand the interest of the notion it is good to study first the power spectra of observables in regular, *i.e.* quasi periodic, motions.

(A) Diophantine quasi periodic spectra.

A quasi periodic motion, as already discussed in the previous chapters, develops over a regular ℓ–dimensional torus in phase space, *i.e.* on a surface with parametric equations $\underline{\varphi} \to u = U(\underline{\varphi})$ with parameters $\underline{\varphi} = (\varphi_1, \ldots, \varphi_l)$ given by an ℓ–ple of angles varying on a standard ℓ–dimensional torus (*i.e.* on $[0, 2\pi]^\ell$ with opposite sides identified), and U is a C^∞ periodic function.

Moreover the motion is a uniform rotation of the ℓ angles: $\underline{\varphi} \to \underline{\varphi} + \underline{\omega} t$ where $\underline{\omega} = (\omega_1, \ldots, \omega_\ell) \in R^\ell$ is a "rotation vector" with *rationally independent* components, in the sense that if: $\sum_i \omega_i n_i = 0$, then $\underline{n} \equiv \underline{0}$.

Hence *all observables* O defined on M will have a quasi periodic history F, *i.e.* a history that can be represented as $t \to F(t) = \Phi(\underline{\omega} t)$ with $\Phi(\underline{\varphi})$ a regular periodic function of the ℓ angles $\underline{\varphi}$.

Note that if $\ell = 1$ the motion is in fact periodic.

From the theory of Fourier transforms we immediately deduce the existence of coefficients $\Phi_{\underline{n}}$ such that

$$F(t) = \sum_{\underline{n}} \Phi_{\underline{n}}\, e^{i\underline{\omega}\cdot\underline{n}t} \tag{5.1.4}$$

where $\underline{n}$ is an integer components vector and the coefficients $\Phi_{\underline{n}}$ have a rapid decrease as $\underline{n} \to \infty$.[2]

Suppose that the vector $\underline{\omega}$ is *Diophantine*, *i.e.* that there exist two constants C, τ such that

$$|\underline{\omega} \cdot \underline{n}|^{-1} < C\, |\underline{n}|^\tau \qquad \text{for each } \underline{0} \neq \underline{n} \in Z^\ell \tag{5.1.5}$$

where Z^ℓ denotes the integer components vectors in R^ℓ. It is known that all vectors $\underline{\omega} \in R^\ell$ are Diophantine *except a set of zero volume*, *c.f.r.* problem [5.1.4].

Computing the integral (5.1.1) in the case of (5.1.4) we get

$$\begin{aligned}
A(p) &= \lim_{T\to\infty} A_T(p) = \\
&= \lim_{T\to\infty} \frac{1}{2} \sum_{\underline{n},\underline{n}'\in Z^\ell} \Phi_{\underline{n}} \overline{\Phi}_{\underline{n}'}\, \tfrac{1}{T} \tfrac{e^{i(\underline{\omega}\cdot\underline{n}-p)T}-1}{i(\underline{\omega}\cdot\underline{n}-p)} \tfrac{e^{-i(\underline{\omega}\cdot\underline{n}'-p)T}-1}{-i(\underline{\omega}\cdot\underline{n}'-p)} = \\
&= \lim_{T\to\infty} 2 \sum_{\underline{n},\underline{n}'\in Z^\ell} \Phi_{\underline{n}} \overline{\Phi}_{\underline{n}'}\, e^{i\underline{\omega}\cdot(\underline{n}-\underline{n}')\frac{T}{2}}\, \tfrac{1}{T} \tfrac{\sin(\underline{\omega}\cdot\underline{n}-p)\frac{T}{2}}{i(\underline{\omega}\cdot\underline{n}-p)} \tfrac{\sin(\underline{\omega}\cdot\underline{n}'-p)\frac{T}{2}}{-i(\underline{\omega}\cdot\underline{n}'-p)} = \\
&= \pi \sum_{\underline{n}} |\Phi_{\underline{n}}|^2\, \delta(p - \underline{\omega}\cdot\underline{n})
\end{aligned} \tag{5.1.6}$$

where δ is Dirac's delta function, and the limits are intended in the sense of distributions, *c.f.r.* [5.1.5]. This essentially follows from the well known

[2] The torus has regular equations, by assumption: hence the function $\Phi(\underline{\varphi})$ is as regular as the torus and the Fourier coefficients of Φ decay very rapidly as $\underline{n} \to \infty$. If the torus is C^k–regular and if F is C^k–regular then $\Phi_{\underline{n}}$ decays at least as $|\underline{n}|^{-k}$.

relation: $\frac{1}{T}\big(\frac{\sin xT/2}{x}\big)^2 \xrightarrow[T\to\infty]{} \frac{\pi}{2}\delta(x)$.[3] Hence

Proposition I: *(nature of quasi periodic spectra) Systems with attracting sets E, which attract exponentially points close to them*[4] *and that are ℓ dimensional tori on which motions are quasi periodic, yield power spectra that are sums of Dirac's deltas centered on the "frequencies" $p = \underline{\omega}\cdot\underline{n}/2\pi$ where $\underline{n}\in Z^\ell$ and $\underline{\omega}$ is the "rotation spectrum" of the quasi periodic motion (or $\underline{\omega}/2\pi$ is its "frequency spectrum").*

Remarks:
(1) In practice, in experimental observations, we can only observe the quantity $A_T(p)$ with $T<\infty$ and, in this case, the functions δ are "rounded" and the graph of the power spectrum appears as a sequence of "peaks" around the values $p=\underline{\omega}\cdot\underline{n}$.
(2) We must also remark that if the $\underline{\omega}$'s are rationally independent (as we suppose with no loss of generality, *c.f.r.* [5.1.1]), and if $\ell>1$, then the values taken by $\underline{\omega}\cdot\underline{n}$ as $\underline{n}$ varies in Z^ℓ densely fill the line p. But the amplitude $|\Phi_{\underline{n}}|^2$ tends to zero very rapidly as $\underline{n}\to\infty$, hence the height of the various peaks is very unequal: for all but a finite number of them *it is not observable* if measurements are performed with a prefixed limited precision. Hence what one really observes in experiments is only a (small) number of peaks which in the graphs of the power spectrum as function of p emerge over a "background noise" representing errors and other fluctuations.
(3) The case of a quasi periodic motion with only one frequency ($\ell=1$), *i.e.* a periodic motion, is obviously special because in this case the peaks are isolated from each other and equispaced on the p axis, although still (and for the same reasons) only a finite number of them will be really distinguishable from the background noise when measurements are performed with a limited precision.
(4) The case of three or more independent frequencies must be considered "rare" when motion is generated by a differential equation describing a fluid, because it would be a non generic behavior: *c.f.r.* §4.2, §4.3 but always keep in mind the remark, in item (I) of §4.3, that a non generic phenomenon can nevertheless take place on sets of parameters, in the space of parameters that control the equations, that may be in some sense very large sets.

---

[3] A relation known from the elementary theory of Fourier series, *c.f.r.* [Ka76] p.12 (*"Fejér's theorem"). It is obtained by remarking that $\delta_N(x) = (2\pi)^{-1}\int_{-N}^{N} e^{ikx}dk$ tends to $\delta(x)$ for $N\to\infty$; hence also its average $T^{-1}\int_0^T dN\,\delta_N(x)$ tends to $\delta(x)$ for $T\to\infty$ and $N\to\infty$.*

[4] *i.e.* such that data initially close enough to E get closer to E and their distance to E tends to zero exponentially fast in time.

(B) Continuous spectrum.

Analysis of the power spectra can be an effective method to detect quasi periodicity of a motion. In the fluidodynamic applications, however, we must expect that the power spectrum will usually reveal one or two fundamental frequencies at most, as seen when discussing the scenarios for the onset of chaos. When the complexity of the motion increases the nature of the spectrum will change and this makes the power spectra effective indicators of the development of motions of other types.

Recalling that we have called "observables" the regular functions on phase space, that we now suppose to be a regular surface $M \subset R^n$, it is convenient to set the following definition

3 Definition *(motions with continuous spectrum):*
(1) Given a motion $t \to x(t)$ suppose that there is an observable whose power spectrum $A(p)$ is, for p in an interval $[p_1, p_2], p_1 < p_2$, a nonzero function which is continuous, or at least integrable. We then say that the motion has a "spectrum with a continuous component", or that the motion has "infinitely many" time scales.
(2) Continuity of the power spectrum may depend on the particular observable considered: a motion is "chaotic" if there exists at least one observable whose spectrum on the selected motion has a continuous component.
If there is a class $\mathcal{F}$ of observables which is dense in $L_2(M)$ (where L_2 is intended with respect to the surface measure on M) and all observables in $\mathcal{F}$ are such that their spectrum $A(p)$, c.f.r. (5.1.3), is a function in L_1 for $|p| > 0$ the motion $t \to x(t)$ is said to have "continuous spectrum" or that it is "completely chaotic".
(3) A system has continuous spectrum with respect to a random choice of initial data with distribution μ if with μ–probability 1 any initial datum generates a motion over which all observables of a family $\mathcal{F}$ dense in $L_2(M)$ have spectrum in L_1 for $|p| > 0$. In such a system, when the initial data are chosen randomly with distribution μ, all observables give rise to time histories which are chaotic.

Remarks:
(i) In item (2) of the definition it is necessary to exclude explicitly $p = 0$ because the trivial observations (*i.e.* the constants) have a power spectrum proportional to the Dirac's delta $\delta(p)$; hence if $\Omega(\vartheta) \xrightarrow[\vartheta\to\infty]{} \Omega(\infty)$ we shall find that $A_T(p)$ is the sum of $\pi\,\Omega(\infty)\,\delta(p)$ and of a distribution that, in motions with continuous spectrum, vanishes at $p = 0$ while it is a summable (or continuous) function elsewhere.
(ii) If all observables in $\mathcal{F}$ admit a an average value on the considered motion, in the sense that the limit $\langle F\rangle = \lim_{T\to\infty} T^{-1}\int_0^T F(t), dt$ exists, and if for each F it is $\Omega(\infty) = \langle F\rangle^2$ (we shall see that this is the case for most motions at least in the case of "mixing systems", *c.f.r.* §5.4) we could equivalently say that the motion under study has continuous spectrum when all observables are such that the spectrum of $F - \langle F\rangle$ is locally in L_1 for all

$F \in \mathcal{F}$.

(C) An example: non Euclidean geometry.

The mathematically simplest example of a system with motions having a continuous spectrum is surprisingly complicated, although extremely interesting: it is provided by a light ray that moves in a closed region Δ of a half plane ($y > 0$ for instance) with refraction index $n(x, y) = \frac{\lambda}{y}$, where λ is a constant length scale set to 1 below.

On this half plane we imagine "*light rays*" that proceed at constant velocity $c = 1$ in the sense of "geometrical optics": *i.e.* are such that in time dt they go through a distance ds such that $n(x, y)\, ds = dt$ (*i.e.* the speed of light is $v = n^{-1}(x, y)$ at the point (x, y)).

We shall suppose that the rays propagate in accordance with Fermat's principle, *i.e.* a light ray that passes through two points P and Q goes, at velocity 1, through a curve γ that joins P with Q on which the "optical path", defined by the line integral $\int_\gamma n\, ds$ with ds being the element of ordinary length $ds = \sqrt{dx^2 + dy^2}$, is stationary.

The light rays can be thought of as vectors of unit length with respect to the metric $d\ell^2 = (dx^2 + dy^2)/y^2$ and their trajectories are identified with the geodesics of the "geometry" associated with the metric in question. The metric $y^{-1}ds$ is called the "*Lobachevsky metric*" or, in the above interpretation of light rays, the "*optical metric*" for the medium with refraction index $n(x, y) = y^{-1}$.

It can be shown, see problems, that given two points in the half plane $y > 0$ there is a *unique* light path that joins them and (hence) the light paths in this particular medium with refraction index y^{-1} minimize the optical path not only for close enough points P, Q but no matter at which distance they may be. Thinking of the optical paths as straight lines for the metric $n\, ds \equiv y^{-1}ds$ this is analogous to the property of the "straight lines" in planar Euclidean geometry. Given a complete optical path γ (*i.e.* a path extended to infinite length on both sides of its points) and a point ouside it there are infinitely many complete optical paths through this point that do not intersect γ: *i.e.* in this geometry there are *several parallels* to a given straight line.

It is well known, see problems, that the geodesics of the metric $y^{-1}\, ds$ are all semi circles with center on the line $y = 0$, including the straight lines $y > 0, x = cost$. In other words they are the semi circles orthogonal to the axis $y = 0$ (that in this metric is at infinity, *i.e.* it has infinite distance from an arbitrary point with $y > 0$).

The motion of the light rays can then be seen as a map S_t that acts on a light ray initially in (x, y) with velocity (of unit modulus in the optical metric) directed so that it forms an angle ϑ with the $y = 0$ axis ($\vartheta \in [0, 2\pi]$): see Fig. (5.1.2) below.

It transforms the above light ray, that we can denote (x, y, ϑ), into a ray

(x', y', ϑ') where
(a) (x', y') is a "suitable" point on the semi circle Γ centered on the axis $y = 0$ and radius such that it passes through (x, y) with tangent in the direction ϑ and
(b) ϑ' is the direction of the tangent to this semi circle in (x', y') giving the semicircle the same orientation as ϑ.
(c) The "suitable" point (x', y') is the point that is reached starting from (x, y) and moving along Γ following the orientation pointed by ϑ and going on a distance which, measured in the optical metric, is precisely t.

In other words the light rays motions are described by what in geometry can be called a "geodesic motion" (with respect to the optical metric)..

This dynamical system is, all things considered, quite simple and of moderate interest because it is clear that motions "go towards infinity", similarly to the behavior of the uniform rectilinear motions in Euclidean geometry.

However, as in planar Euclidean geometry, one can imagine of "confining" the motions to a finite region Δ (*i.e.* such that the largest distance between any two of its points, measured in the metric $ds^2 = y^{-2}(dx^2 + dy^2)$, is finite).

This can be made in a way similar to the one used to confine the uniform rectilinear motions in Euclidean geometry. Imagine imposing "periodic" conditions on the boundary of a region Δ by identifying points of the boundary of Δ that can be transformed into each other by means of suitable maps, in analogy with the translations by integer vectors of a unit square in R^2 which are used to identify the opposite sides in order to turn it into a torus, *i.e.* into a smooth bounded surface without any boundary.

In this way Δ can be thought of as a smooth surface without boundary: we shall see that an example is provided (in suitable coordinates, *c.f.r.* problems betwen [5.1.32] and [5.1.37]) by the figures (5.1.1), (5.1.2) below. The region Δ itself as well as the maps that identify opposite sides of Δ cannot, however, be arbitrary (note that they are not arbitrary already in the planar case).

If we want to obtain a surface which is really smooth and without boundary then it is almost evident that it is necessary that the maps that identify pairs of points on the boundary of Δ form a discrete subgroup of the group $\mathcal{G}$ of the maps of the half plane that conserve *both* the angles between directions coming out of the same points *and* the length of the optical path of the infinitesimal curves. This group $\mathcal{G}$ of maps exists [5] and the maps are called

[5] Such group is non trivial and consists of the maps that have the form $z' = \frac{az+c}{bz+d}$, if $z = x + iy$ and $z' = x' + iy'$ and $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is a real matrix with determinant 1, *c.f.r.* problem [5.1.30].
One can verify (*c.f.r.* [5.1.34] and following) that there exist closed regions Δ whose boundary points can be identified via the action of a suitable subgroup of $\mathcal{G}$ so that they generate regions analogous to the tori in the plane Euclidean geometry: such regions can be constructed so that their boundary consists in suitable arcs of circle, *c.f.r.* [5.1.34]%[5.1.37].

rigid movements: together with metric $ds^2 = y^{-2}(dx^2 + dy^2)$ they form what is called the *geometry of Lobatchevsky*.

The elements of the subgroup must *furthermore* have the property that an infinitesimal line element passing through a point $(x, y) \in \partial\Delta$, on the boundary of Δ and directed in a direction ϑ, "*exiting out of* Δ" is transformed into an infinitesimal line element through a point $(x', y') \in \partial\Delta$ and in a direction ϑ' "*entering* Δ". Only in this way, indeed, the dynamical system generated by S_t can be thought of as a geodesic motion on a closed surface without boundary obtained "by folding" Δ on itself.

Certainly one could even doubt that it is possible to construct domains Δ and subgroups of the group of rigid motions of the geometry of Lobatchevsky with the properties required above: after all in the case of planar geometry the conditions are so restrictive that they single out rectangles with periodic boundary conditions.

However it is not difficult to see that not only these conditions are not incompatible but they even lead to a very wide family of really different domains Δ. These are domains with a boundary that can be divided into arcs whose points can be identified through the action of movements of the geometry so that Δ becomes a smooth surface *without boundary*. The matter is discussed in the problems from [5.1.33] on, where examples are provided.

We can now formulate the statements about the continuous spectrum

Proposition II *(noneuclidean geometry and continuous spectrum):*
Consider one such domain Δ, see Fig. (5.1.2) below. Consider the geodesic evolution $t \to S_t(x, y, \vartheta)$ on the 3–dimensional space of the "light rays trapped within Δ by the boundary conditions". The motions thus constructed are the analogues for the geometry of Lobatchevsky of the quasi periodic motions of Euclidean geometry. They have the property of having continuous spectrum, with the exception of a set of initial data (x, y, ϑ) that have zero volume.

A proof that this system has actually continuous spectrum can be based on the theory of the representations of the group $SL(2, R)$, *c.f.r.* [GGP69], [CEG84]: more details can be found in the problems [5.1.30] and following.

The example is certainly not easy if one does not have a minimum familiarity with the non Euclidean geometry of Lobatchevsky: but it is otherwise elementary and intimately related to the efforts to understand the postulate of parallelism that consumed generations of scientists and, hence, it has a very particular flavor and interest.

(D) A further example: the billiard.

The relative complexity of the preceding example should not lead one to believe that motions with continuous spectrum are rarely met in physically

---

relevant situations. The simplest example, from the physics viewpoint, is in fact immediately formulated: but the mathematical theory and the proof of the continuity of the spectrum is far more difficult than the one of the preceding example (which rests on elementary properties of non Euclidean geometries and group theory).

The example is the *billiard*: the phase space M is the product of the square $Q = [0, L]^2$ deprived of a disk ("table with an obstacle") times the unit circle ("direction of the velocity of the billiard ball"). Hence a point of phase space is x, y, ϑ with $(x, y) \in Q$ and $\vartheta \in [0, 2\pi]$.

The dynamical system is (M, S_t) with the transformation S_t mappimg (x, y, ϑ) into (x', y', ϑ') defined by the position and velocity that the "ball" takes after time t if it moves as a free point mass starting from the position (x, y) with velocity in the direction ϑ and proceeds at unit speed on a straight line unless a collision occurs within time t; if, within time t, collisions occur between the particle and the obstacle the ball is reflected by the obstacle, elastically, reaching in time t the final position at (x', y') with velocity in the direction ϑ'.

The system so described has singular points corresponding to the points in phase space representing collisions with the obstacles (where the velocity can be discontinuous), *c.f.r.* [5.4.20], [5.4.21] of §5.4. Technically (M, S_t) defines a dynamical system more general than the ones considered so far which have always been assumed to be smooth. We shall briefly formalize the notion of non smooth dynamical system in §5.4: however the notion of continuous spectrum extends unaltered to such systems.

The main result is

Proposition III *(billiards and continuous spectrum (Sinai)):*
In this system almost all initial data, with respect to the Liouville measure $\mu(dx\,dy\,d\vartheta) = dx\,dy\,d\vartheta$, *generate motions with continuous spectrum.*

This is the content of a theorem by Sinai. The formulation of the example is therefore very easy: unlike the proof of the statements in proposition III, [Si70], [Si79], [Si94], [Ga75].

Problems: Ergodic theory of motions on surfaces of constant non positive curvature.

[5.1.1]: Let u a C^∞ function on the torus T^ℓ and let $\underline{\omega} \in R^\ell$ be a vector with rationally dependent components. Let $u_{\underline{n}}$ be the Fourier transform of u. Check that $u(\underline{\omega} t)$ can be written as $v(\underline{\omega}' t)$ with $v \in C^\infty(T^{\ell'})$, on a torus $T^{\ell'}$ with lower dimension ($\ell' < \ell$), and find an expression for the Fourier coefficients $\Phi_{\underline{n}}$ in (5.1.4) for the $F(t) = v(\underline{\omega}' t)$ in terms of the Fourier transform of u. (*Idea:* Let $\underline{\omega}'_0 = (\omega'_1, \ldots, \omega'_{\ell'})$ be a maximal subset of rationally independent components of $\underline{\omega}$. Then there is an integer N such that $\omega_j = \frac{1}{N} \underline{n}'_j \cdot \underline{\omega}'_0$ with $\underline{n}'_j$, $j = 1, \ldots, \ell$, suitable integer components vectors. Let $\underline{\omega}' = \frac{1}{N} \underline{\omega}'_0$ and write the Fourier series for $u(\underline{\varphi})$ replacing $\underline{\varphi}$ with $\underline{\omega} t$ and then expressing $\underline{\omega}$ as $\omega_j = M_{ji} \omega'_i$, $M_{ji} = (\underline{n}'_j)_i$).

[5.1.2]: Consider the interval $[0, 1]$ and let $r_1, r_2, \ldots$ be the rational numbers between 0 and 1 arbitrarily numbered. Let O_i be the open interval centered at r_i with length $\varepsilon 2^{-i}$, with $\varepsilon > 0$ prefixed. Check that $\cup_i O_i$ is open, dense, but it has measure not larger than ε. Consider only the rationals in $[0,1]$ that have finitely many non zero digits in their binary representation and numbering them $r_1, r_2, \ldots$ arbitrarily consider the intervals of length $\varepsilon = 2^{-k-j}$ centered around r_j. This is an open dense set: find a rule to build the binary expansion for a point that is not in this set.

[5.1.3]: Show that the function $A_T(p)$ in (5.1.6) tends to 0 for almost all p. (*Idea:* Consider p's in an interval $I = [-a, a]$. Then the intervals of the variable p for which $|p - \underline{\omega}\cdot\underline{n}| < 1/D|\underline{n}|^{\ell+1}$, for any nonzero integer components vector $\underline{n}$, have total length $\le \frac{c}{D}$ for c suitable (*e.g.* one can take $c = \sum_{|\underline{n}|>0} |\underline{n}|^{-\ell-1}$). If p is outside all such intervals the sum over $\underline{n}$ in (5.1.6) is bounded by: $2T^{-1}c^2D^{-2}\sum_{\underline{n},\underline{n}'} |\Phi_{\underline{n}}|\,|\Phi_{\underline{n}'}|\,|\underline{n}|^{\ell+1}|\underline{n}'|^{\ell+1}$; hence it tends to zero as $T \to \infty$).

[5.1.4]: Consider the unit ball S_1 in R^ℓ. Show that, fixed $\varepsilon > 0$, the set $S_1(\underline{n})$ of the $\underline{\omega} \in S_1$ such that $|\underline{\omega}\cdot\underline{n}| < C^{-1}|\underline{n}|^{-\ell-\varepsilon+1}$ has volume $\le C^{-1}B_\varepsilon$ for a suitable constant B_ε. Deduce from this that the volume of the points $\underline{\omega} \in S_1$ satisfying a Diophantine property with constants $C < \infty$ and $\alpha = \ell + \varepsilon$ has complement with *vanishing* volume. (*Idea:* Note that the volume of the points in the ball $S_1(\underline{n})$ such that $|\underline{\omega}\cdot\underline{n}| < C^{-1}|\underline{n}|^{-\ell-\varepsilon+1}$ is $\le \Omega_\ell\, C^{-1}|\underline{n}|^{-\ell-\varepsilon}$ if Ω_ℓ is twice the volume of the unit ball in $\ell - 1$ dimensions, (4π if $\ell = 3$, 4 if $\ell = 2$).)

[5.1.5]: Let $\underline{\omega}$ be a "Diophantine" vector, *c.f.r.* (5.1.5). Consider the function like $A(p)$ in (5.1.6) but with the summations over $\underline{n}, \underline{n}'$ constrained by $\underline{n} \ne \underline{n}'$: show that it tends to zero in the sense of distributions. (*Idea:* Bound $I = \frac{1}{t}\int_a^b \frac{\sin(p-x)t/2}{p-x}\frac{\sin pt/2}{p}dp$ can be bounded by using $|\sin x| \le B_\varepsilon |x|^\varepsilon$ for a suitable constant $B_\varepsilon > 0$ and for every x and every $\varepsilon \in (0,1]$. This implies that

$$I \le B_\varepsilon^2 \frac{1}{(|x|t)^{1-2\varepsilon}} \int_{-\infty}^{\infty} \frac{dz}{(|1-z|\,|z|)^{1-\varepsilon}}$$

Hence, choosing $\varepsilon < \frac{1}{2}$, one sees that the generic term of the series in (5.1.6) with $\underline{n} \ne \underline{n}'$, *multiplied* by a rapidly decreasing C^∞ function $f(p)$ and *integrated* over p, can be bounded by the above inequality (with $x = \underline{\omega}\cdot(\underline{n} - \underline{n}')$) by $T^{-(1-2\varepsilon)}\,||f||_{C^1}\, I\, C|\underline{n} - \underline{n}'|^{(1-2\varepsilon)\gamma}|\Phi_{\underline{n}}|\,|\Phi_{\underline{n}'}|$ hence it tends to zero because the $\Phi_{\underline{n}}$ tend to zero faster than any power, *c.f.r.* (5.1.4)).

- *Ergodic theory of quasi periodic motion*, *i.e.* of geodesic motion on a torus.

[5.1.6]: *(ergodicity of irrational rotations)* Let r be irrational and let $S\,x = x + r \bmod 1$ be a map of $M = [0,1]$ into itself. Show that for all C^∞ periodic functions f on M and for all $x \in M$ the following limit holds

$$\lim_{N\to\infty} \mathcal{M}_N(f)(x) = \lim_{N\to\infty} N^{-1}\sum_{j=0}^{N-1} f(x + j\,r) = \int_0^1 f(y)\,dy \stackrel{def}{=} \overline{f}$$

Infer that, therefore, if f is just summable, *i.e.* $f \in L_1(M, dx)$ the same limit relation holds.
(*Idea:* One notes that $\xi(\alpha) \stackrel{def}{=} N^{-1}\sum_{j=0}^{N-1} e^{i\,\alpha\,j} = N^{-1}(e^{i\,\alpha\,N} - 1)/(e^{i\,\alpha} - 1)$ if $\alpha \ne 2\pi\,k$ and $\xi(\alpha) = 1$ otherwise, and $|\xi(\alpha)| \le 1$ in any case. If $\hat f_\nu$ is the Fourier transform of f one notes that

$$\mathcal{M}_N(f)(x) = \sum_{\nu=-\infty}^{\infty} \hat f_\nu\, e^{2\pi\,i\,\nu\,x} N^{-1}\,(e^{2\pi\,i\,\nu\,N\,r} - 1)/(e^{2\pi\,i\,\nu\,r} - 1)$$

and the ν–th term is bounded by $|\hat f_\nu|$ and at the same time tends to 0 as $N\to\infty$ if $\nu\ne 0$ is kept fixed: hence the limit of $\mathcal{M}_N(f)$ is $\hat f_0$.
Setting $||f||_{L_1}\stackrel{def}{=}\int_0^1|f(x)|$ we get $||\mathcal{M}_N(f)||_{L_1}\le||f||_{L_1}$ and we reca;ll that the C^∞–periodic functions are dense on L_1 so that if $||\mathcal{M}_{N_j}(f)-\overline{f}||_{L_1}\ge\varepsilon>0$ for a sequence $N_j\to\infty$ we could approximate f by a C^∞ periodic funcion φ so that $||f-\varphi||_{L_1}<\varepsilon/3$. Then the relation $\mathcal{M}_{N_j}(f)-\overline{f}||_{L_1}\le||\mathcal{M}_{N_j}(\varphi)-\overline{\varphi}||_{L_1}+2\varepsilon/3$ would yield a contradiction.)

The role of the following classical problems is to make clear that one thing is to say and prove that a system (M,S,μ) is ergodic and a completely different and deeper matter is to understand the asymptotic behavior of motions generated by the evolution.

[5.1.7]: An irrational number r can be uniquely represented by its *continued fraction*, *i.e.* as the limit for $k\to\infty$ of

$$R_k=a_0+\cfrac{1}{a_1+\frac{1}{a_2+\frac{1}{\dots+\frac{1}{a_k}}}}\stackrel{def}{=}(a_0,a_1,\dots,a_k)$$

where $a_j\ge 1$ are positive integers.
Check that if $(a_1,\dots,a_k)=\frac{p'}{q'}$ then $(a_0,a_1,\dots,a_k)=(a_0p'+q')/p'$ and infer that if $\underline{v}_k=(p_k,q_k)\in Z^2_+$ is such that $R_k=\frac{p_k}{q_k}$ then a possibility for $\underline{v}_k$ is

$$\underline{v}_k=\begin{pmatrix}a_0&1\\1&0\end{pmatrix}\begin{pmatrix}a_1&1\\1&0\end{pmatrix}\dots\begin{pmatrix}a_k&1\\1&0\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}$$

The pairs thus constructed are called the *convergents* of the continued fraction of r. (*Idea:* If $[x]$ denotes the integer part of x then $a_0=[r]$, $a_1=[(r-a_0)^{-1}]$ *etc*).

[5.1.8]: From [5.1.7] deduce that $\underline{v}_k=a_k\underline{v}_{k-1}+\underline{v}_{k-2}$, *i.e.*

$$\begin{aligned}p_k=&a_kp_{k-1}+p_{k-2}\qquad k>1\\ q_k=&a_kq_{k-1}+q_{k-2}\qquad k>1\end{aligned}$$

and check that q_k,p_k increase with k and $q_k\ge 2^{(k-1)/2}$ for $k\ge0$ and $p_k\ge 2^{(k-1)/2}$ for $k\ge1$. Or, better, if $c\equiv\min_{i\ge1}a_i\ge1$ and $c_k=\max_{1\le i\le k}a_i\ge1$ then $(1+c)^{(k-1)/2}\le p_k,q_k\le(1+c_k)^{(k-1)/2}$. (*Idea:* $\begin{pmatrix}1&1\\1&0\end{pmatrix}^k\begin{pmatrix}1\\0\end{pmatrix}$ and $\begin{pmatrix}a&1\\1&0\end{pmatrix}^k\begin{pmatrix}1\\0\end{pmatrix}$ provide us with two pairs which bound below and above the pair $\begin{pmatrix}p_k\\q_k\end{pmatrix}$).

[5.1.9]: Check that the recurrence relation in [5.1.8] implies

$$\begin{aligned}&q_kp_{k-1}-p_kq_{k-1}=-(q_{k-1}p_{k-2}-p_{k-1}q_{k-2}=(-1)^k),\qquad k\ge2\\ &q_kp_{k-2}-p_kq_{k-2}=a_k(q_{k-1}p_{k-2}-p_{k-1}q_{k-2})=(-1)^{k-1}a_k\qquad k\ge2\end{aligned}$$

hence

$$\frac{p_{k-1}}{q_{k-1}}-\frac{p_k}{q_k}=\frac{(-1)^k}{q_kq_{k-1}},\qquad\frac{p_{k-2}}{q_{k-2}}-\frac{p_k}{q_k}=\frac{(-1)^{k-1}}{q_kq_{k-2}}a_k$$

[5.1.10]: The statement in [5.1.9] implies

$$\frac{p_0}{q_0}<\frac{p_2}{q_2}\dots<r<\dots\frac{p_3}{q_3}<\frac{p_1}{q_1},\qquad\Big|r-\frac{p_k}{q_k}\Big|<\frac{1}{q_kq_{k+1}}$$

Check that:

$$\frac{1}{q_k(q_k+q_{k+1})}<\Big|r-\frac{p_k}{q_k}\Big|<\frac{1}{q_kq_{k+1}}$$

(*Idea:* If $a,b,c,d>0$ and $\frac{a}{b}<\frac{c}{d}$ then $\frac{a+sc}{b+sd}$ increases with s for $s\ge0$, while if $\frac{a}{b}>\frac{c}{d}$ then it decreases. Hence if k is even

$$\frac{p_{k-2}+sp_{k-1}}{q_{k-2}+q_{k-1}}$$

increases with s and for $s=a_k$ it becomes $\frac{p_k}{q_k}$ which is $<r<\frac{p_{k-1}}{q_{k-1}}$ hence

$$\frac{p_k-2}{q_{k-2}}<\frac{p_{k-2}+p_{k-1}}{q_{k-2}+q_{k-1}}<r$$

from which

$$\Big|r-\frac{p_{k-2}}{q_{k-2}}\Big|>\Big|\frac{p_{k-2}+p_{k-1}}{q_{k-2}+q_{k-1}}-\frac{p_{k-2}}{q_{k-2}}\Big|\equiv\frac{1}{q_k(q_{k-2}+q_{k-1})}$$

and analogously for k).

[5.1.11]: Check that the convergents of a continued fraction of an irrational number r, p_n,q_n, are relatively prime for every n. (*Idea:* True for p_0,q_0. Assuming it true for p_k,q_k, $k=0,1,2,\ldots,n-1$ and for every irrational r, note that if p',q' are convergents of the continued fraction $[a_1,a_2,\ldots,a_n]$ then they are relatively prime. Hence if j divided q_n and p_n then it would divide p',q'! Alternatively check that $q_kp_{k-1}-p_kq_{k-1}=(-1)^k$, see [5.1.9], implies directly what requested).

Definition: A rational number p/q is an optimal approximation for the irrational r if for every pair p',q' with $q'<q$ it is $|q'r-p'|>|qr-p|$.

[5.1.12]: Let p,q positive integers and r irrational. Let j be odd, $\alpha=p/q$, $\alpha_j=p_j/q_j$, and suppose that $\alpha_{j-1}>\alpha>\alpha_{j+1}$; then $q>q_j$. (*Idea:* $\alpha_{j-1}>\alpha>\alpha_{j+1}>r>\alpha_j$ hence $(q_jq_{j-1})^{-1}>|\alpha_{j-1}-r|>|\alpha_{j-1}-\alpha|=|p_{j-1}q-q_{j-1}p|/qq_{j-1}\ge1/qq_{j-1}$, because $|p_{j-1}q-q_{j-1}p|\ge1$. Formulate and check the analogous statement for j even showing that the two results can be summarized by saying that if p/q is between two convergents of orders $j-1$ and $j+1$ then $q>q_j$).

[5.1.13]: In the context of problem [5.1.12] show that if the rational α is not a convergent and $\alpha_{j-1}<\alpha<\alpha_{j+1}$ then $|r-\alpha_j|<|r-\alpha|$; a similar result holds for j even. (*Idea:* : $q|\alpha-r|>q|\alpha-\alpha_{j+1}|=q|pq_{j+1}-qp_{j+1}|/qq_{j+1}\ge1/q_{j+1}\ge q_j|\alpha_j-r|$).

[5.1.14]: Show that [5.1.11], [5.1.12], [5.1.13] imply that if p/q is an approximation to an irrational r such that $|q'r-p'|>|qr-p|$ for all values of $q'<q$ then $q=q_j$, $p=p_j$ for some j. In other words every optimal approximant is a convergent.

[5.1.15]: Show that if r is irrational every convergent is an optimal approximant. (*Idea:* Otherwise for some n numbers p and $q<q_n$ would exist with $|rq-p|<|rq_n-p_n|=\varepsilon_n$; let $\bar p,\bar q$ be the pair minimizing the expression $|q'r-p'|$ with $q'<q_n$; if $\bar\varepsilon$ is the value of the minimum it is: $\bar\varepsilon<\varepsilon_n$; hence $\bar p/\bar q$ is an optimal approximation: therefore $\bar p=p_s,\bar q=q_s$ for some $s<n$ and $1/(q_s+q_{(s+1)})\le|q_sr-p_s|\le|q_nr-p_n|<1/q_{n+1}$, *i.e.* $q_s+q_{s+1}>q_{n+1}$ contradicting $q_{n+1}=a_{n+1}q_n+q_{n-1}$).

[5.1.16]: A necessary and sufficient condition in order that a rational approximation to an irrational number be an optimal approximation is that it is a convergent of the continued fraction of r. (*Idea:* Summary of the preceding problems).

[5.1.17]: Show that if $q_{n-1}<q<q_n$ then $|qr-p|>|q_{n-1}r-p_{n-1}|$. (*Idea:* Otherwise if the minimum $\bar\varepsilon=\min|qr-p|$ over $q_{n-1}<q<q_n$ and over p was reached for some

$\bar{q},\bar{p}$ then $\bar{p}/\bar{q}$ would be an optimal approximation). Check that this can be interpreted by saying that the graph of the function $\eta(q)=\min_p|qr-p|$ is above that of the function $\eta_0(q)=\varepsilon_n=|q_nr-p_n|$ for $q_n\le q<q_{n+1}$.

[5.1.18]: Let n even: the point $q_nr\,\mathrm{mod}\,1$ can be represented, thinking of the interval $[0,1]$ as a circle of radius $1/2\pi$, as a point shifted by ε_n to the right of 0, while $q_{n-1}r$ can be thought of as a point shifted by ε_{n-1}. Show that the points qr with $q_n<q<q_{n+1}$ are *not* in the interval $[0,\varepsilon_{n-1}]$ unless q/q_n is an integer $\le a_{n+1}$. Moreover check that the points $q_{n-1}r+aq_nr$ get closer to 2π by ε_n as a increases and stay on the same side of 2π as $q_{n-1}r$ until $a=a_{n+1}$, which therefore gives the next closest approach $\varepsilon_{n+1}<\varepsilon_n$. Finally check that this gives us a natural interpretation of the entries a_j of the continued fraction of r regarded as the angle of a rotation of the circle $[0,1]$, and at the same time it yields a geometric interpretation of the relation $a_{n+1}q_n+q_{n-1}=q_{n+1}$, and compare it with the construction in [5.1.28] of the rotation number of a circle map..

[5.1.19]: Check that the function $\varepsilon(T)=$ *maximum interval between points having the form* $n\,r\,\mathrm{mod}\,1,\ n=1,0,\ldots,,T$ is:

$$\begin{array}{rlr} q_n \le & T<q_n+q_{n-1} & \varepsilon_{n-1}\\ q_n+q_{n-1}\le & T<2q_n+q_{n-1} & \varepsilon_{n-1}-\varepsilon_n\\ \ldots & & \\ (a_{n+1}-1)q_n\le & T<a_{n+1}q_n+q_{n-1}\equiv q_{n+1} & \varepsilon_{n-1}-(a_{n+1}-1)\varepsilon_n \end{array}$$

and as an application draw (qualitativelt, for a generic r) the graph of $\varepsilon(T)$ and of the inverse function $T(\varepsilon)$ for the golden number, *i.e.* the number with $a_j\equiv1$. Draw (qualitatively) the graph of $-\log\varepsilon(T)$ as a function of $\log T$). (*Idea:* Reinterpret [5.1.18]).

[5.1.20]: Check that if the entries a_j of the irrational r are uniformly bounded by N then the growth of q_n is bounded by an exponential in n (and we can estimate q_n with a constant times $[(N+(N^2+4)^{1/2})/2]^n$). However an exponential estimate on q_n can even hold if the continued fraction entries are not uniformly bounded. (*Idea:* The continued fraction with $a_j\equiv N$ yields convergents which are an upper bound to q_n; for the converse use the recursion in [5.1.8]).

[5.1.21]: Show that if the inequality: $|q_nr-p_n|>1/Cq_n$ holds for all n and for a suitable constant C then q_n cannot grow faster than an exponential in n. (*Idea:* [5.1.10] implies the inequality $1/Cq_n<1/q_{n+1}$).

[5.1.22]: If an irrational number has a continued fraction whose entries are eventually periodically repeated then it is a number verifying a quadratic equation with integer coefficients: one says that the irrational is *quadratic*.

[5.1.23]: (Euclid) Suppose that r is a quadratic irrational, *i.e.* for certain integers a,b,c it is $ar^2+br+c=0$. Note that [5.1.8] shows that the number $r_n=[a_n,a_{n+1},\ldots]$ verifies $r=(p_{n-1}r_n+p_{n-2})/(q_{n-1}r_n+q_{n-2})$. Inserting the latter expression in the equation for r one finds that r_n verifies an equation like: $A_nr_n^2+B_nr_n+C_n=0$. Check, by direct calculation of A_n,B_n,C_n, that

$$\begin{aligned} &A_n=ap_{n-1}^2+bp_{n-1}q_{n-1}+cq_{n-1}^2\\ &C_n=A_{n-1}\\ &B_n^2-4A_nC_n=b^2-4ac \end{aligned}$$

Check also that $|A_n|,|B_n|,|C_n|$ are uniformly bounded (in n) by $H=2(2|a|r+|b|+|a)+|b|$. (*Idea:* It suffices to find a bound for $|A_n|$. Write $A_n=q_{n-1}^2(a\,(p_{n-1}/q_{n-1})^2+b\,(p_{n-1}/q_{n-1})+c)$ and note that $|r-p_{n-1}/q_{n-1}|<1/q_{n-1}^2$ and $ar^2+br+c=0$).

[5.1.24]: (Euclid) Show that the entries of the continued fraction of a quadratic irrational are eventually periodic because, in virtue of the preceding problem results, the numbers r_n can only take a finite number of values. Check that, if H is the constant introduced in [5.1.23], the length of the period of the continuous fraction can be bounded by $2(2H+1)^3$ and that the periodic part must begin from the j–th entry with $j \le 2(2H+1)^3$.

[5.1.25]: Let $\underline{\omega} \in R^l$ be an angular velocity vector verifying the *Diophantine property*: $|\underline{\omega}\cdot\underline{n}| > C_0^{-1}|\underline{n}|^{-\tau}$ for all non zero $\underline{n}\in Z^l$ and for suitable constants C_0,τ. Show that there is a constant γ_l such that the quasi periodic motion on T^l defined by $t\to\underline{\omega}t$ will have visited, after a time $T<\gamma_l C_0\varepsilon^{-(l+\tau)}$ *all* boxes with side ε. Show that we can interpret this by saying that C_0 is a *reference time scale* for the filling of T^l by the motion $t\to\underline{\omega}t$; while τ is related to the number of units of C_0 necessary to the motion to visit all square boxes of size ε.
(*Idea:* Let $\chi(\underline{x})$ be a non negative C^∞ function on R^l vanishing outside the unit square and with integral 1. Define the function χ on T^l: $\chi_\varepsilon(\underline{\psi}) = \varepsilon^{-l}\chi(\underline{\psi}\varepsilon^{-1})$ for $|\psi_j| < \varepsilon \bmod 2\pi$ and 0 otherwise. Then χ_ε is a C^∞ periodic function on T^l. Its Fourier transform is $\hat\chi(\varepsilon\underline{n})$ if $\hat\chi(\underline{w})$ is the Fourier transform of χ regarded as a function on R^ℓ; the function $\hat\chi$ is such that $\hat\chi(\underline{0})\equiv 1$ and $|\hat\chi(\underline{w})|\le\Gamma^\ell_\alpha|\underline{w}|^{-\alpha}$ for all $\underline{w}$, $|\underline{w}|>1$ and for all $\alpha\ge 0$, with Γ_α constants. Then note that a box around $\underline{\psi}_0$ will have been certainly visited before the time T if $I = T^{-1}\int_0^T \chi_\varepsilon(\underline{\omega}t-\underline{\psi}_0)\,dt > 0$. Thus we can exploit that

$$I \equiv 1 + T^{-1}\sum_{\underline{n}}\hat\chi(\varepsilon\underline{n})\frac{e^{i\underline{\omega}\cdot\underline{n}\,T}-1}{i\underline{\omega}\cdot\underline{n}} \ge 1 - 2T^{-1}C\sum_{\underline{n}\ne\underline{0}}|\hat\chi(\varepsilon\underline{n})||\underline{n}|^\tau$$

and we can bound the number of $\underline{n}$ such that $k-1<\varepsilon|\underline{n}|\le k$ by $\Omega_\ell\varepsilon^{-\ell}k^\ell$ and, for such $\underline{n}$'s it is $|\underline{\omega}\cdot\underline{n}|^{-1} < Ck^\tau\varepsilon^{-\tau}$. We see that a lower bound on I is $I > 1 - 2T^{-1}\,\varepsilon^{-\ell-\tau}\,CG^\ell_{\ell+2+\tau}\sum_{k>0}k^{\ell+\tau}k^{-(l+2+\tau)}$. This means that the time T can be taken as asked above.)

[5.1.26]: Compute the continued fraction of the positive solution of $x=\frac{1}{1+x}$ (*golden number*) and that of $\sqrt{2}$. Estimate in terms of $\varepsilon>0$ the value of N necessary in order that any point in the interval $[0,1]$ has within a distance $\varepsilon>0$ a point of the sequence $x_n=[nr]$ with $n\le N$ with r equal to the golden number or to the Pythagoras' number ($r=\sqrt{5}-1$ or $r=\sqrt{2}$ respectively).

[5.1.27]: Show that, in [5.1.25], $l+\tau$ can be replaced in the estimate for T by $(\ell-1)+\tau$. This means that the time T in [5.1.25] can be taken shorter: $T=\gamma_l C_0\varepsilon^{-(\ell-1+\tau)}$.
(*Idea:* Imagine to center a cylinder at the origin of R^ℓ with basis a disk perpendicular to $\underline{\omega}$ and radius 1, and height also 1 along the axis parallel to $\underline{\omega}$. If $\underline{\psi}\in R^\ell$ we denote $\underline{\psi}^\perp,\psi^\parallel$ the orthogonal projections of $\underline{\psi}$ on the plane perpendicular to $\underline{\omega}$ and along $\underline{\omega}$, respectively. Let χ^0,χ^1 be two C^∞ functions on $R^{\ell-1},R$ that vanish outside the unit ball, are positive and have integral 1. Then

$$\vartheta_\varepsilon(\underline{\psi}) = \varepsilon^{-(\ell-1)}\,\chi^0(\underline{\psi}^\perp)\,\chi^1(\psi^\parallel)$$

can be regarded, for ε small enough, as periodic functions on the torus $[-\pi,\pi]^\ell$ vanishing outside the cylinder with axis parallel to $\underline{\omega}$ and height 1, and having as bases disks orthogonal to $\underline{\omega}$ of radius ε. If $\hat\chi^0(\underline{w}),\hat\chi^1(w)$ are the Fourier transforms of χ^1,χ^0 regarded as functions on $R^{\ell-1},R$ then the Fourier transform of ϑ_ε will be $\hat\vartheta_\varepsilon(\underline{n}) = \hat\chi^0(\varepsilon\underline{n}^\perp)\,\hat\chi^1(n^\parallel)$ if $\underline{n}$ is an integer components vector and $\underline{n}^\perp$ and $n^\parallel$ are its components orthogonal and parallel to $\underline{\omega}$. Then we note, as in the hint to [5.1.25], that $I > 1-2T^{-1}C_0\sum_{\underline{n}\ne\underline{0}}|\hat\chi^0(\varepsilon\,\underline{n}^\perp)|\,|\hat\chi^1(n^\parallel)|\,|\underline{n}|^\tau$. Furthermore the sum can be bounded proportionally to $\varepsilon^{-(\ell-1)-\tau}$ because, fixed k,h integers, the number of $\underline{n}$'s with $k-1<\varepsilon|\underline{n}^\perp|\le k$ and $h-1<|n^\parallel|\le h$ is bounded proportionally to $\varepsilon^{-(\ell-1)}k^{\ell-1}h$ and $|\underline{n}|^\tau$ is bounded proportionally to $\varepsilon^{-\tau}k^\tau+h^\tau$ and, analogously to the case in [5.1.25], the functions $\hat\chi^0(\underline{w})$ and $\hat\chi^1(w)$ decay faster than any power at ∞.)

Remark: The estimate in [5.1.27], $\overline{T} > O(\varepsilon^{-\tau-(d-1)})$, really deals with a quantity *different* from the minimum time of visit. It is an estimate of the minimum time beyond which all cylinders with height (along $\underline{\omega}$) 1 (say) and basis of radius ε have not only been visited but they have been visited with a frequency that is, for all of them, larger than $\frac{1}{2}$ of the asymptotic value (proportional to $\varepsilon^{\ell-1}$): we can call the latter time the *first large frequency-of-visit time*. I *think* that $O(\varepsilon^{-\tau-(d-1)})$ is also optimal as an estimate of the first large frequency of visit time. It is known that the optimal result for the first time of visit is much shorter, [BGW98], and of order $\varepsilon^{-\tau}$: check this statement for the case $\ell = 2$ by making use of the above theory of continued fractions and see p. 496 in [BGW98].

[5.1.28]: (*rotation number of a circle map*) Devise an algorithm, analogous to the above for building the continued fraction of a number, to construct the rotation number of a map of the line $\alpha \to g(\alpha)$ with g increasing, continuous and such that $g(\alpha + 2\pi) = g(\alpha) + 2\pi$, so that it can be regarded as a circle map, *c.f.r.* problem [4.3.3]. (*Idea:* Suppose that $0 < g(0) < 2\pi$ for simplicity. We construct a sequence $q_{-1} = 0, q_0 = 1, q_1, \ldots$ as follows. Let $a_1 \geq 1$ be the largest integer such that $g^{aq_0+q_{-1}}(0) \equiv g^a(0) < 2\pi$ for $a = 1, 2, \ldots, a_1$: we say that $x_1 = g^{a_1q_0+q_{-1}}(0)$ is close to 2π to order 1. Then we set $q_1 = a_1 + q_0$ and we start from $g^{q_0}(0)$ and apply to it g^{q_1}: and its iterates which will fall in the arc $[0, g^{q_0}(0)]$ until a maximal value $a_2 \geq 1$ of iterations is reached (a check that $a_2 \geq 1$ is necessary). Then we say that $x_2 = g^{a_2q_1+q_0}(0)$ is close to 2π to order 2 and set $q_2 = a_2q_1 + q_0$. Likewise starting from $g^{q_1}(0)$ we apply iterates of g^{q_2} until we reach a point x_3 that will be close to 2π to order 3. And we continue: the process will never stop *unless* the point $x_{k-1} = g^{q_{k-1}}(0)$, close to 2π to order $k-1$ for some k, will have $a_{k+1} = +\infty$. In this case $g^{aq_k}(x_{k-1})$ has a limit as $a \to \infty$ which is easily seen to be a periodic point with period q_k. Let $a_1, a_2, \ldots$ be the sequence constructed in this way (which is infinite unless one of its entries is $a_k = +\infty$ for some k). And let us define $\rho = (0, a_1, a_2, \ldots)$ the value of the continued fraction with the a_j as entries (having set $a_0 = 0$), see [5.1.5]. Then by the properties of the continued fractions discussed above it follows that also ωq_k is a sequence of approximants of 2π and $|g^{q_l} - q_k 2\pi\rho| < x_0 < 2\pi$ so that $(2\pi\, q_k)^{-1} g^{q_k}(0) \xrightarrow[k\to\infty]{} \rho$ and ρ is the rotation number of the map g.)

[5.1.29]: In the context of the previous problem show that the rotation number $\rho(g)$ of a circle map g with the properties in [5.1.28], *i.e.* a map often called a *circle homeomorphism*, is continuous in g in the sense that if $d(g_n, g) = ||g_n - g|| \stackrel{def}{=} \max_{\alpha\in[0,2\pi]} |g_n(\alpha) - g(\alpha)| \xrightarrow[n\to\infty]{} 0 = 0$ then $\rho(g_n) \to \rho(g)$. (*Idea:* This is implied by the construction of $\rho(g)$ in [5.1.28].)

Ergodic theory of geodesic flows on surfaces of constant negative curvature.

[5.1.30]: Check that the transformations of the complex plane $z = x+iy \to z' = x'+iy'$ defined by $z' = \frac{az+c}{bz+d}$ with a, b, c, d complex and $ad - bc = 1$ (*planar homographies*) transform circles into circles and preserves angles. Furthermore if a, b, c, d are real it maps the Lobatchevsky plane onto itself and preserves the Lobatchevsky distances, *i.e.* ${y'}^{-1}|dz'| = y^{-1}|dz|$. (*Idea:* Note that the homographies are conformal, being holomorphic; furthermore use direct calculation to check metric invariance. Then note that

$$z' = \frac{a}{b}\frac{(zb + bc/a)}{zb + d} = \frac{a}{b} - \frac{(ad - bc)/b^2}{z + d/b} \tag{$*$}$$

hence it suffices to check that the map $z' = R^2/z$, $R > 0$, transforms circles into circles).

[5.1.31]: A *reflection* with respect to a circle C centered at O and with radius R is the map transforming a point P at distance d from O into P' on the half line $OP\infty$ at distance $d' = R^2/d$ from O. Check that each planar homography is a map that can be represented as the composition of a reflection with respect to a circle C_1 and of another with respect to a circle C_2. One of the two circles can always be chosen "infinite", *i.e.* a straight line. (*Idea:* The map $z' = R^2/z$ is the composition of the reflection $z' = -\overline{z}$ with respect to the imaginary axis and and of the reflection $z' = R^2(\overline{z})^{-1}$ with respect to the circle of radius R centered at the origin, (see *) in [5.1.30]).

[5.1.32]: Check the following statements. The homography $z' = \frac{z-i}{z+i}$ maps the upper half plane into the unit disk, mapping i into the origin and the real axis on the unit circle $|z'| = 1$. The metric $\frac{dx^2+dy^2}{y^2}$ (called the *Lobatchevsky metric*) becomes in the new coordinates $\frac{dx^2+dy^2}{(1-x^2-y^2)^2}$ (called the *Poincaré metric*). The group G of the homographies $z' = \frac{az+c}{\overline{c}z+\overline{a}}$, with a, c complex and $|a|^2 - |c|^2 = 1$, is the group of "rigid motions" of the Poincaré's metric. Hence the group $SL(2,R)$ and the group G are isomorphic: find explicitly a realization of the isomorphism. (*Idea:* Denote $z' = (az+c)/(bz+d)$ as zg if $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Denote $\Gamma = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}$ and note that the map of z in the Lobatchevsky plane to z' in the Poincaré disk can be written $z' = z\Gamma$, if the action of the homographies $g \in SL(2,R)$ on the Lobatchevsky plane is $z \to zg$, see [5.1.30]: hence the homographies with matrices $g' = \Gamma^{-1} g \Gamma$ with $g \in SL(2,R)$ will map the Poincaré disk into itself preserving the metric, because g preserves the Lobatchevsky's metric while Γ^{-1} changes the Poincaré disk into the Lobatchevski plane (carrying along the metric) and Γ does the opposite operation. Hence the group of the movements of the Poicarés geometry on the disk is the group $G = \Gamma^{-1} SL(2,R) \Gamma$, isomorphic to $SL(2,R)$).

[5.1.33]: Check that the the Lobatchevsky metric geodesics are half circles orthogonal to the real axis while those of the Poincaré's metric are circles orthogonal to the unit circle. (*Idea:* The half lines orthogonal to the real axis are obviously geodesics for the Lobatchevsky metric: and the real planar homographies, *i.e.* the elements of $SL(2,R)$, leave the Lobatchevsky metric invariant and map half lines orthogonal to the real axis into circles orthogonal to the real axis; in the case of the Poincaré's metric the role of the half lines orthogonal to the real axis is taken by the diameters of the unit disk).

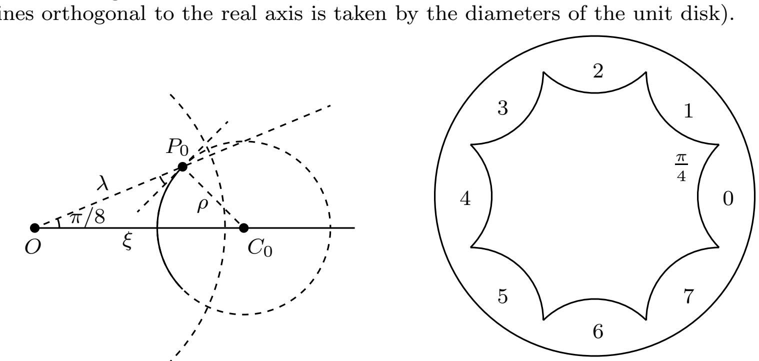


Fig. (5.1.1) *Illustration of the drawings proposed in* [5.1.34] and [5.1.35].

[5.1.34]: Divide the plane into eight sectors separated by the radial lines out of the origin O at angles $\pi/8 + j\pi/4$ with the real axis, (Fig. 5.1.1). On each radial line mark a point P_j at distance λ from the origin. At an angle $\pi/8 + \pi/2$ from the radial lines draw a segment joining P_j with a point C_j on the bisectrix of the j–sector (hence adjacent to the line); call ξ the distance of C_j from O. From each C_j draw a circle with radius ρ equal to the length of the segment P_jC_j. Note that the circle with center O and radius $R = \sqrt{\xi^2 - \rho^2}$ is orthogonal to the eight circles already drawn. Compute λ, ρ, ξ *if* $R = 1$. (*Idea:* $\lambda = 2^{-1/4}$, $\rho = \lambda\,(2+\sqrt{2})^{-1/2}$, $\xi = \lambda\,(\frac{2+\sqrt{2}}{2})^{1/2}$).

[5.1.35]: Consider the octagonal figure Σ_8 cut in the circle of radius $R = 1$ in [5.1.34] Enumerate its arcs consecutively from 0 to 7 and note that by reflecting the figure around one of its "sides" L_j (in the sense of [5.1.31]) one obtains an octagonal figure with angles at the vertices still of $\pi/4$ and contained inside the unit circle: furthermore vectors "exiting" from the octagon become "exiting " vectors from the new figure. Construct with compass and ruler the initial octagonal figure. Write a computer program that draws it and its reflections around the eight sides that constitute its boundary. (*Idea:* Σ_8 is the well known magic figure drawn in the front cover of this book. Upon reflection of it around the side L_j one get a small bug resembling (but different from it) the one in Fig. (5.1.2) adjacent to the side L_j and outside the octagon, but inside the circle of radius 1.)

[5.1.36]: By reflecting the octagonal figure Σ_8 obtained in [5.1.35] with respect to the diameter "parallel" to the side L_j and then reflecting the result with respect to the circle containing the side of the octagon opposite to L_j one obtains a planar homography (because it is a composition of two reflections around circles) that allows us to *identify* the considered side and the one opposite to it (note that the only reflection *is not sufficient*, although it even maps the octagon into itself because it is not a homography and *worse* because a vector exiting from L_j becomes a vector *also exiting* from the opposite side). If we consider the subgroup of the homographies, obtained by performing at each side the reflections considered above, we obtain a subgroup Γ_8 of the group of the isometries of the Poincaré metric. Identifying points of the boundary of the octagon Σ_8, modulo the transformations of the group Γ_8, and imagining Σ_8 as a surface with the metric induced by the Poincaré's metric, the octagon becomes a closed, smooth, boundaryless, surface. Check that topologically this surface is a "donut with two holes", see also [Gu90].

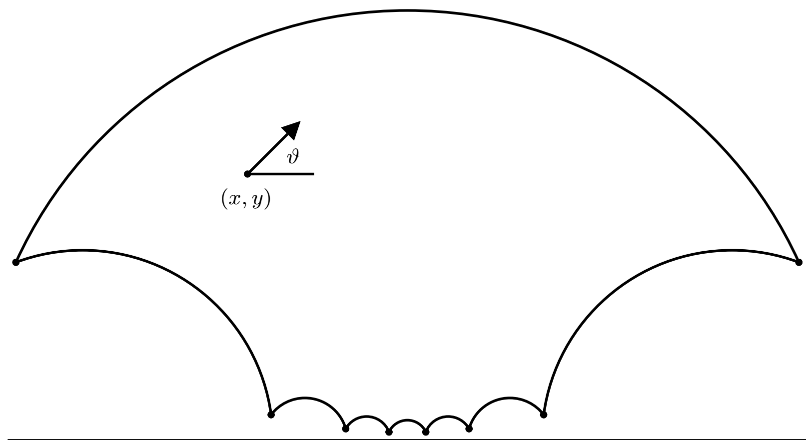


FIg. (5.1.2): *A fundamental domain in the Lobatchevsky plane: it is the image of the domain in the front cover. The (x, y, ϑ) denotes a light ray at (x, y) and direction ϑ.*

[5.1.37]: The octagon Σ_8 of the above problems and its group Γ_8 can be regarded as a figure in the half plane of Lobachevsky and, respectively, as a subgroup of the group $SL(2, R)$ of the rigid motions of this geometry (simply by using the isometry between the Poincaré disk and the Lobachevsky plane discussed in [5.1.28]). We shall call them with the same names Σ_8 and Γ_8. Draw the octagon Σ_8 as a figure in the Lobachevsky plane (with compass and ruler, or with a computer (see Fig. (5.1.2) and admit that it is less fascinating than Fig. (5.1.1))).

[5.1.38]: (a) Let $GL_2(R)$ be the group of the real matrices $g = \begin{pmatrix} p & p' \\ q' & q \end{pmatrix}$ with determinant $\det g = pq - p'q' > 0$ Consider the variables p, q and p', q' as canonically conjugate variables. Consider the flow on $GL_2(R)$ generated by the Hamiltonian $H(g) = \frac{1}{8}(\det g)^2$ is $t \to g(t) = g\, e^{-\frac{1}{4}(\det g)\,\sigma t}$ if $\sigma = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Set $x(t) + iy(t) \equiv z(t) \stackrel{def}{=} i\, g(t)^{-1}$ and
(b) check that $z(t)$ is a motion that runs over a geodesic at velocity $v = v_x + iv_y = \dot z(t)$ and constant speed $|v| = \frac{1}{2}\det g$ (measured in the Lobatchevsky metric as $|v|^2 = (v_x^2 + v_y^2)/y^2$, *c.f.r.* §2 of [CEG84]). And

(c) Check that the motion $t \to z(t)$ and its velocity $\dot z(t)$ with $z(t) = i\,g(t)^{-1}$ is such that the relazion between (z,v) and g is

$$z = i\,g^{-1}, \qquad v = \frac{(\det g)^2}{2}\,\frac{i}{(-p'\,i+p)^2} \tag{!}$$

and the latter relation can be inverted determining g up to a sign. (*Idea:* It is $z(t) = ie^{\frac14(\det g)\,\sigma t}g^{-1} = (e^{\frac12 t\det g}\,i)\,g^{-1}$. Therefore the motion $z(t)$ is the image under the mouvement g^{-1} of the simple motion $t \to e^{\frac12 t\det g}\,i$ which has a trajectory in space which is the imaginary half axis, *i.e.* a geodesic of the Lobatchevsky plane, and it goes over it at speed $\dot y(t)/y(t) = \frac12\det g$. Hence it is a geodesic motion with speed $2^{-1}\det g$. The item (c) is simply obtained by differentiation of $z(t)$.)

[5.1.39]: Consider the surface Σ_8 obtained by identifying the points of the Lobatchevsky plane L modulo the mouvements of the group Γ_8 introduced above, in the sense that z, z' will be considered the same point if there is $g_0 \in \Gamma_8$, $g_0 = \begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}$ such that $z' = z\,g_0$. By the previous problem, see equation (!), we know that we can use the matrices of $PGL_2(R)$, defined as the group $GL_2(R)$ with the identification of its elements differing by just a sign, as "coordinates" for the pairs (z,v) of points in position–velocity space of the geodesic motions on L. Show that the matrices of $PGL_2(R)/\Gamma_8$, *i.e.* the space of the 2×2–matrices with positive determinant that are *identified* modulo $\pm\Gamma_8$ in the sense that g and g' are considered equal if $g' = \pm g_0\,g$ for some $g_0 \in \Sigma_8$, *can be cosidered a system of smooth coordinates* for the geodesic motions on the surface Σ_8. In other words check that if (z,v) and (z',v') are such that $z' = z\,g_0$ and $v' = v/(\beta z+\delta)^2$, *i.e.* if (z,v) and (z',v') represent the same position and velocity of a motion on Σ_8 then the matrices g, g' that represent (z,v) and (z',v') as in (c) of [5.1.38] are *also* necessarily related by $g' = \pm g_0^{-1}g$. (*Idea:* If a motion on Σ_8 is in z with velocity $v = \dot z$ then the same motion can be described as being in $z' = z\,g_0$ with velocity $v' = v/(\beta z+\delta)^2$ because of our identifications: however this means that the pair (z',v') is described by the (equivalent by our definitions, as g_0^{-1} also is in Γ_8) matrix $g_0^{-1}g$ as one checks by expressing the derivative of $ig^{-1}g_0$ in terms of the derivative $\dot z$ of $ig^{-1} = z$ when z varies as a function of a parameter (*e.g.* of time). One simply uses the easily checked "composition rule": given two matrices g_1, g_2 and setting $j(z,g) = (bz+d)$ then $j(z, g_1g_2) = j(z,g_1)j(zg_1,g_2)$).

[5.1.40]: Hence, by the previous problem, if $g_0 \in \Gamma_8$ and if the points $g \in PGL_2(R)$, g_0g are identified ("identification modulo Γ_8") one obtains a surface that can be identified with the space of the pairs (z,v) of points z in Σ_8 and vectors v tangent to the surface Σ_8: the geodesic flow on Σ_8 is in the new coordinate given by the matrix elements of g simply $g \to ge^{\frac12\sigma t}$, see also [Gu90]. Furthermore the surface of energy 1 can be identified with the matrices $g \in SL(2,R)$ "modulo Γ_8" with determinant 2 which in turn can be identified with those with determinant 1 always "modulo Γ_8" (simply by scaling them by $1/\sqrt2$). This surface will be denoted $\Sigma_{8,1}$. Check that also the "Liouville measure on the phase space $PGL_2(R)/\Gamma_8$" $\mu(dg) = \delta(\det g - 1)dpdqdp'dq'$, is an invariant measure with respect to the action of $SL(2,R)$: $\mu(E) \equiv \mu(Eg)$ for every $g \in SL(2,R)$. (*Idea:* Just note that locally on the space of matrices the geodesic evolution is Hamiltonian, by [5.1.38].) *All the above leads one to strongly suspect that if $p_x = v_x/y, p_y = v_y/y$ then the map $(x,y,p_x,p_y) \longleftrightarrow (q,q',p,p')$ is a canonical map: this is indeed true, see [CEG84], [Ga83].*

[5.1.41]: Let $F \in L_2(\Sigma_{8,1},\mu)$, with μ defined in [5.1.40] (note that if F is regular it must be $F(g_0g) = F(g)$ for all elements $g_0 \in \Gamma_8$). We can define a representation of the group $SL(2,R)$ on L_2 associated with $\Sigma_{8,1}$ via: $g \to U(g)$ with $U(g)F(g') = F(g'g)$. Show that this transformation is unitary because the measure μ is invariant. Hence the evolution of the geodesic flow can be seen as a unitary transformation on $L_2(\Sigma_{8,1},\mu)$: check that, indeed, $F \to F(ge^{\frac12\sigma t})$ is unitary.

[5.1.42]: More generally define $U(g)f(g') = f(g'g)$ for $f \in L_2(\Sigma_{8,1},\mu)$ and check that this is a unitary representation of $SL(2,R)$. One could, also, check that the compactness of $\Sigma_{8,1}$ implies that this representation is *reducible*, *i.e.* that it is decomposable into a direct sum of irreducible unitary representations, *c.f.r.* [GGP69]. Since all the representations of $SL(2,R)$ are well known one sees immediately from their description, see for instance [GGP69],[CEG84], that there is no one, other than the trivial representation, which contains a function invariant under the action of the operators $U(e^{\frac{1}{2}\sigma\, t})$ for all t.

[5.1.43]: (*Ergodicity of the geodesic flow on a surface of constant negative curvature*) Show that among the irreducible components of the representation of $SL(2,R)$, existing by [5.1.42], on $L_2(\Sigma_{8,1},\mu)$ described in [5.1.41] the trivial representation has necessarily multiplicity 1. Togheter with the classic result on the structure of the irreducible representations of $SL(2,R)$ quoted in [5.1.42] this implies that in $L_2(\Sigma_{8,1},\mu)$ there cannot be functions invariant for the geodesic flow: *this property is one of the many definitions of ergodicity* of a flow, see also §5.3. (*Idea:* If $f\in L_2(\Sigma_{8,1},\mu)$ transforms according to the trivial representation it is $U(g)f(g')=f(g'g)$ and note that $g'g$ takes all possible values as g varies while g' is kept fixed.)).

[5.1.44]: (*Spectrum of the geodesic flow on a surface of constant negative curvature*) The theory of the power spectrum, and the proof of its continuity, can be reduced to simple problems of the theory of the representations of $SL(2,R)$ (which is also called "Fourier analysis" for $SL(2,R)$) which play the role of the complex exponentials of the ordinary Fourier transform for the analysis of the geodesic flows on tori (*i.e.* the quasi periodic motions). We prefer to refer the interested reader to the literature for such developments, [GGP69],[CEG84].

Bibliography:

See [AA68],[Ga81],[Ka76]. The problems on continued fractions are taken mainly form [Ki63], [Ga83]. The theory of geodesics on the "octagon" is taken from [CEG84]. The analysis of the ergodicity of the geodesic flow on the octagon based on the theory of the representations of $SL(2,R)$ and sketched in the problems following [5.1.38] is rather simple. A geometric theory, independent of representations, is possible. And, in fact, the first proof of the ergodicity of the geodesic flows on compact smooth surfaces of negative curvature (constant or not) has been of purely geometric nature and is due to Hopf (see [Ga74] for a rapid exposition of the main idea): and it had major influences on the development of the theory of hyperbolic flows and maps, (see *c.f.r.* §5.4, §5.7).

§5.2 Timed observations. Random data.

Motion, whether of a fluid or of a general dynamical system described by a differential equation in R^ℓ, is usually studied by observing the evolution at discrete times.

Not only because of the intrinsic impossibility of performing observations on a continuum set of times, but also to avoid performing trivial measurements (consisting essentially in useless repetitions of measurements just past): it is indeed clear that if one performs measurements too frequently in time the results that one obtains differ from each other in a way difficult to appreciate and of little relevance.

Informations about a physical process, *i.e.* about the very existence of motion and the variety of its properties, can only be obtained by comparing measurements done over lapses of time that differ "by the time scale" over which the evolution is noticeable: one should recall, in this context, the *Zeno's paradox*.

Such time scales can vary enormously from system to system and from phenomenon to phenomenon. For instance the tropical eddie that originates the red spot on Jupiter requires observations on a time scale of hundreds of years, or so, to notice its evolution; a terrestrial tropical eddie requires only a few hours; precession of the Earth axis required to Hypparchus hundreds of years of observations to be seen; the precession of the axis of Hyperion (Saturn's satellite) demands a few days, [Wi87]; the variation of the axis of Mars requires hundreds of thousands of years, [LR93].

Therefore in all experiments one chooses to make observations at significant instants, well separated from each other. For instance if a system is subject to a periodic force it is often convenient to perform measurements at intervals "*timed*" on the period of the driving force (*i.e.* at times that are multiples of the period). Measurements consisting in taking "movies" may seem an exception: however even movies are timed observations, because they consist in a rapid succession of photograms taken at constant pace; and one can think of time t itself as an observable that takes values on a circle, like the dial of a clock, and perform measurements every time that the "arm" of the clock indicates a certain point.

More generally if $\mathcal{P}$ is a selected property of the system, one could time observations by making them every time the property $\mathcal{P}$ is verified. For instance the property $\mathcal{P}$ could be the equality of the values of two observables F_1 and F_2 (*i.e.* every time the phase space point u crosses the surface $F_1(u) = F_2(u)$); or the property that an observable F assumes a local maximum, or other. Observations timed on the period of a periodic driving force, or executed at equal time intervals in systems not subject to time dependent forces, can be considered as timed observations as well.

The phase space in which vary the data relevant for a certain timed observation has usually dimension $\ell - 1$, if ℓ is the dimension of the phase space in which the system is described by an autonomous differential equation. This is true, in a way, also in the cases of observations that are executed at equal time intervals and on autonomous systems: at least if we think of time t as of a coordinate, thus increasing by 1 the dimension of the system. As said in §4.2 we shall always suppose that the dimension ℓ of phase space is $< \infty$: in the case of fluid motions this means that we shall consider as a good model a finite model of the equations of the fluid, see §2.1%§3.2 and §4.1.

If observations are timed the dynamics is replaced by a map S of the set $M \subset R^\ell$, (usually with $\ell - 1$ dimensions), of the phase space points on which the chosen timing property $\mathcal{P}$ is verified. This map is defined by associating with a point $u \in M$ the point $u(t_0) \in M$, that is the point into which u evolves at the instant t_0 when, for the first time, the motion in continuous time $t \to u(t)$ enjoys again the property $\mathcal{P}$.

Note that the map S can be considered as an extension of the notion of "Poincaré map", c.f.r. definition 1 in §4.3, (E), *relative to the surface consisting in the points enjoying the property $\mathcal{P}$.*

We shall still call M the "phase space", adding the qualifier of *discrete* when a distinction between the two phase space notions becomes necessary. The space R^ℓ or the manifold in R^ℓ in which the system is described by a differential equation will be called, by contrast, the "*continuum* phase space". The map S, "timed dynamics", transforms M in itself.

The pair (M, S) will be an example of a general notion of "discrete dynamical system", that we shall have to introduce a little later in a more formal way, *c.f.r.* definition (2) of §5.3; here discreteness refers to the fact that dynamics is described by iterates of a map S and motion is a sequence $x \to S^n x$, rather than a continuous curve, because time is now "discrete".

The map S is usually *less regular* than the continuous family of transformations S_t that associate with a given initial point u the point $u(t) = S_t u$ into which u evolves at time t (*c.f.r.* [5.3.4], [5.3.5] of §5.3). It can also happen that S is *not* defined for all points that have the property $\mathcal{P}$: because there can be some data u that have the property $\mathcal{P}$ but that evolve without ever acquiring it again, (*c.f.r.* [5.3.4], [5.3.5], §5.3, for a typical example).

Obviously timed observations can only be useful for studying motions that acquire infinitely often the property $\mathcal{P}$ that is used for the timing. The property $\mathcal{P}$ is then said to be *recurrent* on such motions. And it is tautological that the map S, and its iterates S^n, $n \ge 0$ integer, contain "all informations concerning the property $\mathcal{P}$" that can be found in the transformations S_t, with $t \ge 0$ continuous variable.

In particular if A is an attracting set for the evolution S_t then $A_{\mathcal{P}} = \{$ set of the points of A that enjoy property $\mathcal{P}\}$ is an attracting set for S and "viceversa" the union of the trajectories with initial data on an attracting set $A_{\mathcal{P}}$ for S is an attracting set A for S_t. But one should not understand that the study of S is simpler: the study of *all* properties of S on $A_{\mathcal{P}}$ is equivalent to that of *all* those of S_t on A.

It will be convenient to fix, from now on, our attention upon timed observations, *i.e.* on motions described by the iterations of a map S defined over a set of points $M \subset R^{\ell-1}$ of dimension $\ell - 1$ that enjoy a prefixed property $\mathcal{P}$.

In terms of timed observations it is in fact natural to set up rather general method, quite homogeneous and without too many "technical exceptions" referring to the peculiarities of various systems, for the classification and description of the qualitative and quantitative properties of motion of a fluid or of a general system. Of course one should also keep in mind that the study of timed observations is, or can be, often more adherent to reality: experimental observations, as already said, are almost invariably observations performed at discrete intervals of time.

It is convenient to start by establishing some contact with §5.1 about continuum spectrum: indeed that notion was only introduced with reference to observations supposed taken in continuous time.

Consider a "dynamical system" (with discrete time) (M,S), an observable f and its history on the motion that begins in x: $k\to f(S^kx)$. Define:

1 Definition *(power spectrum for discrete time evolutions):*
*(1) We define the "*discrete autocorrelation function*" of the observable f on the motion $x\to S^nx$ to be the average $\Omega(k,x)$ of the product $f(S^{h+k}x)$ times $f(S^hx)$ intended as an average over the variable h at fixed k. We define then the (discrete) "*power spectrum*" by the Fourier transform*

$$A(\omega)=\frac{1}{2}\sum_{k=-\infty}^{\infty}\Omega(k,x)e^{i\,\omega\,k},\qquad \omega\in R \tag{5.2.1}$$

of the autocorrelation (c.f.r. (5.1.3)).
(3) The observable f is said to have continuous spectrum *on the motion $x\to S^nx$ if the (discrete) power spectrum of f on the motion is a continuous function of ω, or at least an L_1 function, on some interval $[\omega_1,\omega_2]$.*
*(4) A motion $n\to S^nx$ is said to "*have continuous spectrum*" if there is at least one observable which has continuous spectrum.*
*(5) a system (M,S) is said to have "*continuous spectrum*" with respect to a random choice of initial data with a probability distribution μ if, with μ–probability 1, any initial datum generates a motion over which all observables of a family $\mathcal{F}$, dense in $L_2(M)$ (see §5.1, definition 3), have a function $A(\omega)$ which is L_1 for $|\omega|>0$.*

Compare the above definition with the analogous definition 3 of §5.1.

As in §5.1 the (discrete) power spectrum of the observable f on the motion $x\to S^nx$ is, setting aside problems (often nontrivial) of exchange of limits, the limit

$$\lim_{N\to\infty}\frac{1}{2N}\Big|\sum_{k=0}^{N-1}f(S^kx)\,e^{ipk}\Big|^2 \tag{5.2.2}$$

as it can be checked along the same lines of the analogous case of §5.1, *c.f.r.* (5.1.1), (5.1.2)).

In reality the latter definition is an "improvement" over the one in §5.1 because it is based on a quantity more directly measurable: if the "output" of an experimental apparatus yields data $f(S^kx)$ then such data can be *directly* sent to a computer for performing "on line" the sum in (5.2.2).

An illustration of the type of questions that we encounter in trying to establish a relation between (5.2.2) and (5.2.1) is discussed in the problems [5.2.3] through [5.2.12].

If the function $k\to f(S^kx)$ is quasi periodic, *i.e.* if $f(S^kx)=\varphi(k\underline{\gamma})$ with φ a periodic function (with period 2π) of $\ell-1$ angles, with $\underline{\gamma}\in R^{\ell-1}$ for some ℓ, and if $\underline{\omega}=(\underline{\gamma},2\pi)$ verifies a Diophantine property, *c.f.r.* (5.1.5), then one can check that the discrete power spectrum of f is formed by a sum of delta functions concentrated on points $\omega = 2\pi n_0 + \underline{\gamma}\cdot\underline{n}$ where $n_0, \underline{n}$ are ℓ integers. This check is analogous to the corresponding one seen in §5.1: hence it is clear that we can adapt the comments of §5.1 to the case of discrete time power spectra.

If a motion $t \to S_t(x)$ does not have continuous spectrum it may happen that this is so because there are some observables whose evolution is periodic.

One can think, for instance, of a system to which we add "a clock" and consider the observable "position of the arms" on the dial. Then by timing the observations on the period of this observable we see that the position of the clock arm becomes a constant of motion *on every motion* of the system and the arm does not matter any more for the purposes of deciding whether a given motion has continuous spectrum.[1]

Thus we see that by timing the observations we can obtain the result of eliminating uninteresting informations, *i.e.* observables whose evolution can be regarded as trivial. This is the rule when systems driven by periodic forces are considered (in which, for instance, the observable force is trivially periodic).

It is for this reason that in definition 3 in §5.1 we defined a motion $t \to S_t x$, $t \in R$, with "continuous spectrum" if there is *at least one* observation (even very special) with respect to which the motion has continuous spectrum. This is also reflected in the above definition, item (4).

Examples of discrete dynamical systems with continuous spectrum for almost all initial data chosen with a distribution absolutely continuous with respect to the volume measure on phase space can be built in a rather simple way; we give here a list of particularly remarkable ones and we refer to the problems for further details.

(1) The map $S : x \to 4x(1-x)$ of the interval $[0,1]$ into itself: almost all initial data (with respect to the length) generate motions with continuous spectrum, see [5.2.1], [5.2.7], [5.2.8]. This map has a class of regularity C^∞, but it is not invertible; furthermore it conserves the probability distribution $\mu(dx) = dx/\pi\sqrt{x(1-x)}$.[2]

(2) The map $S :\, x \to 2x$ if $x < 1/2$ and $S :\, x \to 2(1-x)$ if $x \ge 1/2$ of the interval $[0,1]$ into itself: almost all initial data (with respect to the measure dx) generate motions with continuous spectrum. This map, often called the *tent*, because of the shape of the graph of Sx, is not invertible and it is only

[1] If one studied the evolution of the water of Niagara falls by observing *only* the position of the arms on the dial of a *Swatch* floating in them one could end up thinking that the motion is not turbulent.

[2] A map S, invertible or not, of a topological space M *conserves a measure* μ if for every measurable set E the set $S^{-1}E$, *i.e.* the set of the points x such that $Sx \in E$, has the same measure of E: $\mu(S^{-1}E) = \mu(E)$.

piecewise regular; furthermore it conserves the probability distribution μ given by the Lebesgue measure $\mu(dx) = dx$.

(3) The map of the bidimensional torus $T^2 = [0, 2\pi]^2$ into itself defined by $\underline{\varphi} = (\varphi_1, \varphi_2) \to \underline{\varphi}' = (\varphi_1', \varphi_2') \bmod 2\pi$ with $\varphi_1' = \varphi_1 + \varphi_2 \bmod 2\pi$ and $\varphi_2' = \varphi_1 + 2\varphi_2 \bmod 2\pi$ is such that almost all points, with respect to the area measure, of T^2 generate motions with continuous spectrum, for details see problems. Note that this map is invertible and has C^∞ regularity together with its inverse, in spite of the apparent discontinuity when one of the angles is multiple of 2π; it conserves the Lebesgue area measure $\mu(d\underline{\varphi}) = d\underline{\varphi}/(2\pi)^2$. This map plays an important role being a paradigm of chaotic discrete dynamical systems. It is sometimes called the *Arnold cat map* because of a well known illustration that Arnold gave of the action of this transformation, [AA68].

(4) Consider the transformation S of the preceding example. Define a 3–dimensional dynamical system acting on the points of $M = T^2 \times [0, L]$ as follows $S_t(\varphi_1, \varphi_2, z) = (\varphi_1', \varphi_2', z')$ where $\varphi_1' = \varphi_1, \varphi_2' = \varphi_2, z' = z + vt$ until $z' \le L$ and $(\varphi_1', \varphi_2') = S(\varphi_1, \varphi_2)$ and $z' = z - L + vt$ if $L < vt \le 2L$, *etc*: *i.e.* the point proceeds with constant velocity v along the z axis until it "collides" with the plane $z = L$; at collision it reappears on the plane $z = 0$ but with new φ_1, φ_2 coordinates (in other words boundary conditions identifying $(\varphi_1, \varphi_2, L)$ with $(S(\varphi_1, \varphi_2), 0)$ are imposed). See figure (5.2.1) below in which $x = (\varphi_1, \varphi_2), Sx = (\varphi_1', \varphi_2')$.

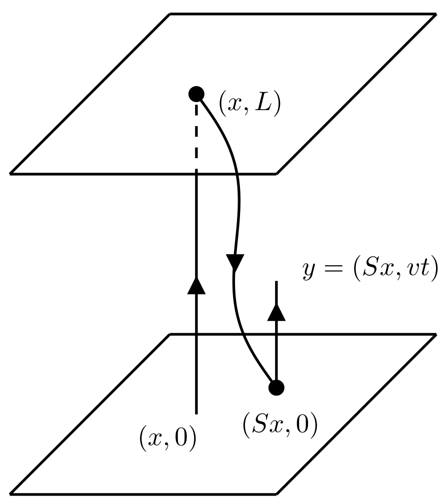


Fig.(5.2.1) *A continuous dynamical system with an embedded "Arnold's cat". The curved line is just a visual aid to the identification of (x, L) with $(Sx, 0)$.*

The dynamical system generated by the action of S_t on M conserves the measure $\mu(d\underline{\varphi}\, dz) = d\underline{\varphi}\, dz/(2\pi)^2 L$ and it *does not* have continuous spectrum; but by performing observations timed at intervals L/v, evidently, one

obtains a dynamical system for which almost all motions have a continuous spectrum, because it essentially coincides with the preceding example (3). With reference to the note [1] one can say that this is a chaotic system which "swallowed a clock".
(5) Evidently the construction of the example (4) is generalizable in various ways. It is thus possible to construct, even starting from the examples (1),(2),(3), other examples of continuous dynamical systems with non continuous spectrum, but such that suitable timed observations exhibit a continuous spectrum on most data.

(6) It is easily realized that in the preceding examples it is *always* necessary to say "almost all points": indeed in each of the examples one can find particular motions whose evolution *does not* have continuous spectrum. Think to $x = 0$ in the cases (1),(2) and to $\underline{\varphi} = \underline{0}$ in the case (3). This *is not* an accident. We shall see, *c.f.r.* §5.7, that the existence of motions with continuous spectrum, *i.e.* of chaotic motions, is almost inevitably accompanied by the existence of (unstable) periodic motions. Often such periodic orbits stem out of a set of initial data that is *dense* in phase space, but that has zero measure with respect to most invariant probability distributions μ that can be defined, as it happens for the probability distributions μ (which, in the above examples, are absolutely continuous with respect to the Lebesgue measure).

Having introduced the spectrum of an observable on a continuous or on a timed motion we should proceed to the analysis of the important notion of *statistics* of a motion observed either continuously in time or via timed observations and generated from a *randomly chosen* initial datum. For this purpose it is convenient to delay the analysis of the notion of statistics and to discuss once and for all, in the rest of this section, what "chosen at random" will precisely mean here.

A *random choice of data* is intended to be an algorithm that produces the $\ell-1$ coordinates of an initial datum u, that is then evolved in time building its *random history*, *i.e.* the sequence $S^n u$, $n = 0, 1, 2, \ldots$.

2 Definition: *(random choice of data):*
Consider an algorithm P_0 that produces, starting from an a priori given "seed", i.e. from a real number $s_0 \in [0,1]$,[3] *an ℓ–ple $\xi_1, \ldots, \xi_{\ell-1}, s_1$ of real numbers in $[0,1]$. The first $\ell-1$ numbers are collected into a $(\ell-1)$-ple that provides, in this way, a point x_1 in the square $Q = [0,1]^{\ell-1}$. The ℓ–th number s_1 will be a new seed to use to repeat the algorithm and generate again a new point in Q and a new seed, etc.*
The sequence $x_1, x_2, \ldots$ of points in Q produced by the algorithm P_0 and seed s_0, will be called a "sequence of random data" in Q with generator (P_0, s_0)
If $g : Q \to U$ is a regular function with values in an open set U on an

[3] Rational because irrational numbers are an abstraction of little utility when real calculations are performed.

arbitrary regular surface, the sequence $u_1 = g(x_1)$, $u_2 = g(x_2), \ldots$ *will be called a sequence of random data in* U *produced with the algorithm* $P = (P_0, s_0, g)$ *obtained "by composition" of* g *with the algorithm* P_0*. If* g *is the identity it will be omitted.*

Remarks:

(i) To choose a sequence of random data with respect to the algorithm P, means to build the sequence $u_1, u_2 \ldots$. Variations on this method for constructing random sequences with given distribution are possible but we shall not discuss them here.
(ii) One sees that there is *nothing random* in the sequence u_i: these are points constructed according to a precise rule *that yields always the same results* (unless computational errors are made, a possibility not to be discarded because the algorithms under consideration are generated, typically, on computers).
(iii) *Nevertheless* we conventionally call the sequences so constructed "random" and their "distribution" (see below) is defined by the algorithm $P = (P_0, \sigma_0, g)$ that generates them.
(iv) One can imagine algorithms so simple to produce sequences that no one would call random. On a computer that represents reals with 16 binary digits, and that performs multiplications by truncating the mantissa at 16 digits one can digitally program the iterations of the tent map S in example (2) above and define the algorithm with seed σ_0 by setting $x_0 = \sigma_0, \sigma_1 = S\sigma_0, x_1 = \sigma_1, \sigma_2 = S\sigma_1, x_2 = \sigma_2, \ldots$ (or in other words $x_k = Sx_{k-1}$, $k \geq 1$). Or one can consider the even simpler algorithm defined in the same way but with the map $S' : x \to 2x \bmod 1$ of $[0, 1]$ into itself which is very similar to the tent map and, like the latter, has also continuous spectrum with respect to the distribution dx and it is among the simplest maps with continuous spectrum. Hence one might expect that the algorithms generate rather random sequences. But of course, and on the contrary, the maps S generate, from an arbitrary seed s_0, a sequence $x_0 = \sigma_0$ and $x_k = Sx_{k-1}$, $k \geq 1$ which after 16 iterations becomes identically $x_k = 0$! because any seed σ_0 will be represented by just 16 digits, the others being implicitly 0.

The above observations show that it becomes necessary to attempt a qualitative and quantitative formulation not only of what we mean by "probability distribution" of random numbers generated by an algorithm P, but also in which sense the numbers generated by P can be really considered as "random". The following definition puts the matter into a quantitative form

3 Definition *(approximate random number generator):*
A sequence of N *random numbers in* $Q = [0, 1]^{\ell-1}$*, generated as described by an algorithm* $P = (P_0, s_0)$*, has a probability distribution* μ *absolutely continuous with some density* ρ*, within a precision* ε *and with respect to a given n-ple of test functions* $f_1, \ldots, f_n$ *defined on* Q*, if*

$$|\frac{1}{N}\sum_{i=1}^{N} f_k(x_i) - \int_Q \rho(x)\, f_k(x)\, dx| < \varepsilon \quad \text{for each} \quad k = 1, \ldots, n \qquad (5.2.3)$$

where $\mu(dx) \stackrel{def}{=} \rho(x)dx$ is an absolutely continuous probability distribution. The quantities $(P, f_1, \ldots, f_n; \rho, \varepsilon, N)$ will be called an "approximate discrete model of random distribution" with density ρ on the cube Q; or they will also be called an "approximate random numbers generator in Q". The generator is characterized by the algorithm $P = (P_0, \sigma_0)$, by the precision ε, by the control functions $(f_1, \ldots, f_n)$, by the density ρ of the distribution, by the statistical size N.

It will happen that the "goodness" of an algorithm $P = (P_0, s_0)$ depends on the seed s_0 and on the complexity and number of functions $(f_1, \ldots, f_n)$, although algorithms that exhibit such dependence in a too sensible way for N not too large cannot (obviously) be considered as good random numbers generators. Like the trivial example mentioned in the remark (iv) to definition 3 in which all sequences generated by any seed and tested on any family of test functions show poor quality for all sizes N, at least when the algorithm is (naively) implemented on a computer.

4 Definition *(non absolutely continuous random data):*
A sequence of N random numbers in $Q = [0, 1]^{\ell-1}$, generated by an algorithm P, has a probability distribution μ on Q, within a precision ε and with respect to a given n–ple of functions $f_1, \ldots, f_n$ defined on Q, if

$$\Big|\frac{1}{N}\sum_{i=1}^{N} f_k(x_i) - \int_Q \mu(dx)\, f_k(x)\Big| < \varepsilon \quad \text{for each} \quad k = 1, \ldots, n \qquad (5.2.4)$$

where $\mu(dx)$ is a probability distribution. The quantities $(P, f_1, \ldots, f_n; \mu, \varepsilon, N)$ yield a "discrete approximate" model of the distribution μ on Q, which is also referred to as a random numbers generator in Q with distribution μ. The generator is characterized by the precision ε, by the control functions $(f_1, \ldots, f_n)$, by the distribution μ and by the statistical size N.

Remarks:

(i) The choice of Q as a unit cube is not restrictive because given a random number generator on Q one can build another one with values in an open set U, contained in a piecewise regular surface M and which is the image of Q under a piecewise regular map F: one simply sets $u = F(x)$ for $x \in Q$. If F is regular (*e.g.* invertible and with non zero Jacobian determinant) then absolutely continuous distributions retain this property, *i.e.* they are described by a density function with respect to the area measure defined

on U. If, instead, F is regular but the distribution μ is not absolutely continuous then also the one generated on U will not be.
(ii) Thinking of generating random numbers on the basis of the time marked by a clock when one decides to look at it, or by opening phone books, or other of the kind is an *uncontrolled, not reproducible, subjective* procedure, hence unscientific (and also ugly). It has to be avoided as it is always better to know what is being done.

If one could build infinite sequences then one could eliminate from the definition of random generator the intrinsically approximate nature of it (due to the finite number n of control functions, to the finiteness of the statistical size N and to the positivity of accuracy ε).

This being *impossible* perfect random generators do not exist. It is nevertheless often useful to consider abstractly perfect generators:

5 Definition *(ideal random number generator):*
We shall define "ideal random number generator with distribution μ in Q with respect to a family $\mathcal{F}$ of functions on Q" an algorithm (P_0, s_0) which out of a seed s_0 chosen in a non empty set $\Sigma \subset [0,1]$ of "admissible seeds" produces a sequence $(x_1, x_2, \ldots)$ such that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} f(x_i) = \int_Q f(x) \quad \text{for each} \quad f \in \mathcal{F} \tag{5.2.5}$$

i.e. (5.2.4) "holds exactly" with $\varepsilon = 0$ in the limit $N \to \infty$.

Given a random generator and a dynamical system, *e.g.* a fluid model, "choosing a random initial datum" means generating x_1; choosing successively k data means letting the generator run producing $x_1, x_2, \ldots, x_k$.

For (*metaphysical*) reasons usually one considers random generators that produce distributions absolutely continuous (within an approximation judged convenient) with respect to the volume measure dx on phase space, *i.e.* distributions of the form $\rho(x)\,dx$ where $\rho(x) \ge 0$ is a smooth function. But one should not think that it is not equally easy to build generators (ideal or real) that produce distributions which are not absolutely continuous.

As a simple example consider an ideal generator (P_0, σ_0) that produces sequences of points x in $[0,1]$ that have the distribution $\mu(dx) = dx$ ("generator of the Lebesgue measure"). Let $x_1, x_2, \ldots$ be the sequence of numbers $x \in [0,1]$ produced by the generator (P_0, σ_0): write each of them in base 2 but interpret the result as the digital representation in base 3 of a sequence of numbers $x'_1, x'_2, \ldots$ (hence a sequence of numbers whose base 3 representation is such that the digit 2 never appears). This new sequence identifies a probability distribution μ' on $[0,1]$ which gives probability 1 to the Cantor set C_3, *i.e.* to the set of reals in $[0,1]$ which when written in base 3 do not have the digit 2. This Cantor set has zero length (*i.e.* zero Lebesgue measure) hence it has zero probability with respect to the distribution $\mu(dx) = dx$, or with respect to any other distribution absolutely continuous with respect to it.

Hence we see how distributions singular with respect to the Lebesgue measure can be easy to generate and therefore we see that their study can be as significant.

Problems.

[5.2.1]: Check that the coordinate change $y = 2\pi^{-1}\arcsin\sqrt{x}$ transforms the tent map S defined by $y \to 2y$ if $y < 1/2$ and $y \to 2(1-y)$ if $y \ge 1/2$, of $[0,1]$ into itself, into the map $\tilde S$ of $[0,1]$ into itself, defined by $x \to 4x(1-x)$ and the probability distribution dy into $\mu(dx) = dx/\pi\sqrt{x(1-x)}$.

[5.2.2]: Check that the probability distribution $\mu'(dy) = dy$ on $[0,1]$ is invariant with respect to the map S of [5.1.1], in the sense that $\mu'(S^{-1}E) = \mu'(E)$ in spite of the fact the the map evidently multiplies by 2 the lengths of (most) infinitesimal intervals. Deduce that the distribution $\mu(dx)$ in [5.2.1] is invariant with respect to the action of the map $x \to 4x(1-x)$. (*Idea:* S is not invertible, see footnote [2]).

For the discussion of the abstract properties of dynamical systems it is useful to keep in mind Birkhoff's theorem, also called additive ergodic theorem, that we quote here deferring its simple proof to [5.4.2]

Theorem: Let (A,S) be a dynamical system and let μ be an invariant probability distribution (i.e. such that $\mu(S^{-1}E) = \mu(E)$ for every measurable set E). Let $f \in L_1(\mu)$ then, with μ–probability 1 on the choices of $u \in A$, the average value $\lim_{T\to\infty} T^{-1}\sum_{k=0}^{T-1} f(S^k u) = \overline{f}(u)$ exists, and $\overline{f}(Su) \equiv f(u)$ μ–almost everywhere.

[5.2.3]: Consider the map S of the torus T^2 into itself described in the example (3) of the text (*Arnold's cat*) via the matrix $C = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$: $S\underline{\varphi} = C\underline{\varphi} \bmod 2\pi$ and check that if f is a regular function on T^2, and $f_{\underline{n}}$ is its Fourier transform, then

$$\Omega(k) \equiv \lim_{N\to\infty} \frac{1}{2N}\sum_{h=0}^{N-1} f(C^{k+h}\underline{\varphi}) f(C^h\underline{\varphi}) = \frac{1}{2}\sum_{\underline{n}} f_{\underline{n}} f_{C^{-k}\underline{n}} \qquad (a)$$

for *almost all* $\underline{\varphi} \in T^2$ (with respect to the (invariant) area measure). (*Idea:* Note that

$$\frac{1}{2N}\sum_{h=0}^{N-1} f(C^{k+h}\underline{\varphi}) f(C^h\underline{\varphi}) = \sum_{\underline{n}}\sum_{\underline{n}'} f_{\underline{n}} f_{\underline{n}'} \frac{1}{2N}\sum_{h=0}^{N-1} e^{i(C^h(C^k\underline{n}+\underline{n}')\cdot\underline{\varphi})}$$

and furthermore if $\langle\cdot\rangle$ denotes the average over T^2 (*i.e.* the integral over $\mu(d\underline{\varphi}) = (2\pi)^{-2}d\underline{\varphi}$) then

$$\lim_{N\to\infty}\langle\frac{1}{N}\sum_{h=0}^{N-1} e^{iC^h\underline{n}\cdot\underline{\varphi}}\rangle = 0, \qquad \lim_{N\to\infty}\langle\frac{1}{N^2}|\sum_{h=0}^{N-1} e^{iC^h\underline{n}\cdot\underline{\varphi}}|^2\rangle = 0 \qquad (b)$$

if $\underline{n} \ne \underline{0}$. Combine this with the Birkhoff theorem just stated and deduce that for each regular function g the limit for $N \to \infty$ of $N^{-1}\sum_{h=0}^{N-1} g(S^h\underline{\psi}) = \langle g\rangle$ for almost all $\underline{\psi}$ ("*ergodicity of Arnold's cat*"). (*Idea:* The second relation in (b) holds because after integrating over $\underline{\varphi}$ only N of the N^2 terms in the square of the sum give a non zero contribution. By Birkhoff's theorem the average in (a) exists μ–almost everywhere and the average $\overline{F}(\underline{\varphi})$ of the function

$$F(\underline{\varphi})=\frac{1}{2}\,\Big(f(C^k\underline{\varphi})\,f(\underline{\varphi})-\sum_{\underline{n}}f_{C^{-k}\underline{n}}\,f_{\underline{n}}\Big) \tag{c}$$

exists and, by the second of (b), it is such that the μ–integral $\langle\overline{F}^2\rangle$ vanishes; hence $\overline{F}(\underline{\varphi})=0$ too, μ–almost everywhere and this proves (a).)

[5.2.4]: In the context of [5.2.3] check that the function $\Omega(k)$ is rapidly decreasing as $k\to\infty$ if f has zero average and is differentiable at least three times. Make use of the property that the eigenvectors of the matrix $\begin{pmatrix}1&1\\1&2\end{pmatrix}$ are $\underline{v}_\pm=(\frac{1}{2}(-1\mp\sqrt{5}),1)\gamma_\pm$ (where $\gamma_\pm$ is the normalization constant), with eigenvalues $\lambda_\pm=\frac{1}{2}(3\pm\sqrt{5})$. (*Idea:* Since the ratio between the two components of $\underline{v}_\pm$ is a quadratic irrational, the Diophantine property holds, (*c.f.r.* (5.1.5)): $|n_1v_{\pm,1}+n_2v_{\pm,2}|\ge D/(|n_1|+|n_2|)$ for some $D>0$. Hence $|C^k\underline{n}|\ge\lambda_+^k|\underline{n}\cdot\underline{v}_+|\ge D\lambda_+^k/(|n_1|+|n_2|)$. Hence we have $|f_{\underline{n}}|\le\frac{F}{|\underline{n}|^3}$ (because f is C^3) and $|f_{C^k\underline{n}}|\le\frac{|\underline{n}|^3F}{(\lambda_+^kD)^3}$, and also $|f_{C^k\underline{n}}|<F'$ if F' is the maximum of $|f|$.
It follows that $|f_{\underline{n}}f_{C^k\underline{n}}|\le\frac{F}{|\underline{n}|^3}F'^{1-\varepsilon}(|\underline{n}|^3FD^{-3}\lambda_+^{-3k})^\varepsilon)$ for any $\varepsilon\in(0,1)$; finally the series $\sum|\underline{n}|^{-3+3\varepsilon}$ converges for ε small and $\Omega(k)\le cost\,\lambda_+^{-3\varepsilon k}$).

[5.2.5]: The result of [5.2.4] holds also if we only suppose that f is of class C^2. (*Idea:* Note that the set of the $\underline{n}$ for which $(C^k\underline{n})\cdot\underline{v}_+$ can be small is in reality "one dimensional" consisting only of the vectors $\underline{n}$ close to the line through the origin and orthogonal to $\underline{v}_+$.)

[5.2.6]: Consider the dynamical system in [5.2.3] and show that the definition 1 of continuous spectrum of a function $f\in C^\infty(T^2)$ and the definition based on (5.2.2) coincide if the limit as $N\to\infty$ of (5.2.2) is intended in the sense of distributions, *c.f.r.* §1.6. (*Idea:* Let $g(p)=e^{-ipk_0}$ be a test function and, calling $A_N(p)$ the function in (5.2.2), evaluate $\int_0^{2\pi}g(p)A_N(p)\frac{dp}{2\pi}$. It will turn out

$$\frac{1}{N}\sum_{h=0}^{N-1}\sum_{h'=0}^{N-1}f(S^h\underline{\psi})f(S^{h'}\underline{\psi})\delta_{h'=h-k_0}=\frac{1}{N}\sum_{\substack{h=0\\0\le h-k_0\le N-1}}^{N-1}f(S^h\underline{\psi})f(S^{h-k_0}\underline{\psi})$$

that has the same limit as $N^{-1}\sum_{h=0}^{N-1}f(S^h\underline{\psi})f(S^{h-k_0}\underline{\psi})$. By Birkhoff's theorem the latter limit exists *almost everywhere* and by the hint to [5.2.3] (*i.e.* by the ergodicity of Arnold's cat map), equals

$$\int\frac{d\underline{\psi}'}{(2\pi)^2}\,f(\underline{\psi}')f(S^{-k_0}\underline{\psi}')\stackrel{def}{=}\Omega(k_0)=\sum_{\underline{\nu}}f_{\underline{\nu}}\,\overline{f_{S^{k_0}\underline{\nu}}}=|f_{\underline{0}}|^2+\Omega_0(k_0)$$

and $\Omega_0(k_0)$ tends to zero rapidly as $k_0\to\infty$. Hence the Fourier transform of $\Omega(k)$ is the sum of a Dirac delta on the origin and of a C^∞ function: the spectrum of every non constant regular observable over almost all motions is therefore continuous.)

[5.2.7]: Consider the map $S:x\to 2x \bmod 1$ and the correspondence $I:x\to\underline{\sigma}(x)$ that to each point of $[0,1]$ associates the sequence $\underline{\sigma}$ of its binary digits $0,1$. Given n digits $\overline{\sigma}_1,\ldots,\overline{\sigma}_n$ show that the set $I(\overline{\sigma}_1,\ldots,\overline{\sigma}_n)$ of the points x whose first n binary digits are $\overline{\sigma}_1,\ldots,\overline{\sigma}_n$ is an interval of length 2^{-n}. The interval $I(\overline{\sigma}_1,\ldots,\overline{\sigma}_n)$ is called a *dyadic interval.*

[5.2.8]: Check that the map S of [5.2.7]acts on the dyadic intervals in [5.2.7] so that $S^{-h}I(\overline{\sigma}_1,\ldots,\overline{\sigma}_n)\cap I(\tilde\sigma_1,\ldots,\tilde\sigma_m)$ has length 2^{-m-n} if $h>m$. (*Idea:* Note that S

acts on x to generate a number that in binary representation has the same digits of x translated by one unit to the left, after erasing the first digit).

[5.2.9]: Consider the family $\mathcal{F}$ of functions that are piecewise constant on a finite number of dyadic intervals (an example is the characteristic function of any dyadic interval, *c.f.r.* [5.2.7]). Show that if $f \in \mathcal{F}$

$$\Omega(k) = \lim_{N\to\infty} \int_0^1 \frac{1}{N}\sum_{h=0}^{N} f(S^{h+k}x) f(S^k x)\, dx = (\int_0^1 f(x)\, d\,x\,)^2$$

if k is large enough, *i.e.* $\Omega(k) - \Omega(\infty) \equiv 0$ for all k_0 large enough. Thus all functions in $\mathcal{F}$ have continuous spectrum on motions generated by almost all data x. (*Idea:* Consider first the case in which f is a characteristic function of a dyadic interval and explicitly compute the integral.)

[5.2.10]: Adapt the above analysis to the case of the "tent map" in [5.2.1].

[5.2.11]: Proceeding as in [5.2.3] deduce from the result of the preceding problems, [5.2.9] and [5.2.10], that almost all initial data x for the dynamical systems (M, S) of problems [5.2.7] and [5.2.1] have continuous spectrum. (*Idea:* Just apply [5.2.9] and definition 3 in §5.1.)

[5.2.12]: Show that the "Arnold cat map" is a map, *c.f.r.* [5.2.3], of $[0, 2\pi]^2$ with a dense set of periodic points, all of which unstable. (*Idea:* The points with coordinates that are rational multiples of 2π are periodic points.)

[5.2.13]: Show that the map $\underline{\varphi} \to \underline{\varphi}' = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \underline{\varphi}$ of the torus T^2 into itself shares the "same" properties exhibited by the "Arnold's cat" in [5.2.1], [5.2.2], [5.2.3]. (*Idea:* This map could be called the "square root of the cat".)

Bibliography: [AA68], [Ga81].

§5.3 Dynamical systems types. Statistics on attracting sets.

We can now introduce the notion of *statistics of an attracting set A* for an evolution S defined on a phase space M.

In consideration of the already noted possible non regularity of S, when S is a transformation associated with timed observations of solutions of a differential equation (even if regular), we shall not suppose that S is regular, but only that it is "piecewise regular". This means that we shall suppose that M is a piecewise regular surface closed and bounded (one also says "compact") of dimension $\ell - 1$ and that inside each piece (*i.e.* inside every regular portion of M) S is C^∞.

We shall use the geometric objects that we call "*portions of regular surfaces*", or "*piecewise regular surfaces*" or "*differentiable surfaces* or *manifolds*" M in an intuitive sense, but of course a precise definition is possible.[1]

If S is a map of M in itself we shall say that S is piecewise regular on M if M can be realized as a union of portions of regular surfaces $M'_1 \cup \ldots \cup M'_s$ with only boundary points in common and such that the restriction of S to the interior of M'_j is of class C^∞, with Jacobian determinant $\det \partial S$ not zero, and furthermore it is extendible by continuity (together with its derivatives) to the whole M'_j (without necessarily coinciding with S also on the boundary points).[2] The set N of the boundary points of the portions M'_j will be called the set of the *singularity points* of S.

The value of S on the singular points is in a sense arbitrary: but it is convenient to imagine that S *is defined without exceptions*.

On a piecewise regular surface M it is therefore possible to define the "*volume measure*" which with every measurable set E, $E \subset M$, associates its $(\ell-1)$–dimensional surface area $\sigma(E)$.[3] This measure plays a particular role only because we assumed that M is piecewise regular, thus giving a special

---

[1] A *piece* or *portion* of regular surface of dimension n immersed in R^m $(m \ge n)$ is the image of a region $D \subset R^n$ that is bounded and simply connected and that is the closure of its interior points D^0 (*i.e.* $D = \overline{D^0}$). The image is realized by an analytic map with an analytic inverse, defined in a neighborhood of D. Moreover one supposes (inductively) that the boundary of D is as well a union of a finite number of portions of regular surface of one less dimension with, possibly, only boundary points in common: so that if a point is by definition a regular surface of dimension 0 we have set up a recursive definition. If $m = n$ the protion of surface will be called a *domain*.
A piecewise regular n–dimensional surface M, immersed in R^m, is a closed set that can be realized as a union of a finite number of portions of regular surface of dimension n having in common at most boundary points and the intersections between the portions are again a piecewise regular surface.
A regular surface M can be represented as a union of regular portions in several ways: if for every point $x \in M$ there is a representation of the surface as union of portions of regular surface one of which contains x *in its interior* then we say that the surface is a "*differentiable surface*".

[2] The regularity in analytic class for the portions of regular surfaces of M and in class C^∞ for the portions of regularity of S could be changed, with obvious modifications to the definitions just set up, and adjusted to $C^{(k)}$ and $C^{(h)}$ regularity respectively, with $k \ge h$: we shall occasionally refer to such less smooth cases, but for our purposes the above notions will usually suffice.

[3] We recall that Borel sets are defined as the sets of the smallest class of sets containing the open sets and which are closed with respect to denumerable combinations of operations of set union and set complementation, *c.f.r.* also the remark to (3.4.1) in §3.4. The Borel sets are a convenient class of sets that are measurable with respect to "*all*" measures in the sense that we only consider measures, *i.e.* countably additive functions, defined at least over the Borel sets. Given such a measure μ one then calls μ–measurable also the sets that differ from a Borel set by a set N with null external measure with respect to μ: *i.e.* by a set N that, although possibly being not borelian, is contained into the sets of a sequence of Borel sets whose measure tends to 0. Such sets are occasionally called "*μ–measurable* mod 0".
A function f with values in R is called μ–measurable if there is a set N with null external measure such that $f^{-1}(E)/N$ is μ–measurable mod 0 for all Borel sets $E \subset R$.
Finally: the difference between measurable and non measurable sets is by no means negligible: it *cannot be ignored* unless one has at least enough self control to avoid using the sinister axiom of choice: see problems [5.3.7], [5.3.8].

role to the metric structure induced on M by the Euclidean metric that we imagine defined on phase space. The latter is in turn privileged because we suppose that in the ℓ–dimensional phase space motion is described by regular differential equations. An hypothesis that is *not* justified other than by our metaphysical conceptions on the possibility of describing natural phenomena with "regular" equations and objects of elementary geometry.[4]

Let A be an attracting set for the map S and let $U \supset A$ be an open set in the global attraction basin of A. It is convenient to fix a precise definition of attracting set (so far used in an intuitive sense, because it hardly needs a definition) because it is a notion that it is natural to set up differently in different contexts and that we shall use in a precise sense from now on so that not defining it would lead to ambiguities.

1 Definition*(attracting sets and general dynamical systems):*
(1) A dynamical system (M, S) is defined by a piecewise regular surface M, and by a piecewise regular map S of M into itself; if S is invertible and (M, S^{-1}) is a dynamical system then (M, S) will be called an invertible dynamical system. The points x such that $S^k x$ is not a singularity point for all $k \ge 0$ form the set of regular points; if (M, S) is invertible then regular points are those for which $S^k x$ is not singular for $k \in (-\infty, +\infty)$.
(2) Given a dynamical system (M, S) we say that a closed invariant set A (i.e. $A \supseteq SA$) is an attracting set if there exists a neighborhood $U \supset A$ whose points $x \in U$ evolve so that their distance to A tends to zero: $d(S^n x, A) \xrightarrow[n\to+\infty]{} 0$.
(3) The union of all neighborhoods U of A that have this property is the global basin of attraction of A, and each of them is a "basin of attraction".
(4) An attracting set is "minimal" if it does not contain properly other attracting sets.

The above notion of dynamical system extends the one that we have been using so far in this chapter, because it allows singularities, even discontinuities, on the phase space M and/or on the evolution map S. When singularites are absent, *both on M and in S*, the system is called a "*differentiable dynamical system*", while if just discontinuities in S are absent the system is a very special case of a "*topological dynamical system*". Formally:

2 Definition: *(topological and differentiable dynamical systems):*
(1) A "topological dynamical system" (M, S) consists of a metric compact phase space M and of a continuous map S of M into itself. If S has a continuous inverse then also (M, S^{-1}) is a topological dynamical system

[4] It should not surprise that such a conception of the world, acceptable or, in fact, accepted without further discussion by most physicists, can appear unnatural or even repellent to many mathematicians, who often treat the arguments of interest here without giving a privileged role to the probability distributions that are absolutely continuous on phase space. Nothing can be done to convince someone giving to the volume on phase space a privileged role: this is not a scientific matter but it has a purely metaphysical nature, hence it has no general interest.

and (M, S) *is said "invertible".*
(2) If M *is a compact differentiable surface and* S *is a differentiable map of* M *into itself then* (M, S) *is a "differentiable dynamical system"; if* S *is invertible with a differentiable inverse also* (M, S^{-1}) *is a differentiable dynamical system and we say that* (M, S) *is invertible or also that* S *is a diffeomorphism of* M*.*

Imagine selecting randomly an initial datum u in the basin of attraction U of an attracting set A. By this we mean that we consider an ideal generator P of random numbers producing the initial datum with a probability distribution μ_0 on U, *c.f.r.* §5.2, definition 3.
Given an observable F defined on phase space consider the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} F(S^k u) = \langle F\rangle(u) \tag{5.3.1}$$

where u is a random initial datum, *i.e.* an element of the sequence $u_1, u_2, \ldots$ generated in U by the random number generator.
In general the average (5.3.1) depends on the point u used as initial datum (although we saw in §5.2 that there are simple cases in which it does not depend on it, aside of a set of zero volume); hence if (M, S) is a dynamical system and A an attracting set with basin U we set

3 Definition *(statistics of a random motion):*
Consider a sequence of randomly chosen initial data $u_1, u_2, \ldots$ *with a probability distribution* μ_0 *and consider an observable* F *and its time averages* $F_j = \langle F\rangle(u_j)$ *over the motions starting at* u_j, $j = 1, 2, \ldots$.
(1) Suppose that the F_j *are "essentially independent of the data* u_j*", i.e. that there exists a value* m_F *such that* $F_j = m_F$*, except for a number of values of the labels* j *that has zero density.*[5]
(2) Assume also that this happens for all observables of some prefixed large class $\mathcal{F}$ *(by large we mean that any continuous function can be approximated uniformly by sequences of functions in* $\mathcal{F}$*; e.g. all continuous functions, for instance).*
Then we shall say that the distribution μ_0 *produced by the random data generator* P *has a well defined statistics with respect to* S *and* $\mathcal{F}$*.*

The number m_F can then be written as an integral

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} F(S^k u) = \int_A F(v)\mu(dv) \quad \text{for } F \in \mathcal{F} \tag{5.3.2}$$

where μ is a probability distribution on M supported by A (*i.e.* $\mu(A) = 1$).[6]

[5] The density is defined naturally as the limit as $K \to \infty$ of the number of $j \le K$ that have the property in question divided by K itself.

[6] The correspondence $F \to \langle F\rangle = m_F$ is linear and continuous with respect to the uniform

The distribution μ_0 is not (generally) an invariant probability distribution and it must not be confused with the distribution μ: which is invariant, see remark (ii) below.

Remarks:

(i) The distribution μ will be called the *statistics* associated with the random generator P on the attracting set A or, equivalently, with the choice of the initial data with distribution μ_0. Such statistics (*if it exists*) is, with μ_0–probability 1, independent of the choice of the initial datum.
(ii) An important property of μ is that $\mu(E) = \mu(S^{-1}E)$, for each $E \subset A$ that is μ-measurable: this is the "*S–invariance*" of the statistics μ.
(iii) If P is only an approximate generator of random data the definition can be naturally extended by adding, wherever necessary, the attribute of "approximate".
(iv) If μ_0 itself is an invariant distribution, in general, the limit (5.3.1) exists *for all* functions F that are μ_0–integrable and for μ_0–almost all choices of u: this is Birkhoff's theorem, *c.f.r.* problems of §5.2 and §5.4. Nevertheless the value of the limit can be a nontrivial function (*i.e.* really depending on u and not μ_0–almost everywhere constant). So that μ_0 will not turn out to be the statistics of μ_0–almost all motions, and there will be several possible statistics for such motions.
(v) One could ask whether the invariance property of μ_0 is really necessary for the purposes of the remark (iv) which uses the invariance because it relies on Birkhoff's theorem which has it as a key assumption, except perhaps to exclude cases of only mathematical interest.
This is asking whether it can be said under very general grounds that μ_0–almost all random initial data are such that the limit (5.3.1) exists, even accepting that it might be a nontrivial function of u. However this is *not true* in simple and interesting cases, so that it is necessary to be careful if assuming *a priori* the above property. An example can be exhibited in which μ_0 even has a density with respect to the volume measure and yet the limit (5.3.1) does not exist for many functions F and for u in an open set of initial data: the matter is discussed in problem [5.3.6].
(vi) The limit (5.3.2) could fail to exist either because there is dependence on the initial datum u or because it really does not exist for some F and for many initial data u.

It will be useful to formalize a further variation of the preceding notions of dynamical systems: we present it below, after listing a few more concepts

---

convergence of sequences of observables; it is positivity preserving (*i.e.* $\langle F\rangle \geq 0$ if $F \geq 0$) and $\langle 1\rangle = 1$: this is one of the possible definitions of the notion of probability distribution.
The measure of an open set E can be defined as the supremum of the values of $\mu(F)$ as F varies among the continuous nonnegative functions, ≤ 1, vanishing outside E; this measure can be extended to the closed sets E setting $\mu(E) = 1 - \mu(M/E)$ and to all *Borel–sets*, *c.f.r.* the remark to (3.4.1) in §3.4 and footnote [3] above.

at the cost of appearing pedantic.

(a) A map S of a separable metric space M on which a Borel measure μ is defined, *c.f.r.* footnote [3], is said to be *μ–measurable* mod 0 if there exists a set of zero μ–measure N out of which S is everywhere defined and with values out of N, and furthermore for every Borel set $E \subset M/N$ the set $S^{-1}E$ is borelian (see footnote [3]).

(b) A μ–measurable mod 0 map is "*invertible* mod 0" if the set of zero μ–measure N can be so chosen that S is invertible on M/N and its inverse S^{-1} is μ–measurable mod 0.

(c) If S is a μ–measurable map and invertible mod 0 then such are S^n for $n \in Z$ integer; and these maps form a μ–measurable *group* of maps of M. If, instead, S is only μ–measurable but not invertible the maps $(S^n)_{n\in Z_+}$ form a *semigroup μ–measurable* mod 0.

(d) Likewise let $(S_t)_{t\in R}$ be a family of μ–measurable mod 0 maps, with $S_0 = 1$ and suppose that the set N of zero measure out of which the S_t are defined can be chosen t–independent for each *denumerable* family of values of t (but possibly dependent on the considered family). If, furthermore, outside a set $N_{t,t'}$ of zero μ–measure it is $S_tS_{t'} = S_{t+t'}$ for $t, t' \in R$ then $(S_t)_{t\in R}$ is a "*μ–measurable flow*" on M, denoted (M, S_t, μ). If instead this happens only for the maps of the family $\{S_t\}$ with $t \geq 0$ then $(S_t)_{t\geq 0}$ is a "*μ–measurable semiflow*". A flow or a semiflow will also be called a continuous dynamical system (M, S_t, μ).

In the following we shall often consider flows or groups of maps (or semiflows or semigroups of maps) which are μ–measurable mod 0 with respect to some measure μ. Therefore the following definition is useful

4 Definition *(metric dynamical systems, discrete and continuous):*
(1) a triple (M, S, μ) consisting of a closed set M, a probability distribution μ on the Borel sets of M and of a μ–measurable μ–invertible mod 0 *map S of M which leaves μ invariant*[7] *will be called a* discrete and bilateral metric dynamical system *(if instead S is not invertible then it will be called* unilateral*).*
(2) A triple (M, S_t, μ) consisting of a closed set M, a probability distribution μ on the Borel sets of M and of a group or semigroup $t \to S_t$ of maps of M, μ–measurable mod 0*, with respect to which μ is invariant,*[7] *will be called a* continuous metric dynamical system*, or a* continuous metric flow *(or* semiflow*).*

Remarks:

(i) By timing the observations of a continuous metric dynamical systems at constant time intervals we obtain a discrete metric dynamical system.

(ii) If the timing intervals are positive and not constant but they are timed by an event $\mathcal{P}$ we also obtain a discrete metric dynamical system to which a

[7] This means that $\mu(S^{-1}E) = \mu(E)$ for each measurable set E.

natural invariant measure ν is associated. Call in fact $\tau(x)$ the time interval between a timed observation of x (which therefore enjoys the property $\mathcal{P}$) and the successive observation that will take place at $Sx = S_{\tau(x)}x$. We see that the points y of phase space can be identified by giving (x, τ) if $\tau \in [0, \tau(x)]$ is the time that elapses between the realization of the event $x \in \mathcal{P}$ that in its motion precedes y and the moment in which the motion reaches $y = S_\tau x$.[8] Then the invariant measure for the continuous system can be written in the form $\mu(dy) = \nu(dx)\, d\tau$, where ν is a measure on the space of the points x enjoying property $\mathcal{P}$, and one realizes that ν is an invariant measure on the space of the timing events.

Hence if P is an ideal random generator and μ is the statistics of the motions generated from the sequence of random initial data produced by P and by a map S, the triple (M, S, μ) is an example of a "*metric dynamical system*", *c.f.r.* §5.2, definition 5. From the theory of dynamical systems we can immediately import a few qualitative well studied notions among which the notions of "*ergodicity*", "*mixing*", "*continuous spectrum*", "*isomorphism with a Bernoulli shift*" and others like that of "*positive entropy*" (*c.f.r.* §5.6). Correspondingly the generator P of random data and its statistics, when existent, will be called "ergodic", "mixing", "with continuous spectrum", "Bernoulli", "with positive entropy", *etc.* For instance, if (M, S, μ) is a metric dynamical system we set

5 Definition: *(ergodicity)*
Given a metric dynamical system (M, S, μ) let F be a bounded measurable constant of motion, i.e. let F be a μ–measurable (see footnote[3]) bounded function on M such that $F(u) = F(Su)$ outside of a zero μ–probability set of values of u. If such an F is necessarily constant (μalmost everywhere) then the invariant distribution μ is called ergodic *or* indecomposable.

Remarks:
(i) By the general Birkhoff theorem (*c.f.r.* problems [5.2.2] of §5.2, and [5.4.2] of §5.4) we deduce that given a metric dynamical system (M, S, μ) the asymptotic average $\overline{F}(u) \equiv \lim_{T\to\infty} T^{-1} \sum_{k=0}^{T-1} F(S^k u)$ exists "with μ-probability 1" (*i.e.* for $u \notin N_F$ with $\mu(N_F) = 0$). If μ is ergodic then, since the average $\overline{F}$ is obviously a constant of motion, it follows that $\overline{F}$ is independent of u and, in this case,

$$\overline{F}(u) \equiv \lim_{T\to\infty} T^{-1} \sum_{k=0}^{T-1} F(S^k u) = \int_M \mu(dv) F(v) \tag{5.3.3}$$

for μ–almost all u ("μ–almost everywhere"): the second relation in (5.3.3) is obtained from the first by integrating both sides with respect to μ and using that μ is by assumption an invariant distribution and, *moreover*, that

[8] Suppose here that all data y follow in their motion a timing event x and that are followed by another such event $x' = Sx$.

$\overline{F}(u)$ *does not* depend on u, *c.f.r.* problem [5.1.44].
(ii) Viceversa if for each F in L_1 (or in a set of functions dense in L_1) the limit relation (5.3.3) holds almost everywhere (*i.e.* outside a set N_F of μ–probability 0) then there are no nontrivial constants of motion and the statistics μ is ergodic.
(iii) If P is a random data generator with S–invariant distribution μ and if (M, S, μ) is ergodic then (5.3.3) can be read also as follows: "μ–almost all initial data admit a statistics and the statistics is precisely μ itself".
(iv) The requirement that F be measurable is *not* a subtlety, see problems [5.3.7], [5.3.8] and the bibliographic comment.
(v) The problems of §5.1 and those of this section give a few important examples of ergodic metric dynamical systems.

6 Definition *(mixing):*
Let (M, S, μ) be a metric dynamical system and suppose that for each pair F, G of continuous observables it is

$$C_{F,G}(k) \equiv \int_M \mu(du)\, F(S^k u) G(u) \xrightarrow[k\to\infty]{} \Big(\int_M \mu(du)\, F(u) \Big) \Big(\int_M \mu(du)\, G(u) \Big) \tag{5.3.4}$$

then the distribution μ is said "mixing". The function $C_{F,G}(k)$ is called the correlation function between F and G in the dynamical system (M, S, μ).

Remarks:
(i) One checks that, given (M, S, μ), if the statistics μ is mixing then (M, S, μ) is also ergodic. Indeed the density in L_1 of piecewise regular functions implies the validity of (5.3.4) for all bounded functions $F, G \in L_1$. Consider the function $\overline{F}(u)$ defined by the limit (5.3.3): it is such that $\overline{F}(u) \equiv \overline{F}(Su)$ and one can apply (5.3.4) to $F = G = \overline{F}(u)$ finding: $C_{\overline{F},\overline{F}}(k) \equiv \int_M \mu(du)\, \overline{F}(u)^2 = \big(\int_M \mu(du)\, \overline{F}(u) \big)^2$. Hence $\overline{F}(u)$ is constant μ–almost everywhere and (M, S, μ) is ergodic.
(ii) An attracting set that admits a mixing statistics is called "*chaotic*": this is just one among many alternative definitions of chaos; we have met some and we shall meet others. Usually various definitions of "chaos" are not mathematically equivalent, even though in practice one never encounters systems that are chaotic in a sense and not in another, other than for trivial reasons like, for instance, the case of a system composed by two separate chaotic systems (which, in fact, does not even have a dense orbit).
(iii) A question that seems intereseting is whether eq. (5.3.4) is implied by its validity for $F = G$, see also problem [5.3.9].

The following is an abstract generalization of a definition already discussed in §5.1, §5.2

7 Definition *(continuous spectrum for a metric dynamical system):*
A metric dynamical system (M, S, μ) has continuous spectrum if there exists a family $\mathcal{F}$ dense in $L_2(\mu)$ and for each $F \in \mathcal{F}$ the Fourier transform of the

function $C_{F,F}(k)$ in (5.3.4) has the form

$$\hat{C}_{F,F}(\omega) = (\int F\,d\mu)^2 \delta(\omega) + \Gamma_F(\omega), \qquad \omega \in [-\pi, \pi] \tag{5.3.5}$$

where $\delta(\omega)$ is Dirac's delta function and $\Gamma_F(\omega)$ is summable with respect to the Lebesgue measure $d\omega$ on $[-\pi, \pi]$.

Remarks:
(i) It can be checked that if a metric dynamical system has continuous spectrum then it is mixing, see problem [5.3.9].
(ii) One could also say, imitating definition 1 of §5.2 or definition 3 of §5.1 that the system has continuous spectrum if μ–almost all data generate a motion over which all observables $F \in \mathcal{F}$ have continuous spectrum: this can be checked to be equivalent to the above definition.
(iii) In all examples (that I can think of) the family $\mathcal{F}$ can be chosen $L_2(\mu)$. Usually the notion of continuous spectrum is given with $\mathcal{F} = L_2(\mu)$.

For the time being we skip the definitions of isomorphy and of systems with positive entropy, *c.f.r.* §5.6.

In practice one can try to determine the statistics of an attracting set by experimentally computing the averages $\langle F \rangle$ of the most important or simplest observables F. This can be done by using a sampling of the initial data according to an approximate (*c.f.r.* §5.2) statistics μ_0: a correct analysis of the results should, however, be always accompanied by a description of the chosen random generator.

We conclude by quoting an interesting result that follows from general theorems of ergodic theory: the possible statistics μ, associated with the various distributions μ_0 corresponding to different random generators, can be expressed in terms of the *ergodic statistics*, *i.e.* of the statistics μ_i such that the dynamical system (M, S, μ_i) is ergodic, at least if such statistics form a finite or denumerable set of probability distributions. One finds

$$\mu(E) = \sum_i \alpha_i \mu_i(E), \qquad \alpha_i \ge 0, \quad \sum_i \alpha_i = 1 \tag{5.3.6}$$

for all Borel sets E and this decomposition is *unique*. It is called *baricentric* or *ergodic decomposition* for obvious reasons and the uniqueness induces to say that the set of all statistics is a *simplex*[9] and that the ergodic statistics are its extremal points.

Often, however, the ergodic statistics form a *non denumerable* set: in such cases and quite generally (*e.g.* if M is a separable compact set and S is continuous) a formula like (5.3.6) holds with "the sum replaced by an integral" in a sense that we shall not discuss here, *c.f.r.* (5.6.7).

[9] A convex geometric figure in R^n is called a *simplex* if every interior point can be represented as the center of mass of a unique distribution of masses located at the extremal points of the figure: hence a segment is a simplex in R^1, a triangle is such in R^2, a tetrahedron in R^3 *etc.*

We shall not discuss the proof; not because it is difficult, *c.f.r.* [Ga81], but because this result will have here only a conceptual and philosophical relevance.

The (5.3.6) (or its analogue in integral form in the non denumerable case) says that essentially the only possible, pairwise distinct, statistics are the ergodic ones, since all others can be interpreted as a *statistical mixture* of them.

Problems.

[5.3.1]: *(cats)* Consider the dynamical system (M, S), called "Arnold cat map", *c.f.r.* problem [5.2.3], or its "square root", *c.f.r.* problem [5.2.12], and check that M itself is a minimal attracting set. (*Idea:* Attractivity is clear since M is the whole phase space. Minimality follows, for instance, from the property that the average of every smooth observable is almost everywhere equal to the integral over the torus M, *c.f.r.* [5.2.6], hence almost all points have dense trajectories.)

[5.3.2]: Show that the statistics of the choice of initial data with distribution $\mu_0(d\underline{\psi}) = d\psi_1 d\psi_2/(2\pi)^2$ in the dynamical system of [5.3.1] is $\mu = \mu_0$. (*Idea:* The probability distribution μ_0 is invariant and ergodic, *c.f.r.* problem [5.2.4].)

[5.3.3]: *(ergodic cats)* Show that the "Arnold cat", *c.f.r.* [5.3.1], is such that *not all* points have statistics $\mu(d\underline{\psi}) = d\psi_1\, d\psi_2/(2\pi)^2$ although almost all, with respect to the area distribution, do. (*Idea:* For instance $\underline{\psi} = \underline{0}$ has a different statistics (which?) and so happens for any point with rational coordinates, *c.f.r.* problem [5.2.11]).

[5.3.4]: *(recurrence and all that)* Consider point mass on a square Q of side $2L$ with periodic boundary conditions. At the center O of Q is centered a C^∞ circularly symmetric decreasing potential $v(\underline{x}) \geq 0$ and different from 0 only in a neighborhood U_ε of radius $\varepsilon \in (0, L/2)$. The potential is assumed to have nonzero derivative away from O wherever it is not zero. The phase space is the surface $\dot{\underline{x}}^2/2m + v(\underline{x}) = E$ with $E > 0$ fixed. The points of phase space can be described by the two Cartesian coordinates of $\underline{x}$ and by the angle φ formed by the velocity with the axis of the abscissae. The motion of an initial datum $(\underline{x}, \varphi)$ will be denoted $(\underline{x}', \varphi') = S_t(\underline{x}, \varphi)$ and it defines a Hamiltonian dynamical system that conserves the Liouville measure μ (which up to a constant factor is $d\underline{x} d\varphi$ if $\underline{x}$ is *out* of the circle U_ε of radius ε, where the potential is not zero).
Consider observations timed by the event "the point *enters* the circle U_ε". The "collisions" (soft) with U_ε can be characterized by the parameter $s \in [0, 2\pi\varepsilon]$ giving the abscissa on the circle at the entrance point and the angle of incidence $\vartheta \in [\frac{\pi}{2}, \frac{3\pi}{2}]$ formed between the external normal to the circumference in the collision point and the direction of entrance.
(a) If (s', ϑ') denotes the collision following a given collision (s, ϑ) show that the map $S(s, \vartheta) = (s', \vartheta')$ is well defined for *all* data (s, ϑ) (*i.e.* for all motions that *start with a collision*).
(b) Show that there exist data that do not undergo any collision and check that they have zero measure.
(c) Show that the map S conserves the measure $\nu(ds\, d\vartheta) = -\cos\vartheta\, ds\, d\vartheta$; and justify why it is given the name of "Liouville measure", showing that it can be built starting from the Liouville measure on the continuous phase space as described in the remark following definition 3.
(d) Finally check that the transformation S is singular in the points (s, ϑ) followed by a tangent collision.

[5.3.5]: *(recurrence again)* Imagine now, in the context of the preceding problem, that on the square Q a closed curve γ is drawn and that it does not intersect the circle U_ε, and to fix ideas assume that this curve is $x_1 = a = const$, $(2\varepsilon < |a| < L/2)$. Let Γ be a band around γ. Assume that in the region Γ a conservative force acts that attracts towards γ (for instance with a smooth potential $w(x_1, x_2) = f(x_1)$ decreasing from the boundary of Γ towards γ and with a quadratic minimum on γ). Finally suppose that inside the region Γ the motion of the point is also affected by a friction that opposes the motion with a force $-\lambda(x_1)\dot{x}_1$, with $\lambda(x_1) \ge 0$, that is not zero only in Γ and near γ has a constant value $\lambda > 0$. Show that almost all initial data evolve tending asymptotically to a periodic motion that develops on γ. Hence the above is an example of a system in which observations timed to the collisions with the circle U_ε are not very significant because such collisions are not (all) recurrent. Show that nevertheless there is a non empty set of zero measure in phase space consisting of points for which the collision with U_ε is recurrent.

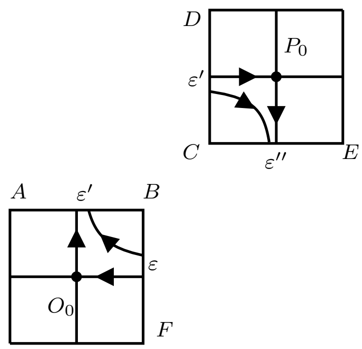


Fig. (5.3.1) *Illustration of the construction in* [5.3.6].

[5.3.6]: *(an example of an open set of points without statistics)* Consider a differential equation on the plane R^2 that in the neighborhoods Q_{O_0}, Q_{P_0} of side 2ℓ of $O_0 = (0,0)$ and $P_0 = (x_0, y_0)$ (with $x_0, y_0 > 4\ell$)) has, respectively, the form

$$\begin{cases} \dot{x} = -\lambda_1 x \\ \dot{y} = \lambda_2 y \end{cases} \qquad \begin{cases} \dot{x} = -\lambda_1 (x - x_0) \\ \dot{y} = \lambda_2 (y - y_0) \end{cases}$$

with $\lambda_1 > \lambda_2 > 0$. Imagine continuing the vector field $(f_1(x,y), f_2(x,y)$ defining the right hand side of the equation so that, considering the Fig. (5.3.1), the points of the side AB evolve in a time τ into the corresponding of side CD without expansion nor contraction, and the side CE evolves in BF in the same way. Show that if F is a function that assumes value a in Q_O and $b \ne a$ in Q_{P_0} then the average value of F on a trajectory that starts on BF with $y > 0$ is $\sim (a\lambda_2 + b\lambda_1)/(\lambda_1 + \lambda_2)$ if evaluated at the times when the point crosses BF while it has value $\sim a\lambda_1 + b\lambda_2/(\lambda_1 + \lambda_2)$ if evaluated at the times when the point crosses CD.
Deduce that, therefore, all initial data $u = (x, y)$ in Q_0 with $y > 0$ are such that the limit (5.3.1) does not exist for the considered functions F.
(*Idea:* The example, adapted from Bowen, is constructed so that the times T_j, T'_j defined below can be easily computed exactly. If motion starts on BF, for instance, at $y = \varepsilon_0 > 0$ the time that elapsed at the n–th arrival on BF is $T_0 + T'_0 + \ldots + T_{n-1} + T'_{n-1} + (\tau + \tau')(n-1)$ if T_k is the time necessary to go from BF to AB after visiting k times BF and T'_k is the time ecessary to go from DC to BF, and τ is the time spent going from AB to DC, τ' that from CE to BF, which are independent from ε_0.
The time $T_0 + T_1 + \ldots + T_{n-1}$ is spent within Q_{0_0} and $T'_0 + \ldots + T'_{n-1}$ in Q_{P_0}. Likewise at the n–th arrival on CD the elapsed times are $T_0 + \ldots + T_{n-1}$ in Q_0 and $T'_0 + \ldots + T'_{n-2}$ in Q_{P_0}. Since $T_k = (\lambda_1/\lambda_2)^{2k} \lambda_2^{-1} \log \varepsilon_0^{-1}$, $T'_k = (\lambda_1/\lambda_2)^{2k+1} \lambda_2^{-1} \log \varepsilon_0^{-1}$, the average at arrival in BF is asymptotically as $n \to \infty$ $(a + b\,\lambda_1/\lambda_2)/(1 + \lambda_1/\lambda_2)$ while if computed upon arrival on DC it is, instead, $(a + b\,\lambda_2/\lambda_1)/(1 + \lambda_2/\lambda_1)$).

[5.3.7]: *(non measurable sets and functions)* Consider the map $Sx = x + r \bmod 1$ of $M = [0,1]$ into itself and the metric dynamical system (M, S, μ) where $\mu(dx) = dx$. Take r to be irrational. Consider the set of all the distinct trajectories ξ that are generated by the dynamics S. Out of each trajectory ξ we choose a point $x(\xi)$ and consider the set E

of the values of $x(\xi)$ as ξ varies among the trajectories. The set E can be regarded as a set of distinct labels for the trajectories ξ.
(a) show that the set E is not measurable with respect to the measure dx (*i.e.* E it is neither a Borel set nor it differs from such a set by a set of zero external measure, *c.f.r.* footnote [3]).
(b) define the function $F_E(x)$ that associates with x the label of the trajectory in which x lies. Show that this function is a constant of motion but it is not measurable with respect to the measure dx.
(c) Check that in the construction of E (hence of F_E) the sinister axiom of choice has been used.
(*Idea:* The sets $E_n = E + n\,r \bmod 1$, $n \in Z$, are pairwise disjoint and $\cup_{n=-\infty}^{\infty} E_n = [0,1] = M$: hence if E was measurable it would have to have measure > 0 otherwise the measure of M would be zero; but if it has > 0 measure then $M = [0,1]$ would have infinite measure, because the measure dx is S–invariant, while it has measure 1).

[5.3.8]: Check that the analysis in [5.3.7] can be *verbatim* applied to any ergodic system (M, S, μ) with μ which gives 0 measure to individual points and with M containing a continuum of points. In other words labeling each trajectory by one of its points in an ergodic system defines a set E of labels which is not measurable, except in trivial cases.

[5.3.9]: (*continuous spectrum implies mixing*) Prove that a continuous spectrum dynamical system in the sense of definition 7 is mixing (in the sense of definition 6). (*Idea:* take the Fourier tansform of equation (5.3.5) and check that it implies (5.3.4) for $F \equiv G$ by the elementary Lebesgue theorem on Fourier transforms. The case $F \neq G$ is not elementary and relies on the spectral theory of unitary operators, *c.f.r.* [RS72].)

Bibliography: [AA68], [Ga81]. Without using the sinister axiom of choice it is impossible to construct an example of a non measurable set in $[0,1]$ (say): so one might reject the axiom of choice and take the (equally sinister) axiom that all sets in $[0,1]$, or on a manifold, are measurable with respect to the Lebesgue measure. It would be wiser to reject both (no bad nor noticeable consequences would ensue): an example of the damage that the axiom of choice can do is its use in the very influencial review [EE11], see notes #98 and #99, p. 90, to reject the ergodic hypothesis of Boltzmann, for a critique see [Ga99a], Ch. 1.9.

§5.4 Dynamical bases and Lyapunov exponents.

On an attracting set A for an evolution S there can be a large number of invariant ergodic probability distributions. This is in particular true for *strange attracting sets* (see §4.2) that can also be defined simply as attracting sets capable of supporting a large number of ergodic statistics.

We have seen however the privileged role plaid (for reasons of "cultural tradition") by initial data randomly selected, within the basin of an attracting set, with a probability distribution μ_0 admitting a density with respect to the volume measure.

In principle such random choices could fail to define a statistics or they could lead, as well as the more general ones, to a nonergodic statistics, see problems [5.3.6], [5.4.28] for an example. However usually the statistics associated with such μ_0 is *also* ergodic and independent on the particular density function that is used to select initial data (assuming that the random choice is ideal in the sense of §5.2, *c.f.r.* the discussion of (5.2.5)).

Therefore it is convenient to formalize the following definition

1 Definition: *(SRB statistics, normal attracting sets)*
Suppose that initial data of a motion for the dynamical system (M, S)*, c.f.r. §5.3 definitions 1,2, are chosen in the basin of an attracting set* A *with distribution* μ_0 *absolutely continuous with respect to volume.*
If data so chosen admit, with μ_0*–probability* 1*, a statistics independent of the data themselves then we say that the attracting set* A *is* normal *and that it admits a natural statistics* μ *to which we give the name of SRB* statistics.

Remarks:

(i) Given the importance of random choices of initial data with a probability distribution with density with respect to the volume measure (occasionally called a distribution "*absolutely continuous with respect to the Lebesgue measure*", or somewhat ambiguously just "*absolutely continuous*"), it is of great interest to find cases in which such distribution can be effectively studied.
(ii) However the importance attributed to data choices with absolutely continuous distributions is quite unsatisfactory: hence it is *also* interesting to study families of statistical distributions on A associated with random choices of initial data with distributions *different* from those absolutely continuous continuous with respect to the volume.
(iii) SRB are initials for Sinai, Ruelle, Bowen.

There is (only) one general family of dynamical systems (M, S) for which one can study wide classes of invariant probability distributions and characterize among them the *SRB* distribution. These are the *hyperbolic systems*: which are, in a certain sense, also the most intrinsically "chaotic" systems.

It is convenient to dedicate this section to a general discussion of notions and properties related to hyperbolicity and to delay to the coming §5.5 the discussion of the specific properties of the SRB distributions.

2 Definition *(hyperbolic, Anosov and axiom A attracting sets):*
Given a d*–dimensional differentiable dynamical system* (M, S) *with* M *connected, let* X *be an attracting set for* S *which we suppose with an inverse* S^{-1} *differentiable in the vicinity of* X*, and satisfying*
(1) (hyperbolicity) *For each* $x \in X$ *there is a decomposition of the plane* T_x *tangent to* M *as* $T_x = R^s(x) \oplus R^i(x)$ *in two transversal planes (i.e. linearly independent and with dimensions* $d_s > 0$ *and* $d_i > 0$ *with* $d_s + d_i = d$*) which vary continuously with respect to* $x \in X$*, and such that there exists constants* $\lambda > 0, C > 0$ *for which, for all* $n \ge 0$

$$\begin{aligned} |S^n(x+dx) - S^n(x)| \equiv& |\partial S^n(x) \cdot dx| < C e^{-\lambda n} |dx|, \quad \text{if} \quad dx \in R^s(x) \\ |S^n(x+dx) - S^n(x)| \equiv& |\partial S^n(x) \cdot dx| > C e^{\lambda n} |dx|, \quad \text{if} \quad dx \in R^i(x) \end{aligned} \tag{5.4.1}$$

(2) (completeness of periodic motions) The points of the periodic orbits in X are a dense set in X.
Then if X is a regular connected surface we say that the attracting set "verifies the Anosov property". More generally X will just be a closed set and we shall say that the map S "verifies axiom A" on X.

Remarks: (i) The conditions (5.4.1) imply *covariance* of the decomposition $T_x = R^s(x) \oplus R^i(x)$, *i.e.* $R^\alpha(Sx) = \partial S_x R^\alpha(x)$ for $\alpha = u, s$.
(ii) These systems are particularly interesting because their attracting set is normal, *i.e.* there is an SRB distribution on it, in the sense of definition 1, *c.f.r.* [Ru76]: but in reality systems with attracting sets verifying axiom A as just described are *very rare*, and their importance is, in the end, in being examples of systems with normal attracting sets, *i.e.* of systems in which a well defined statistics describes the asymptotic behavior of motions with data randomly chosen with a distribution absolutely continuous with respect to the volume measure.

In a dynamical system we distinguish the *wandering points* from the *non wandering* ones. The first are the points x with a neighborhood $\mathcal{N}$ such that $S^n\mathcal{N} \cap \mathcal{N} = \emptyset$ for all $|n| > 0$. The other points are nonwandering.
The qualification of wandering might be misleading as one could think that a point which never, in its future evolution, returns close to the initial position is wandering: this is *not* necessarily true because, although it might go away and never come back, it can still happen that, having arbitrarily prefixed a time, points initially close to it do come back close to it again after this time has elapsed. It is essential to keep this in mind to understand properly the notion of axiom A systems. Something unexpected from the literal meaning of the word "wandering" happens considering the familiar pendulum motion and taking as phase space the set of points with energy $\le E$ where E is larger than the separatrix energy: then all points on the separatrix wander away from their initial position; nevetheless it is easy to see that they are non wandering, see also [5.4.28].
A simple example relevant for the notion of wandering point is a Hamiltonian system: if M is an energy surface and S is a canonical map on it, then all points are non wandering as a consequence of Poincaré's recurrence theorem. The following definition is quite natural

3 Definition *(Axiom A and Anosov systems):*
(1) A differentiable dynamical system (M, S) on a connected phase space M is said to verify axiom A *if the set Ω of non wandering points is hyperbolic, i.e. it verifies property (1), and periodic points are dense, i.e. it verifies property (2) of definition 2 without being necessarily an attracting set, c.f.r.* [Sm67], p. 777.
(2) An Anosov system *is a differentiable dynamical system (M, S) with M connected and S a diffeomorphism which is hyperbolic in every point of M, i.e. verifies (1) of definition 2 on the whole M.*

Remarks (a list of related results):
(i) A simple example of a system verifying axiom A is described in the example (4) following the definition 1 of §5.2, illustrated in Fig. (5.2.1). Examples, less simple but important, can be found among the problems of §5.5.
(ii) If (M, S) verifies the axiom A then the set Ω of its nonwandering points may fail to contain points with dense orbits on Ω. However its nonwandering points can be divided into closed disjoint invariant sets $C_1, C_2, \ldots$, *finitely many* of them, on each of which there is a dense orbit: the latter sets are called *basic sets* (this is a theorem by Smale or *spectral decomposition theorem*), *c.f.r.* [Sm67]. One says that the set Ω of nonwandering points of an Axiom A system is the union of a finite number of "components" C on which the action of S is *topologically transitive*.
(iii) The set of nonwandering points will contain a basic set C that is an attracting set: *i.e. all* points of the phase space M close enough to C are such that $d(S^n x, C) \xrightarrow[n\to\infty]{} 0$. A minimal attracting set X in an axiom A system *is necessarily a basic set*. Hence any axiom A system contains at least one axiom A attracting set. And in fact every $x \in M$ evolves towards *only one* among the basic sets, *c.f.r.* [Ru89b], p. 169, (possibly not attracting).[1]
(iv) A notion similar to topological transitivity (*i.e. existence of a dense orbit*) is that of *topological mixing*: if C is an invariant closed set we say that the action of the transformation S is *topologically mixing on* C, if given any two sets U and V in C, relatively open in C, there is an integer $\overline{n}$ such that for $n > \overline{n}$ it is $S^n U \cap V \neq \emptyset$. Evidently topological mixing implies topological transitivity.
(v) Topological mixing is related to systems that verify axiom A because one proves that every basic set C_j (*c.f.r.* (3)) can be represented as a union of a finite number p_j, $C_j^1, C_j^2, \ldots, C_j^{p_j}$, of closed pairwise disjoint sets on each of which the map S^{p_j} is topologically mixing, [Ru89b] p. 157. Furthermore on each of them the stable and unstable manifolds are dense.
(vi) It is *not impossible* that in an Anosov system the manifold M itself is necessarily identical to its non wandering set Ω, *c.f.r.* [Ru89b] p. 171. This property is however not proven although, so far, no counterexamples are known. If an Anosov system admits a dense orbit then every point is nonwandering and $M = \Omega$; the same can be said if the system is topologically mixing. This implies that the periodic points are dense (see p. 760 in [Sm67]): hence transitive Anosov systems verify axiom A. It would clearly be particularly interesting that $M = \Omega$ in general. In fact the stable and unstable manifolds of the sets C_j^k of (v) above are dense in C_j^k, [Sm67], p. 783, so that if $M = \Omega$ there will be only one basic set (M itself) and the action of S on M will be transitive and mixing with stable and ustable manifolds dense in M. If $M = \Omega$ in general one could also say that Anosov

[1] Axiom A excludes the existence of basic sets C with points in their vicinity which, while evolving so that $d(S^n x, C) \xrightarrow[n\to\infty]{} 0$, get away from C by a distance $\delta > 0$ which can be chosen independently on how is y near to C, *c.f.r.* [Ru89b] p. 167; see the example in [5.4.28].

systems are smooth Axiom A attracting sets.
(vii) If an Anosov system (M, S) is such that the map S leaves invariant the volume measure μ then (M, S, μ) is ergodic, p. 759 in [Sm67], hence there is a dense orbit and $M = \Omega$. In this case the periodic points are dense and the stable and unstable manifolds of each point are dense.

It might seem that the structure of the basic sets will not change much if one varies by a little the map S. However in order for this to be true further conditions must be added: a sufficient condition is that the system enjoys the following property

4 Definition *(axiom B systems): Suppose that the invertible dynamical system (M, S) verifies axiom A and, calling Ω the nonwandering set,*
(1) Denote, for $x, x' \in \Omega$, W^i_x (respectively $W^s_{x'}$) the points z such that $d(S^n z, S^n x) \xrightarrow[n\to-\infty]{} 0$ (respectively $d(S^n z, S^n x') \xrightarrow[n\to+\infty]{} 0$) for $x, x' \in \Omega$: such sets are locally manifolds tangent to the unstable and stable tangent planes that in (1) of Definition 1 are denoted $R^i(x)$ and $R^s(x')$; they are called the global unstable manifold *of x and the* global stable manifold *of x' and they will be considered more formally below.*
(2) It can be shown that if (M, S) verifies axiom A, ([Ru89b], p. 169*), for each point $y \in M$ there is a pair of nonwandering points x, x' such that $y \in W^i_x \cap W^s_{x'}$: suppose that in y the surfaces $W^i_x, W^s_{x'}$ intersect transversally, i.e. the smallest plane that contains their tangent planes at y is the full tangent plane T_y in y.*
Then we shall say that the dynamical system (M, S) verifies axiom B*.*[2]

Remarks:
(i) The example quoted in comment (i) to definition 3 of a system that satisfies axiom A is also a simple example of a system satisfying axiom B.
(ii) If a system verifies axiom B and one perturbs S by a small enough amount one obtains a map S' which can be transformed back to S via a continuous change of coordinates, on M: this is a non trivial "*structural stability*" result that says that if S and S' are close enough in C^1 (*i.e.* if they are close together with their derivatives) then there exists a continuous map h of M into itself with continuous inverse (*i.e.* a "*homeomorphism*)" such that $hS = S'h$, *c.f.r.* [Ru89b] p. 170. *Note that h needs not be differentiable.*

The principal difficulty, concerning the generality, of the definitions just given is tied to the requests of regularity of S on A and of continuity of the derivatives of S on A. The attempts to get free from assigning a privileged role to absolutely continuous (with respect to the volume) distributions and, more generally, the attempts of not assigning a privileged role to a special

[2] Here we have taken the liberty of calling "*axiom B*" what is called in [Ru89b], p. 169, *strong transversality*: the original axiom B definition of Smale is *somewhat weaker*, *c.f.r.* [Sm67] p. 778. The notion used here seems to be the natural one if one takes into account the conjectures of Palis–Smale and the theorems by Robbin–Robinson and Mañé on the stability of such systems *c.f.r.* [Ru89b], p. 171.

system of coordinates have led to various results.

It is convenient to fix once and for all a "minimal " regularity requirement on S that permits us to expose various results that are collected under the name of *thermodynamical formalism*: the name motivation should become clear in the following.

The temptation is strong to demand simply that S is analytic (or C^∞) as a map on a phase space M, itself an analytic (or C^∞) ℓ–dimensional surface. However we have already mentioned that usually dynamical systems (M,S) are generated by timed observations on smooth dynamical flows. In such cases often M *will not result* an analytic surface, but a piecewise analytic surface with a boundary. Hence *it is not wise to require that S and M are subject to a global analyticity or smoothness requirement.*

The following notion of *regular dynamical system* with respect to a probability distribution μ on M provides us with a sufficient generality for our purposes (and it is difficult to imagine other cases in which it would not be sufficient).

5 Definition: *(regular singularities, regular metric dynamical systems): Let (M,S) be a dynamical system in the sense of the definition 1 of §5.3, and let μ be a probability distribution on M.*

(a) We shall say that (M,S) is μ–regular, with parameters $C,\gamma>0$, if

(1) The set N of the singularity points of S and, if (M,S) is invertible, the set N' of the singularities of S^{-1} are such that $\mu(U_\delta(N)),\mu(U_\delta(N'))<C\delta^\gamma$, where $U_\delta(X)$ denotes the set of points at a distance $<\delta$ from X.

The points $x\in M/(\cup_{j=-\infty}^{\infty}(S^{-j}N))$, or $x\in M/(\cup_{j=-\infty}^{\infty}(S^{-j}N\cup S^jN')$ if (M,S) is invertible, will be called "μ–regular points" for (M,S).

(2) If $\delta(x,y)$ is the minimum between the distances of x and y from N and if $d(x,y)$ is the distance between x and y in the metric of M, the derivative $\partial^{\underline{a}}S^{\pm1}(x)$ satisfies

$$|\partial^{\underline{a}}S^{\pm1}(x)-\partial^{\underline{a}}S^{\pm1}(y)|\le\underline{a}!\,D\,C^{|\underline{a}|}\frac{d(x,y)}{\delta(x,y)^{|\underline{a}|+1}}\qquad\text{if}\quad d(x,y)<\frac12\delta(x,y)\tag{5.4.2}$$

where $\underline{a}=(a_1,\ldots,a_\ell)$ are ℓ (not negative) differentiation labels, and $|\underline{a}|\equiv\sum_i a_i$, $\underline{a}!=\prod a_i!$; C,D are constants >0. And, if (M,S) is invertible, we require the analogous relation for S^{-1} (with N' in place of N).

(3) The Jacobian matrix $\partial S^{\pm1}(x)$ verifies[3]

$$||\partial S(x)\cdot(\partial S(y))^{-1}-1||<C\frac{d(x,y)}{\delta(x,y)}\tag{5.4.3}$$

and the analogous property is requested for S^{-1} if (M,S) is invertible. If the properties (1), (2), (3) above hold we say that the system has "μ–regular singularities".

---

[3] If T is a matrix mapping the tangent plane to M at the point x into the tangent plane to M at the point y, $||T||$ is the maximum of $|Tu|$ on the unit vectors u, where the length of the tangent vector u is measured in the metric of M at x while the length of Tu is measured in the metric at y.

(b) A triple (M, S, μ) with a μ–regular (M, S) and an S–invariant μ will be called a "μ–regular metric dynamical system".

Remarks:

(i) It is clear that if (M, S) is μ–regular then the singularity set N for S has zero μ–measure. Property (2) says that S is piecewise analytic and that its singularity on N is "fairly controlled" because the radius of convergence of its Taylor series tends to $\to 0$ at most as the distance from N.
(ii) If S is μ–regular and if μ is S–invariant, *i.e.* $\mu(E) = \mu(S^{-1}E)$, let $\delta_n = n^{-\gamma'}$ with $\gamma' > \gamma^{-1}$. Then $\sum_n \delta_n^\gamma < +\infty$, *i.e.* $\sum_n \mu(U_{\delta_n}(N)) < \infty$. But invariance of μ implies that the set Δ_n of the points x such that $S^n x \in U_{\delta_n}(N)$, *i.e.* $\Delta_n \equiv S^{-n} U_{\delta_n}(N)$, has the same measure of $U_{\delta_n}(N)$ and hence $\sum_n \mu(\Delta_n) < \infty$. Then (by the Borel–Cantelli theorem, *c.f.r.* philosophical problems [5.4.1] and [5.4.2]) μ–almost all points can only belong to a finite number of sets Δ_n: which means that, for μ–almost all x, a constant $C(x)$ exists such that the distance of $S^n x$ from N stays larger than $C(x)\delta_n$, for all n.
(iii) Hence the preceding remark stresses the main property of the maps S that enjoy μ–regularity: watching the evolution of μ–almost all points we see that they *do not get close* to the singularities faster than a power of the elapsed time.
(iv) By definition of dynamical system (M, S), *c.f.r.* definition 1 of §5.3, the singularities of S are on a finite number of piecewise regular surfaces of dimension inferior to ℓ, the full dimension of phase space. If μ is a measure absolutely continuous with respect to volume the property (1) is then satisfied. The other two demand a special case by case analysis.
(v) However a map S can be regular with respect the volume but not with respect to another (or more) invariant distribution μ, or viceversa.
(vi) In applications the hypotheses of regularity are often trivially verified. This happens, for instance, if S and M are analytic (without any singularities) and S has a non singular Jacobian matrix ∂S. However whenever such properties are not trivial the analysis of the regularity properties is invariably difficult and interesting.
(vii) The μ–regularity definition has been set up so that systems like "billiards" are μ–regular if μ is the volume measure on their energy surface, see problems [5.4.22], [5.4.23].

It is also convenient to examine explicitly cases in which S is *single valued*, or "*injective*", *i.e.* $Sx = Sy$ implies $x = y$. Or, more generally, "injective mod 0" with respect to a measure μ: *i.e.* S is μ–measurable mod 0 and, outside a set N of μ–measure zero, the relation $Sx = Sy$ implies $x = y$.

We suppose in the remaining part of this section that S is an injective map and, occasionally that S is also invertible: the difference is important only in the cases of discrete dynamical systems generated by irreversible equations, like the examples (4.1.20), (4.1.28), (4.1.30) which, although verifying the uniqueness property for the solutions do not allow (in general) their existence backward in time.[4]

We now discuss various qualitative notions necessary to the formulation of the simplest results whose collection usually sets the general frame of the qualitative theory of chaotic motions. Among these the notions of *Lyapunov exponents* and of "*dynamical bases*" are important.

6 Definition *(Lyapunov exponents and dynamical bases of a trajectory): If (M, S) is an invertible dynamical system and x is a regular point (c.f.r. definition 1, §5.3) we say that on the trajectory $k \to S^k x$ a* dynamical base *for S is defined and denoted $W_1(x), \ldots, W_n(x)$, if it is possible to define $n = n(x)$ linearly independent planes $W_1(x), \ldots, W_n(x)$, spanning the full tangent plane T_x, and such that*

$$\lim_{k\to\pm\infty} \frac{1}{k} \log \frac{|\partial S^k d\xi|}{|d\xi|} = \lambda_i(x) \qquad \text{if} \quad 0 \neq d\xi \in W_i(x) \tag{5.4.4}$$

where
(1) $\lambda_1 > \lambda_2 > \ldots > \lambda_n$ are called Lyapunov exponents *associated with the trajectory,*
(2) the integers $m_j = \dim(W_j)$, $j = 1, \ldots, n$ are the respective multiplicities.
(3) If no one of the exponents λ_j vanishes the point x is called "hyperbolic".
(4) If $\lambda_j > 0$ and $\lambda_{j+1} < 0$ the plane $V_s(x) \equiv W_{j+1}(x) \oplus \ldots W_n(x)$ is called the stable direction.
(5) The plane $\tilde{V}_i(x) = W_1(x) \oplus \ldots \oplus W_j(x)$, unstable direction, *is analogously defined as stable direction for S^{-1}.*
(6) More generally one defines the "characteristic planes for S" *the planes $V_j(x) = W_j(x) \oplus \ldots \oplus W_n(x)$ and those for S^{-1} as $\tilde{V}_j(x) = W_1(x) \oplus \ldots \oplus W_j(x)$.*
In the case of noninvertible systems the same definition makes sense, after some obvious modifications, if there is a sequence of nonsingular points, $k \to x_k$ with $k \in (-\infty, +\infty)$, such that $S x_{k-1} = x_k$ and $x_0 = x$.

Remarks:
(i) The $\lambda_i(x)$ and $m_i(x)$ *do not depend* from the particular point x chosen on one and the same trajectory.
(ii) The spaces $W_i(x)$ are "covariant" in the sense that $\partial S\, W_i(x) = W_i(Sx)$.
(iii) Existence of a dynamical base is not so rare as one might fear. This is illustrated by the following theorem (Oseledec, Pesin):

I Theorem *(existence of Lyapunov exponents): Let (M, S, μ) be an invertible regular metric dynamical system (in the sense of definition 5, (b), with*

[4] If M is a ball of large enough radius the evolution generated by these equations keeps data inside the ball M, as seen in §4.1: however if evolution is regarded backwards in time motions "leave" the ball, *i.e.* one cannot in general suppose that phase space is bounded.

M not necessarily containing a smaller attracting set A); then μ-almost every point x generates a trajectory whose points admit a dynamical base. Moreover the Lyapunov exponents are constants of motion and so are their multiplicities.

Remarks:
(i) Suppose that the distribution μ used for selecting random initial data on an invariant set $A \subseteq M$ (attracting or not) is ergodic. Then μ–almost all its points admit dynamical bases; the corresponding Lyapunov exponents and their multiplicities not only are constants of motion (a property which follows simply from the fact that they depend on the trajectories and not on the points on them) but they are *also constants on* A, with probability 1 with respect to μ.
(ii) Although this is implicit in what said above it is good to stress that the independence of the Lyapunov exponents $\lambda_i(x)$ from the points on an invariant set A holds with μ–probability 1 if x is chosen with a distribution μ which is *S–invariant* and *ergodic* on A. *This does not mean that there cannot be points x of A with different Lyapunov exponents.*
(iii) *Indeed* in general such points exist and they can form dense sets on A. Changing the distribution for the random choices from μ to μ', we shall in general obtain *new* values for the Lyapunov exponents, also constant with probability 1 with respect to the new distribution μ', if the latter is also ergodic.
(iv) Hence Lyapunov exponents *are not a purely dynamical property*, *i.e.* they are not a property of the map generating the dynamics only, but they must be regarded as a *joint* property of the map *and* of distribution μ for the random selection of the initial data.
If we wish, instead, to consider *all* initial data, allowing no exceptions, then the Lyapunov exponents will in general depend on the point, and there can even be points to which one cannot associate a dynamical base, nor any Lyapunov exponents (*i.e.* these are points x which move so irregularly that the dynamics S does not have well defined properties of expansion and contraction of the infinitesimal tangent vectors that S can be thought of as carrying along with x).
(v) Oseledec's theorem (see [5.4.11]) is also called *multiplicative ergodic theorem* and it generalizes the Birkhoff's ergodic theorem (*c.f.r.* problems of §5.2 and [5.4.2], [5.4.3]), that we can also call the *additive ergodic theorem.* What at first sight seems surprising is the generality under which the theorem holds. A generality that can be further extended because the regularity hypothesis of S on M can be weakened by only demanding little more than S be differentiable μ-almost everywhere (which is the minimum necessary to give a meaning to the action $d\xi \to \partial S(x)d\xi$ of S on the infinitesimal tangent vectors and, hence, to the statements of the theorem).
(vi) In fact the statement that brings to light this absolute generality is: "let (M, S, μ) be an ergodic dynamical system for which the action of S on

infinitesimal vectors is meaningful, *i.e.* $\partial S(x)$ is defined μ–almost everywhere, and such that $\int_M \log(1+||\partial S(x)||)\,\mu(dx) < +\infty$, then at μ–almost every point of M the map S has a well defined asymptotic action of expansion and contraction, with Lyapunov exponents well defined and, with μ-probability 1, independent of the initial data".)

(vii) So general a theorem (practically without assumptions) *must* have a simple proof. This does not mean that its proof did not require major effort: in fact it escaped to several people looking for it. This is somewhat similar to what happened in the case of the (pointwise) additive ergodic theorem whose proof escaped to the search by many mathematicians (among which Von Neumann). See the problems for a guide to the proof.

(viii) It is, however, also clear that the theorem, because of its generality, will have to be *nonconstructive* and therefore it will not be too useful, like Birkhoff's theorem: in applications the true problem is, given (M, S, μ), how to identify the μ–almost all, *but not all*, points that have well defined Lyapunov exponents and how to find their dynamical bases; or how to find the points which *do not* have well defined Lyapunov exponents (which, although exceptional, can exist). Such questions have great importance and usually generate problems which are very interesting both mathematically and physically and which must be studied on a case by case basis, see [5.4.27] for a still somewhat abstract but important example.

(ix) In many cases the map S is not invertible. Definition 4 can then be adapted to define the system of characteristic planes, *c.f.r.* definition 4, $V_1(x) \supset \ldots \supset V_{n(x)}(x)$ via the property

$$\lim_{k\to+\infty} \frac{1}{k} \log|\partial S^k(x)u| = \lambda_j(x) \qquad dx \in V_j(x)/V_{j+1}(x) \tag{5.4.5}$$

while the planes $\tilde V_j(x)$, see item (6) in definition 6, cannot evidently be defined (because every vector that expands with a given exponent can be altered, by adding to it a vector that dilates more slowly, obtaining a vector that still expands in the same way). Here we cannot exchange the role of "dilating" and that of "contracting" because we do not have the inverse map S^{-1}.

(x) Nevertheless even though S^{-1} is not defined (as in the mentioned case of the dynamics generated by the Navier–Stokes truncated equation, see footnote[4]) it can happen that the Jacobian matrix ∂S is invertible in each point of the trajectory of x and that S also is "invertible on the trajectory of x", *i.e.* there is a unique sequence x_h such that $x_0 = x$ and $Sx_h = x_{h+1}$, $h \in (-\infty, \infty)$ for μ–almost all x (this happens if S is injective or if it is invertible on the support of the distribution μ, *i.e.* on a closed set with μ–probability 1).

In this case it is still possible to define characteristic planes "for S^{-1}", $\tilde V_1 \supset \tilde V_2 \supset \ldots \supset \tilde V_n$ and also dynamical bases. The theorem given above extends to such situation: *i.e.* given (A, S, μ) and if S, S^{-1} are differentiable on A then μ–almost all points of A admit also a dynamical base for S^{-1}, [Ru78]. See the problems for a precise formulation.

(xi) Sometimes rather than specifying the multiplicity of the Lyapunov exponents of a metric dynamical system (A, S, μ) it is convenient to simply repeat them according to multiplicity. In this way a ℓ degrees of freedom system admits ℓ Lyapunov exponents $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_\ell$ on every trajectory with a dynamical base.
(xii) The ℓ Lyapunov exponents of a point x with a dynamical base can also be defined in terms of the action of the map S on infinitesimal surfaces tangent at x, rather than in terms of the action of S on just the infinitesimal vectors. It is not difficult to check, in the frame of the proof of theorem I (see problem [5.4.12]) that if x admits a dynamical base it is

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \log \frac{|\partial S^n d\xi_1|}{|d\xi_1|} = \lambda_1(x), \\
&\lim_{n\to\infty} \frac{1}{n} \log \frac{|\partial S^n d\xi_1 \wedge d\xi_2|}{|d\xi_1 \wedge d\xi_2|} = \lambda_1(x) + \lambda_2(x), \\
&\lim_{n\to\infty} \frac{1}{n} \log \frac{|\partial S^n d\xi_1 \wedge d\xi_2 \wedge d\xi_3|}{|d\xi_1 \wedge d\xi_2 \wedge d\xi_3|} = \lambda_1(x) + \lambda_2(x) + \lambda_3(x), \quad \ldots
\end{aligned} \tag{5.4.6}$$

where $d\xi_1 \wedge d\xi_2 \wedge \ldots$ denotes the infinitesimal surface element delimited by the infinitesimal vectors $d\xi_1, d\xi_2, \ldots$, and the limit relations in (5.4.6) *do not* hold for all choices of the infinitesimal vectors but for *almost all* choices of their orientations on the unit sphere of the tangent space. In fact consider, for instance, the case in which all exponents are pairwise distinct: then the first relation in (5.4.6) holds only if $d\xi_1 \not\in V_2$, the second if $d\xi_1 \not\in V_2$ and $d\xi_2 \not\in V_3$, *etc.*

Therefore given an S–invariant indecomposable distribution μ[5] on a set A, attracting with respect to a map S which is differentiable in the vicinity of A, we can associate with it the ℓ Lyapunov exponents, and consider the action of S on the surface elements of various dimensions.
(xiii) The equation (5.4.6) yields a method of wide use in numerical experiments for computing Lyapunov exponents. In fact one chooses randomly, with uniform distribution, on the unit sphere of tangent vectors, m independent tangent vectors $d\xi_1, d\xi_2 \ldots, d\xi_m$ and one measures the logarithm of asymptotic expansion of the surface element that they delimit: such expansion is $\lambda_1 + \ldots + \lambda_m$, with probability 1 with respect to the choices of $\xi_1, \ldots, \xi_m$, and of $x \in M$, *c.f.r.* [BGGS80].
(xiv) We should stress that if the initial data are chosen with an invariant ergodic distribution μ and (M, S) is μ–regular then, as it follows from the definitions, *c.f.r.* problems, the Lyapunov exponents for S are *opposed* to those of S^{-1}.
But this *does not mean* that "to compute the minimum Lyapunov exponent of a map S it suffices to compute the maximum one of the inverse map", not even in the simple case in which the system admits a unique attracting set.

[5] *i.e.* such that (A, S, μ) is ergodic, *c.f.r.* definition 5 in §5.3.

Indeed the map S in general will have an attracting set A *different* from the one, A_-, for S^{-1} (that we suppose also unique, for simplicity). If then the initial data are chosen randomly with the distribution μ and if we make even the smallest round–off error in the calculation on the initial datum, or on one of those into which it evolves, we obtain data x that are not exactly on the attracting set on which the distribution μ is concentrated and, hence, the trajectory of x under the action of S^{-1} *does not stay close to* A *but evolves towards* A_-*! and therefore its stability properties will have nothing to do with those of the exact motion of the initial datum.*

And even if $A_- = A$ the relation between Lyapunov exponents for S and S^{-1} will not always hold: because the statistics observed for S^{-1} with data randomly chosen with distribution μ will differ from that for S if round–off or computational errors are present. Such errors will likely imply that the choices of the initial data are performed with an effective distribution that is absolutely continuous: therefore, if the statistics SRB for S^{-1} is different from that for S, (*this is the usual case, see example in problem* [5.4.20]), the observed Lyapunov exponents will be different from the ones for S changed of sign: they will be, instead, those of the SRB distribution for S^{-1}.

The statistics of the motions for the two evolutions can be *completely different* and concentrated on very different sets: in the sense that the first may attribute probability 1 to a set to which the second attributes probability 0 and viceversa (*even though the attracting sets* A *and* A_- *may coincide* between themselves and with the full phase space M). Hence there will not be, in general, *any relations* between the Lyapunov exponents of the forward and of the backward motions.

(xv) Consider the very special but very interesting cases in which the system is *reversible*, *i.e.* in the cases in which there is a regular isometry I of M in itself and such that $IS = S^{-1}I$, *c.f.r.* §7.1. In this case the time reversal symmetry implies *equality* (and not "oppositeness") of the Lyapunov exponents of S and of S^{-1} for initial data randomly chosen with absolutely continuous distributions with respect to volume (if the attracting sets for S and S^{-1} are normal in the sense of the definition 1).

(xvi) The remark (xiv) is a manifestation of the phenomenon of the *irreversibility* of motions; and remark (xv) shows that there is *no direct relation between irreversibility of motion and time reversal symmetry* in spite of the fact that usually time reversal symmetry of the equations of motion is called (improperly) "reversibility", *c.f.r.* §7.1 for a further discussion.

(xvii) Finally one can prove (Pesin, [Pe76]), that if the $\lambda_j > 0$ and $\lambda_{j+1} < 0$ the planes $\tilde V^{(i)} = W_1 \oplus W_2 \oplus \cdots W_j$ are "integrable", *i.e.* for μ–almost all points x there is a vicinity of x and inside it a regular surface $\overline{\tilde V}^{(i)}$ consisting of points y such that

(a) $d(S^{-k}y, S^{-k}x) \le C(x)e^{-\lambda_j k}$, $k \ge 0$ for some $C(x) > 0$ and

(b) with $\tilde V^{(i)}(y)$ being the tangent plane to $\overline{\tilde V}^{(i)}$ at y for almost all x.

Likewise there is, in the same vicinity of x, a regular surface $\overline{V}^{(s)}$ which

(a) has the plane $V^{(s)}(y) = W_{j+1} \oplus W_2 \oplus \cdots W_n$ as tangent plane in y for

almost all x and
(b) consists of points y such that $d(S^k y, S^k x) \le C(x)e^{-\lambda_j k}$, $k \ge 0$.

Philosophical problems.

The following examples of abstract thought are (rightly) famous although it does not seem that they have applications other than that of permitting us a formulation, general and without too many exceptions and *distinguo*, of a formalism and of a conceptual framework for the qualitative theory of motion.

[5.4.1]: (*Borel–Cantelli theorem*) If μ is a Borel measure on R^n and if Δ_n is a sequence of measurable sets such that $\sum_n \mu(\Delta_n) < +\infty$ then μ–almost all points are contained in at most a finite number of sets of the sequence. (*Idea:* The set of points that are *not* in a finite number of sets Δ_n is $N = \cap_{k=1}^\infty(\cup_{h=k}^\infty \Delta_h)$: but $\mu(N) \le \sum_{h=k}^\infty \mu(\Delta_h)$ for each k hence $\mu(N) = 0$, because the series converges.)

[5.4.2]: In the case of the observation (ii) to definition 5 every point x outside all Δ_n except a finite number of them is such that $d(x, N) \ge n^{-\gamma'}$ for all n except a finite number of them. Let the U be the set of the points x such that $d(S^n x, N) > 0$: note that U has complement U' with zero measure and infer that for almost all points $x \in U$ there is a constant $C(x)$ for which $d(S^n x, N) > C(x)n^{-\gamma'}$. (*Idea:* Apply [5.4.1]).

[5.4.3: (*Garsia's maximal average theorem*) Let (M, S, μ) be an invertible dynamical system. Let f be a funxction (μ–measurable, of course) such that $|f(x)| < K$, μ–almost everywhere. Then if D_n is the set of points $x \in M$ for which some average of f over a time $\le n$ is not negative (*i.e.* $m^{-1}\sum_{j=0}^{m-1} f(S^j x) \ge 0$ for some $m \le n$, $1 \le m \le n$), then $\int_{D_n} f(x)\mu(dx) \ge 0$. (*Idea:* if $n = 1$ the condition defining D_1 is simply $f(x) \ge 0$ and there is nothing to prove. If $n = 2$ the condition defining D_2 is $f(x) \ge 0$ or $f(x) < 0$ and $f(x) + f(Sx) \ge 0$. Then the new points, *i.e.* points in D_2 outside D_1 are points x with $f(x) < 0$ but which can be paired with points Sx of D_1 so that the sum of the values of f on such pairs is ≥ 0. Hence we can subdivide D_2 into pairwise disjoint sets $(D_2/D_1 \cup S(D_2/D_1)) \cup D_1/S(D_2/D_1)$: on the third set $f(x) \ge 0$, and such is the integral of f over it; while the integral over the first two can be written *by the S–invariance of μ*, as $\int_{D_2/D_1}(f(x) + f(Sx))\mu(dx)$, which is tehrefore not negative. The case $n = 3$ is only slightly more involved and, after analyzing it, the general case becomes crystal clear).

[5.4.4]: (*Birkhoff's ergodic theorem*) Show that [5.4.3] implies immediately the validity, under the same assumptions of [5.4.3] of the following theorem, see [5.2.2]. Given f as in [5.4.3], μ–almost everywhere there is the limit: $\overline{f}(x) = \lim_{n\to\infty} n^{-1}\sum_{j=0}^{n-1} f(S^j x)$. (*Idea:* Let $f_{sup}(x) = \limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(S^j x)$ and likewise define $f_{inf}(x) = \liminf_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(S^j x)$; if $f_{sup}(x)$ and $f_{inf}(x)$ were not equal amost everywhere there would be two constants $\beta > \alpha$ such that the set D of the points where $f_{sup}(x) > \beta$ and $f_{inf}(x) < \alpha$ would be an invariant set (because $f_{sup}(x) = f_{sup}(Sx)$, μ–almost everywhere and so also f_{inf}) and the measure of D will be $\mu(D) > 0$. By definition of D the functions $f - \beta$ and $\alpha - f$ would have on D necessarily some *non negative* averages (of some finite order n): hence by the theorem in [5.4.1] they would have a non negative integral over D. Hence $\int_D(f - \beta)d\mu \ge 0$ and $\int_D(\alpha - f)d\mu \ge 0$: and summing these two inequalities one would conclude that $(\alpha - \beta)\mu(D) \ge 0$ which is impossible because $\alpha - \beta < 0$ and $\mu(D) > 0$).

[5.4.5]: Show that the boundedness assumption $|f(x)| < K$ in [5.4.4] is not necessary and it can be replaced by $f \in L_1(\mu)$. Furthermore

$$|\overline{f}|_{L_1} \equiv \int_M \mu(dx)\,|\overline{f}(x)| \le \int_M \mu(dx)|f(x)| \equiv |f|_{L_1}$$
$$\int_M \overline{f}(x)\mu(dx) = \int_M f(x)\,\mu(dx)$$

hence, since f is a constant of motion ($f(x) = f(Sx)$), if μ is ergodic it will be $\overline{f}(x) \equiv \int_M f(y)\mu(dy)$ μ–almost everywhere.

[5.4.6]: (*subadditive ergodic theorem of Kingman*)
Let (M, S, μ) be an ergodic dynamical system and $n \to f_n(x)$ a subadditive sequence of μ–measurable functions

$$|f_1(x)| < K, \qquad f_{n+m}(x) \le f_n(x) + f_m(S^n x) \qquad \mu\text{−almost everywhere}$$

for all $m, n > 0$. By applying the ergodic theorem [5.4.4],[5.4.5] show that the limit $\lim_{n\to\infty} \frac{1}{n} f_n(x) = \overline{f}(x)$ exists μ–almost everywhere.. (*Idea:* The functions

$$f_{sup}(x) = \limsup_{n\to\infty} n^{-1} f_n(x), \qquad f_{inf}(x) = \liminf_{n\to\infty} n^{-1} f_n(x)$$

are such that $f_{sup}(Sx) \ge f_{sup}(x)$ and $f_{inf}(Sx) \ge f_{inf}(x)$, because $f_n(x) \le f_1(x) + f_{n-1}(Sx)$. Therefore the invariance of μ implies that these functions are constants of motion, (indeed $\int (f_{sup}(Sx) - f_{sup}(x))d\mu = 0$). Then, by the assumed ergodicity, there exist two constants $\alpha < \beta$ such that $f_{sup}(x) = \beta$ and $f_{inf}(x) = \alpha$ μ—almost everywhere. Contemplate the case $\beta > \alpha$ and let $\eta > 0$ such that $\alpha + 2\eta < \beta$.
If we define Δ_n as the set of points x for which $f_m(x) \le (\alpha+\eta)m$ for at least one $0 < m \le n$ it is clear that $\mu(\Delta_n) \xrightarrow[n\to\infty]{} 1$. Hence given $\varepsilon > 0$ with $K\varepsilon < 1$ there exists n_ε such that $\mu(\Delta^c_{n_\varepsilon}) < \varepsilon$ (where a label c denotes the complementary set).
For μ–almost all $x \in M$ one can suppose that the frequency of visit of x to $\Delta^c_{n_\varepsilon}$ is smaller than ε, because such frequency is the average value (over j) of the function $\chi_{\Delta^c_{n_\varepsilon}}(S^j x)$, if χ_Δ denotes the characteristic function of Δ.
Hence if $j_1 < j_2 < \ldots$ is the sequence of times for which, instead, $S^{j_k}x \in \Delta^c_{n_\varepsilon}$ one deduces that the number p of the times $j_k \le T$ is such that $p/T < \varepsilon$ for T large enough (because $\lim_{T\to\infty} p/T = \mu(\Delta^c_{n_\varepsilon}) < \varepsilon$ by the ergodicity assumption)).
Let $k_0 \ge 0$ be the first time $\le T$ when $S^{k_0}x \in \Delta_{n_\varepsilon}$, and let k'_0 be the largest integer $\le T$ such that $f_{k'_0-k_0}(S^{k_0}x) \le (\alpha+\eta)(k'_0-k_0)$: there exists at least one which is $\le k_0+n_\varepsilon$; let $k_1 > k'_0, k_1 \le T$ be the first time successive to k'_0 and different from $j_1, j_2, \ldots$; and let then $k'_1 \le T$ be the largest successive time for which $f_{k'_1-k_1}(S^{k_1}x) \le (\alpha+\eta)(k'_1-k_1)$ and so on. We thus define a sequence of time intervals $[k_0, k'_0], [k_1, k'_1], \ldots, [k_r, k'_r] \subset [0, T]$. Each value k'_j must be one of the j_i's because, *e.g.* considering k'_0, the point $S^{(k'_0-k_0)}S^{k_0}x = S^{k'_0}x$ we see that it cannot be in Δ_{n_ε} otherwise there would be $0 < m \le n_\varepsilon$ such that $f_m(S^{k'_0}x) \le (\alpha+\eta)m$ hence $f_{k'_0+m-k_0}(S^{k_0}x) \le f_{k'_0-k_0}(S^{k_0}x) + f_m(S^{k'_0-k_0}S^{k_0}x) \le (\alpha+\eta)(k'_0-k_0+m)$ and k'_0 would not be maximal. The value of k'_r must be closer to T than a quantity $\le n_\varepsilon$. It is therefore clear that the points outside such intervals $[k_j, k_{j+1}]$ are at most $p + n_\varepsilon$. Hence

$$\frac{1}{T} f_T(x) \le \frac{1}{T}\Big((p+n_\varepsilon)\,K + (\alpha+\eta)T\Big) = K\frac{p+n_\varepsilon}{T} + (\alpha+\eta) \xrightarrow[T\to\infty]{} \alpha + 2\eta < \beta$$

which end the necessity of contemplating the case $\beta > \alpha$, by the manifest contradiction, and implies $\alpha = \beta$).

[5.4.7]: Show that, setting $\overline{f}(x) = \lim_{n\to\infty} n^{-1} f_n(x)$, [5.4.4] implies that $\overline{f}$ is constant μ–almost everywhere and μ–almost everywhere it is

$$\lim_{n\to\infty} n^{-1} f_n(x) = \lim_{n\to\infty} n^{-1} \int_M f_n(x)\mu(dx) = \inf_n n^{-1} \int_M f_n(x)\mu(dx)$$

(*Idea:* $n^{-1} f_n(x) \le n^{-1}(K + f_{n-1}(Sx))$ hence $\overline{f}(x) \le \overline{f}(Sx)$ therefore, analogously to [5.4.6], $\int(\overline{f}(x) - \overline{f}(Sx))\mu(dx) = 0$ implies $\overline{f}(x) = \overline{f}(Sx)$ μ–almost everywhere hence, by the assumed ergodicity, $\overline{f}$ is μ–almost everywhere constant. By the dominated convergence theorem we derive the first relation. On the other hand the function $\langle f_n\rangle = \int f_n(x)\,\mu(dx)$ is *subadditive* (*i.e.* $\langle f_{n+m}\rangle \le \langle f_n\rangle + \langle f_m\rangle$ and bounded by K and, by an elementary argument, $n^{-1}\langle f_n\rangle \xrightarrow[n\to\infty]{} \inf_n n^{-1}\langle f_n\rangle$).)

[5.4.8]: Show that the hypothesis $|f_1(x)| < K$ in [5.4.6],[5.4.7] is not necessary and it can be replaced by a summability hypothesis on the positive part of $f_1^+(x) \equiv \max(0, f_1(x))$ which demands $f_1^+ \in L_1(\mu)$. (*Idea:* Just examine carefully the proofs of [5.4.6],[5.4.7]).

[5.4.9]: Supponse S and (M, S, μ) μ–regular and apply the ergodic theorem to show that if $\mu(N) = 0$ and if $S^j N$ is μ–measurable for $j \ge 0$, also $\mu(SN) = 0$. (*Idea:* The frequency of visit $\varphi_x(SN)$ to SN by the trajectory starting in x is equal to that to the set N itself: $\varphi_x(SN) = \varphi_x(N)$; but by the ergodic theorem $\mu(SN) = \int \varphi_x(SN)d\mu \equiv \int \varphi_x(N)d\mu = \mu(N) = 0$).

[5.4.10]: *(equality of future and past averages for metric dynamical systems)* Given (M, S, μ) suppose that S is invertible (μ–almost everywhere). Define, via the ergodic theorem, the "future" average $f^+(x) = \lim n^{-1}\sum_{j=0}^{n-1} f(S^j x)$ and the "past average" $f^-(x) = \lim n^{-1}\sum_{j=0}^{n-1} f(S^{-j}(x))$. Show that $f^+(x) \equiv f^-(x)$ μ–almost everywhere. Likewise if $f_n(x)$ is bounded and subadditive $\lim_{n\to]io} n^{-1} f_n(x) = \lim_{n\to]io} n^{-1} f_n(S^{-n}x)$, μ–almost everywhere. (*Idea:* Let D, α, β be such that $\mu(D) > 0$ and $f^+(x) > \beta > \alpha > f^-(x)$ for $x \in D$. Let D_n^+ be the set of points $x \in D$ for which $\frac{1}{m}\sum_{j=0}^{m-1} f(S^j x) > \beta$ for all values $m \ge n$; and let D_n^- be the analogous set where $\frac{1}{m}\sum_{j=0}^{m-1} f(S^{-j}x) < \alpha$. Then if n is large enough $S^{-(n-1)}D_n^- \cap D_n^+ \ne \emptyset$, because $\mu(D_n^-) \equiv \mu(S^{-(n-1)}D_n^-)$ and $\mu(D_n^+)$ are both very close to $\mu(D)$ for n large, hence: $\alpha > n^{-1}\sum_{j=0}^{n-1} f(S^{-j}x) \equiv n^{-1}\sum_{j=0}^{n-1} f(S^j S^{-(n-1)}x) > \beta$; impossible).

Further problems, theorems of *(Oseledec, Raghunathan, Ruelle.)*

Given (M, S, μ) suppose in the following problems that S is μ–regular. It will make sense to set $T^n(x) \equiv \partial S^n(x)$, for $n \in (-\infty, +\infty)$ which, (*c.f.r.* comment (8) to definition 3), is a function defined μ–almost everywhere. We shall always suppose that (M, S, μ) is an ergodic system, and $M \subset R^\ell$.

[5.4.11]: *(Oseledec theorem)* Check that $T_n[x] = T(S^{n-1}x)\cdot\ldots\cdot T(Sx)\cdot T(x)$ deducing that if $(\log||T||)^+ \in L_1(\mu)$ then there exists the limit $\lambda_1(x) = \lim_{n\to\infty} n^{-1}\log||T(x)||$, μ–almost everywhere (the norm is defined by thinking that on R^ℓ is defined the euclidean scalar product and $||T||$ is the maximum of the length $||Tu||$ as u varies with $||u|| = 1$). (*Idea:* The function $f_n(x) = \log|T_n(x)|$ is subadditive in the sense of [5.4.6], see also [5.4.8]).

[5.4.12]: Imagine the vectors $u \in R^\ell$ as functions $i \to u_i$ on the finite space $L = \{1, 2, \ldots, \ell\}$. Let $(R^\ell]^{\wedge q}$ be the space of the antisymmetric functions on L^q (*i.e.* functions $u_{i_1\ldots i_q}$ antisymmetric in $i_1 \ldots i_q$). On this linear space one can define a natural scalar product and $((u, v) = \sum u_{i_1,\ldots,i_q} v_{i_1,\ldots,i_q}$, and therefore a length of the vectors in $(R^\ell)^{\wedge q}$). Define the matrix $T_n^{\wedge q}$ acting on $(R^\ell)^{\wedge q}$ as

$$(T_n^{\wedge q}(x)u)_{i_1\ldots i_q}=\sum_{j_1\ldots j_q}^{1,\ell}(T_n(x))_{i_1j_1}\ldots(T_n(x))_{i_nj_n}u_{j_1\ldots j_n}$$

and show that if $(\log||T_1||)^+\in L_1(\mu)$ the limit $\lambda_q(x)=\lim_{n\to\infty}n^{-1}\log||T_n^{\wedge q}(x)||$ exists μ–almost everywhere (the norm of an operator O on $(R^\ell)^{\wedge q}$ is defined via the above length notion). (*Idea:* This is implied by the ergodic subadditive theorem, as in the preceding case because $\log||T_n^{\wedge q}(x)||$ is also subadditive).

[5.4.13]: Consider a sequence of real operators (*i.e.* matrices with real matrix elements) T_j on R^ℓ. Let $T^n=T_n\cdot T_{n-1}\cdot\ldots\cdot T_1$ and assume the existence of the limits

$$\limsup_{n\to\infty}n^{-1}\log||T_n||\le0,\qquad\lim_{n\to\infty}n^{-1}\log||T^{n\wedge q}||=\lambda_q\qquad q=1,2,\ldots,\ell\qquad(*)$$

If $\Lambda_n\equiv(T^{n*}T^n)^{1/2n}$ let $t_n^{(1)}\ge\ldots\ge t_n^{(\ell)}$ be the ℓ eigenvalues of Λ_n. Show that the limits $\lim_{n\to\infty}n^{-1}\log t_n^{(i)}$ exist. (*Idea:* Note that $t_n^{(1)}t_n^{(2)}\ldots t_n^{(q)}\equiv||T^{n\,\wedge q}||$. The first hypothesis in (*) is not necessary, and is made only for later reference.)

[5.4.14]: In the context of [5.4.13] let $\mu^{(1)}>\ldots>\mu^{(s)}$ be the possible distinct values of the limits $\lim_{n\to\infty}n^{-1}\log t_n^{(i)}$: we call multiplicity m_i of $\mu^{(i)}$ the number of values of q for which $\lim\frac1k\log t_k^{(q)}=\mu^{(i)}$. Let $U_k^1,U_k^2,\ldots,U_k^s$ be the linear spaces spanned by the first m_1 eigenvectors of Λ_k, by the subsequent m_2 eigenvectors, …, by the last m_s eigenvectors: realize that it is natural to say that m_i is the asymptotic multiplicity of the eigenvalues μ_i. Define $r(p)=i$ if $\frac1k\log t_k^{(p)}\xrightarrow[k\to+\infty]{}\mu^{(i)}$.

[5.4.15]: *(orthogonality properties)* In the context of the preceding problem show that the notion of asymptotic multiplicity is even more justified since the following orthogonality property holds. Given $\delta>0$ there is $C>0$ such that for each pair $u\in U_n^{(r)}$ and $u'\in U_{n+k}^{(r')}$, with $1\le r,r'\le s$, it is

$$|(u,u')|\le Ce^{-(|\mu^{(r)}-\mu^{(r')}|-\delta)n}$$

if $||u||=||u'||=1$. (*Idea:* The case $r=r'$ is obvious. Distinguish the "trivial" case $r'>r$ from $r'<r$. If $r'>r$ consider first $k=1$ and note the following chain of relations (based on the spectral theorem for Λ_{n+1}, which allows us to write $\Lambda_{n+1}=\sum_i t_{n+1}^{(i)}P_i$ where P_i is the projection operator on the eigenplane associated with $t_{n+1}^{(i)}$), given $\delta_1>0$,

$$\begin{aligned}|(u,u')|\le&\max_{all\ u'',||u''||=1}e^{-(n+1)\mu^{(r')}}|(u,\sum_{r(i)\ge r'}e^{(n+1)\mu^{(r')}}P_iu'')|=\\=&e^{-(n+1)\mu^{(r')}}\max_{all\ u'',||u''||=1}|(\sum_{r(i)\ge r}e^{(n+1)\mu^{(r')}}P_iu,u'')|\le\\\le&e^{-(n+1)\mu^{(r')}}||\sum_{r(i)\ge r}e^{(n+1)\mu^{(r')}}P_iu||\le\\\le&e^{-(n+1)\mu^{(r')}+(n+1)\delta_1}||\sum_{r(i)\ge r}t_{n+1}^{(i)}P_iu||\le\\\le&e^{-(n+1)\mu^{(r')}+(n+1)\delta_1}||(T^{(n+1)*}T^{(n+1)})^{1/2}u||\end{aligned}$$

if n is so large that $|\frac{1}{m}\log t_m^{(j)} - \mu^{(i)}| < \delta_1/2$ for $m \ge n$ and for all labels j for which this is true: *i.e.* for the labels j such that $\lim \frac{1}{m}\log t_m^{(j)} = \mu^{(i)}$ (or $r(j) = i$ with the notations of [5.4.12]).
Then since the last quantity is $||T^{n+1}u||$, because

$$||(A^*A)^{1/2}u||^2 = ((A^*A)^{1/2}u, (A^*A)^{1/2}u) = (u, A^*Au) = ||Au||^2,$$

deduce, if n is so large that for $m \ge n$ it is $\frac{1}{m}\log ||T_m|| < \log 2$ (using here for the first time the first hypothesis in (*) of [5.4.13]).

$$\begin{aligned}|(u,u')| \le & e^{-\mu^{(r')}(n+1)+\delta_1(n+1)}||T^{n+1}u|| = e^{(n+1)(-\mu^{(r')}+\delta_1)}||T_nT^nu|| \le \\ & \le e^{(n+1)(-\mu^{(r')}+\delta_1)}2\,||T^nu|| \le 2\,e^{-\mu^{(r)}}\,e^{(n+1)(-\mu^{(r')}+\delta_1)}e^{(n+1)(\mu^{(r)}+\delta_1)}\end{aligned}$$

because $u \in U_n^{(r)}$. Hence, given $\delta_1 > 0$, there is $C_1 > 0$ such that the conclusion holds, if $r' > r$ and if $k = 1$. The same argument (of course) applies if $u' \in U_{n+1}^{(1)} \oplus \ldots U_{n+1}^{r+1}$ and $u \in U_n^{(r)} \oplus \ldots$. The geometrical meaning (consider first the 2–dimensional case with $r' = 1$ and $r = 2$) of the above is that the "angle" α_j between the orthogonal complement of $U_{n+j+1}^{(1)} \oplus \ldots \oplus U_{n+j+1}^{r+1}$, *i.e.* $U_{n+j+1}^{(r)} \oplus \ldots$, and $U_{n+j}^{(r)} \oplus \ldots$ is bounded by $C'e^{-(\mu^{(r)}-\mu^{(r+1)})(n+j)}$ for some C'. Therefore the angle α between $U_n^{(r)} \oplus \ldots$ and $U_{n+k}^{(r)} \oplus \ldots$ is bounded by $|\alpha| = |\sum_j \alpha_j| \le Ce^{-(\mu^{(r+1)}-\mu^{(r)})n}$. This proves the result if $r' = r+1$ and therefore for $r' > r$.
The conceptually nontrivial part is the case $r' < r$. Given the data u and u' imagine to define two orthonormal bases F and F' containing respectively u and u' among their elements. Let the orthonormal bases be chosen so that the first m_1 vectors of F are in $U_n^{(1)}$, the successive m_2 in $U_n^{(2)}$, ..., the last m_s in $U_n^{(s)}$; and the vectors of F' have the analogous property with respect to the eigenspaces $U_{n+k}^{(i)}$.
Consider the orthogonal matrix $U_{\alpha\alpha'} = (u_\alpha, u'_{\alpha'})$. By what seen in the preceding case we can say that (with the definition of $r(\alpha)$ given in [5.4.14])

$$\begin{aligned}&|(u_\alpha, u'_{\alpha'})| \le C_1 e^{-\left(\mu^{(r(\alpha'))}-\mu^{(r(\alpha))}-3\delta_1\right)n} && \text{if} \quad u_\alpha \in U_n^{(r(\alpha))}, u'_{\alpha'} \in U_{n+k}^{(r(\alpha'))}\\ &|(u_\alpha, u'_{\alpha'})| \le 1 && \text{for all} \quad \alpha,\alpha'\end{aligned}$$

and the *key remark, basis and foundation of the theorem*, is simply that if $r(\alpha) < r(\beta)$ then we can use orthogonality of U which implies $(u_\alpha, u'_\beta) = U_{\alpha\beta} \equiv \overline{U^{-1}_{\beta\alpha}}$ where the bar denotes complex conjugation. The latter quantity is the determinant of the matrix obtained by erasing the row α and the column β of the matrix U (because U is orthogonal). The evaluation of this determinant consists in the sum of the $(\ell-1)!$ products: in performing such products we shall pick a certain number of factors equal to the matrix elements $\gamma\gamma'$ "above the diagonal" for which $r(\gamma') > r(\gamma)$ which are bounded by $C_1e^{-\Delta\mu n + 2l\delta_1 n}$ and one checks that the sum of the $\Delta\mu$ that one finds is necessarily $\ge \mu^{(r)} - \mu^{(r')}$ (to understand this property consider first the nondegenerate case in which all Lyapunov exponents are distinct, hence $r(\alpha) \equiv \alpha$). The other matrix elements are bounded by 1 hence one sees that in this case $|(u_\alpha, u'_\beta)| \le (\ell-1)!C_1^{\ell-1}e^{-|\mu^{(r)}-\mu^{(r')}|n+2l\delta_1 n}$ concluding the analysis).

[5.4.16]: Check that the result in [5.4.15] implies that $U_n^{(i)}$, thought of as a plane in R^ℓ tends, for $n \to \infty$ to a limit plane $U^{(i)}$ for each i. (*Idea:* The planes $U_{n+k}^{(1)}$ and $U_{n+h}^{(1)}$ must form with $\bigoplus_{r>1} U_n^{(r)}$ an angle closer to 90^o than a prefixed quantity, if n is large enough, independently of $h, k \ge 0$. Hence the planes $U_n^{(1)}$ form "a Cauchy sequence"; for the other planes the argument is identical).

[5.4.17]: From [5.4.16] and from the hypothesis (*) in [5.4.13] deduce that the limit $\lim_{n\to+\infty}(T^{n*}T^n)^{1/2n}$ exists and is a matrix Λ with eigenvalues $e^{\mu^{(i)}}$ with corresponding eigenspaces $U^{(i)}$, for $i=1,\ldots,s$.

[5.4.18]: Deduce that $\lim n^{-1}\log|T^n u|=\mu^{(i)}$ if $u\in U^{(i)}$.

[5.4.19]: Check that the results of problems [5.4.11]÷[5.4.18] imply theorem I. (*Idea:* By the same arguments we can construct $\hat U^{(1)},\ldots,\hat U^{(s)}$ using S^{-1} so that, by problem [4.5.10], $\lim_{k\to\infty}k^{-1}\log|S^{-k}dx|/|dx|=-\mu^{(i)}$ if $dx\in\hat U^{(i)}$. Then we can set $W^{(i)}=\Big(U^{(s)}\oplus U^{(s-1)}\oplus\ldots U^{(i)}\Big)\cap\Big(\hat U^{(i)}\oplus\hat U^{(i-1)}\oplus\ldots\hat U^{(1)}\Big)$).

[5.4.20]: Consider the dynamical system $\mathcal{C}=[-1,1]\times T^2$, where T^2 is the bidimensional torus. Let $x=(z,\varphi_1,\varphi_2)$ be a point in $\mathcal{C}$. Let $z\to f(z)$ be a regular map such that $f^n(z)\xrightarrow[n\to\pm\infty]{}\pm1$ for all $|z|\ne1$, and let $\nu(z)=2$ if $z>0$ and $\nu(z)=1$ if $z<0$; define

$$x'=(z',\varphi_1'\varphi_2')=\begin{cases}f(z)\\ \varphi_1+\nu(z)\varphi_2 \mod 2\pi\\ \nu(z)\varphi_1+(\nu(z)^2+1)\varphi_2 \mod 2\pi\end{cases}$$

and check that the statistics of initial data randomly chosen with distribution $\mu=dz\,d\varphi_1\,d\varphi_2/(2\pi)^2$ are *different* for $n\to\pm\infty$.

[5.4.21]: Compute the Lyapunov exponents in the example of problem [5.4.20]; and find the dynamical bases (of the points that admit them) and the systems of planes relative to S and S^{-1} (of the points that admit them).

[5.4.22]: Consider the dynamical system (M,S_t) obtained by considering the bidimensional torus T^2 deprived of a finite number of pairwise disjoint circular regions $C_1,\ldots,C_s$ (or more generally convex) and a point that moves among them with unit speed and elastic collisions. Check that the maps S_t on phase space of the (x,α), where $x\in T^2/\cup_j C_j$ and $\alpha\in[0,2\pi]$ denote position and velocity direction, with $S_t(x,\alpha)=(x',\alpha')$ giving the coordinates at time t of the point that initially was in (x,α). Check that the maps S_t have singularities but they leave the Liouville measure $dxd\alpha$ on M invariant.

[5.4.23]: In the context of [5.4.22] define on every boundary ∂C_j a coordinate $r\in[0,\ell_j]$ giving the curvilinear abscissa of the generic point of ∂C_j. Consider a *collision* (x,α) with $x\in\cup_j\partial C_j$ and α such that velocity forms an angle $\varphi\in[\frac12\pi,\frac32\pi]$ with the external normal to the point x of the obstacle C_j. Then the set of the *collisions* is parameterizable with (j,r,φ) and the set of such parameters will be denoted $\mathcal{C}$. Imagine to define the timing events as the events of collisions. Let S be the map of a collision (j,r,φ) into the successive (j',r',φ'). Check that S conserves the measure $\mu=\sin\varphi d\varphi dr$ and that S is regular on $\mathcal{C}$ with respect to the volume and that it is μ–regular. The dynamical system $(\mathcal{C},S)$ is called a *billiard ball system* or imply *billiard*.

[5.4.24]: (*Oseledec*) Note that in the above theory it is not necessary that the matrix $T(x)$ be the Jacobian of the map S: exactly the same conclusions hold if $T(x)$ is supposed to be a matrix valued function on M such that $\int\log_+||T(x)||\,\mu(dx)<+\infty$. Therefore the above theory can be viewed as a theory of products of random matrices.

[5.4.25]: Let $M=[0,2\pi]$ and $\mathcal{M}$ be defined by

$$\mathcal{M}(\varphi)=\begin{pmatrix}\lambda\cos\varphi\,-\,E & -1\\ 1 & 0\end{pmatrix}$$

where λ,E are real numbers. Let $\mathcal{M}_N(\varphi)=\mathcal{M}(\varphi+(N-1)\omega)\ldots\mathcal{M}(\varphi+\omega)\mathcal{M}(\varphi)$ where $\omega/2\pi$ is *irrational*: check that the maximum Lyapunov exponent, in the sense of the extension in [5.4.24], can be expressed as:

$$\Lambda_{\max}(\varphi)=\lim_{N\to\infty}\frac{1}{2N}\log||\mathcal{M}_N(\varphi)^*\mathcal{M}_N(\varphi)||=$$
$$=\lim_{N\to\infty}\frac{1}{2\pi}\int_0^{2\pi}d\psi\,\frac{1}{2N}\log||\mathcal{M}_N(\psi)^*\mathcal{M}_N(\psi)||$$

which exists and is φ–independent μ–almost surely. (*Idea:* This is Oseledec theorem rephrased and taking into account the extension [5.4.24] and the ergodicity of the irrational rotations of the circle, see problem [5.1.6]).

[5.4.26]: In the context of the above [5.4.24], [5.4.25], note that if $z \stackrel{def}{=} e^{-i\varphi}$ then

$$\Lambda_{\max}(\varphi) = \lim_{N\to\infty} \frac{1}{2\pi\, N} \int_0^{2\pi} \max_{||\underline{v}||=1} \log ||\mathcal{M}_N(\varphi)\underline{v}||\, d\varphi =$$
$$= \lim_{N\to\infty} \frac{1}{2\pi\, N} \int_0^{2\pi} \max_{||\underline{v}||=1} \log ||\tilde{\mathcal{M}}_N(z)\underline{v}||\, d\varphi$$

where

$$\tilde{\mathcal{M}}(ze^{-in\omega}) = \begin{pmatrix} \frac{\lambda}{2} + \frac{\lambda}{2} z^2 e^{-2in\omega} - E\, z\, e^{-in\omega} & z e^{-i\omega n} \\ -ze^{-i\omega n} & 0 \end{pmatrix} \qquad \textit{and}$$
$$\tilde{\mathcal{M}}_N(z) = \tilde{\mathcal{M}}(ze^{-i(N-1)\omega}) \ldots \tilde{\mathcal{M}}(z)$$

Show that the function $\tilde{\mathcal{M}}_N(z)\underline{v}$ as a function of z is analytic for all $\underline{v}$ and for $|z| \le 1$. Deduce that $||\tilde{\mathcal{M}}_N(\varphi)\underline{v}||$ and hence $\max_{||\underline{v}||=1} ||\tilde{\mathcal{M}}_N(\varphi)\underline{v}||$ are subharmonic functions of z and so is the logarithm $\log ||\tilde{\mathcal{M}}_N(\varphi)||$. (*Idea:* Just recall the definition of a subharmonic function: f is subharmonic in a domain D if the value of f at every point z is $\le$ than the average of f on a circle in D centered at z: then apply Cauchy's theorem on holomorphic functions and use the triangular inequality and the concavity of the logarithm).

[5.4.27]: *(Herman's theorem)* Show that the largest Lyapunov exponent of the product of matrices in [5.4.25], [5.4.26] is

$$\Lambda_{\max} \ge \frac{1}{2}\lambda \qquad \text{for almost all } \varphi$$

Show that, as a consequence, the recurrence relation (*Schöedinger quasi periodic equation*)

$$-(x_{n+1} + x_{n-1}) + \lambda \cos(\varphi + n\omega)\, x_n - E x_n = 0$$

is such that for $\lambda > 2$ and any E there is a set $\Delta(\lambda, E)$ of φ's in $[0, 2\pi]$ of zero measure such that, if $\varphi \notin \Delta(\lambda, E)$, there exist initial data (x_1, x_0) generating, under the above recurrence, a trajectory (x_n, x_{n+1}) diverging exponentially as $n \to +\infty$. (*Idea:* Since the function $\log ||\mathcal{M}_N(z)||$, as a function of z, is subharmonic the integral $\frac{1}{2\pi}\int_0^{2\pi} \log ||\tilde{\mathcal{M}}_N(z)||\, d\varphi$ with $z = e^{-i\varphi}$ is larger or equal than $\log ||\tilde{\mathcal{M}}_N(0)|| \equiv \log(\lambda/2)^N$. The question on the recurrence relation generates exactly the problem on products of matrices solved in the previous two problems.)

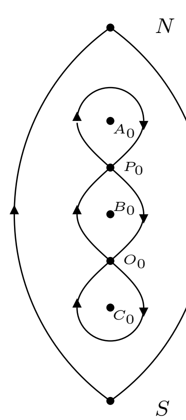


Fig. (5.4.1) Illustration of the one of the four quarters of the sphere in the construction of a non normal attracting set, *c.f.r.* problem [5.4.28].

[5.4.28]: *(a non normal attracting set)* Consider a sphere and draw two orthogonal maximal circles on it. Let N, S be the poles where they cross. We label the four "meridians" $0, 1, 2, 3$ counterclockwise; and we imagine that N and S are hyperbolic fixed points with the meridians $0, 2$ being stable manifolds for N and unstable for S while the meridians $1, 3$ are the unstable manifolds of N and the stable ones for S. Thus the sphere is partitioned into four sectors. In each sector i we put two points O_i, P_i and we imagine that they are hyperbolic points for the map and that their stable and unstable manifolds link the two points in a three loop fashion (*i.e.* in a double 8 shape), see Fig. (5.4.1). At the centers of the loops $\mathcal{L}_i$ we imagine three points A_i, B_i, C_i which are also hyperbolic fixed points (with complex Lyapunov exponents). We now define arbitrarily the map S elsewhere so that it pushes away from the meridians and from the fixed points A_i, B_i, C_i and pushes towards the loops $\mathcal{L}_i$. Clearly P_i, O_i behave like the points O_0, P_0 in the problem [5.3.6]. Note that the system *does not* verify the axiom A and identify the reasons. Show that the nonwandering set consists in the lines joining N, S and O_i, P_i, just described, and of the points A_i, B_i, C_i, while the attracting sets are the loops $\mathcal{L}_i$: check that no point outside the above lines has a well defined statistics, because of the mechanism of [5.3.6]. The periodic points are *not* dense and the attracting sets are not hyperbolic. And the SRB distribution is undefined. (*Idea: c.f.r.* [5.3.6].)

Bibliography: [Sm67], [Ru89b], [Ru79]; the philosophical problems are taken from [Ga81] and the theory of Oseledec is taken from the extension [Ru89] of the original articles [Os68], [Ra79]. The result on the lower bound on the maximum Lyapunov exponet for the lattice Shrödinger equation exposed in the last three problems is taken from [He83]. For a general theory of random matrices and their products see [FK60] and [Me90].

§5.5 SRB Statistics. Attractors and attracting sets. Fractal dimension.

We now go back to systems with an attractive set verifying the axiom A, in the sense of definition 2 §5.4. The aim is the anlysis of the results mentioned in the previous sections, about the SRB statistics generated by trajectories obtained by randomly selecting initial data with a distribution absolutely continuous with respect to the volume on phase space.

(A) "Physical" (i.e. SRB) probability distributions.

It is convenient to begin with an intuitive description of the SRB distribution and of its main properties. The difficulty is in the correct visualization of an attractive hyperbolic set A which in general is a set that is not a regular surface, not even piecewise, and it has a fractal nature. See problems [5.5.8], [5.5.9] for examples of attractive fractal sets on which the evolution map acts verifying the axiom A.

To develop the intuition it is, however, useful to consider a simple case in which A is a surface. Consider the dynamical system with a 3–dimensional phase space $M = T^2 \times [-1, 1]$ whose points are described by (x, y, z) with x and y defined $\bmod 2\pi$, so that they must be thought of as coordinates on a bidimensional torus, *c.f.r.* remark 2 to definition 2 of §5.4. The map S defining the dynamics will be

$$S(x,y,z)=(x',y',z')=\begin{cases} x'=x+y & \mod 2\pi \\ y'=x+2y & \mod 2\pi \\ z'=\frac{1}{2}z \end{cases} \tag{5.5.1}$$

In this case the set $A=T^2\times\{0\}$ is a global hyperbolic attractive set because from every point $x\in A$ emerge three vectors

$$\begin{aligned} \underline{e}_+ =&(1,\frac{1}{2}(1+\sqrt{5}),0)\equiv(\underline{v}_+,0) \\ \underline{e}_- =&(1,\frac{1}{2}(1-\sqrt{5}),0)\equiv(\underline{v}_-,0), \qquad \underline{e}_3=(0,0,1) \end{aligned} \tag{5.5.2}$$

where $\underline{v}_\pm$ are eigenvectors (not normalized) of the matrix $\begin{pmatrix}1&1\\1&2\end{pmatrix}$ relative to the eigenvalues $e^{\lambda_+}=(3+\sqrt{5})/2$ and $e^{\lambda_-}=e^{-\lambda_+}$. An infinitesimal vector $d\xi$ in the direction of $\underline{e}_+$ expands under the action of S exponentially

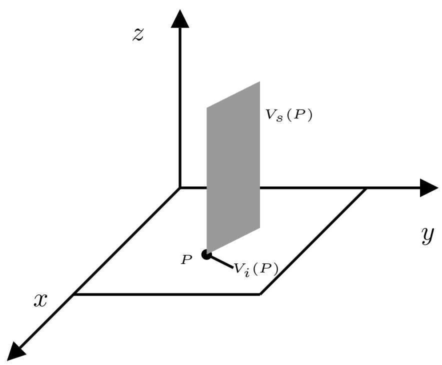


Fig. (5.5.1): The grey surface is a piece of the stable plane $V_s(P)$ of P while the segment is a piece of the unstable plane $V_i(P)$.

with progression e^{λ_+}, while one in the direction of $\underline{e}_-$ contracts with $e^{-\lambda_+}$ and one in the direction $\underline{e}_3$ contracts with $1/2$.

The 2–dimensional linear space spanned by $\underline{e}_-,\underline{e}_3$ is the space V_2 of definition 4 of §5.4, while the space V_3 is the line parallel to $\underline{e}_3$. Likewise the space spanned by $\underline{e}_3,\underline{e}_+$ is $\tilde V_2$ of the definition 4, §5.4, for S^{-1} and the space parallel to $\underline{e}_+$ is the space $\tilde V_3$ of the definition 4,§5.4. The stable and unstable directions of a point $x\in A$ are, respectively, $V_s(x)\equiv R_2$ (plane $\underline{e}_-,\underline{e}_3$) and $V_i(x)\equiv\tilde R_3$ (line $\underline{e}_+$). See the above Fig. (5.5.1) for a graphical representation of such geometrical objects.

This particularly simple example illustrates some general properties, not always easy to prove or to see in less simple dynamical systems (M,S).

(1) The families of planes $V_s(x)$ and $V_i(x)$ *are integrable*, this means that inside a ball centered at x of radius δ small enough (compared to the diameter and to the curvature of M) one can define a portion of regular surface

$W_x^{\delta,s}$ passing through x and having at each of its points y in A the plane $V_s(y)$ as tangent plane. Therefore $W_x^{\delta,s}$ is a graph over the ball (a disk in the example) of radius δ around x in the plane $V_s(x)$, *i.e.* locally $W_x^{\delta,s}$ and $V_s(x)$ are identical up to second order in the distance from x.

Likewise one can define a portion of regular surface $W_x^{\delta,i}$ having everywhere $V_i(y)$ as tangent plane at each of its points y, and $W_x^{\delta,i}$ is a graph over the ball (a segment in the example) of radius δ around x in the plane $V_i(x)$. See Fig. (5.5.1).

In the example under analysis such surfaces are "trivial": they are in fact the (disk) intersection between the sphere of radius δ centered at x and the plane $\underline{e}_-\underline{e}_3$ through x and, respectively, the (segment) intersection between the same sphere and the line through x parallel to $\underline{e}_+$: *provided* r is small compared to the dimension of M, *i.e.* with respect to 2π. If δ is too large the disk and the segment must be thought of as "wrapped" on phase space (since x, y are defined $\mathrm{mod}\, 2\pi$) and they cannot any longer be regarded as graphs.

(2) The surfaces $W_x^{\delta,s}$ and $W_x^{\delta,i}$ are such that if $y \in W_x^{\delta,s}$ then $|S^n x - S^n y| \xrightarrow[n\to+\infty]{} 0$ exponentially (with progression constant at least equal to the weakest contraction of the tangent vectors); while if $y \in W_x^{\delta,i}$ one has $|S^{-n} x - S^{-n} y| \xrightarrow[n\to+\infty]{} 0$ exponentially (with progression at least equal to the weakest expansion of the tangent vectors). The surfaces $W_x^{\delta,s}, W_x^{\delta,i}$ can be called *r–local stable and unstable manifold* of x.

(3) At every point x the two surfaces are "transverse and independent" *i.e.* their tangent planes span the whole tangent space and form a nonzero angle (equal to the angle between $V_i(x)$ and $V_s(x)$ which, in the natural metric on M, in the present case, is constant and equal to 90^o). An important general result (Pesin, *c.f.r.* theorem I, §5.4 and the following remark (xvii)) shows that in the hypotheses of theorem I, §5.4, and fixed arbitrarily a regular invariant probability distribution μ (defined on M) one has that through μ–almost every point of M pass portions of regular surfaces $W_x^{\delta(x),s}$ and $W_x^{\delta(x),i}$ with $\delta(x) > 0$ that *integrate* the planes $V_i(x)$ and $V_s(x)$ of the dynamical base: hence this *is not a peculiarity* of the example.

One can define the "global stable and unstable manifolds" of x as described in definition 4 of §5.4: namely as the set of points y such that $S^n y$ get exponentially fast close to $S^n x$ as $n \to +\infty$ or, respectively, as $n \to -\infty$.

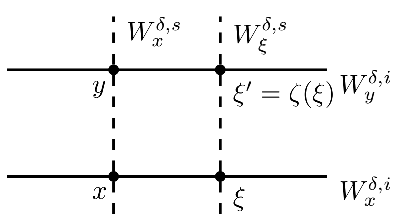


Fig. (5.5.2): The correspondence ζ between two nearby portions $W_x^{\delta,i}$ and $W_y^{\delta,i}$ of unstable manifolds established by the stable manifolds $W_\bullet^{\delta,s}$ that intersect them. The points x, y and ξ, ξ' are two pairs of corresponding points. The segment $\delta^o_x = x\xi$ is mapped by ζ into the segment $\delta^o_y = y\xi'$

(4) In the example (5.5.1) we see that if δ is small enough the regularity of these surfaces (clear in this case because they are portions of flat surfaces) allows us to define, given two close enough points x, y, a *correspondence* between points $\xi \in W^{\delta,i}_x$ and $\xi' \in W^{\delta,i}_y$ putting into correspondence ξ and ξ' if $\xi, \xi' \in W^{\delta,s}_\xi$. The correspondence is illustrated in Fig. (5.5.2) (in which $\xi, \xi' = \zeta(\xi)$ are in correspondence, like x and y themselves).

Corresponding points evolve in the future getting exponentially close and, although roaming on the attractive set, they are dynamically indistinguishable. This happens *even if x and y are on the same unstable manifold* (which is possible if, as in the example, the global unstable manifolds are dense): *i.e.* if $W^{\delta,s}_x$ continued outside the ball of radius δ contains $W^{\delta,s}_y$.
Moreover corresponding infinitesimal arcs δ^0_x of $W^{\delta,i}_x$ and δ^0_y of $W^{\delta,s}_y$ have comparable lengths, *i.e.* lengths with finite ratio (1 in the example): this, [Pe76], remains true in a suitable sense, in general, under the hypotheses of the theorem I of §5.4. The latter property has the name of *absolute continuity* of the unstable manifold with respect to the stable one. Naturally, at the same time, the symmetric property holds: with interchanged roles between stable and unstable manifold.

In general, under very weak assumptions, it is possible to define the stable and unstable manifolds of μ–almost all points of an invariant set with respect to the action of a μ–regular (*c.f.r.* definition 2, §5.4) map S: however the definition has an apparently very technical character unless one has in mind concrete examples (it is in fact entirely inspired by the properties of the billiards, *c.f.r.* §23,§5.4, and it is an "abstract version" of it and a generalization). We rather devote here our attention to the much simpler case of differentiable systems (M, S), *i.e.* with M analytic (or C^∞) and S an analytic diffeomorphism (or C^∞) such that (M, S) is a dynamical system that verifies axiom A (*c.f.r.* §5.4, definition 2 and remarks following it). For such a system the following theorem holds

I Theorem *(existence of stable and unstable manifolds and their absolute continuity):*
Let (M, S) be a differentiable dynamical system verifying axiom A and let A be an attractive set on which S is topologically transitive (which is not very restrictive, c.f.r. §5.4, remarks to definition 2). Suppose that S is invertible with differentiable inverse in the vicinity of A.[1] *Then*
(i) The stable and unstable planes $V_i(x)$ and $V_s(x)$ are Hölder continuous as $x \in A$ varies; furthermore they are integrable in the sense that through each $x \in A$ pass portions of surfaces of class C^∞ denoted $W^{\delta,s}_x$ and $W^{\delta,i}_x$. Such surfaces at each of their points y in A are tangent to the plane $V_s(y)$

[1] Or at least suppose that every point of A be part of a sequence $\ldots, x_{-2}, x_{-1}, x_0 = x$ of points of A such that $Sx_{k-1} = x_k,\ k \le -1$.

and $V_i(y)$ respectively, provided δ is small enough ("existence of the δ–local stable and unstable manifolds on A").

(ii) There exist $C, \lambda > 0$ for which if $y \in W_x^{\delta,s}$ then $d(S^n y, S^n x) \le Ce^{-\lambda n}$. Furthermore $W_x^{\delta,i} \subset A$ and if $y \in W_x^{\delta,i}$ then $d(S^{-n}y, S^{-n}x) \le Ce^{-\lambda n}$.[2]

(iii) Let δ be small enough so that the surfaces $W_x^{\delta,i}$ exist; and consider, for each $y \in W_x^{\delta,s} \cap A$, the surface $W_y^{\delta,i}$: see Fig. (5.5.2). Given x and $y \in W_x^{\delta,s}$ with $y \in A$ close enough to x, for each $\xi \in W_x^{\delta/2,i}$ the $W_\xi^{\delta,s}$ intersects the $W_y^{\delta,i}$ in a point ξ'. One thus establishes a correspondence $\xi' = \zeta(\xi)$ "along the stable manifold" $W_\xi^{\delta,s}$ between points of two close δ–local unstable manifolds, $W_x^{\delta,i}$ and $W_y^{\delta,i}$. The correspondence is defined for each ξ close enough to x and it enjoys the properties

(a) it is Hölder continuous and

(b) the ratio between corresponding surface elements $d\sigma$ and $d\sigma'$ is a non vanishing function $\rho_{x,y}(\xi)$.

Remarks:

(1) Property (b) is called *absolute continuity* of the unstable manifold. Item (ii) allows us to define the *global stable manifolds* $W_s^g(x)$ simply as the set of the y such that $d(S^n x, S^n y) \xrightarrow[n\to+\infty]{} 0$. And one can analogously define the *global unstable manifolds* $W_i^g(x)$. In the considered example they are the plane $\underline{e}_-, \underline{e}_3$ and the line $\underline{e}_+$ through x naturally regarded as wrapped on the torus, since the coordinates x, y are defined mod 2π. As in the case of equation (5.5.1), the global manifolds of each point are dense on the attractive set. This is a *general* property, *c.f.r.* [Ru89b] p. 157, in the case that the attractive set is topologically mixing which, in turn, is a property that is not substantially restrictive (by the remark (4) to definition 2 of §5.4). Hence we shall use it to build intuition on the nature of an attractive set.

(2) Property *(iii)* shows that the absolute continuity remains true, although in a weaker local sense, much more generally than in the example (5.5.1).

(B) Structure of axiom A attractors. Heuristic considerations.

(1) In the topologically mixing case the attractive set A is the closure of the unstable manifold of one of its points, arbitrarily chosen. Hence it is also the closure of the unstable manifold of an arbitrary fixed point, or of an arbitrary periodic orbit, lying on it (for instance of $\underline{0}$ in the above example in equation (5.5.1)). The stable manifold of a point $x \in A$ consists of a "negligible" part (that we can call *wandering* or *errant*), consisting in the points "really" outside of the attractive set,[3] and in a part on A itself which we can call "non negligible part".

---

[2] In the non invertible case the "close enough pair" $S^{-n}x, S^{-n}y$ must be replaced by pairs y_{-n}, x_{-n} such that $S^n y_{-n} = y$, $S^n x_{-n} = x$: one shall have that $d(y_{-n}, x_{-n}) \le Ce^{-\lambda n}$.

[3] In the example (5.5.1) this "negligible part" consists in the points that have a coordinate $z \ne 0$ and evolve towards the attractive set while the part on the attractive set consists of the points with $z = 0$ located on the line parallel to $\underline{e}_-$ which covers densely A.

In many cases the distinction, between negligible part or non negligible part, which is here sharp, is not so clear and the negligible part may even be empty, *c.f.r.* examples in the problems [5.5.7], [5.5.8]. What is however always correct is to think of the attractive set as the closure of the unstable manifold of any of its fixed or periodic point (excluding non topologically mixing maps which, however, would only trivially modify the picture). These manifold are wrapped on themselves, but two of its close points get far away from each other, with exponential progression, under the evolution by S, *provided* their distance is measured *along the manifold itself.*[4]
(2) Hence it is convenient to think the attractive set as the unstable manifold of a fixed or periodic point x_0 "developed" over phase space. In the considered example this means that the line $\mathcal{L}$ through $x_0 = (\underline{0}, 0)$ parallel to $\underline{e}_+$ is not thought of as as a line wrapped on the torus (and dense there): it is rather thought of as an unfolded line $\mathcal{L}$ on the plane (*i.e.* we do not think of it as defined $\bmod 2\pi$). Motion is then of great simplicity: initial data that are in a 3-dimensional neighborhood U (for instance a little cube of side h with center in $x_0 \in A$ and sides parallel to $\underline{e}_+, \underline{e}_-, \underline{e}_3$) evolve becoming confused with the line $\mathcal{L}$; and U is deformed into a very long and thin parallelepiped.
(3) With this heuristic representation of motion in mind we can see easily what happens if we choose $\mathcal{N}$ points $x_1, \ldots, x_{\mathcal{N}} \in U$ with distribution absolutely continuous with respect to the volume measure; for instance with constant density $\rho = h^{-3}$. It is clear that the average $\langle f \rangle$ of the values of an arbitrary C^∞ observable f on phase space is computable, over a large finite time, by simply averaging it on the set $S^N U$ with the density $h^{-3} |(\det \partial S^N)_i^{-1}|$, where $(\partial S^N)_i$ is the Jacobian matrix of the map S^N *as a map of the line* $\mathcal{L}$ (which is the unstable manifold, whence the notation with the index i), rather than of the entire M:

$$\langle f \rangle = \int_{S^N U} f(x)\, \rho\, |\det(\partial S^{-N}(x)_i)|\, dx \tag{5.5.3}$$

Since the region $S^N U$ is very thin around a bounded portion $\mathcal{L}_N$ of the line $\mathcal{L}$ what really counts is the linear density on $\mathcal{L}_N$. Thus the average of f can be simply computed by integrating its values on $\mathcal{L}_N$ with respect to the arc length ds with a linear density *proportional* to $\Lambda_i(S^{-N} x(s); N)^{-1}$ if $\Lambda_i(x; N)$ is the expansion coefficient of the line element on $\mathcal{L}$ that is initially inside U and arrives in $x(s)$ at time N. If $\lambda_i(x)$ denotes the "local Lyapunov exponent", $\lambda_i(x) \equiv \log \Lambda_i(x; 1)$, then by the composition of differentiations, it is $\Lambda_i(x, N) = \prod_{j=0}^{N-1} e^{\lambda_i(S^j x)}$.
In general the unstable manifold has more than one dimension and the dilatation coefficient $\Lambda_i(x; 1)$ has to be replaced by the dilatation of the area element under the action of the restriction of S to the manifold $\mathcal{L}$,

---

[4] Of course since the manifold is bounded the real distance cannot grow indefinitely, *unlike* the distance measured along the shortest path contained in the manifold itself.

i.e. $|\det(\partial S)_i|$ and the linear density should now become a surface density with respect to the area element of the unstable manifold.
(4) What just said does not yet allow us to describe which is the SRB distribution associated with the attractive set: to obtain it we should eliminate the dependence of the results from the line $\mathcal{L}$ that represents the "developed" attractive set, because obviously such an ideal geometric object should not appear in the results. Therefore we recall that points that are very far on $\mathcal{L}$ may in fact be close as points on A. For instance if x and y are two points of W^i_x close on M and such that $y \in W^s_x$ then we can easily compare the mass near x and that near y: indeed let $\delta^0_x \subset \mathcal{L}$ be a (infinitesimal) segment of unstable manifold around x and δ^0_y be the *corresponding* segment around y obtained via the correspondence illustrated in Fig. (5.5.2) where δ^o_x is represented by $x\xi$ and δ^o_y by $y\xi'$.

Imagine N' very large and note that the mass of $S^{N'}U$ which is found around x or y becomes more and more rarefied as N grows because $S^{N'}U$ will stretch more and more the mass originally in the vicinity U of the fixed point O; the ratio of the masses in x and y tends exactly to

$$\begin{aligned}\frac{\mathrm{mass}(\delta^0_x)}{\mathrm{mass}(\delta^0_y)} &\stackrel{def}{=} \lim_{N'\to\infty} \frac{\delta^0_x \prod_{j=0}^{N'} e^{-\lambda_i(S^{-j}x)}}{\delta^0_y \prod_{j=0}^{N'} e^{-\lambda_i(S^{-j}y)}} \equiv \\ &\equiv \lim_{\substack{N\to\infty\\ N'\to\infty}} \frac{\prod_{j=-N'}^{N} e^{-\lambda_i(S^j x)}\delta^N_{S^N x}}{\prod_{j=-N'}^{N} e^{-\lambda_i(S^j y)}\delta^N_{S^N y}} \equiv \prod_{j=-\infty}^{\infty} \frac{e^{-\lambda_i(S^j x)}}{e^{-\lambda_i(S^j y)}}\end{aligned} \tag{5.5.4}$$

where the first relation requires considering the limit as $N' \to \infty$ because the expansion and contraction of S near the fixed point are not constant, in general. In the second relation $\delta^N_{S^N x}$ is the segment of unstable manifold image of δ^0_x under the action of S^N, and a similar notation is used for $\delta^N_{S^N y}$.
The two last segment are "practically identical" since (*c.f.r. (i)* of theorem I) the unstable manifolds of a pair of points that are on the *same* stable manifold are Hölder continuous functions, with some exponent $\alpha > 0$, of the distance between x and y (and hence also $\lambda_i(x)$ depends on x so that it is Hölder continuous with exponent $\alpha > 0$).
Finally the infinite product of the third relation converges because $S^j(x)$ and $S^j(y)$ have distance that tends to zero exponentially both in the future and in the past: in the future because they are situated on the same stable manifold, by assumption, and in the past because they are on the same unstable manifold.
All this is trivial in the case of the example because $\lambda_i(x) \equiv \lambda_+$ is constant; and also $\delta^0_x/\delta^0_y = 1$ because the lines that establish the correspondence between the segments $\delta^0_x\ \delta^0_y$ are perpendicular to the segments themselves (parallel to $\underline{e}_-$).

Let us introduce, inspired by the last remark (4), the notion of distribution absolutely continuous on the unstable manifold:

1 Definition *(absolute continuity along the unstable manifold):*
In the hypotheses of theorem I consider the δ–local manifolds $W_x^{\delta,i}, W_x^{\delta,s}$ on an attractive set A. Let the set Δ be a "parallelogram" consisting in the union of the surfaces $W_\xi^{\delta,i}$ that pass through the points $\xi\in W_x^{\delta/2,s}$ (the geometry is illustrated in Fig. (5.5.3)). We consider for an arbitrarily fixed smooth f the integral $\int_\Delta f(y)\mu(dy)$.
We shall say that the measure μ defined on the Borel sets of A is absolutely continuous on the unstable manifold *if for all the points $x\in A$ it happens that the integral can be computed as*

$$\int_\Delta f(y)\mu(dy)=\int_{W_x^s\cap A}\nu(d\xi)\int_{W_\xi^i} f(\xi,\sigma)\rho_\xi(\sigma)d\sigma \tag{5.5.5}$$

where ν is a suitable probability distribution concentrated on $W_x^{\delta/2,s}\cap A$, $d\sigma$ is the surface element on $W_\xi^{\delta,i}$, a point $y\in\Delta$ located on the surface element $d\sigma\in W_\xi^{\delta,i}$ is denoted (ξ,σ) and $\rho_\xi(\sigma)$ is a suitable function called the density on the unstable manifold.

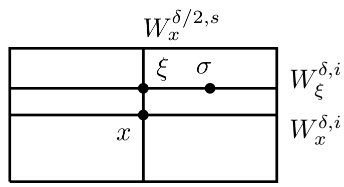


Fig.(5.5.3): *The rectangle represents Δ and the figure illustrates (5.5.5). The horizontal segments represent portions of unstable manifolds and the vertical ones portions of stable manifolds.*

Heuristic remarks:

(1) The (5.5.5) is a generalization of Fubini's formula for ordinary integrals and differs from it because the measures μ and ν *are not* necessarily given by a density, *i.e.* they are not necessarily absolutely continuous. In the example considered in (5.5.1), ν is *not* absolutely continuous (trivially) because if $x=(\xi,\sigma,z)$ is a generic point it is $\nu(d\xi\,dz)=d\xi\,\delta(z)\,dz$.
(2) The formula (5.5.4) is quite interesting: if the attractive set is visualized as an unstable manifold then we can say that the SRB distribution has a density with respect to the area measure on the unstable manifold and the density is inversely proportional to the expansion coefficient of the manifold itself. In other words one can think that the dynamical system restricted to the attractive set has a statistics which is "an equilibrium Gibbs distribution" with energy function formally equal to $H=cost+\sum_{j=-\infty}^{+\infty}\log\lambda_i(S^j(x))$ and inverse temperature $\beta=1$: in such distribution every configuration x has a "weight" proportional to $e^{-\beta H(x)}$. This is obviously improper because the surface is infinite, since the unstable

manifold "has no boundary", and H is given by a divergent sum. Nevertheless the product $e^{-H}d\sigma$ can make sense and it can be interpreted as the measure μ.

We recall that in statistical mechanics we say that the probability distribution that describes the equilibrium state of a mechanical system of particles in R^3 (*gas*) is the Gibbs measure $\mu = Ce^{-H(\underline{p},\underline{q})}d\underline{p}d\underline{q}$: this is also a statement that does not make sense, unless we intend it as a limit relation in a suitable sense. The sense in which we should understand the statement made on H and μ in the case of an attractive set is analogous.

One should interpret the area measure on the unstable manifold of the fixed point as analogous to the Liouville measure: so that the part of unstable surface $S^N W_O^{\delta,i}$, of the chosen fixed point O, with N large will be analogous to the container of large size in the case of the gas. And $const\, e^{-\sum_{j=0}^N \lambda_i(S^{-j}x)}d\sigma$, *c.f.r.* (5.5.4), is analogous to the Gibbs measure at finite volume; the limit $N\to\infty$ plays the role of thermodynamic limit.

These analogies are the true reason behind the otherwise surprising similarity between the language used to describe qualitative theory of motions with attractive sets and the language and the ideas of statistical mechanics.

(3) Clearly (5.5.5) does not determine uniquely ν and ρ_ξ: if there is such a pair we can multiply $\nu(d\xi)$ by some function of ξ and divide ρ_ξ by the same function obtaining another pair ν',ρ' that still verifies the (5.5.5). There is however a natural choice for ν and ρ that is suggested by the argument leading to (5.5.4).

One can define $\nu(d\xi)$ so that $\int_{\xi_1}^{\xi_2}\nu(d\xi) =$ mass of U that is transformed by S^N and ends up in the band Δ between $W^i_{\xi_1}$ and $W^i_{\xi_2}$.

This choice determines uniquely ν,ρ. In fact assume for simplicity that Δ is infinitesimal so that the stable manifolds (whose slope changes in a Hölder continuous fashion) can be regarded as parallel and we can consider that a surface element $d\sigma$ on $W_x^{\delta,i}$ and its image under the map ζ on $W_y^{\delta,i}$ have equal area. We can therefore denote both as $d\sigma$: *i.e.* we imagine that σ is a horizontal coordinate. Then the heuristic argument leading to equation (5.5.4) shall say that the ratio $\frac{\rho_\xi(\sigma)}{\rho_{\xi'}(\sigma)}$ is precisely given by (5.5.4).

Hence, fixed Δ and the value of $\rho_\xi(\sigma)$ in an arbitrary point $x=(\xi,\sigma)$ in Δ (and on the support of μ) one determines the values of ρ in the points $x'=(\xi,\sigma')$ and $y=(\xi',\sigma)$ by

$$\frac{\rho_\xi(\sigma)}{\rho_{\xi'}(\sigma)} = \prod_{j=-\infty}^{\infty}\frac{e^{-\lambda_i(S^{-j}x)}}{e^{-\lambda_i(S^{-j}y)}}, \quad \text{and} \quad \frac{\rho_\xi(\sigma)}{\rho_\xi(\sigma')} = \prod_{j=0}^{\infty}\frac{e^{-\lambda_i(S^{-j}x)}}{e^{-\lambda_i(S^{-j}x')}} \qquad (5.5.6)$$

where the second relation simply expresses that if $d\sigma$ and $d\sigma'$ are surface elements of equal area on the *same* portion $W_x^{\delta,i}$ then the mass around them came from two slices of the vicinity U of O of sizes proportional to $S^{-N}(d\sigma)$ and $S^{-N}(d\sigma')$. The (5.5.6) determine uniquely $\rho_\xi(\sigma)$ in Δ up to a factor (because the value of ρ in the selected point x is left undetermined).

Considering (5.5.6) for each of the various choices of Δ one finds that in reality there is only the arbitrariness of a *global* factor because, being it possible to cover the attractive set by sets Δ, the arbitrary factors in each Δ are all determined except one. The latter is however determined by the normalization condition that the total μ–measure of the attractive set should be 1. Hence (5.5.5),(5.5.6) determine ν, ρ in each Δ that we wish to consider.

The heuristic considerations that led to definition 4 induce, therefore, also the following conjecture, [Ru80],[Ru89]:

Conjecture *(natural probability distribution): If (M, S) is a dynamical system (in the sense of §5.3) the SRB distribution associated with an attractive set A and describing the statistics of almost all points in the vicinity of A exists and is unique; it is a probability distribution μ such that*

(i) for μ–almost all points of A it makes sense to define a regular (here we mean "smooth") local stable manifold and regular local unstable manifold.

(ii) μ is absolutely continuous on the unstable manifold of each point $x \in A$ and its restriction to an infinitesimal region Δ like the one shown in Fig. (5.5.3) has a density that verifies the (5.5.5) and (5.5.6) provided reasonable *hypotheses hold (!).*

Remarks:

(1) The necessity of "reasonable hypotheses", and the precise meaning to give to the notions that intervene in the conjecture, is due to the otherwise easy construction of counterexamples, *c.f.r.* problems. Which could such hypotheses be is a part of the problem posed by the conjecture. *If A is an attractive set that verifies axiom A and S is mixing on A the conjecture holds true,* (Sinai, Ruelle, Bowen)*, hence the name "*SRB *distribution".* Furthermore, in this case, μ is such that the dynamical system (A, S, μ) is ergodic, mixing and each observable has continuous power spectrum (in the sense described at the beginning of §5.2).

(2) Hence if A is an attractive set verifying axiom A (*c.f.r.* definition 2 of §5.4), and if S is mixing on A then a SRB distribution exists and the problem is completely solved, because one can also show that it coincides with the unique invariant distribution μ verifying (5.5.5).

(3) If the dynamical system (M, S, μ) is regular in the sense of definition 3 of §5.4 and A is an attractive set on which the invariant distribution μ is concentrated, then by theorem I of §5.4 the property (i) of the conjecture holds: for this reason, [ER81], one sometimes adopts a definition of SRB distribution, *different from the one that we set and that we shall use here*, calling SRB distribution for (M, S) every distribution μ such that (M, S, μ) is a regular dynamical system (*c.f.r.* definition 5 in §5.4) with μ absolutely continuous along the unstable manifold. The regularity requisite, *i.e.* the condition $(\log ||T||)^+ \in L_1(\mu)$ of problem [5.4.11], is also posed in a weaker form compared to our definition 3 of §5.4.

In connection with the observation (2) on attractive hyperbolic sets it is interesting to mention that, for such attractive sets, other properties hold which one thinks are valid possibly more generally. We quote the example

II Theorem *(periodic orbit representation of SRB distributions): If A is an attractive set verifying axiom A the SRB distribution defined on it can be computed as follows. Let* $\mathrm{per}_n(A)$ *be the set of the points $x \in A$ that generate a periodic motion with period n and let $\Lambda_i(x) \equiv \det \partial S^n(x)|_{W^i(x)}$, i.e. the determinant*[5] *of the matrix ∂S^n thought of as a map acting on the unstable direction $V_i(S^{-n/2}x)$ (with values in $V_i(S^{n/2}x)$).*
Let us define

$$\langle f\rangle_n = \frac{\sum_{x\in \mathrm{per}_n(A)} \Lambda_i(x)^{-1} f(x)}{\sum_{x\in \mathrm{per}_n(A)} \Lambda_i(x)^{-1}} \tag{5.5.7}$$

then, for each regular observable f, it is

$$\lim_{n\to+\infty} \langle f\rangle_n \equiv \int_A \mu(dx)\, f(x) \tag{5.5.8}$$

where μ is the SRB distribution.

Remarks:

(1) Hence we see, again, that the logarithm of the local expansion coefficient $\log \Lambda_i(x)$ around a point x on a time n, plays the role of the energy function in statistical mechanics, the “observation” time n plays the role of volume of the dynamical system. The limit $n \to \infty$ plays the role of the “thermodynamic limit”.
(2) By this remark we see, in an alternative way with respect to what already noted in remark (2) to definition 1, that it is possible to formulate several qualitative properties of the dynamical system temporal averages by using ideas and methods of statistical mechanics. Usage of this analogy gives rise to the so called “thermodynamic formalism” for describing motions on strange attractive sets.
(3) The relation (5.5.7) is an immediate consequence of the theory of the SRB distribution, [Si72], [Si77], and has been very closely studied in [Ru79] (*c.f.r.* [Ga81], see in particular p.254 Eq. (21.18). For a further discussion see the following §5.7, (D)).
(4) The (5.5.7), (5.5.8) follow from Sinai's theory based on Markovian pavements (*c.f.r.* §5.7, and [Ga81] §21, XLIII): but in the original works this fact is so obvious that it has not been explicitly underlined. Nevertheless equation (5.5.7) “filtered” outside (of the theory) and it continues, surprisingly, to be “rediscovered”: it is known as the *periodic orbits development* of the SRB distribution.

One should attribute the popularity of this representation of the SRB distribution to the fact that periodic orbits are a very simple concept, even

[5] This will also be called the “expansion coefficient” of the area element on $V_i(x)$.

elementary if compared with geometric notions on which the theory of Sinai is based (*c.f.r.* §5.7). However it should be stressed that it is precisely upon notions of Sinai's theory "judged" more involved ("Markov pavements", *c.f.r.* §5.7), that ultimately, at least in the well understood cases, rests the validity of (5.5.8) *which is not equivalent* to the complete theory (discussed in §5.7).

(C) Attractive sets and attractors. Fractal dimensions.

At this point it is convenient to introduce the notion of *attractor* and of its *fractal dimension*.

Let (M, S) be a dynamical system with an attractive set A.[4] The definition 1 in §5.4 hints at the existence on A of several S–invariant probability distributions , μ which can be thought of as statistics of various distributions μ_0 defined on the basin U of attraction of A (see problems below for actual examples).

Such distributions μ are usually ergodic and, hence, to say that μ and μ' are not identical means to say that $\mu(N) = 0$ and $\mu'(N) = 1$ for some (S–invariant) Borel set $N \subset A$. In this case the distributions μ_0 and μ'_0, of which μ and μ' are the statistics, have the same attractive set but with a different attractor. Hence we give a formal definition

2 Definition *(attractors and information dimension):*
Given a regular dynamical system (M, S) and a probability distribution μ_0 on M we shall say that A_{μ_0} is an attractor *for (M, S, μ_0) if μ_0 has a statistics μ and*
(i) A_{μ_0} is invariant.
(ii) A_{μ_0} has μ–probability 1: $\mu(A_{\mu_0}) = 1$.
(iii) A_{μ_0} has fractal Hausdorff dimension d_I (c.f.r. §3.4, (A)) which is minimal between those of the sets that verify (i,ii). The dimension d_I is called information dimension *of* (M, S, μ_0) and it will be denoted $d_I(\mu_0)$ when it will be necessary to stress its dependence on the distribution μ_0.

Remarks:

(1) A_{μ_0} is *not*, in general, unique. For instance if the individual points of A_{μ_0} have zero μ–probability and $x \in A_{\mu_0}$ then the set obtained by taking out of A_{μ_0} the entire orbit of x still enjoys the properties (i, ii, iii).
(2) If μ_0 is a probability distribution that attributes probability 1 to a basin U of attraction for an attractive set A (*i.e.* "if μ_0 is concentrated on the basin of attraction of an attractive set A") then A_{μ_0} can be chosen to be contained inside A.
(3) An attractor for (M, S, μ_0) is, in general, *not a closed set.*

[4] *i.e.* a closed invariant set such that for all points x of a neighborhood of it U it is $d(S^n x, A) \xrightarrow[n\to+\infty]{} 0$, and that furthermore does not contain subsets with the same properties, *c.f.r.* definition 1 in §5.3.

(4) The notion of attractive set is natural for dynamical systems in which M is regular and S is at least continuous. The notion of attractor for (M, S, μ_0) makes instead sense, and is interesting, also when M and S have singularities on sets of zero μ_0–measure. And even if M is much more general than a manifold: for instance if M has infinite dimension or is just a metric space, *c.f.r.* [DS60], p. 174.

(5) In general if A is a set that has probability 1 for a probability distribution μ we say that μ is *concentrated* or *has support* on A. Hence one can say that *an attractor for* (M, S, μ_0) *is an invariant set of minimal Hausdorff dimension on which the statistics of* μ_0 *can be concentrated*.

(6) The heuristic discussion leading to the conjecture in (B) above and to equation (5.4.6) allow us to establish another interesting notion of fractal dimension related to, but different from, that of Hausdorff dimension.

Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_l$ be the Lyapunov exponents, counted with their multiplicities, of an ergodic dynamical system (M, S, μ) in which μ is the SRB distribution on an attractive set $A \subset M$. Then equation (5.4.6) shows that the volume elements of dimension d with d such that $\lambda_1 + \lambda_2 + \ldots + \lambda_d > 0$ "generically" expand and, hence, we could expect that an attractor has information dimension (*in general different from that* d_H *of* A) at least d.

If $\lambda_1 + \lambda_2 + \ldots + \lambda_{d+1} \leq 0$ the information dimension of the dynamical system "should be" $\leq d + 1$ and if ε is such that $\lambda_1 + \lambda_2 + \ldots + \lambda_d + \varepsilon\lambda_{d+1} = 0$ then the information dimension should be $d_L = d + \varepsilon$: a heuristic argument, of restricted validity, can be found in [KY79]. It is an interesting problem to find conditions implying that $d + \varepsilon$ is really equal to the information dimension.

Since it can differ from the information dimension d_I, the quantity $d_L = d + \varepsilon$ is called *Lyapunov dimension* of the dynamics S observed with the distribution μ on the attractive set A: the notion is due to Kaplan and Yorke, *c.f.r.* [ER81], and sometimes it is denoted $d_L(\mu)$ to stress its dependence on the distribution μ. See (G) in §6.2 and problems to §6.2 for an interesting technique to bound the dimension d_L of an attracting set which has remarkable applications to the Navier–Stokes equation or to the Rayleigh equations (*c.f.r.* §1.5) for convection.

(7) It is known that the Lyapunov dimension d_L of an invariant ergodic distribution μ on an attractive set A for an analytic dynamics S on an analytic manifold M is not smaller than the information dimension d_I of (M, S, μ), [Le81], *whether* μ *admits a SRB statistical distribution or not*: $d_I \geq d_L$ (*Ledrappier inequality*). Furthermore if the dimension of M is 2 and if μ is absolutely continuous on the unstable manifold then the Lyapunov dimension of (M, S, μ) and the information dimension coincide, [Yo82], (*Young theorem*). See problems [5.7.9]÷[5.7.11] of §5.7 for a heuristic analysis of the theorem of Young.

In absolute generality it is easy to find counterexamples to the equality $d_I = d_L$, *c.f.r.* problem [5.5.10], other examples can be found in [ER81]. Therefore it will be always convenient to think as *distinct* the notions of information dimension and of Lyapunov dimension.

Appendix: The theory of Pesin.

The following theorem, *c.f.r.* [Os68],[Pe76],[Pe92], holds

I' Theorem *(existence of stability manifolds and Lyapunov exponents for general dynamical systems):*
If (M,S) is an invertible dynamical system and (M,S,μ) is a μ–regular ergodic dynamical system (c.f.r. definitions 1 in §5.3, and 5 in §5.4):
(i) There is a set X with $\mu(X)=1$ whose points admit a dynamical base for S.
(ii) The Lyapunov exponents $\lambda_1>\lambda_2>\ldots>\lambda_s$ for S and their multiplicity are constants on X and the exponents of S^{-1} are opposite to those of S.
(iii) If for $x\in X$ no Lyapunov exponent vanishes and if $V_s(x)$, $V_i(x)$ are the stable and unstable planes for S, then such planes are integrable in the sense that exist $\delta(x)>0$ and two portions of regular surface, denoted $W_x^{\delta(x),s}$ and $W_x^{\delta(x),i}$, that contain x and are contained in the ball of radius $\delta(x)$ centered at x and
(1) At each of their points $y\in X$ there is a dynamical base and the planes $V_s(y)$ and $V_i(y)$ are tangent respectively to $W_x^{\delta(x),s}$ and $W_x^{\delta(x),i}$.
(2) if $y\in W_x^{\delta(x),s}\cap X$ or $y\in W_x^{\delta(x),i}$ it is $\delta(y)\ge\frac{1}{2}\delta(x)$.
(iv) If all the nonpositive Lyapunov exponents of $x\in X$ are $<-\lambda<0$ then for each $y\in W_x^{\delta(x),s}$ it is $d(S^nx,S^ny)\le C(x)e^{-\lambda n}$ for $n\ge0$, likewise if all the nonnegative exponents are $>\lambda>0$ then for each $y\in W_x^{\delta(x),i}$ it is $d(S^{-n}x,S^{-n}y)\le C(x)e^{-\lambda n}$ for $n\ge0$, with a suitable $C(x)<\infty$.

Remarks:
(1) The property (iv) allows us to extend the definition of the *global stable manifolds* $W_s^g(x)$, *c.f.r.* observation (1) to theorem I, simply by defining it as the collection of points y such that $d(S^nx,S^ny)\xrightarrow[n\to+\infty]{}0$: however in general such global manifolds will not be globally smooth and will contain singularity points. Analogously we can define the *unstable global manifolds* $W_i^g(x)$. The manifold $W_i^g(x)$ is sometimes dense on the support of μ and the theorem can be used to construct the intuition on the nature of an attractive set in this more general situation, in analogy with what said in the case of the attractive sets verifying axiom A, see the heuristic remark (2) following definition 1 above.
(2) Also the correspondence between stable manifolds via their intersections with the unstable manifolds, *c.f.r. (iii)* of the Theorem I and Fig. (5.5.2), Fig. (5.5.3), is generalizable to the cases of the theorem I'. But the fact that in general the local manifold can have size $\delta(x)$ variable as a function of x renders a precise formulation somewhat heavy and we do not discuss it explicitly.
(3) The reason at the root of the validity of this theorem is that on the one hand regular points get close to the singularities of S or S^{-1} (*c.f.r.* observation (2) to definition 3 of §5.4) with "at most polynomial" speed while, on the other hand, the properties of the stable and unstable manifolds

are derived on the basis of their exponential contraction. Hence on regular points "things go" as if the dynamical system had no singularity.
(4) The proof of this theorem is the basis of a rather satisfactory extension of the theorem II to dynamical systems (M, S) regular with respect to the volume measure μ_0 in the sense of the definition 3, §5.4, *c.f.r.* [Pe92]. The theory in [Pe92] extends (5.5.7) and mainly the (5.7.4),(5.7.8) following it (of which the (5.5.7) is a consequence): *c.f.r.* Ch. VII, §7.1÷§7.4.

Problems

[5.5.1]: Consider a pendulum and its separatrix in phase space. Show that it is possible to modify the equations of motion away from the unstable fixed point so that the separatrix attracts the points of its neighborhood, but so that the motion on the separatrix remains unaltered. Show that the attracting set thus obtained (for the evolution timed at constant time intervals, for instance) is not hyperbolic (find at least a reason different from the absence of periodic dense points). Furthermore show that on the attracting set there is a probability distribution which describes the statistics of almost every point near it but it is not absolutely continuous on the unstable manifold (*i.e.* on the separatrix). (*Idea:* The invariant distribution is simply a Dirac delta on the unstable equilibrium point.)

[5.5.2]: Modify the example in problem [5.5.1] to build an attractive set that does *not* have a SRB distribution. (*Idea:* Imagine the circle on which the pendulum rotates as "double", *i.e.* the angle varies between 0 and 4π so that the dynamical system acquires two fixed unstable distinct points, but "related by the same separatrix" then, because of the same mechanism seen in problem [5.3.7], the time spent near each of the two unstable fixed points is not a well defined fraction of the total elapsed time T, not even in the limit $T \to \infty$.)

[5.5.3]: Consider the the example in equation (5.5.1) and write $S = S_0 \times 2^{-1}$ where S_0 is the arnoldian map defined on the torus T^2 by the first two relations on the r.h.s. of equation (5.5.1)). Assuming that the distribution $\mu = dx\,dy/(2\pi)^2$ is not the only invariant ergodic distribution for the map S_0 which gives positive probability to all open sets of T^2, see §5.7 for examples, construct an example of an attractive set which, with respect to two methods of random choice of the initial data, generates different attractors, with different information dimensions. (*Idea:* Let $\mu' \neq \mu$ be another invariant distribution; then the random choices of initial data with distribution $\rho(x, y, z)\,dx\,dy \times dz$ lead, for any choice of $\rho > 0$ to a statistics $\mu(dx\,dy) \times \delta(z)\,dz$ while random choices of initial data with distribution $\rho(x, y, z)\,\mu'(dx\,dy) \times dz$ lead to a statistics $\mu'(dx\,dy) \times \delta(z)\,dz$. The information dimension will be $d_I(\mu) = 2$ in the first case and $d_I(\mu')$ in the second and, see §5.7 for examples, μ' can be so chosen that $d_I(\mu')$ is as small as wished.)

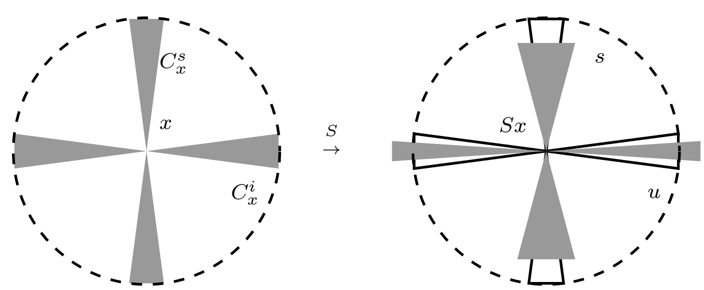


Fig. (5.5.4): *Symbolic representation of the cone property of the map S for the stable (vertical) manifold and for the unstable (horizontal) manifold.*

[5.5.4]: Let (M, S) be a hyperbolic dynamical system, in the sense of the definition 2 of §5.4, then it is possible to define in the tangent space T_x of each point x two cones[6] of tangent vectors C^α_x, $\alpha = s, i$ of opening $\rho > 0$ and axes (R^s_x, R^i_x) or, respectively, (R^i_x, R^s_x) such that there is an integer $m > 0$ and a coefficient $\kappa > 1$ for which

$$\begin{aligned} &\partial S^m C^i_x \subset C^i_{S^{\pm m}x} \\ &\partial S^m C^i_x \oplus C^s_{S^m x} = T_{S^m x} \\ &\|\partial S^m \xi\| \ge \kappa \|\xi\|, \qquad && \xi \in C^i_x \\ &\|\partial S^m \xi\| \le \kappa^{-1} \|\xi\|, \qquad && \xi \in C^s_x \end{aligned} \tag{!}$$

for each $x \in M$. Interpret the above figure as an illustration of the definition

[5.5.5]: Given the dynamical system (M, S) suppose that X is an invariant closed subset and that in its vicinity S, S^{-1} are invertible and C^∞. Suppose that in each point $x \in X$ the tangent space T_x is decomposed as $T_x = E^+_x \oplus E^-_x$ (with a not necessarily invariant decomposition) and that there exist functions $\rho^\pm(x)$ (not necessarily continuous) and $\kappa > 1$, $m > 0$ and two cones C^+_x and C^-_x of axes (E^+_x, E^-_x) and (E^-_x, E^+_x), and opening $\rho^+(x)$ and $\rho^-(x)$, respectively. If the cone property (!) in [5.5.4] holds then we say that the dynamical system (M, S) is *approximately hyperbolic* on X, *c.f.r.* [Ru89b] p. 95.

[5.5.6]: If (M, S) is approximately hyperbolic, in the sense of [5.5.5], on X then the set X is hyperbolic in the sense of definition 2 of §5.4. See [Ru89b] p. 125 and p. 126 for a simple and rapid proof.

[5.5.7]: If the dynamical system dynamical (M, S) is regular with respect to the volume μ_0 in the sense of the definition 3 of the §5.4 and if X is an invariant set of regular points on which the properties approximate hyperbolicity property holds in the sense of [5.5.5] but *without the condition that* X *is closed* then for each $x \in X$ there exist planes R^s_x and R^i_x such that $T_x = R^i_x \oplus R^s_x$ that verify (5.4.1). Prove this along the ideas of [5.5.6].

[5.5.8]: Consider a "filled" bidimensional torus M, *i.e.* a filled cylinder with the bases identified (commonly called a "doughnut", to be thought uncooked in the construction that follows). Imagine to compress by a factor $\lambda > \sqrt{2}$ the section and to scale, at the same time, the circumference with a factor 2. We obtain a doughnut long and narrow that can be deformed until it acquires a form of "an eight" (without however superposing it, not even partially, to itself: the doughnut is therefore thought of as impenetrable) and hence imagine inserting it in the space already occupied by the doughnut at the beginning of the described stretchings. One thus defines a map S of M into itself. The attractor of this dynamics verifies the axiom A but it is not a regular surface. For reasons that escape me this classic example is called "*the solenoid*".

[5.5.9]: The solenoid of the problem [5.5.8] can be thought of as $C \times \{-1, 1\}^{Z_+}$, product of a circle $C = [0, 2\pi]$ times space $\{-1, 1\}^{Z_+}$ of the sequences $\underline{\sigma}$ of digits ± 1 identifying $(0, \underline{\sigma})$ and $(2\pi, \underline{\sigma}')$ with $\underline{\sigma}'$ suitably defined in terms of $\underline{\sigma}$. Show also that the Hausdorff dimension of the solenoid is $\alpha = 1 + \frac{\log 2}{\log \lambda}$. (*Idea:* Note that the section of the doughnut after a map S^n consists of 2^n circles of radius λ^{-n}, then proceed as in the analysis of the Hausdorff dimension of the Cantor set, *c.f.r.* §3.4, (A).)

[5.5.10] Consider the example (5.5.1) but with z replaced by $\underline{z} = (z_1, \ldots, z_k)$, $z_j \in [-1, 1]$ and with the map S defined by the arnoldian cat on (x, y) and by

$$z'_j = e^{-\varepsilon} z_j, \qquad \varepsilon > 0,\ j = 1, \ldots, k$$

for the other variables. Show that *all the invariant distributions* have the same exponents of Lyapunov namely $\lambda = \log \frac{1}{2}(3 + \sqrt{5})$ and $-\lambda$ with multiplicity 1 and $-\varepsilon$ with multiplicity k. Show also that the SRB distribution is absolutely continuous on the unstable

[6] A cone in x with axes (W_x, Y_x) and opening ρ is a set of tangent vectors of the form $w + y$ with $|y| < \rho |w|$ and $w \in W_x, y \in Y_x$ and $W_x \oplus Y_x = T_x$. We recall that the sum $+$ of two tangent planes denotes the tangent plane spanned by two addends while the sum $\oplus$ denotes that the two planes are also linearly independent.

manifold and $\mu(dxdydz_1\ldots) = \frac{dxdy}{(2\pi)^2}\prod_j \delta(z_j)dz_j$. Show also that the attractive set has Hausdorff dimension 2, the attractor has also information dimension 2, while the Lyapunov dimension is $2+k$ if $\varepsilon = \lambda/k$. (*Idea:* Lebesgue measure is absolutely continuous on the unstable manifold (and also on the stable) by Fubini's theorem.)

[5.5.11] Find an example of a smooth dynamical system with an attracting set A on which there are several different distributions $\mu, \mu', \ldots$ which are such that almost all points chosen near A with a distribution that gives positive probability to any open set have a statistics which is one among $\mu, \mu', \ldots$. (*Idea:* Consider the dynamical system in problem [5.5.3] and, using the notations of the suggestion to [5.5.3], let $\overline{\mu} = (\mu+\mu')/2$ and $\mu_0(dx\,dy\,dz) = \overline{\mu}(dx\,dy)\cdot dz$; then $\overline{\mu}$–almost all points in $M = T^2 \times [0,1]$ have a statistics which is, however, with μ_0–probability 1/2 the distribution $\mu \times \delta(z)dz$ and with μ_0–probability 1/2 it is the distribution $\mu' \times \delta(z)dz$.)

Bibliography: [Si94],[Ru79], [Ga81], [Ru89b] p.94. For (5.5.8) see for example [Ru78], (7.20). For the detailed analysis of various notions of dimension *c.f.r.* [Pe84]. See [ER81] and [Pe92], [Ru89b] for the problems.

§5.6 Ordering of Chaos. Entropy and complexity.

Another important qualitative property of chaotic motions is the their *complexity*: this is a notion that can be made quantitative and leads to interesting new dynamical ideas.

(A) Complete observations and formal symbolic dynamics.

Given a regular metric dynamical system (M,S,μ) with S a μ–regular map, in the sense of the definition 5 of §5.4, we can imagine the following "model of observations": let $\mathcal{P} = (P_0, P_1, \ldots, P_{n-1})$ be a "pavement" of the surface M (phase space) with n domains (*c.f.r.* footnote [1] in §5.3) which, pairwise, have in common at most boundary points (which we shall suppose "regular", *e.g.* piecewise C^∞). The interiors of the sets P_i, $i = 0,\ldots,n-1$, represent the "states" x for which the results of a certain observation yield a given value: if $x \in P_i$ then the result of the observation is i; and if x is common to two or more sets of $\mathcal{P}$ then it represents a state on which the observation is "not precise" and provides an ambiguous result.

One can, for instance, think that phase space is subdivided into regions consisting of points on which a certain regular function $f(x)$ bounded between 0 and $n+1$ takes values $f(x) \in [i, i+1]$ with $i = 0,1,\ldots,n-1$. If $P_i = f^{-1}([i,i+1])$, with $i = 0,\ldots,n-1$, and if $x \in P_i$ we shall say that the result of the observation is i.

With every point x we can associate its *S–history*, or simply its *history*, $\underline{\sigma}(x)$ *with respect to the observation* $\mathcal{P}$: it is the sequences $\underline{\sigma}(x) = \{\sigma_k\}_{k=0,\infty}$ such that $S^k x \in P_{\sigma_k}$. The history tells us in which set P_σ the point x can be found after a time k.

Since the elements of the pavement $\mathcal{P}$ are not necessarily disjoint (because of possible boundary points in common) it may be that we could associate several histories to the same x: but this can happen only if for some k the point $S^k x$ belongs to the boundary of two different elements of the pavement and hence it will usually be an exceptional event. It is better to think that the S–history is just not defined in such exceptional cases in which it would be ambiguous (see also definition 1 in §5.7).

The history of a point is a way to introduce coordinates that represent it in a *non conventional* way. One can imagine that the symbols σ_k, that appear in the history of the point x, are a *generalization* of the representation on base n of the value of a coordinate of the system, or even of all the coordinates necessary to describe the state of the system.

To clarify this concept it is useful to refer to an example: let $M = [0, 1]$, and let S be the map $Sx = 10\,x \bmod 1$. Then imagine dividing the interval $[0, 1]$ into 10 intervals $P_0 = [0, 0.1]$, $P_1 = [0.1, 0.2]$, $\ldots$, $P_9 = [0.9, 1]$. If $x = 0.\sigma_0\sigma_1\ldots$ is the representation in base 10 of $x \in M$ it is immediate to verify that a possible history of x on the just described pavement $\mathcal{P}$ is precisely $\sigma_0, \sigma_1, \ldots$. Hence we see in which sense the history of a point may be a way to describe by coordinates the point itself.

We also see that in this case few points can have ambiguous histories: they are "only" the points that in base 10 are represented by sequences of digits eventually equal to 0 or eventually equal to 9. Such points are a denumerable set (*although dense*).

If the map S is invertible the history of a point usually does not determine the point itself: in such case one introduces the *bilateral history i.e.* the sequence $\underline{\sigma} = \{\sigma_k\}_{k=-\infty,\infty}$ defined by $S^k x \in P_{\sigma_k}$ for $k \in (-\infty, \infty)$. The bilateral history may determine the point even though the unilateral history does not.

1 Definition *(complete observation, generating pavement):*
We say that an observation $\mathcal{P}$ is complete *with respect to S and to the S–invariant distribution μ if for μ–almost all points x the history $\underline{\sigma}(x)$, (bilateral if S is invertible and unilateral otherwise), is unique and determines uniquely x.*[1] *Equivalently we say that $\mathcal{P}$ is a generating pavement.*

Obviously an invertible map can also be considered just for positive times and as a non invertible map: hence is possible that the same pavement is a complete observation if the map is thought of as invertible, and it is not complete if, instead, the map S is only considered for times ≥ 0, *i.e.* it is considered as not invertible.

Not all maps admit generating pavements: for example the identity map $x \to x$ obviously *does not* admit one (unless M is a set with a finite number of points). But, at least if M is a surface (of finite dimension) and S is regular on M, the cases in which complete observations do not exist are very special.

Let (M, S, μ) be a regular metric dynamical system (*c.f.r.* definition 3 (c) of the §5.3) admitting a complete observation $\mathcal{P} = (P_0, P_1, \ldots, P_{n-1})$. It

[1] Or, in other terms, if a set N of zero μ–probability, $\mu(N) = 0$, exists such that if $x, x' \notin N$ and $\underline{\sigma}(x) = \underline{\sigma}(x')$ then $x = x'$ and, furthermore, if $x \notin N$ then the history of x on $\mathcal{P}$ is unique, *i.e.* in its motion x does not hit the boundaries of the elements of $\mathcal{P}$.

is clear that the action of S is represented in a trivial way in terms of the history of $x \in M$. If N is the set of zero measure outside which the history determines uniquely the point that generates it and is determined by it, then the history of Sx is related to that of x simply by $\sigma_k(Sx) \equiv \sigma_{k+1}(x)$. In other words in the invertible case the application of S is equivalent to the *translation* of the history by one unit to the left; in the non invertible case it is, instead, equivalent to the translation by one unit to the left followed by the *removal* of the first symbol.

Since the correspondence point–history leads to a universal and simple representation of the dynamics, namely a translation in a space of sequences, *it must be* in general very difficult to construct the *code* that associates with $x \in M$ its history on a partition $\mathcal{P}$ (by the conservation of difficulties).

The first interesting question is "which is the minimum number m of symbols necessary in order that a pavement with m symbols be possibly generating for a given dynamical system (M, S, μ)?".

This simple question leads to the concept of *complexity* or *entropy* of the dynamical system (M, S, μ). To introduce it we first define the the notion of *distribution of frequencies* of symbols of a history or, more generally, of a sequence of digits.

2 Definition *(frequencies of strings in a sequence of symbols):*
If $k \to \sigma_k$, $k = 0, 1, \ldots$, is a sequence of digits with $\sigma_k = 0, \ldots, n-1$ we say that $\underline{\sigma} = \{\sigma_k\}_{k\ge 0}$ has well defined frequencies if, for every finite string of digits $(\hat\sigma_0, \ldots, \hat\sigma_p)$, the limits

$$\begin{aligned}\nu(\hat\sigma_0, \ldots, \hat\sigma_p) = \lim_{N\to+\infty} N^{-1} \big(&\text{number of values of } h \text{ such} \\ &\text{that } \sigma_h = \hat\sigma_0, \ldots \sigma_{h+p} = \hat\sigma_p \text{ with } h \le N\big)\end{aligned} \tag{5.6.1}$$

exist, i.e. if the frequencies $\nu(\hat\sigma_0, \ldots, \hat\sigma_p)$ with which the string $(\hat\sigma_0, \ldots, \hat\sigma_p)$ "appears" in the sequence $\underline{\sigma}$ are well defined .

Remarks:
(1) Given a μ-regular *ergodic* dynamical system (M, S, μ), *c.f.r.* §5.4 definition 5 and §5.3 definition 5, and a complete observation (or a generating pavement) $\mathcal{P}$, we can define for μ–almost every point x the history $\underline{\sigma}(x)$ on $\mathcal{P}$. This is a consequence of Birkhoff's ergodic theorem, see [5.4.4]. Thus the family $\{\nu(\hat\sigma_0, \ldots, \hat\sigma_p)\}$ of the frequencies of the finite strings $(\hat\sigma_0, \ldots, \hat\sigma_p)$, with $\hat\sigma_i$ taking the values $0, 1, \ldots, n-1$, is well defined. Such frequencies are obviously constants of motion (because x and Sx have histories with equal frequencies) hence, by the assumed ergodicity, they must be μ–almost everywhere independent by x.
(2) Then by our definitions, in the ergodic case of the preceding remark, if we set

$$E^{\,0\ \ 1\ \ \ldots\ \ p}_{\,\hat\sigma_0\,\hat\sigma_1\,\ldots\,\hat\sigma_p} \equiv \bigcap_{j=0}^{p} S^{-j} P_{\hat\sigma_j} \tag{5.6.2}$$

we see that $E{\,}^{0\ 1\ \dots\ p}_{\hat\sigma_0\,\hat\sigma_1\,\dots\,\hat\sigma_p}$ is the set of points which visit the sets $P_{\hat\sigma_0},\dots P_{\hat\sigma_p}$ at times $0,\dots,p$, so that we should have

$$\nu(\hat\sigma_0,\hat\sigma_1,\dots,\hat\sigma_k)=\mu\big(E{\,}^{0\ 1\ \dots\ p}_{\hat\sigma_0\,\hat\sigma_1\,\dots\,\hat\sigma_p}\big) \tag{5.6.3}$$

(3) It can be shown, on the basis of the fundamental definitions of measure theory, that if $\mathcal{P}$ is complete (*c.f.r.* definition 1), then the measure μ is uniquely determined by the frequencies (5.6.3).
(4) If $\mathcal{P}$ is a complete observation, every function $x\to f(x)$ on M (which is μ–measurable) can be thought of as a function on the phase space $\mathcal{S}$ of the unilateral or bilateral sequences $\underline{\sigma}$ of digits in the non invertible case or, respectively, in the invertible case. Furthermore the remarks (2) and (3) show that the frequencies of a sequence $\underline{\sigma}$ of digits can be interpreted as a probability distribution $\mu_{\underline{\sigma}}$, on the phase space $\mathcal{S}$, invariant with respect to the translations τ. We just set

$$\mu_{\underline{\sigma}}\big(C{\,}^{0\ 1\ \dots\ p}_{\hat\sigma_0\,\hat\sigma_1\,\dots\,\hat\sigma_p}\big)=\nu(\hat\sigma_0,\dots,\hat\sigma_p) \tag{5.6.4}$$

where $C{\,}^{0\ 1\ \dots\ p}_{\hat\sigma_0\,\hat\sigma_1\,\dots\,\hat\sigma_p}$ is a *cylinder* in $\mathcal{S}$, *i.e.* it is the set of sequences of $\mathcal{S}$ such that $\sigma_0=\hat\sigma_0,\dots,\sigma_p=\hat\sigma_p$; or, more formally,

$$C{\,}^{0\ 1\ \dots\ p}_{\hat\sigma_0\,\hat\sigma_1\,\dots\,\hat\sigma_p}=\bigcap_{j=0}^{p}\tau^{-j}C{\,}^{0}_{\hat\sigma_j} \tag{5.6.5}$$

The definition in equation (5.6.4) can be extended by "additivity" so that a meaning is given to the probability $\mu_{\underline{\sigma}}(E)$ for every set E representable as finite or denumerable union of disjoint cylinders of the form (5.6.5). We thus obtain a probability distribution $\mu_{\underline{\sigma}}$ defined on the smallest family of sets containing the cylinders and closed with respect to operations of complementation, denumerable union and intersection (*c.f.r.* [DS58], p. 138).

On the space $\mathcal{S}$ of sequences of digits it is natural to introduce the following notion of convergence: we shall say that $\underline{\sigma}_n$ converges to $\underline{\sigma}$ as $n\to\infty$, writing $\underline{\sigma}_n \xrightarrow[n\to\infty]{} \underline{\sigma}$, if for every j it is *eventually* $(\sigma_n)_j\equiv\sigma_j$ (*i.e.* this relation holds for $n>n_j$ and a suitable n_j). The topology on $\mathcal{S}$ associated with this notion of convergence and the probability distribution on the cylinders (5.6.5) play the role of ordinary topology and of Lebesgue measure on the subintervals of $[0,1]$. Hence the probability distribution $\mu_{\underline{\sigma}}$ can be thought of as defined on the *Borel sets* of the natural topology of $\mathcal{S}$.

Thus given the ergodic dynamical system (M,S,μ) the frequencies of μ–almost all points in M define[2] a metric dynamical system $(\mathcal{S},\tau,\mu_{\underline{\sigma}})$, where

[2] In the invertible case also the probability of the cylindrical sets $C{\,}^{-p\ \dots\ 0\ 1\ \dots\ p}_{\hat\sigma_{-p}\,\dots\,\hat\sigma_0\,\hat\sigma_1\,\dots\,\hat\sigma_p}$ are needed to determine μ: such measures will be naturally defined by their probabilities $\mu_{\underline{\sigma}}(C{\,}^{0\ \dots\ p+1\ p+2\ \dots\ 2p}_{\hat\sigma_{-p}\,\dots\,\hat\sigma_0\ \ \hat\sigma_1\ \dots\,\hat\sigma_p})$ and this implies their translation invariance.

τ is the translation, which is "equivalent" to (M, S, μ): the equivalence (called *isomorphism*) is a notion that we shall not define formally and that, in this case, is established by the correspondence "point" $\longleftrightarrow$ "history on $\mathcal{P}$". Such correspondence associates with the cylinder $C^{\,0\ \ 1\ \dots\ p}_{\,\hat\sigma_0\,\hat\sigma_1\,\dots\,\hat\sigma_p}$ the set $E^{\,0\ \ 1\ \dots\ p}_{\,\hat\sigma_0\,\hat\sigma_1\,\dots\,\hat\sigma_p}$, (5.6.2).

3 Definition *(symbolic dynamics):*
(1) A symbolic dynamical system is a pair $(\mathcal{S}, \tau)$ with $\mathcal{S}$ the space of sequences of a finite number of symbols, τ is the translation of the sequences by one step (to the left).
(2) Given a sequence of n symbols $\underline{\sigma}$ with defined frequencies we shall define a symbolic metric dynamical system associated with the sequence $\underline{\sigma}$ as $(\mathcal{S}, \tau, \mu_{\underline{\sigma}})$, where $\mathcal{S}$ is the space of the sequences with n symbols, τ is the translation to left of the sequences and $\mu_{\underline{\sigma}}$ is the probability distribution generated, on the cylinders of $\mathcal{S}$, by the frequencies of the finite strings of digits (c.f.r. (5.6.1), (5.6.4)).

Remarks:
(1) The dynamical system receives the attribute of symbolic because in this case $\mathcal{S}$ is not a surface, but it is just a compact metric space (in the sense that from every sequence of points it is possible to extract a convergent subsequence) if we define the distance between two sequences as $d(\underline{\sigma}', \underline{\sigma}'') = 2^{-N}$ when N is the largest integer for which $\sigma'_i = \sigma''_i$ for $|i| \le N$: note that with this definition of distance the translation τ is a *continuous* map.[3] Compare this definition of symbolic metric dynamical system with the definition 4 of the §5.3 of metric dynamical system (of which it is a particular case), and with the definition 2 of the §5.3 of topological dynamical system (to which it adds the distribution $\mu_{\underline{\sigma}}$). It is therefore natural to attribute to sequences properties with the same name of properties that we have so far attributed to dynamical systems, *c.f.r.* definitions 5,6,7 of §5.3, as follows.

(2) A sequence $\underline{\sigma}$ is said *ergodic* if the dynamical system $(\mathcal{S}, \tau, \mu_{\underline{\sigma}})$ is ergodic, *i.e.* if every function $\mu_{\underline{\sigma}}$–measurable f such that $f(\underline{\sigma}') = f(\tau\underline{\sigma}')$, for $\mu_{\underline{\sigma}}$–almost all sequences $\underline{\sigma}'$, is necessarily constant.

(3) A sequence is said *mixing* if every pair of continuous functions f, g on $\mathcal{S}$ is such that

$$\Omega_{f,g}(n) = \int_{\mathcal{S}} \mu_{\underline{\sigma}}(d\underline{\sigma}')\, f(\underline{\sigma}') g(\tau^n \underline{\sigma}') \xrightarrow[n\to\infty]{} \langle f\rangle_{\underline{\sigma}} \langle g\rangle_{\underline{\sigma}} \tag{5.6.6}$$

where $\langle\cdot\rangle_{\underline{\sigma}}$ denotes integration with respect to $\mu_{\underline{\sigma}}$, *i.e.* $(\mathcal{S}, \tau, \mu_{\underline{\sigma}})$ is mixing.

[3] Here 2 can be replaced by an arbitrary number > 1. This notion of distance is very natural if we think to what it becomes in the case in which the sequence consists of two symbols only, $0, 1$, and σ_i, σ'_i with $i \ge 0$ are interpreted as the binary development of two real numbers x, x', while σ_i, σ'_i with $i < 0$ are interpreted as binary development of two numbers y, y'. We see easily that 2^{-N} is essentially $\max(|x - x'|, |y - y'|)$.

(4) A sequence $\underline{\sigma}$ *has a component with continuous spectrum* if there exists a function f such that the Fourier transform $\hat{\Omega}_f(k)$, $k \in (-\pi,\pi)$, of the correlation $\Omega_f(n)$ is locally a L_1-function for $k \neq 0$. We say that the system has *continuous spectrum* if this happens for all not constant functions.

(B) Complexity of sequences of symbols. Shannon–McMillan theorem.

For symbolic dynamical systems on a space $\mathcal{S}$ of sequences of finitely many symbols a property of *ergodic decomposition* holds, *c.f.r.* observations conclusive to §5.3: *i.e.* if $(\mathcal{S},\tau,\mu)$ *is not ergodic*, then μ–almost all points $s \in \mathcal{S}$ have an ergodic statistics and if $\mathcal{S}^e$ denotes the set of the points s with ergodic statistics μ_s, we can write

$$\int_{\mathcal{S}} \mu(dx)\, f(x) = \int_{\mathcal{S}^e} \nu(ds) \int_{\mathcal{S}} \mu_s(dx)\, f(x) \tag{5.6.7}$$

here ν is a suitable probability distribution on $\mathcal{S}^e$.

If an evolution S observed on a pavement $\mathcal{P}$ generates a sequence of digits $\underline{\sigma}$ as history of a point *randomly chosen* with an S–invariant distribution, then equation (5.6.7) shows that lack of ergodicity of $\underline{\sigma}$ is probabilistically impossible. For this reason we shall confine ourselves to considering the notion of complexity *only for sequences with ergodic statistics*, *c.f.r.* [Ga81].

We then set the following definition of *entropy* or *complexity* of an ergodic sequence

4 Definition *(complexity of a sequence): Given a sequence of digits $\underline{\sigma}$ with n symbols and with an ergodic statistics, consider the strings $(\hat{\sigma}_1,\ldots,\hat{\sigma}_p)$ that appear in $\underline{\sigma}$ with a positive frequency. Let us divide such strings in "ε–frequent" strings and "ε–rare" strings; the set $\mathcal{C}^{rare}_{\varepsilon,p}$ is a family of strings of p digits with frequencies whose sum ("total frequency") is $< \varepsilon$. Given $\varepsilon > 0$ there exist, in general, several ways of collecting the (at most n^p) strings of p digits into ε–rare and ε–frequent groups. Let $\overline{\mathcal{C}}^{rare}_{\varepsilon,p}$ be a choice for which the correspondent number of ε–frequent strings is minimal: denote such number of ε–frequent strings with $\mathcal{N}_{\varepsilon,p}$ (it is $\mathcal{N}_{\varepsilon,p} \leq n^p$). We define (when the limit exists)*

$$s(\underline{\sigma}) = \lim_{\varepsilon\to 0^+} \lim_{p\to\infty} \frac{1}{p} \log \mathcal{N}_{\varepsilon,p} \leq \log n \tag{5.6.8}$$

and call $s(\underline{\sigma})$ the entropy *or* complexity *of $\underline{\sigma}$.*

Remarks:
(1) Hence $s(\underline{\sigma})$ counts the number of strings of p digits "really existent" asymptotically as $p \to \infty$, *if we are willing to ignore* a class of strings of total frequency $< \varepsilon$.
(2) Obviously if s is the entropy of a book (voluminous, so that we can consider it as an infinite sequence of letters of the alphabet, spaces, punctuation

and accents included) its complexity is $s \le \log 73$.[4]

The *Shannon–McMillan theorem* establishes existence of the limit in (5.6.8) for any ergodic sequence; the theorem can also be refined by adding to it the statement that the frequent strings have "about" the same frequency (at equality of length); *i.e.* fixed $\varepsilon > 0$ and all $(\hat\sigma_1, \ldots, \hat\sigma_p) \notin \overline{\mathcal{C}}^{rare}_{\varepsilon,p}$ the frequency $\nu(\hat\sigma_1, \ldots, \hat\sigma_p)$ satisfies

$$e^{-(s(\underline{\sigma})+\varepsilon)p} < \nu(\hat\sigma_1, \ldots, \hat\sigma_p) < e^{-(s(\underline{\sigma})-\varepsilon)p} \tag{5.6.9}$$

for p large enough. The ergodicity of the distribution $\mu_{\underline{\sigma}}$ implies also that $\mu_{\underline{\sigma}}$–almost all strings have the *same statistics*, and hence the same complexity, of $\underline{\sigma}$ itself.

Furthermore if C is a *code* which codes sequences of digits $\underline{\sigma}$ into sequences of digits $\underline{\sigma}' = C\underline{\sigma}$, *i.e.* it is a map of $\mathcal{S} = \{0, \ldots, n\}^Z$ with values in $\mathcal{S}' = \{0, \ldots, m\}^Z$ which commutes with τ (translation of the digits of the sequences: $C\tau\underline{\sigma} \equiv \tau C\underline{\sigma}$; *i.e.* it commutes with "the writing" of the strings), then the dynamical system $(\mathcal{S}, \tau, \mu_{\underline{\sigma}})$ is transformed by C into the dynamical system $(\mathcal{S}', \tau', \mu')$ where μ' is the measure image of $\mu_{\underline{\sigma}}$ via the "change of coordinates" C. And the complexity of μ'–almost all sequences $\underline{\sigma}' \in \mathcal{S}'$ can be shown to be $\le s(\underline{\sigma})$, [AA68]. Such complexities are in fact equal if C is invertible.[5]

Hence by changing representation of the sequences, *i.e.* "by translating them into another language" *one cannot increase their complexity.* If the translation is perfect (*i.e.* C is invertible) then the complexity remains the same. This and other "thermodynamical" properties of the complexity explain why it is also called "entropy" or "information".

For instance we can interpret this remark, referring also to the preceding observation (2), as a formalization of the empirical fact that by simply translating a book from a language to another we cannot increase the information it contains. And if the new language has an alphabet of k symbols with $\log k$ smaller than the complexity s of the book, then *it will not be possible* to translate the book into one of equal length in the new language.

Nevertheless it will be possible to translate it into any alphabet, possibly at the cost of an increase in the text length: for instance into a binary alphabet (*i.e.* a two symbols alphabet as in digital books, by now common). The lengthening will be at least by a factor p'/p such that $2^{p'} = e^{ps}$: *i.e.* $p'/p = s/\log 2$.

---

[4] 73 letters: these are 25 lower case letters, 25 capitals, 10 numbers, 6 punctuation signs, 2 two accents, a sign for word splitting, a sign for blank space, two parentheses and a newline. If mathematics books are included imagining them digitized in TEXthen the number of necessary characters grows to almost all the 128 `ascii` characters.

[5] It is possible to give counterexamples to the natural guess that μ' is simply $\mu_{C\underline{\sigma}}$; it can even be that the sequence $C\underline{\sigma}$ does not have well defined frequencies. This pathology does not arise if the map C is a finite code, *i.e.* it is such that $(C\underline{\sigma}')_i$ is a function of $(\underline{\sigma}')_j$ only with $|j-i| < M$, where the "memory" M of the code C is an integer.

Since in general $s < \log n$ it may even be possible to reduce the length of a (long) text, written in an alphabet with n characters, by using the *same* alphabet: the reduction will be from a length p to p', but obviously it will never be a reduction by more than a factor $s/\log n$, *i.e.* $n^{p'} = e^{ps}$. For example the set of software manuals can be rewritten in a drastically shorter form or, a further example, the book obtained by merging all astrology books, having entropy $s = 0$, can even be translated into a text of length 1 consisting in the symbol of "space". For a general theory of the transmission of information see, [Ki57].

This shows the immediate informatic interest of the entropy notion; in fact researches dedicated to make more precise the notion of "best possible codification" led Shannon to the notion of complexity, or information or entropy of a sequence of symbols. We should therefore appreciate the theoretical interest of the notion: but it is *very difficult* to estimate the complexity of objects as structured as the English language or just the complexity of a long text.

A way to estimate it can be to examine the length of the codes that bring books into digital form, by now quite common also commercially. Such codes are written (some of them, at least) trying to take advantage of the redundancies of the language to reduce the space used by the text. Although the profit is not optimal, the length p' of the text binarily coded compared to the length p_0 of the same text "in clear form" will be such that $p' \log 2 > p_0 s$ so that $\frac{p'}{p_0} \log 2$ will provide an upper estimate of s.

Because of the great progresses of electronics, it has become "not really necessary" to try to code books in a very astute way: because they can be digitized via photographical procedures, quite inefficient from an informatics viewpoint (so that it will usually happen that $p' \log 2 \gg p_0 \log n$), but such that the entire Encyclopedia Britannica can be written on a few "compact disks". A codification that kept into account the redundancies in a more effective way, would not only require an important editorial work but mainly would require a major and very interesting study of the structure of the English language.

On the other hand it is possible to develop codes that result remarkably effective by profiting only of the simplest redundancies. For instance I quote the code called "*zip*" that, when applied to common texts (such as this book) reduces their length "by a factor $\sim \frac{4}{9}$". Indeed *zip* translates an ordinary English text (written in an alphabet of 73 characters, *c.f.r.* note [4]), into a text written in an *extended ascii* alphabet of 256 characters by reducing it by a factor approximately equal to 3. This means that if s is the entropy of the text then $s < \frac{1}{3} \log 256$ and hence if we retranslated[6] into the original alphabet with 73 characters we could reduce a text of length p by a factor $x = \frac{1}{3} \frac{\log 256}{\log 73} \simeq \frac{4}{9}$. But the zip algorithm (*c.f.r.* the file `algorithm.doc`

[6] This is not so strange an operation because about half of the 256 characters normally cannot be correctly transmitted by the simplest communication software: hence there are programs that retranslate "zipped" texts into texts with only the 128 ascii characters, like the unix *uuencode*.

that on appears *internet* together with the sources of zip) is quite simple (although very clever) and (hence) it does not take into full account the syntactic structure of English language.

(C) Entropy of dynamical a system and the Kolmogorov–Sinai theory.

Having introduced the notion of complexity of a sequence we should recall that sequences are generated, essentially always, as histories of the evolution of a point observed on a pavement $\mathcal{P}$, or better on a *partition* $\mathcal{P}$, of the phase space of a metric dynamical system (A, S, μ).

A partition $\mathcal{P} = (P_0, \ldots, P_n)$ differs from a pavement because we require that the sets P_i are pairwise *disjoint*. One obtains a partition from a pavement simply by "deciding" to which sets the points common to the boundary of two elements belong. But partitions are more general because we allow the possibility that their elements be just Borel sets, without any smoothness. Obviously the notion of history is set up in the same way as for pavements: but in this case the history of every point is unique. A partition will be called generating with respect to a map S and to a probability distribution μ if there exists a set N with $\mu(N) = 0$ such that the histories on $\mathcal{P}$ of the points outside of N determine uniquely the points themselves.

It is then natural to define the complexity of a metric dynamical system (A, S, μ) as the *largest complexity* of the sequences that can be obtained with nonzero probability as histories of a point on a arbitrary partitions $\mathcal{P}$,[7] *i.e.* as the largest complexity of a sequence that can be generated by the map S after selecting the initial datum at random with respect to the distribution μ.

Obviously we could fear that such complexity is ∞ because of the observation in the footnote [7] above; or because we can construct arbitrarily fine partitions which contain a large number of elements and, from the examination of only a single symbol of the history of a point on a very fine partition, one can obviously obtain "a lot of information" since one determines almost completely the point itself. Think of the extreme case in which one represents a book (of a large library) with a single symbol: the book itself so that the library is coded by its catalogue (a code understandable only by very learned readers).

The following theorem is, therefore, very important (Sinai), *c.f.r.* [AA68]:

---

[7] It is essential to exclude sequences generated by data with zero probability. Otherwise given a sequence $\underline{\sigma} = (\sigma_1, \ldots)$ generated by a Bernoulli scheme with N equally probable events and, hence, ergodic (*c.f.r.* problem [5.6.1] below) and with complexity $\log N$ the following unpleasant construction would be possible. Given a nonperiodic point x define P_j, $j = 1, \ldots, N$, as the set (at most denumerable) of the points $S^t x$ with t such that $\sigma_t = j$ and let P_0 be the set of the other points (hence of almost all points, if μ attributes zero probability to individual points). By construction the history of x on the partition so constructed is $\underline{\sigma}$: it has well defined frequencies, is ergodic, and it has complexity $\log N$, and N is arbitrary. We then see that the history in question is precisely excluded by the hypothesis that we only consider histories of points that have μ–probability not zero to be randomly chosen.

I Theorem *(maximal complexity): If (A, S, μ) is an ergodic dynamical system then μ–almost all points x generate, on a given partition $\mathcal{P}$, histories with a complexity $s(\mathcal{P})$ independent of x. The complexity $s(\mathcal{P})$ is called complexity of the partition $\mathcal{P}$ in the system (A, S, μ). Furthermore all generating partitions (if they exist) have equal complexity s and $s \geq s(\mathcal{P})$.*

Hence the sequences that can be obtained by observing the evolution of a point randomly chosen in A with distribution μ *cannot* have arbitrarily large complexity. In fact maximal complexity sequences are precisely obtained by using, to construct them, a generating partition (or a generating pavement if the boundary of the pavement elements have 0 μ–probability).

Hence it is not worth the effort to refine observations beyond a certain finite limit, because the information that we get out of them cannot increase, at least when the initial data are randomly chosen with distribution μ and if we are willing to perform *several* observations: in other words with several not too refined observations we can obtain the same information that we get with few very refined ones. This appears, if presented in this way, obvious. But it is interesting that by the notion of complexity and by the above theorems, we can put these statements into a quantitative form that eludes the difficulty mentioned in the footnote [7] and that, as noted, has the "flavor of thermodynamics".

Finally what can be said when the history $\underline{\sigma}$ of a point, *i.e.* the system $(A, S, \mu_{\underline{\sigma}})$, is not ergodic (while still having defined frequencies)? The following theorem (Kolmogorov–Sinai) holds

II Theorem *(average entropy): If (M, S, μ) is a metric dynamical system and $\mathcal{P}$ is a partition of A into μ–measurable sets consider the quantities $\mu(\cap_{i=0}^{p} S^{-j} P_{\hat{\sigma}_j})$, which if μ is ergodic are the frequencies of the strings of digits $\hat{\sigma}_0, \ldots, \hat{\sigma}_p$ in the history on $\mathcal{P}$ of almost all initial data. Then the limit*

$$s(\mathcal{P}, \mu, S) = \lim_{p\to\infty} -\frac{1}{p} \sum_{\hat{\sigma}_0,\ldots,\hat{\sigma}_p} \mu(\bigcap_{i=0}^{p} S^{-j} P_{\hat{\sigma}_j}) \log \mu(\bigcap_{i=0}^{p} S^{-j} P_{\hat{\sigma}_j}) \qquad (5.6.10)$$

exists and is called the average entropy *of the partition $\mathcal{P}$.*
Furthermore if $\mathcal{P}$ is generating $s(\mathcal{P}, \mu, S)$ takes the value of its lowest upper bound computed on all partitions $\mathcal{P}'$: $s(\mu, S) = \sup_{\mathcal{P}'} s(\mathcal{P}', \mu, S)$. This lowest upper bound is called the average entropy *of the dynamical system (M, S, μ), or the* Kolmogorov–Sinai invariant *of (M, S, μ). Finally, if (M, S, μ) is an ergodic system and $\mathcal{P}$ is generating, $s(\mu, S)$ coincides with the entropy of the histories on $\mathcal{P}$ of μ–almost all the points of A.*

Remarks:

(1) It is natural to define the entropy of a dynamical system, ergodic or not, as $s(\mu)$, supremum of the complexity of the histories that can be constructed by randomly extracting (with distribution μ) a point x and by following its

history on a partition $\mathcal{P}$. If (M, S, μ) is not ergodic the entropy and the average entropy are, in general different.

(2) This theorem provides us also with a method to compute complexity. For instance if $(\mathcal{S}, S, \mu)$ is a Bernoulli scheme, $(p_1, \ldots, p_n)$, *i.e.* it is the dynamical system in which $\mathcal{S}$ is the set of the bilateral sequences with n symbols, S is the translation to the left and μ is the probability distribution that assigns to every cylinder in $\mathcal{S}$ a probability equal to the product of the probability $(p_1, \ldots, p_n)$ of the n symbols, then

$$\mu(C\,{}^{0\ \ 1\ \ \ldots\ \ p}_{a_0\ a_1\ \ldots\ a_p}) = \prod_{i=0}^{p} p_{a_i} \tag{5.6.11}$$

It is an immediate consequence of the theorem that

$$s(\mu) = -\sum_i p_i \log p_i \tag{5.6.12}$$

(3) Let $(\mathcal{S}, S, \mu)$ be a Markov process (p_{ij}) with n states, *i.e.* $\mathcal{S}$ is the space of the sequences with n symbols, S is the translation to the left and p_{ij} is the probability of transition from the state i to the state j (hence such that $\sum_j p_{ij} = 1$), which we suppose to be a "mixing matrix", *i.e.* such that the matrix elements of a suitable power of it are all strictly positive: a Markov process with this property is called a *mixing Markov process*.

Then the probability of a cylinder is defined by

$$\mu(C\,{}^{0\ \ 1\ \ \ldots\ \ p}_{a_0\ a_1\ \ldots\ a_p}) = \pi_{a_0} \prod_{i=0}^{p-1} p_{a_i a_{i+1}} \tag{5.6.13}$$

where π_i, $i = 1, \ldots, n$ is such that $\sum_{i=1}^{n} \pi_i p_{ij} = \pi_j$ is the left eigenvector of the matrix p_{ij} with eigenvalue 1.

By applying the theorem we immediately see that:

$$s(\mu, S) = -\sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i\, p_{ij} \log p_{ij} \tag{5.6.14}$$

Problems.

[5.6.1] Show that a Bernoulli scheme $(p_1, \ldots, p_n)$, *i.e.* the dynamical system $(\mathcal{S}, \tau, \mu)$ on the space $\mathcal{S}$ of the bilateral strings of n digits in which the evolution τ is the translation of the sequence by one unit to the left and μ is defined by the probability of the cylinders, (5.6.11), is a metric (in the sense of definition 4 in §5.4) ergodic dynamical system. (*Idea:* Let $\chi_{\hat\sigma_1, \ldots, \hat\sigma_n}(\underline{\sigma})$ be the characteristic function of the cylinder in (5.6.5), *i.e.* of the set of the sequences of digits such that the digits with label $1, \ldots, n$ coincide with the digits $\hat\sigma_1, \ldots, \hat\sigma_n$. Define

$$\Delta_N(\underline{\sigma}) = N^{-1} \sum_{j=0}^{N-1} \chi_{\hat\sigma_1, \ldots, \hat\sigma_n}(\tau^j \underline{\sigma}) \;-\; \prod_{j=1}^{n-1} p_{\hat\sigma_j}$$

and note that, by the Birkhoff theorem, the limit as $N \to \infty$ of $\Delta_N(\underline{\sigma})$ exists with μ–probability 1 and in $L_1(\mu)$, *c.f.r.* problems [5.4.3], [5.4.4]. Then $\int \Delta_N(\underline{\sigma})^2\, d\mu$ is:

$$N^{-2} \sum_{j,j'=0}^{N-1} \int d\mu\, (\chi_{\hat\sigma_1, \ldots, \hat\sigma_n}(\tau^j \underline{\sigma}) \;-\; \prod_{j=1}^{n-1} p_{\hat\sigma_j})(\chi_{\hat\sigma_1, \ldots, \hat\sigma_n}(\tau^{j'} \underline{\sigma}) \;-\; \prod_{j=1}^{n-1} p_{\hat\sigma_j})$$

and if $|j-j'|>2n$ the integral is obviously 0 so that $\lim_{N\to\infty}\int\Delta_N(\underline{\sigma})^2\,d\mu=0$ and it follows that $\lim_{N\to\infty}\Delta_N(\underline{\sigma})=0$ with μ–probability 1. This means that μ–almost all sequences $\underline{\sigma}$ of $\mathcal{S}$ have defined frequencies and such frequencies are *independent* of the initial datum $\underline{\sigma}$.
By the density in $L_1(\mu)$ of the functions that are finite linear combinations of characteristic functions of cylinders it follows that the averages $\lim\frac{1}{N}\sum_{j=0}^{N-1}f(\tau^j\underline{\sigma})$ exist and are almost everywhere constant. Hence every constant of motion is almost everywhere constant and the system is ergodic. See also problem [5.2.3].)

[5.6.2]: Show that the Bernoulli schemes are dynamical systems with continuous spectrum. Show the same for the Markov processes which are transitive (*i.e.* with a compatibility matrix $M_{\sigma\sigma'}$ with an iterate with all the elements of matrix positive). (*Idea:* Compute the correlation function between two cylindrical functions, *i.e.* between two functions that depend only on a finite number of digits of the sequences $\underline{\sigma}$ and apply the definition 1 of the §5.2).

[5.6.3]: Show, by applying directly definition 4, that the complexity of a sequence $\underline{\sigma}$ with defined frequencies and such that $\nu(\hat\sigma_1,\ldots,\hat\sigma_r)=\prod_{i=1}^r p_{\hat\sigma_j}$, *c.f.r.* (5.6.1), is $s=-\sum_i p_i\log p_i$. Hence the complexity of a almost any sequence chosen with the distribution of a Bernoulli scheme is given by (5.6.12) (one also says that a "typical" sequence generated by a Bernoulli scheme has complexity S). (*Idea:* First study the probability, in a bernoulli scheme with two symbols $0,1$ with probabilities p and $1-p$, of the string with k symbols 0, which is $\binom{n}{k}p^k(1-p)^{n-k}$, showing that it is bounded, for all $0\le k\le n$ and for all n large, above or below ($\pm$ respectively) by

$$n^{\pm1/2}e^{-(s(k/n)+(k/n)\log p+(1-k/n)\log(1-k/n)\log(1-p))n}$$

where $s(x)=-x\log x-(1-x)\log(1-x)$: this is a consequence of Stirling's formula for evaluating the factorials in $\binom{n}{k}$. Since the function in the exponent has a maximum in k at $x=k/n=p$ with second derivative $p^{-1}(1-p)^{-1}$ it follows that the total probability of the strings such that $|k/n-p|>\delta$ is bounded above by

$$n\sqrt{n}\exp\left(-2^{-1}p^{-1}(1-p)^{-1}\delta^2+O(\delta^3)\right)n$$

Fix $\varepsilon,\delta>0$ and n so that the latter expression is $<\varepsilon$. Then the optimal decomposition $\mathcal{C}^\varepsilon_{rare},\mathcal{C}^\varepsilon_{freq}$ must be such that all $\mathcal{C}^\varepsilon_{freq}$ are strings with $|k/n-p|\le\delta$. Hence their number is $\le\sum_{|k/n-p|\le\delta}\binom{n}{k}\le\exp(s(p)+O(\delta))n$. But since the probability of each such string is bounded below by $e^{-(s(p)-O(\delta))n}$ their number cannot be less than $e^{(s(p)-O(\delta))n}$ if δ is small enough, hence $s=s(p)$. The general case of Bernoulli schemes with more than 2 symbols is reduced to the case just treated.).

[5.6.4]: As problem [5.6.2] but for a sequence $\underline{\sigma}$ with defined frequencies as in the (5.6.13), *i.e.* "typical for a Markov process".

[5.6.5]: Consider a differentiable dynamical system (M,S) with dimension of M equal to d and let $\mu(dx)$ be the volume measure and suppose that μ is S–invariant. Let $\mathcal{P}$ be a partition obtained from a pavement $\mathcal{P}=(P_0,\ldots,P_n)$ with domains with piecewise regular boundary ∂P_σ by deciding (arbitrarily, because we shall see that this is irrelevant) to which element belong the points common to two boundaries. The boundary of $\cap_{j=0}^{k-1}S^jP_{\sigma_j}$ will consist of several elements $\mathcal{F}^j_{\sigma_0,\ldots,\sigma_{k-1}}\subset S^j\partial P_{\sigma_j}$ with, at most, parts of their boundaries in common. The union of all these boundary elements is $\cup_{j=0}^{k-1}S^j\cup_\sigma\partial P_\sigma$. Assuming that $\lambda=\max_{x,\xi,\delta=\pm1}|\partial S^\delta_x\xi|/|\xi|$ is the largest coefficient of expansion of the infinitesimal vectors ξ tangent in x (*i.e.* the largest expansion of the line elements), show that for a suitable constant C:

$$\sum_{j=0}^{k-1}\text{area}\,(\mathcal{F}^j_{\sigma_0,\ldots,\sigma_{k-1}})\le C\frac{\lambda^k}{\lambda-1}$$

(*Idea:* Note that the largest expansion of a $(d-1)$–dimensional surface element is $\le \gamma\lambda$ (*and not* λ^{d-1}) for a suitable constant γ because the volume is by hypothesis conserved).

[5.6.6]: With reference to the definition 4 and in the context of the preceding problem consider, fixed $\eta > 0$, the class $\mathcal{C}_1(k)$ of the strings $i \equiv \sigma_0, \ldots, \sigma_{k-1}$ such that

$$p_i \stackrel{def}{=} \mu(\cap_{j=0}^{k-1} S^{-j} P_{\sigma_j}) > e^{-k\eta}\lambda^{-dk}$$

and show, exploiting [5.6.5], that the set $\mathcal{C}_2(k)$ *complementary* of $\mathcal{C}_1(k)$ has total probability $X \le C' e^{-\eta k/d}$ for some constant C'. (*Idea:* By the isoperimetric inequality in R^d the volume of a region E with fixed surface $|\partial E| = \text{area}(\partial E)$ is largest for the sphere and, hence, $\mu(E) \le \Gamma |\partial E|^{d/(d-1)}$. Then one notes that:

$$\begin{aligned}
X &= \sum\nolimits_{i\in\mathcal{C}_2(k)} \mu(\cap_{j=0}^{k-1} S^{-j} P_{\sigma_j}) \equiv \sum\nolimits_{i\in\mathcal{C}_2(k)} p_i = \\
&= \sum\nolimits_{i\in\mathcal{C}_2(k)} \big(\mu(\cap_{j=0}^{k-1} S^{-j} P_{\sigma_j})\big)^{\frac{d-1}{d}+\frac{1}{d}} \le \\
&\le \sum\nolimits_{i\in\mathcal{C}_2(k)} \Big(\big(\Gamma |\partial \cap_{j=0}^{k-1} S^{-j} P_{\sigma_j}|\big)^{\frac{d}{d-1}}\Big)^{\frac{d-1}{d}} e^{-\eta k/d}\lambda^{-k} \le C' e^{-\eta k/d}).
\end{aligned}$$

[5.6.7]: In the context of [5.6.5], [5.6.6] show that a differentiable dynamical system (M, S, μ) with S a map that conserves μ = volume measure on M cannot have entropy larger than $d \log\lambda$ if λ is the largest expansion coefficient of the line elements (*Kouchnirenko theorem*), (*c.f.r.* [Ga81] p. 127). (*Idea:* Let $i \equiv (\sigma_0, \ldots, \sigma_{k-1})$ and $p_i = \mu(\sigma_0, \ldots, \sigma_{k-1})$; then

$$\begin{aligned}
&-\sum\nolimits_i p_i \log p_i = -\sum\nolimits_{i\in\mathcal{C}_1(k)} p_i \log p_i - \sum\nolimits_{i\in\mathcal{C}_2(k)} p_i \log p_i \le \\
&\le (d\log\lambda + \eta)k + X\big(\sum\nolimits_{i\in\mathcal{C}_2(k)} \tfrac{-p_i}{X} \log \tfrac{p_i}{X}\big) + \log X \le \\
&\le (d\log\lambda + \eta)k + X\, k\, \log(n+1) + X\, \log X
\end{aligned}$$

and, via the definition 4 and the theorem II, use the arbitrariness of η).

Bibliography: [Ki57], [AA68], [Ga81] §12. Further applications of the Shannon–McMillan theorem and remarkable refinements of the theorem of Kouchnirenko (the formula of Pesin and the theorems of Ledrappier–Young) will be found among the last few problems of §5.7.

§5.7 Symbolic dynamics. Lorenz model. Ruelle's principle.

Coming back to the question, posed before definition 2 of the §5.6, about the necessarily difficult construction of a code describing the history of points of a dynamical system on a generating pavement $\mathcal{P}$, we shall discuss some cases in which this is possible.

(A) Expansive maps on $[0,1]$. *Infinity of the number of invariant distributions.)*

The simplest case is provided by the *expansive maps of the interval* $I = [0,1]$. Such maps are not invertible and have a singularity in their derivative (at least a discontinuity); they are defined by a function f that is regular on the intervals $[0, a_0], [a_0, a_1], \ldots, [a_{n-2}, 1]$, with $0 \equiv a_{-1} < a_0 < \ldots < a_{n-1} \equiv 1$, except at their extremes and transforms each of these intervals into the *whole* interval $[0, 1]$.

Let us suppose that f is expansive, *i.e.* that $|f'(x)| > \gamma > 1$ for a certain $\gamma > 1$, and that in every portion (a_i, a_{i+1}) the function f can be extended to a regular function on *all* $[a_i, a_{i+1}]$.

If we draw a graph of this function we see that its derivative should be discontinuous at the points a_i, extremes of the intervals of the pavement $\mathcal{P}$ of I with the sets $P_0 = [0, a_0]$,$P_1 = [a_0, a_1]$,.., $P_{n-1} = [a_{n-2}, 1]$.

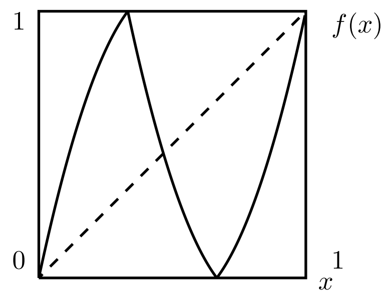


Fig. (5.7.1): *A continuous expansive map of the interval* $[0, 1]$.

Therefore the function f consists of n regular functions f_i, with f_i defined and regular on $[a_{i-1}, a_i]$ and with the absolute value of the derivative larger than $\gamma > 1$, see Fig. (5.7.1).

We shall suppose that $f(x) = f_i(x)$ if $x \in [a_{i-1}, a_i)$ and $f(1) = 1$ and that the f_i are of class $C^\infty([a_i, a_{i+1}])$. At the points a_i it can happen that the function f is discontinuous (*i.e.* $f_i(a_i) = 0$ and $f_{i+1}(a_i) = 1$ or viceversa). The pair (I, f) is a *dynamical system* in the general sense of definition 1 in §5.3.

It is convenient to examine in some detail a simple example. Consider, for instance, $f(x) = 10\, x \bmod 1$ defined by setting $a_i = \frac{i}{10}$ and $f_i(x) = 10\,(x - a_{i-1})$ for $a_{i-1} \le x < a_i$ and $f(1) = 1$: the points a_i with $0 < i < 10$ are discontinuity points. In this case the history of a point is simply the sequence of the digits of its decimal representation $x = 0.\sigma_1\sigma_2\sigma_3\ldots$. Note that there exist exceptional points whose history is ambiguous because in their evolution they fall on one among the a_i; according to the conventions of §5.6 the history of such points will be ambiguous.[1]

[1] By the choices made in defining f at the extremes of the intervals (and the property $n \bmod 1 = 0$) it is clear that if one insists in defining the history also for points which in their evolution visit the $a_i's$ then sequences eventually equal to 9 are not possible, except the sequence consisting just of 9's that represents 1. This happens in spite of the fact that they can be thought of as a decimal development of a number and represent a point of the interval: the right extreme points of the intervals have histories that could be also represented, defining differently the value of f on the singularity points, with sequences eventually equal to 9, for example 0.1 could have as history $0.0999\ldots$. This explains why it is inconveniient to define the histories of the exceptional point.

It is clear that *all histories* except a denumerable infinity, *i.e.* all sequences $\underline{\sigma}$ with 10 symbols except a denumerable infinity of them, correspond in a one to one way to a point of I. The correspondence is one to one between the sequences that are neither eventually 0 nor eventually 9, and the points of the subset of the interval obtained by taking away from it the set D of the "decadic" numbers, that in base 10 have a representation with period 0 (*i.e.* are represented by a finite number of significant digits), or with period 9.

One realizes that the set of exceptional points, obviously denumerable, consists of the points of $[0,1]$ such that $10^k x \bmod 1$ is one of the points a_i for some integer k ("decadic points").

If we select the initial data randomly with a distribution that attributes zero probability to all individual points, hence in particular to all denumerable sets, we can imagine that the decadic points "do not exist", as far as the study of motions that we shall be able to observe is concerned. Then the correspondence point–history would be one to one and described by a simple code.

It is interesting to remark that starting with (I, f) as above, *i.e.* $f(x) = 10\,x \bmod 1$, one can construct a metric dynamical system (I, f, μ) where μ is the uniform distribution on I (*Lebesgue measure*), *c.f.r.* definition 3 (c) of the §5.4. Indeed the inverse image of an arbitrary segment (hence of a measurable set) E has the *same length* of the initial set.[2]

We can ask what becomes of the uniform distribution μ, once "coded" on the space of the histories by identifying a point x with its history $\underline{\sigma}(x)$: it is immediate to check that it simply becomes the unilateral Bernoulli scheme with 10 symbols, each with probability 1/10. Hence to follow the history of a point randomly chosen with distribution μ produces a history that is not distinct from the sequence of the results of the tossings of an *equitable dice* with 10 faces. We conclude that the map f well deserves the title of "chaotic". It could even be used to generate random numbers.[3]

The latter remark also makes clear that the dynamical system (I, f) admits *several* other invariant distributions. For instance by thinking of the points as represented by the their histories on $\mathcal{P}$ (*i.e.* in decimal representation, and hence by identifying I with the space of the unilateral sequences with 10 symbols) one can define on the space of the sequences the Bernoulli distribution $\tilde{\mu}$ in which the symbols have probability $p_0, \ldots, p_9$ different from 1/10 (*unfair dice*). For example $p_i = 1/9$ for $i \neq 1$ and $p_1 = 0$.

This example shows that one can define on I *infinitely many f–invariant distributions*: one of them is the uniform distribution, and the others are

---

[2] Because it consists of 10 small segments each long 1/10 of the initial segment.

[3] This is, in reality, illusory because our representation in base 10 of the reals is such that any number we arbitrarily choose can necessarily be considered decadic, *i.e.* represented by a finite number of not zero decimal digits followed by infinitely many 0's, and hence after a (small) number of iterations of the map the history terminates in a boring sequence of 0's. Here we see the deep difference between a "theoretically" good random number generator, such as $x \to 10\,x \bmod 1$, and a really good generator, that obviously *should not have* the property of becoming rapidly trivial.

not uniform and they are, rather, concentrated (*c.f.r.* definition 2, §5.5, observation (5)) on sets of zero length and fractal dimension that can be < 1. In the just given example the code transforms the Bernoulli measure into a measure $\tilde{\mu}$ on I concentrated on the set of the numbers that do not have 1 in the their decimal representation and that therefore gives measure 1 to a Cantor set whose Hausdorff dimension is $\log 9/\log 10$, *c.f.r.* §3.4.

Bernoulli schemes, as metric dynamical systems, are ergodic *i.e.* they generate with probability 1 sequences with ergodic frequencies, *c.f.r.* problem [5.6.1]. Therefore a sequence generated as a f–history on $\mathcal{P}$ starting from an initial datum randomly chosen with a distribution on I that is image of a Bernoulli scheme via the decadic code, is an ergodic sequence (*c.f.r.* definition 3 of the §5.6, observation (2)). Furthermore we can compute its complexity by appealing to the theorems of §5.6.

For example by randomly extracting a datum with uniform distribution μ we produce sequences with complexity $\log 10$, while extracting them with the second distribution introduced above, $\tilde{\mu}$, we shall obtain histories with lower complexity: namely with complexity $\log 9$.

Obviously other invariant distributions are possible: it suffices to define on the space of sequences an arbitrary[4] translation invariant distribution and then transform it, via the decadic code, into a distribution on I. For example one can consider any Markov process with 10 states, *c.f.r.* §5.6, and thus obtain a new invariant distribution on I.

What just said does not substantially change if the function f is an arbitrary expansive map, *i.e.* if the expansivity is not constant (10 in the preceding example), *c.f.r.* problems for an essentially complete theory.

From the viewpoint of symbolic dynamics there is not much difference between the dynamics $10\,x \bmod 1$ and any other expansive dynamics. The correspondence between point and history will be always one to one with a denumerable infinity of exceptions (related to the fact that the intervals $[a_i, a_i+1]$ have extremes in common). Bernoulli schemes, Markov processes and other distributions on the space of sequences (that assign zero probability to each single sequence) will be coded into invariant probabilities on I (that assign zero measure to all points a_i and to their images and inverse images with respect to f).

Among the invariant distributions there will be one (*c.f.r.* problems) which has a density (not uniform, in general) while all the others are concentrated on sets of zero length, except trivial cases like the probability distributions that have the form $\mu' = \alpha\mu + (1-\alpha)\nu$ where μ is a distribution with density while ν does not have a density, and $0 < \alpha < 1$ (note that this can only happen in cases in which μ' is not ergodic).

The fact that not all points have a (well defined) history on the pavement $\mathcal{P}$ is clearly "unpleasant" and it is due to the fact that a pavement is not

[4] But still attributing 0 probability to single points to avoid problems due to the decadic numbers.

a partition: this is not really a problem as the points for which a history cannot be defined (being ambiguous as they fall, in their evolution, on boundaries common to more than one element of the pavement) usually form invariant sets of 0 probability with respect to the random choices of inititial data that one is interested to consider. For a system (M, S) *without* singularities it is however convenient to introduce the notion of "*compatible sequence* with respect to a pavement $\mathcal{P}$", defined without exceptions:

1 Definition *(compatible sequence): Let* (M, S) *be a dynamial system without singularities. A sequence* $\underline{\sigma}$ *is* S*–compatible with the pavement* $\mathcal{P}$ *of* M *if* $\cap_i S^{-i} P_{\sigma_i} \neq \emptyset$. *We shall say that any point* $x \in \cap_i S^{-i} P_{\sigma_i}$ *has the sequence* $\underline{\sigma}$ *as "possible history".*

Thus, if $\mathcal{P}$ is generating under S, to every point x we can associate a compatible sequence and if μ is a S–invariant distribution which attributes 0 probability to the boundaries ∂P_σ of the elements of the pavement there is one and only one compatible sequence (*i.e.* possible history) associated with each point with the exception of a set N of points with zero probability. A S–compatible sequence coincides with the S-history of the point x that it determines if x outside the exceptional set N.

(B) Application to the Lorenz model.

An important application of the above observations has been developed to show that in the Lorenz model, after the second bifurcation, in which time independent points eventually lose stability, *c.f.r.* §4.4, a motion appears that is chaotic.

Observing motions with timing events given by the successive local maxima of the coordinate z (*c.f.r.* §4.1), Lorenz remarked, in fact, that "with good approximation", after an initial transient, the value of the coordinate x *determined* the coordinate x' corresponding to the next observation (and the same happened for the y coordinate). Hence, calling x_{min} and x_{max} the minimal and maximal values of the coordinates x at the times of the observations, the evolution could be modeled by a map f of the interval $[x_{min}, x_{max}]$ into itself. And from the numerical data obtained from his experiment he could draw the graph of this function f, *c.f.r.* [Lo63].

The graph, after a suitable rescaling and translation to transform the interval $[x_{min}, x_{max}]$ in $[0, 1]$, turns out to be that of an expansive map based on 2 intervals, *i.e.* with $n = 2$, and "essentially" similar to the map whose graph has the form of a"tent": $x \to 2x$ if $x \in [0, \frac{1}{2})$ and $x \to 2(1 - x)$ if $x \in [\frac{1}{2}, 1]$, see Fig. (5.7.2).

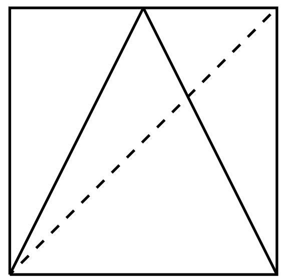


Fig. (5.7.2): *The graph of the "tent map".*

Hence he deduced that the system possessed chaotic motions; in fact it could be thought of as a generator of sequences of random numbers, as described in (A) above (*c.f.r.* however §5.3).

(C) Hyperbolic maps and markovian pavements.

The examples considered in (A) of maps generating chaotic motions concern maps (of the interval $[0,1]$) that are not invertible. Analogous examples can also be obtained for invertible maps. The paradigmatic case is provided by the map S of the torus T^2 defined by the matrix $g = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$, considered in §5.2: $S\underline{\varphi} \equiv g\underline{\varphi} \mod 2\pi$ ("Arnold cat"). In this case we note that the map is hyperbolic in the sense of §5.4, point (c) of definition 2. From every point $\underline{\varphi}$ of T^2 come out the straight lines parallel to the eigenvectors $\underline{v}_\pm$ of g: they have the interpretation of stable and unstable manifolds of the hyperbolic motion generated by g on T^2.

Such straight lines are dense on T^2 and they can be visualized as wrapped on the torus T^2 regarded as a square with periodic boundary conditions: the eigenvectors $\underline{v}_\pm$ have indeed irrational slope, *c.f.r.* (5.5.2). We shall denote $e^{\lambda_+}, e^{\lambda_-} \equiv e^{-\lambda_+}$ the corresponding eigenvalues.

Let us construct a pavement $\mathcal{P}$ of T^2, generating with respect to the dynamics S and to the area measure μ (*c.f.r.* definition 1, §5.6). The construction that follows is important because it can be repeated almost *verbatim* in the case of more general topologically mixing Anosov maps on 2–*dimensional manifolds*: the only minor difference will be that the stable and unstable manifolds of a fixed point will be curved manifolds and not simple straight lines.

We draw the portions of length L, L' of the stable manifold and of the unstable manifold of the origin (which is a fixed point) in the positive directions, for example (*i.e.* we "continue" the vectors $\underline{v}_\pm$).

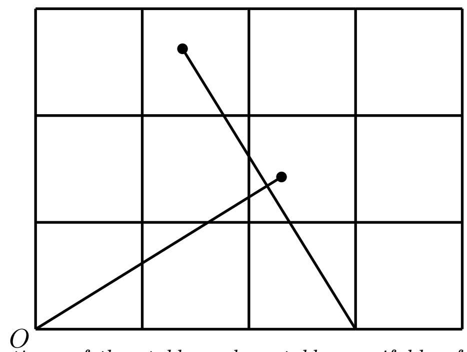


Fig. (5.7.3): *Portions of the stable and unstable manifolds of the Arnold's cat map drawn after representing T^2 as a lattice of copies of the square $[0,2\pi]^2$ with opposite sides identified.*

Since the manifolds, thought of as wrapped on the torus, are dense it is clear that by drawing them we delimit on T^2 a large number of small rectangles, see Fig. (5.7.4) below, of diameter that can be made as small as wished by taking L, L' large enough. In reality not quite all rectangles are "complete": indeed the origin and the extremes of the drawn portions of stable and unstable manifolds will end up ("surely" if drawn with a "randomly chosen" length) in the middle of two rectangles (or perhaps inside the same one) without reaching the side opposite to that of entrance.

In Fig. (5.7.3) the stable and unstable manifolds of the origin in the direction of the eigenvectors $\underline{v}_\pm$ of the matrix g are drawn. For clarity the figure has been drawn imagining the torus "unwrapped" and repeated periodically: but corresponding points on the various squares must be identified. Performing the identification we obtain the *not dashed* lines in Fig. (5.7.4).

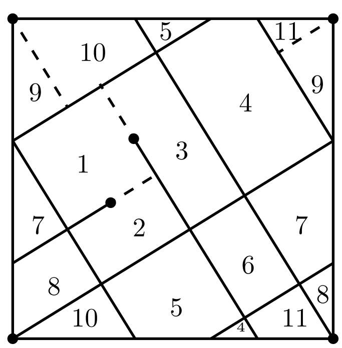


Fig. (5.7.4): *The continuation parts are marked as dashed and come out of the marked points; note that the four vertices of the square are in reality the same point (i.e. the origin) because of the periodic conditions that must be imagined at the boundaries.*

Then we imagine to continue the portion of stable manifold until it meets the side, opposite to the one already crossed, in the rectangle where it ends up. And then we imagine repeating the same operation on the portion of unstable manifold (the order of the operations is arbitrary even though the result depends, in general, on it as exemplified in Fig. (5.7.4)). We also continue the stable and unstable manifolds of the origin in the direction *opposite* to the eigenvectors $\underline{v}_\pm$ of g until they meets the already drawn portions of manifolds. This is illustrated in Fig. (5.7.4) above (see the four dashed lines) and it is unfolded in Fig. (5.7.6) at the end of the section.

We obtain in this way a pavement of the torus with elements P_i of diameter smaller than a prefixed quantity δ, provided we choose L, L' large enough.

The boundaries of the rectangles (thee are eleven rectangles in the above figure, if counting mistakes are avoided) are *bidimensional analogues of the points a_i of the expanding maps of the interval*. They form a set of measure zero with respect to the uniform distribution (area measure) on the torus.

Therefore, if we are only interested in the evolution of initial data randomly chosen with distributions that attribute zero measure to all segments in T^2, hence to the their denumerable unions, we see that every such initial datum has a well defined history that in turn determines it uniquely.

Indeed if two points had the same history σ_i for $0 \le i \le N$ their distance in the direction $\underline{v}_+$ would be smaller than the largest side δ of the rectangles *divided by the expansion* $e^{\lambda_+ N}$ that this side undergoes in N iterations of the map S (*i.e.* "in time N"). This is true if δ is small enough with respect to the side of T^2, *i.e.* with respect to 2π.[5]

Likewise if the two points have the same history for $-N \le i \le 0$, then the two points must have, in the direction $\underline{v}_-$, distance smaller than $\delta e^{-|\lambda_-|N}$.

The above remarks imply that if two points have the same history then they must coincide, provided δ is small enough (one finds, by performing the analysis in a formal way, that it is sufficient that $\delta e^{\lambda_+} < 2\pi/4$).

As discussed in §5.6 we should remark that there are points whose history is ambiguous (because $\mathcal{P}$ is a pavement and not a partition). Since the map is continuous it makes sense to consider S–compatible sequences: then to each point corresponds at least one compatible sequence in the sense of definition 1 above; and for almost all points with respect to the (invariant) area measure there is only one compatible sequence which coincides with the history of te point.

A difference with respect to the case of expansive maps of the interval is that now *not all sequences* with n digits, if n is the number of rectangles in $\mathcal{P}$, are compatible.

This can be understood by considering the case in which δ is very small. In this case it is clear that the S–image of a rectangle *cannot intersect* all other rectangles, simply because the side of its S–image will be still very short, even after the amplification of the factor e^{λ_+}. Hence symbolic dynamics corresponding to this pavement deals with sequences subject to important constraints, that cannot be eliminated by simply deciding to ignore a set

[5] We should indeed avoid the cases in which the rectangles can be so large to "wrap around the torus" allowing the image SQ, of a rectangle Q of $\mathcal{P}$, *to intersect another rectangle Q' in disconnected parts*: as it can be seen by a drawing (a little hard as one should take into account the periodic geometry), this is possible only if the rectangles are so large, that their images under the map S can even appear as disconnected when drawn on the torus represented as a square of side 2π).

of zero area consisting in a denumerable infinity of segments: for example if σ_i and σ_{i+1} are two symbols that follow each other in the history of a point x neither of which belongs to the boundaries of the P_σ nor to the images of such boundaries with respect to iterates of S, then it should be $SP^0_{\sigma_i} \cap P^0_{\sigma_{i+1}} \neq \emptyset$, if P^0_σ denotes the internal part of P_σ.

The latter remark leads us to the general notion of *S–compatibility by nearest neighbour* of a sequence with respect to a given pavement. We define, for this purpose

2 Definition *(compatibility matrix of symbolic dynamics):*
*Let $\mathcal{P} = (P_1, \ldots, P_n)$ be a pavement of phase space of the dynamical system (X, S) and let P^0_σ denote the internal part of $P_\sigma \in \mathcal{P}$. The matrix $M_{\sigma\sigma'}$ with matrix elements $M_{\sigma\sigma'} = 1$ if $SP^0_\sigma \cap P^0_{\sigma'} \neq \emptyset$ and $M_{\sigma\sigma'} = 0$ otherwise will be called "*compatibility matrix*" of the pavement $\mathcal{P}$ with respect to (X, S). A sequence $\underline{\sigma}$ will be called "*compatible by nearest neighbours*" if $M_{\sigma_i,\sigma_{i+1}} \equiv 1$ for $i \in (-\infty, \infty)$ (or, in the non invertible cases, for $i \in [0, \infty)$).*

A sequence $\underline{\sigma}$ can possibly be the history of a point randomly chosen with a distribution μ on T^2, that attributes measure zero to segments, only if it is compatible by nearest neighbours, *i.e.* only if the "compatibiliy condition" $\prod_i M_{\sigma_i\sigma_{i+1}} = 1$ holds.

If we decide to consider only histories compatible by nearest neighbours the ambiguity of the correspondence between points and their possible histories (see definition 1 above) is to a large extent eliminated: indeed the possible histories which are compatible with $\mathcal{P}$ are much less than the possible ones, in general.[6] Nevertheless there still remains an ambiguity which is usually *finite*: for instance in the case of the example of the expansive maps it is 2 at most, and in the case of the Arnold cat it is at most 4, (a simple check).

The general problem which we face when trying to define a digital code (that may also be called a *symbolic dynamics* representation) for the histories of points observed on given pavements is that, even if we only consider histories compatible by nearest neighbors, as it is natural and as we shall do from now on, *in general we shall have to impose many more compatibility conditions in order that a history compatible by nearest meighbors does really* correspond to the history of a point.

In fact nothing guarantees, in general, that the compatibility between symbols that immediately follow each other is sufficient, to generate compatible sequences that are really histories of some point. For example in order that the symbol σ can be followed by σ' and then by σ'' it will be necessary not only that the pairs $\sigma\sigma'$ and $\sigma'\sigma''$ are compatible but also that $S^2P_\sigma \cap P_{\sigma''} \neq \emptyset$, *etc.* Hence we should expect that there exist *infinitely many, further, compatibility conditions* involving arbitrary numbers of elements of the history.

[6] For instance if a periodic point x is in $\partial P_i \cap \partial P_j$ for some $i \neq j$ the number of possible histories is not denumerable but the number of histories possible *and* compatible by nearest neighbors does not exceed n and it is 2 if i, j are the only pair of elements of $\mathcal{P}$ that contain x.

Therefore in general, as we already expected, the "point–history" code will be unpractically difficult making the definition of Markov pavement for a dynamical system (X, S) interesting in spite of its very special character:

3 Definition *(Markov pavements): A generating pavement* $\mathcal{P}$[7] *for the dynamical system* (X, S) *such that every sequence compatible by nearest neighbours determines (uniquely) a point* $x \in X$, *i.e. if the condition* $\prod_i M_{\sigma_i \sigma_{i+1}} = 1$ *implies that* $\underline{\sigma}$ *is a possible history of* x, *will be called a markovian pavement for the dynamical system* (X, S)*: this is a rare case but for this reason an interesting one.*

An example of markovian pavement is provided by the expansive maps of $[0, 1]$ discussed in (A), in which *all* sequences are compatible by nearest neighbors,[8] and are possible histories in the sense of definition 1 (the compatibility matrix is $M_{\sigma\sigma'} \equiv 1$).

A more interesting example is provided by the pavements, just constructed on the torus T^2, related to the Arnold cat map S above.

(D) Arnold cat map as paradigm for the properties of Markov pavements.

(1) To check the latter statement the key property to note is that if one applies S to one of the rectangles P_σ of the pavement $\mathcal{P}$ constructed in (C) above (see Fig. (5.7.4)) then the rectangle is deformed along the directions of the eigenvectors of the matrix g: namely it is dilated in the direction of $\underline{v}_+$ and compressed in the direction of $\underline{v}_-$, besides being naturally displaced elsewhere.

Consider the sides that have become shorter: they form a collection of segments parallel to $\underline{v}_-$ *each of which must necessarily be contained in the union of the sides of the rectangles of the pavement and parallel to* $\underline{v}_-$: because the union of all sides of the original pavement $\mathcal{P}$ parallel to $\underline{v}_-$ is *by construction* a *connected* portion (approximately[9] of length L) of the stable manifold of the origin. But the latter, under the action of S, will contract by a factor $e^{-|\lambda_-|}$ becoming a subset of itself (because the origin is a fixed point). And this just means that "no new boundaries parallel to $\underline{v}_-$ are created" if one collects the boundaries parallel to $\underline{v}_-$ of $S\mathcal{P}$ (*i.e.* the union of the boundaries parallel to $\underline{v}_-$ of $S\mathcal{P}$ is *entirely* contained inside the union of the boundaries of $\mathcal{P}$).

Analogously we consider the boundary parallel to $\underline{v}_+$ of a rectangle P_σ; acting on it with the map S^{-1}, we see that it is transformed into a subset of the union of all sides of the rectangles of $\mathcal{P}$ parallel to $\underline{v}_+$.

---

[7] *c.f.r.* definition 1 §5.6.

[8] And each of them determines uniquely a point in $[0, 1]$ which in turn, with the exception of a denumerable family of sequences, determines uniquely the sequence.

[9] Recall that the extremes might have undergone a small stretching, in the initial construction, by at most δ as represented by the dashed lines in Fig. (5.7.4).

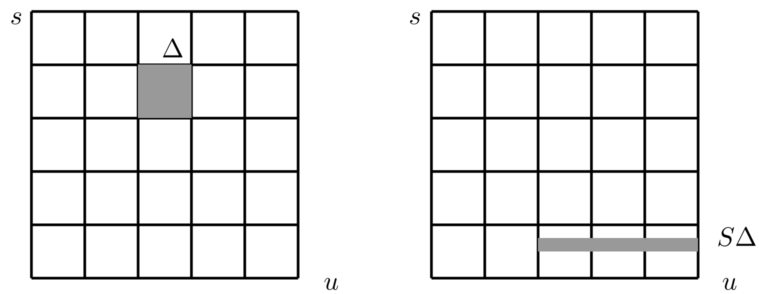


Fig. (5.7.5): *The figures illustrate, symbolically as squares, a few elements of a markovian pavement. An element Δ of it is transformed by S into $S\Delta$ in such a way that the part of the boundary that contracts ends up exactly on a boundary of some elements of $\mathcal{P}$.*

As illustrated in Fig. (5.7.5), if $M_{\sigma\sigma'}=1$, the image SP^0_σ of the interior P^0_σ of P_σ intersects $P^0_{\sigma'}$ hence, by the construction of $\mathcal{P}$, SP_σ *entirely crosses* $P^0_{\sigma'}$ in the sense that the image of every segment of P_σ parallel to the expansive side and with extremes on the "contracting bases" (*i.e.* parallel to $\underline{v}_-$) is transformed by S into a longer segment that cuts either *both bases* of $P_{\sigma'}$ or *none* (if it is external to $P_{\sigma'}$).

Were it not so the boundary parallel to $\underline{v}_-$ of SP_σ would fall in the interior of $P_{\sigma'}$ and hence it would not be contained in the union of the boundaries parallel to $\underline{v}_-$ of the rectangles of the original pavement $\mathcal{P}$. This can be well realized by trying to draw what described above, see the following figure.

An analogous geometric property holds for S^{-1}, exchanging the roles of $\underline{v}_+$ and $\underline{v}_-$.

(2) It is now sufficient a moment of thought to understand that this means that, if a sequence verifies the compatibility property by nearest neighbors then *it is really* a possible history of a point x and of only one point. Furthermore supposing that S^kx visits, for at least one k, the boundaries of the elements of the pavement one realizes that only finitely many different possible histories compatible by nearest neighbours can be associated with the same point x: in the example considered above (cat, see Fig. (5.7.4)) we see that there are at most 4 possible histories compatible by nearest neighbors: furthermore it is possible to show that such points are representable symbolically by sequences in which every element consists of 2, 3 or 4 symbols σ subject to a compatibility constraint which is described by a suitable compatibility matrix, (*Manning theorem*). However the correspondence between point and history will be one to one apart from a set of zero area.

(3) The above remark (1) also means that the set of the points whose history (or possible history) between $-N'$ and N coincide is, geometrically, a small rectangle. Furthermore if $\sigma_{-N'},\ldots,\sigma_N$ is the history in question and if δ^s_σ denotes the side parallel to $\underline{v}_-$ of P_σ and δ^e_σ denotes the side parallel to $\underline{v}_+$, then the side parallel to $\underline{v}_-$ of this small rectangle is $\delta^s_{\sigma_{-N'}}e^{-N'\lambda}$ and

the one parallel to $\underline{v}_+$ is $\delta^e_{\sigma_N} e^{-N\lambda}$ where $\lambda = \lambda_+ = -\lambda_-$ are the exponents of expansion and contraction of the map S (opposite of each other because S conserves the area, as the determinant of g equals to 1).

(4) Hence the small rectangle $\cap_{k=-N'}^{N} S^{-k} P_{\sigma_k}$ coincides with the set of the points whose history between $-N'$ and N is $\sigma_{-N'}, \ldots, \sigma_N$; and it has area

$$\delta^s_{\sigma_{-N'}} \delta^e_{\sigma_N} e^{-(N+N')\lambda} \prod_{i=-N'}^{N-1} M_{\sigma_i \sigma_{i+1}} \tag{5.7.1}$$

which is automatically zero in case of incompatibility.

Since performing the union over the values of the label σ_N (or of the label $\sigma_{-N'}$) of the sets $\cap_{k=-N'}^{N} S^{-k} P_{\sigma_k}$, one gets $\cap_{k=-N'}^{N-1} S^{-k} P_{\sigma_k}$ (or, respectively, $\cap_{k=-N'+1}^{N} S^{-k} P_{\sigma_k}$) it must be

$$e^{-\lambda} \sum_{\sigma} M_{\sigma\sigma'} \delta^e_{\sigma'} \equiv \delta^e_{\sigma}, \qquad e^{-\lambda} \sum_{\sigma'} \delta^s_{\sigma'} M_{\sigma'\sigma} \equiv \delta^s_{\sigma} \tag{5.7.2}$$

i.e. the sides parallel to $\underline{v}_+$ and $\underline{v}_-$ of the rectangles of $\mathcal{P}$ *can be interpreted as components of the right or left eigenvector,* respectively, *with eigenvalue* e^{λ} *of the matrix* M.

(5) The density on the torus T^2 of the stable and unstable manifolds of the origin implies the existence of a power k such that M^k has all matrix elements positive. And this simply means that every rectangle of $\mathcal{P}$ is so stretched, by the repeated action of S, in the expansive direction to eventually intersect the internal parts of *all* other small rectangles of the initial pavement $\mathcal{P}$. Hence, from the elementary theory of matrices, we deduce (*Perron–Frobenius theorem*) that the eigenvalue 1 of $e^{-\lambda}M$ is simple (*i.e.* the right and left eigenvectors are unique).

(6) Since the union of all the rectangles of $\mathcal{P}$ is the entire torus, it follows that $(2\pi)^{-2} \sum_\sigma \delta^e_\sigma \delta^s_\sigma = 1$. Hence we can imagine defining a square matrix, whose labels are the digits σ that distinguish the elements of $\mathcal{P}$, by setting $\pi_{\sigma\sigma'} = e^{-\lambda} (\delta^e_\sigma)^{-1} M_{\sigma\sigma'} \delta^e_{\sigma'}$ and $p_\sigma = (2\pi)^{-2} \delta^s_\sigma \delta^e_\sigma$ and then the (5.7.1) becomes

$$\mu(\cap_{k=-N'}^{N} S^{-k} P_{\sigma_k}) = p_{\sigma_{-N'}} \prod_{k=-N'}^{N-1} \pi_{\sigma_k \sigma_{k+1}} \tag{5.7.3}$$

showing us that *the uniform distribution* μ on the torus *is coded into a Markov process*, *c.f.r.* eq. (5.6.13), with states equal in number to the number of rectangles constituting the pavement $\mathcal{P}$ of T^2 and with transition probability given by $\pi_{\sigma\,\sigma'}$.[10]

---

[10] See (5.6.13), and note that by definition $\sum_{\sigma'} \pi_{\sigma\sigma'} = 1$ and p is an eigenvector of π, with eigenvalue 1, normalized to $\sum_\sigma p_\sigma = 1$.

(7) Hence the histories on $\mathcal{P}$ of points randomly chosen with uniform distribution have well defined frequencies which can be interpreted as the frequencies of the symbols of a *Markov process*. By applying formula (5.6.14) we see that the complexity s of the histories is λ (where e^λ is the expansion of the unstable direction: $\lambda \equiv \log[(3+\sqrt{5})/2]$).

We have already seen (*c.f.r.* [5.6.2]) that, with probability 1 with respectto a choice of the initial data with uniform distribution (*c.f.r.* §5.2), a sequence generated by a motion has a continuous spectrum.

Now we see that the motion is *chaotic* also in the sense that it is not distinct from a Markov process (for data chosen with probability 1 with respect to the uniform distribution μ). In other words *motions are not different from a sequence of tossings of a finite number of distinct dice* (with equal number of faces distinguished by some label σ which is used *also* to distinguish the dice), obtained by selecting for the successive tossing the dice σ, if σ is the result obtained at the last tossing, while the first extraction is instead performed with a distribution attributing to each face its frequency on a large number of extractions.[11]

(8) Naturally one can define several other Markov processes with the same states and the same compatibility matrix. Interpreted as probability distributions on T^2, via the *point–history* code, such Markov processes define invariant distributions *different* from the uniform distribution and not expressible via a density function. By making use of Markov processes that give zero probability to some symbol we shall get distributions concentrated on fractal sets of dimension lower than 2 (*i.e.* lower than that of the torus).

This is an important characteristic of "chaotic systems": they admit several, infinitely many in fact, invariant distributions which are interesting and not equivalent. This is, however, the same as saying that there exist very many really different motions.

(E) More general hyperbolic maps and their Markov pavements.

The constructions and remarks in (D) can be extended, without great difficulty to systems verifying *axiom A*, and that admit a fixed point, or a periodic orbit, with stable and unstable manifold dense on phase space (Anosov system) or on an attracting set ("topologically mixing attracting set that verifies axiom A").[12] See the definition 2 in §5.4.

All such systems admit Markov pavements $\mathcal{P}$, c.f.r. definition 2, with elements P_σ of diameter smaller than a prefixed quantity: this is an important theorem of Sinai, [Si70]

Its proof in the 2–dimensional case is a repetition, with obvious modifications, of the construction just discussed in the case of the arnoldian cat. One choses a fixed point (or, if none exists, a periodic point) and draws a long

[11] That can be defined by starting the choices with an arbitary dice and does not depend on which one starts with.

[12] The case of a periodic motion with period k is reduced easily to the case of a fixed point, by studying the map S^k instead of the map S itself.

portion of its stable and unstable manifolds completing the "rectangles" in a way analogous to the one illustrated in Fig. (5.7.4). For the general proof in arbitrary dimension see [Si70], [Bo70] and for an extension of the idea of the proof presented here for the 2–dimensional case to higher dimension, see [Ga95c].

Having seen, *c.f.r.* §4.1, the ubiquity of the chaotic systems we ask the question whether there exists a natural invariant distribution, among the many that they do possess.

Answering this question in systems more general than the Arnold cat greatly clarifies the nature of the problem because in the simple case of the cat map the phase space volume is invariant and ergodic, hence it is trivially its own statistics. This is no longer so simple in more general systems because not only volume is no longer invariant but in general no invariant distribution μ exixts which is absolutely continuous with respect to the volume μ_0. So the very existence of an SRB distribution, *i.e.* a distribution that is the statistics of all data apart from a set of zero volume (in the basin of the attracting set) is *a priori* not clear. The result states existence and uniqueness, already mentioned several times in previous sections, of the statistics of the volume measure μ_0.

Consider a dynamical system (M, S) verifying axiom A with an *attracting set* A on which S acts in a topologically mixing way. which *is not necessarily invariant* with respect to S and that is concentrated on the basin of attraction of the attracting set A.

Then we can ask whether μ_0 has a statistics on A, *i.e.* if A is a *normal attracting set* for μ_0 in the sense of definition 1 of §5.4:

I Theorem *(variational principle for SRB distribution): Suppose that (M, S) verifies axiom A and that A is an attracting set on which S is topologically mixing. Suppose that the measure μ_0 has a density*[13] *with respect to the volume measure on the basin of attraction of A.*
(1) There exists a statistics μ for μ_0–almost all points of the basin of attraction of A; and it is a statistics independent on the initial points (with μ_0–probability 1). Such statistics, that we called SRB in §5.5, will generate a dynamical system (A, S, μ) which is ergodic, mixing, with continuous spectrum, and isomorphic to a Bernoulli scheme and
(2) The distribution μ verifies a variational principle (Ruelle's principle) being the unique S–invariant distribution making largest the following function defined on the set $\mathcal{M}$ of the S–invariant probability distributions ν on A

$$s(\nu) - \int_A \nu(dx)\,\log \Lambda_e(x) \tag{5.7.4}$$

where $s(\nu)$ is the average entropy of the distribution ν with respect to S (c.f.r. (5.6.10)) and $\Lambda_e(x) = |\det(\underline{\partial} S)_e|$ is the determinant of the Jacobian matrix of the map S considered as restricted to the unstable manifold (a

[13] Usually this is expressed also by saying, *if μ_0 is absolutely continuous with respect to the Lebesgue measure.*

quantity whose logarithm is occasionally called the "sum of the local unstable exponents of Lyapunov").

Remarks
(1) This theorem or, better, the existence (for the systems considered in the theorem) of markovian pavements (on which its proof rests), is the basis of the proof of the theorem II of §5.5. And it is the basis of the conjecture that, *apart from exceptional cases*, initial data chosen randomly with an absolutely continuous distribution on the basin of an attracting set, even if not hyperbolic, admit "in cases of physical interest" a statistics that is obtained by solving a variational problem like (5.7.4), [Ru80].

(2) Applications of the conjecture are difficult: but if it is coupled with its natural consequence given by theorem II described in the following point (F), it has nevertheless an applicative value that in some cases seems remarkable, [GC95a],[GC95b], and in the future it might reveal itself to be quite important and even emerge as a new universal principle of the type of the *Gibbs principle* that postulates the statistics of Boltzmann–Gibbs as the correct statistics for the computation of the equilibrium properties of systems in statistical mechanics.

We shall describe some applications in the Ch. VII below. At the moment we limit ourselves to discussing a fundamental theorem on the structure of the SRB distributions.

(F) Representation of the SRB distribution via markovian pavements.

The proof of the theorem I of the point (E) is based on the esistence of markovian pavements (see remarks (1)), *i.e.* on symbolic dynamics; and on symbolic dynamics is also based the *expansion in periodic orbits* (5.5.8) of the SRB distribution.

If $\mathcal{E}$ is a generating Markov pavement for (M,S) one can define the finer pavement $\mathcal{E}_T = \cap_{k=-T/2}^{T/2} S^{-k}\mathcal{E}$. Its elements can be denoted E_j if $j = (\sigma_{-T/2},\ldots,\sigma_{T/2})$ and

$$E_j \stackrel{def}{=} E_{\sigma_{-T/2},\ldots,\sigma_{T/2}} = \cap_{k=-T/2}^{T/2} S^{-k} E_{\sigma_k} \tag{5.7.5}$$

and $(\sigma_{-T/2},\ldots,\sigma_{T/2})$ is a string of digits compatible by nearest neighbors, see definition 2, (*i.e.* if M is the compatibility matrix of the pavement $\mathcal{E}$ one has $M_{\sigma_i,\sigma_{i+1}} = 1$ for $i = -T/2,\ldots,T/2-1$).

In each of these sets E_j, $j = (\sigma_{-T/2},\ldots,\sigma_{T/2})$, one can select a point $x_j \in E_j$ by continuing the string j to an infinite bilateral sequence which is a possible history of a point x_j. We assign to every digit σ an infinite sequence $\sigma_1^+,\sigma_2^+,\ldots$ such that $\sigma,\sigma_1^+,\sigma_2^+,\ldots$ is compatible by nearest neighbors and a second sequence $\ldots,\sigma_{-2}^-,\sigma_{-1}^-$ such that $\ldots,\sigma_{-2}^-,\sigma_{-1}^-,\sigma$ is also compatible by nearest neighbors. We shall say that $\ldots,\sigma_{-2}^-,\sigma_{-1}^-$ is an "extension to left" of σ and that $\sigma_1^+,\sigma_2^+,\ldots$ is an "extension to the right".

A *standard extension* is such a pair of functions $\sigma \to (\sigma_1^+, \sigma_2^+, \ldots)$ and $\sigma \to (\ldots, \sigma_{-2}^-, \sigma_{-1}^-)$.

Given a standard extension consider, as $j = (\sigma_{-T/2}, \ldots, \sigma_{T/2})$ varies, the bilateral sequences

$$\underline{\sigma}_j \equiv \ldots, \sigma_{-2}^-, \sigma_{-1}^-, \sigma_{-T/2}, \ldots, \sigma_{T/2}, \sigma_1^+, \sigma_2^+, \ldots \tag{5.7.6}$$

where $\sigma_1^+, \sigma_2^+, \ldots$ is the right extension of $\sigma_{T/2}$ and $\ldots, \sigma_{-2}^-, \sigma_{-1}^-$ the left extension of $\sigma_{-T/2}$. The points x_j that have $\underline{\sigma}_j$ as possible history will be called the *centers* of E_j (with respect to the given right and left extensions).

There are many centers of E_j: one for every standard extension in the just defined sense. The set of points that can be centers of E_j is even dense in E_j (note that there exist infinitely many possible standard extensions and hence, for every j, infinitely many centers).

Extensions of $\sigma_{-T/2}, \ldots, \sigma_{T/2}$ which are *not standard* are, for instance, sequences compatible by nearest neighbors whose symbols with labels external to $[-T/2, T/2]$ *not only depend* on $\sigma_{-T/2}$ and $\sigma_{T/2}$ (as in a standard extension) but *also depend* on the values σ_j with $|j| < T/2$.

Let $\Lambda_{e,T}(x), \Lambda_{s,T}(x)$ be the determinants of the Jacobian matrix os S^T considered as a map of the unstable or stable (respectively) manifold at x to the corresponding manifold at $S^T x$. Let $\lambda_e(\underline{\sigma}) = \log|\Lambda_{e,1}(x(\underline{\sigma}))|$, $\lambda_s(\underline{\sigma}) = \log|\Lambda_{s,1}(x(\underline{\sigma}))|$ if $\underline{\sigma}$ is the history of x.

Fixed T and a standard extension, and hence the family of the centers of the sets E_j for every j, consider the following probability distributions $\mu_T^+, \mu_T^-, \mu_T^0$ that are defined on M by assigning to each set E_j a weight given by

$$\begin{cases} Z_T^+(x_j) \equiv \Lambda_{e,T}^{-1}(x_j) = \exp - \sum_{k=-T/2}^{T/2-1} \lambda_e(S^k \underline{\sigma}_j) & \text{for } \mu_T^+ \\ Z_T^-(x_j) \equiv \Lambda_{s,T}(x_j) = \exp \sum_{k=-T/2}^{T/2-1} \lambda_s(S^k \underline{\sigma}_j) & \text{for } \mu_T^- \\ Z_T^0(x_j) \equiv \exp(\sum_{k=-T/2}^{-1} \lambda_s(S^k \underline{\sigma}_j) - \sum_{k=1}^{T/2-1} \lambda_e(S^k \underline{\sigma}_j)) & \text{for } \mu_T^0 \end{cases} \tag{5.7.7}$$

So that, for $\alpha = \pm, 0$, we set

$$\mu_T^\alpha(F) \stackrel{def}{=} \frac{\sum_{\sigma_{-T/2}, \ldots, \sigma_{T/2}} F(x_j) Z_T^\alpha(x_j)}{\sum_{\sigma_{-T/2}, \ldots, \sigma_{T/2}} Z_T^\alpha(x_j)} \tag{5.7.8}$$

by definition.

The fundamental theorem, of which theorem I of point (E) and the (5.5.8) are corollaries, is (Sinai):

II Theorem *(SRB and volume measures): The limits as $T \to \infty$ of $\mu_T^\pm(F)$ and $\mu_T^0(F)$ exist for every regular function F and are independent on the choice of the centers x_j. If we denote $\mu^\pm, \mu^0$, respectively, the limit distributions as $T \to \infty$ one finds that $\mu^\pm$ are the SRB distributions for S and*

S^{-1} while μ^0 is proportional to the (non invariant) volume measure on M, i.e. it is absolutely continuous.

The proof of this theorem, and hence of theorem I and of (5.5.8) ([Si94]), will not be discused here, see also [Ga81],[Ga95].

Theorem II not only implies existence of the SRB distribution, *i.e.* the normality of the attracting mixing sets that verify axiom A, but *it provides a useful expression* for the SRB distributions μ^-, μ^+ ("for the past" and "for the future", *i.e.* for S and S^{-1}) and for a distribution μ_0 absolutely continuous with respect to the volume (*c.f.r.* footnote [13]).

This is our endpoint of the discussion on the *kinematics* (or "structure") of chaotic motions. In the successive Chap. VI and VII we shall illustrate some general, but concrete, applications of this qualitative conception of motions.

A detailed analysis of the proof of the above theorems and of the construction of the Markov pavements for two dimensional Anosov maps can be found in [GBG04]

Problems.

Problems [5.7.1] through [5.7.14] provide a rather general theory of the density function $\rho(x)$ for the invariant distribution $\mu(dx) = \rho(x)dx$ of an expansive map S of the interval, *c.f.r.* definition 1 above. The theory is taken from [Ru68] and is an extension of the Perron–Frobenius theorem on matrices with positive elements.

[5.7.1]: Let S be a continuous expansive map of $[0, 1]$, defined by s maps f_i : $[a_i, a_{i+1}] \longleftrightarrow [0, 1]$ of class C^∞ and derivatives $|f_i'| \ge \lambda > 1$, *c.f.r.* (A), as $S(x) = f_i(x)$ for $x \in [a_{\sigma_i}, a_{\sigma_i+1}]$. If $\underline{\sigma} = (\sigma_0, \sigma_1, \sigma_2, \ldots)$ is an arbitrary sequence of digits with s values there exists a point $x = x(\underline{\sigma}) \in [0, 1]$ such that $S^j x \in [a_{\sigma_j}, a_{\sigma_j+1}]$. The correspondence $x \longleftrightarrow \underline{\sigma}$ between the space $[0, 1]$ and the space $\mathcal{S}$ of all sequences of digits with s symbols is one to one except for a denumerable infinity of points x to which correspond 2 sequences. The same happens if S is discontinuous at some a_i, $0 < i < s$ and its value at a_i is fixed arbitrarily to be the left limit of f_i or the right limit of f_{i-1}. (*Idea:* There can be only one time k such that $S^k x = a_i$ for some $0 < i < s$; and this can only happen for countably many distinct points x.)

[5.7.2]: In the context pf problem [5.7.1], let $\lambda(\underline{\sigma}) = \log |S'(x(\underline{\sigma})|$ then $\lambda(\underline{\sigma})$ is such that $|\lambda(\underline{\sigma}) - \lambda(\underline{\sigma}')| < Ce^{-\lambda n}$ if $\sigma_j = \sigma_j'$ for $j = 0, \ldots, n-1$ and $\lambda = \min |\lambda(\underline{\sigma})|$, $C = \max_x |S"(x)|/|S'(x)|$. Set $d(\underline{\sigma}, \underline{\sigma}') = e^{-n}$ if $\sigma_j = \sigma_j'$, $j = 0, \ldots, n-1$ but $\sigma_n \ne \sigma_n'$ then: $|\lambda(\underline{\sigma}) - \lambda(\underline{\sigma}')| < Cd(\underline{\sigma}, \underline{\sigma}')^\lambda$ (hence we say that $\lambda(\underline{\sigma})$ is *Hölder continuous*).

[5.7.3] *(expansive interval maps: equation for the invariant density):* In the context pf problem [5.7.1], suppose that $n(dx) = \rho(x)dx$ is a S–invariant probability distribution then:

$$\rho(x) = \sum_{Sy=x} |S'(y)|^{-1} \rho(y) \stackrel{def}{=} L\rho(x)$$

where the operator L is defined by this relation. Show also that L has the property $\int_0^1 (L^k\, 1)(x)\, dx \equiv 1$, for $k \ge 1$. (*Idea:* look for the geometical meaning of $|S'(y)|^{-1}dy$.)

[5.7.4] *(expansive interval maps: symbolic equation for the invariant density and transfer matrix operator):* Let $\alpha\underline{\sigma} \stackrel{def}{=} (\alpha, \sigma_0, \sigma_1, \ldots)$; write the equation in [5.7.3] for the density $h(\underline{\sigma})=\rho(x(\underline{\sigma}))$ and check that it becomes:

$$h(\underline{\sigma})=\sum_\alpha e^{-\lambda(\alpha\underline{\sigma})}h(\alpha\underline{\sigma})\stackrel{def}{=}\mathcal{L}h(\underline{\sigma})$$

where the operator $\mathcal{L}$ is defined by this relation.

[5.7.5] *(expansive interval maps: transfer matrix operator)*: Show that if $1(\underline{\sigma})\stackrel{def}{=}1$ is the function which is identically 1 on the sequences $\underline{\sigma}$ then the operator $\mathcal{L}$ defined in [5.7.4] verifies

$$B^{-1}<\frac{(\mathcal{L}^n1)(\underline{\sigma})}{(\mathcal{L}^n1)(\underline{\sigma}')}<B$$

for $B=\exp(Ce^{-\lambda}/(1-e^{-\lambda}))$ and C,λ as in [5.7.2]. Show that there exist $\underline{\sigma}_0,\underline{\sigma}_1$ such that $\mathcal{L}^n\,1(\underline{\sigma}_0)\le 1$ and $\mathcal{L}^n\,1(\underline{\sigma}_1)\ge 1$. (*Idea:* Note that:

$$\frac{\sum_{\alpha_1,\ldots,\alpha_n}e^{-\lambda(\alpha_1,\ldots,\alpha_n,\sigma_0,\ldots)-\lambda(\alpha_2,\ldots,\alpha_n,\sigma_0,\ldots)-\ldots-\lambda(\alpha_n,\sigma_0,\ldots)}}{\sum_{\alpha_1,\ldots,\alpha_n}e^{-\lambda(\alpha_1,\ldots,\alpha_n,\sigma_0',\ldots)-\lambda(\alpha_2,\ldots,\alpha_n,\sigma_0',\ldots)-\ldots-\lambda(\alpha_n,\sigma_0',\ldots)}}$$

and bound above this ratio with the max of the ratios of the corresponding terms, using [5.7.2]. Note that if $n_0(dx)=dx$ and $\nu_0(d\underline{\sigma})$ is the correspondent probability distribution on sequences $\underline{\sigma}\in\mathcal{S}$ then by [5.7.13] $\nu_0(\mathcal{L}^k1)=1$ for all $k\ge 1$.)

[5.7.6]: If n is an invariant distribution $n(dx)=\rho(x)\,dx$, *c.f.r.* [5.7.3], then for every $f\in L_1(n)$ it is $n(Lf)\equiv n(f)$. If ν is the correspondent probability distribution on the sequences $\underline{\sigma}\in\mathcal{S}$ induced by the code $\underline{\sigma}\longleftrightarrow x(\underline{\sigma})$ as image of n then: $\nu(\mathcal{L}F)\equiv\nu(F)$ for every continuous function F (with respect to the distance $d(\underline{\sigma},\underline{\sigma}')$ in [5.7.2]) on $\mathcal{S}$.

[5.7.7] *(uniform continuity of the transfer operator for interval maps):* Show that if $d(\underline{\sigma},\underline{\sigma}')$ is defined as in [5.7.2] and if $\mathcal{L}$ is defined as in [5.7.4] it is

$$|(\mathcal{L}^n1)(\underline{\sigma})-(\mathcal{L}^n1)(\underline{\sigma}')|<Dd(\underline{\sigma},\underline{\sigma}')^\lambda$$

(*Idea:* Develop the idea of [5.7.5]).

[5.7.8] *(interval maps, existence of an invariant density)*: Note that [5.7.5] and [5.7.7] imply that the sequence $n\to\mathcal{L}^n1$ is an equicontinuous and equibounded sequence on $\mathcal{S}$ and hence such is also the sequence of the "*Cesaro averages*" $n\to n^{-1}(1+\mathcal{L}1+\ldots+\mathcal{L}^{n-1}1)$. Show that every accumulation point h of this last sequence is a fixed point for $\mathcal{L}$. Show that this means that $\nu\stackrel{def}{=}h\,\nu_0$ (ν_0 is defined in the hint to [5.7.5]) is a distribution invariant under translations $\tau:(\sigma_0,\sigma_1,\ldots)\to(\sigma_1,\ldots)$ and deduce that its image measure on $[0,1]$ via the code $\underline{\sigma}\longleftrightarrow x(\underline{\sigma})$ is a distribution ρdx invariant under S. The proof just given of the existence of the invariant distribution *is not* constructive: can it be improved by making it constructive? (*c.f.r.* following problems). (*Idea:* Indeed the proof makes use of the theorem of Ascoli–Arzelá on equicontinuous equibounded sequences of functions, which is not constructive.)

[5.7.9]: Let f be a continuous function on $\mathcal{S}$ and $f\in\Gamma^k=$ (*space of the functions $f(\underline{\sigma})$ that depend only on $\sigma_0,\ldots,\sigma_{k-1}$*), also called the space of the *cylindrical functions* or *local functions* on $\mathcal{S}$. Show that $f\ge 0$ implies that $\mathcal{L}^kf>0$ and, if $||f||$ denotes the maximum of f and $\nu=h\nu_0$ is the measure defined in problem [5.7.8], $||\mathcal{L}^kf||\ge B^{-1}\nu(\mathcal{L}^kf)=B^{-1}\nu(f)$. (*Idea:* Proceed as in [5.7.5] and show that $\mathcal{L}^pf(\underline{\sigma})\ge B^{-1}\mathcal{L}^p(\underline{\sigma}')$ for $p\ge k$ and hence integrate both sides with respect to $\nu(d\underline{\sigma})$ using that $\nu(\mathcal{L}^k1)=1$, for all $k\ge 1$.)

[5.7.10]: Consider the function $g=1-h$, *c.f.r.* [5.7.8]: then $\nu(g)=0$. Let $g_k=1-h_k$ with $h_k=h_k^0-\nu(h_k^0)+1$ and $h_k^0(\underline{\sigma})=h(\sigma_0,\ldots,\sigma_{k-1},0,0,\ldots)$. Show that $||h-h_k||<$

$Ee^{-\lambda k}$ and $||g_k|| < E$, for a suitable $E > 0$ and for $k, n \geq 0$. (*Idea:* Use the Hölder continuity of h implied by [5.7.7]).

[5.7.11]: If $f \in C(\mathcal{S})$ and $f_\pm = (|f| \pm f)/2$ realize, in the context of [5.7.10], the validity of the following relations for $n \geq k$, as consequences of the inequality in [5.7.9]:

$$\begin{aligned}
&\nu(|\mathcal{L}^n g_k|) = \nu(|\mathcal{L}^k \mathcal{L}^{n-k} g_k|) = \nu(|\mathcal{L}^k \Big((\mathcal{L}^{n-k} g_k)_+ - (\mathcal{L}^{n-k} g_k)_-\Big)|) = \\
&\nu\Big(\Big|\mathcal{L}^k \Big((\mathcal{L}^{n-k} g_k)_+ - \frac{1}{B}\nu((\mathcal{L}^{n-k} g_k)_+) - (\mathcal{L}^{n-k} g_k)_- \frac{1}{B}\nu((\mathcal{L}^{n-k} g_k)_-)\Big)\Big|\Big) \leq \\
&\nu(\mathcal{L}^k \Big((\mathcal{L}^{n-k} g_k)_+ - \frac{1}{B}\nu((\mathcal{L}^{n-k} g_k)_+) + (\mathcal{L}^{n-k} g_k)_- - \frac{1}{B}\nu((\mathcal{L}^{n-k} g_k)_-)\Big)) = \\
&= (1 - B^{-1})\nu(|\mathcal{L}^{n-k} g_k|)
\end{aligned}$$

(*Idea:* $\nu((\mathcal{L}^{n-k} g_k)_+) \equiv \nu((\mathcal{L}^{n-k} g_k)_-)$ because $\nu(\mathcal{L}^{n-k} g_k) = 0$; furthermore it is $(\mathcal{L}^{n-k} g_k)_\pm - \frac{1}{B}\nu((\mathcal{L}^{n-k} g_k)_\pm) \geq 0$.)

[5.7.12]: Evince from [5.7.11] that $\nu(|\mathcal{L}^n g_k|) \leq (1 - B^{-1})^{n/k} \nu(|g_k|)$ if n multiple of k.

[5.7.13]: Show that [5.7.12] and [5.7.10] imply that $\nu(|\mathcal{L}^n(1-h)|) \xrightarrow[n\to\infty]{} 0$ bounded by $ae^{-b\sqrt{n}}$. (*Idea:* Note that $g - g_k$ is bounded proportionally to $e^{-\lambda k}$ and choose $k = \sqrt{n}$ in [5.7.12].)

[5.7.14] *(interval maps, fast convergence of the approximants of the invariant density):* Check that [5.7.13] implies that $\lim_{n\to\infty} \mathcal{L}^n 1 = h$ with an error that can be bounded by $ae^{-b\sqrt{n}}$ and with a, b computable constants. Hence the proof of the existence of an S–invariant absolutely continuous probability distribution has been made constructively.

[5.7.15]: Consider a transitive bidimensional Anosov system (M, S) with a fixed point. Construct a Markov pavement $\mathcal{E}$ with the method of the point (C) starting from the stable and unstable manifolds of the fixed point. (*Idea:* The construction in (C) does not really ever use the special form of the stable and unstable manifolds of the origin.)

[5.7.16] *(Markov partitions of a perturbation of a 2–dimensional Anosov map)*: Perturb in class C^∞ the system in [5.7.15] obtaining the dynamical system (M, S_ε). Show that if the perturbation is small enough, with ε measuring of its size, the new system admits a markovian pavement $\mathcal{E}_\varepsilon$ with the same number of elements of $\mathcal{E}$. The elements of $\mathcal{E}_\varepsilon$ are obtained by small deformations of those of $\mathcal{E}$ and the compatibility matrices of $\mathcal{E}$ in (M, S) and of $\mathcal{E}_\varepsilon$ in (M, S_ε) *are identical*. (*Idea:* Note that for the construction of the pavement $\mathcal{E}$ we only use *finite portions* of the stable and unstable manifolds of the fixed point. Furthermore the fixed point "survives to the perturbation" if ε is small enough, by the implicit functions theorem (being hyperbolic its Jacobian is not zero)).

[5.7.17]: Establish a one to one correspondence between the points of M by defining $C_\varepsilon x$ as the point that, on the new pavement $\mathcal{E}_\varepsilon$ has with respect to the new dynamics S_ε the same history of x with respect to the unperturbed dynamics S. Show that this correspondence is continuous, and in fact Hölder continuous. (*Idea:* The points are determined with exponential precision in terms of the specified number of digits of their history.)

[5.7.18] *(theorem of structural stability)*: Check that the correspondence C_ε verifies $S_\varepsilon C_\varepsilon \equiv C_\varepsilon S$. Furthermore if x and x' have S_ε–histories eventually equal in the future it is $d(S_\varepsilon^n x, S_\varepsilon^n x') \leq Ce^{-\lambda' n}$ with $0 < \lambda' < \lambda$ and C suitable. Likewise if x and x' have histories eventually equal in the past it is $d(S_\varepsilon^n x, S_\varepsilon^n x') \leq Ce^{-\lambda' n}$. Deduce that through every point x pass two surfaces $W_x^{\varepsilon,s}, W_x^{\varepsilon,u}$ that vary with Hölderian continuity as x varies. (*Idea:* One has $W_x^{\varepsilon,\alpha} = C_\varepsilon W_x^\alpha$ for $\alpha = u, s$.)

[5.7.19]: Consider a dynamical system (M, S) of class C^∞ with S a mixing Anosov diffeomorphism and M a 2–dimensional surface. If $\mathcal{P}$ is a generating Markov pavement consider the pavement $\cap_{-n_-}^{n_+} S^{-j}\mathcal{P}$ with (within one unit) $n_+\lambda_+ = n_-|\lambda_-|$ and $\lambda_\pm$ the Lyapunov exponents of an ergodic invariant distribution μ (not necessarily the SRB distribution of (M,S)). A support A of the distribution μ (*i.e.* any invariant set with μ–probability 1) is covered by the elements of $\cap_{-n_-}^{n_+} S^{-j}\mathcal{P}$: give a heuristic argument to infer that in reality "only" $\sim e^{(n_++n_-)s}$, if s is the entropy of μ with respect to S, such elements suffice to cover A, for n large. (*Idea:* Make use of the theorem of Shannon–McMillan of the §5.6).

[5.7.20] *(Kaplan–Yorke formula for 2–dimensional Anosov maps)*: On the basis of [5.7.19] and observing that the elements of $\cap_{-n_-}^{n_+} S^{-j}\mathcal{P}$ are small rectangles of almost equal size in the stable and unstable directions (by the choice of $n_\pm$) infer, (always heuristically), that the minimal Hausdorff dimension of A is $d_A(\mu) = (1/\lambda_+ + 1/|\lambda_-|)\, s$. (*Idea:* It is $n_+ + n_- = n_+(1+\lambda_+/|\lambda_-|)$ while the linear dimension of the "important" elements (*i.e.* those $P_{\sigma_{-n_-},\ldots,n_+} = \cap_{-n_-}^{n_+} S^{-j}P_{\sigma_j}$ defined by strings $\sigma_{-n_-,\ldots,n_+}$ of digits "frequent" in the sense of the theorem of Shannon–McMillan) is $e^{-n_+\lambda_+}$. Hence for (5.6.8) and (5.6.9) we should have:

$$e^{n_+(1+\lambda_+/|\lambda_-|)s} = e^{-n_+\lambda_+\alpha}$$

if α is the Hausdorff dimension of A.)

[5.7.21] If μ is the SRB distribution for the system considered in [5.7.19], [5.7.20] then $s = \log\lambda_+$: give a heuristic argument. Deduce, from [5.7.20], that in the system in question the dimensions of information and of Lyapunov relative to the invariant distribution μ are equal. The discussion is in reality more general and the conclusions of [5.7.20] hold under the only hypothesis that (M,S) are of class C^2 and that μ is an ergodic distribution. And in 2 dimensional systems the identity between $d_L(\mu)$ and $d_I(\mu)$ holds under the only hypothesis that μ is a SRB distribution in the sense (weaker than that we use in this volume) of the observation (3) to the conjecture in (B), §5.5: *theorem of Young*, *c.f.r.* [Yo82].

[5.7.22] *(Pesin formula)***:** Show, at least heuristically, that if (M,S) is a mixing Anosov system and if μ is its SRB distribution then its entropy s is given by the sum of the logarithms of the positive Lyapunov exponents of (M,S,μ) (*Pesin's formula*). It is this a particular case of a theorem of Ledrappier–Young, *c.f.r.* [ER81], p. 639, that requires only that (M,S) is a dynamical system with M,S of class C^2 and that ρ is an ergodic distribution and SRB in the weak sense of the observation (3) to the conjecture of the §5.5. (*Idea:* Find the geometric interpretation in terms of Markov partitions of the relation between of the third expression in (5.7.7) and the area of the sets E_j in (5.7.5): it will appear that the quantities $Z_p^0(\sigma_{-p},\ldots,\sigma_p)$ differ by a factor the can be bounded above and below by a p-independent constant from the area of $\cap_{i=-p}^p S^{-j}P_{\sigma_j}$. Since the analysis in problems [5.7.4] through [5.7.14] depends only on the Hölder continuity of $\lambda(\underline{\sigma})$ it can be applied to the two factors defining Z^0 to infer that $\cap_0^p S^{-j}P_{\sigma_j}$ have area bounded above and below by a p-independent constant by $e^{-\sum_{j=0}^p \lambda_e(S^j\underline{\sigma})}$ hence the sum $C = \sum_{\sigma_0,\ldots,\sigma_p} e^{-\sum_{j=0}^p \lambda_e(S^j\underline{\sigma})}$ is uniformly bounded as $p\to\infty$, because $\sum_{\sigma_0,\ldots,\sigma_p} \mu_0(\cap_{i=-p}^p S^{-j}P_{\sigma_j}) = 1$, so that by the first of (5.7.7):

$$\lim_{p\to\infty} -p^{-1} \sum_{\hat\sigma_0,\ldots,\hat\sigma_p} \mu(\cap_{i=0}^\infty S^{-j}P_{\hat\sigma_j}) \log\mu(\cap_{i=0}^\infty S^{-j}P_{\hat\sigma_j}) =$$

$$= \lim_{p\to\infty} \Big[+p^{-1}\log C - p^{-1} \sum_{\sigma_0,\ldots,\sigma_p} C^{-1} e^{-\sum_{k=0}^p \lambda_e(S^k\underline{\sigma})} \big(-\sum_{k=0}^p \lambda_e(S^k\underline{\sigma})\big)\Big] =$$

$$= \int \lambda_e(x)\mu(dx)$$

that just shows what desired. For a rigorous control of the errors one should make use of the theory in the problems [5.7.1]÷[5.7.14].)

Bibliography: [AA68], [Ga81], [Ga95], [ER81] . The constructive proof of the existence of the invariant distribution is taken from Ruelle, *c.f.r.* [Ga81], and it can be adapted to provide constructive proofs of existence of various equations (*c.f.r.* for instance: [Ga82]): several problems on the elliptic equations studied with methods variational (not constructive), *c.f.r.* problems of §2.2, can be studied with the methods of this section). As noted by Ruelle this is of a remarkable generalization of a well known theorem on matrices with all matrix elements positive (it is the *theorem of Perron–Frobenius* according to which a matrix M with positive elements has the eigenvalue of largest absolute value that is positive and simple and with the corresponding eigenvector with positive components): here the role of the matrix is plaid by the operator $\mathcal{L}$.
A more elementary proof, *remarkably* more general, of the existence of a probability distribution invariant for an expansive map can be found in [LY73] (*theorem of Lasota–Yorke*).
The construction in (C) and its extension in [5.7.16] were suggested to me by M. Campanino. The proof of the structural stability (partial because we have not shown the regularity of the manifolds, *i.e.* the existence of the tangent plane at each of their points) is more difficult in dimension > 2, *c.f.r.* [Ga95] for an attempt of extension along the lines of [5.7.15]÷[5.7.18]; see the appendix of [Sm67] for a general proof due to Mather.
Symbolic dynamics for a geodesic flow on surfaces of constant negative curvature, like the one relative to the octagon group, *c.f.r.* problems of §5.1, is also possible although much more involved if compared to that relative to the arnoldian cat: it will not fit into this volume, *c.f.r.* [AF91].
See, for instance, [GBG04] for detailed analysis of all the problems in Chapter VI and for proofs of most statements left unproved here.

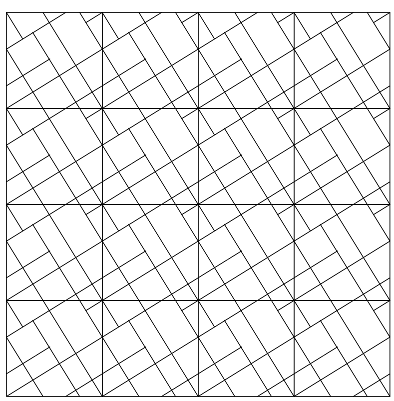


Fig. (5.7.6) *Markovian pavement of §5.7*

CHAPTER VI

Developed turbulence

§6.1 Functional integral representation of stationary distributions.

From a qualitative viewpoint the onset of turbulence, *i.e.* the birth of chaos, is rather well understood, as analyzed in Chap. IV.

Here we shall try to begin an analysis devoted to understanding and organizing the properties of developed turbulence, "at large Reynolds number".

One could think that the only difference is the large dimension of the attracting set A (or of the attracting sets, in the hysteresis cases *c.f.r.* §4.3), and that the statistics that describes motions following random choices of initial data is determined by Ruelle's principle, *c.f.r.* §5.7, at least if initial data are chosen within the basin of the attracting set and with a distribution absolutely continuous with respect to volume.

The main difficulty in considering such viewpoint as satisfactory, even only approximately, is that usually we do not have any idea whatsoever of the nature and location in phase space of the attracting set. Hence it is difficult, if not impossible, to apply the principle as we lack, in a certain sense, the "raw material".

One can begin by trying to study probability distributions, defined on the space of the of velocity fields, and invariant with respect to the evolution defined by NS equations, incompressible to fix the ideas. Such distributions can be candidates to describe the statistics of motions developing on a some attracting set, *i.e.* stationary states of the fluid.

The approach is suggested and justified by the success of equilibrium statistical mechanics where there exists a natural invariant distribution, proportional to the Liouville measure, which in fact describes the statistical and thermodynamic properties of most systems.

For the NS equation the problem is, however, difficult because there are no obvious invariant distributions when the external force that keeps the fluid in motion is constant.

A better situation is met, instead, in cases in which the external force is random; and it is useful to quote briefly the basic formal results, warning immediately that one is (still) unable to go much further than formal discussions.

Consider the equation on $\Omega=[0,L]^3$ with periodic conditions

$$\begin{aligned}&\dot{\underline{u}}+\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{u}=\nu\Delta\underline{u}-\underline{\partial}p+\underline{f}(t)\\&\underline{\partial}\cdot\underline{u}=0,\qquad\int_\Omega\underline{u}\,d\underline{x}=\underline{0}\end{aligned}\tag{6.1.1}$$

where $t\to\underline{f}(t)$ is a randomly chosen volume force.

We shall suppose that the force $\underline{f}$ has a *Gaussian distribution*, determined, therefore, by its *covariance*.[1] For example if we set

$$\underline{f}(\underline{x},t)=\sum_{\underline{k}\neq\underline{0}}e^{i\underline{k}\cdot\underline{x}}\,\underline{f}_{\underline{k}}(t)=\sum_{\underline{k}\neq\underline{0}}\int dk_0\,e^{i\,(\underline{k}\cdot\underline{x}+k_0t)}\,\underline{f}_{\underline{k}}\tag{6.1.4}$$

an interesting covariance for the variable $f_{\underline{k},\alpha}$, *i.e.* for the component of mode $\underline{k}$ and label $\alpha=1,2,3$, of the forcing (interesting because it is an hypothesis "maximum simplicity") can be

$$\langle\,\overline{f_{\underline{k},\alpha}(t)}f_{\underline{h},\beta}(t')\rangle=\delta_{\underline{h}\,\underline{k}}\,g_{\underline{k},\alpha,\beta}(t-t')\tag{6.1.5}$$

with $g_{\underline{k}}$ proportional to $\gamma\delta_{\alpha\beta}\delta(t-t')$ and given by:

$$g_{\underline{k},\alpha,\beta}(t-t')=\frac{\gamma\delta_{\alpha\beta}}{|\underline{k}|^a}\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{ik_0(t-t')}\,dk_0\tag{6.1.6}$$

---

[1] A *Gaussian process* is a probability distribution $\overline{P}$ for random choices of functions $\xi\to f_\xi$, defined on a *finite* space of N labels ξ such that every linear functional having the form $\sum_\xi c_\xi\,f_\xi$ with c_ξ arbitrary constants is a random variable with Gaussian distribution. This simply means that one can find a positive definite symmetric matrix A such that the probability of the infinitesimal volume element $df=\prod_\xi df_\xi$ is

$$\overline{P}(df)\equiv e^{-\frac12(Af,f)}\frac{\prod_\xi df_\xi}{(\pi^N\det A^{-1})^{1/2}}\tag{6.1.2}$$

The distribution $\overline{P}$ is therefore uniquely defined by the matrix A or, what is the same, by A^{-1}. Since one can check that $\langle f_\xi f_\eta\rangle\stackrel{def}{=}\int\overline{P}(df)f_\xi f_\eta\equiv(A^{-1})_{\xi\eta}$ it is convenient to say that the stochastic process is defined by the *covariance matrix* A^{-1}, simply called *covariance*. Note that the average values of the products $\prod_{i=1}^{2n}f_{\xi_i}$ are easily expressible in terms of the covariance ("*Wick theorem*"). This follows by differentiating both sides of the identity

$$\int\overline{P}(df)e^{\sum c_\xi f_\xi}=e^{\frac12(A^{-1}c,c)}\tag{6.1.3}$$

with respect to the parameters c_ξ and then setting them equal to zero.

If the space of the labels not is finite, then a gaussian process can be defined in the same way in terms of the covariance. This means that one assigns a positive definite quadratic form denoted $(A^{-1}f,f)$ and one defines the integrals of the products $\prod_{i=1}^{2n}f_{\xi_i}$ by suitable differentiations via (6.1.3).

Note that if $\overline{P}$ is a gaussian process then every linear operator $f\to O(f)$ defines a gaussian random variable.

(having represented $\delta(t-t')$ as an ntegral over an auxiliary variable k_0) for some constants $\gamma>0, a$ that are called the *intensity* and the *color* of the random force $\underline{f}$. The constant a will be taken ≥ 0 but it could be taken <0 as well. Indeed the case $a=-2$ is of particular interest as, for instance, argued in [YO86]: it should give the statistical properties of a fluid in thermal equilibrium at temperature T if $\gamma=2\nu k_B T/\rho$ with k_B being Boltzmann's constant and ρ the fluid density. It is believed that the most relevant case for our purposes is $a=3$, see [YO86].

Imagining, for simplicity, that the fluid is initially motionless at time $-\Theta$ (and therefore $\underline{u}(-\Theta)=\underline{0}$), we shall have

$$\begin{aligned}&\underline{u}_{\underline{k}}(t)=\\&=\int_{-\Theta}^{t} e^{-\nu\underline{k}^2(t-\tau)}\Big(\Pi_{\underline{k}}\underline{f}_{\underline{k}}(\tau)-i\sum_{\underline{k}_1+\underline{k}_2=\underline{k}}\underset{\sim}{u}_{\underline{k}_1}(\tau)\cdot\underset{\sim}{k}_2\,\Pi_{\underline{k}}\underline{u}_{\underline{k}_2}(\tau)\Big)\,d\tau\end{aligned}\tag{6.1.7}$$

where (*c.f.r.* (2.2.7)) $\Pi_{\underline{k}}\equiv\Pi_{\rm rot}$ denotes the projection on the plane orthogonal to $\underline{k}$ and we have *ignored* the uniqueness problems of the solutions of (6.1.7) (which we do not know how to solve, something that should not be forgotten).

Let us proceed heuristically by asking ourselves which could be the probability distribution of the field $\underline{u}(t)$ for t large and write symbolically the equation (6.1.7) as:

$$u=v+Tu\tag{6.1.8}$$

where v depends *only on the external force* $\underline{f}$ and $Tu\equiv Q(u,u)$ is the nonlinear term in (6.1.7) thought of as a function defined on the fields $u=\underline{u}(\underline{x},t)$ with zero divergence. Hence, always formally

$$v=(1-T)u,\qquad u=(1-T)^{-1}v\tag{6.1.9}$$

The probability distribution of u is then a distribution μ such that the μ–probability that u is in "an infinitesimal volume element du" is

$$\mu(du)=P_0(dv)\tag{6.1.10}$$

where P_0 is the probability distribution of v (inherited directly from that of $\underline{f}$, assumed known *c.f.r.* (6.1.6)) and dv is the image of the volume element du under the map $(1-T)$: symbolically $dv=(1-T)du$.

Note that in the example (6.1.5) the distribution $P_0(dv)$ is the distribution of $v=\int_{-\Theta}^{t}\Gamma_{t-\tau}*\Pi_{\rm rot}\,\underline{f}(\tau)d\tau$ (a relation following from first term in the r.h.s. of (6.1.7), where Γ is the periodic heat equation Green's function, cf. (3.3.17)) and it is a Gaussian distribution. Hence if we write, for the purpose of a formal analysis, $P_0(dv)=G(v)dv$ we find

$$\mu(du)=G(v)\frac{dv}{du}du=G((1-T)u)\det\frac{\partial(1-T)u}{\partial u}du\tag{6.1.11}$$

Obviously this formal computation has to be interpreted. A possible way is to discretize "everything": *i.e.* to fix an interval of time t_0, to set $\tau = jt_0$ and to write the integral in (6.1.7) as a finite sum $t_0\sum_{j=0}^{t/t_0-1}$, and furthermore to truncate also the sum over $\underline{k}_1, \underline{k}_2$ at a cut–off value N: $|\underline{k}| < N$. Here t_0, N are regularization parameters introduced for the purposes of our calculations and that should be let eventually to 0 and ∞, respectively.

If this viewpoint is adopted we see that the jacobian matrix $\frac{\partial(1-T)u}{\partial u}$ has a triangular form with 1 on the diagonal, because the times in the left hand side of the discretized (6.1.7) are all $< t$ (because jt_0 goes up to $t - t_0$ *excluded*).[2]

Hence we obtain from the (6.1.11), *c.f.r.* [Gh60], the *Ghirsanov formula*

$$\mu(du) = G((1-T)u)du \tag{6.1.12}$$

where $P_0(dv) \stackrel{def}{=} G(v)dv$ is the distribution of $v = \int_{-\Theta}^{t} \Gamma_{t-\tau} * \Pi_{\rm rot}\, \underline{f}(\tau)d\tau$.

With the notation $(\underline{u}, \underline{v}) = \int \underline{u}(\underline{x}, t)\cdot\underline{v}(\underline{x}, t)\, d\underline{x}\, dt/2\pi L^3$ that defines a scalar product between *zero divergence* fields $\underline{u}, \underline{v}$ we want to suppose that the external force field $\underline{f}$ is a random force with a gaussian distribution.

Therefore v, being a linear function of $\underline{f}$ *i.e.* $v = \int_{-\Theta}^{t} \Gamma_{t-\tau} * \Pi_{\rm rot}\, \underline{f}(\tau)\, d\tau$, will also have a gaussian distribution and it will be possible to write (formally) its distribution as $G(v)\, dv = \mathcal{N} \exp -(Av, v)/2\, dv$ where $\mathcal{N}$ is a constant of normalization and (Av, v) is a suitable quadratic form. We shall also write $Tu \equiv Q(u, u)$, to remind us that T is a nonlinear operator quadratic in u (*c.f.r.* (6.1.7)). With the latter notations we see that eq. (6.1.12) becomes

$$\begin{aligned}\mu(du) =&\mathcal{N} e^{-\frac12 (Au,u)} e^{-\frac12 (AQ(u,u),Q(u,u))-(Au,u),u)} du \equiv\\ \equiv& P_0(du)\, e^{-\frac12 (AQ(u,u),Q(u,u))-(AQ(u,u),u)}\end{aligned} \tag{6.1.13}$$

and the gaussian process $P_0(dv) = G(v)dv$ has covariance, *c.f.r.* the *first* term of the right hand side of (6.1.7):

$$\begin{aligned}&\int \overline{u}_{\underline{k},\alpha}(t) u_{\underline{h},\beta}(\tau) dP_0 = \int_{-\Theta}^{t} d\vartheta \int_{-\Theta}^{\tau} d\vartheta' e^{-\nu \underline{k}^2(t-\vartheta)} e^{-\nu \underline{h}^2(\tau-\vartheta')}\cdot\\ &\cdot \int dP_0\, \delta_{\underline{h}\,\underline{k}}\, (\Pi_{\underline{k},\alpha\alpha'}\, \overline{f}_{\underline{k}\alpha'}(\vartheta))(\Pi_{\underline{h},\beta\beta'} f_{\underline{h}\beta'}(\vartheta')) =\\ &= \delta_{\underline{h}\,\underline{k}} \int_{-\Theta}^{t} \int_{-\Theta}^{\tau} d\vartheta d\vartheta' e^{-\nu \underline{k}^2(t+\tau-2\vartheta)} \Pi_{\underline{k},\alpha\alpha'} \Pi_{\underline{h},\beta\beta'} \delta_{\alpha'\beta'} \frac{\gamma}{|\underline{k}|^a} \delta(\vartheta - \vartheta') =\end{aligned}$$

---

[2] One could object that all this ignores that although in the discretization of the integral we wrote it as a sum up to $j = t/t_0 - 1$ we could *equally well* have written it as sum up to $j = t/t_0$. And in this second way the elements on the diagonal *would not have all been* equal to 1. But they would have differed from 1 by a quantity proportional to t_0 that, formally, in the limit $t_0 \to 0$, would vanish, except possibly the term linear in t_0, if the determinant was computed as series in t_0. The linear term in t_0 however *is zero* because the sum on $\underline{k}_1$ and $\underline{k}_2$ escludes always that $\underline{k}_1 = \underline{k}$ or $\underline{k}_2 = \underline{k}$. Such last terms are those that would contribute to the value of the determinant at first order in t_0.

$$= \frac{\gamma}{|\underline{k}|^a}\delta_{\underline{k}\,\underline{h}}\Pi_{\underline{k}\,\alpha\beta}\int_{-\Theta}^{\min(t,\tau)} d\vartheta e^{-\nu\underline{k}^2(t+\tau-2\vartheta)} \equiv$$
$$\equiv \frac{\gamma}{|\underline{k}|^a}\delta_{\underline{k}\,\underline{h}}\,(\delta_{\alpha\beta}-\frac{k_\alpha k_\beta}{|\underline{k}|^2})\frac{e^{-\nu\underline{k}^2|t-\tau|}-e^{-\nu\underline{k}^2|t+\tau+2\Theta|}}{\nu\underline{k}^2} \tag{6.1.14}$$

with the convention of summation on repeated labels.

The gaussian distribution P_0 depends on Θ and we shall consider the limit P_∞ as $\Theta\to\infty$ of P_0. This will be a (formally) stationary gaussian distribution which should not be confused with the limit as $\Theta\to\infty$ of the non gaussian distribution μ in (6.1.10) or (6.1.13). It is now interesting to study first the process P_∞. More precisely we consider the process $\underline{u}(\underline{x},t)$ for $\Theta\to\infty$.

Such process is defined by its covariance

$$\int \overline{u}_{\underline{k},\alpha}(t)u_{\underline{h},\beta}(\tau)dP_\infty \stackrel{def}{=} \lim_{\Theta\to\infty}\int \overline{u}_{\underline{k},\alpha}(t)u_{\underline{h},\beta}(\tau)dP_0 \tag{6.1.15}$$

and we see that for $\Theta\to\infty$ the distribution converge, in the sense that its covariance $\langle\,\overline{u}_{\underline{k},\alpha}(t)u_{\underline{k},\beta}(\tau)\rangle$ tends to a limit that is interpreted as covariance of the gaussian process (invariant under time translation, obviously) with covariance, *c.f.r.* (6.1.14)

$$\langle\,\overline{u}_{\underline{k}\alpha}(t)u_{\underline{k}\beta}(\tau)\rangle = \frac{\gamma}{|\underline{k}|^a}\delta_{\underline{k}\,\underline{h}}(\delta_{\alpha\beta}-\frac{k_\alpha k_\beta}{|\underline{k}|^2})\frac{e^{-\nu\underline{k}^2|t-\tau|}}{\nu\underline{k}^2}\equiv$$
$$\equiv \frac{2\gamma}{|\underline{k}|^a}\delta_{\underline{k}\,\underline{h}}(\delta_{\alpha\beta}-\frac{k_\alpha k_\beta}{|\underline{k}|^2})\int_{-\infty}^{\infty}\frac{dk_0}{2\pi}\frac{e^{ik_0(t-\tau)}}{k_0^2+(\nu\underline{k}^2)^2} \tag{6.1.16}$$

as we deduce simply by computing the integral with the method of the residues. Introducing the Fourier transform $u_{\underline{k},\alpha,k_0}$ in time of $u_{\underline{k},\alpha}(t)$ defined by setting $u_{\underline{k},\alpha}(t)=\int_{-\infty}^{\infty}dk_0\,u_{\underline{k},\alpha,k_0}\,e^{-ik_0\,t}$ the (6.1.16) can also be written as an equality between the covariance $\langle\,\overline{u}_{\underline{k},\alpha,k_0}u_{\underline{h},\beta,h_0}\rangle$ and $\gamma\pi^{-1}\delta_{\underline{k}\,\underline{h}}(\delta_{\alpha,\beta}-k_\alpha k_\beta/\underline{k}^2)\,\delta(k_0-h_0)/(k_0^2+(\nu\underline{k}^2)^2)$.

Hence if A is the matrix of the quadratic form (Au,u) defined by

$$\frac{2\pi}{2\gamma}\int_\Omega\int_{-\infty}^{\infty}\frac{d\underline{x}\,dt}{2\pi L^3}(-\Delta)^{a/2}(\partial_t^2+\nu^2\Delta^2)\underline{u}(\underline{x},t)\cdot\underline{u}(\underline{x},t)=(Au,u) \tag{6.1.17}$$

i.e. :

$$A=\frac{\pi}{\gamma}(-\Delta)^{a/2}(\partial_t^2+\nu^2\Delta^2)\equiv\frac{\pi}{\gamma}(-\Delta)^{a/2}(\partial_t-\nu\Delta)(-\partial_t-\nu\Delta) \tag{6.1.18}$$

we see that $P_\infty(du)$ is a Gaussian process defined by the operator A on the space of the *zero divergence velocity fields.*

Turning our attention to the actual velocity field which we have also called u, *c.f.r.* (6.1.13), it follows that the asymptotic distribution of the field u is, *c.f.r.* (6.1.13)

$$\mu(du)=P_\infty(du)\,e^{-\frac{1}{2}(AQ(u,u),Q(u,u))-(AQ(u,u),u}} \tag{6.1.19}$$

that provides us with a first *formal expression* for μ.

The quadratic form $Q(u,u)$ can be expressed, in the limit $\Theta\to\infty$, as

$$\begin{aligned}Q(\underline{u},\underline{u})_{\underline{k}}(t)=&\int_{-\infty}^{t} d\tau e^{-\nu\underline{k}^2(t-\tau)}\,\Pi_{\rm rot}\left(\underset{\sim}{\underline{u}}(\tau)\cdot\underset{\sim}{\partial}\,\underline{u}(\tau)\right)_{\underline{k}}\equiv\\ \equiv&\int_{-\infty}^{\infty}d\tau\int_{-\infty}^{\infty}\frac{dk_0}{\pi}\frac{e^{ik_0(t-\tau)}}{ik_0+\nu\underline{k}^2}\,\Pi_{\rm rot}\left(\underset{\sim}{\underline{u}}(\tau)\cdot\underset{\sim}{\partial}\,\underline{u}(\tau)\right)_{\underline{k}}=\\ =&\Big(\frac{1}{(\partial_t-\nu\Delta)}\Pi_{\rm rot}\,\underset{\sim}{\underline{u}}\cdot\underset{\sim}{\partial}\,\underline{u}\Big)_{\underline{k}}\end{aligned}\tag{6.1.20}$$

because if $\tau>t$ the integral over k_0 in the intermediate term of (6.1.20) vanishes (by integrating on k_0 with the method of the residues); hence

$$\begin{aligned}&(AQ(u,u),Q(u,u))=\frac{\pi}{\gamma}\Big((-\Delta)^{a/2}(\underset{\sim}{\underline{u}}\cdot\underset{\sim}{\partial}\,\underline{u}),\Pi_{\rm rot}\,(\underset{\sim}{\underline{u}}\cdot\underset{\sim}{\partial}\,\underline{u})\Big)\\ &(AQ(u,u),u)=\frac{\pi}{\gamma}\Big((-\Delta)^{a/2}(\partial_t-\nu\Delta)(\underset{\sim}{\underline{u}}\cdot\underset{\sim}{\partial}\,\underline{u}),\underline{u}\Big)\end{aligned}\tag{6.1.21}$$

meaning that this is a function on the zero divergence fields $\underline{u}(\underline{x},t)$ defined for $\underline{x}\in\Omega$, $t\in(-\infty,+\infty)$.

The conclusion is that the probability distribution on the space of velocity fields with zero divergence

$$\begin{aligned}\mu(d\underline{u})=&\mathcal{N}\,e^{-\frac{(2\pi L^3)^{-1}}{2\gamma/\pi}\int d\underline{x}dt\big((-\Delta)^{a/2}(\partial_t^2+\nu^2\Delta^2)u,u\big)}\cdot du\cdot\\ &\cdot e^{-\frac{(2\pi L^3)^{-1}}{\gamma/\pi}\int d\underline{x}dt\big[\big(\frac12(-\Delta)^{a/2}\Pi_{\rm rot}\,(\underset{\sim}{\underline{u}}\cdot\underset{\sim}{\partial}\,\underline{u}),(\underset{\sim}{\underline{u}}\cdot\underset{\sim}{\partial}\,\underline{u})\big)-\big((-\Delta)^{a/2}(-\partial_t-\nu\Delta)(\underset{\sim}{\underline{u}}\cdot\underset{\sim}{\partial}\,\underline{u}),\underline{u}\big)\big]}\end{aligned}\tag{6.1.22}$$

is a formally invariant measure for the NS equation subject to a random force, with covariance in space–time

$$\frac{\gamma/\pi}{(-\Delta)^{a/2}(k_0^2+\nu^2\Delta^2)}\delta_{\underline{k}\,\underline{k}'}(\delta_{\alpha\beta}-\frac{k_\alpha k_\beta}{|\underline{k}|^2})\,\delta(k_0-k_0')\tag{6.1.23}$$

and the case $a=0$ corresponds to a *white noise*, while the cases $a\neq0$ will be called *colored noises* of color $a/2$.

If *a is even and positive* we have a very interesting case in which the functional integral (6.1.23) is, from a formal viewpoint, a *local field theory*. In this case the quadratic forms in (6.1.22) indeed are expressed as integrals of products of velocity fields and their derivatives: all of them computed at the *same point* of space–time.[3] This no longer true if $a/2$ is not an integer ≥1 (because $(-\Delta)^{a/2-1}$ is not a local operator). Here one should keep in mind that heuristic arguments suggest that a most interesting case should be the odd color $a=3$.

[3] Note that $(-\Delta)^a\Pi_{\rm rot}$ is an ordinary differential operator if a is even ≥2 because its Fourier transform is the matrix $|\underline{k}|^a\delta_{\alpha\beta}-k_\alpha k_\beta|\underline{k}|^{a-2}$, $\alpha,\beta=1,2,3$.

Here, as in all formal field theories, it is necessary to clarify what one should exactly understand when considering the functional integral (6.1.22). And this must be done even if we wish to, or are willing to, forget the mathematical check that it *really* defines a probability distribution invariant for the evolution generated by the NS equation (with noise).[4]

Before attempting to set up this ("overwhelming") task, it is good to stop to assess the situation. Precisely we must ask which could be the interest of what we are saying and writing (or perhaps reading).

Why to modify and widen the problem by considering random volume forces, when what interests us is the case of constant volume forces?

First of all it is easy to add a constant force $\underline{g}$ (that we shall suppose with zero divergence (without affecting the generality). It suffices indeed to replace in (6.1.7) $\underline{f}_{\underline{k}}$ with $\underline{f}_{\underline{k}} + \underline{g}_{\underline{k}}$ and proceed, without hesitation, in the same way. We see immediately that in such case the formally invariant distribution, that we denote $\mu_{\underline{g}}(du)$, is written

$$\mu_{\underline{g}}(du) = \mu(du)\, e^{-\nu^{-1}(\Delta^{-1}A\underline{g},u-Q(u,u))} \tag{6.1.24}$$

where $\mu(du)$ is the (6.1.22): $\mu \equiv \mu_{\underline{0}}$.
Hence an invariant distribution in presence of external force $\underline{g}$ and *absence* of noise can be obtained as the *limit* ($\gamma \to 0$) of a field theory with Lagrangian $\mathcal{L}(\underline{u})$

$$\begin{aligned}
&-\frac{\gamma}{\pi}\mathcal{L}(\underline{u},\dot{\underline{u}}) = \int \frac{d\underline{x}dt}{2\pi L^3}\Big(-\frac{1}{2}|(-\Delta)^{a/4}(\partial_t+\nu\Delta)\underline{u}|^2-\\
&-\frac{1}{2}|(-\Delta)^{a/4}\Pi_{\rm rot}\,\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{u}|^2-\big((-\Delta)^{a/2}(\partial_t-\nu\Delta)(\underset{\sim}{u}\cdot\underset{\sim}{\partial}\,\underline{u})\big)\cdot\underline{u}\Big)+\\
&+\nu^{-1}([(-\Delta)^{-1+a/2}(-\partial_t^2+(\nu\Delta)^2)\underline{g}],\underline{u})-\\
&-\nu^{-1}\big([(-\Delta)^{a/2}(\partial_t-\nu\Delta)\underline{g}],(\underset{\sim}{u}\cdot\underset{\sim}{\partial}u)\big)
\end{aligned}\tag{6.1.25}$$

This limit could be interpreted as a probability distribution concentrated on the fields that minimize the $\mathcal{L}$ in (6.1.25). They are solutions of the differential equations that are obtained by imposing a stationarity condition on the functional (6.1.25). *The color parameter can be chosen arbitrarily in the sense that it will be possible, for any value a, to interpret every limit as a distribution invariant for the evolution of NS.*
It is, however, difficult to see what has been gained.
The extremes of the action built with the lagrangian $\mathcal{L}$ are, formally, invariant measures but they obey to differential equations (if a is even, at least) that must be equivalent to the initial equation of NS, or better to its

[4] Which would lead us back to the theory of exixtence and uniqueness of the solution of the NS equation which we have already seen to be still in dire need of new ideas.

"dual" version as equation of evolution for probability distributions on the space of velocity fields. We shall not write it, but it is known that searching for solutions of this type did not yet turn out to be more generous in terms of results than the search, also quite inconclusive, of solutions of the NS equation itself.

In reality the interest of the above analysis lies, rather, in the possibility of using what has been recently understood in the field theory, after the development of the methods of analysis based on the *renormalization group*, in the case $\gamma > 0$, for instance *c.f.r.* [BG95].

The possibility arises that for $\nu \to 0$ and $\underline{g}$ fixed, *i.e.* as the Reynolds number tends to infinity, the results could become *independent* of the value of γ, at least some of them among those of physical relevance and at least for what concerns the distribution of the components $\underline{u}_{\underline{k}}$ with $\underline{k}$ not too large.

At $\underline{g}$ fixed, $\nu \to 0$ and γ small we see that one studies the problem of what happens at large Reynolds number in presence of a small noise. In a certain sense *we study, therefore, properties of the attractors which are stable with respect to (certain) random perturbations.* Hence we see that, if we succeeded in giving some answer to the question of the existence and properties of the distributions (6.1.24) we would obtain the answer to a problem that perhaps is even the most interesting among those that we can pose.

Until now research in the direction of studying the integral (6.1.24) has not borne relevant or *unambiguous* fruits. But the matter continues to attract the attention of many. Among the attempts we quote here [YO86] where functional integration is not explicitly used because the work relies on the different but equivalent approach of perturbation expansions (we could say via an explicit evaluation of Feynman diagrams, although the diagrams are not even mentioned in the quoted paper): approximations must be made and so far it is not yet clear how to estimate the neglected "marginal terms", *c.f.r.* p. 48 in [YO86]. Hence it is important to try to find on heuristic bases what we could "reasonably" expect from a future theory. Hopefully we shall reach some better understanding within a not too remote future.

The treatment parallel to what has been discussed above in the much simpler case of the *Stokes equation*

$$\begin{aligned} \dot{\underline{u}} &= \nu\Delta\underline{u} - \underline{\partial}p + \underline{g} + \underline{f}(t) \\ \underline{\partial}\cdot\underline{u} = 0, &\qquad \int \underline{u} = \underline{0} \end{aligned} \tag{6.1.26}$$

is left as a problem for this section (see problems below): this is a case which is easy but very interesting and instructive: its simplicity makes useless a formal discussion.

Problems.

[6.1.1]: Study the invariant distributions of the evolution defined by the equation, defined on the circle $x \in [0, 2\pi]$, $\dot{u} = \partial_x^2 u + f(t)$ with $f(t)$ a noise of color $c = a/2$ and average zero and with initial data with zero average, along the lines of what has been discussed above. (*Idea:* Note that the theory is purely Gaussian. Compute the covariance.)

[6.1.2]: Study formally along the lines discussed for the NS equation with random noise the invariant distributions for the evolution defined by the equation on the line $\dot{u} = -u + \partial_x^2 u - u^3 + f(t)$ with $f(t)$ a noise of color $a/2$. Derive the lagrangian of the field theory generated by the applicaton of Ghirsanov's formula (6.1.12) and the lagrangian corresponfing to (6.1.15).

[6.1.3] Formulate the theory of the invariant distributions for the equation of Stokes (6.1.26), along the lines of what discussed above. (*Idea:* Note that the theory is purely Gaussian and compute the covariance of the velocity field distribution).

[6.1.4] Compute, in the case of the problem [6.1.1], the behavior of the function of $\underline{x}, \underline{y}$ defined by $\langle(\underline{u}(\underline{x}, t) - \underline{u}(\underline{y}, t))^2\rangle$ when $|\underline{x} - \underline{y}| \to 0$, as a function of the color of the noise: the average is understood over the stationary distribution for the equations (6.1.26) constructed in the problem [6.1.3]. (*Idea:* Study the covariance of the Gaussian distribution in [6.1.3]).

[6.1.5] As in the problem [6.1.4] but for the function $\langle|\underline{u}(\underline{x}, t) - \underline{u}(\underline{y}, t)|^3\rangle$.

[6.1.6] Using the technique of problem [2.4.7] in §2.4 (Wiener theorem) discuss the class of regularity in $\underline{x}$ at fixed t of the samples of the gaussian process in problem [6.1.3] showing that they are Hölder continuous fields with an exponent that can be bounded below in terms of the color of the noise $a/2$ and that becomes positive for a large enough, say $a \geq a_0$ and estimate a_0. (*Idea:* Remark that the covariance becomes regular if a is large enough and then proceed as in the quoted problem of §2.4 to prove the Hölder continuity, with probability 1 of the Brownian motion.)

Bibliography: The analysis in this section is classical, see for example [An90]. For further developments see, for instance, [YO86]. A quote from the latter reference, p. 47, can be interesting in order to develop some intuition about the philosophy behind the more recent attempts at using the renormalization group ("RNG") to study turbulence statistics (however the quote can be better understood after reading the next section where the notion of inertial range is introduced): "*The RNG method developed here* [in [YO86]] *is based on a number of ideas. First there is the correspondence principle, which can be stated as follows. A turbulent fluid characterized in the inertial range by scaling laws can be described in this inertial range by a corresponding Navier–Stokes equation in which a random force generates velocity fluctuations that obey the scaling of the inertial range of the original unforced system. Second*" ... "*We believe that the results of the RNG fixed–point calculations can be applied to any fluid that demonstrates Kolmogorov–like scale–invariant behavior in some range of wavevectors and frequencies. This situation resembles the theory of critical phenomena in the sense that the critical exponents computed at the fixed points are approximately valid in the vicinity of the critical point where* $(T - T_c)/T_c \ll 1$."

Very recent results on the existence of stationary distributions have been obtained in the 2–dimensional case, [KS00], [BKL00]. These works essentially solve completeley some of the problems posed in this section.

§6.2 Phenomenology of developed turbulence and Kolmogorov laws.

> *Perhaps the most striking of all properties of three–dimensional turbulence is that decrease of the viscosity, with no change of the conditions of generation of the turbulence, is accompanied by increase of the mean square vorticity (the process of magnification of vorticity by extension of vortex lines then being restrained by viscous diffusion of vorticity), and that in the limit $\nu \to 0$ the rate of energy dissipation is conserved (Batchelor),*[Ba69].

Consider a Navier–Stokes fluid in a cube $\Omega = [0, L]^d$, $d = 2, 3$, with periodic boundary conditions and subject to a constant force $\underline{g}$, regular enough so that only its Fourier components $\underline{g}_{\underline{k}}$ with $|\underline{k}| \approx 2\pi/L$ are appreciably different from 0. We say that the system is "forced on the *scale* L of the container", or that the fluid receives energy from the outside "on scale" L. We shall study the stationary state of the fluid as we decrease the value ν of the viscosity.

(A): Energy dissipation: inertial and viscous scales.

The constant external force sustains motion and not only it injects into the fluid an energy ε_ν, in average, per unit time and unit volume, but it also has the effect that the velocity components which are appreciably not zero increase in number while the viscosity ν decreases. Hence the components of the Fourier transform of the velocity field which are appreciably not zero will be those with $|\underline{k}| < k'_\nu$ where k'_ν is a suitable scale (with dimension of an inverse length). The length scale $(k'_\nu)^{-1}$ will be called the *viscous scale*. We recall that our conventions on the Fourier transform are such that $\underline{u}(\underline{x})$ and its transform $\underline{u}_{\underline{k}}$ are related by $\underline{u}(\underline{x}) = \sum_{\underline{k}} e^{i\,\underline{k}\cdot\underline{x}}\,\underline{u}_{\underline{k}}$. We shall also fix the density ρ to be $\rho = 1$.

In the NS–equation appear both the *transport term* also called *inertial term*, $\underset{\sim}{u} \cdot \underset{\sim}{\partial}\,\underline{u}$, and the *friction term*, $\nu\,\Delta\,\underline{u}$, also called *viscous term*: the latter contributes to the $\underline{k}$–mode of the Fourier transform of the equation a quantity of size $\nu\,\underline{k}^2\,|\underline{u}_{\underline{k}}|$; while the inertial term shall contribute to the same Fourier component a quantity of order of magnitude $\simeq |\underline{k}||\underline{u}_{\underline{k}}|^2$ at least. Then the friction terms become dominant over the inertial ones at large $|\underline{k}|$: the cross over takes place at values of $\underline{k}$ where (in average over time) $\nu|\underline{k}|^2\,|\underline{u}_{\underline{k}}| \approx |\sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \underset{\sim}{u}_{\underline{k}_1} \cdot \underline{k}_2 \Pi_{\underline{k}} \underline{u}_{\underline{k}_2}|$, which is a criterion to determine what we shall call the "*viscous scale*" k'_ν. The transition will not be sharp hence we introduce another scale $k_\nu < k'_\nu$ *of the same order of magnitude* so that for $\underline{k}$ in the buffer region between k_ν and k'_ν the inertial and the viscous terms will be comparable (in average over time). It will be supposed that $\underline{u}_{\underline{k}} \equiv \underline{0}$ for $|\underline{k}| > k'_\nu \xrightarrow[\nu\to 0]{} \infty$. We shall set $g = \max_{\underline{k}} |\underline{g}_{\underline{k}}|$, keeping in mind that the forcing $\underline{g}_{\underline{k}}$ differs from $\underline{0}$ sensibly only if $|\underline{k}|$ is close to its minimum value $|\underline{k}| \simeq 2\pi L^{-1} \stackrel{def}{=} k_0$.

If the external force is fixed the low $|\underline{k}|$ modes are certainly not viscous. In fact we could expect that (in the average over time) $|\underline{u}_{\underline{k}}| \gg \nu|\underline{k}|$ for ν small and $|\underline{k}| \approx k_0$ because the external force will be able to produce on the Fourier components of the velocity on which it acts directly velocities of order $|\underline{g}_{\underline{k}}|L^2/\nu$ (because $\nu\underline{k}^2\underline{u}_{\underline{k}} \approx \underline{g}_{\underline{k}}$ would be the time independent solution in absence of inertial terms and $\underline{g}_{\underline{k}}$ is different from $\underline{0}$, appreciably, only if $|\underline{k}| \approx 2\pi/L$).

However $gL^2\nu^{-1}$ is not the only magnitude with the dimensions of a velocity that can be formed with the parameters of the problem (*i.e.* g, L, ν): another one is $\sqrt{gL}$. It seems from the experimental data that, as ν tends to 0, the stationary state of the fluid dissipates per unit time and volume a quantity $\eta = \int \underline{g}\cdot\underline{u}\,d\underline{x} \simeq L^3 g\,\langle|\underline{u}_{\underline{k}_0}|\rangle$ of energy that behaves as a power, negative or possibly 0, of ν: hence we must think that the average velocity on the macroscopic scale ($\langle|\underline{u}_{\underline{k}}|\rangle$ with $|\underline{k}| \simeq k_0 = 2\pi L^{-1}$) also behaves as a power of ν. If it approaches a limit value as $\nu \to 0$ (*i.e.* the exponent of the power of ν is 0), *c.f.r.* [Ta35] and [Kr75a] p. 306, then such limit must be of order $\sqrt{gL}$, which is the only ν–independent velocity that can be formed with the available parameters.

A theoretical confirmation of the existence of a "saturation value" of the dissipation per unit time, as viscosity tends to zero, has been obtained in some cases: *e.g.* in the case (*different from the above* because there is no volume force) of the NS equation for a fluid between two parallel plates in motion with parallel (but different) velocities: *c.f.r.* [DC92]. A simple (not optimal) technique for estimating the dissipation rate, as $\nu \to 0$, in a forced system with constant volume force, is in problems [6.2.1], [6.2.2].

(B): Digression on the physical meaning of "$\nu \to 0$".

It is convenient to make a short digression on the question of viscosity dependence of various quantities (like the average dissipation) in the limit in which the latter vanishes.

It is clear that it is not possible in experiments, other than numerical, to vary viscosity so that it tends to zero. But experiments on Navier–Stokes fluids (*i.e.* that remain well described by NS equations as the experimental parameters vary) can be used to infer what would happen if ν was as variable a parameter as wished. Let us examine two interesting cases.

Consider a fluid in a cubic container of side L, with horizontal *periodic* conditions but with conditions of adherence to the upper and lower walls. The first is supposed to move with constant velocity U while the second is fixed. This is a model (*"shear flow"*) simpler than that of a fluid between two cylinders one of which is fixed and the other rotating (*"Couette flow"*)

Then the Reynolds number is $R = UL\nu^{-1}$. Setting $\underline{u}(\underline{x},t) = \frac{\nu}{L}\underline{w}(\frac{\underline{x}}{L},\frac{\nu t}{L^2})$, *c.f.r.* §1.3, one checks that the equation for $\underline{w}$ becomes that of a fluid with $\nu = 1, L = 1$ and with shear velocity $U' = R$. Hence we see how a real experiment on a fluid with $\nu = 1, L = 1$ and shear velocity $U' = R \to \infty$, conceivable as a laboratory experiment, can give informations on a fluid in which U is fixed and $\nu \to 0$. The relation between the dissipation per unit time and volume of the two fluids, related by the described rescaling of variables, is simple and it is $\eta = \nu L^{-3}\int |\underset{\sim}{\partial}\underline{u}|^2 d\underline{x} = \nu(\nu L^{-1})^2\eta_0(R)$, if $\eta_0(R)$ is the dissipation of the "rescaled" fluid (*i.e.* with unit volume and viscosity).

Therefore if the dissipation η is independent of ν in the limit $\nu \to 0$ this means that $\eta_0(R) \propto R^3$ for $R \to \infty$: in [DC92] the dissipation η is *bounded* above independently of ν in the above shear flow problem.

A second case is a fluid in a periodic cubic container of side L subject to a force $F\underline{g}(\underline{x}/L)$ with $\underline{g}$ fixed (the dependence via $\underline{x}/L$ means that the force acts on "macroscopic scales"). The rescaling of the variables described in the preceding case leads us to say that the rescaled fluid flows in a container of side $L = 1$, with viscosity $\nu = 1$ and subject to a force $FL^3\nu^{-2}\underline{g}(\underline{x}) \equiv R^2\underline{g}(\underline{x})$, where $R = \sqrt{FL^3\nu^{-2}}$. The relation between the dissipation $\eta_0(R)$ in the rescaled fluid and that in the original fluid is still $\eta = \nu^3L^{-4}\eta_0(R)$, and hence the ν–independence (if verified) corresponds to a proportionality of $\eta_0(R)$ to R^3.

The theory that follows will not be based on any hypothesis on the (average) dissipation per unit time and volume as $\nu \to 0$: the validity of any assumption on this quantity (*e.g.* that it tends to a constant) cannot be taken for granted and it should be analyzed on a case by case basis as we are far from a full understanding of it.

(C): The K41 tridimensional theory.

> *The physical picture is that large–scale motions should carry small eddies about without distorting them. It is not obvious that this need be true, but the idea is certainly intuitively plausible, (Kraichnan)* [Kr64].

Coming back to the theme of this section we note that the presence of the inertial terms causes an appreciable dispersion of energy that, in their absence, would stay confined to the modes on which the external force acts: but there is no reason that the average (in time) amplitude of $|\underline{u}_{\underline{k}}|$ for $\underline{k}$ on scale $k_0 = 2\pi L^{-1}$ is not, in a stationary regime, monotonically increasing as ν decreases, and that at the same time the average size of $|\underline{u}_{\underline{k}}|$ is decreasing to 0 as $\underline{k}$ increases to ∞ (at fixed ν). This, alone, is already sufficient to state that it will be possible to neglect friction forces until down to scales $k_\nu \xrightarrow[\nu\to 0]{} \infty$, with k_ν smaller than the momentum k'_ν of the *viscous* scales (where viscous terms become dominant): $k_0 \ll k_\nu < k'_\nu$. In fact, as noted already the ratio between inertial and friction terms is, dimensionally, $|\underline{u}_{\underline{k}}|/\nu\,|\underline{k}|$.

The scale k_ν will be called *Kolmogorov scale*, the scale k'_ν will be called *viscous scale*, but what said until now *does not, yet, allow* us to determine the size of the scales k_ν, k'_ν. They will be determined below on dimensional grounds after discussing the key hypothesis of the theory, namely the "*homogeneity*" of turbulence, *c.f.r.* (6.2.6).

Hence we get to the point of regarding turbulent motion as well described by the equations (if $k_0 \equiv 2\pi L^{-1}$)

$$\begin{aligned}
\dot{\underline{u}}_{\underline{k}} &= -\,i \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \underset{\sim}{u}_{\underline{k}_1}\cdot \underset{\sim}{k}_2 \Pi_{\underline{k}} \underline{u}_{\underline{k}_2} + \underline{g}_{\underline{k}} && k_0 \le |\underline{k}| \le k_\nu\\
\dot{\underline{u}}_{\underline{k}} &= -\,\nu \underline{k}^2 \underline{u}_{\underline{k}} - i \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \underset{\sim}{u}_{\underline{k}_1}\cdot \underset{\sim}{k}_2 \Pi_{\underline{k}} \underline{u}_{\underline{k}_2} && k_\nu < |\underline{k}| < k'_\nu \qquad (6.2.1)\\
\dot{\underline{u}}_{\underline{k}} &= -\,\nu \underline{k}^2 \underline{u}_{\underline{k}} && |\underline{k}| > k'_\nu
\end{aligned}$$

that are read by saying that the motion of a fluid at high Reynolds number, *i.e.* at small viscosity, is described by the Euler equation on the scales of length larger than the Kolmogorov scale but it is, instead, described by the Stokes equation on scales smaller than the viscous scale.

One says that

(1) the scales of length are *inertial* for $k_0 \ge |\underline{k}| \ge k_\nu$ and *viscous* if $|\underline{k}| > k'_\nu$. The intermediate modes $k_\nu < |\underline{k}| < k'_\nu$ can be called *dissipative modes* because it is "in them" that the dissipation of the energy provided by the forcing takes place, and

(2) "turbulence" is due to the inertial modes, while the motion of the dissipative and viscous modes is always "laminar". The motion of the viscous modes is trivial and in fact we can set $\underline{u}_{\underline{k}} = \underline{0}$ for the viscous components, *c.f.r.* (6.2.1).

(3) The "role" of the dissipative modes is to absorb energy from the inertial modes and dissipate it.

The problem is now to determine k_ν, k'_ν. Imagine drawing in the space of the modes[1] $\underline{k}$ the sphere of radius $\kappa < k_\nu$. We can then define, from (6.2.1), the energy E^κ per unit time that "exits from the sphere" (of momenta) of radius κ by setting

$$E^\kappa \stackrel{def}{=} \frac{1}{2}\frac{d}{dt} L^3 \sum_{|\underline{k}|<\kappa} |\underline{u}_{\underline{k}}|^2 = \mathcal{L} - \mathcal{E}_{\kappa,\kappa'} - \mathcal{E}_{\kappa,\kappa',\infty} \qquad (6.2.2)$$

[1] One often confuses the "mode" $\underline{k}$, that is a label with the dimension of an inverse length, and the Fourier transform component $\underline{u}_{\underline{k}}$ that is, instead, a dynamical observable with the dimension of a velocity. I will not make here any distinction, because it is not possible (reasonably) to be worried by the ambiguity that one could fear: "temer si dee di sole quelle cose c' hanno potenza di fare altrui male; dell' altre no, chè non son paurose" (II, 88, *Inferno*), *we should fear only about that which can cause harm, not of that which cannot as it is not scary.*

where $\mathcal{L}$ is the work done by the external force $\underline{g}$ per unit time ($\mathcal{L} = L^3 \sum_{\underline{k}} \overline{\underline{g}_{\underline{k}}} \cdot \underline{u}_{\underline{k}}$ equal to $2L^3 \mathrm{Re}\, \overline{\underline{g}_{\underline{k}_0}} \cdot \underline{u}_{\underline{k}_0}$ if $\underline{g}$ has only the component $\underline{k}_0$ in its Fourier transform), and

$$\mathcal{E}_{\kappa,\kappa'} = iL^3 \sum_{|\underline{k}_3|<\kappa} \sum^{*}_{\underline{k}_1+\underline{k}_2+\underline{k}_3=\underline{0}} \underset{\sim}{u}_{\underline{k}_1} \cdot \underset{\sim}{k}_2\, \underline{u}_{\underline{k}_2} \cdot \underline{u}_{\underline{k}_3} \tag{6.2.3}$$

and $\sum^*$ means that $|\underline{k}_2|$ is in the interval $[\kappa,\kappa')$ and $|\underline{k}_1|$ is in $[k_0,\kappa')$. Finally $\mathcal{E}_{\kappa,\kappa',\infty}$ is given by a similar expression with $|\underline{k}_2|$ in $[\kappa',\infty)$ or $|\underline{k}_2| \in [\kappa,\kappa')$ and $|\underline{k}_1| \in [\kappa',\infty)$. The apparently missing terms have zero sum because $\underline{k}_3 + \underline{k}_2 = -\underline{k}_1$ is orthogonal to $\underline{u}_{\underline{k}_1}$, as already seen several times (*c.f.r.* (3.2.14), for example).

One can read (6.2.2) by saying that the energy on scales $< \kappa$ changes because of the work "done by the external force" (expressed by $\mathcal{L}$), because of the work that "the modes with $|\underline{k}| < \kappa$ perform on those with $|\underline{k}| \in [\kappa,\kappa')$ "directly" (expressed by $\mathcal{E}_{\kappa,\kappa'}$) and because of the work that "the modes with $|\underline{k}| < \kappa$ perform on those with $|\underline{k}| \in [\kappa',\infty)$" (expressed by $\mathcal{E}_{\kappa,\kappa',\infty}$).

Hypothesis *(homogeneous turbulence): The fundamental hypothesis of the Kolmogorov's theory is that $\mathcal{E}_{\kappa,\kappa',\infty}$ is, "in average", zero if κ and $\kappa' = 2\kappa$ are $< k_\nu$, at least for the motions in a asymptotic regime which, therefore, develop on an attracting set.*

Hence it really makes sense to say that the energy is dissipated only on scales $|\underline{k}| > k_\nu$, "cascading" without sensible dissipation from the large length scales to the small ones.

The hypothesis, which has to be understood as an asymptotic property as $R \to \infty$, does not have a justification other than its simplicity and elegance (besides being, perhaps, very natural) and, *a posteriori*, its eventual self consistency and, even better, its adherence to experimental observations. We shall see, indeed, that it has remarkable implications that can be tested experimentally.

Alternative hypotheses, apparently also reasonable, are possible and lead to *qualitatively different* results, *c.f.r.* (4.1) and (4.3) in [Kr64], see also §6.3.

Note that the assumption is indedependent on whether the dissipation per unit time and volume tends to a constant as $R \to \infty$ or behaves as a power of ν or R.

A way to read the hypothesis is the following: "in the inertial regime ($|\underline{k}| < k_\nu$) we neglect energy exchanges between scales of momentum which are *not* contiguous" (*locality of the energy cascade*), *c.f.r.* [Kr64], [Ga99d]. Or "the smaller vortices are passively transported by the larger ones that contain them": in this form the hypothesis does not seem well in agreement with the visual experience, at least not with the experience on turbulent motions in which the small vortices are such only in one or two dimensions while in the other, or in the other two, they have dimensions of many orders of magnitude larger and hence their motion is "distorted" by what happens on very different ("not contiguous") length scales. The phenomenon is sometimes called *intermittency.* Therefore we should expect that the theory developed upon the above homogeneous turbulence hypothesis will need corrections.

On the basis of the above heurisic analysis we shall suppose that $\varepsilon_\nu \equiv \varepsilon \equiv L^{-3}E^\kappa$ is constant for κ in the range $k_0 \ll \kappa < k_\nu$. We must now remark that, for what concerns the average properties of the Fourier components of the velocity field with mode $\underline{k}$ of any order κ with $\kappa \in (k_0, k_\nu)$, it *must be equivalent* to force the system on a scale of length L or on another scale, smaller but still $\gg \kappa^{-1} \gg k_\nu^{-1}$, *provided* the quantity of energy dissipated per unit time and mass is always ε.

It follows that ε must be expressible, on all scales $\kappa < k_\nu$, only in terms of quantities that pertain the scale $l = \kappa^{-1}$ *excluding* viscosity (that does not enter in the first of (6.2.1) that describes the motion of the components on scales $\kappa < k_\nu$). Hence if v_l is a velocity variation characteristic of the scale l we must have, for dimensional reasons

$$\varepsilon \simeq v_l^3 l^{-1} \tag{6.2.4}$$

because this is the only quantity with the dimension of energy dissipation per unit volume and unit time that can be formed with quantities characteristic of the length scale l; equation (6.2.4) must hold, in particular, on scale $\simeq L$.

Let us define v_l precisely as a time average

$$v_l^2 \stackrel{def}{=} \langle \big(|\Delta|^{-1} \int_\Delta (\underline{u}(\underline{x}) - \underline{u}(\underline{x}_0))d\underline{x}\big)^2 \rangle \tag{6.2.5}$$

where Δ is a small cube arbitrarily centered and side $l \equiv \kappa^{-1}$; and (in the stationary states) assume that

(1) the components of the Fourier transform of the velocity relative to different modes $\underline{k}$ are statistically independent,
(2) the velocity variations in disjoint little cubes Δ are statistically independent,
(3) the variables $\underline{u}_{\underline{k}}$, components of the Fourier transform of $\underline{u}$, are essentially equally distributed if $|\underline{k}|$ has a given order of magnitude $k_\nu \gg |\underline{k}| \gg k_0$.

We then see that the energy "contained in the scale k", *i.e.* the energy of the modes $\underline{k}$ such that $k_0 \ll k < |\underline{k}| < 2k \ll k_\nu$, is approximately $l^3 v_l^2$ times the number of the small cubes of side $l = k^{-1}$ that pave Ω (a number that is of order $(kL)^3$, having made use of the hypothesis (2)).

On the other hand, from hypothesis (1) and with the normalizations in (2.2.2) on the Fourier transforms, we see that $v_l^2 \simeq \langle (\sum_{k<|\underline{k}|<2k} \underline{u}_{\underline{k}})^2 \rangle$ is proportional to the number of modes $\underline{k}$ between k and $2k$ (*i.e.* $\propto (kL)^3$), by the central limit theorem and by assumptions (1) and (3). Hence

$$v_l^2 \sim (kL)^3 \langle |\underline{u}_{\underline{k}}|^2 \rangle \qquad \text{and} \qquad v_l^3 = \varepsilon l, \qquad l = |\underline{k}|^{-1} \tag{6.2.6}$$

if $\langle|\underline{u}_{\underline{k}}|^2\rangle$ is the average quadratic value (with respect to time or to the statistics of the attracting set that describes the asymptotic motion) of a single Fourier component of the velocity, (such average does not depend on $\underline{k}$ for $k<|\underline{k}|<2k$, by the hypothesis (3)).

Hence the average energy $L^3K(k)dk$ contained between k and $k+dk$ is (if $l=k^{-1}$) is given by

$$L^3K(k)dk=L^3\sum_{k<|\underline{k}|<k+dk}\langle|\underline{u}_{\underline{k}}|^2\rangle=\sum_{k<|\underline{k}|<k+dk}\frac{v_l^2}{k^3}=\frac{1}{k^3}\big(\frac{\varepsilon}{k}\big)^{2/3}\frac{4\pi k^2\,dk}{(2\pi/L)^3} \tag{6.2.7}$$

because $4\pi k^2\,dk/k_0^3$, $k_0=2\pi/L$, is the number of modes between k and $k+dk$. It follows that

$$K(k)=cost\;\varepsilon^{2/3}\,k^{-5/3},\qquad k_0\ll k\ll k_\nu \tag{6.2.8}$$

is the *energy density* per unit of $|\underline{k}|$ and per unit of mass: "*law* 5/3 *of Kolmogorov*".[2]

We also say that the energy spectrum (*i.e.* $K(k)$) is concentrated at small k, while the vorticity spectrum (*i.e.* $K(k)k^2$) is concentrated at large k.[3] This has been observed by Taylor in 1938, beginning the chain that led Kolmogorov, in 1941, to formulate the theory that we are explaining (*c.f.r.* [Ba70], p. 112) also called *K41 theory*.

The Reynolds number of the fluid is $R=v_LL\nu^{-1}$: and we can introduce the more general notion of *Reynolds number on scale* l as $R_l\equiv v_ll\nu^{-1}$. Therefore it can be computed by using the basic assumption $v_l^3/l=v_L^3/L$ (*c.f.r.* (6.2.4)), as: $R_l=v_LL\nu^{-1}(l/L)^{4/3}\equiv(l/L)^{4/3}\,R$.

After the above discussion of the implications of the homogeneous turbulence hypothesis *the Kolmogorov scale k_ν can be naturally identified as the only length scale that can be formed with ε and ν*, or as the scale on which the Reynolds number becomes of order $O(1)$. The latter is the scale $l_\nu\equiv k_\nu^{-1}$ such that $R_{l_\nu}=1$ or $k_\nu=L^{-1}R^{3/4}$, while the first scale is

$$k_\nu\equiv l_\nu^{-1}=\big(\varepsilon/\nu^3\big)^{1/4}=L^{-1}R^{3/4} \tag{6.2.9}$$

where the last equality follows taking, still, into account (6.2.4): the two definitions lead to the same result. We see that, actually, k_ν tends to ∞ as

---

[2] The argument according to which, for dimensional reasons, the energy on scales k, contained between k and $2k$, is $v_l^2L^3$ and hence $L^3K(k)k\sim v_l^2L^3$ leads, combined with (6.2.4), directly to (6.2.8), with less assumptions; it allows us to suppose only independence of the $\underline{u}_{\underline{k}}$'s with $\underline{k}$ relative to different scales. However if we only supposed this we would not understand how it could be that $(\sum_{k<|\underline{k}|<2k}\underline{u}_{\underline{k}})^2$, square of the sum of $\sim(kL)^3$ random variables, could have order $(kL)^3$, as instead it follows from the independence of the $\underline{u}_{\underline{k}}$ with different $\underline{k}$.

[3] Because $K(k)$ is summable for $k\to\infty$, but $k^2K(k)$ is not.

$\nu \to 0$ if ε stays fixed or does not tend to 0 too fast (in any event *it is not at all reasonable* to think that ε tends to 0).

Wishing to express everything in terms of the intensity g of the force (*e.g.* $g = (L^{-3} \int \underline{g}(x)^2 dx)^{1/2}$) acting on the fluid and assuming that as the viscosity decreases the energy dissipated per unit time *reaches finite saturation value*, as it is observed in several experiments on systems forced by boundary forces, one should think in the present case that, for dimensional reasons, the asymptotic value (for $\nu \to 0$) of the dissipation ε must be proportional to $L^{1/2}g^{3/2}$, if L is the side of the container. We mention here that the existence of a finite saturation value for the energy dissipation when friction tends to 0 has far reaching consequences: for instance it led Onsager to propose that also the Euler equations did admit stationary states in which energy remains constant in spite of the action of an external force and of the absence of friction in the equations. *c.f.r.* [Ta35],[On49],[Kr75a],[Ey94],[CWT94]).

Note that $v_L \sim (\varepsilon L)^{1/3}$ and $R = \nu^{-1}(\varepsilon L)^{1/3}$ (and if we accept that the asymptotic dissipation is $\propto L^{1/2}g^{3/2}$ then also $R = (gL)^{1/2}L\nu^{-1}$). It follows that the number of Fourier components with $k < k_\nu$ is of order

$$N_\nu \sim \frac{8\pi}{3}\frac{k_\nu^3}{(2\pi L^{-1})^3} = cost\, R^{9/4} \qquad (6.2.10)$$

that gives the order of magnitude of the *apparent number of degrees of freedom*, probably proportional to the *information dimension* of the attractor. See §6.3 for a model simpler than the NS equation, but similar to them, in which the fractal dimension of the attractor turns out to be half of the apparent dimension.

We can ask whether the Kolmogorov law is exactly true already for $k \sim k_0 = 2\pi/L$. In reality we must expect deviations at least on scales comparable to that on the which the force acts (*i.e.* L); because on such scales the details of the structure of the force must be important. Hence we must think that what said on the law 5/3 is true in an interval of scales $k_0' \ll k \ll k_\nu$ with $k_0 \ll k_0'$ and $k_\nu = k_0 R^{3/4}$. This interval of scales is called *inertial domain*, or *inertial field*, and in it we have *homogeneous turbulence*, in which homogeneous universal laws hold for the energy distribution (*c.f.r.* (6.2.8)) and for other quantities (*c.f.r.* problems).

Let us now determine the viscous scale $k_\nu' > k_\nu$ that provides us with a natural *ultraviolet cut–off* for the NS equation: it is interesting here to recall the discussion in §2.2 where the ultraviolet cut–off has been introduced without discussing it physical meaning to formulate an empirical algorithm of solution of the NS equation.

By what said above the value of k_ν' must be determined from the condition that the viscous terms, which become comparable with the inertial ones on scales $k \sim k_\nu$, dominate on the inertial ones for $k > k_\nu'$.

We have seen that the condition is the equality of the averages (in time) $\nu|\underline{k}|^2 \langle|\underline{u}_{\underline{k}}|\rangle \approx \langle| \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} \underset{\sim}{u}_{\underline{k}_1} \cdot \underset{\sim}{k}_2 \Pi_{\underline{k}} \underline{u}_{\underline{k}_2}|\rangle$, and the sum can be roughly estimated as $|\underline{k}|\langle|\underline{u}_{\underline{k}}|^2\rangle(|\underline{k}|L)^{\frac{3}{2}}$ by using the hypothesis (1), which means that it is a sum of $O((|\underline{k}|L)^3)$ independent random variables with zero average (since $\underline{k}_1$ cannot be parallel to $\underline{k}_2$ by the zero divergence property of $\underline{u}$), and $\langle|\underline{u}_{\underline{k}}|^2\rangle^{1/2}$ can be evaluated as $v_l(|\underline{k}|L)^{-\frac{3}{2}}$ by using (6.2.6) (with $l = k^{-1}$); hence we see that this condition leads to $k'_\nu \sim (\varepsilon\nu^{-3})^{\frac{1}{4}} = k_\nu$. Here it is useful to keep in mind that k_ν was defined, heuristically and independently, on purely dimensional grounds, *c.f.r.* equation (6.2.9).

Hence the scales k'_ν and k_ν have the same order of magnitude, a not surprising consistency check: this means that the arguments given here are not sufficient to determine the ratio k'_ν/k_ν and they only say that it will be ≈ 1; a value of $k'_\nu = 2k_\nu$ (for example) could be acceptable for possible numerical experiments or further theoretical deductions (and the correct value of the ratio would be determined by finding at which magnitude of the ratio the results become independent from value of the ratio itself).

We can formulate a more general hypothesis (*c.f.r.* [Ba70], p. 114) that for *all the scales k, also "close"(enough) to k_0 or "beyond" k_ν*, the spectrum of energy depends only on ε and ν: the hypothesis is that $K(k) = v_c^2 l_c K_{univ}(l_c k)$, with v_c, l_c "characteristic" velocity and length scales only dependent on ε, ν, *i.e.* $l_c = (\nu^3\varepsilon^{-1})^{1/4}$ and $v_c = (\nu\varepsilon)^{1/4}$, and with K_{univ} a suitable universal function of its argument.

This implies that if k is in the inertial domain, in which the (average) energy density spectrum depends only on ε and is a given by a power law, the universal function $K_{univ}(x)$ must be $\propto x^{-5/3}$, because only the combination $v_c^2 l_c l_c^{-5/3} = \varepsilon^{2/3}$ depends only on ε.

The domain of validity of the more general formula should be more extended than that in which $K_{univ} \propto k^{-5/3}$ and the question of which are the scales k_1, k'_0, k_2 that delimit

(a) the inertial domain (k'_0, k_ν) and

(b) the larger domain (k_1, k_2), with $k_0 < k_1 < k'_0 < k_\nu < k_2 < k'_\nu$, in which the more general law holds

is a question which is not soluble within the present framework and it requires a more detailed understanding of the problem.

Particular interest has the question whether it is $k_0/k_1 \xrightarrow[\nu\to 0]{} 0$ and/or $k_\nu/k_2 \xrightarrow[\nu\to 0]{} 0$: the idea of the existence of the universal function suggests that $k_0/k_1 \to 0$ and k_1 should have order of magnitude such that $l_c k_1 \ll 1$. It seems (without any surprise) that the experiments are compatible *both* with $k_0/k_1 \sim O(1/\log R)$ *and* with $O(1)$.

Finally if the external force acts on scales $k_{in} \gg k_0$ but $k_{in} \ll k_\nu$ we expect to have the law 5/3 for $k > k_{in}$ with a distortion around a $k = k_{in}$. For $k_0 \ll k \ll k_{in}$ we expect equipartition of energy among the modes.

Finally some of the above results can be put in a remarkably rigorous setting: see the discussion in (G) below

(D): Bidimensional theory.

I propose to adopt the hypotheses that material lines are extended in two–dimensional turbulence, that there is a cascade process of transfer of mean square vorticity to higher wavenumbers, and that the limiting value of the rate of dissipation of mean–square vorticity as $\nu \to 0$ is nonzero. (Batchelor), [Ba69].

If $d = 2$ we could repeat an analogous argument: but this time there is also, if $\nu = 0$, the first integral of the vorticity, *enstrophy*, that could be treated in the same way as the energy integral in the 3–dimensional case. This means that we could suppose that, for $\kappa < \kappa'$, the quantity $\mathcal{S}_{\kappa,\kappa',\infty}$, defined by replacing in the l.h.s. of (6.2.2) $|\underline{u}_{\underline{k}}|^2$ with $\underline{k}^2|\underline{u}_{\underline{k}}|^2$ and working out (starting from the NS equations) the r.h.s. to achieve a similar decomposition into three terms as in (6.2.2), vanishes. Hence the enstrophy would be transfered *in a local way* (in the sense discussed in (C)) from the scales of large length to those of small length where it would be dissipated.

The energy and enstrophy dissipations per unit volume and time are $\varepsilon = \nu L^{-3} \int_\Omega |\underline{\partial} \wedge \underline{u}|^2 d\underline{x}$ and $\sigma = \nu L^{-3} \int_\Omega |\Delta u|^2 d\underline{x}$: their dimensions are of a velocity cube over a length and of a velocity cube over a length cube (*i.e.* $[v^3/l]$ and $[v^3/l^3]$) respectively.

The above hypothesis is natural if we imagine that for $\nu \to 0$ the energy and enstrophy dissipation, per unit time and volume, have finite limits denoted ε and σ respectively, and $\varepsilon = 0$.

Taking $\varepsilon = 0$ requires a justification, at least on heuristic grounds. Indeed if $E = \int d\underline{x}\, |\underline{u}(\underline{x})|^2$ denotes the energy and $\omega^2 = \Omega = \int d\underline{x}\, |\underline{\partial} \wedge \underline{u}(\underline{x})|^2$ the enstrophy it is, by the NS equations:

$$\frac{1}{2}\dot{E} = -\nu\omega^2 + \mathcal{L}, \qquad \frac{1}{2}\dot{\Omega} = -\nu L^3 \sum_{\underline{k}} |\underline{k}|^4 |u_{\underline{k}}|^2 + \tilde{\mathcal{L}}$$

with $\mathcal{L}$ defined in (6.2.3), and $\tilde{\mathcal{L}} = L^3 \sum (i\underline{k} \wedge \overline{\underline{g}}_{\underline{k}}) \cdot \underline{u}_{\underline{k}}$, that implies $|\tilde{\mathcal{L}}| \le g\omega_0 \le g\omega$ if $g = ||\underline{g}||_2$ and ω_0 is the vorticity on scale $\sim k_0$.

We see that if in the stationary state vorticity is concentrated at large values of $\underline{k}$, *i.e.* for $|\underline{k}| \sim k_\nu =$ where k_ν is the Kolmogorov scale $k_\nu = (\sigma\nu^{-3})^{1/6}$ (which is the only L–independent inverse length that one can form with the parameters σ, ν), then $L^3 \sum_{\underline{k}} |\underline{k}|^4 |u_{\underline{k}}|^2 \simeq k_\nu^2 \omega^2$.

Denoting time averages by $\langle\cdot\rangle$ stationarity implies that the time average $2^{-1}\langle\dot{\Omega}\rangle$ vanishes so that $-\nu k_\nu^2 \langle\omega^2\rangle = -\langle\tilde{\mathcal{L}}\rangle < g\langle\omega\rangle$ *i.e.* $\langle\omega\rangle < g\,/(\nu k_\nu^2)$.

Therefore $\varepsilon = \nu\langle\omega^2\rangle < g^2/\nu k_\nu^4 \xrightarrow[\nu\to0]{} 0$ because k_ν^4 tends to ∞ as ν^{-2}: we see that for $\nu \to 0$ the system (formally verifying the Euler equation) *conserves* *the energy in the inertial range, namely* $\varepsilon = 0$.

The enstrophy, in the inertial range, is not conserved[4] and one can assume that it "cascades" through the inertial range, at rate σ.[5]

We can easily repeat the dimensional analysis of the K41 theory with enstrophy replacing energy: the dimensions of the various quantities change and the results are somewhat different. If σ denotes the enstrophy that is communicated from the external force to the system per unit time and volume, we see that on scale l it would be $\sigma = v_l^3 l^{-3}$, hence the energy between k and $k + dk$ would be $L^2 K(k) dk = \sum \frac{v_l^2}{k^2}$ and hence, since $v_l = \sigma^{1/3} k^{-1}$, and proceeding as in (6.2.7)

$$K(k) = cost\,\sigma^{2/3}\,k^{-3}, \quad k_0 \ll k \ll k_\nu \tag{6.2.11}$$

that is a scaling law that turns out to be *summable* for k large and hence it means that, in a regime of developed turbulence, *the energy remains concentrated on the large scales* (while enstrophy is distributes on the whole inertial domain and in fact the quantity of enstrophy is asymptotically concentrated on modes with $k \sim k_\nu$ because $k^2 K(k)$ is not summable, for k large, while $K(k)$ is summable).

If $d = 2$ we can, therefore, say that, in the stationary state, enstrophy is concentrated on modes with large k, *i.e.* on the Kolmogorov scale, although the phenomenon is less pronounced than the correspondent phenomenon at $d = 3$ because, if $d = 2$, the enstrophy integral is only logarithmically divergent at large k (while if $d = 3$ it diverges as $k^{4/3}$). Energy remains concentrated on the modes at scales of order $k_0 = 2\pi L^{-1}$.

Note that if $d = 2$ the Reynolds number on scale l is given by $R_l = (\frac{l}{L})^2 R$, *i.e.* it depends from the Reynolds number R via the power 2 instead of 4/3; the *Kolmogorov scale* and the number N_ν of apparent degrees of freedom are, therefore

$$k_\nu = L^{-1} R^{1/2}, \qquad N_\nu = R \tag{6.2.12}$$

respectively.

We can also consider the case in which the external force acts on a scale of momentum $k_{in} \gg k_0$. In this case the stationary distribution of the energy depends on value of k with respect to a k_{in}. For $k > k_{in}$ the (6.2.11) should hold while for $k_0 \ll k \ll k_{in}$ the energy should be equidistributed

[4] This does not contradict the regularity theorems for the solutions of the Euler equations for $d = 2$ because it is a property of the motions that develop on the attractor for the NS evolution. We lack any real knowledge of this set: which could consist of fields that, although approximable with very regular functions, are rather singular so that one cannot conclude that enstrophy (which is formally conserved if $\nu = 0$) is *really* conserved. This is very similar to what has been said about the energy in the case $d = 3$.

[5] Which, perhaps, tends to become ν–independent if an analogy with the corresponding energy cascade in the $d = 3$ case is correct (which is not clear).

between the modes $k_0 < |\underline{k}| < k_{in}$ and its value should be such that the energy density (here proportional to k) matches with the one in (6.2.11) at $|\underline{k}| = k_{in}$.

(E): Remarks on the K41 theory.

The discussion of Kolmogorov's scaling laws, certainly rather heuristic and questionable under many aspects, may appear not too convincing. Why indeed should we suppose that only if $d = 2$ the enstrophy dominates the cascade at short length scales (k large)? In reality, in absence of viscosity, also at $d = 3$ one has vorticity conservation in the form of Thomson's theorem; and we did not take this into account. It is legitimate to think that, succeeding in taking that into account "correctly", we could obtain different results even in 3 dimensions, (and, more generally, dimension dependent results at all dimensions because an analogue of the theorem of Thomson holds at all dimensions). The difficulty in taking into account these conservation laws does not seem a sufficient reason for not considering them as relevant.

It remains, therefore, to see if at least the experiments agree with the Kolmogorov scaling law. The answer seems strongly positive if $d = 3$, [Ba70]: but since we do not have an idea of how to estimate the corrections, we cannot be really certain of the quality of the result and some doubts stand. More, and very careful, research on the theme is certainly desirable and explains the interest that have raised "alternative" methods, such as the functional method of the §6.1.
Recently, [VW93], experimental evidence accumulated on deviations from the law 5/3, (6.2.8). One of the mechanisms of such deviations, which also allows us to keep the essential part of Kolmogorov ideas, the *multifractality*, will be discussed in the §6.3 in a simplified model.

In the problems below we examine other simple consequences of the Kolmogorov hypothesis that lead to interesting statements on the quantity

$$\langle \prod_{i=1}^{n} (u_{\alpha_i}(\underline{x}_i, T + t_i) - u_{\alpha_i}(\underline{y}_i, T + t'_i)) \rangle \tag{6.2.13}$$

where the average is understood as average over the time T; or, assuming that the motion for large times is described by an attracting set with a (ergodic) statistics, the average can be understood with respect to the statistics of the attracting set (assuming T large).

For example we must have, and this is a possible precise definition [5] of v_l, that $\langle(\underline{u}(\underline{x})-\underline{u}(\underline{y}))^2\rangle$ be $cost\, l^{2/3}$ if $|\underline{x}-\underline{y}|=l$ (provided $L\gg l\gg k_\nu^{-1}$). And this is an interesting property of the statistics of the velocity field in the inertial domain.

It implies, indeed, that the velocity field has fractal nature, in a certain sense analogous to that of the *Brownian motion*: the latter provides us with trajectories that are fractal in the sense that the increments of the position as time increases are proportional to the power 1/2 of the increment of time. Fluids in states of homogeneous turbulence seem to provide examples of velocity fields which, for instance, are fractal in space, in the sense that the velocity increments are proportional to the power 1/3 of the spatial variations (and they also seem to provide examples of fields which are fractal in time, *c.f.r.* problems).

A simple heuristic argument is the following. Suppose that $(x-y)k_c\simeq O(1)$, *i.e.* suppose that we look at the velocity field on a small scale of the order of the Kolmogorov length. Then $\underline{u}(\underline{x})-\underline{u}(\underline{y})=\sum_{|\underline{k}|\le k_c}(e^{i\underline{k}\cdot\underline{x}}-e^{i\underline{k}\cdot\underline{y}})\,\underline{u}_{\underline{k}}$. Therefore $|\underline{u}(\underline{x})-\underline{u}(\underline{y})|^2$ has an average value (under the assumption that $\underline{u}_{\underline{k}}$ are random independent variables with variance $\langle|\underline{u}_{\underline{k}}|^2\rangle=v_\ell^2\,(|\underline{k}|L)^{-3}$) given by

$$\begin{aligned}&\langle|\underline{u}(\underline{x})-\underline{u}(\underline{y})|^2\rangle\simeq\sum_{\underline{k}}|e^{i\underline{k}\cdot\underline{x}}-e^{i\underline{k}\cdot\underline{y}}|^2\langle|\underline{u}_{\underline{k}}|\rangle^2\le\\ &\le\ const\,\frac{L^3}{(2\pi)^3}\int_0^{k_c}k^{2\alpha}|\underline{x}-\underline{y}|^{2\alpha}\frac{v_\ell^2}{(k\,L)^3}\,4\,\pi\,k^2\,dk=\\ &=\ const\,|\underline{x}-\underline{y}|^{2\alpha}\int_0^{k_c}\frac{d\,k}{k}\,k^{2\alpha}\,(\frac{\varepsilon}{k})^{2/3}\end{aligned}\tag{6.2.14}$$

where α can be taken any number in $(0,1)$ having used, if C_α is a suitable constant and if $k_c|\underline{x}-\underline{y}|\le O(1)$, the bound $|e^{i\underline{k}\cdot(\underline{x}-\underline{y})}-1|\le C_\alpha\,(|\underline{k}|\,|\underline{x}-\underline{y}|)^\alpha$.

So if $\alpha<1/3$ the integral converges and we see that $|\underline{u}(\underline{x})-\underline{u}(\underline{y})|\propto|\underline{x}-\underline{y}|^{1/3}$ uniformly in the Reynolds number provided we look at velocity increments over distances of the order of the Kolmogorov length. If $\alpha\ge1/3$ a uniform bound of "Hölder type" on the velocity increment is not possible along the above lines.

Although this does not prove that a better uniform bound is not actually possible it clearly shows the special role plaid by the value 1/3. And if we try to bound the velocity increment proportionally to $|\underline{x}-\underline{y}|^\alpha$ with $\alpha>1/3$ then we expect to see that the constant grows with k_c as $k_c^{\alpha-1/3}$ or, *c.f.r.* (6.2.9), as the $(\alpha-1/3)/4$ power of the Reynolds number, $R^{(\alpha-1/3)/4}$. The behavior with exponent 1/3 of the Hölder continuity extends, beyond the region $k_\nu|\underline{x}-\underline{y}|=O(1)$, to cover increments over scales into the whole inertial range as discussed in problem [6.2.7], see [LL71].

[5] Which should be equivalent to (6.2.5).

The interest of such random fields, for probability theory, has been stressed by Taylor in 1935 (*c.f.r.* [Ba70], p.8): it is even increased by another of their properties, theoretically predicted by Taylor in 1938 and observed experimentally later by Steward, 1951 (*c.f.r.* [Ba70], p. 171): it is the *skewness* of the distribution of $\underline{\delta}(\underline{\rho}) = \underline{u}(\underline{x}) - \underline{u}(\underline{x}+\underline{\rho})$. The observable $\underline{\delta}(\underline{\rho})$ has zero *average* in space and in time for each $\underline{\rho}$, but it has third moment *different from zero*, so that it is not a centered random variable (and obviously it is not Gaussian).

Kolmogorov's theory of homogeneous turbulence provides the second non trivial example of fractal properties that become manifest in situations with a direct physical meaning, after Einstein's theory of Brownian motion.

In the next section we shall try to make clear the deep difference with respect to Brownian motion, that allows us to say that the statistics of developed turbulence has *multifractal* character: here this means that the time average of $\delta_n(\underline{\rho}) = |\underline{u}(\underline{x}) - \underline{u}(\underline{x}+\underline{\rho})|^n$ *does not behave* like $\langle\delta_2(\underline{\rho})\rangle^{n/2}$ as $\underline{\rho} \to \underline{0}$, not even in the inertial domain $k_0^{-1} \ll l \ll k_\nu^{-1}$, but rather like $\langle\delta_2(\underline{\rho})\rangle^{\zeta_n/2}$ with $\zeta_n \ne n/2$ a *nonlinear* function of n.

(F): The dissipative Euler equation.

We shall call *dissipative Euler equation* the equation for an incompressible fluid

$$\partial_t \underline{u} + \underset{\sim}{u} \cdot \underset{\sim}{\partial} \underline{u} = -\underline{\partial} p - \chi \underline{u} + \underline{g}, \qquad \underline{\partial} \cdot \underline{u} = 0 \tag{6.2.15}$$

in a container Ω that is a torus or that has perfect walls so that a *no slip* boundary condition $\underline{u} \cdot \underline{n} = 0$ holds, if $\underline{n}$ is the external normal.

The constant χ will be called *sticky viscosity* and it does not corresponds to a constitutive equation of the type considered in §1.1,§1.2. It rather corresponds to a perfect fluid that flows with friction over a background: the model can be physically interesting mainly in 2–dimensional cases, when one often imagines that the fluid flows on a "rough" surface . But in order that a bidimensional fluid that "flows on a table" be well modeled by (6.2.15) it would be necessary that the thickness of the fluid is less than the Kolmogorov length k_ν^{-1} if ν is its true viscosity: this makes it difficult to find real applications of the equation. Here we consider it only as a mathematical model.

Is it possible to conceive a theory analogous to the theory K41? The following few comments are an attempt at posing the problems rather than at suggesting their solutions.

Let $\eta = \chi L^{-3} \int_\Omega |\underline{u}|^2 d\underline{x}$ be the energy dissipated per unit volume and per unit time. Hence the length scale that can be formed with the quantities η and χ is $l_\chi = (\eta\chi^{-3})^{1/2}$, and $k_\chi = l_\chi^{-1} = (\chi^3\eta^{-1})^{1/2}$.

There is a difference between the present case and the NS case: namely the friction is "weaker" at large $\underline{k}$ and therefore the inertial terms dominate, at least dimensionally, at large $\underline{k}$ rather than at small $\underline{k}$. Note also that if η is constant then $k_\chi \xrightarrow[\chi\to 0]{} 0$ "*unlike*" k_ν (which diverges as $\nu \to 0$) in the NS case.

Therefore if k_0 is the momentum scale on which the force $\underline{g}$ acts and if we assume that the energy "cascades" above the forcing scale, see the homogeneous turbulence hypothesis in (C) above, then for $d = 3$

(1) If $k_\chi < k_0$ we should have an inertial domain above k_0. However this would mean that energy is never dissipated which is impossible because the energy is *a priori* bounded. It seems that waiting an infinite time the energy will simply extend further and further in $\underline{k}$–space and if there is an ultraviolet cut–off K_{uv} it would tend to become equipatitioned between all modes with $k_0 < |\underline{k}| < K_{uv}$ with the exception of the energy contained in the modes close to k_0, "directly" forced and which seem likely to hold a finite fraction of the total energy. This situation is the more likely as we expect that $k_\chi \to 0$ as $\chi \to 0$ since it does not seem possible that $\eta \xrightarrow[\chi\to 0]{} 0$ faster than χ, see problem [6.2.1]: therefore eventually $k_\chi < k_0$.

(2) if $k_\chi \gg k_0$ we should have again equipartition (in presence of an ultraviolet cut–off) among the modes above k_χ while in the range between k_0 and k_χ, if large enough, we should have a scaling inertial range with the 5/3–law for the energy distribution.

The $d = 2$ case cannot be discussed similarly: if we assume that, as in the NS case, there is an enstrophy cascade and if σ is the enstrophy dissipated per unit time and volume then σ has the dimension of inverse time cube; since χ has also dimension of inverse time we *cannot form any length scale* with χ and σ. This might mean that in $d = 2$ there is always equipartition (in presence of a cut–off). Note, however, that the basis for an enstrophy cascade assumption is unclear in this case because the argument of (D) cannot be adapted to the present situation.

In all cases we see that it is necessary to impose *a priori* an ultraviolet cut–off K_{uv} on the ED equations so that all this has a meaning. Otherwise, in absence of a cut–off, the system will necessarily tend towards a stationary state in which every mode, but the ones directly forced and their neighbors, has zero energy (*i.e.* equidistribution among infinitely many modes).

The situation recalls the black body problem and, as in that case, we can think that it might be unnecessary to introduce an ultraviolet cut–off, *c.f.r.* [Ga92], because the time scales needed to transfer appreciable amounts of energy to the modes of large $|\underline{k}|$ may increase very rapidly with the value of the momentum $\underline{k}$ itself. An hypothesis which, in the case of the black body, was advanced by Jeans in his attempts of a classical explanation of the black body spectrum. But as in the case of the Jeans' problem the question is very difficult to analyze. Since long times are involved direct numerical tests are not possible. And a possible (if existent) effective ultraviolet cut–off K must be, on dimensional grounds, a momentum formed with the quantities $k_0, k_{in}, \varepsilon, \chi, g$: with such quantities we can form several momenta, $k_0, k_{in}, \sqrt{\chi^2/\varepsilon}, \chi/\sqrt{gL}$, hence K cannot be determined without a detailed theory.

(G) The Ruelle–Lieb bounds and K41 theory

We conclude this section by quoting certain really remarkable rigorous results due to Ruelle and Lieb which, although not well known as they deserve, are closely related to the K41 theory.

The results that we select concern an estimate of the Kaplan–Yorke dimension of the attractor for the $d = 2, 3$ NS equations or for the $d = 2$ Rayleigh equations. Calling $\varepsilon = \nu(\underset{\sim}{\partial}\underline{u})^2$ the energy dissipation per unit volume and assuming that $\varepsilon \in L_{1+d/2}(\Omega)$ then the Kaplan–Yorke and the Hausdorff dimensions of the attractor associated with any ergodic invariant measure μ for the incompressible NS flow under a smooth constant forcing is bounded by

$$K\,R^{3d/4} \tag{6.2.16}$$

where $R = v_L L \nu^{-1}$ is the Reynolds number defined in terms of the "velocity on the scale L of the forcing" by setting $||\varepsilon||_{L_{1+d/2}} \stackrel{def}{=} v_L^3 L^{-1}$.

The constant K depends on the boundary conditions and can be explicitly estimated to be a constant of order 1. The result, and many technical ideas to obtain it, was proposed for $d = 3$ in [Ru82] (*c.f.r.* equation (2.8) of this reference) where it was proved subject to further assumptions; it was successively fully proved in [Li84] (*c.f.r.* equation (43) of this reference) which also extended it to $d = 2$.[6]

The same idea was then applied in [Ru84] to the analysis of the Rayleigh model for convection, *c.f.r.* §1.5, obtaining a bound on the dimension of the attracting set for the convection model as

$$K\,R\,(1 + R_{Pr}^{-1})\,a^2/H^2 \tag{6.2.17}$$

for large Rayleigh number R ans small Prandtl number R_{Pr}, where a is the spatial period of the flow and H is the height (see §1.5, (1.5.17), and §4.1 for the notations); K is a (dimensionless) constant.

Here we should note the *remarkable agreement* between (6.2.16) and the K41 prediction (6.2.9): in the case in which we can regard ε as constant (as empirically assumed in the K41 theory) (6.2.16) is a precise and rigorous statement identical to (6.2.9): this is one of the achievements of the mathematical theory of fluids.

The shortcoming of the result is that we do not know *a priori* that on the attractor the dissipation per unit volume is bounded in $L_{1+d/2}$ if $d = 3$. We do not know that in $d = 2$ either: in spite opf the fact that we have

[6] In fact what one really needs is a bound on $||\varepsilon||_{L_{1+d/2}}$ which has a finite average with respect to the considered invariant distribution μ.

(*c.f.r.* Chap. III) a rather good existence and regularity theory of the NS equations (with no results, however, imply a bound global in time on the $L_{1+d/2}$ norm of the dissipation ε per unit volume). Furthermore in the $d=2$ case we would like to have a bound in terms of the enstrophy dissipation rate rather than in terms of the energy dissipation rate as it emerges from the analysis in (D) above. Such a bound is possible but in this case we lack *a fortiori* an estimate (global in time) of the enstrophy dissipation on the attracting set.

A rough idea of the arguments in [Ru82], [Li84], [Ru84], is presented in problems [6.2.11] and [6.2.12].

Problems: *Dissipation and attractor dimension as $\nu \to 0$. Kolmogorov's skew correlations.*

We define $\langle X\rangle$ the time average of an observable X as seen on a "typical" motion of the fluid, that develops on an attracting set A on which it is described by a statistics (such that the average with respect to the statistics of X coincides with $\langle X\rangle$). The following problems closely summarize the analysis of [LL71].

[6.2.1]: (*Dissipation rate as $\nu \to 0$*) Consider a NS fluid with viscosity ν in a periodic container of side L and subject to a volume force of density $F\sqrt{\nu}\underline{g}(\underline{x}/L)$ (with $\underline{g}$ admensional and fixed). Then the average dissipation $\langle\eta\rangle$ is bounded uniformly in the Reynolds number $R=\sqrt{F\sqrt{\nu}L^3/\nu^2}$ (*c.f.r.* the discussion at the beginning of the section). Suppose (as always in this section) that the solutions of the NS equation are C^∞. Show that the same argument implies a uniform bound on the dissipation in the case of the ED equation (*i.e.* $\chi\langle E\rangle \le (F||\underline{g}||_2)^2$). (*Idea:* (NS case) If $\underline{u}=\sum_{\underline{k}} e^{i\underline{k}\cdot\underline{x}}$ and $E=L^3\sum_{\underline{k}}|\underline{u}_{\underline{k}}|^2=\int|\underline{u}|^2 d\underline{x}$ and $S=L^3\sum_{\underline{k}}\underline{k}^2|\underline{u}_{\underline{k}}|^2=2^{-1}\int|\underline{\partial}\,\underline{u}|^2 d\underline{x}$ one finds, immediately (or *c.f.r.* (3.2.15))

$$\dot E/2=-\nu S+\sqrt{\nu}F\int \underline{u}\cdot\underline{g}\,d\underline{x}$$

Hence averaging over time the left hand side, which is a time derivative of a bounded quantity (if $\nu>0$, *c.f.r.* §3.2 for instance) and therefore has zero time average, we get $0=-\nu\langle S\rangle+F\sqrt{\nu}\int\langle\underline{u}\rangle\cdot\underline{g}\,d\underline{x}$. It follows that: $0\le-\nu\langle S\rangle+||\underline{g}||_2\sqrt{\langle E\rangle}\,F\sqrt{\nu}$ *i.e.* (noting, as usual, that $E\le k_0^{-2}S$, if $k_0=2\pi/L$) it is $\nu\langle S\rangle\le(F\,||\underline{g}||_2k_0^{-1})^2$.)

[6.2.2]: (*dissipation rate as $\nu\to 0$*) If the bound in [6.2.1] was optimal (*i.e.* equality holds) then it not would be true that η is independent of ν in the case considered [6.2.1]. Check this statement. Suppose that in [6.2.1] the force has the form $F\nu^\alpha\underline{g}$ and that for $\alpha=0$ it is $\eta\xrightarrow[\nu\to 0]{}\infty$: show that there is a value of α ($\le\frac12$) such that η has an nonzero (upper) limit for $\nu\to 0$. (*Idea:* In the present notations the case discussed at the beginning of the section would be written as a fluid subject to a force of the form $F\underline{g}$, *i.e.* $F\nu^\alpha\underline{g}$ with $\alpha=0$ so that the bound of [6.2.1] would be a bound proportional to ν^{-1} which, if optimal, gives a divergent η. The second part is, given the result of [6.2.1], a continuity statement in α.)

[6.2.3]: *(Scaling properties of velocity correlations in the inertial range)* Define

$$V_{\alpha\beta}(\underline{\rho})=\langle u_\alpha(\underline{x})u_\beta(\underline{x}+\underline{r})\rangle,\quad V_{\alpha\beta,\gamma}(\underline{\rho})=\langle u_\alpha(\underline{x})u_\beta(\underline{x})u_\gamma(\underline{x}+\underline{r})\rangle$$

and suppose that the length $|\underline{r}|\equiv\rho\ll L$, so that the scale of length ρ is in the inertial domain or viscous, where these tensors are, by hypothesis, rotation and translation invariant functions (of $\underline{x}$). Show that, by this invariance:

$$
\begin{aligned}
V_{\alpha\beta}(\underline{r}) =& A(\rho)\delta_{\alpha\beta} + B(\rho)\frac{\rho_\alpha\rho_\beta}{\rho^2} \qquad \rho \ll L \\
V_{\alpha\beta,\gamma}(\underline{r}) =& C(\rho)\delta_{\alpha\beta}\frac{\rho_\gamma}{\rho} + D(\rho)(\delta_{\alpha\gamma}\frac{\rho_\beta}{\rho} + \delta_{\beta\gamma}\frac{\rho_\alpha}{\rho}) + E(r)\frac{\rho_\alpha\rho_\beta\rho_\gamma}{\rho^3}
\end{aligned}
$$

where A, B, C, D, E are suitable functions. (*Idea:* The tensors are the only ones that can be formed with the vector $\underline{r}$ alone).

[6.2.4]: Define the tensors

$$
\begin{aligned}
B_{\alpha\beta}(\underline{r}) =& \langle (u_\alpha(\underline{x}) - u_\alpha(\underline{x}+\underline{r}))(u_\beta(\underline{x}) - u_\beta(\underline{x}+\underline{r}))\rangle \\
B_{\alpha\beta\gamma}(\underline{r}) =& \langle (u_\alpha(\underline{x}) - u_\alpha(\underline{x}+\underline{r}))(u_\beta(\underline{x}) - u_\beta(\underline{x}+\underline{r}))(u_\gamma(\underline{x}) - u_\gamma(\underline{x}+\underline{r}))\rangle
\end{aligned}
$$

and show that if $\rho \ll L$

$$
\begin{aligned}
B_{\alpha\beta}(\underline{r}) =& 2(V_{\alpha\beta}(\underline{r}) - V_{\alpha\beta}(\underline{0})) \\
B_{\alpha\beta\gamma}(\underline{r}) =& 2(V_{\alpha\beta\gamma}(\underline{r}) + V_{\gamma\alpha\beta}(\underline{r}) + V_{\beta\gamma\alpha}(\underline{r}))
\end{aligned}
$$

(*Idea:* If $\rho \ll L$ the results of [6.2.3] hold).

[6.2.5]: Show that if $|\underline{r}| = \rho \ll L$ it is

$$
\partial_\gamma V_{\gamma\beta,\alpha}(\underline{r}) + \partial_\gamma V_{\gamma\alpha,\beta}(\underline{r}) - 2\nu\Delta V_{\alpha\beta}(\underline{r}) + \langle g_\alpha(\underline{x})u_\beta(\underline{x}+\underline{r})\rangle + \langle u_\alpha(\underline{x})g_\beta(\underline{x}+\underline{r})\rangle = 0 \quad (!)
$$

as a consequence of the law of evolution of the field $\underline{u}$ according to the NS equation. (*Idea:* Note that

$$
\begin{aligned}
\partial_t V_{\alpha\beta} =& -\langle (u_\gamma(\underline{x})\partial_\gamma u_\alpha(\underline{x}))u_\beta(\underline{x}+\underline{r})\rangle - \langle u_\alpha(\underline{x})(u_\gamma(\underline{x}+\underline{r})\partial_\gamma u_\beta(\underline{x}+\underline{r}))\rangle + \\
&+ \nu\langle \Delta u_\alpha(\underline{x})u_\beta(\underline{x}+\underline{r})\rangle + \nu\langle u_\alpha(\underline{x})\Delta u_\beta(\underline{x}+\underline{r})\rangle + \\
&+ \langle (-\partial_\alpha p(\underline{x}) + g_\alpha(\underline{x}))u_\beta(\underline{x}+\underline{r})\rangle + \langle u_\alpha(\underline{x})(-\partial_\beta p(\underline{x}+\underline{r}) + g_\beta(\underline{x}+\underline{r}))\rangle
\end{aligned}
$$

and the hypothesis of independence of the velocity fields in the various points of the fluid allows us to think the averages over t also as averages over $\underline{x}$ and hence, after some integrations by parts with respect to $\underline{x}$, we obtain the result because we see that the pairs of terms on the second and third rows are equal; and furthermore $\langle p(\underline{x})u_\alpha(\underline{x}+\underline{r})\rangle$ must be a vector with zero divergence formed with the only vector $\underline{r}$ and it must hence have the form $t(\rho)\rho_\alpha/\rho$ with $\frac{d}{d\rho}\rho^2 t(\rho) = 0 \;\longleftrightarrow\; t(\rho) = c\rho^{-2} = 0$, (because $c = 0$ since t must be regular for $\underline{r} \to \underline{0}$). The first member has zero average by the hypothesis of stationarity and $\langle u_\alpha(\underline{x}+\underline{r})u_\beta(\underline{x}+\underline{r})u_\gamma(\underline{x})\rangle = -V_{\alpha\beta,\gamma}$).

[6.2.6]: Compute the trace of the tensor (identically zero) in [6.2.5], equation (!), and evince that

$$
\partial_\gamma V_{\gamma\alpha\alpha}(\underline{r}) - \nu\Delta V_{\alpha\alpha}(\underline{r}) + \varepsilon = 0
$$

(*Idea:* Note that $\langle \underline{g}(\underline{x}) \cdot \underline{u}(\underline{x})\rangle$ is the power dissipated per unit volume, being the averages on $\underline{x}$ equal to those on t; hence $\langle \underline{g}(\underline{x}) \cdot \underline{u}(\underline{x}+\underline{r})\rangle \xrightarrow[\underline{r}\to\underline{0}]{} \varepsilon$).

[6.2.7]: If $\underline{r}$ is in the inertial domain the vector $V_{\gamma\alpha\alpha}(\underline{r})$ must depend only on ε and $\underline{r}$, hence

$$
V_{\gamma\alpha\alpha}(\underline{r}) = \Gamma\varepsilon r_\gamma \qquad \text{for} \quad k_\nu^{-1} \ll \rho <\ll L
$$

where Γ is a suitable (universal) constant. Hence $\partial_\gamma V_{\gamma\alpha\alpha} = 3\Gamma\varepsilon$. We deduce, from [6.2.6], that $\Gamma = -1/3$ because in the inertial domain the term $-\nu\Delta V_{\alpha\alpha}(\underline{r})$ is negligible with respect to ε. (*Idea:* To check this note that $\Delta V_{\alpha\alpha}(\underline{r}) \equiv \Delta(V_{\alpha\alpha}(\underline{r}) - \Delta V_{\alpha\alpha}(\underline{0})) = \frac{1}{2}\Delta B_{\alpha\alpha}(\underline{r})$

(by [6.2.4]) and $B_{\alpha\alpha}(\underline{r}) = cost\,\varepsilon^{2/3}\rho^{2/3}$ in this domain; hence $\Delta V_{\alpha\alpha} \propto \varepsilon^{2/3}\rho^{-4/3}$. Imposing that $\nu\varepsilon^{2/3}\rho^{-4/3} \ll \varepsilon$ one finds $\rho \gg l_\nu$, *c.f.r.* Eq. (6.2.9), *i.e.* the condition that ρ is in the inertial domain).

[6.2.8]: *(Skewness in the inertial range)* Check that [6.2.7], *i.e.* $\Gamma = -1/3$, implies immediately that the quantity $B_{rrr} \equiv B_{\alpha\beta\gamma}(\underline{r})\, r_\alpha\, r_\beta\, r_\gamma\, \rho^{-3}$ is $-4\varepsilon\rho/5$ in the inertial domain. This is the result on the skewness of the distribution of the velocity field in the inertial domain (due to Kolmogorov). (*Idea:* See following problem)

[6.2.9]: Consider the tensors of [6.2.3] and verify, by direct computation, that the incompressibility conditions $\partial_\beta V_{\alpha\beta}(\underline{r}) = 0$ and $\partial_\gamma V_{\alpha\beta,\gamma}(\underline{r}) = 0$ imply the following relations between the coefficients $A, B, \ldots$:

$$\begin{aligned}
&A' + B' + \frac{2B}{\rho} = 0, \quad C' + \frac{2(C+D)}{\rho} = 0, \quad 2D' - 2\frac{D}{\rho} + E' = 0\\
&3C + 2D + E = 0\\
&B(0) = C(0) = D(0) = E(0) = 0, \quad A(0) = \frac{1}{3}\langle \underline{u}^2\rangle\\
&A(\rho) - A(0) = -\frac{1}{2\rho}\frac{d}{d\rho}\rho^2 B(\rho), \qquad D(\rho) = -\frac{1}{2\rho}\frac{d}{d\rho}\rho^2 C(\rho)
\end{aligned}$$

(*Idea:* Perform explicitly the derivatives, then impose the vanishing of the coefficient w of the vector $\partial_\beta V_{\alpha\beta} = w(\rho) r_\alpha$ and of the coefficients of the tensors $\delta_{\alpha\beta}$ and $r_\alpha r_\beta$ in terms of which one expresses $\partial_\gamma V_{\alpha\beta\gamma}$. Rather than $3C + 2D + E = 0$ one will find $\frac{d}{d\rho}\rho^2(3C + 2D + E) = 0$ that, however, implies the preceding relation because the field $\underline{u}$, and hence the its averages, are regular functions for $\rho \to 0$, *i.e.* for ρ smaller than the Kolmogorov scale. For the same reason B, C, D, E must vanish for $\rho = 0$. The relations imply $3C = -2D$, $E = 0$, and $B_{rrr} = -12C$, $V_{\gamma\alpha,\alpha} = -5\frac{C}{\rho}$: $\Rightarrow$, by [6.2.7], $B_{rrr} = \frac{4}{5}\varepsilon\rho$.)

[6.2.10]: *(Regularity in the dissipative range)* Imagine now that ρ is in the dissipative domain ($\rho \ll l_\nu$ and all the fields are regular, differentiable functions) and, hence, $B_{\alpha\alpha}(\underline{r}) = a\rho^2$; show that the constant a that, for dimensional reasons, must have the form $a = c\,\varepsilon\nu^{-1}$ is such that $c = 1/15$. (*Idea:* From [6.2.2] deduce $V_{\alpha\beta}(\underline{r}) = V_{\alpha\beta}(\underline{0}) - \frac{1}{2}B_{\alpha\beta}(\underline{r})$; and from the incompressibility condition of [6.2.9], $A' + B' + \frac{2B}{\rho} = 0$ deduce that $\frac{1}{2}B_{\alpha\alpha}(\underline{r}) = -a\rho^2\delta_{\alpha\beta} + \frac{a\rho^2}{2}\frac{\rho_\alpha}{\rho}\frac{\rho_\beta}{\rho}$. Hence differentiating $V_{\alpha\beta}$ one finds:

$$\langle \partial_\beta u_\alpha \partial_\beta u_\alpha\rangle = 15a, \qquad \langle \partial_\beta u_\alpha \partial_\alpha u_\beta\rangle = 0$$

but $\varepsilon = \frac{1}{2}\nu\,\langle(\partial_\alpha u_\beta + \partial_\beta u_\alpha)^2\rangle = \nu\,\langle(\partial_\alpha u_\beta)^2 + \partial_\alpha u_\beta \partial_\beta u_\alpha\rangle = 15a\nu$).

[6.2.11] *(Bounds on the dimension of an attractor) (Ruelle)* Let $\dot{\underline{u}} = \underline{f}(\underline{u})$ be a differential equation with an *a priori* bounded attracting set (*i.e.* such that $||\underline{u}||_{L_2} \le B$ if $\underline{u}$ is on the attracting set). Let $\underline{u}_0$ be a solution and let

$$\dot{\underline{\delta}} = \frac{\partial}{\partial \underset{\sim}{\underline{u}}}\underline{f}(\underline{u}_0)\underset{\sim}{\underline{\delta}} \stackrel{def}{=} H_{\underline{u}_0}\underline{\delta}$$

Suppose that the eigenvalues $\lambda_j^{(\underline{u}_0)}$ of the *Hermitian part* of $H_{\underline{u}_0}$ are ordered in a non decreasing way and that they are such that for some N, *independent of* $\underline{u}_0$, it is $\sum_{j=1}^N \lambda_j^{(\underline{u}_0)} \le 0$. Then, no matter which invariant distribution μ we consider on the attracting set it follows that the attracting set cannot have Lyapunov dimension $> N$ (*c.f.r.* comment (6) to definition 2 in §5.5). (*Idea:* Let $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_N$ be a basis for the first N eigenvalues (assumed distinct for simplicity) then the parallelogram $\underline{e}_1 \wedge \underline{e}_2 \wedge \ldots \wedge \underline{e}_N$

must contract in volume because

$$\frac{d}{dt}\frac{1}{2}\log||\underline{e}_1\wedge\underline{e}_2\wedge\ldots\wedge\underline{e}_N||^2=$$
$$=\sum_{i=1}^{N}\frac{(\underline{e}_1\wedge\underline{e}_2\wedge\ldots\wedge\underline{e}_N,\,\underline{e}_1\wedge\underline{e}_2\wedge\ldots\wedge H_{\underline{v}_0}\underline{e}_k\wedge\ldots\wedge\underline{e}_N)}{||\underline{e}_1\wedge\underline{e}_2\wedge\ldots\wedge\underline{e}_N||^2}\le\sum_{j=1}^{N}\lambda_j^{(\underline{v}_0)}$$

see [Ru82].)

[6.2.12] (*estimate of the dimension of the NS attractor (Ruelle–Lieb)*) Consider the 2 or 3 dimensional NS equation in a domain Ω with periodic boundary conditions and side L. Let $\underline{u}_0$ be in an attractive set and consider the linearization operator acting on divergenceless fields $\underline{\delta}\in L_2$ with $\underline{\partial}\cdot\underline{\delta}=0$ defined by

$$H_{\underline{u}_0}\underline{\delta}\stackrel{def}{=}\nu\,\Delta\,\underline{\delta}-(\underset{\sim}{u}_0\cdot\underset{\sim}{\partial}\underline{\delta}+\underset{\sim}{\delta}\cdot\underset{\sim}{\partial}\underline{u}_0)-\underline{\partial}p'$$

where p' is such that the r.h.s. has zero divergence. We order the eigenvalues $\lambda_j^{(\underline{u}_0)}$ of the hermitian part of $H_{\underline{u}_0}$ (as an operator densely defined on $C^\infty(\Omega)$ and extended to a self–adjoint operator via the quadratic form that it defines in $L_2\cap C^\infty$) by increasing order. Then if for some N, *independent of* $\underline{u}_0$, it is $\sum_{j=1}^N\lambda_j^{(\underline{u}_0)}\le 0$ the Lyapunov dimension of the attracting set does not exceed N. (*Idea:* Remark that the attracting sets only contain fields $\underline{u}_0$ verifying the *a priori* bound $||\underline{u}_0||_{L_2}\le E$ with $E=const\,||\underline{g}||_{L_2}L^2/\nu$, *if* $\int_\Omega|\underset{\sim}{\partial}\underline{u}|^{2+d}\,d\underline{x}<\infty$ *c.f.r.* (3.2.17). The estimate on the sum $\sum_{j=1}^N\lambda_j^{(\underline{u}_0)}$ is done by bounding this quantity by the corresponding quantity for the Schrödinger operator $\nu\Delta+w(\underline{x})$ where $w(\underline{x})^2=\nu\,(\underset{\sim}{\partial}\underline{u}+\underset{\sim}{\partial}\underline{u})^2$ as an ordinary Shrödinger operator (acting on vector fields not necessarily divergenceless). An estimate can be done and it yields for $d=3$ (6.2.16) [Ru82], [Li84] extendible to $d=2$ as in (6.2.16), [Li84]: the details are nontrivial and we refer to the original works. The method can be further developed to be applicable to the Rayleigh equations in $d=2$, *c.f.r.* (6.2.17), see [Ru84].)

Bibliography: The theory K41 is taken from [LL71] and [Ba70]. The bidimensional theory is inspired by [Ba69] where it is exposed in detail, see also [Kr67],[Kr75b]. The reference [DC92] inspires the first two problems, the others are taken from [LL71]. A recent review of the scaling laws in developed turbulence, which includes some of the main new results of the late 1990's is in [Ga99d]. The results in (G) come from the works [Ru82], [Li84], [Ru84].

§6.3 The shell model. Multifractal statistics.

Having seen the heuristic theory of homogeneous turbulence we can try to avoid the study of the functional integrals of section §6.1 by making an hypothesis on the structure of the attracting set A and then choose the invariant distribution according to Ruelle's principle of §5.7.

We have seen that the Kolmogorov theory of homogeneous turbulence supposes that there is no interchange of energy between the *shell* of the modes $\kappa<|\underline{k}|<\kappa'$ and the shell of the modes $\kappa_1<|\underline{k}|<\kappa_2$, if the shells $[\kappa,\kappa']$

and $[\kappa_1, \kappa_2]$ are separated by at least "an order of magnitude" (for instance $\kappa_1 > 2\kappa'$).

Imagine dividing the modes $\underline{k}$ into shells of different orders of magnitude, *i.e.* so that the "*n-th shell*", or "*shell of scale n*", is defined by

$$\Delta_n : \qquad 2^{n-1}k_0 < |\underline{k}| < 2^n k_0 \tag{6.3.1}$$

where k_0 is the "scale of the container":*i.e.* $k_0 = 2\pi L^{-1}$ is the minimum of the values possible for $|\underline{k}|$ and depends only on the length $L \equiv l_0$ of the side of the container, imagined as a cube Ω with periodic boundary conditions.

The absence of energy interchange between distant shells, or distant scales, will be then *imposed* by writing the evolution equation as the NS equation modified as follows, *c.f.r.* [Kr64], if $\underline{k} \in \Delta_n$:

$$\dot{\underline{u}}_{\underline{k}} = -\nu \underline{k}^2 \underline{u}_{\underline{k}} - i \sum_{\substack{\underline{k}_1+\underline{k}_2=\underline{k} \\ \underline{k}_2\in\Delta_{n\pm1}}} \underset{\sim}{u}_{\underline{k}_1} \cdot \underset{\sim}{k}_2 \, \Pi_{\underline{k}} \underline{u}_{\underline{k}_2} + \underline{g}_{\underline{k}} \tag{6.3.2}$$

In this way the absence of direct energy interchange between shells separated (by at least 2 units of scale) is guaranteed *a priori* , and hence by a stronger reason the laws of Kolmogorov should hold.

The model (6.3.2) can be *further simplified* by turning it into a model in which one thinks of replacing the set of approximately 2^{3n} modes of the n-th shell with a *single mode* $\underline{u}_n = (u_{n1}, u_{n2})$ and write, having set $k_n = 2^n k_0$, for $\alpha, \beta, \gamma = 1, 2$

$$\begin{aligned} \dot{u}_{n\alpha} =& -\nu k_n^2 u_{n,\alpha} + k_n \sum_{\beta\gamma} \Big(C^\alpha_{\beta\gamma}(n) u_{n-1,\beta} u_{n+1,\gamma} + \\ & + D^\alpha_{\beta\gamma}(n) u_{n+2,\beta} u_{n+1,\gamma} + E^\alpha_{\beta\gamma}(n) u_{n-2,\beta} u_{n-1,\gamma}\Big) + \underline{g}_n \end{aligned} \tag{6.3.3}$$

where $C^\alpha_{\beta\gamma}(n), D^\alpha_{\beta\gamma}(n), E^\alpha_{\beta\gamma}(n),\ \alpha,\beta,\gamma = 1,2$ are constants, symmetric in $\beta \longleftrightarrow \gamma$, defined so that the sum:

$$\begin{aligned} &\sum_{\alpha\beta\gamma;n} k_n \Big(C^\alpha_{\beta\gamma}(n) u_{n-1,\beta} u_{n+1,\gamma} + D^\alpha_{\beta\gamma}(n) u_{n+2,\beta} u_{n+1,\gamma} + \\ &+ E^\alpha_{\beta\gamma}(n) u_{n-2,\beta} u_{n-1,\gamma}\Big) u_{n\alpha} \equiv 0 \end{aligned} \tag{6.3.4}$$

that guarantees conservation of the energy $\mathcal{E} = \frac{1}{2}\sum_n |\underline{u}_n|^2$. The function $\underline{g}_n$ is the "external force" and is imagined different from 0 only for the first values of n (for instance only for $n = 4$).

The condition (6.3.4) can be imposed by requiring that

$$\frac{1}{2} C^\beta_{\alpha\gamma}(n+1) + \frac{1}{4} D^\gamma_{\alpha\beta}(n+2) + E^\alpha_{\beta\gamma}(n) = 0 \tag{6.3.5}$$

where, for example, $E(n) \equiv E$, $C(n) \equiv C$ and $D(n) \equiv D$ are tensors suitably fixed. A simple choice is provided by the model of Gledzer, Ohkitani and Yamada, that is called GOY*model*. We set $u_n = u_{n,1} + iu_{n,2}$ and consider the equation:

$$\dot{u}_n = -\nu k_n^2 u_n + ik_n\big(-\frac{1}{4}\,\overline{u}_{n-1}\,\overline{u}_{n+1} + \overline{u}_{n+1}\,\overline{u}_{n+2} - \frac{1}{8}\,\overline{u}_{n-1}\,\overline{u}_{n-2}\big) + g\delta_{n,4} \tag{6.3.6}$$

where we suppose that the components u_n with $n = -1, 0$ vanish by definition, *c.f.r.* [BJPV98] p. 55%56.[1]

The equations (6.3.4) and (6.3.6) have been intensely studied from a numerical point of view: in performing the truncations (a necessary step in any numerical computation) the equations truncated continue to conserve the energy (if $\nu = 0$): provided, if N is the value of the truncation over n, one requires that u_{N+1}, u_{N+2} are interpreted as $\equiv 0$. It becomes then possible to obtain an *a priori* bound on the energy $\mathcal{E} \equiv \frac{1}{2}\sum |u_n|^2$: for instance for (6.3.6) it is

$$\frac{1}{2}\frac{d}{dt}|u_n|^2 = -\nu\sum_n k_n^2|u_n|^2 + \sum_n \overline{u}_n g_n \tag{6.3.8}$$

valid also for the truncated equations (in which case the sums extend until N).

A second constant of motion is the quantity

$$\mathcal{H} = \sum_n (-1)^n k_n |u_n|^2 \tag{6.3.9}$$

that plays the role of the *elicity*, *c.f.r.* §2.2, of the equations of NS in $d = 3$ although it is unsatisfactory that the terms with n even and those with n odd play a different role.[2]

In this model u_n can be thought of as defining a function with real values

$$u(x) = \sum_{n=1}^{\infty}\Big(u_n e^{i2^n x k_0} + \overline{u}_n e^{-i2^n x k_0}\Big) \tag{6.3.10}$$

---

[1] Another, between the infinite choices that satisfy the (6.3.5), is

$$\dot{u}_n = -\nu k_n^2 u_n + ik_n(4\,\overline{u}_{n-1}\,\overline{u}_{n+1} - 4\,\overline{u}_{n+1}\,\overline{u}_{n+2} - \overline{u}_{n-2}\,\overline{u}_{n-1}) + g_n \tag{6.3.7}$$

Furthermore we could ask why the external force acts on the component $n = 4$. The reason rests probably on the property of the NS equations in 2 dimensions, in which one has (linear) stability of the motion subject to a force acting on only one component, $g_n = \delta_{nn_0} g$, for *all* values of ν, if the component n_0 is that with k_n minimum. Furthermore the "laminar" motion , $u_n = (\,k_{n_0}^2)^{-1} g\delta_{nn_0}$ is, *c.f.r.* problem [4.1.13] and [Ma86], globally attracting for all values of g_{n_0} (or ν). Although this *is not* so for (6.3.6) if $n_0 = 1$, *c.f.r.* problem [6.3.1], nevertheless in numerical experiments one often chooses $n_0 > 3$.

[2] We could remedy this by further complicating the model and introducing 2 complex components for every k_n that play the role of the components of different elicity seen in §2.2, *c.f.r.* [BJPV98]. However it is better to recognize that the GOY model is useful because it can illustrate several behaviors that could appear also in solutions of the NS model but that cannot be taken too seriously as a model with direct physical interest.

and the Kolmogorov analysis can be repeated, obtaining that the energy "on scale $k_n = l_n^{-1} = 2^n k_0$" is $\mathcal{E}_n = l_0^3 v_n^2$ and such that $v_n^3/l_n = \varepsilon$ with ε a parameter that depends by g and is independent on n and ν for $n \gg 1$.

Hence $v_n = (\varepsilon k_n^{-1})^{1/3}$ and $R_n = (k_n/k_0)^{4/3}\, R$ if $R = v_0 l_0/\nu$, and therefore the energy on scale k_n, the scale $k_{n_\nu} \equiv l_{n_\nu}^{-1}$ of Kolmogorov[3] and the apparent number of degrees of freedom N_{n_ν} will be, respectively,

$$\mathcal{E}_n = (\varepsilon k_n^{-1})^{2/3}, \quad k_{n_\nu} = (\varepsilon^{1/3}\nu^{-1})^{3/4}, \quad N_\nu = \frac{3}{4}\log R \tag{6.3.11}$$

The average value of the variation of the velocity will be still

$$\langle |u(x) - u(y)|^2 \rangle \propto |x-y|^{2/3} \qquad k_\nu^{-1} \ll |x-y| \ll 1 \tag{6.3.12}$$

For dimensional reasons we could believe that $\varepsilon = (\sqrt{gl})^3 l^{-1}$: in fact it appears likely that also in this model the dissipation ε has a behavior for $\nu \to 0$ that saturates tending to a limit value (*c.f.r.* §6.2, comment to the (6.2.9)).

The advantage of this model is its simplicity, that allows us to perform a more detailed numerical analysis which can serve as benchmark for checking various ideas on which Kolmogorov theory rests. The disadvantage is that it does not allow us to check the relevance of the Thomson law for the theory of Kolmogorov, because such law is not simulated in the GOY model, not even in a simplified form (unlike the energy conservation). Indeed the existence of the integral of motion (6.3.9) is only poorly analogous to a *partial* consequence of the theorem of Thomson.[4]

Furthermore the GOY model *presupposes* the phenomenon of the energy "cascade", *i.e.* that energy is not transmitted directly from one scale to another scale of different order of magnitude: the latter, however, is one of the key points to understand because it repflects an important feature of the structure of the NS equations.

Coming back to the analysis of the GOY shell model we see that, if we can take as valid a representation of Kolmogorov type for the stochastic process describing the statistics of the attracting set, then at large Reynolds number we obtain an interesting example of a statistical distribution that produces, with probability 1, samples $u(x)$ that are Hölder continuous functions with exponent 1/3. Such a field appears to be very interesting also from the

[3] Defined by $v_{n_\nu} k_{n_\nu} \nu^{-1} = 1$.

[4] It would be vane, then, to think of simulating conservation of vorticity by requiring that between the C, D, E a second relation holds implying an analogous conservation, if $\nu = 0$, of $\sum_n k_n^2 |\underline{u}_n|^2$. This model would be indeed analogous to NS in 2 dimensions, in which the vorticity is a scalar quantity. Instead the characteristic of the dimensions > 2 is precisely that the vorticity *not* is a scalar and its conservation, if $\nu = 0$, is only expressed by the theorem of Thomson. We appreciate therefore the *intrinsic* perfidy of the principle of conservation of difficulties. In fact precisely in order not to fall back on a bidimensional problem we shall have to impose that C, D, E *are not* such that they would guarantee that $\sum_n k_n^2 |\underline{u}_n|^2$ is conserved if $\nu = 0$.

point of view mathematical probability because we must expect that it is a nontrivial distribution (for instance *skew*, *c.f.r.* problems of the§6.2).

An interesting question that can be examined is whether the Kolmogorov law could be "violated", for instance if the distribution of the field u could deviate from the naive expectation according to which the correlation defined by the average (over the time) of $\delta_p(r) = |u(x) - u(x+r)|^p$ is proportional to $r^{p/3}$, as one thinks in the theory of Kolmogorov (*c.f.r.* the theory of the skewness in the problems of section §6.2, for an example).

In general, denoting $\langle\cdot\rangle$ the time average, we can define ζ_p via

$$\langle|\delta_p(r)|\rangle \propto r^{\zeta_p} \tag{6.3.13}$$

to leading order as $r \to 0$ and postulate, following [BPPV93], p.164, a distribution P_r for the random variable $a = a(u)$ defined (as a function of the sample u of the field) by setting $|u(x) - u(x+r)| = \overline{u}\,(k_0 r)^a$ where $\overline{u}$ is a typical velocity variation. The assumption is that the distribution P_r has, for r small, taking $k_0 = 1$, the form:

$$P_r(a)da = r^{g_r(a)}\rho(a)da \tag{6.3.14}$$

with $g_r(a)$ (and ρ) "slightly" dependent on r; in this case it is

$$\zeta_p = \min_a(pa + g(a)) \tag{6.3.15}$$

if $g(a) = \lim_{r\to 0} g_r(a)$.

It is easy to verify, *c.f.r.* problem [6.3.3], that in the case of the Brownian motion the analogous distribution posed in the form (6.3.14) leads immediately to $\zeta_p \equiv \frac{p}{2}$, *i.e.* the minimum is always (for any p) at $a = \frac{1}{2}$ (which is possible only if $g(a)$ vanishes for $a \neq 1/2$ and is $-\infty$ for $a = 1/2$, *e.g.* $g(a) \sim -\log\delta(a - 1/2)$)).

Because of this one says that the samples of Brownian motion are *simple fractals*. If, instead, ζ_p does not turn out to be linear in p we shall say that the distribution of u has *multifractal* samples, *c.f.r.* [FP84].

And the question that arises is whether the "Kolmogorov fields", *i.e.* the samples of the velocity fields distributed with the stationary distribution of a fluid at large Reynolds number R are, in the limit $R \to \infty$, simple fractals with $\zeta_p = \frac{1}{3}p$ (as the result on the skewness of §6.2 suggests) or whether they are multifractal, as suggested in [FP84].

The study of this problem is very difficult but its analogue in the case of the GOY model has been analyzed from a numerical point of view and the results *suggest* precisely that the distribution *is multifractal*. If this was true also for the statistics of the equation of NS then we should think that the laws of Kolmogorov are a first approximation, *even in the hypothesis of locality of the energy cascade*, *c.f.r.* §6.2, and that in principle they require corrections that bring up the true multifractal structure.

Recent experiments on turbulence that develops as air passes through a grid in a wind tunnel seem to indicate that actually we can observe, also in real fluids, multifractal corrections, *c.f.r.* [VW93].

The first remark towards a more detailed study in the case of the GOY model is that (6.3.6) says that, for any $\varepsilon > 0$ and any real α

$$u_n = 2\varepsilon^{1/3} k_n^{-1/3} e^{i\alpha} \tag{6.3.16}$$

is a "solution" of the (6.3.3) for $n \ge 5$ and $\nu = 0$. This shows that the special role of the exponent $1/3$ *is intrinsic*, right from the beginning in equation (6.3.3); and tells us that the law $5/3$ represents an "*exact*" solution (and a "*trivial*" one) of the GOY equation.[5] This is so at least in the inertial domain: because it is neither a solution for the first values of n, ($n \le 4$, where one cannot neglect the forcing g), nor for n too large (where one cannot even neglect the friction terms).

The study can proceed in two directions: by making use of the functional method of the §6.1, or by trying to apply the *principle of Ruelle* of §5.7. The second viewpoint is useful if one succeeds in identifying the attracting set, or at least some invariant probability distributions.

In a sense to make precise, the solution that we can call "Kolmogorov solution" appears as a *fixed point* in phase space for the evolution of the GOY model at zero viscosity. And to make since this *rigorously* true we shall imagine to modify the function g_n in such a way that the (6.3.6) is *exactly* solved by (6.3.16) for $k_0 \ge k_n \ge k_{n_\nu}$, *i.e.* also for the extreme values of n close to 1 or n_ν (therefore g_n must now be thought of as defined in terms of the parameter ε in (6.3.16) via the (6.3.11)).

We can expect that this fixed point is *unstable*. It is then natural to think that its *unstable manifold* generates an attracting set that really attracts motions.

In this scenario the possible multifractality of the statistics of the motion would be due precisely to the extension in phase space of the attracting set, *i.e.* to the fact that it is not the single point (6.3.16).

Let us proceed to rewrite the equation (6.3.16) in dimensionless form: we choose as time unit the time scale associated with the Kolmogorov length l_{n_ν}, (6.3.11) and set:

$$t' = t k_{n_\nu}^{2/3} \varepsilon^{1/3} \tag{6.3.17}$$

One verifies immediately, via (6.3.11), that $k_{n_\nu}^{2/3}\varepsilon^{1/3} \equiv \nu k_{n_\nu}^2$: *i.e.* one dimensionless time unit t' corresponds to a physical unit of time associated with the scale of Kolmogorov, *i.e.* it is the *fastest* (nontrivial) time of the motion.

Changing notation so that the components $n = 0, 1, \ldots$ are counted starting from the Kolmogorov scale downwards, *i.e.* setting $u_n(t) \equiv k_n^{-1/3}\varepsilon^{1/3}\varphi_{n_\nu - n}(t')$, one finds the dimensionless equation

$$\dot\varphi_k + 2^{-2k}\varphi_k = i 2^{-2k/3}\big(-\frac{1}{2}\,\overline\varphi_{k-1}\,\overline\varphi_{k+1} + \overline\varphi_{k-2}\,\overline\varphi_{k-1} - \frac12\,\overline\varphi_{k+1}\,\overline\varphi_{k+2} + \gamma_k(\alpha)\big) \tag{6.3.18}$$

[5] We can say that the (6.3.11) is the law 5/3 for the GOY model because it predicts that the energy contained in the shell $(k_n, 2k_n)$ is proportional to $\varepsilon^{2/3}k_n^{-2/3}$ as it would follow from (6.2.8) integrated between k_n and $2k_n$.

with $k = 0, 1, \ldots, n_\nu - 1$ and $\gamma_n(\alpha)$ *defined so that* $\varphi_k \equiv e^{i\alpha}$ (*i.e.* the (6.3.16) in the new variables) makes $\dot\varphi_k$ to vanish *exactly* in the (6.3.18).

Note that, as already observed, this last condition *implies* that only the components with n of order 1 and n_ν of $\gamma_n(\alpha)$ are "appreciably" not zero: indeed the $\varphi_h = e^{i\alpha}$ already makes exactly zero the sum of the first three terms of the right hand side of (6.3.18) and hence $\gamma_k(\alpha)$ must "only" compensate the term $2^{-2k}\varphi_k$ that tends rapidly a 0 as k increases (which is, hence, present only for k small) and the original force (present only for k near n_ν). In this way the $\varphi_k \equiv e^{i\alpha}$ corresponds to an exact "Kolmogorov solution" (6.3.16).

The necessity of the term $\gamma_k(\alpha)$ and its dependence on α arises only to impose that (6.3.16) makes $\dot\varphi_k$ vanish when inserted in (6.3.18) *also for the extreme values of* k: hence we can say that the *GOY* equation has a "symmetry" that manifests itself through the fact that the existence of a one parameter family of fixed points is only forbidden by the "boundary conditions".

This means that one must imagine that γ_k is fixed, and hence α is fixed, because we must fix once and for all the equation that we want to study. Since γ_k differs from zero only for few extreme values of k one can wonder whether in the limit in which $\nu \to 0$ (and hence $n_\nu \to \infty$) the symmetry is *restored* in the minimal sense that there exist exact solutions of (6.3.18) having essentially the form $e^{i\alpha'}$ for every α' *at least* for $1 \ll n \ll n_\nu$.

From the results of [BPPV93] evidence emerges, both experimental and theoretical, that in reality the rotation *symmetry* of the parameter α is spontaneously broken and *is not* restored, not even in the limit $\nu \to 0$: precisely the attracting set of (6.3.6) is determined by the (6.3.18) with $\alpha = \pi/2$. And the statistics of (6.3.18) with $\alpha \neq \pi/2$ is *also* identical to the statistics of (6.3.6) with $\alpha = \pi/2$.

If one assumes this (experimentally suggested) property one can then remark that (6.3.18) with $\alpha = \pi/2$ is *universal*: it does not depend any more by any parameter. Suppose that the solutions of (6.3.6) have the property that there exists a scale k'_ν above which "nothing interesting happens" so that it will be possible to replace (6.3.6) with the same equations truncated at n'_ν, depending on g (that defines a suitable value of ε). Then it becomes tautological to say that, if the statistics of the motion will be universal and determined from the statistics of (6.3.18).

Setting $\alpha = \pi/2$ the SRB statistics of the field u is related to that of the solution of (6.3.18) through the relation:

$$u_n(t) = 2\varepsilon^{1/3} k_n^{-1/3} \varphi_{n_\nu - n}(t k_{n_\nu}^{2/3} \varepsilon^{1/3}) \qquad (6.3.19)$$

and $\varphi_k(t)$ is distributed according to the SRB statistics of the equation (6.3.18) with $\alpha = \pi/2$.

If the solution $\varphi_k \equiv i$ of the (6.3.18) was stable we would have the exact validity of the law 5/3 of Kolmogorov and the statistics $\mu(du)$ should be a Dirac delta concentrated on u given by (6.3.19) with $\alpha = \frac{\pi}{2}$.

It is convenient therefore to examine the question of the stability, setting $\varphi_k = \rho_k e^{i\vartheta_k}$ and writing the equation for ρ_k, ϑ_k and linearizing them near the stationary solution $\varphi_k \equiv i$. The equations for ρ_k, ϑ_k are:

$$\begin{aligned}\dot\rho_k + 2^{-2k}\rho_k =& - 2^{-3k/2}\big(\rho_{k-2}\rho_{k-1}\cos\Delta_{k-2}-\\ &- \frac12\rho_{k-1}\rho_{k+1}\cos\Delta_{k-1} - \frac12\rho_{k+1}\rho_{k+2}\cos\Delta_k\big) + r_k\\ \dot\vartheta_k + 2^{-2k} =& 2^{-3k/2}\rho_k^{-1}\big(\rho_{k-2}\rho_{k-1}\sin\Delta_{k-2}-\\ &- \frac12\rho_{k-1}\rho_{k+1}\sin\Delta_{k-1} - \frac12\rho_{k+1}\rho_{k+2}\sin\Delta_k\big) + \sigma_k\end{aligned} \tag{6.3.20}$$

and $\Delta_k \equiv \vartheta_k + \vartheta_{k+1} + \vartheta_{k+2}$ and r_k, s_k are such that $\rho_k \equiv 1$ and $\vartheta_k \equiv \pi/2$ is an exact stationary solution.

The linearization of the (6.3.20) around to the exact stationary solution gives the following equations for η_k, δ_k, obtained by setting $\Delta_k = 3\pi/2 + \delta_k$ and $\rho_k = 1 + \eta_k$ and developing in series of the increments η_k,

$$\begin{aligned}\dot\eta_k + 2^{-2k}\eta_k =& - 2^{-3k/2}\big(\eta_{k-2} + \frac12\eta_{k-1} - \eta_{k+1} - \frac12\eta_{k+2}\big)\\ \dot\delta_k =& + 2^{-3k/2}\big(\delta_{k-2} + \frac12\delta_{k-1} - \delta_{k+1} - \frac12\delta_{k+2}\big)\end{aligned} \tag{6.3.21}$$

The *particularity* of the choice $\alpha = \pi/2$ is seen by noting that $\alpha = \pi/2$ is the *only choice* of α in (6.3.18) such that the linearized equations (6.3.21) for η and δ are separated.

It is easy to realize that the matrix M defining the equation for δ is a real traceless matrix: hence it must have *some* eigenvalues with positive real part unless they are all purely imaginary. But the matrix M is the product of a diagonal matrix times a matrix of Toeplitz and is easy see that it admits eigenvalues with nonzero real part, *c.f.r.* [BPPV93].

We deduce the *instability of the Kolmogorov solution* for the GOY model. Thus the problem arises of trying to understand how is the attracting set made and which is the correspondent statistics.

The matrix $M'_{kk'}$ that defines the stability equation for η_k is different from the $-M'_{k'k}$ because of the addition of a diagonal term $2^{-2k}\delta_{kk'}$. We can see that this matrix M' has all the eigenvalues with negative part real. Therefore one gets is led to think that the attracting set for (6.3.18) could be such that on it $\rho_k \stackrel{def}{=} (1+\eta_k)$ can be expressed in terms of the Δ_k. It is natural to suppose, [BPPV93], that in the inertial domain the relation expressing the ρ's in terms of the Δ's is essentially the condition of vanishing of the term in parenthesis in the first of (6.3.20).

We shall write the latter condition, implicitly defiining F, as

$$\rho_{k+1} = \rho_k\, e^{F(\Delta_{k+1},\Delta_k,\Delta_{k-1},\rho_k,\rho_{k-2},\rho_{k-3})} \tag{6.3.22}$$

but on the Δ_k (or the δ_k) the instability does not allow us to make any statement other than these variables will probably have a random distribution on an interval around a $3\pi/2$ (or, respectively, around a 0).

In such case the attracting set, as a geometric locus in the space of the $(\underline{\eta}, \underline{\delta})$ variables, would have (6.3.22) as equation; *and if the parameters* Δ_k *were independently distributed* then the attracting set would have dimension n_ν (*i.e.* half of the dimension $2n_\nu$ of the phase space). Furthermore the (6.3.22) show that the variables ρ_k would have a statistics that is well described by a random process with short memory.

If the eq. (6.3.22) is assumed as the equation for the attractor then the SRB distribution is concentrated on the attractor and it should *therefore* be a probability distribution μ, on the sequence of the phases $\{\delta_k\}$, characterized by Ruelle's variational principle (5.7.4): however the motion on the attractor is controlled by an equation for the δ-phases of which the second of (6.3.21) is not a good approximation away from the unstable fixed point. Therefore it is not possible, with information obtained so far, to apply the principle because it would require determining the Jacobial determinant along the unstable manifold for the evolution on the attractor. Since there is no reason to think that the phases δ_k should assume privileged values it is a reasonable guess to identify the SRB distribution with the uniform independent distribution on the parameters δ_k. From the point of view of the variational principle this is equivalent to assuming that the dynamics preserves the volume in the space of the phases $\{d_k\}$ and that the expansion along the unstable manifold is essentially constant so that the variational principle will give that the SRB distribution just maximizes the entropy and coincides with the distribution attributing a uniform and independent distribution between 0 and $2p$ to each phase d_k.

This seems to give a qualitative explanation of why a violation of the Kolmogorov law due to the instability of the solution $\varphi_k \equiv i$ leads to a multifractal distribution of the u_n.

To clarify the mechanism of this last property consider the simple case in which the Δ_k are *really independently distributed* and the relation (6.3.22) is replaced by the relation $\rho_{k+1} = \rho_k \exp f(\Delta_k)$ and the distribution of each Δ_k is $\pi(\Delta)d\Delta$, with a suitable density function π. Then under this assumption[6]

$$|u_n| = 2k_n^{-1/3}\varepsilon^{1/3}\rho_n = 2\rho_1 k_n^{-1/3}\varepsilon^{1/3}e^{\sum_{j=1}^{n-1} f(\Delta_j)} \tag{6.3.23}$$

and hence, recalling that $k_n = 2^n$:

$$\frac{\langle |u_n|^p\rangle}{\langle |u_1|^p\rangle} = 2^{-np/3}\int \prod_{j=1}^{n-1} \pi(\Delta_j)\, e^{p\sum_j f(\Delta_j)}\, d\Delta_j \tag{6.3.24}$$

[6] Which has illustrative character only, because this is obviously not the case if ρ_k is as in (6.3.22) where the variables $\Delta_{j-1}, \Delta_j, \Delta_{j+1}$ influence each other distribution. However a nearest neighbor dependence turns the distribution of Δ_j from "Bernoullian" to "markovian" and it could be equally well treated with minor technical problems to solve.

We write this as a single integral with a density function $e^{G(p)}$ defined be $e^{G(p)} \equiv \int \pi(x)\,dx\,e^{pf(x)}$, furthermore we set $l_n = k_n^{-1}$ and writing $e^{nG} = 2^{nG/\log 2} = \big(l_0/l_n\big)^{G/\log 2}$, we realize that

$$\frac{\langle |u_n|^p\rangle}{\langle |u_1|^p\rangle} = (\frac{l_n}{l_0})^{p/3-G(p)/\log 2} \qquad \to \qquad \zeta_p = \frac{1}{3}p - \frac{G(p)}{\log 2} \tag{6.3.25}$$

thus we see a concrete mechanism that can lead to a multifractal distribution, *c.f.r.* [BPPV93].

The application to the GOY model of the functional integration method of §6.1 is also promising, but it has not yet been really considered in the literature. We must note that in the present situation we can apply the method in a somewhat different version: *i.e.* to the *universal* equation (6.3.18). In this way the problem of field theory that one has to consider is a problem of scales decreasing from 1 to 0, *i.e.* it is an "infrared" problem, rather than a ultraviolet problem as in §6.2, and as it would also be in the present case if we proceeded exactly in the same way followed in §6.2. But this is not the appropriate place to attack a problem which is still so little considered in the literature.

Problems.

[6.3.1]: Study the stability of the "laminar" solution $u_n = (g/\nu k_n^2)\,\delta_{nn_0}$ of the (6.3.6) with $n_0 = 1$ and show that becomes unstable for ν small enough. (*Idea:* Check first that, as a consequence of the negative signs in the terms with n odd in (6.3.9) the argument used in problem [4.1.13] in the case of the two dimensional Navier Stokes equation does not apply to the present case.)

[6.3.2]: As [6.3.1] for the equation (6.3.7) in footnote [1].

[6.3.3]: Check that if $t \to \omega(t)$, $t \in [0,\infty)$, is a sample of a Brownian motion then $\langle|\omega(t)-\omega(t')|^{2p}\rangle, \langle|\omega(t)-\omega(t')|^2\rangle^p, |t-t'|^p$ are proportional with t,t'–independent proportionality constant: hence ζ_p in (6.3.13) is $\zeta_p = \frac{p}{2}$ (*Idea:* By definition in the Brownian motion the increments $\delta = \omega(t)-\omega(t')$ have a Gaussian distribution $e^{-\delta^2/2|t-t'|}d\delta\,/\sqrt{2\pi|t-t'|}$.)

Bibliography: The discussion of the attracting set in the GOY model is taken from [BPPV93] where one finds a complete discussion of the theory and of the experimental results on the GOY model.

CHAPTER VII

Statistical properties of turbulence

§7.1 Viscosity, reversibility and irreversible dissipation.

It is now convenient to reexamine some questions of fundamental nature with the purpose of analyzing the possible consequences of Ruelle's principle introduced in §5.7. We shall make frequently reference to the general description of motions given in Chap. 5 in a context in which we imagine that the considered motions are attracted by some attracting set in phase space, which will have zero volume when energy dissipation occurs in the system. The main purpose of this section and of the following ones is to analyze consequences of Ruelle's principle, see §5.7, with particular attention to fluid motions.

(A): Reversible equations for dissipative fluids.

First of all we must stress (again) that the derivation of the NS equations presented in §1.1, §1.2 was based on the empirical assumption that there was a viscous force opposing gliding of adjacent layers of fluid (*c.f.r.* the tensor denoted, in (1.1.17), as $\underset{\approx}{\tau}\,'$).

It is difficult to imagine how the reversible microscopic dynamics could generate a macroscopic dynamics in which time reversal symmetry is completely absent. We recall, *c.f.r.* remark (xv) to theorem I in §5.4, that if we consider a time evolution described by a differential equation of which S_t is the solution flow (so that $t \to S_t x$ is the motion with initial datum x, *c.f.r.* §5.3 Definition 3) or a discrete evolution S (associated with a timed observation, *c.f.r.* §5.2) a *time reversal symmetry* is (any) isometric map i such that

$$i^2 = 1 \quad and \quad iS_t = S_{-t} i \qquad or \qquad Si = S^{-1} i \tag{7.1.1}$$

respectively.

Note that this definition is *more general* than the often used and more common definition which takes i to be the "*velocity reversal with unchanged positions*" map, which in the case of simple fluids becomes simply velocity reversal and which for clarity of exposition will be called here, perhaps be more appropriately, *velocity reversal* symmetry. It is clear that while Newton's equations are reversible in the latter sense the NS equations do not have such velocity reversal symmetry and for this reason they are called irreversible. In principle a system may admit time reversal symmetry in the sense of (7.1.1) even though it does not admit the special velocity reversal symmetry: we shall see some interesting examples below.

The negation of above notion of reversibility is not "irreversibility": *it is instead the property that a map i does not verify* (7.1.1). This is likely to generate misunderstandings as the word irreversibility usually refers to lack of velocity reversal symmetry in systems whose microscopic description is or should be velocity reversal symmetric.

To understand how it is possible that a reversible microscopic dynamics, in the sense of velocity reversal or in the more general sense in (7.1.1), is compatible with irreversible macroscopic equations (as the NS equations manifestly are) we must think that several scales of time and of space are relevant to the problem.

The macroscopic equations are approximations apt to describe properties on "large spatial scale" and "large time scale"[1] of the solutions of reversible equations. The typical phenomenon of reversibility (*i.e.* the indefinite repetition, or "*recurrence*", of "*impossible*" states) should indeed manifest itself, but on time scales much longer and/or on scales of space much smaller than those interesting for the class of motions considered here: which must be motions in which the system could be considered as a continuous fluid.

We have already seen in §1.3 and mainly in §1.5 how the equations can change aspect if one is interested in studying a property that becomes manifest in particular regimes. For example in the theory of the Rayleigh equations of §1.5 we have seen that "in the Rayleigh regime" the general equations (1.2.1) simplify and become the equations (1.5.14) in which the generation of heat due to viscous friction between layers of fluid (last term in (1.5.6)) is absent, *c.f.r.* the third of the (1.5.14) or the comment I preceding (1.5.8).

This does not mean that friction does not generate heat. It only means, as it turned out from the analysis of §1.5, that on a time scale in which the (1.5.14) can be considered a good approximation (*c.f.r.* (1.5.12)) the quantity of heat generated is *negligible*.

The same mechanism is, or at least is believed to be, at the basis of the derivation of the equations (1.2.1) from atomic dynamics, [EM94].

This immediately makes us understand that it should be possible to express[2] the phenomenological coefficients of viscosity or thermal conductivity in terms of *averages*, over time and space, of microscopic quantities which are more or less rapidly fluctuating.

---

[1] Compared to the atomic scales where every motion is reversible.

[2] As indeed elementary kinetic theory strongly suggests that this is possible, *c.f.r.* problems [1.1.4], [1.1.5].

We deduce that the transport coefficients (such as viscosity or conductivity or other) *do not have a fundamental nature*: they must be rather thought of as macroscopic parameters that measure the disorder at molecular level.

Therefore it should be possible to describe in different ways the same systems, simply by replacing the macroscopic coefficients with quantities that vary in time or in space but rapidly enough to make it possible identifying them with some average values (at least on suitable scales of time and space). *The equations thus obtained would then be equivalent to the previous.*

Obviously we can *neither hope nor expect* that by modifying the equations (1.2.1) into equations in which various constant are replaced by variable quantities we shall obtain simpler or easier equations to study (on the contrary!). However imposing that equations that should describe the same phenomena do give, actually, the same results can be expected to lead to *nontrivial relations* between properties of the solutions (of both equations).

This is a phenomenon quite familiar in statistical mechanics of equilibrium where one can think of describing a gas in equilibrium at a certain temperature as a gas enclosed into an adiabatic container with perfect walls or as enclosed in a thermostat at the same temperature.

The two situations are described respectively by the microcanonical distribution and by the canonical distribution. Such distributions are *different*: for example the first is concentrated on a surface of given energy and the other on the whole phase space. A sharp difference, indeed, being the surfaces of constant energy sets with zero volume.

Nevertheless the physical phenomena predicted in the two descriptions must be the same: it is well known that from this Boltzmann derived the *heat theorem*, [Bo84],[Ga95c] p. 205,[Ga99a], *i.e.* a proof of the second law of equilibrium thermodynamics, and the general theory of the *statistical ensembles*.

Hence providing different descriptions of the same system is not only possible but it can even lead to laws and deductions that would be impossible (or at least difficult) to derive if one did confine himself to consider just a single description of the system.

What just said *has not been systematically applied to the mechanics of fluids*, although by now there are several deductions of macroscopic irreversible equations starting from microscopic velocity reversible dynamics, to begin with Lanford's derivation of the Boltzmann equation, [La74]. In the remaining sections I try to show that the above viewpoint is at least promising in view of its possible applications to the theory of fluids.

It is well known that Boltzmann was drawn into disputes, [Bo97], occasionally quite animated, to defend his theory of irreversibility. His point has been that one should make a distinction between reversibility of motion and irreversibility of the phenomena that the accompany it.

The works of Sinai on Anosov systems, *c.f.r.* [Si94], [Ru79], show why it is not necessary, in order to make this distinction clear, to deal with systems of very many particles, to which Boltzmann was always making reference. Systems with few degrees of freedom, even 2 degrees of freedom, are sufficient at least for illustration purposes and show irreversible phenomena although their motions are governed by a reversible dynamics.

This phenomenon has been empirically rediscovered, independently by several experimental physicists:[3] some hailed it as the "solution of the Loschmidt paradox",[HHP87], [Ho99], correctly seeing its relation with the Boltzmanian polemics (without perhaps taking into account that, at least in the case of Boltzmann, the question had been already solved and precisely in the same terms posed by Boltzmann himself, in papers that few had appreciated, [Bo84],[Bo97], see [Le93], [Ga99a]).

Therefore keeping in mind the above considerations we shall imagine other equations that should be "equivalent" to the Navier–Stokes incompressible equation (in a container Ω with periodic boundary conditions).

Note that a forced fluid has an average energy and an average dissipation that rapidly end up fluctuating around an average value depending only on the acting force (I simplify here a little, to avoid trivial digressions needed to take into account situations in which there are important *hysteresis phenomena*, *i.e.* several attracting sets, *e.g.* [FSG79], [FT79]).

In situations in which viscosity is small (*i.e.* the Reynolds number is large) the theory K41 suggests that essentially the fluid flows subject to Euler's equations (*i.e.* with zero viscosity acting on the "important" degrees of freedom in number of $O(R^{9/4})$, *c.f.r.* §6.2,) but *with energy dissipation* rate constant in time. The rate of energy dissipation in an incompressible Navier–Stokes fluid[4] is

$$\eta = \nu \int_\Omega (\underline{\partial} \wedge \underline{u})^2 d\underline{x} \tag{7.1.2}$$

(this quantity is called ε in §6.2). Clearly the way in which dissipation takes place is not properly accounted for by Euler's equation and the phenomenon of heat production due to friction can possibly be well described only from a really microscopic viewpoint, *c.f.r.* the problems of the §1.1. However apart from the heat production the equations of Euler supplemented with some mechanism that takes away the energy supplied to the system by the forcing forces should properly describe the flows, *c.f.r.* §6.2.

Then if the only difference between the Euler equations and the NS equations at low viscosity is the existence of dissipation we can imagine another equation that has the same properties; *i.e.* the Euler equation with the addition of a force that performs work on the system but in such a way to absorb (in the average) a constant quantity of energy per unit time.

[3] Long after the early works on the SRB distributions, [Lo63], [Si68], [Bo70], [Ru76].

[4] In which (7.1.2) equals $\nu \int_\Omega dx\, (\underline{\partial} \underset{\sim}{u} + \underset{\sim}{\partial} \underline{u})^2/2$, *c.f.r.* (1.2.7).

It is interesting that already Gauss posed the problem of which would be the *minimum* force necessary to impose a constraint (whether *holonomic* or *anholonomic*). The principle of Gauss, *c.f.r.* problems, applied to an ideal fluid subject to the constraint of dissipating energy at constant rate leads to the following equations

$$\dot{\underline{u}} + \underset{\sim}{u} \cdot \underset{\sim}{\partial}\, \underline{u} = -\frac{1}{\rho}\underline{\partial} p + \underline{g} + \alpha\, \Delta\, \underline{u}, \qquad \underline{\partial} \cdot \underline{u} = 0 \tag{7.1.3}$$

where $\alpha = \alpha(\underline{u})$ *is not* the viscosity but, rather, it is the multiplier necessary to impose the constraint that (7.1.2) is a constant of motion for the equation (7.1.3). A simple computation provides us with:

$$\alpha(\underline{u}) = \frac{\int \big(\underline{\partial} \wedge \underline{g} \cdot \underline{\omega} + \underline{\omega} \cdot\, (\underset{\sim}{\omega} \cdot \underset{\sim}{\partial}\, \underline{u})\big)\, d\underline{x}}{\int (\underline{\partial} \wedge \underline{\omega})^2\, d\underline{x}} \tag{7.1.4}$$

The idea that is suggested by the analysis developed until now is precisely that the equations (7.1.3), (7.1.4) are *equivalent* to the NS equation. Suppose that the solutions of the NS equation with given viscosity and force admit an attracting set on which the average energy dissipation has a certain value $\eta(\nu)$ and imagine that the r.h.s. of the relation (7.1.2) is fixed to be *precisely equal* to η via a Gaussian constraint of the type described by (7.1.3) and (7.1.4). Then the average values of the observables *may turn out to be the same* with respect to the statistics of the motions on the attracting set for the NS equations and for the equations (7.1.3) and (7.1.4). Such identity will certainly be approximate if the Reynolds number, *i.e.* the intensity of the force, is finite but one can conjecture that it can become more and more exact as R increase.

We shall call the (7.1.3) and (7.1.4) *GNS equations*, or *Gaussian Navier–Stokes equations*. And the just proposed conjecture is, formally

Equivalence conjecture: *Consider the GNS equations (7.1.3), (7.1.4) with initial data, in which the quantity η in (7.1.2) is fixed equal to the average value of the same quantity with respect to the SRB distribution, that we denote $\mu_{\nu,ns}$, for the NS equations with viscosity ν. Let $\mu_{\eta,gns}$ be the SRB distribution for the GNS equations thus defined.*
(i) Then the distributions $\mu_{\nu,ns}$ and $\mu_{\eta,gns}$ assign, in the limit in which the Reynolds number tends to infinity,[5] *equal values to the same observables $F(\underline{u})$ that are "local" in the momenta, i.e. that depend only on the Fourier components of the velocity field $\underline{\gamma}_{\underline{k}}$ with $\underline{k}$ in a finite interval of values of $|\underline{k}|$.*
(ii) In such conditions $\langle\alpha\rangle_{\eta,gns} = \nu$, if $\langle\cdot\rangle_{\eta,gns}$ and $\langle\cdot\rangle_{\nu,ns}$ denote the average values with respect to the distributions $\mu_{\eta,gns}$ and $\mu_{\nu,ns}$ respectively.

[5] *e.g.* $\nu \to 0$ at fixed external force density and at fixed container size.

Remarks:

(1) The conjecture is closely related, and in a way it is a natural extension, of similar conjectures, [Ga99a], of equivalence between different thermostats in particle systems. For the latter systems it was clearly formulated and pursued, in particle systems, in several papers by Evans and coworkers starting witht he early 1980's: for a more recent review see [ES93] where a proof under suitable assumptions is presented. For fluids it was proposed in [Ga96] but in a sense it was already clear from the paper [SJ93] and it is somewhat close to ideas developed in [Ge86], see also [GPMC91] and the review [MK00].
(2) This conjecture proposes therefore that *for the purpose of computing average values* the NS and GNS equations are *equivalent* provided, of course, the free parameters in the two equations are chosen in a suitable relation.
(3) It will not escape to the reader that the described correspondence is very analogous to the equivalence, so familiar in statistical mechanics, between the statistical ensembles that describe equilibrium of a single system,[Ga95c], [Ga95b], [Ga99a].
(4) We see therefore how this conjecture expresses that different equations can describe the same phenomena: and in particular the *reversible* GNS equation and the NS equation (much better known and *irreversible*), describe the same physical phenomenon at least for R large.
(5) A formal way to express the point (i) of the conjecture is that

$$\lim_{R\to\infty} \frac{\langle F\rangle_{\nu,ns}}{\langle F\rangle_{\eta(\nu),gns}} = 1 \tag{7.1.5}$$

for all local observables $F(\underline{u})$ whose average $\langle F\rangle_{\nu,ns}$ does not tend to 0 as $R\to\infty$.

(B) Microscopic reversibility and macroscopic irreversibility.

The question is then how could this dual reversible and irreversible nature of the phenomena be possible?

The crucial point to remark is *that there is no relation between irreversibility, understood in the common sense of the word, and lack of reversibility of the equations that describe motions.*

At first sight this appears paradoxical: but this is a fact that becomes substantially already clear to anyone who studies, even superficially, the disputes on irreversibility between Boltzmann and his critics, [Bo97],[Ga95c],[Ga95b]. We therefore examine this "paradox" in more detail.

A reversible dynamical system (M, S), see (7.1.1), in general, shall have attracting sets A that can *fail to be invariant under time reversal*: $iA \neq A$.

The case in which the sets A and iA are really different is clearer and we shall discuss it first, keeping in mind, however, that it is a case that arises in systems that are strongly out of equilibrium. Indeed for systems in equilibrium we imagine, at least when their evolution is chaotic, that the attracting set is the whole phase space (a consequence of the "ergodic hypothesis"): and this remains true also if the system is subject to small external forces that keep it out of equilibrium, (this is a *structural stability* theorem in the case of simple chaotic motions, *c.f.r.* §5.7).

Let us suppose, for simplicity, that there is a unique attracting set A. Then the set iA is a *repelling set*, *i.e.* it is an attracting set for motions observed *backwards in time* (described by S^{-1}).

Motions starting in the basin of attraction of A develop, after an initial transient, essentially on the attracting set A. Hence they *no longer exhibit any symmetry with respect to time reversal*. By "motions" we mean here "typical' motions", *i.e.* all of them except a set of 0–volume in phase space or except a set of 0–probability with respect to a distribution μ_0 which is "absolutely continuous" with respect to volume.

Strictly speaking such motions do not start, with μ_0–probability 1, do not start exactly on A but get close to A with exponential rapidity, as time runs past, *without ever getting into it*. If we could really distinguish the point x reached after a long time from a point x' *near it but located exactly* on the attracting set then, by proceeding backwards in the time, the point x will go back and in the long run it will get, with exponential rapidity, very close to iA, *i.e.* to the repelling set, because the latter is an attracting set for the motions that are observed backwards in time. A completely different behavior will be that of the point x', once close to x *but lying on the attractor*, because it will stay forever on A.

However the time reversal invariance *does not refer to the difference between these two motions* followed backwards and foward in time, rather it usually "only" says that if we inverted all velocities of the points of the system (such is the effect of the map i in most cases) then the system would go through a trajectory that, *as time increases*, coincides in position space with that followed until the instant of the reversal, while in velocities space it has velocities systematically opposite to those previously assumed when occupying the same positions.

By applying the time reversal map to a state x of the system very near to A (*but not on* A) one finds a state ix very close to iA (*but not on* iA) and hence "atypical":[6] but proceeding in the motion, as the time n increases, *also* $S^n\,i\,x$ will *again* evolve getting close to A, very rapidly so: hence the initial state ix would appear as an atypical fluctuation.

This *would not happen* if instead of considering a state x very near A one considered one *exactly* on A:[7] the state ix would be exactly on iA and

---

[6] For example, in a system of charged particles which is out of equilibrium because of an field electric acting on it, such state would have a current opposite to field.

[7] A 0–probability event if initial data are randomly chosen with a distribution μ_0 abso-

then, proceeding in time *both towards the future and towards the past*, this point ix would move on iA without ever leaving it. And this evolution, *anomalous for an observer unaware of the particularity of the initial state*, would indefinitely continue on $i\,A$.

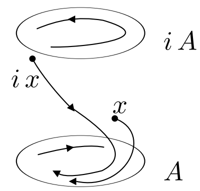


Fig. (7.1.1): *Illustration of a "strongly forced" system with attracting set A different from its time reversal image iA. The forward evolution of a point x and of its time reversal image ix. Both x and ix eventually approach A in their forward evolution. Motions starting on A or iA stay there in the forward as well as in the backwards evolutions.*

An observer could say that the system behaves in a irreversible way in which starting from a configuration x it reaches asymptotically a stationary state μ, always the same whatever the initial state is (with μ_0–probability 1) and *the same* for x and ix. Furthermore proceeding backwards in time the system reaches *instead* (always with μ_0–probability 1) a stationary state μ_- different[8] from the state μ but still *the same* for x and ix.

Starting instead with an initial datum randomly chosen precisely on the 0–volume attracting set A and *with the asymptotic and stationary statistics* μ one generates a motion that, *both towards the future and towards the past*, has the same statistics μ: *i.e.* starting with an initial datum typical of the stationary state and proceeding backwards *there is no way to reach* states which are not typical of the stationary state (such as the states close to the repelling set iA).

The question is, as we see, rather delicate: macroscopic irreversibility is, in this case, the manifestation of the existence of an attracting set $A \neq iA$ while the microscopic reversibility only implies that to every attracting set must correspond a repelling one that is substantially a copy of it. This implies that except for an initial transient the property of the motions towards the future are the same for all x's and $i\,x$'s (unless one succeeds in choosing an initial datum *exactly* on the sets of zero phase space volume A or iA).

But if one was able to measure and fix exactly all coordinates of the system then the transient time could be made as long as wanted: it would be enough to invert exactly all the velocities at a given time and the motion of a system that has been observed as absolutely "normal" for a prefixed time T would develop *in a way now absolutely strange* "backwards" for the same time T: and then, at least if the initial data were randomly chosen with a distribution

lutely continuous with respect to the volume

[8] Although "isomorphic" to μ because of time reversal symmetry, *i.e.* a state that can be transformed into μ via the map i.

μ_0 proportional to the volume of phase space, it would proceed again in a normal way *and for ever so.*

What said would seem to give rise to problems in the case in which the attracting set A is the whole space (accessible compatibly with the initial value of the possible constants of motion) and hence it coincides with iA: in fact this happens in systems in thermodynamical equilibrium or slightly away from it.

In reality no new problem arises because, upon an attentive exam, the attracting set A whether equal or different from iA will be such that, obviously, not all points of A are equally probable with respect to the statistics μ on A.

We have seen in §5.5 that the SRB statistics is "in some sense" concentrated on the more stable motions: (5.5.8) says indeed that, if the system is "approximated" by the set of its periodic motions of period n with n very large, then the statistical weight, for the purpose of the SRB averages, of the periodic motions is inversely proportional to their "instability", *i.e.* to the product of the eigenvalues larger than 1 (in average) of their stability matrix (*c.f.r.* the factor $\Lambda_e(x)^{-1}$ in eq. (5.5.8)).

Hence even when it is $A = iA$ it remains nevertheless possible that the properties of the motions towards the future are, *for possibly very long transient times*, very different from the average ones and this can be seen by the behavior of the fluctuations and is essentially the only form in which irreversibility can manifest itself in systems in equilibrium.

What can then be said about the cases in which $A = iA$ but the system *is not* in equilibrium? (Think of a gas of charged particles in a toroidal container subjected to an axially directed weak, but not vanishingly small, electric field).

(C) Attractors and attractive sets.

In reality the notion of attracting set, as a closed set A to which the motions starting near enough get nearer and nearer, is too rough to describe what happens and it is not adequate when $A = i\,A$.

In §5.5 we introduced the distinction between attractive set and attractor for data randomly chosen with distribution μ_0 with density with respect to volume on phase space, defining (we recall):

Definition: *(reminder of the notion of attractor):*
Given the dynamical system (M, S) an attractor for motions with initial data chosen with a distribution μ_0 in the vicinity of an attracting set A is any invariant set $A_0 \subset A$ which has probability 1 with respect to the statistics μ of these motions and at the same time has minimum Hausdorff dimension. The value of this minimum is the "information dimension" of the attractor. If μ_0 is absolutely continuous with respect to volume the information dimension is also called the dimension of the system (M, S) on

A.

This is a much more precise notion than that of attracting set and it allows us to distinguish A_0 from iA_0 also when the closures A and iA of these sets are the same. It allows us, furthermore, to define naturally the dimension of the attractor contained in A, so that this notion is not trivial even when A_0 is dense in phase space and A is therefore the whole phase space.

For example in conservative systems it is usually (believed to be) true that the attracting set is the entire surface of given energy ("*ergodic hypothesis*"). And in such cases it also happens that there exists *only one* attracting set for the motion towards the future and towards the past of data chosen randomly with a distribution proportional to the Liouville measure μ_0.

Forcing these systems and providing, at the same time, also some dissipation mechanism allowing them to keep constant (or bounded) the energy and hence to reach a stationary state, we obtain systems that still have the *whole phase space* as an *attracting set* at least if the force is not too large (structural stability of chaotic motions, *c.f.r.* §5.7): but this time the *attractor* for the motion towards the future and that towards the past will be different.

In the sense that it will be possible to find two sets, *both dense on the full phase space*, A_0 and A'_0 to which the SRB statistics μ_+ and μ_-, for the motions towards the future and those towards the past, attribute probability $\mu_+(A_0) = 1 = \mu_-(A'_0)$ while $\mu_+(A'_0) = \mu_-(A_0) = 0$.

The above analysis of the distinction between microscopic reversibility and macroscopic irreversibility can be repeated and remains substantially unchanged. *Hence also the discussion made in point (B) can be essentially repeated with A_0 and A'_0 playing the role of A and iA*: we see that microscopic reversibility and macroscopic irreversibility are compatible also in this case.

In strongly out of equilibrium systems (*e.g.* a fluid at large Reynolds number) we expect (as said above) that the attracting sets for motions towards the future and towards the past are different: this can be interpreted as a *spontaneous breaking* of time reversal symmetry and, as discussed in (B), it provides us with a simpler version of the mechanism that shows the compatibility between microscopic reversibility and macroscopic irreversibility. A mechanism that, as we have seen, is more hidden in the cases of equilibrium or close to equilibrium because of the identity of the closures of the attractor and of the repeller.

We conclude the discussion by discussing, to provide a concrete example, various aspects of the abiove analysis in the case of the well known case of the expansion of a gas in a (perfect) container of which it initially occupies only a half, we can distinguish the adiabatic expansion from the isothermal expansion (*i.e.* with the system in thermal contact with a heat reservoir keeping fixed its temperature identified with its average kinetic energy).

In the first case the system is Hamiltonian and it will evolve towards an attractor A_0 that is dense on the whole surface of energy E (with E equal to the initial energy of the gas) and, in the statistics μ of the motion, the initial state will appear as a very rare fluctuation. In the second case, instead, the attractor will be determined by a microscopic mechanism of interaction between gas and thermostat (and it will appear to have an energy which will be distributed with very unlikely fluctations around its *average* determined by the temperature of the heat reservoir).

In both cases the attractor will still be symmetric with respect to time reversal and dense on the available phase space; the initial state will still appear as a rare fluctuation.

If now the gas is imagined to be made of charged particles on which an electric nonconservative field (*i.e.* an electromotive force) acts, the system will reach a stationary state only in presence of a mechanism of interaction with a thermostat or with external bodies that absorb the energy generated by the work of the field. If the field is different from zero the *attractors* A_0 and $A_0' = iA_0$ *become different* and of 0–volume in phase space: and if the field is sufficiently strong then it will become possible that *even* the attracting set A, closure of A_0, becomes smaller than the whole phase space and different from A', closure of A_0'. This would be a case of "*spontaneous breakdown*" of time reversal symmetry: it will be discussed again later.

Problems:

[7.1.1]: Let $\varphi(\underline{\dot{x}},\underline{x}) = 0$, $\underline{x} = \{\underline{\dot{x}}_j,\underline{x}_j\}$ be a general anholonomic constraint for a mechanical system. Let $\underline{R}(\underline{\dot{x}},\underline{x})$ be the constraint reaction and $\underline{F}(\underline{\dot{x}},\underline{x})$ be the active force. Consider all possible accelerations compatible with the constraints when the system is in the state $\underline{\dot{x}},\underline{x}$. We say that $\underline{R}$ is *ideal* or *verifies the principle of least effort* if the actual acceleration due to the forces $\underline{a}_i = (\underline{F}_i + \underline{R}_i)/m_i$ minimizes the *effort*: $\mathcal{E} = \sum_{i=1}^N \frac{1}{m_i}\underline{R}_i^2 \equiv \sum_{i=1}^N (\underline{F}_i - m_i\underline{a}_i)^2/m_i$, *i.e.*

$$\sum_{i=1}^N (\underline{F}_i - m_i\underline{a}_i)\cdot\delta\underline{a}_i = 0$$

for all the possible variations of the accelerations $\delta\underline{a}_i$ compatible with the constraints φ at *fixed* velocities and positions. Show that the possible accelerations, in the configuration $\underline{\dot{x}},\underline{x}$, are those such that: $\sum_{i=1}^N \partial_{\underline{\dot{x}}_i}\varphi(\underline{\dot{x}},\underline{x})\cdot\delta\underline{a}_i = 0$.

[7.1.2]: Show that, thanks to the observations in [7.1.1], the condition of minimum constraint becomes:

$$\underline{F}_i - m_i\underline{a}_i - \alpha\,\partial_{\underline{\dot{x}}_i}\varphi(\underline{\dot{x}},\underline{x}) = \underline{0}$$

$$\alpha = \frac{\sum_i(\underline{\dot{x}}_i\cdot\partial_{\underline{x}_i}\varphi + \frac{1}{m_i}\underline{F}_i\cdot\partial_{\underline{\dot{x}}_i}\varphi)}{\sum_i m_i^{-1}(\partial_{\underline{\dot{x}}_i}\varphi)^2}$$

which is the analytic expression of Gauss' principle, *c.f.r.* [LA27], [Wi89]. Of course the definition of effort $\mathcal{E}$ is quite arbitrary and modifying it leads to different analytic expressions.

f**[7.1.3]:** Check that if the constraints are holonomic then Gauss' principle reduces to the principle of D'Alembert. (*Idea:* Note that the velocities permitted by the holonomic constraint $\varphi(\underline{x}) = 0$ are $\underline{\dot{x}}\cdot\underline{\partial}\varphi(\underline{x}) = 0$ and hence a holonomic constraint can be thought of as a constraint anholonomic having the form special: $\underline{\dot{x}}\cdot\underline{\partial}\varphi(\underline{x}) = 0$.)

[7.1.4]: Consider a system of N particles subjected to a conservative force with potential energy V. Consider the system of points subject to the force $\underline{f} = -\underline{\partial} V$ and to a Gaussian constraint imposing that $T = const$ (where T is the kinetic energy $T = \sum_i \underline{p}_i^2/2m$). Verify that the equations of motion are (by [7.1.3]):

$$m\underline{\dot{x}}_i = \underline{p}_i,\ \underline{\dot{p}}_i = -\underline{\partial}_{\underline{q}_i} V - \alpha\, \underline{p}_i \stackrel{def}{=} \underline{F}_i, \qquad \alpha = -\frac{\sum_i \underline{\partial}_{\underline{q}_i} V \cdot \underline{p}_i}{\sum_i \underline{p}_i^2}$$

Show that for arbitrary choices of the function $r(T)$ the probability distribution on phase space with density: $\rho(\underline{p}, \underline{q}) = r(T) e^{-\beta V(\underline{q})}$ is invariant if, defined ϑ as $3N\, k_B \vartheta/2 \stackrel{def}{=} T$: $\beta = 3N - 1/(3Nk_B\, \vartheta)$. (*Idea:* The continuity equation is indeed $\partial_t \rho + \sum_i \underline{\partial}_{\underline{p}_i} (\rho \underline{F}_i) + \sum_i \underline{\partial}_{\underline{q}_i} (\rho\, \underline{p}_i/m) = 0$).

For Gauss principle applications to fluid mechanical equations see problems in §7.4.

Bibliography: [GC95a], [GC95b],[Ga95],[Ga95b],[Ga96],[LA27].

§7.2 Reversibility, axiom C, chaotic hypothesis.

I do not intend to claim that the description that I have given of the soul and of its functions is exactly right – a wise man could not possibly say that. But I claim that, once immortality is accepted as proved, one can think, not improperly nor lightheartedly, that someting like that is true.
(words of Socrates (in Phaedon)).

To study in more detail some of the problems posed in the discussion in §7.1 we shall refer, assuming it *a priori*, to Ruelle's principle of Sec. §5.7.

(A) The SRB distribution, and other invariant distributions.

Suppose that the dynamical system (M, S) has an attracting set A verifying axiom A (*c.f.r.* definition 2, §5.4): it will then be possible to describe the points of A via the *symbolic dynamics* associated with a Markovian pavement $\mathcal{P}$, *c.f.r.* §5.7 (C).

If the logarithm of the determinant $\Lambda_e(x)$ of the matrix ∂S^n, thought of as a map acting on the unstable manifold of $x \in A$, is considered as a function $\lambda_e(\underline{\sigma})$ of the history $\underline{\sigma}$ of x then, since the function $\Lambda_e(x)$ is "regular" (Hölder continuous), the dependence of $\lambda_e(\underline{\sigma})$ on the "far" digits of $\underline{\sigma}$ is also exponentially small.[1]

In this context we can make use of the expression (5.7.8) for the SRB distribution in terms of the code $x = X(\underline{\sigma})$ between points and histories.

If ϑ denotes the translation to the left of the histories and if μ is the SRB distribution on A the equations (5.7.7), (5.7.8) yield the following expression for μ:

$$\int f(y)\mu(dy) = \lim_{N\to\infty} \frac{\sum_{\sigma_{-N/2},\ldots,\sigma_{N/2}} e^{-\sum_{j=-N/2}^{N/2} \lambda_e(\vartheta^j \underline{\sigma})} f(X(\underline{\sigma}))}{\sum_{\sigma_{-N/2},\ldots,\sigma_{N/2}} e^{-\sum_{j=-N/2}^{N/2} \lambda_e(\vartheta^j \underline{\sigma})}} \qquad (7.2.1)$$

where $\underline{\sigma}$ is an infinite compatible sequence in the sense of the histories on Markovian pavements, *c.f.r.* §5.7, (C), obtained extending the string of digits $\sigma_{-\frac{1}{2}N}, \ldots, \sigma_{\frac{1}{2}N}$ in "a standard way", *c.f.r.* §5.7, (D).

If we replace, in formula (7.2.1), the function $\lambda_e(\underline{\sigma})$ with an *arbitrary* function $\rho(\underline{\sigma})$ that has a very weak dependence on the digits σ_i with large label i (for example it depends only on the digit with label 0) the new formula still defines an invariant distribution μ' on A which, however, in general is *completely different from* μ.[2]

Just as the probability distributions of two different Bernoulli schemes can be chosen[3] to be different even if they have the same space of states: in fact their attractors consist in the sequences that have given frequencies of appearance of given symbols; but such frequencies are *different* in the two cases: hence the attractors are *different and disjoint* sets. Nevertheless both attractors are dense in the space of all sequences[4] and the space of all sequences is therefore the attracting set for both cases, while the attractors are different (and can be chosen so that they do not have points in common).

This example lets us well appreciate the difference between the two possible notions of attracting set and of attractor. It makes us, furthermore, see that in a system with an attracting set A verifying axiom A *there exist infinitely many other invariant distributions besides* the SRB distribution which can be very different from, or just about equal to, the SRB. In fact

---

[1] Indeed the digits of the history $\underline{\sigma}$ determine the point x with exponential rapidity, *i.e.* the distance between two points whose histories coincide between $-N$ and N tends to zero as $e^{-\lambda N}$, *c.f.r.* §5.7: hence the values of $\lambda_e(\underline{\sigma})$ and $\lambda_e(\underline{\sigma}')$ differ by $O(e^{-\alpha\lambda N})$, if the sequences $\underline{\sigma}$ and $\underline{\sigma}'$ coincide between $-N$ and N and if α is the exponent of Hölder continuity of $\Lambda_e(x)$.

[2] The (7.2.1), even if modified in this way, can be interpreted as the definition of the thermodynamic limit of an unidimensional "Ising model" with a short range interaction, see also §9.6 in [Ga99a]; hence this property is a well known result: the technique for the proof is illustrated in detail in the problems [5.7.1]÷[5.7.16] of §5.7 taken from [Ga81].

[3] We recall that while an attracting set is closed and uniquely characterized by the dynamics, an attractor is only defined up some (trivial) ambiguity, *c.f.r.* definition in (C) of §7.1 and §5.7.

[4] If the distance between two sequences is, as usual, defined as e^{-N}, where N is the largest value for which $\sigma_i = \sigma'_i$ for $|i| \le N/2$.

given an arbitrary *finite number* of observables one can define an invariant probability distribution μ' concentrated on A (*i.e.* giving probability 1 to A: $\mu'(A) = 1$) and attributing to the chosen observables average values close to those attributed by the SRB distribution within a prefixed approximation (one obtains μ' by simply modifying by a small quantity the functions $\lambda_e(\underline{\sigma})$ in (7.2.1)).

(B) Attractors and reversibility. Unbreakability of the time reversal symmetry.

Dynamical systems can be reversible, and in general their attracting sets A differ from the (repelling) sets iA that are their images under the time reversal map i, *c.f.r.* §7.1.

This can be interpreted, as already observed in §7.1, as a phenomenon of *spontaneous symmetry breaking*: one can think that phase space is precisely the attracting set A and ignore, for the purpose of studying the statistical properties of motions, the points outside of A. Limiting ourselves, for simplicity, to the case of an attracting set A that is a regular surface, we see that as far as asymptotic (in the future) observations are concerned the dynamical system is *de facto* the system (A, S).

Obviously this system *is no longer reversible* if time reversal is performed with the map i (which *cannot* even be thought of as a map of A into itself, because $iA \neq A$).

We shall see however, in what follows, that reversible systems with an attracting set that coincides with the whole space have extremely interesting properties. In this respect we can ask whether we could define another map $i^* : A \longleftrightarrow A$ that anticommutes with the evolution, *i.e.* $i^*S = S^{-1}i^*$ *while leaving invariant the attracting set* A and squaring to the identity, $i^{*2} = 1$.

If, with some generality, it was possible to define i^* we could say that the dynamical system is still reversible, although the time reversal symmetry is now i^* and not the original i. In this way, in the same generality, reversible systems endowed with an attractive set that is *smaller* than the entire phase space could also be considered as systems enjoying time reversal symmetry and endowed with an attractive set that *coincides* with the whole phase space.

In other words time reversal symmetry would be unbreakable. When spontaneously violated it would spawn an analogous symmetry on every non symmetric attracting set!

A first example of unbreakability of time reversal symmetry, or at least an example of a very similar phenomenon, can be found even in fundamental Physics. In relativistic quantum theories that should describe our Universe time reversal symmetry T, that we could think as valid at a fundamental level, is spontaneously broken, as it is well known, *c.f.r.* [Al93] p. 241. But a symmetry which anticommutes with time evolution continues to exist as a symmetry of the dynamics of our Universe: it is the TCP symmetry.

By applying to our universe the map T we obtain another universe, absolutely different from ours, but equally possible.

One can think that the fundamental equations are symmetric with respect to the map T, *i.e.* reversible, but dissipative. Therefore our Universe would evolve towards an attracting set smaller than the whole phase space and it would no longer be time reversal invariant. Nevertheless if time reversal symmetry was actually unbreakable (in the above sense) then the motion would be still symmetric with respect *to another operation* that inverts the sign of time, that could be TCP.

To support what just said one can fear that it would be necessary to think that there exists a dissipative mechanism in the dynamics of the Universe: nothing more unsatisfactory. However the dissipation of which we talk here *would not be* the empirical dissipation to which we are perhaps used: *because the fundamental equations would remain reversible.* One could rather think of a level higher than the one accessible directly to us, a "Universe of Universes", that acts on our evolution like a reversible thermostat acts on the evolution of a gas or of a fluid out of equilibrium (*i.e.* as a force that absorbs heat without breaking time reversal, like the forces that impose the constant energy or constant dissipation in the ED or GNS equations in (7.1.1)). This would be sufficient and it would allow us to think that the Universe evolved rapidly, ending up on an attracting set that does not have any more the symmetry T but "only" TCP.

Or, with a larger conceptual economy and without crossing (as done above) the border of science fiction, one could think that it is our same Universe to act, in a reversible though dissipative way, as a thermostat on the world of elementary particles generating the symmetry breaking of their dynamics that we observe experimentally. The asymmetry observed in the weak interactions could be a trace in the subatomic world of the asymmetry that we observe between past and future, at macroscopic level. It would certainly be important to produce a concrete and credibile model for the mechanism of interaction between the macroscopic (atomic) world and the microscopic (subatomic) one: by imagining a lagrangian for the description of the weak interactions, *c.f.r.* [Al93], that is *a priori* not time reversal invariant we, perhaps, only take into account the atomic–subatomic interaction phenomenologically. This vision does not seem absurd to me, although admittedly is very daring.

(C) An example.

Coming back to our much more modest analysis of the motion of a gas or of a fluid, let us consider the structure of the motions of a reversible system endowed with an attracting hyperbolic set A, that we think as a *regular* surface of dimension lower than that of phase space.

The points of A will have stable and unstable manifolds: the unstable will be entirely contained in A, and the stable will consist of a part on A and a part *outside* A, precisely because A is attractive.[5]

We shall assume, for simplicity, that the dynamical system has only two invariant closed sets of *nonwandering* points, A and its image iA, that we shall call *poles*, *c.f.r.* §5.4 observation (2) to definition 2.

The stable manifold of the points of A extends out of A reaching the set iA: *i.e.* if x is a point near A (*but not on it*) and if we follow it backwards in time we see that $S^{-n}x$ tends to the set iA, that repells, *i.e.* that attracts motions seen backwards in time.

Therefore the stable manifold in question extends until iA and we must expect that it is "dense" on this set (meaning that the closure of such manifold will contain iA). It can *a priori* behave, in its vicinity, in several ways: for example it could *wrap* around iA.

Nevertheless the *simplest geometric hypothesis* is that the manifold reaches the surface $i\,A$ "cutting" this surface in a "transversal" way: this is possible even though, obviously, no point of the manifold can belong to $i\,A$.

To understand how to interpret the latter geometric property we first discuss a paradigmatic example, see Fig.(7.2.1) below, with the aim of a later abstraction of a general formulation.

Referring to Fig.(7.2.1) below, the poles $A_+ = A$ and $A_- = iA$ are, in the example in Fig.(7.2.1), proposed below, two regular closed and bounded surfaces identical, copies of a surface M^*.

The map S on the *whole* phase space will be defined by

$$S(x,z) = (S_* x, \tilde S z) \tag{7.2.2}$$

where the generic phase space point is a pair (x,z) with $x \in M^*$ and z is a set of transversal coordinates that tell us how far away the point (x,z) is from the attracting set.

The point z will be imagined as a point on a smooth manifold Z and $\tilde S$ will be an evolution on Z that has two fixed points, one z_- (unstable) and another z_+ (stable). Furthermore $\tilde S$ will be supposed to evolve all points $z \neq z_\pm$ in such a way that all motions tend to z_+ in the future and to z_- in the past. This can be realized in many ways and we select arbitrarily one of them. By construction the sets A_+ and A_- are the sets with $z = z_+$ and, respectively, $z = z_-$.

The coordinate x identifies a point on the compact surface M^* on which a reversible map S_* acts: we suppose the system (M^*, S_*) is an Anosov

[5] This means that at every point of A the stable manifold has a tangent plane which can be decomposed into a direct sum of two independent planes one of which is tangent to A and one which is transversal to it. Since A *in general is not a smooth surface*, and it might have fractal structure, this has to be made more precise; supposing for simplicity that A is a set immersed in R^d, for some d, the "tangent" part of the stable plane at x will be such that the maximum distance to A of points y on it and with $d(y,x) \le r$ is $< O(r^2)$ while the maximum distance to A of points y on the part of the stable plane outside A will be $> O(r)$. The second might be empty.

system. For example M^* can be the bidimensional torus $\mathcal{T}^2$ and S_* could be the "Arnold cat", *c.f.r.* §5.3

$$S_*(\psi_1\,,\,\psi_2)=\begin{pmatrix}1&1\\1&2\end{pmatrix}\begin{pmatrix}\psi_1\\\psi_2\end{pmatrix}\quad\mod\,2\pi \tag{7.2.3}$$

which is reversible if one defines time reversal as the map that permutes the coordinates of $\underline{\psi}$ and changes the sign to the first, *i.e.* $i'x=i'(\psi_1,\psi_2)=(-\psi_2,\psi_1)$.

For instance Z could be a circle $z=(v,w)$ with $v^2+w^2=1$. The evolution $\tilde{S}$ could be the map that one obtains by considering at time $t=1$ the point $\tilde{S}(v,w)$ into which (v,w) evolves according to the differential equation $\dot{v}=-\alpha v,\ \dot{w}=E-\alpha w$ with $\alpha=Ew$. Such evolution has precisely the property $\tilde{S}^n z\xrightarrow[n\to\pm\infty]{}z_\pm$ with $z_+=(v_+,w_+)\equiv(0,1)$ and $z_-=(v_-,w_-)\equiv(0,-1)$. With the above choices for $S_*,\tilde{S}$ the map S, in (7.2.2), is *reversible* if we define time reversal by $i(x,z)=(i'x,-z)$.

Therefore we see that the system $(M^*\times Z,S)$ is a system endowed with an attracting set and a repulsive set both hyperbolic, respectively given by $M^*\times\{z_+\}$ and $M^*\times\{z_-\}$.

The two poles $A_\pm=M^*\times\{z_\pm\}$ are transformed into each other by the symmetry i that, obviously, *is not* a symmetry for motions that develop on them.

The dynamical system is chaotic, having an attractive set $(M^*,\{z_+\})$ on which the evolution enjoys the Anosov property: in the full system (M,S) time reversal symmetry is spontaneously violated (in the above sense).

Nevertheless we see that if one defines the map $i^*A_+\longleftrightarrow A_+$ as: $i^*(x,z_+)=(i'x,z_+)$ then i^* anticommutes with the evolution S *restricted* to A_+ (and a map i^* can be defined also on A_- in a analogous way and, analogously, it anticommutes with S on A_-).

Therefore in this case i^* is a "time reversal" map defined only "locally" (*i.e.* on the poles of the system) "inherited" from the global symmetry i: the symmetry i, however, is not a local symmetry, *i.e.* it cannot be restricted to the poles, because the poles are not symmetric and they are not i–invariant.

We now examine in which cases the construction just described is generalizable.

(D) The axiom C.

First of all we give a formal description of the geometric property introduced in [BG97] and called there *axiom C* property of a dynamical system endowed with hyperbolic poles.

In the observations to definition 2 of §5.4 we noted that any axiom A dynamical system (M,S) is a system whose nonwandering points can be decomposed into a finite number of sets, called *basic sets* or *poles*, densely covered by periodic orbits and on which there exists a dense orbit (this is a theorem by Smale, *c.f.r.* §5.4).

Not all poles are attracting sets: if a system is reversible then every pole A of attraction has a time reversed "image" $i\,A$ that is a repulsive pole.

Given a pole Ω (attractive, repulsive or other) one defines $W^s(\Omega)$ as the set of the points that evolve towards Ω for $t \to +\infty$ and $W^e(\Omega)$ as the set of points that evolve towards Ω for $t \to -\infty$.

When Ω is attractive the set $W^s(\Omega)$ is the basin of attraction of Ω, while $W^e(\Omega)$ is Ω itself. In general Ω is neither attractive nor repulsive and the two sets $W^s(\Omega)$ and $W^e(\Omega)$ are both nontrivial. For simplicity we restrict the following discussion to the case in which the system has at most two poles Ω_+, Ω_-, one attractive and one repulsive.

It is convenient to define the "distance" $\delta(x)$ of a point x from the poles. If d_0 is the diameter of the phase space M and $d_\Omega(x)$ is the ordinary distance (in the metric of M) of the point x from the pole Ω

$$\delta(x) = \min_{i=\pm} \frac{d_{\Omega_i}(x)}{d_0} \tag{7.2.4}$$

We say that two manifolds intersect transversally if the plane spanned by their tangent planes at a point of intersection has dimension equal to that of the whole phase space, *c.f.r.* §5.4. The latter notion of transversality is then useful to fix the notion of "*axiom B system*" or of system that "*verifies axiom B*". It is (rephrasing here definition 4 of §5.4) a system that verifies the axiom A with the further property that if $W^s(\Omega_i)$ has a point y in common with $W^e(\Omega_j)$, hence $y \in W^s_x \cap W^e_{x'}$ for some pair $x \in \Omega_i$ and $x' \in \Omega_j$, then the intersection between W^s_x and $W^e_{x'}$ is *transversal* in y, *c.f.r.* §5.4 observation 3 to the definition 2.

The structures now described are interesting because the systems that verify them are *stable*: if a system verifies axiom B then, by perturbing the map S in class C^∞, one generates a new system that, via a *continuous* (but, in general, not differentiable and hence not necessarily regular) coordinates change can be transformed into the original one.

The latter is a deep result (*Robbin theorem*), [Ru89b] p. 170. The converse statement is a conjecture (conjecture of Palis–Smale); "in class C^r" for $r \ge 1$ and for $r = 1$ it is already a theorem (*Mañé theorem*), *c.f.r.* [Ru89b], p. 171 for a precise formulation: see, also, the comments to definition 4 in §5.4.

The example given at point (C) obviously verifies the axioms A and B. It verifies furthermore the property that in [BG97] has been called axiom C

Definition *(axiom C): A dynamical system $(\mathcal{C}, S)$ verifies axiom C if it is a mixing Anosov system or at least it verifies axiom B and if in the latter case*

(i) It admits only one attractive pole and only one repulsive pole, A_+ and A_-, with basins of attraction for A_+ and of repulsion for A_- open and with complement with zero volume, (globality property of the attracting and of repulsive set). The poles are, furthermore, regular surfaces *on which S acts in topologically mixing way (hence $(A_\pm, S)$ are mixing Anosov systems).*

(ii) For every $x \in M$, the tangent plane T_x admits a Hölder–continuous decomposition as a sum of three planes T_x^u, T_x^s, T_x^m such that[6]

$$\begin{array}{llll} a) & dS\,T_x^\alpha &= T_{Sx}^\alpha & \alpha = u,s,m \\ b) & |dS^n w| &\le Ce^{-\lambda n}|w|, & w \in T_x^s,\ n \ge 0 \\ c) & |dS^{-n} w| &\le Ce^{-\lambda n}|w|, & w \in T_x^u,\ n \ge 0 \\ d) & |dS^n w| &\le C\delta(x)^{-1}e^{-\lambda|n|}|w|, & w \in T_x^m,\ \forall n \end{array}$$

where the dimensions of T_x^u, T_x^s, T_x^m are > 0 and $\delta(x)$ is defined in (7.2.4).
(iii) if x is on the attractive pole A_+ then $T_x^s \oplus T_x^m$ is the tangent plane to the stable manifold in x; viceversa if x is on the repulsive pole A_- then $T_x^u \oplus T_x^m$ is the tangent plane to the unstable manifold in x.

Remarks:

(1) Although T_x^u and T_x^s are not uniquely determined for general x's the planes $T_x^s \oplus T_x^m$ and $T_x^u \oplus T_x^m$ are uniquely determined for all $x \in A_+$ and, respectively, for all $x \in A_-$.
(2) It is clear that an axiom C system verifies also, necessarily, axiom B. The possibility that every axiom B reversible system that has only two poles, one attractive and one repulsive, verifies necessarily axiom C is not remote, possibly with the help of some additional ("natural") hypothesis.
(3) (3) If we drop the condition that the poles are smooth and just require that motion on them is topologically mixing then the resulting weaker notion is structurally stable: this is implied by the quoted theorem by Robbin, [BG97].
(4) The hypothesis that there are two poles is posed here only for simplicity and probably one could dispense of it, [BG97].
(5) Also the requirement that the poles be regular surfaces, is probably not always necessary for the purposes of the discussions that follow.

It is possible to visualize the axiom C property via the following figure (commented below)

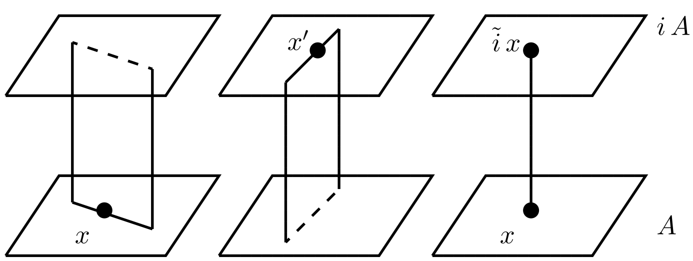


Fig. (7.2.1): *Illustration of the example, preceding the definition, of axiom C system. The vertical direction represents z on which the map $\tilde{S}$ of the example acts; the horizontal plane is a symbolic representation of the torus on which S_* acts; see (7.2.2). See below for a detailed interpretation of the figure.*

[6] One could prefer C^∞ or C^p regularity, with $1 \le p \le \infty$: but this would exclude most cases. On the other hand Hölder continuity could be equivalent to simple continuity C^0, as in the case of Anosov systems, *c.f.r.* [AA68], [Sm67].

Axiom C property, [BG97], is stronger than axiom B property and it is a *structurally stable* notion, *i.e.* it remains a true property also if the systems obtained with small enough perturbations from an axiom C system still verify axiom C.

Informally, in systems endowed only with an attracting and a repelling set, without other invariant sets, axiom C says that the stable manifold of the points on A_+ arrives "*transversally*" on the repelling set A_-, rather than "*wrapping around it*" and, furthermore, $A_\pm$ are smooth surfaces.

The first figure in Fig.(7.2.1) illustrates a point $x \in A_+$ and a local part of its stable manifold that extends until the set A_- intersecting it in the hatched line (that is a stable manifold for the motion *restricted* to the surface A_- but that *is not part* of the stable manifold of x). Likewise the second figure describes a point x' on A_- with a local part of its unstable manifold.

The third figure in Fig.(7.2.1) shows the intersection between the stable manifold of a point $x \in A_+$ and the unstable manifold of the point $\tilde{x} \in A_-$: in the figure such intersection is a unidimensional curve that connects x with $\tilde{x}$ (that is uniquely determined by x, *c.f.r.* following) *establishing the correspondence defining* $\tilde{i}$.

In this case the stable manifold of x is the sum $T^s_x \oplus T^m_x$ if T^m_x is the vertical direction, while the unstable manifold of x' is $T^u_x \oplus T^m_x$. Here T^s_x and $T^u_{x'}$ are parallel to the stable and unstable manifolds (the solid horizontal lines) of the map S_*. The intersection of the two manifolds is a line T^m_x in the vertical direction.

The points "between the two surfaces" $A_\pm$ represent most of the points of phase space, but they are *wandering points*, *c.f.r.* §5.4.

Let $\delta_\pm$ the dimension of the surfaces $A_\pm$ and let $u_\pm, s_\pm$ the dimension of the stable and unstable manifolds of the dynamical systems $(A_\pm, S)$ respectively ($\delta_\pm = u_\pm + s_\pm$). It is $s_+ = u_-, u_+ = s_-$ and hence $\delta_+ = \delta_- = \delta$ and the total dimension of phase space is $\delta + m$ with $m > 0$.

The dimension of the stable manifold of $x \in A_+$ in the original dynamical system (*i.e.* describing dynamics not restricted to A_+) is, therefore, $m+s_+$, because such manifold "sticks outside of A_+" (because A_+ attracts) and that of the unstable manifold of $x' \in A_-$ is $m+u_-$. Hence the dimension of their intersections is m and such surface intersects A_+ and A_- in *two points* that we can call x and $\tilde{x}$, thus allowing us to define $\tilde{i}x = \tilde{x}$. We see that such surface is a *wire* that joins points of A_+ to points of A_- (defining $\tilde{i}$): hence the representation in Fig.(7.2.1) is an accurate representation, in spite of it being schematic.

Axiom C, that forbids the stable manifold of the points of A_+ and the unstable manifold of the points of A_- to "wrap" around A_- or, respectively, around A_+, can be seen as a hypothesis of maximum simplicity on the geometry of the system. The interest of this geometric notion lies in the

Theorem *(axiom C and time reversal stability): Let the dynamical system* (M, S) *be reversible and verifying axiom C. Then there exists a map* i^*

defined on the poles of the system that leaves them invariant, squares to the identity, and that anticommutes with the time evolution.

In the case illustrated in figure the i^* is indeed the composition $i\cdot\tilde{i}$, [BG97]. Hence this theorem shows that in systems verifying axiom C time reversal symmetry is *unbreakable*: if its spontaneous breaking occurs as some parameters of the system vary (with appearance of attractive sets smaller than the whole phase space, and not invariant for the global time reversal i) the attractive pole always admits a symmetry i^* that inverts time, *i.e.* that anticommutes with time evolution. The unsatisfactory aspect of the above analysis is that in general a perturbation of a system verifying Axiom C will have all the properties of the unperturbed system with the possible exception that the poles might lose the property of being smooth surfaces and a better understanding of this is desirable.

(E) The chaotic hypothesis.

The structural stability properties of axioms A, B and C systems has been one of the reasons of the following reinterpretation and extension of Ruelle' principle of §5.7, called *chaoticity principle* or the *chaotic hypothesis*. It has not been, however, the main reason because the chaoticity principle has been reality "derived" on the basis of the interpretation of experimental results, [ECM93], [GC95].

Chaotic Hypothesis: *A mechanical system, be it a particles system or a fluid, in a chaotic stationary state* [7] *can be considered as a mixing Anosov system for the purpose of the computation of the macroscopic properties. In the case of reversible systems a map i^* of the attractive sets into themselves exists which anticommutes with the evolution and squares to the identity.*

In the case the system has an attracting set A which is smaller than phase space the hypothesis has to be interpreted as saying that the attracting set A is a smooth surface and the time evolution map S is such that (A,S) is a mixing Anosov system which enjoys a time reversal symmetry: both assumptions are true if the system verifies axiom C, by the theorem in subsection (D) above. Therefore it is not restrictive to suppose that the attracting set is the entire phase space and that on it a time reversal symmetry exists.

This principle, together with the considerations developed at the point (A), will allow us to propose rather detailed properties of the Navier–Stokes equation, in the §7.4.

We conclude with some comments on the meaning of the chaotic hypothesis. The hypothesis has to be interpreted in the same way one interprets

[7] We shall understand by *chaotic* any stationary distribution with *at least* one positive Lyapunov exponent which is not close to zero: this is necessary, like in equilibrium, because near a non chaotic system one may have phenomena reminiscent of those that arise in equilibrium theory with systems that are close to integrable ones.

the ergodic hypothesis in statistical mechanics. *One must not intend that a system of physical interest "really" verifies axiom C*, rather one must intend that this property holds *only for the purpose* of the computation of the average values of the few interesting observables with respect to the distribution SRB, *i.e.* the average with respect to the statistics of motions that follow initial data randomly chosen in phase space with some distribution proportional to volume, *c.f.r.* §5.7,.

We must also remark that this principle is stronger than the ergodic hypothesis: indeed it applies to non equilibrium systems and one can show, [Si94], that if the dynamical system is Hamiltonian and if it is a mixing Anosov system, then the SRB distribution is precisely the *Liouville distribution* on the constant energy surface: hence the ergodic hypothesis holds.

The ergodic hypothesis implies classical thermodynamics, even when applied to systems that manifestly are not ergodic, like the perfect gas. Likewise one has to understand that chaotic hypothesis *cannot be generally true*, strictly speaking, for many systems of interest for physics: sometimes because of the trivial reason that the evolution of these systems is described by maps S that *are not regular everywhere* but only piecewise so, *c.f.r.* Definition 5 in §5.4.

The idea is that the chaotic hypothesis could allow us to establish *relations* between physical quantities without really computing the value of any of them. In the same way as Boltzmann deduced the heat theorem (*i.e.* the equality of the derivatives of the functions of state expressing that the differential form $(dU + pdV)/T$ is exact) from a formal expression for the equilibrium distribution of a gas.

Now the role of the formal expressions for the Gibbs distributions in equilibrium statistical mechanics will be plaid, for the SRB distributions, by the formula of Sinai (7.2.1).

Bibliography: [Ru79], [Si94], [Ga81], [Ga95], [GC95].

In the original work [GC95] the chaotic hypothesis has been formulated by requiring that the dynamical system be a mixing Anosov system (that in the notations of the work was, somewhat improperly with respect to the current terminology, called transitive). As explained in [BGG97] this statement has to be interpreted, to be in agreement with experiments relative to situations *strongly* outside of equilibrium, in the sense that the system is of Anosov mixing type if *restricted* to the attracting set (*i.e.* the attracting set must be a regular surface on the which S acts in a mixing way). But in this last case it also became necessary to add the hypothesis that the time reversal was "unbreakable". The search of a geometric condition that guaranteed *a priori* the unbreakability of time reversal and that was *a priori* stable led in [BG97] to formulate axiom C.

The end point of this chain of refinements, and in a certain sense, of simplifications of the original hypothesis, *still requires* in order to "be completely satisfactory" the elimination of the hypothesis that the poles are regular manifold.

§7.3 Chaotic hypothesis, fluctuation theorem and Onsager reciprocity, entropy driven intermittency.

We consider now a rather general dynamical system: however we keep in mind the reversible NS equations introduced in §7.1 as the example to which we would like to apply the following ideas. We study motions on a regular surface V described by an equation

$$\dot{\underline{x}} = \underline{f}(\underline{x}) \tag{7.3.1}$$

whose solutions $t \to \underline{x}(t) = S_t\underline{x}$, with initial datum $\underline{x}$ admit a time reversal symmetry i. This means, see §7.1, that S_t anticommutes with the isometric operation[1] i: $iS_t = S_{-t}i$, and $i^2 = 1$.

First of all we look at motions through timed observations. This means that we imagine that our system is observed at discrete times, *c.f.r.* §5.2. Although not really necessary, this simplifies a little the discussion and reduces by one unit the dimension of phase space and allows us to consider the evolution as a map S. The phase space $M \subset V$ on which S acts can be considered as a piecewise regular surface, possibly made of various connected parts and everywhere transversal to the trajectories of the solutions of (7.3.1), see [Ge98].

The map S is related to the flow S_t by the relation $Sx = S_{t(x)}x$ if $t(x)$ is the time that elapses between the "timing event" $x \in M$ and the successive one.

The dynamical system that we study will then be (M, S) and we shall call, as usual, μ_0 a probability distribution endowed with a density with respect to the volume measure on M. Time reversal invariance becomes $iS = S^{-1}i$, with i isometry of M and $i^2 = 1$: if the continuous time evolution S_t is time reversal symmetric then the timed observations will also be such for a suitably defined i.[2]

(A) The fluctuation theorem.

On the basis of the chaotic hypothesis we imagine that the attractive set for the evolution S is the whole phase space M, without loss of generality.

---

[1] Supposing, under our hypotheses, that i is an isometry is not restrictive: it suffices to redefine suitably the metric so that the distance $d(x, y)$ between two close points becomes $(d(x, y) + d(ix, iy))/2$.

[2] For instance if the system is a "billiard" and the observations are timed at the collisions with the obstacles then a possible time reversal maps a collision c into the new collision ic obtained by considering the result c' of the collision c, which is no longer a collision "being a vector that comes out of the obstacle", and changing the sign of the velocity obtaining again a collision, *i.e.* "a vector entering the obstacle", that defines ic.

The distribution SRB shall have the form (7.2.3): the important point is that the combination of (7.2.3) with the time reversal symmetry is rich of consequences, surprising at least at first sight.

Note that we can suppose without loss of generality that the Markovian pavement $\mathcal{P}$, used to represent the SRB distribution via Sinai's formula, *c.f.r.* (7.2.1), could be chosen *invariant under time reversal*: *i.e.* such that if $\mathcal{P} = (P_1, \ldots, P_n)$ then iP_σ is still an element $P_{i\sigma}$ of $\mathcal{P}$ with $i\sigma = \sigma'$ suitable. In fact from the definition of Markovian pavement, *c.f.r.* §5.7 (C), it follows that

(1) by intersecting two Markovian pavements $\mathcal{P}$ and $\mathcal{P}'$ we obtain a third Markovian pavement: *i.e.* the pavement whose elements are the sets $P_i \cap P'_j$ is still Markovian.
(2) by applying the map i to the elements of a Markovian pavement $\mathcal{P}$ we obtain a Markovian pavement $i\mathcal{P}$: this is so because time reversal i transforms the stable manifold $W^s(x)$ and the unstable manifold $W^e(x)$ into $W^e(ix)$ and $W^s(ix)$, respectively.

Hence intersecting $\mathcal{P}$ and $i\mathcal{P}$ we obtain a time reversal symmetric pavement. If $iP_\sigma \stackrel{def}{=} P_{i\sigma}$ is the correspondence between elements of the pavement established by the action of i we see that i is therefore represented as the map that acts on the sequence of symbols $\underline{\sigma} = \{\sigma_k\}$ by transforming it into $\underline{\sigma}' = \{\sigma'_k\}$ with $\sigma'_k = i\,\sigma_{-k}$.

Furthermore it is not difficult to verify that this implies that a standard extension of the compatible strings $\sigma_{-\tau/2}, \ldots, \sigma_{\tau/2}$, *c.f.r.* (7.2.1) and §5.7, can be performed so that *if x_j is the center of $E_j = \cap_{k=-\tau/2}^{\tau/2} S^{-k} P_{\sigma_k}$ then ix_j is the center of iE_j.*

Let us denote by $J_\tau(x)$ the Jacobian matrix of the map S^τ, where τ is an even integer, as a map of $S^{-\tau/2}x$ to $S^{\tau/2}x$; and denote with $J_{e,\tau}(x)$ and $J_{s,\tau}(x)$ the Jacobian matrices of the same maps thought of as maps of $W^e_{S^{-\tau/2}x}$ to $W^e_{S^{\tau/2}x}$ or, respectively, of $W^s_{S^{-\tau/2}x}$ to $W^s_{S^{\tau/2}x}$. Then one can establish simple relations between the determinants of these matrices.

If $\alpha(x)$ is the angle formed, in x, between the stable and the unstable manifolds[3] and if we denote, respectively, $\Lambda_\tau(x) = |\det J_\tau(x)|$, $\Lambda_{s,\tau}(x) = |\det J_{s,\tau}(x)|$, $\Lambda_{e,\tau}(x) = |\det J_{e,\tau}(x)|$ then, noting that such determinants are related to the expansion or contraction of the elements of surface of the manifolds M, W^s_x and, respectively, W^e_x, it follows that

$$\Lambda_\tau(x) = \Lambda_{s,\tau}(x)\Lambda_{e,\tau}(x)\frac{\sin\alpha(S^{\tau/2}x)}{\sin\alpha(S^{-\tau/2}x)} \tag{7.3.2}$$

[3] The angle between two planes that have in common only one point can be defined as the minimum angle between non zero vectors lying on the two planes attached to the common point. The angle between two manifolds that locally have only one point in common is defined as the angle between their tangent planes.

Time reversal symmetry (and its isometric character) implies that

$$\Lambda_\tau(ix) = \Lambda_\tau(x)^{-1}, \qquad \Lambda_{e,\tau}(ix) = \Lambda^{-1}_{s,\tau}(x), \qquad \Lambda_{s,\tau}(ix) = \Lambda^{-1}_{e,\tau}(x) \quad (7.3.3)$$

and if $\lambda_\tau(x) = -\log \Lambda_\tau(x)$, $\lambda_{e,\tau}(x) = -\log \Lambda_{e,\tau}(x)$, $\lambda_{s,\tau}(x) = -\log \Lambda_{s,\tau}(x)$, therefore

$$\lambda_\tau(ix) = -\lambda_\tau(x), \quad \lambda_{e,\tau}(x) = -\lambda_{s,\tau}(ix), \quad \lambda_{s,\tau}(x) = -\lambda_{e,\tau}(ix) \quad (7.3.4)$$

note that the quantity $\lambda_\tau(x)$ is simply related to the divergence $\delta(x)$ of the equation (7.3.1), $\delta(x) = -\sum_j \partial_j f_j(x)$. If, as above, $t(x)$ denotes the time interval between the timed observation producing the result x and the next one it is $\lambda(x) = \int_0^{t(x)} \delta(S_t x) dt$. Then

$$\lambda_\tau(x) = \sum_{j=-\frac12\tau}^{\frac12\tau-1} \lambda(S^j x) \qquad (7.3.5)$$

We shall call *entropy creation* on τ timing events the contraction of the volume of phase space (that could *also be negative*, *i.e.* in fact an expansion), which is the quantity

$$\sigma_\tau(x) = \sum_{j=-\frac12\tau}^{\frac12\tau-1} \lambda(S^j x) \stackrel{def}{=} \tau\,\langle\lambda\rangle_+ p \qquad \text{if } \langle\lambda\rangle_+ \neq 0 \qquad (7.3.6)$$

where $\langle\lambda\rangle_+$ denotes the average value of the function $\lambda(x)$ with respect to the SRB distribution of the system (M, S), *c.f.r.* §7.2, §5.7 and p is a variable (that depends on τ and x) on the phase space M and that we can call the (adimensional) average *rate of creation of entropy* on τ events around x.

Whether the name of *entropy*, [An82] and Sec. 9.7 in [Ga99a], is properly used here, or not, is debatable. In reality we are interested in cases in which the quantity $\langle\lambda\rangle_+$ is not zero: such cases will be called *dissipative*. In this respect one should note a theorem that says that if such average is not zero then it is necessarily positive and this is a property that, without doubt, is certainly desired from a definition of rate of entropy creation, [Ru96].

Here we cannot invoke, to justify the use of the name "entropy", independent definitions of such notion: simply because the notion of entropy *has never been well defined* in cases of systems outside of equilibrium.

We shall adopt this name also because we shall see that this quantity has various other desirable properties that help making the notion a satisfactory one, see [An82],

The first important property is that the dimensionless rate of entropy creation p, *c.f.r.* (7.3.6), is a variable that has a probability distribution $\pi_\tau(p)$ with respect to the stationary statistics SRB that describes the asymptotic properties of motions. By definition $\langle p\rangle_+ \equiv 1$ (in the dissipative cases, of course).

We shall set:

$$\zeta(p) = -\lim_{\tau\to\infty} \frac{1}{\tau} \log \pi_\tau(p) \tag{7.3.7}$$

where the existence of the function $\zeta(p)$ (in the case of a mixing Anosov system, *i.e.* if the chaotic hypothesis holds) is proved by a theorem of Sinai, [Si72], [Si77]. Then the function $\zeta(p)$ verifies, [GC95a], [GC95b],the following fluctuation theorem

I Theorem *(fluctuation theorem): The "rate function" $\zeta(p) \ge -\infty$ has odd part verifying*

$$\zeta(-p) = \zeta(p) - \langle\lambda\rangle_+\, p \tag{7.3.8}$$

for all p.

An illustration is provided by the Fig. (7.3.1).

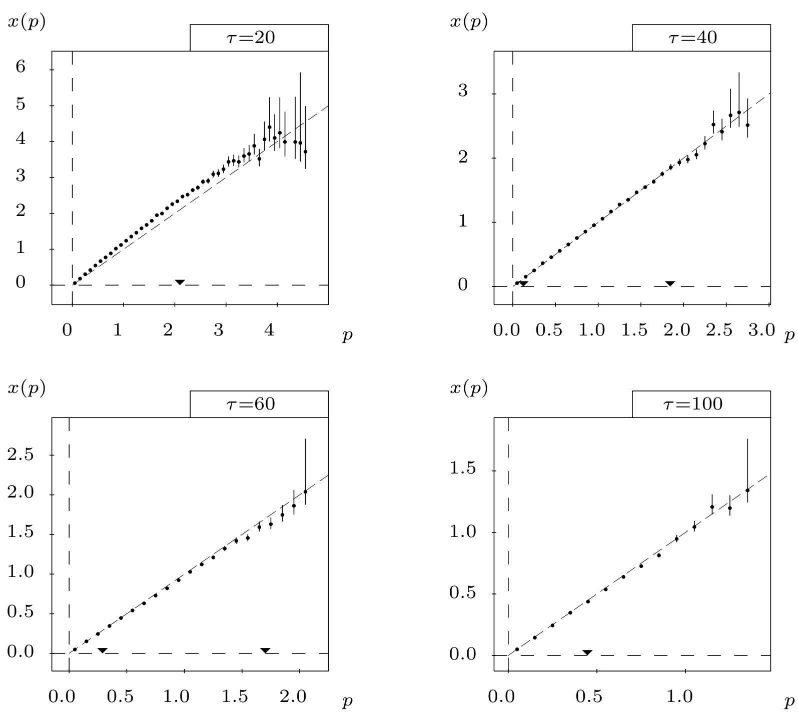


Fig. (7.3.1) *Illustration of the fluctuation theorem of §7.3 which gives $x(p) = \frac{\zeta(-p)-\zeta(p)}{\langle\sigma\rangle_+} = p$ in the limit $\tau\to\infty$, for an electrical conduction model in a very strong electromotive field, taken from the experiment in [BGG97]. The dashed graph is $x(p) = p$ while the four graphs correspond to the choices $\tau = 20, 40, 80, 100$.*

The proof is simple: informally it is the following. One can compute the ratio $\pi_\tau(p)/\pi_\tau(-p)$ by (7.2.1)

$$\frac{\pi_\tau(p)}{\pi_\tau(-p)}=\frac{\sum_{\sigma_\tau(x_j)=\langle\lambda\rangle_+\tau p}\Lambda_{e,\tau}(x_j)^{-1}}{\sum_{\sigma_\tau(x_j)=-\langle\lambda\rangle_+\tau p}\Lambda_{e,\tau}(x_j)^{-1}} \tag{7.3.9}$$

where the sum runs over the centers x_j of the partition elements $\mathcal{E}_\tau = \cap_{-\tau/2}^{\tau/2} S^h\mathcal{E}$, *c.f.r.* (5.7.8) (*i.e.* on the points whose history, with respect to the compatibility matrix of the Markovian pavement, is a compatible sequence $\sigma_{-\tau/2},\ldots,\sigma_{\tau/2}$ extended in "standard way" to an infinite compatible sequence, *c.f.r.* (7.2.1), (5.7.8)).

In the discussion that follows *we do not take into account* that (7.3.9) is not correct and that the correct formula, *c.f.r.* (5.7.8), should be

$$\frac{\pi_\tau(p)}{\pi_\tau(-p)}=\lim_{T\to\infty}\frac{\sum_{\sigma_\tau(x_j)=\langle\lambda\rangle_+\tau p}\Lambda_{e,T}(x_j)^{-1}}{\sum_{\sigma_\tau(x_j)=-\langle\lambda\rangle_+\tau p}\Lambda_{e,T}(x_j)^{-1}} \tag{7.3.10}$$

where the sum runs on the elements of the partition $\cap_{-T/2}^{T/2} S^{-k}\mathcal{E}$, *i.e.* on the sets $E_j = E_{\sigma_{-T/2},\ldots,\sigma_{T/2}}$. In other words we should first let T to ∞ and then $\tau\to\infty$. *The eq. (7.3.9), instead, considers* $T=\tau$.

Evidently by using (7.3.9) instead of the correct (7.3.10) one commits *errors* that we could fear to be unrepairable. But it is not so and the error can be bounded, *c.f.r.* [GC95a],[GC95b], and one can show (easily) that the correct ratio between $\pi_\tau(p)$ and $\pi_\tau(-p)$ is bounded from above and from below by the r.h.s. of equation (7.3.9) respectively *multiplied or divided by a factor a which is* τ*–independent*. Since we are only interested in the limit (7.3.7) we see that such an error has no influence on the result.

The possibility of this bound is obviously *essential* for the discussion: it is actually easy, but it rests on the deep structure of symbolic dynamics and on well known properties of probability distributions on spaces of compatible sequences, the reader is referred to [Ga95a] or [Ru99c].

Accepting (7.3.9) one remarks that the sum in the denominator can be rewritten by making use of the fact that if x_j is the center of E_j and it has adimensional rate of entropy creation p, then ix_j is center of iE_j (*c.f.r.* observations at the beginning of the section) and by time reversal symmetry, see (7.3.3), rate $-p$: hence the (7.3.9) is rewritten as

$$\frac{\pi_\tau(p)}{\pi_\tau(-p)}=\frac{\sum_{\sigma_\tau(x_j)=\langle\lambda\rangle_+\tau p}\Lambda_{e,\tau}(x_j)^{-1}}{\sum_{\sigma_\tau(x_j)=\langle\lambda\rangle_+\tau p}\Lambda_{e,\tau}(ix_j)^{-1}} \tag{7.3.11}$$

We can now remark that (7.3.3) allows us to rewrite this identity as

$$\frac{\pi_\tau(p)}{\pi_\tau(-p)}=\frac{\sum_{\sigma_\tau(x_j)=\langle\lambda\rangle_+\tau p}\Lambda_{e,\tau}(x_j)^{-1}}{\sum_{\sigma_\tau(x_j)=\langle\lambda\rangle_+\tau p}\Lambda_{s,\tau}(x_j)} \tag{7.3.12}$$

so that numerator and denominator are sums over an equal number of terms. We also see that corresponding terms (*i.e.* terms with the same label j) have ratio $\Lambda_{e,\tau}(x_j)^{-1}\Lambda_{s,\tau}(x_j)^{-1}$ *and this ratio is equal to* $\Lambda_\tau(x_j)$ apart from the factors that come from the ratios of the sines of the angles, *c.f.r.* (7.3.2), $\alpha(S^{-\tau/2}x_j)$ and $\alpha(S^{\tau/2}x_j)$: such ratios will however be bounded from below and from above by a^{-1} and a, for a suitable a (because the angles $\alpha(x)$ are bounded away from 0 and π by the assumed hyperbolicity of A, *c.f.r.* §5.4, and furthermore by taking the products of the fractions in (7.3.2) all sines simplify "telescopically" and one is left only with the first numerator and the last denominator).

We note that what said is rigorously correct only if the system (M, S), restricted to the attracting set, is really a mixing Anosov system (or just only transitive, *c.f.r.* §5.4).[4] We make use of the chaotic hypothesis when we suppose that the properties used are "in practice" true at least for the purposes of computing quantities of interest, like precisely $\zeta(p)$.

By definition of p it is $\Lambda_\tau^{-1}(x_j) = e^{p\,\tau\,\langle\lambda\rangle_+}$, for all choices of j, so that

$$\frac{1}{a}e^{-p\tau\langle\lambda\rangle_+} < \frac{\pi_\tau(p)}{\pi_\tau(-p)} < a e^{-p\tau\langle\lambda\rangle_+} \tag{7.3.13}$$

and (7.3.8) follows in the limit $\tau\to\infty$.

More generally let $\kappa_1(x),\ldots,\kappa_n(x)$ are n functions on phase space such that

$$\kappa_j(ix) = -\kappa_j(x) \tag{7.3.14}$$

i.e. they are *odd* under time reversal, define

$$\kappa_{\tau,j}(x) \stackrel{def}{=} \sum_{r=-\tau/2}^{\tau/2}\kappa_j(S^r x) = q_j\,\tau\,\langle\kappa_j\rangle_+ \tag{7.3.15}$$

Consider the joint probability, with respect to the SRB statistics, of the event in which the variables $p, q_1,\ldots,q_n$ have given values, and denote such probability by $\pi_\tau(p,q_1,\ldots,q_n)$. Let

$$\zeta(p,q_1,\ldots,q_n) = -\lim_{\tau\to\infty}\frac{1}{\tau}\log\pi_\tau(p,q_1,\ldots,q_n) \tag{7.3.16}$$

then the same argument exposed above to prove theorem I implies (obviously)

II Theorem: *(extended fluctuation theorem): The large deviations functions* $\zeta(p,q_1,\ldots,q_n)$ *verify*

$$\zeta(-p,-q_1,\ldots,-q_n) = \zeta(p,q_1,\ldots,q_n) - \tau\,p\,\langle\lambda\rangle_+ \tag{7.3.17}$$

---

[4] A property used several times, when using the formula (7.2.1) for the SRB distribution and now when saying that the ratios between the sines of the angles above introduced are uniformly bounded from below and from above.

if $p, q_1, \ldots, q_n$ *is a point internal to the domain* D *in which the variables* $p, q_1, \ldots, q_n$ *can assume values.*

It is very interesting, as we shall see in the successive point (B), that the last term in (7.3.17) *does not depend on the variables* q_j.

Remark: It is important to note that the fluctuation theorem can also be formulated in terms of properties of the not–discretized system, *i.e.* in terms of the quantity $\delta(x) = -\delta(ix)$, divergence of the equations of the motion, that we call rate of entropy creation per unit of time (instead of "per timing event"). And possibly in terms of the other quantities $\gamma_j(x) = -\gamma_j(ix)$ odd with respect to time reversal. The theorems are stated in the same way provided we modify the definitions of p, q_j by replacing the sums in (7.3.15) and (7.3.6) as

$$\sigma_t = p\langle\delta\rangle_+ \int_{-t/2}^{t/2} dt'\, \delta(S_{t'}x) \qquad \kappa_{t,j} = q\,\langle\gamma\rangle_+ \int_{-t/2}^{t/2} dt'\, \gamma_j(S_{t'}x) \tag{7.3.18}$$

The extension requires a detailed analysis, *c.f.r.* [Ge98].

The above fluctuation theorems are remarkable because they can be considered *laws of large deviations* in the probabilistic sense of the term (namely they give a property of the probabilities of deviations away from the average of a sum of τ random variables and such deviations have magnitude $2p\tau$ or $2q_j\tau$: hence they have *order of magnitude much larger than the "normal size"* of fluctuations *i.e.* $\gg \sqrt{\tau}$, if p or q_j are close to their typical value ~ 1).
It is a result that can be accessible to experimental checks in many non-trivial cases: indeed the experimental observation, *c.f.r.* [ECM93], of the validity of the (7.3.8) in a special case has been the origin and the root of the development of the chaotic hypothesis and of the derivation of the above theorems. Successively is has been reproduced in several different experiments, *c.f.r.* [BGG97], [BCL98].

One should be careful to understand properly the nature of the fluctuation theorem: it involves considering non trivial limits which *cannot* be light-heartedly interchanged as discussions in the literature have amply shown, [CG99].

The interest of the relations (7.3.8) and (7.3.17) is increased because it has been noted that can they can be considered a *generalization to systems outside equilibrium* of Onsager's reciprocity relations, *c.f.r.* §1.1, and of the Green–Kubo formulae for transport coefficients, *c.f.r.* [Ga96a].

(B) Onsager's reciprocity and the chaotic hypothesis.

We shall study a typical system of N particles subjected to internal and external conservative forces, with potential $V(\underline{q}_1, \ldots, \underline{q}_N)$, and to not conservative external forces $\{\underline{F}_j\}$, $j=1,\ldots,N$, of intensity measured by parameters $\{G_j\}$, $j=1,\ldots,N$. Furthermore the system will be subject also to forces $\{\underline{\varphi}_j\}$, $j=1,\ldots,N$, that have the role of absorbing the energy given to system by the nonconservative forces and, hence, of allowing the attainment of a stationary state. Let μ_+ be the statistics of the motion of an initial datum randomly chosen with a distribution μ_0, with density with respect to volume in phase space.

The equations of motion will have, therefore, the form

$$m\underline{\dot q}_j=\underline{p}_j,\qquad \underline{\dot p}_j=-\partial_{\underline{q}_j}V(\underline{q}_j)+\underline{F}_j(\{G\})+\underline{\varphi}_j \tag{7.3.19}$$

with $m=$ mass of the particles: they will be supposed reversible for all $\underline{G}$.

If $O(\{\underline{q},\underline{\dot q}\})$ is an observable and if S_t is the map that describes the evolution, the distribution μ_+ is defined by

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T O(S_tx)\,dt=\int_M O(y)\mu_+(dy)\stackrel{def}{=}\langle O\rangle_+ \tag{7.3.20}$$

for all $x\in M$ except a set of zero μ_0–volume on M.

We shall suppose also that the rate of entropy generation $\delta(x)$, *c.f.r.* (7.3.5), which however we shall continue to denote $\sigma(x)$ (but recall that it now represents a rate of entropy generation per unit time and not per timing event) has the form:

$$\sigma(x)=\sum_{i=1}^{s}G_iJ_i^0(x)+O(G^2) \tag{7.3.21}$$

an assumption that, in fact, only sets the restriction that in absence of not conservative forces it is $\sigma=0$.

Following Onsager, one defines the *thermodynamic current* associated with the force G_i as $J_i(x)=\partial_{G_i}\sigma(x)$. Onsager's relations concern the *transport coefficients* defined by

$$L_{ij}=\partial_{G_i}\langle J_j\rangle_+\big|_{\underline{G}=0} \tag{7.3.22}$$

and establish the *symmetry* of the matrix L.

We want to show that the fluctuation theorem (7.3.8), (7.3.17) can be considered an extension to nonzero values $\underline{G}$ of the external forces (*"thermodynamic forces"*) of the reciprocity relations.

This will be obtained by computing $\zeta(p),\zeta(p,q_1,\ldots)$ for $\underline{G}$ small, up to infinitesimals of order larger than $O(|\underline{G}|^3)$ (and it will result that $\zeta(p)$ is an infinitesimal of second order in $\underline{G}$, so that in the computation infinitesimals of the third order will be neglected). The expression obtained will be compared with the fluctuation theorem and the *relations of Onsager* will follow, together with the *formulae of Green–Kubo* (also called at times *fluctuation dissipation theorem*) that imply them.

Therefore we shall say that the fluctuation theorem is a proper extension to nonzero fields of Onsager's reciprocity relations, because indeed it is valid without the conditions $\underline{G} = \underline{0}$ characteristic of the classical Onsager relations and *exactly* (if one supposes the chaotic hypothesis): a property that is also characteristic of Onsager's relations.

For simplicity we shall refer to the continuous time version of the fluctuation theorem described in connection with equation (7.3.18)..

The proof of the above statements is put in the appendix being of rather technical nature.

(C) A fluidodynamic application.

As an application we can consider the equation

$$\dot{\underline{u}} + (\underset{\sim}{u} \cdot \underset{\sim}{\partial}\, \underline{u})^{(\kappa)} = -\underline{\partial}\, p \tag{7.3.23}$$

with $\underline{u} = \sum_{\kappa<|\underline{k}|<2\kappa} \underline{u}_{\underline{k}} e^{i\underline{k}\cdot\underline{x}}$, $\underline{u}_{\underline{k}} = \overline{\underline{u}}_{-\underline{k}}$, $\underline{k}\cdot\underline{u}_{\underline{k}} = 0$; and with $f^{(\kappa)}$ we denote the truncation of the Fourier series of the function f to the modes $\underline{k}$ such that $\kappa < |\underline{k}| < 2\kappa$. We can interpret this equation as an equation describing the motion of a "single inertial shell" of Fourier modes in the Navier Stokes equation in the sense of §6.2,§6.3.

Let us suppose also that the energy $E = L^3 \sum_{\kappa<|\underline{k}|<2\kappa} |\underline{u}_{\underline{k}}|^2$ (conserved by the dynamics of (7.3.23)) is $E = C\kappa^{-2/3}$: *i.e.* is given by the energy content, in the Kolmogorov distribution, of the shell of momenta in $(\kappa, 2\kappa)$: *c.f.r.* (6.2.8) with C correspondent to a given value of ε.

We now ask which is the response of the system to the switching on of of an infinitesimal force $\underline{g}_{\underline{k}}$ acting on the mode $\underline{k}$, while the system is kept at *constant energy* E by means of a force defined by the Gauss' principle, *i.e.* assuming that the system is governed (in presence of forces) by an equation:

$$\dot{\underline{u}} + (\underset{\sim}{u} \cdot \underset{\sim}{\partial}\, \underline{u})^{(\kappa)} = -\underline{\partial} p - \kappa^2 \alpha \underline{u} + \underline{g}_{\underline{k}} \tag{7.3.24}$$

with

$$\alpha = \frac{1}{\kappa^2} \frac{\sum_{\underline{k}} \overline{\underline{u}}_{\underline{k}} \cdot \underline{g}_{\underline{k}}}{\sum_{\underline{k}} |\underline{u}_{\underline{k}}|^2} \tag{7.3.25}$$

This equation is reversible and is forced by the external force $\underline{g}$. The entropy production in this equation is zero if $\underline{g} = \underline{0}$ and hence we are in the situation of the point (B).

It follows that in this regime we shall have

$$L_{\underline{k},\beta;\underline{k}',\beta'} = \partial_{g_{\underline{k},\beta}} \langle \gamma_{\underline{k}',\beta'} \rangle_+ |_{\underline{g}=\underline{0}} = L_{\underline{k}',\beta';\underline{k},\beta} \tag{7.3.26}$$

because from the (7.3.25) we see that $\partial_{g_{\underline{k},\beta}} \sigma = \gamma_{\underline{k},\beta}$.

This also shows that if the conjecture of §7.1 on the equivalence of the statistical ensembles could be interpreted in a "suitably wide" sense one could perhaps deduce reciprocity relations for Navier–Stokes fluids in a regime of developed turbulence, *c.f.r.* [Ga97]. And it appears even possible that such relations could be experimentally checked both in real and numerical experiments. But setting these predictions in a mathematically and physically cleaner form, susceptible of checks, requires further analysis and ideas. I shall try to expose them in the §7.4.

(D) Physical interpretation of the fluctuation relations. Onsager–Machlup fluctuations. Entropy driven intermittency.

An important question is, naturally, "which is the physical interpretation" of the fluctuation theorem?

A simple extension of the theorem holds under the same hypotheses (*i.e.* chaotic hypothesis and time reversibility). It can be regarded as an extension of the Onsager–Machlup theory of fluctuation patterns, [OM53]. Let F, G be observables that, for simplicity, we suppose odd under time reversal, *i.e.* such that: $F(ix) = -F(x), G(ix) = -G(x)$; and let $h, k : [-T/2, T/2] \to R^1$ be two real valued functions or "*patterns*".

We call $h'(t) = -h(-t)$, $k'(t) = -k(-t)$ the "*time–reversed patterns*" or "*antipatterns*" of the patterns h, k. If $F(S_t x) = h(t)$ for $t \in [-T/2, T/2]$ we say that F follows the pattern h around the reference point x in the time interval $[-T/2, T/2] \stackrel{def}{=} W_T$. Then the following theorem can be proved in the same way as the above theorems I, II, see [Ga99c],

Theorem III *(extension of Onsager–Machlup fluctuations theory): The probabilities of the patterns h, k conditioned to a T–average dimensionless entropy production p, see (7.3.6), denoted $\pi\big(F(S_t\cdot)\big) = h(t),\, t \in W_T \,\big|\, p)$ and $\pi\big(G(S_t\cdot)\big) = k(t),\, t \in W_T \,\big|\, p)$ respectively, verify*

$$\frac{\pi\big(F(S_t\cdot) = h(t),\, t \in W_T \,\big|\, p\big)}{\pi\big(F(S_t\cdot) = -h(-t),\, t \in W_T \,\big|\, -p\big)} \;=\; 1 \tag{7.3.27}$$

and (consequently)

$$\frac{\pi\big(F(S_t\cdot) = h(t),\, t \in W_T \,\big|\, p\big)}{\pi\big(G(S_t\cdot) = k(t),\, t \in W_T \,\big|\, p\big)} = \frac{\pi\big(F(S_t\cdot) = -h(-t),\, t \in W_T \,\big|\, -p\big)}{\pi\big(G(S_t\cdot) = -k(-t),\, t \in W_T \,\big|\, -p\big)}$$

Hence relative probabilities of patterns observed in a time interval of size T and in presence of an average entropy production p are the same as those of the corresponding antipatterns in presence of the opposite average entropy production rate.

Remark: an equivalent way to write (7.3.27) is

$$\frac{\pi\big(F(S_t\cdot) = h(t),\, t \in W_T,\, p\big)}{\pi\big(F(S_t\cdot) = -h(-t),\, t \in W_T,\, -p\big)} \;=\; e^{-p\langle\lambda\rangle_+ t} \tag{7.3.28}$$

expressing the same relation as (7.3.27) in terms of joint probabilities rather than probabilities conditined to the event that in the time interval it is $\sigma_t = \langle\lambda\rangle_+ \, p\, T$

In other words it "suffices" to change the sign of the entropy production to reverse the arrow of time. In a reversible system the quantity $\zeta(p)$ measures the degree of irreversibility of a motion observed to have the value p of dimensionless entropy creation rate during an observation time of size T: if we observe patterns over time intervals of size T then the fraction of such intervals in which we shall see an entropy production p rather than 1 (which is the most probable value) will be proportional to

$$e^{(\zeta(p)-\zeta(1))T} \tag{7.3.29}$$

This tells us that normally we shall see an entropy production $p = 1$ but occasionally, with a frequency in time proportional to

$$e^{-\langle\lambda\rangle_+ \, T} \tag{7.3.30}$$

an entropy production $p = -1$ will be seen, *c.f.r.* (7.3.8): and it will be accompanied by very unexpected behavior of the time evolution of most observables, *c.f.r.* (7.3.28).

This is an *entropy driven intermittency* phenomenon. "Intermittency" seems to be a notion defined on a case by case basis (we have met it in a sense similar to the present in §4.3 and also in a somewhat different sense in §6.2) and we interpret here as a the phenomenon of rare, randomly spaced, interval of time during which some observables behave in avery different way with respect to their average.

In applying the above analysis to a fluid other difficulties arise due to the fact that we expect that $\langle\lambda\rangle_+$ becomes very large when the forcing, hence the Reynolds number, becomes large so that such strong fluctuations become practically unobservable.

However one can regard continua as composed by small macroscopic systems which are identified with volume elements of the continuum so that one can imagine that each such small system is a system in a stationary state transported around by the fluid motion. Therefore by making local observations on small volume elements we can imagine that fluctuations are more frequent there than in the whole fluid and that in each volume element the evolution takes place as a thermostatted evolution with chaotic fluctuations governed by a fluctuation theorem in which the rate function $\zeta_V(p)$ and the local entropy creation rate $\langle\lambda\rangle_+$ are proportional to the volume under observation $\zeta_V(p) = V\,\overline{\zeta}(p)$ and $\langle\lambda\rangle_+ = V\;\overline{\sigma}_+$ if p is the dimensionless local entropy creation rate and $\overline{\sigma}_+$ is the entropy production rate per unit volume. In other words we want think of the fluid in a stationary state as an ensemble of copies of small systems in the same conditions each of which can be described as a system evolving in presence of a thermostat.

It is difficult to get a more precise picture of the above situation and to check it in concrete model. So far there are examples of systems in which a local entropy creation rate $p\,\overline{\sigma}_+$ per unit volume and a local rate function $\overline{\zeta}(p)$ can be defined and verify a local fluctuation theorem in the sense that the rate function *per unit volume* can be properly defined and verifies

$$\overline{\zeta}(-p) = \overline{\zeta}(p) - p\,\overline{\sigma}_+ \tag{7.3.31}$$

In such systems one has a "*spatio–temporal intermittency* in the sense that the fraction of time intervals of size T in which we shall observe p in a given box of size V_0 will be $e^{(\overline{\zeta}(p)-\overline{\zeta}(1))\,V_0\,T}$. This same quantity will be the fraction of boxes in which we shall observe, within a given time interval of size T, entropy production p.

Normally we shall see $p = 1$ in a fixed box V_0 but "seldom" we shall see $p = -1$ and then, by the above extension of the Onsager–Machlup theory, *everything will look wrong*: every improbable pattern will appear as frequently as we would expect its (probable) antipattern to appear. This will last only for a moment and then things will return normal for a very long time (as the fractions of time in which this can happen in a given box is $e^{-\overline{\sigma}_+ V_0 T}$). Furthermore, fixed a time interval of size T, we shall also see intermittency, in the form of a reversed time arrow, *happening somewhere* in a small volume V_0 in the volume V of the system, provided

$$V\,{V_0}^{-1} e^{-\overline{\sigma}_+ V_0\,T} \simeq 1 \tag{7.3.32}$$

Hence, in such cases, there will be a simple relation between fraction of volumes and fraction of times where time reversal occurs: namely they are equal and directly measured by $e^{-\overline{\sigma}_+ V_0\,T}$, *i.e.* by the average entropy creation rate. And the situation looks more promising from an experimental viewpoint (on real fluids) because we can imagine taking V_0 and T not too large so that the fluctuations will not be so rare to be unobservable and we find ourselves in a situation analogous to the one we meet when we try to observe density fluctuations in a rarified gas. The latter can be seen only in small volumes and intermittently in space: but the rate function that controls the fluctuations is proportional to the volume in which they are observed. The above considerations are at the basis of attempts to intepret certain experimental results, [CL98], [Ga00].

Appendix: Onsager reciprocity as a consequence of the fluctuation theorem.

The computation of $\zeta(p)$ for $\underline{G}$ small will be performed by means of a development in series. As often in statistical mechanics it is useful to first compute the Laplace transform of the probability distribution $\pi_\tau(p) = e^{-\tau\zeta(p)}$:

$$\begin{aligned} e^{\tau\lambda(\beta)} &= \int e^{\beta\tau\,(p-1)\langle\sigma\rangle_+ - \tau\zeta(p)} dp = \\ &= \int d\mu_+(x) e^{\beta\sum_{-\tau/2}^{\tau/2}(\sigma(S^j x)-\langle\sigma\rangle_+)} \end{aligned} \tag{7.3.33}$$

(where now τ is a continuous variable) and then deduce $\zeta(p)$ via a suitable anti-transform. Which is, as it is well known, the Legendre transform of the function λ: $\zeta(p) = \max_\beta\,(\beta(p-1) - \lambda(\beta))$.

Taking the logarithm of (7.3.33) and developing in series the result one finds

$$\lambda(\beta) = \frac{1}{2!}\beta^2 C_2 + \frac{1}{3!}\beta^3 C_3 + \ldots \tag{7.3.34}$$

where the coefficients C_j are combinations of average values of products of $\sigma(S^j x)$ computed at various values of j. In the limit $\tau \to \infty$, and provided the integrals

$$C_j = \int_{-\infty}^{\infty} \langle \sigma(S_{t_1}\cdot)\sigma(S_{t_2}\cdot)\ldots\sigma(S_{t_{j-1}}\cdot)\sigma(\cdot)\rangle_+^T\, dt_1 \ldots dt_{j-1} \tag{7.3.35}$$

converge absolutely, if $\langle\ldots\rangle_+^T$ denote precisely the "suitable combinations of products". Such combinations are called *cumulants* of the distribution of $\sigma(\cdot)$ and for example (as one verifies directly):

$$C_2 = \int_{-\infty}^{\infty} (\langle\sigma(S_t\cdot)\sigma(\cdot)\rangle_+ - \langle\sigma(\cdot)\rangle_+\langle\sigma(\cdot)\rangle_+)\, dt \tag{7.3.36}$$

The convergence of the integrals is a consequence of the chaotic hypothesis that implies that the dynamical system (M, S, μ_+) is mixing and mixes with exponential velocity the correlations between regular observables (as the $\sigma(x)$).

Hence the computation to second order, for which the (7.3.36) suffices, tells us that $\lambda(\beta) = \frac{1}{2}\beta^2 C_2$ and hence, inverting the transform as said after (7.3.33), we deduce an expression for $\zeta(p)$

$$\zeta(p) = \frac{1}{2}\frac{\langle\sigma\rangle_+^2}{C_2}(p-1)^2 + O((p-1)^3 G^3) \tag{7.3.37}$$

Comparing with the fluctuation theorem (*i.e.* $\zeta(-p) - \zeta(p) = p\langle\sigma\rangle_+$) and imposing the compatibility between the two relations we get

$$\langle\sigma\rangle_+ = \frac{1}{2}C_2 + O(G^3) \tag{7.3.38}$$

If we now develop the left hand side in series of $\underline{G}$ around $\underline{G} = \underline{0}$ one finds, *to second order in* $\underline{G}$ and abridging from now on ∂_{G_i} with ∂_i:

$$\langle\sigma\rangle_+ = \frac{1}{2}\sum_{ij} G_i G_j [\partial_i\partial_j\langle\sigma\rangle_+]_{\underline{G}=\underline{0}} \tag{7.3.39}$$

But the quantity $\partial_i\partial_j\langle\sigma\rangle_+$ is the sum of three terms

$$\begin{aligned}&\int \mu_+(dx)\,\Big(\partial_i\partial_j\sigma(x)\Big) + \int \Big(\partial_i\partial_j\mu_+(dx)\Big)\,\sigma(x) +\\ &+\Big[\int \Big(\partial_i\sigma(x)\Big)\Big(\partial_j\mu_+(dx)\Big) + (i \longleftrightarrow j)\Big]\end{aligned} \tag{7.3.40}$$

where the first two terms obviously vanish if $\underline{G} = \underline{0}$. The first because if $\underline{G} = \underline{0}$ the distribution μ_+ is invariant by time reversal (and coincides with the Liouville distribution μ_0 because $\sigma = 0$) and σ is odd by time reversal; the second because $\sigma = 0$ if $\underline{G} = \underline{0}$. Hence

$$\partial_i\partial_j\langle\sigma\rangle_+|_{G=0} = \Big(\partial_j\langle J_i^0\rangle_+ + \partial_i\langle J_j^0\rangle_+\Big)\Big|_{G=0} \tag{7.3.41}$$

Let us remark now that if $\underline{G}=\underline{0}$ it is $\partial_i\langle J_j\rangle_+=\partial_i\langle J^0_j\rangle_+$, because J^0 and J are odd with respect to i and differ by infinitesimals $O(G)$, by (7.3.21).

Hence since $L_{ij}\stackrel{def}{=}\partial_i\langle J_j\rangle_+$, *c.f.r.* (7.3.22), one finds (by equating the coefficients of second order in $\underline{G}$ of the two sides of the (7.3.37) and, to simplify the result, using the fact that $\langle J_i\rangle_+|_{\underline{G}=\underline{0}}=0$)

$$\frac12(L_{ij}+L_{ji})=\frac12\int_{-\infty}^{\infty}\langle J_i(S_t\cdot)J_j(\cdot)\rangle_+|_{G=0}dt \tag{7.3.42}$$

which, setting $i=j$, shows us that the fluctuation theorem (7.3.8) reduces, in the limit in which $\underline{G}\to\underline{0}$ to the fluctuation dissipation theorem *for a single current* (*i.e.* to the simple Green–Kubo formula).

To see that the fluctuation theorem implies more generally the Onsager's reciprocity relations and the general fluctuation dissipation theorem, always in the limit $\underline{G}=0$, it is necessary to make use of its more general formulation in (7.3.17).

One chooses, fixed j, as observable $\kappa_1(x)\equiv\kappa(x)$ the magnitude $\kappa(x)=G_j\partial_j\sigma(x)$ that is "odd" in the sense discussed in connection with (7.3.18), and one defines $q_1\equiv q$ as

$$\int_{-\tau/2}^{\tau/2-1}\kappa(S_tx)\,dt=\tau G_j\langle\partial_j\sigma\rangle_+q=\tau\langle\kappa\rangle_+q \tag{7.3.43}$$

and proceeding as in the already seen case, one computes $\zeta(p,q)$ by computing first the Laplace transform

$$e^{\tau\lambda(\beta_1,\beta_2)}=\int e^{\beta_1(p-1)\langle\sigma\rangle_++(q-1)\langle\kappa\rangle_+-\tau\zeta(p,q)}dpdq \tag{7.3.44}$$

always with the *method of the cumulants* and neglecting the third order in $\underline{G}$. The $\zeta(p,q)$ is then computed by means of a Legendre transform (as after the (7.3.33)) on two variables β_1,β_2.

Comparing the result with the fluctuation theorem one obtains, after elementary computations, analogous to those already described, the relation

$$\langle G_j\partial_j\sigma\rangle_+=\frac12C_{12}+O(G^3) \tag{7.3.45}$$

analogous to (7.3.38). And this relation, *now asymmetric because j plays a special role* having been fixed *a priori*, combined with (7.3.38) is translated into $L_{ij}=L_{ji}$, essentially by repeating the observations that led to the (7.3.42), *c.f.r.* [Ga96a], and, at the same time, into the Green–Kubo relation

$$L_{ij}=\frac12\int_{-\infty}^{\infty}\langle J_i(S_t\cdot)J_j(\cdot)\rangle_+|_{G=0}dt \tag{7.3.46}$$

which is a stronger *nonsymmetric* version of the simpler (7.3.42).

The reversibility assumption used to link the Onsager relations and the fluctuation theorem is supposed for all $\underline{G}$ near $\underline{0}$, However Onsager reciprocity holds more generally under the only assumption of reversibility at $\underline{G}=\underline{0}$ A derivation of the reciprocity solely based upon the latter assumption and on the chaotic hypothesis is possible as well, see [GR97].

Bibliography: [An82],[GC95a],[GC95b][BGG97],[Ga95a],[Ga97], [Ga99a], [Gr97]. The connection between fluctuation theorem and Green–Kubo formulae has been observed empirically in the experiments in [BGG97] where it was correctly interpreted by one of the authors (P.G.) and from it the theory of Onsager's relations of this section started.

§7.4 The structure of the attractor for the Navier–Stokes equations. Dissipative Euler Equations. Barometric formula.

To conclude this work we turn to an attempt at a further understanding of Kolmogorov's theory and to the description of further properties of what we shall call the *Navier–Stokes attractor*, meaning with this elocution the statistical properties of the invariant distribution associated with the NS evolution under constant forcing and giving its statistics.

(A) Reversible and irreversible equations for a real fluid.

We shall consider the following four equations

$$\begin{aligned}
&\dot{\underline{u}} + \underset{\sim}{u} \cdot \underset{\sim}{\partial}\, \underline{u} = -\frac{1}{\rho}\underline{\partial} p + \underline{g} + \nu \Delta \underline{u}, \qquad \underline{\partial} \cdot \underline{u} = 0 \qquad \text{NS}\\
&\dot{\underline{u}} + \underset{\sim}{u} \cdot \underset{\sim}{\partial}\, \underline{u} = -\frac{1}{\rho}\underline{\partial} p + \underline{g} + \beta \Delta \underline{u}, \qquad \underline{\partial} \cdot \underline{u} = 0 \qquad \text{GNS}\\
&\dot{\underline{u}} + \underset{\sim}{u} \cdot \underset{\sim}{\partial}\, \underline{u} = -\frac{1}{\rho}\underline{\partial} p + \underline{g} - \chi \underline{u}, \qquad \underline{\partial} \cdot \underline{u} = 0 \qquad \text{ED}\\
&\dot{\underline{u}} + \underset{\sim}{u} \cdot \underset{\sim}{\partial}\, \underline{u} = -\frac{1}{\rho}\underline{\partial} p + \underline{g} - \alpha \underline{u}, \qquad \underline{\partial} \cdot \underline{u} = 0 \qquad \text{GED}
\end{aligned} \tag{7.4.1}$$

that describe an incompressible fluid in a region Ω that will be a tridimensional torus, possibly deprived of some circular regions (*obstacles*). For simplicity it will be convenient to suppose that the obstacles, if present, are such that by repeating them periodically in space they would "occult" infinity (*i.e.* there is no straight line that can be drawn in space without intersecting the lattice formed by the obstacles and their copies).

On the boundary of the obstacles we shall put *slip boundary conditions* *i.e.* $\underline{u} \cdot \underline{n} = 0$ if $\underline{n}$ is the normal to the obstacles.

The first equation is the NS equation with viscosity ν. The second equation is the Gaussian Navier–Stokes equation, or GNS equation, introduced in §7.1.[1] As seen in §7.1 this means that β is, (7.1.4)

$$\beta(\underline{u}) = \frac{\int_\Omega \left(\underline{\partial} \wedge \underline{g} \cdot \underline{\omega} + \underline{\omega} \cdot \, (\underset{\sim}{\omega} \cdot \underset{\sim}{\partial} \underline{u})\right) d\underline{x}}{\int_\Omega (\underline{\partial} \wedge \underline{\omega})^2 \, d\underline{x}} \tag{7.4.2}$$

[1] The symbol for the multiplier necessary to fix the total vorticity $\eta L^3 = \rho \int \underline{\omega}^2 \, dx$, with $\underline{\omega} = \underline{\partial} \wedge \underline{u}$, is here changed into β, *c.f.r.* (7.1.3).

The third equation, *c.f.r.* §6.2 (D), will be called *dissipative Euler equation*, or ED: and it represents a nonviscous ideal fluid that flows on a "sticky bottom": think to the case $d = 2$ in which the fluid flows on a real surface (*i.e.* a "rough" surface). The constant χ can be called *sticky viscosity*, *c.f.r.* §6.2. In the case $d = 3$, that here interests us, this equation does not seems to be a good model for a real fluid and it will be considered initially only for the purpose of comparison with the Navier–Stokes equation. But it will result, from the discussion, that the connection between the four equations is in reality very strict and they are in a certain sense *equivalent*.

The fourth equation will be called *Gaussian–dissipative Euler equation*, or GED, and here α is a multiplier defined so that the total kinetic energy $\varepsilon L^3 = \frac{\varrho}{2}\int \underline{u}^2\, dx$ is a constant of motion *in spite of* the action of the force $\underline{g}$; this means that α is given by

$$\alpha(\underline{u}) = \frac{\int_\Omega \underline{g}\cdot\underline{u}\, d\underline{x}}{\int_\Omega \underline{u}^2\, d\underline{x}} \tag{7.4.3}$$

A similar equation but with a different constraint has been considered in [SJ93]: the constraint considered there is that the energy content per unit volume and in every shell of momentum, in the sense of the §6.2, is *prefixed and equal to value predicted from the theory K41* (*i.e.* $\int_{k_n}^{k_{n+1}} K(k)dk \propto (\nu\eta)^{2/3}k_n^{-2/3}$, *c.f.r.* (6.2.12), if $k_n = 2^n k_0 = 2^n\, 2\pi/L$). A reversible equation with variable reversible friction appeared earlier in [Ge86], see the review [MK00].

Both the equations GED and the GNS have a symmetry in $\underline{u}$, that makes them *reversible* in the sense that if S_t is the flow that solves the equations (so that $t \to S_t\underline{u} = \underline{u}(t)$ is the solution with initial datum $\underline{u}$), then the map $i : \underline{u} \to -\underline{u}$ *anticommutes* with the time evolution

$$i\, S_t \;=\; S_{-t}\, i \tag{7.4.4}$$

In absence of results on existence and uniqueness for the equations (7.4.1) we shall consider only the truncated equations with momentum cut–off K, *c.f.r.* §2.2, §3.2 and §6.2, so large that it will be possible to suppose heuristically that the solutions of the truncated equations can be a good model for the motion.

The truncation will be performed on a convenient orthonormal base in the space of the zero divergence fields $\underline{u}$: we shall consider natural, on account of the simple boundary conditions chosen, to use the base generated by the *minimax principle*, *c.f.r.* problems of the §2.2, applied to the Dirichlet quadratic form $\int_\Omega(\underset{\sim}{\partial}\, \underline{u})^2 d\underline{x}$ defined on the space of the divergenceless fields $\underline{u} \in C^\infty(\Omega)$ and tangent to the boundaries of the obstacles: $\underline{u}\cdot\underline{n} = 0$ on $\partial\Omega$ and $\underline{\partial}\cdot\underline{u} = 0$ in Ω.

The fields of the base will then verify, *c.f.r.* §2.2, $\Delta\underline{u}_j = -E_j\underline{u}_j + \underline{\partial}_j\mu$, with $\underline{u}_j, \mu_j \in C^\infty$, if μ_j is a suitable multiplier and E_j are eigenvalues.

For example *in the case of a container with no obstacles* let $\underline{u} = \sum_{\underline{k}\ne\underline{0}} \underline{\gamma}_{\underline{k}} e^{i\underline{k}\cdot\underline{x}}$ be the representation of $\underline{v}$ as Fourier series, with $\underline{\gamma}_{\underline{k}} = \overline{\underline{\gamma}}_{-\underline{k}}$ and $\underline{k}\cdot\underline{\gamma}_{\underline{k}} = 0$; here the "momentum" $\underline{k}$ has components that are integer multiples of the "lowest" momentum $k_0 = 2\pi/L$. Then we consider the equation

$$\dot{\underline{\gamma}}_{\underline{k}} = -\vartheta(\underline{k})\underline{\gamma}_{\underline{k}} - i \sum_{\underline{k}_1+\underline{k}_2=\underline{k}} (\underline{\gamma}_{\underline{k}_1}\cdot\underline{k}_2)\,\Pi_{\underline{k}}\underline{\gamma}_{\underline{k}_2} + \underline{g}_{\underline{k}} \tag{7.4.5}$$

in which the $\underline{k}$ takes only values $0 < |\underline{k}| < K$ for some suitably large *cut–off momentum* $K > 0$ and $\Pi_{\underline{k}}$ is the projection on the plane orthogonal to $\underline{k}$. This is an equation that defines a "truncation at scale K" of the equations (7.4.1) if

$$\begin{cases}\vartheta(\underline{k}) = -\nu\underline{k}^2 & \text{NS case}\\ \vartheta(\underline{k}) = -\beta\underline{k}^2 & \text{GNS case}\end{cases}\qquad \begin{cases}\vartheta(\underline{k}) = -\chi & \text{ED case}\\ \vartheta(\underline{k}) = -\alpha & \text{GED case}\end{cases} \tag{7.4.6}$$

We shall suppose, in this case with no obstacles, that the mode $\underline{k} = \underline{0}$ is absent, *i.e.* $\underline{\gamma}_{\underline{0}} = \underline{0}$: this is possible if, as we shall suppose, the external force $\underline{g}$ does not have a component on the Fourier mode $\underline{0}$, (*i.e.* it has average zero).

In the no obstacles case it is also easy to express the coefficients α, β for the truncated equations

$$\begin{aligned}
\alpha &= \frac{\sum_{0<|\underline{k}|<K} \overline{\underline{g}}_{\underline{k}}\cdot\underline{\gamma}_{\underline{k}}}{\sum_{0<|\underline{k}|<K} \underline{\gamma}_{\underline{k}}^2}\\
\beta &= \beta_i + \beta_e, \qquad \beta_e = \frac{\sum_{\underline{k}\ne\underline{0}} \underline{k}^2 \underline{g}_{\underline{k}}\cdot\overline{\underline{\gamma}}_{\underline{k}}}{\sum_{\underline{k}} \underline{k}^4 |\underline{\gamma}_{\underline{k}}|^2}\\
\beta_i &= \frac{-i\sum_{\underline{k}_1+\underline{k}_2+\underline{k}_3=\underline{0}} \underline{k}_3^2\,(\underline{\gamma}_{\underline{k}_1}\cdot\underline{k}_2)\,(\underline{\gamma}_{\underline{k}_2}\cdot\underline{\gamma}_{\underline{k}_3})}{\sum_{\underline{k}} \underline{k}^4|\underline{\gamma}_{\underline{k}}|^2}
\end{aligned} \tag{7.4.7}$$

where the $\underline{k}$ takes only the values $0 < |\underline{k}| < K$, with a cut–off at momentum $K > 0$.

The cases in which the region Ω contains obstacles is very similar, even though we cannot write simple expressions for the fields of the base nor for the truncated equations, that are formally very close to the (7.4.5), *c.f.r.* §2.2, so much that for brevity we shall always refer to (7.4.5)÷(7.4.7) *even* in the cases in which we shall consider other boundary conditions (obviously in such cases we shall have to think that in reality the equations are somewhat different, for example the $|\underline{k}|$ will be in reality $\sqrt{E_j}$ *etc*, but the differences will never be important except when explicitly mentioned).

Let us denote with $S_t^{\nu,ns}\underline{u}, S_t^{\eta,gns}\underline{u}, S_t^{\chi,ed}\underline{u}, S_t^{\varepsilon,ged}\underline{u}$ the solutions of the equations (7.4.5), or of the corresponding ones in the cases with obstacles, corresponding to a given initial datum $\underline{u}$. Or in general

$$S_t^\xi \underline{u}, \quad \xi = (\nu, ns), (\eta, gns), (\chi, ed), (\varepsilon, ged) \tag{7.4.8}$$

The label ξ specifies the model that we consider among the four models $\xi = ns, gns, ed, ged$; η is the total vorticity $\int_\Omega (\underline{\partial}\wedge\underline{u})^2 d\underline{x}$, constant in the GNS evolution, and ε is the total energy $\int_\Omega \underline{u}^2 d\underline{x}$, constant in the ED evolution. Note that, in general, the "phase space" is not the same for the various models, because in some of them (GNS and GED) the velocity fields are subjected to constraints.

Keeping the nonconservative force $\underline{g}$ constant, we shall suppose that for every equation, *i.e.* for every choice of the label ξ, (7.4.8), there is only one stationary distribution μ_ξ that describes the statistics of a given initial $\underline{u}$, chosen with a distribution μ_0 endowed with density on phase space: note that, being $K < \infty$, phase space has finite dimension.

The value of K will be fixed in the case NS setting $K = k_\nu$, *c.f.r.* eq. (6.2.9) and (D) in §6.2; in the case GNS one shall choose $K = K^1$ so large that the average value $\langle\beta\rangle_{\eta,gns} \overset{def}{=} \tilde\nu$ becomes independent of K (we suppose that this is possible) and then we shall choose $K = \max(k_{\tilde\nu}, K^1)$.[2] In the cases ED and GED we shall make analogous choices of K, always under the hypothesis that at $\chi, \underline{g}$ fixed there exists a value of K such that the time averages of the observable α become "effectively" K–independent.

What follows however *does not* depend on the choice made on K: hence one could also avoid presupposing this "ultraviolet" stability hypothesis with respect to the values of K (that could easily appear unreasonable) and it will be enough that one only supposed that K is "large".

We assume existence of the statistics: which means that, given an observable F on phase space $\mathcal{F}$ (of the velocity fields with cut–off to momentum K), for some probability distribution μ_ξ it is

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T F(S_t^\xi \underline{\gamma})dt = \int_{\mathcal{F}} F(\underline{\gamma}')\mu_\xi(d\underline{\gamma}') \overset{def}{=} \langle F\rangle_\xi \tag{7.4.9}$$

for all choices of $\underline{\gamma}$ except for a set of volume zero (with respect to the volume measure on phase space). The latter will be called the *SRB distribution* for the equations (7.4.5),(7.4.6).

A particular role will be plaid, as we can imagine from the analysis of the §7.1, §7.2, §7.3, by the $\langle\eta\rangle_\xi, \langle\varepsilon\rangle_\xi$ and by the averages $\langle\alpha\rangle_\xi, \langle\beta\rangle_\xi$. Together, obviously, with the rate of entropy production $\sigma(\underline{\gamma})$ *defined*, in agreement with what was said at §7.3, by the *divergence* of the r.h.s. of the truncated equations and of its average $\langle\sigma\rangle_\xi$ (*c.f.r.* §7.1 where this quantity has been denoted $\delta(x)$).

Considering explicitly the case with no obstacles let

---

[2] The discussion that follows hints that $K^1 \equiv k_{\tilde\nu}$.

• D_K the number of modes $\underline{k}$ with $0 < |\underline{k}| < K$: so that the number of (independent) components $\{\underline{\gamma}_{\underline{k}}\}$ is $2D_K$ (note that $\underline{\gamma}_{\underline{k}}$ has two independent complex components for each $\underline{k}$ but $\underline{\gamma}_{\underline{k}} = \overline{\underline{\gamma}}_{-\underline{k}}$).
• $2\,\overline{D}_K = \sum_{|\underline{k}|<K} 2\underline{k}^2$ (that in the case with obstacles would become the quantity $2\,\overline{D}_K = \sum_{\sqrt{E_j}<K} E_j)$).

One finds that the phase space contraction per unit time is σ given by

$$\begin{aligned}
\sigma =&2\,\overline{D}_K\nu & \xi = (\nu, ns)\\
\sigma =&2\,\overline{D}_K\beta - \overline{\beta}_e - \overline{\beta}_i & \xi = (\eta, gns)\\
\sigma =&2D_K\chi & \xi = (\chi, ed)\\
\sigma =&2D_K\alpha - \alpha & \xi = (\varepsilon, ged)
\end{aligned} \tag{7.4.10}$$

where $\overline{\beta}_i$, $\overline{\beta}_e$ are suitably defined. For example in the case without obstacles:

$$\overline{\beta}_e = \frac{\sum_{\underline{k}} \underline{k}^2\,\overline{\underline{g}}_{\underline{k}} \cdot \underline{\gamma}_{\underline{k}}}{\sum_{\underline{k}} \underline{k}^4 |\underline{\gamma}_{\underline{k}}|^2} - 2\frac{\big(\sum_{\underline{k}} \underline{k}^2\,\overline{\underline{g}}_{\underline{k}} \cdot \underline{\gamma}_{\underline{k}}\big)\big(\sum_{\underline{k}} \underline{k}^4 \underline{\gamma}_{\underline{k}}^2\big)}{\big(\sum_{\underline{k}} \underline{k}^4 |\underline{\gamma}_{\underline{k}}|^2\big)^2} \tag{7.4.11}$$

hence $\sigma \simeq 2\,\overline{D}_K\beta$ for $\xi = (\eta, gns)$ and $\sigma \simeq 2D_K\alpha$ for $\xi = (\varepsilon, ged)$.

With these definitions (*c.f.r.* §7.1) the following conjecture has been proposed that I will call the *conjecture of statistical equivalence* of the GNS and the NS dynamics

Conjecture *(equivalence betwee NS and GNS): The statistics* $\mu_{\nu,ns}$ *and* $\mu_{\eta,gns}$ *of the NS equation and, respectively, of the GNS equation are equivalent in the limit in which the Reynolds number* R *tends to infinity provided the parameters* η *and* ν *are related so that that* $\langle\sigma\rangle_{\nu,ns} = \langle\sigma\rangle_{\eta,gns}$ *(or, equivalently,* $\nu = \langle\beta\rangle_{\eta,gns}$*).*

Equivalent means that the ratios between average values of the same "local observables" with respect to the two distributions tend to 1 as $R \to \infty$. By *local observable* we mean an observable that depends on the field $\underline{u}$ only through the components of scale contained between two fixed values in the inertial domain: *i.e.* only through the components of the field $\underline{\gamma}_{\underline{k}}$ with "*scale*" $|\underline{k}|$ such that $k_1 < |\underline{k}| < k_2$ with $k_1 \gg k_0$ and $k_2 < \infty$: the "locality" is therefore understood in the "space of momenta"; the Reynolds number is defined here as $R = \eta^{1/3}L^{4/3}\nu^{-1}$, *c.f.r.* comments to the (6.2.9). In the case of nonperiodic boundary conditions the *scale* of a component γ_j of the field will be determined by the value $\sqrt{E_j}$ of the correspondent eigenvector $\underline{u}_j$.

And an analogous equivalence conjecture between statistics can be formulated for the equations ED and GED

Conjecture *(equivalence between ED and GED): The statistics* $\mu_{\chi,ed}, \mu_{\varepsilon,ged}$ *of the equations ED and of the equations GED are equivalent in the limit of large Reynolds number provided the parameters* ε *and* χ *are related so that* $\langle\sigma\rangle_{\chi,ed} = \langle\sigma\rangle_{\varepsilon,ged}$ *(or* $\chi = \langle\alpha\rangle_{\varepsilon,ged}$*).*

Anyone who has some familiarity with statistical mechanics will recognize in the just stated conjectures a strong analogy with the corresponding statements on the equivalence between statistical ensembles, *c.f.r.* [Ga95c]: with the limit $R \to \infty$ that plays the role of the thermodynamic limit.

The idea of the possibility of describing in terms of *statistical ensembles* the statistical properties of systems outside equilibrium has gradually developed in the recent literature and the idea of the possibility of equivalent descriptions in terms of different statistical ensembles emerged at the same time, *c.f.r.* [Ge86], [ES93], [SJ93], [Ga95b],[Ga96], [MK00].

On a heuristic basis the conjectures would be justified *if the rate of entropy creation would reach its average value on a time scale which is rapid with respect to the time scales characteristic of hydrodynamics.* The coefficients $\alpha \simeq (2D_K)^{-1}\sigma$, and $\beta \simeq (2\,\overline{D}_K)^{-1}\sigma$, (7.4.10), could then be identified with their average values $\langle\alpha\rangle_{\varepsilon,gne}$ or $\langle\beta\rangle_{\eta,gns}$ and *hence identified with the viscosity constant* ν or χ.

In this way the GNS and NS equations would be equivalent and both would be macroscopic manifestations of two equivalent microscopic mechanisms of dissipation. One esplicitly specified by the Gaussian constraint of constant total vorticity (and hence total dissipation, by the proportionality between the two quantities), the other one with *a priori* fluctuating dissipation but that can be phenomenologically modeled by means of a constant viscosity.

The same can be said of the relation between the equations ED and GED.

We now look at the problem of understanding how to extract from the conjectures just discussed some consequence observable in experiments. This will be possible by combining what said above with the chaotic hypothesis of the previous sections.

(B) Axiom C and the pairing rule.

Unfortunately it will be still necessary to propose assumptions that it will not be possible to justfy other than by possibly checking some of their consequences. Nevertheless since this is the conclusive section of a "foundations of fluid mechanics" I feel that it will be permitted to discuss them, also because a discussion somewhat out of balance in a really heuristic direction can be stimulating.

What follows has, therefore, to be seen as a collection of ideas that developed naturally while meditating on the many works consulted to accomplish the task that I undertook several years ago, *i.e.* of presenting to the students of my course of fluid mechanics a guided introduction to a very vast field of research.

The main difficulty for the application of the fluctuation theorem to the GNS or GED equations, in spite of their reversibility, is that they are systems *far out of equilibrium* hence we cannot imagine that the attractive set be the whole phase space: typically we expect indeed that the attractor has *finite Hausdorff dimension* (and there exist several arguments in favor of this idea), while the full phase space has *a priori* infinite dimension, [FP67], [Ru82], [Li84]), [Ru84].

Then the chaotic hypothesis tells us that "things proceed as if" the attractor was a smooth, finite dimensional, surface on which dynamics is well modeled by an evolution S that is hyperbolic.

If furthermore we suppose that the dynamics verifies the axiom C, the reversibility of the GED or GNS equations will imply the existence, *c.f.r.* §7.2, of a time reversal map i^* *that leaves invariant the attracting set.*

Hence the fluctuation theorem will hold, *c.f.r.* §7.3. But obviously the contraction of phase space that enters in its formulation *not will be* σ because the latter is the contraction $\sigma(x)$ of the *total* volume and not the contraction $\sigma_0(x)$ of the volume element on the surface of the attracting set.

Of course *one cannot hope to characterize* in a simple way the attracting set for the purpose of computing its element of surface. A theory that made reference to the "equations of the attracting set" would risk strongly to remain totally inapplicable (see, nevertheless, the case of the GOY model of§6.3 in which equations of the attracting set can be proposed and used, *c.f.r.* (6.3.22)).

Help comes from a property whose validity has been slowly changing from a "curiosity" to "interesting but exceptional" to "interesting and often verified". It is a remarkable property of the Lyapunov exponents of chaotic systems related in some way to Hamiltonian systems afflicted by dissipative phenomena.

It has been noted, starting with the work of Dressler, [Dr88], that in certain Hamiltonian systems with ℓ degrees of freedom and subject to particular forms of friction the Lyapunov exponents, arranged as

$$\lambda_\ell^+ \ge \lambda_{\ell-1}^+ \ge \ldots \ge \lambda_1^+ \ge \lambda_1^- \ge \lambda_2^- \ge \ldots \ge \ldots \lambda_\ell^-$$

in decreasing order and adding a superscript $\pm$ to distinguish the ones with the same subscript, are such that

$$\frac{1}{2}(\lambda_j^+ + \lambda_j^-) = constant \qquad \text{for each } j \tag{7.4.12}$$

In fact this property is even true, at least in the first examples in which it was found and if the metric that is used is suitably chosen, for the eigenvalues of the matrices of *local expansion and contraction*, *i.e.* also without considering the limit that appears in the definition of Lyapunov exponents; in such case the constant depends on the phase space point where the expansions and contractions are computed. To be precise we define here what we mean by *local Lyapunov exponents* over a time τ: they are the eigenvalues of the matrix $(J_\tau^T(x)J_\tau(x))^{1/2\tau}$ with $J_\tau(x)$ being the Jacobian matrix of the map S^τ as a map between $S^{-\tau/2}x$ and $S^{\tau/2}x$ (τ even).

The value of the constant in (7.4.12) is (obviously) the average value of the phase space contraction. For an illustration of a numerical check of the rule (in a case it can be proved to be rigorously valid see Fig. (7.4.1)).

The breakthrough on the above "*pairing rule*" has been due to an experimental discovery, [ECM90]: an actual mathematical proof came only later: this is just one example of a property that is mathematically important and yet relatively "easy" to prove but which has been missed by mathematicians. For an illustration of a numerical check of the rule (in a case in which it can be proved to be rigorously valid) see Fig. (7.4.1).

Recently the *pairing rule* (7.4.12), [DM97], has been shown valid, *in the local version*, for rather wide classes of systems (called *isokinetic*) subject to Gaussian dissipative constraints in which forcing takes place through the action of locally conservative external forces (but not globally conservative, like an electromotive electric field). A typical example being a system of particles subject to the Gaussian constraint of keeping total kinetic energy constant and to a force that tends to establish a current of matter in a (not simply connected) container. An important extension is in [WL98].

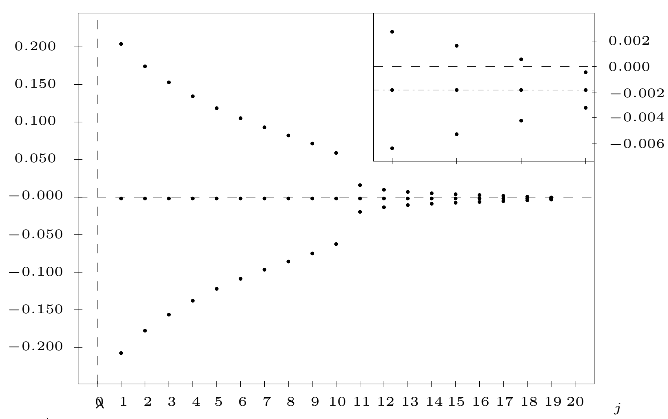


Fig. (7.4.1) *The* 38 *Lyapunov exponents for a model of electrical conduction, in a very large electromotive field; the model has* 19 *degrees of freedom and is governed by isokinetic equations. The small figure is an enlargement of the tail of large one and it shows the pairing rule and the fact that at such field the* 19*–th exponent is slightly negative and hence the attracting set has dimension lower than that of the phase space.* From [BGG97].

The interest of a dissipative reversible system for which the pairing rule holds in a local sense lies in the fact that it seems possible to establish for it in a natural way a relation between the contraction, $\sigma_0(x)$, around x of the surface element of the attracting set and that, $\sigma(x)$, of the volume element with full dimension.

Indeed, following the analysis in [BGG97], if $2N$ is the dimension of the phase space and $2(N-M)$ that of the attracting set (assumed smooth) it will be that the contraction of the phase space around points located on the surface of the attracting set A is $\sigma(x)=\sigma_0(x)+\sigma_\perp(x)$ where σ_0 is the rate of contraction "on" the attracting set and $\sigma_\perp$ that on the part of the stable manifold of the point x which is a manifold that partly "sticks out of A", *c.f.r.* footnote [5] in 7.2.

This can be interpreted by thinking that the tangent space in x consists of $2(N-M)$ directions, $N-M$ of which expansive and $N-M$ contracting, all tangent to A and in $2M$ directions *all contracting* that instead concern the part of stable manifold that sticks out of the attracting set; *furthermore*, it appears natural that the *pairs* of exponents are divided into $N-M$ *pairs* relative to the $2(N-M)$ directions tangent to A and in the $2M$ remaining ones.

The above is not the only possibility, but it is certainly the simplest. And if it is verified then we deduce immediately, assuming the local pairing rule (*i.e.* (7.4.12) with an x–dependent constant) that

$$\sigma_0(x)=\frac{N-M}{N}\sigma(x) \tag{7.4.13}$$

i.e. there is *proportionality between the the total contraction $\sigma(x)$ of phase space and that, $\sigma_0(x)$, of the element of surface on the attracting set.*

Since $\sigma(x)$ is directly accessible, or at least more directly accessible, than the individual Lyapunov exponents we realize the great potential of the pairing rule. For example combined with time reversibility, axiom C and the fluctuation theorem of the §7.3 it tells us that, *c.f.r.* (7.3.8):

$$\zeta(p)-\zeta(-p)=(1-\frac{M}{N})\langle\sigma\rangle_+p \tag{7.4.14}$$

where $\zeta(p)$ is defined by (7.3.7) in terms of the *total* contraction (given by σ) of phase space.

Eq. (7.4.14) gives the result, perhaps surprising at first sight, that the slope of the $\zeta(p)-\zeta(-p)$ as a function of p *diminishes* by $1-\frac{M}{N}$ if the dimension of the attractor diminishes, *i.e.* if the viscous phenomena increase. This result, although still not accurately checked by any experiment, seems at least consistent with the results of the experiments in §6 of [BGG97] (that have inspired it) and with a few others that followed.

What has all this to do with (7.4.1)? First of all if the forcing field $\underline{g}$ is locally conservative (possible only if the container Ω has holes!)[3] one must

[3] Because it is not possible to define on a torus a vector field $\underline{g}$ with zero average locally conservative and not globally such (hence that cannot be trivially absorbed in the pressure term of the equation): in order that such a field exists it is necessary that there be holes, *i.e.* that the regions that the fluid can occupy be not simply connected (*e.g.* periodic).

remark that, *c.f.r.* [Ga96], the systems ED and GED are *naturally* among the cases in which the pairing rule has been proved. The first is a particular case of the theorem in [Dr88] while the second is a particular case of the theorem in [DM97].

The key observation is that the Euler equations can be thought of as *half* of the equations that describe the motion of the fluid: the other half of the equations describes the field of the displacements $\underline{\delta}(x)$ of the points of the incompressible fluid with respect to the positions that they have in a reference configuration

$$\dot{\underline{\delta}}(x)=\underline{u}(\underline{\delta}(x)),\qquad \dot{\underline{u}}(x)+\underset{\sim}{u}\cdot\underset{\sim}{\partial}\underline{u}=-\frac{1}{\rho}\,\underline{\partial}\,p(x) \tag{7.4.15}$$

as it was discussed in detail to §1.7, point (E), (1.7.29).

Having made this observation it is easy to realize that, at least formally, the ED equations are just equations obtained by imposing an external locally conservative force and a friction proportional to the velocity. While the GED case, correspond to imposing the same external force and an isokinetic constraint, *i.e.* the constraint $\int\underline{u}^2dx=constant$, by means of the Gauss' principle.

Hence we get the formal validity of the pairing rule *for the equations (7.4.15) in a phase space with a double number of dimensions*: in which the displacements $\underline{\delta}(x)$ with respect to a reference configuration are also described, see [Ga96] for the details. Naturally the Lyapunov exponents of the GED equation *will not* verify the pairing rule because several of the pairs will consist in pairs of exponents of which one is relative to the GED and the other is relative to the additional degrees of freedom "of displacement".

But an attentive examination of the pairing rules proofs, see [DM97], [WL98], in the cases in which they are known to hold, induce to think that, in spite of the many examples, the pairing rule *is not general* and conclusive numerical evidence has been provided in [BCP98]. For example it *does not seems reasonable* that it holds *in the case of the GNS equation*. Hence to be able to obtain informations, relevant for real fluids, from the fluctuation theorem it is necessary, besides assuming the validity of the axiom C, some extension of the pairing rule that allows us *to establish a relation between the contraction rate* σ *of the whole phase space and that* σ_0 *of the surface of the attracting set*.

The discussion in [Ga96] proposes indeed *to define* in the case of the GNS equation the numbers c_j as

$$c_j=\frac{\lambda_j^+ +\lambda_j^-}{\langle\beta\rangle_+} \tag{7.4.16}$$

where $\langle\beta\rangle_+$ is the average of β, see (7.4.1), with respect to the distribution $\mu_{\eta,gns}$. Here the exponents $\lambda^{\pm})j$ are the ones of the GNS equation *coupled* with the displacement equation $\dot{\underline{\delta}}=\underline{u}(\underline{\delta}(\underline{x}))$.

In the case NS one must instead define c_j by means of the (7.4.16) with ν that replaces $\langle\beta\rangle_+$. And the paper [Ga97] proposes, then, the following generalization of the pairing rule

The local Lyapunov exponents verify (7.4.16) in the average and the average value is reached very rapidly for Reynolds number R large. So that one can consider (7.4.16) as locally true: hence one can repeat the argument that links the contraction of phase space to the contraction of the surface of the attracting set.

The volume contraction in the phase space of the equations in question is the same whether one considers the equations just as equations only for the velocity fields or also for the fields of velocity $\underline{u}$ and of displacement $\underline{\delta}$. This is due to the "triangular" structure of the Jacobian matrix due to the fact that the displacements equation $|\dot{V}\delta(\underline{x}) = \underline{u}(\underline{\delta}(\underline{x}))$ decouples from the equation of the velocity field, *c.f.r.* (7.4.15). We deduce that the fluctuations of the total contraction $\sigma(x)$ of phase space verifies a fluctuation theorem and the quantity $\zeta(p) - \zeta(-p)$ is linear in p.

The coefficient of proportionality is measurable from the statistics of the solutions of the equations for the field $\underline{u}$ and has value $P\langle\alpha\rangle_+$ in the cases of the GED equations and $\overline{P}\langle\beta\rangle_+$ in the case of the GNS equations where P is the number of pairs of Lyapunov exponents with an element > 0 and one < 0 divided by the total number of pairs, while $\overline{P}$ is $\sum_j^* c_j / \sum_j c_j$ where $\sum_j^*$ runs over the only j which corresponds to a pair of Lyapunov exponents of opposite sign.

On the basis of the equivalence conjectures we could hope to translate some predictions on the equations GNS or GED, immediately, into predictions for the equations NS and ED, respectively. Defining α for ED via the equations (7.4.7) and σ via the corresponding fourth of (7.4.10) we could expect that the fluctuations of σ verify a fluctuation relation with slope $P\,\chi$.

Likewise in the NS case we would expect that, defining β via (7.4.7) and σ via the second of (7.4.10), then σ verifies a fluctuation theorem with slope $\overline{P}\,\nu$. *This, of course, when the conditions of equivalence of the conjectures at point (A) are verified.*

What said may however be doubted because of the fact that the fluctuation theorem deals with the quantity σ, divergence of the equations of the motion, that is a "nonlocal" quantity in the space of the momenta (in the sense above indicated): rather it is global. In statistical mechanics the analogous observables are nonlocal observables that are, often, observables that have different distributions *even in statistically equivalent ensembles* (think to the total energy in the canonical ensemble and in the microcanonical ensemble).

Perhaps the mentioned relation is reasonably valid only if applied to the quantity σ_Δ defined as σ but replacing the integrals on Ω in (7.4.2) and (7.4.3) with integrals on a small volume Δ internal to the fluid: unfortunately a satisfactory analysis of this idea, *c.f.r.* [Ga96], [Ga00], is lacking although it is desirable, since it could lead to "very stringent" predictions that could even conceivably be experimentally checkable, [Ga00].

We address now the problem of showing that there exists a strict relation between the four equations (7.4.1). Developing an idea already in [Ga97].

(C) Relation between the NS and ED equations: the barometric formula.

Meditating on the ED or GED equations and on the NS or GNS equations one is led to think that the relation between them may be similar to the relation that one finds in statistical mechanics between the equilibrium distribution of a gas at different heights when the gas is in the field of gravity.

Locally a gas in a field appears simply as a homogeneous gas in equilibrium, but globally (on a length scale H on which the external potential changes substantially: *i.e.* $\beta mgH \sim 1$, if β is the inverse temperature and m the mass of the particles) one shall see that the pressure and the density are not constant. To describe their variations one arrives at the so called *barometric formula*, *c.f.r.* [MP72].

Likewise we can expect that the stationary states of the ED (or equivalently of the GED) are *also* "locally (in momentum space)" equivalent to stationary states for the NS or GNS: in the sense that if we consider observables that depend on the velocity field components $\underline{u}_{\underline{k}}$ with modes $\underline{k}$ on a certain scale $|\underline{k}| \sim \kappa$ then we should essentially see no difference at all provided the amount of energy present in this shell (that depends on some parameters related to the initial data that generate the stationary states for the two equations) is arranged to be the same.

The exact relation that determines κ will be called *barometric formula*: and it should not be difficult to determine the barometric formula on the basis of considerations of dimensional nature. Note that here "*locality*" has to be understood, as in (A),(B) *in the space of the momenta rather that in that of the coordinates*.

The determination of the barometric formula consists, essentially, in the development of a theory analogous to that of Kolmogorov K41 for the equations ED, *c.f.r.* [Ga96].

One can try to develop such theory in the case of a container without obstacles and on the basis of some hypothesis that at the moment seems reasonable to me. By changing the hypothesis the result could change in its analytic form but the fundamental idea on which the derivation that follows is based is independent of the details of the theory proposed as analogous to the theory K41.

We shall suppose, just to be concrete, as "reasonable" that in the case of the ED equations the stationary distribution equipartitions energy between the modes, *i.e.* $\langle|\underline{\gamma}_{\underline{k}}|^2\rangle = \gamma^2$ for all the $\underline{k}$ in the "inertial domain" $L^{-1} \ll |\underline{k}| \ll k_\chi$ where k_χ is the scale where the ultraviolet cut–off, necessary for giving a mathematical meaning to the equations, is performed *c.f.r.* §6.2 (D). Hence $\gamma^2(k_\chi L)^3 = \varepsilon$ will be the total energy.

For purpose of comparison with (D) in §6.2 we note that the quantity there called ε corresponds to $\eta\nu$ here.

In this case the distribution of energy (*i.e.* the amount $K(k)dk$ of energy per unit volume and between k and $k+dk$) is $K(k)=\frac{3\varepsilon}{4\pi}\frac{k^2}{k_\chi^3}$, for $k<k_\chi$: very different from the law $k^{-5/3}$ of Kolmogorov. It is rather analogous to the law of Rayleigh–Jeans of the black body, *c.f.r.* [Ga92].

In the theory K41 a key role is plaid by the quantity $v_\kappa^3\kappa$ that is identical to $\eta\nu$ for all values $k_0\ll\kappa\ll k_\nu$. Therefore let us compute the value of $v_\kappa^3\kappa$ in our case. We find

$$\frac{v_\kappa^3\kappa}{\varepsilon\chi}=\frac{((\kappa L)^3\gamma^2)^{3/2}\kappa}{\varepsilon\chi}=\frac{\big((k_\chi L)^3\gamma^2\big)^{3/2}k_\chi}{\varepsilon\chi}\big(\frac{\kappa}{k_\chi}\big)^{11/2}=\frac{\varepsilon^{3/2}k_\chi}{\varepsilon\chi}\big(\frac{\kappa}{k_\chi}\big)^{11/2} \tag{7.4.17}$$

and we see that the quantity $v_\kappa^3\kappa$ depends on κ in the ED case.

Given κ the SRB statistics for the ED equations in a stationary state with total energy ε attributes to this quantity the same value that it has in the SRB statistics for the NS equations in a stationary state with total vorticity η if

$$\frac{\varepsilon\chi}{\eta\nu}=\frac{\chi}{\sqrt{\varepsilon}k_\chi}\big(\frac{\kappa}{k_\chi}\big)^{-11/2} \tag{7.4.18}$$

provided (naturally) κ is smaller than the "Kolmogorov scales" k_ν, k_χ.

The "barometric formula" is then the statement of equivalence between NS and ED on the scale κ, i.e. if we only look to the properties of the velocity field that depend on $\underline{\gamma}_{\underline{k}}$ for $\frac12\kappa<|\underline{k}|<\kappa$, if (7.4.18) holds and $\kappa\ll k_\nu,k_\chi$.

If we look on a different scale $\kappa'=2^n\kappa$ for some n (large) then we can expect equivalence between ED or (GED) and NS (or GNS), *but the pairs* ε,η will have now to be such that the equation (7.4.18) holds on the new scale.

The analogy with the usual barometric formula for the distribution of Boltzmann–Gibbs in the field of gravity justifies the name given to the (7.4.18). We see that $\eta\,\nu$ play the role of the gravity, $\varepsilon\,\chi$ that of the chemical potential while κ/k_χ that of the height.

The barometric formula is only an example of the consequences that we can draw from the equivalence conjectures between the stationary states for various equations that describe the dynamics of a given system. *It is interesting also because it gives us the possibility of obtaining the statistics of the NS equation on a given scale by simulating a different simpler equation* (as it is an equation in which no second derivatives appear).

The above analysis seems well in agreement with the spirit of the analysis in [SJ93] that first proposed, in a different context and with different perspectives, a vision that has strong similarity with that discussed here.

Nevertheless in order that the above be accepted as a correct, although phenomenological, analysis it is necessary to determine also the constant k_χ:

as is explained in §6.2 (D), we cannot determine it without a much more detailed theory of the equations ED; the multiplicative constant (*i.e.* k_χ) in (7.4.18) cannot be determined and "only" the exponent 11/2 in the barometric formula is heuristically established.

Problems.

[7.4.1] (*Gauss' principle applied to fluids*) Equations derived on the basis of the Gauss principle depend on the *effort function* $\mathcal{E}(\underline{a})$ of the accelerations that one wants to minimize. The choice $\mathcal{E}(\underline{a}) = ((\underline{f} - m\underline{a})/m)^2$ that was used in the problems of §7.1 was just an example. In systems with infinitely many degrees of freedom this ambiguity becomes even more obvious. For instance we could minimize conditioning to a *fixed total energy* (and to incompressibility) the "effort" $\mathcal{E}_1(\underline{a}) \stackrel{def}{=} \big((\underline{a}+\underline{\partial}p-\underline{f}),(\underline{a}+\underline{\partial}p-\underline{f})\big)$ or, subject to the same constraint, $\mathcal{E}_2(\underline{a},s) \stackrel{def}{=} \big((\underline{a}+\underline{\partial}p-\underline{f}),(-\Delta)^{-1}(\underline{a}+\underline{\partial}p-\underline{f})\big)$. Check that, in toroidal geometry (*i.e.* periodic boundary conditions), imposing $\mathcal{E}_1(\underline{a})$ on divergenceless fields $\underline{u}$ with the constraint $\varphi \stackrel{def}{=} \int(\underline{u})^2\,d\underline{x}$, leads to the GED equations while imposing $\mathcal{E}_2$ leads to equations that look like the incompressible NS equations. In the latter case the equations obtained are not the GNS equations of §7.1, or (7.4.1), (7.4.2): they coincide with the second of (7.4.1) but the multiplier β is different. Compute β in the latter case. (*Idea:* β has to be such that energy rather than dissipation stays exactly constant in time.)

[7.4.2] Check that, in toroidal geometry, the GNS equations in (7.4.1), (7.4.2) can be obtained by applying the Gauss principle with effort $\mathcal{E}_1(\underline{a}) \stackrel{def}{=} \big((\underline{a}+\underline{\partial}p-\underline{f}),(\underline{a}+\underline{\partial}p-\underline{f})\big)$ and constraint $\varphi \stackrel{def}{=} \int(\underline{\partial}\,\underline{u})^2 d\underline{x} = const$ on the divergenceless fields $\underline{u}$. (*Idea:* Note that $\varphi = \int \underline{u}\cdot\Delta\underline{u}\,d\underline{x}$.)

Bibliography: [Ga96], [SJ93], [Ga95b],[Ga95]. The importance of the pairing rule in the isokinetic cases for the purpose of the application of the fluctuation theorem has been noted in the course of the work of interpretation of the experimental results in [BGG97] by one of the authors (F.B.).
The fluid equation with constrained constant energy with effort function $\mathcal{E}_1(\underline{a})$ has been considered in [BPV98] with the purpose of checking the conjecture of equivalence of §7.4. The numerical experiment is performed on the GOY model, rather than on the NS equation, and the results seems to indicate that the equation with constrained energy and with effort given by the analogue of $\mathcal{E}_1$ in problem [7.4.1] gives results in agreement with the conjecture. This seems to be *not so* for the equation with constraint of constant dissipation (*i.e.* the analogue of the GNS equation for the GOY model); it is in this work that the the effort function $\mathcal{E}_1$ was introduced thus extending the conjectures. Experiments on 2–dimensional NS with all the above constraints, and more, have been performed in [RS99] with results that seem always compatible with the conjectures: however the latter experiments are performed with rather severe truncations of the NS equations so that they may not be testing the conjectures at really large Reynolds numbers.

Bibliography.

[AA68] **Arnold, V.I., Avez, A.**: *Ergodic problems of classical mechanics*, Benjamin, New York, 1968.

[AF91] **Adler, R., Flatto, L.**: *Geodesic flows, interval maps and symbolic dynamics*, Bullettin of the american mathematical society, **25**, 229–334, 1991.

[Al93] **Altarelli, G.**: *Interazioni deboli*, in *Dizionario delle scienze fisiche*, ed. Enciclopedia italiana, vol. III, Roma, 1993.

[An82] **Andrej, L.**: *The rate of entropy change in non–Hamiltonian systems*, Physics Letters, **111A**, 45–46, 1982. And *Ideal gas in empty space*, Nuovo Cimento, **B69**, 136–144, 1982. See also *The relation between entropy production and K–entropy*, Progress in Theoretical Physics, **75**, 1258–1260, 1986.

[An90] **Alankus, T.**: *An exact representation of the space time characteristic functional of turbulent Navier Stokes flows with prescribed random initial states and driving forces*, Journal of Statistical Physics, **54**, 859–872, 1990.

[Ar79] **Arnold, V.**: *Metodi Matematici della Meccanica Classica*, Editori Riuniti, 1979.

[Ba69] **Batchelor, G.**: *Computation of the energy spectrum in homogeneous two–dimensional turbulence*, Physics of fluids, **12**, (II supplement), 233–239, 1969.

[Ba70] **Batchelor, G.**: *The theory of homogeneous turbulence*, Cambridge University Press, 1970.

[BCL98] **Bonetto, F., Chernov, N., Lebowitz, J.**: *(Global and Local) Fluctuations of Phase Space Contraction in Deterministic Stationary Non-equilibrium*, Chaos, **8**, 823–833, 1998.

[BCM00] **Benedetto, D., Caglioti, E., Marchioro C.**: *On the motion of a vortex ring with a sharply concentrated vorticity*, Mathematical Methods in the Applied Sciences, **23**, 147–168, 2000.

[BCP98] **Bonetto, F., Cohen, E.G.D., Pugh, C.**: *On the Validity of the Conjugate Pairing Rule for Lyapunov Exponents*, Journal of Mathematical Physics, **92**, 587–627, 1998.

[Be64] **Becker, R.**: *Electromagnetic fields and interactions*, Ed. F. Sauter, New York, Blaisdell Pub., 1964.

[BF79] **Boldrighini, C., Franceschini, V.**: *A five dimensional truncation of the plane incompressible Navier Stokes equation*, Communication in Mathematical Physics, 64, 159-170, 1979.

[BG95] **Benfatto, G., Gallavotti, G.**: *Renormalization group*, p. 1–144, Princeton University Press, 1995.

[BG97] **Bonetto, F., Gallavotti, G.**: *Reversibility coarse graining and the chaoticity principle*, Communications in Mathematical Physics, **189**, 263–276, 1997.

[BGG97] **Bonetto, F., Gallavotti, G., Garrido P.**: *Chaotic principle: an experimental test*, Physica D, **105**, 226–252, 1997.

[BGG97] **Bonetto, F., Gallavotti, G., Garrido, P.**: *Chaotic principle: an experimental test*, Physica D, **105**, 226–252, 1997.

[BGGS80] **Benettin, G., Galgani, L., Giorgilli A., Strelcyn J.M.**: *Lyapunov Characteristic Exponents for Smooth Dynamical Systems and for Hamiltonian Systems; a Method for Computing all of Them. Part 1: Theory*, Meccanica, **15**, 9–20, 1980; and *Lyapunov Characteristic Exponents for Smooth Dynamical Systems and for Hamiltonian Systems; a Method for Computing all of Them. Part 2: Numerical Applications.* Meccanica **15**, 21–30 (1980).

[BGW98] **Bourgain, J. Golse, F., Wennberg, B.**: *On the Distribution of Free Path Lengths for the Periodic Lorentz Gas*, Communications in Mathematical Physics, **190**, 491–508, 1998.

[BJPV98] **Bohr, T., Jensen, M.H., Paladin, G., Vulpiani A.**: *Dynamical systems appproach to turbulence*, Cambridge University Press, 1998.

[BKL00] **Bricmont, J. Kupiainen, A., Lefevere R.**: *Ergodicity of the 2D Navier-Stokes Equations with Random Forcing*, in print in Communications in Mathematical Physics, 2000.

[BKM74] **Bakouline, P., Kononovitch, E., Moroz V.**: *Astronomie Générale*, MIR, 1974.

[Bo70] **Bowen, R.**: *Markov partitions for Axiom A diffeomorphisms*, American Journal of Mathematics, **92**, 725–747, 1970. And: *Markov partitions and minimal sets for Axiom A diffeomorphisms*, American Journal of Mathematics, **92**, 907–918, 1970.

[Bo79] **Boldrighini, C.**: *Introduzione alla fluidodinamica*, Quaderno CNR, Roma, 1979.

[Bo84] **Boltzmann, L.**: *Über die Eigenshaften monzyklischer und anderer damit verwandter Systeme*, in "Wissenshafltliche Abhandlungen", ed. F.P. Hasenöhrl, vol. III, Chelsea, New York, 1968, (reprint).

[Bo97] **Boltzmann, L.**: *Zu Hrn. Zermelo's Abhandlung "Ueber die mechanische Erklärung irreversibler Vorgänge"*, engl. trans. in S. Brush, "Kinetic Theory", **2**, 238.

[BPPV93] **Benzi, R., Paladin, G., Parisi, G., Vulpiani A.**: *Multifractal and intermittency in turbulence*, in "Turbulence in spatially extended systems", ed. R. Benzi, C. Basdevant, S. Ciliberto, Nova Science Publishers, Commack (NY), 1993, p. 163– 188.

[BPV98] **Biferale, L., Pierotti, D., Vulpiani, A.**: *Time–reversible dynamical systems for turbulence*, Journal of Physics A, **31**, 21–32, 1998.

[BS98] **Bricmont, J., Sokal, A.**: *Impostures*, Jakob, Paris, 1999.

[CD82] **Calogero, F., Degasperis, A.**: *Spectral transforms and solitons*, Amsterdam, North Holland, 1982.

[CE80] **Collet, P., Eckmann, J.P.**: *Iterated maps on the interval as dynamical systems*, Birkhauser, Cambridge (MA), 1980.

[CEG84] **Collet, P., Epstein, H., Gallavotti, G** : *Perturbations of geodesic flows on surfaces of constant negative curvature*, Communications in Mathematical Physics, **95**, 61–112, 1984.

[CF88a] **Constantin, P., Foias, C.**: *Navier Stokes Equations*, Chicago Lectures in Mathematics series, University of Chicago Press, 1988.

[CF88b] **Celletti, A., Falcolini, C.**: *A remark on the KAM theorem applied to a four vortex system*, Journal of Statistical Physics, **52**, 471–477, 1988.

[CF93] **Constantin, P., Fefferman, C.**: *Direction of vorticity and the problem of global regularity for the Navier–Stokes equations*, Indiana University Journal of Mathematics, **42**, 775–789, 1993.

[CG99] **Cohen, E.G.D., Gallavotti, G.**: *Note on Two Theorems in Nonequilibrium Statistical Mechanics*, Journal of Statistical Physics, **96**, 1343–1349, 1999.

[Ch82] **Chorin, A.**: *The evolution of a turbulent vortex*, Communications in Mathematical Physics, 83, 517- 535, 1982.

[Ch88] **Chorin A.:** *Hairpin removal in vortex interactions*, UCB preprint, nov. 1988.

[CL98] **Ciliberto, S., Laroche, C.:** *An experimental eib.tex-verification of the Gallavotti–Cohen fluctuation theorem*, Journal de Physique, **8**, 215–222, 1998.

[Co65] **Coles, D.**: *Transition in circular Couette flow*, Journal of Fluid Mechanics, **21**, 385–425, 1965.

[CKN82] **Caffarelli, L., Kohn, R., Nirenberg, L** : *Partial regularity of suitable weak solutions of the Navier- Stokes equations*, Communications on pure and applied mathematics, 35, 771- 831, 1982.

[Cv84] **Cvitanovic, P.** *Universality in chaos*, Adam Hilger, Bristol, 1984.

[CW95] **Craig, W., Worfolk, P.**: *An integrable normal form for water waves in infinite depth*, *Physica D* **84** (1995) pp. 513-531. See also Craig, W.: *Birkhoff normal forms for water waves*, Mathematical Problems in Water Waves, Contemporary Mathematics, AMS (1996), pp. 57 - 74.

[CWT94] **Constantin, P., Weinan E., Titi, E. S** : *Onsager's conjecture on the energy conservation for solutions of Euler's equation*, Communications Mathematical Physics, **165**, 207, 1994.

[DC92] **Doering, C. R., Constantin, P.**, *Energy dissipation in sherar driven turbulence*, Physical Review Letters, **69**, 1648–1651, 1992.

[DGM84] **de Groot, S., Mazur, P.**: *Non equilibrium thermodynamics*, Dover, 1984, (reprinted).

[DLZ95] **Dyachenko, Lvov, Y.V., A.I., Zakharov V.E**: *Five-wave interaction on the surface of deep fluid*, Physica D, **87**, 233–261, 1995.

[DM97] **Dettman, C.P., Morriss, G.P.**: *Proof of conjugate pairing for an isokinetic thermostat*, Physical Review **53 E**, 5545–5549, 1996.

[Dr88] **Dressler, U.**: *Symmetry property of the Lyapunov exponents of a class of dissipative dynamical systems with viscous damping*, Physical Review, **38A**, 2103–2109, 1988.

[DS60] **Dunford, N., Schwartz, I.**: *Linear operators*, Interscience, 1960.

[DS94] **Doliwa, A., Santini, P.**: *An elementary geometric characterization of the integrable motion of a curve*, Physics Letters, **185A**, 373–384, 1994.

[DS95] **Doliwa, A., Santini, P.**: *Integrable dynamics of a discrete curve and the Ablowitz–Ladik hierarchy*, Journal of Mathematical Physics, **36**, 1259–1268, 1995.

[DZ94] **Dyachenko, A.I., Zakharov, V.E.**: *Is free surface hydrodynamics an integrable system?*, Physics Letters **A190**, 144–148, 1994.

[Eb77] **Ebin, D.G.**: *The Motion of Slightly Compressible Fluids Viewed as a Motion With Strong Constraining Force*, Annals of Mathematics, **105**, 141–200, 1977.

[Eb82] **Ebin, D.G.**: *Motion of slightly compressible fluids in a bounded domain. I*, Communications on pure and applied mathematics, **35**, 451–485, 1982.

[Ec81] **Eckmann, J. P.**: *Roads to turbulence in dissipative dynamical systems*, Reviews of Modern Physics, 53, 643- 654, 1981.

[ECM90] **Evans, D.J., Cohen, E.G.D., Morriss, G P.**: *Viscosity of a simple fluid from its maximal Lyapunov exponents*, Physical Review, **42A**, 5990–5997, 1990.

[ECM93] **Evans, D.J.,Cohen, E.G.D., Morriss, G P.**: *Probability of second law violations in shearing steady flows*, Physical Review Letters, **71**, 2401–2404, 1993.

[EE11] **Ehrenfest, P., Ehrenfest, T.**: *The conceptual foundations of the statistical approach in Mechanics*, Dover, New York, 1990, (reprint).

[EH69] **Elmore, W.C., Heald, M.A.**: *Physics of waves*, McGraw Hill, New York, 1969, Reprinted by Dover, 1985.

[EM90] **Evans, D.J., Morriss, G.P.:** *Statistical Mechanics of Nonequilibrium fluids*, Academic Press, New York, 1990.

[EM94] **Esposito, R., Marra, R.**: *On the derivation of the incompressible Navier–Stokes equation for Hamiltonian particle systems*, Journal of Statistical Physics, **74**, 981–1004, 1999.

[ER81] **Eckmann, J.P., Ruelle, D.**: *Ergodic theory of chaos and strange attractors*, Reviews of Modern Physics, 57, 617–656, 1981.

[ES93] **Evans, D.J., Sarman, S.**: *Equivalence of thermostatted nonlinear responses*, Physical Review, **E 48**, 65–70, 1993.

[Ey94] **Eyink, G. L.**: *Energy dissipation without viscosity in ideal hydrodynamics, I*, Physica D, **78**, 222–240, 1994.

[Fe78] **Feigenbaum, M.**: *Quantitative universality for a class of non linear transformations*, Journal of Statistical Physics, **19**, 25, 1978.

[Fe80] **Feigenbaum, M.**: *The transition to aperiodic behavior in turbulent systems*, Communications in Mathematical Physics, **77**, 65, 1980.

[FGN88] **Franceschini, V., Giberti, C., Nicolini M.**: *Common periodic behavior in larger and larger truncations of the Navier Stokes equations*, Journal of Statitsical Physics, **50**, 879–896, 1988.

[FK60] **Furstenberg, H., Kesten, H.**: *Products of random matrices*, Annals of Mathematics and Statistics, **31**, 457–469, 1960.

[FP67] **Foias, G., Prodi, C.**, Sur le comportement g;lobal des solutions des équations de Navier–Stokes, Rnendiconti del Seminaro Matematico di PAdova, **39**, 1–34, 1967.

[FP84] **Frisch, U., Parisi, G.**: in *Turbulence and predictability*, p. 84–87, a cura di M. Ghil, R. Benzi, G. Parisi, North Holland, 1984.

[Fr80] **Franceschini, V.**: *A Feigenbaum sequence of bifurcations in the Lorenz model*, Journal of Statistical Physics, **22**, 397– 406, 1980.

[Fr83] **Franceschini, V.**: *Two models of truncated Navier Stokes equations on a two dimensional torus*, Physics of Fluids, **26**, 433– 447, 1983.

[FSG79] **Fenstermacher, F., Swinney, H., Gollub J.**: *Dynamical instabilities and the transition to chaotic Taylor vortex flow*, Journal of Fluid Mechanics, 94, 103–128, 1979.

[FT79] **Franceschini, V., Tebaldi, C.**: *Sequences of infinite bifurcations and turbulence in a five–mode truncation of the Navier Stokes equations*, Journal of Statistical Physics, **21**, 707- 726, 1979: reprinted in [Cv84]. And *A seven–mode truncation of the plane incompressible Navier Stokes equation*, Journal of Statistical Physics, **25**, 397–417, 1981.

[FZ85] **Franceschini, V., Zironi, F.**: *On constructing Markov partitions by computer*, Journal of Statistical Physics, **40**, 69–91, 1985.

[FZ92] **Franceschini, V., Zanasi, R.**: *Three dimensional Navier Stokes equations truncated on a torus*, Nonlinearity, **4**, 189–209, 1992.

[Ga00] **Gallavotti, G.**: *Non equilibrium in statistical and fluid mechanics. Ensembles and their equivalence. Entropy driven intermittency.* archived in xxx@ lanl. gov: physics # 0001071. And *Fluctuations and entropy driven space–time intermittency in Navier–Stokes fluids*, archived in xxx@ lanl. gov: nlin.CD/0001021. See also *Entropy driven intermitency*, nlin/CD #0003025.

[Ga74] **Gallavotti, G.**: *Modern theory of the billiards. An introduction*, in Global Analysis and its applications, vol. II, ed. P. De la Harpe, International Atomic Energy Commission, Vienna, 1974, p. 193–202.

[Ga75] **Gallavotti, G.**: *Lectures on the billiard*, in "Lecture notes in Physics", vol. **38**, ed. J.Moser, Springer Verlag, 1975, pp. 236–295. See also [Ga74].

[Ga81] **Gallavotti, G.**: *Aspetti della teoria ergodica qualitativa e statistica del moto*, "Quaderni dell' Unione Matematica Italiana", vol. 21, Pitagora editrice, Bologna, 1981. Copies can be obtained by writing to "U.M.I., Dip. Matematica, Università di Bologna, P.zza di Porta S. Donato, 5, 40127, Bologna", or to "Editrice Pitagora, Via Zamboni 57, 40127 Bologna" (Lit. 8000).

[Ga82] **Gallavotti, G.**: *The Dirichlet problem and the Perron–Frobenius theorem*, Bollettino Unione Matematica Italiana, (6), **1B**, 1029– 1038, 1982.

[Ga83] **Gallavotti, G.**: *The elements of Mechanics*, 1983, Springer Verlag (Texts and monographs in Physics).

[Ga86] **Gallavotti, G.**: *Quasi integrable hamiltonian systems*, ed. K. Osterwalder, R. Stora, Les Houches, XLIII, "Phenomènes Critiques, Systèmes aleatoires", p. 539-624, North Holland, 1986.

[Ga92] **Galgani, L.**: *The quest for Planck's constant in classical physics*, in Probabilistic methods in mathematical physics (F.Guerra, M.Loffredo, C.Marchioro eds.), World Scientific (singapore, 1992).

[Ga93] **Gallavotti, G.**: *Some rigorous results about 3D Navier Stokes*, in "Turbulence in spatially extended systems", p.45– 74, ed. R. Benzi, C. Basdevant, S. Ciliberto, Nova Science Publishers, Commack (NY), 1993.

[Ga95a] **Gallavotti, G.**: *Reversible Anosov maps and large deviations*, Mathematical Physics Electronic Journal, MPEJ, (http:// mpej.unige.ch), **1**, 1–12, 1995.

[Ga95b] **Gallavotti, G.**: *Ergodicity, ensembles irreversibility in Boltzmann and beyond*, Journal of Statistical Physics. **78**, 1571–1589, 1995.

[Ga95c] **Gallavotti, G.**: *Topics on chaotic dynamics*, in Third Granada Lectures in *Computational Physics*, Ed. P. Garrido, J. Marro, in Lecture Notes in Physics, Springer Verlag, **448**, p. 271–311, 1995. See also *New methods in nonequilibrium gases and fluids*, Open Systems and Information Dynamics, **6**, 101–136, 1999. Also in mp_arc@ math. utexas.edu #96–533 and chao-dyn #9610018.

[Ga96a] **Gallavotti, G.**: *Extension of Onsager's reciprocity to large fields and the chaotic hypothesis*, Physical Review Letters, **77**, 4334–4337, 1996.

[Ga96] **Gallavotti, G.**: *Dynamical ensembles equivalence in fluid mechanics*, Physica D, **105**, 163–184, 1997, and *Equivalence of dynamical ensembles and Navier Stokes equations*, Physics Letters, **223**, 91–95, 1996.

[Ga97] **Gallavotti, G.**: *Chaotic principle some applications to developed turbulence*, Journal of Statistical Physics, **86**, 907–934, 1997.

[Ga98] **Gallavotti, G.**: *A local fluctuation theorem*, Physica A, **263**, 39–50, 1999. And *Chaotic Hypothesis and Universal Large Deviations Properties*, Documenta Mathematica, extra volume ICM98, vol. I, p. 205–233, 1998, also in chao-dyn 9808004.

[Ga99a] **Gallavotti, G.**: *Statistical Mechanics*, Springer–Verlag, 1999.

[Ga99b] **Gallavotti, G.**: *Quasi periodic motions from Hipparchus to Kolmogorov*, chao-dyn #9907004.

[Ga99c] **Gallavotti, G.**: *Fluctuation patterns and conditional reversibility in non-equilibrium systems*, Annales de l' Institut H. Poincaré, **70**, 429–443, 1999.

[Ga99d] **Gawedzki, K.**: *Easy turbulence*, Lectures at the Banff Summer School, august 1999, chao–dyn # 9907024.

[GBG04] **Gallavotti, G., Bonetto, F., Gentile, G.**: *Aslpects of the ergodic, qualitative and statistical properties of motion*, Springer–Verlag, Berlin, 2004.

[GC95a] **Gallavotti, G., Cohen, E.G.D**: *Dynamical ensembles in nonequilibrium statistical mechanics*, Physical Review Letters, **74**, 2694–2697, 1995.

[GC95b] **Gallavotti, G., Cohen, E.G.D**: *Dynamical ensembles in stationary states*, Journal of Statistical Physics, **80**, 931–970, 1995.

[Ge86] **Germano, M.**: *A proposal for a redefinition of the turbulent stresses in the filtered Navier–Stokes equations*, Physics of Fluids A,**5**, 1282–1284, 1986.

[Ge98] **Gentile, G.**: *A large deviation theorem for Anosov flows*, Forum Mathematicum, **10**, 89–118, 1998.

[GGP69] **Gelfand, I., Graev, M., Piateckii-Shapiro I.**: *Representation theory and automorphic functions*, Saunders, Philadelphia, 1969.

[Gh60] **Ghirsanov, V.**: *On trasforming a certain class of stochastic processes by absolutely continuos substitution of measures*, Theory of Probability and Applications, **5**, 285-301 (1960).

[GPMC91] **Germano, M., Piomelli, U., Moin. P., Cabot, W.**: *A dynamic subgrid–scale eddy viscosity model*, Physics of Fluids A, **3**, 1760–1765, 1991.

[GR97] **Gallavotti, G., Ruelle, D.:** *SRB states and nonequilibrium statistical mechanics close to equilibrium*, Communications in Mathematical Physics, **190**, 279–285, 1997.

[Gu90] **Gutzwiller, M.C.**: *Chaos in clasical and quantum mechanics*, Springer–Verlag, 1990.

[GZ93] **Giberti, C., Zanasi, R.**: *Behavior of a three–torus in truncated Navier Stokes equations*: Physica, **65D**, 300–312, 1993.

[Ha72] **Hasimoto, H.**: *A soliton on a vortex filament*, Journal of Fluid Mechanics, **51**, 477–485, 1972.

[He83] **Herman, M.**: *Une méthode pour minorer les exposants de Lyapunov et quelques examples montrant le caractère local d' un théorème d' Arnold et de Moser sur le tore de dimension* 2, Commentarii Mathematici Helvetici, **58**, 453–502, 1983.

[HHP87] **Hoolian, B.L., Hoover, W.G., Posch, H A.**: *Resolution of Loschmidt's paradox: the origin of irreversible behaviour in reversible atomistic dynamics*, Physical Review Letters, **59**, 10–13, 1987.

[Ho99] **Hoover, W. G.**: *Time reversibility Computer simulation, and Chaos*, World Scientific, 1999.

[IM65] **Ito, K., McKean, H. P.**: *Diffusion processes and their sample paths*, Academic Press, 1965.

[Ka67] **Kato, T.**: *On classical solutions of the two dimensional non stationary Euler equation*, Archive for Rational Mechanics and Analysis, **25**, 188–200, 1967.

[Ka76] **Katznelson, Y.**: *An introduction to harmonic analysis*, Dover, 1976.

[Ki49] **Kintchin, A.I.**: *Mathematical foundations of Statistical Mechanics*, Dover, 1949.

[Ki57] **Kintchin, A.I.**: *Mathematical foundations of information theory*, Dover, 1957.

[Ki63] **Kintchin, A.I.**: *Continued fractions*, Noordhoff, Gröningen, 1963.

[KM81] **Kleinerman, S., Majda, A.**: *Singular limits of quasilinear hyperbolic flows with large parameters and the incompressible limit of compressible fluids*, Communications in Pure and Applied Mathematics, **34**, 481– 524, 1981.

[KM82] **Kleinerman, S., Majda, A.**: *Compressible and incompressible flows*, Communications in Pure and Applied Mathematics, **35**, 629– 651, 1982.

[KY79] **Kaplan, J.L., Yorke, J.L.**: *Preturbulence a regime observed in a fluid flow model of Lorenz*, Communications in Mathematical Physics, **63**, 93–108, 1979.

[Kr64] **Kraichnan, R.H.**: *Kolmogorov's hypotheses and eulerian turbulence theory*, Physics of Fluids, **7**, 1723–1734, 1964.

[Kr67] **Kraichnan, R.H.**: *Inertial ranges in two–dimensional turbulence*, Physics of fluids, **10**, 1417–1423, 1967.

[Kr75a] **Kraichnan, R.H.**: *Remarks on turbulence*, Advances in Mathematics, **16**, 305–331, 1975.

[Kr75b] **Kraichnan, R.H.**: *Statistical dynamics of two–dimensional turbulence*, Journal of Fluid Mechanics, **67**, 155–175, 1975.

[KS00] **Kuksin, S., Shirikyan, A.**: *Stochastic dissipative PDE's and Gibbs measures*, in print, Communications in Mathematical ¡ Physics, 2000.

[LA27] **Levi-Civita, T., Amaldi, U.**: *Lezioni di Meccanica Razionale*, Zanichelli, Bologna, 1927 (reprinted 1974), volume 3.

[La32] **Lamb, H.**: *Hydrodynamics*, sixt edition, Cambridge University Press, 1932.

[Le34] **Leray, J.**: *Sur le mouvement d' un liquide visqueux remplissant l' espace*, Acta Mathematica, 63, 193- 248, 1934.

[Le81] **Ledrappier, F.**: *Some relations between dimension and Lyapunov exponents*, Communications in Mathematical Physics, **81**, 229, 1981.

[Le93] **Lebowitz, J.L.**: *Boltzmann's entropy and time's arrow*, Physics Today, 32–38, september 1993. And *Microscopic Reversibility and Macroscopic Behavior: Physical Explanations and Mathematical Derivations*, in 25 years of non-equilibrium statistical me-

chanics, ed. J. Brey, J. Marro, J. Rubi, M. San Miguel, *Lecture Notes in Physics*, **445**, Springer, Berlin, 1995.

[LFL83] **Libchaber, A., Fauve, S., Laroche, C** : *Two parameter study of the routes to chaos*, Physica D, **7**, 73–84, 1983.

[Li84] **Lieb, E.**: *On characteristic exponents in turbulence*, Communications in Mathematical Physics, **92**, 473–480, 1984.

[LL71] **Landau, L., Lifchitz, E.**: *Mécanique des fluides*, MIR, Moscow, 1971.

[LL01] **Lieb, E., Loss, M.**: *Analysis*, American Mathematical Society, second edition, Providence, 2001.

[LM72] **Lyons, RJ.L., Magenes, E.**: *Non-homogeneous boundary value problems and applications*, Springer-Verlag, 1972.

[La74] **Lanford, O.**: *Time evolution of large classical systems*, in "Dynamical systems, theory and applications", p. 1–111, ed. J. Moser, Lecture Notes in Physics, vol. 38, Springer Verlag, 1974

[Lo63] **Lorenz, E.**: *Deterministic non periodic flow*, J. of the Athmospheric Sciences, 20, 130-141, 1963, reprinted in [Cv84].

[LR93] **Laskar, J., Robutel, P.**: *The chaotic obliquity of the planets*, Nature, **361**, 608–612, 1993.

[LY73] **Lasota, A., Yorke, J.A.**: *On the existence of invariant measures for piecewise monotonic transformations*, Transactions American Mathematical Society, **186**, 481–488, 1973.

[Ma86] **Marchioro, C.**: *An example of absence of turbulence for any Reynolds number*, Communications in mathematical Physics, **105**, 99–106, 1986.

[Ma87] **Marchioro, C.**: *An example of absence of turbulence for any Reynolds number: II*, Communications in mathematical Physics, **108**, 647–651, 1987.

[MD00] **Murray, C.D., Dermott, S.F.**: *Solar system dynamics*, Cambridge University Press, 2000.

[Me90] **Metha, M., L.**: *Random matrices*, Academic Press, 1990.

[Me97] **Mermin, D.** : *What's wrong with this reading*, Physics Today, **50**, 11–13, 1997. See also *Sokalratic debate continues, fueled by Latour and Copenhagen ineterpretations*, Physics Today, **52**, 15+82–83, 1999.

[Mi70] **Miranda, C.**: *Partial differential equations of elliptic type*, Springer Verlag, Berlino, 1970.

[MK00] **Meneveau, C., Katz, J.**: *Scale–Invariance and turbulence models for large-eddy simulation.* Annual Review of Fluid Mechanics, **32**, 1–32, 2000.

[MP72] **Marchioro, C., Presutti, E.**: *Thermodynamics of particle systems in the presence of external macroscopic fields: I classical case*, Communications in Mathematical Physics, **27**, 146– 154, 1972.

[MP84] **Marchioro, C., Pulvirenti, M.**: *Vortex methods in two dimensional fluid dynamics*, Lecture N. in Math., vol. 203, Springer Verlag, 1984.

[MP92] **Marchioro, C., Pulvirenti, M.**: *Mathematical theory of incompressible non viscous fluids*, in stampa, 1992.

[MS99] **Mattingly, J.C., Sinai Ya.G.**: *An elementary proof of existence and uniqueness theorem for the Navier Stokes Equations*, Communications in Contemporary Mathematics **1**, 497–516, 1999.

[Ne64] **Nelson, E.**: *Feynman integrals and the Schrödinger equation*, Journal of Mathematcial Physics, **5**, 332–343, 1964.

[OM53] **Onsager, L., Machlup, S.**: *Fluctuations and irreversible processes*, Physical Review, **91**, 1505–1512, 1953. And **Machlup, S., Onsager, L.**: *Fluctuations and irreversible processes*, Physical Review, **91**, 1512–1515, 1953.

[On49] **Onsager, L.**: *Statistical hydrodynamics*, Supplemento Nuovo Cimento, **6**, 279–287, 1949.

[Os68] **Oseledec, V.**: *A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems*, Transactions of the Moscow Mathematical Society, **19** 197–221, 1968.

[Pe76] **Pesin, Y.**: *Lyapunov characteristic exponents and ergodic properties of smooth dynamical systems with an invariant measure*, Soviet Mathematics Doklady, **17**, 196–199, 1976. E più in dettaglio, *Invariant manifold families which correspond to non vanishing characteristic exponents*, Math. USSR Izvestia, **10**, 1261–1305, 1976.

[Pe84] **Pesin, Y.**: *On the notion of dimension with respect to a dynamical system*, Ergodic theory and dynamical systems, **4**, 405–420, 1984.

[Pe92] **Pesin, Y.**: *Dynamical systems with generalized hyperbolic attractors: hyperbolic, ergodic and topological properties*, Ergodic theory and dynamical systems, **12**, 123–151, 1992.

[PM80] **Pomeau, Y., Manneville, P.**: *Intermittent transition to turbulence in dissipative dynamical systems*, Communications in Mathematical Physics, **74**, 189–197 ,1980, reprinted in [Cv84].

[PS87] **Pumir, A., Siggia, E.**: *Vortex dynamics and the existence of solutions to the Navier Stokes equations*, Physics of Fluids, **30**, 1606–1626, 1987.

[Ra79] **Raghunathan, M.**: *A proof of Oseledec's multiplicative ergodic theorem*, Israel Journal of Mathematics, **32**, 356–362, 1979.

[Ri82] **Riela, G.**: *A new six mode truncation of the Navier Stokes equations on a two dimensional torus: a numerical study*, Nuovo Cimento, **69B**, 245, 1982.

[RS72] **Reed, M., Simon, B.**: *Methods of moderm mathematical Physics. Functional analysis*, Academic Press, New York, 1972.

[RS99] **Rondoni, L., Segre, E.**: *Fluctuations in two dimensional reversibly damped turbulence*, Nonlinearity, **12**, 1471–1487, 1999.

[RT71] **Ruelle, D., Takens, F.**: *On the nature of turbulence*, Communications in mathematical Physics, **20**, 167, 1971. And *Note concerning our paper "On the nature of turbulence"*, Communications in Mathematical Physics, **23**, 343–344, 1971: this note is important as it brings to attention several references that were not well known at the time.

[Ru68] **Ruelle, D.**: *Statistical mechanics of one–dimensional lattice gas*, Communications in Mathematical Physics, **9**, 267–278, 1968.

[Ru76] **Ruelle, D.**: *A measure associated with Axiom A attractors*, American Journal of Mathematics, **98**, 619–654, 1976.

[Ru78] **Ruelle, D.**: *Thermodynamic formalism*, Addison Wesley, 1978.

[Ru79] **Ruelle, D.**: *Ergodic theory of differentiable dynamical systems*, Publications Mathématiques de l' IHES, **50**, 27–58, 1979. E: *Characteristic exponents and invariant manifolds in Hilbert space*, Annals of Mathematics, **115**, 243–290, 1982.

[Ru80] **Ruelle, D.**: *Measures describing a turbulent flow*, Annals of the New York Academy of Sciences, **357**, 1–9, 1980.

[Ru82] **Ruelle, D.**: *Large volume limit of the distribution of characteristic exponents in turbulence*, Communications in Mathematical Physics, **87**, 287–302, 1982.

[Ru84] **Ruelle, D.**: *Characteristic exponents for a viscous fluid subjected to time dependent forces*, Communications in Mathematical Physics, **93**, 285–300, 1984.

[Ru89b] **Ruelle, D.**: *Elements of differentiable dynamics and bifurcation theory*, Academic Press, 1989.

[Ru89] **Ruelle, D.**: *Chaotic evolution and strange attractors*, Accademia Nazionale dei Lincei, Roma, 1987 (notes by S. Isola), printed by Cambridge University Press, 1989.

[Ru96] **Ruelle, D.**: *Positivity of entropy production in non equilibrium statistical mechanics*, Journal of Statistical Physics, **85**, 1-25, 1996. And *Positivity of entropy production in the presence of a random thermostat*, Journal of Statistical Physics, **86**, 935–951, 1997.

[Ru97] **Ruelle, D.**: *Entropy production in nonequilibrium statistical mechanics*, Communications in Mathematical Physics, **189**, 365–371, 1997.

[Ru99a] **Ruelle, D.**: *A remark on the equivalence of isokinetic and isoenergetic thermostats in the thermodynamic limit*, Journal of Statistical Physics, **100**, 757–763, 2000.

[Ru99b] **Ruelle, D.**: *Natural nonequilibrium states in quantum statistical mechanics*, Journal of Statistical Physics, **98**, 57–75, 2000.

[Ru99c] **Ruelle, D.**: *Smooth dynamics and new theoretical ideas in non-equilibrium statistical mechanics*, Journal of Statistical Physics, **95**, 393–468, 1999.

[Sa62] **Saltzman, B.**: *Finite amplitude free convection as an initial value problem*, Journal of Atmospheric Science, **19**, 329- 341, 1962.

[Sa86] **Saccheri, G.**: *Euclides ab omni naevo vindicatus*, english translation, Chelsea, New York, 1986.

[Sc77] **Scheffer, V.**: *Hausdorff dimension and the Navier Stokes equations*, Communications in Mathematical Physics, 55, 97- 112, 1977. And *Boundary regularity for the Navier Sokes equation in half space*, Communications in Mathematical Physics, **85**, 275- 299, 1982.

[Se64] **Seeley, R.**: *Extensions of C^∞ functions defined in half space*, Proceedings American Mathematical Society, **15**, 625- 626, 1964.

[SG78] **Swinney, H., Gollub, J., P.**: *The transition to turbulence*, Physics Today, **31**, 41–49, 1978.

[Si68] **Sinai, Y.G.**: *Markov partitions and C-diffeomorphisms*, Functional Analysis and Applications, **2**, 64–89, 1968, n.1; and: *Construction of Markov partitions*, Functional analysis and Applications, **2**, 70–80, 1968, n.2.

[Si70] **Sinai, Y.G.**: *Dynamical systems with elastic reflections. Ergodic properties of dispersing billards*, Russian Mathematical Surveys, **25**, 137–189, 1970.

[Si72] **Sinai, Y.G.**: *Gibbs measures in ergodic theory*, Russian Mathematical Surveys, **27**, 21–69, 1972.

[Si77] **Sinai, Y.G.**: *Lectures in ergodic theory*, Lecture notes in Mathematics, Princeton U. Press, Princeton, 1977.

[Si79] **Sinai, Y.G.**: *Development of Krylov's ideas*, in Krylov, N.S.: *Works on the foundations in statistical physics*, Princeton University Press, 1979, p. 239–281.

[Si91] **Sinai Ya.G.**: *Mathematical problems of statistical mechanics*, World Scientific, 1991.

[Si94] **Sinai. Y.G.**: *Topics in ergodic theory*, Princeton U. Press, 1994.

[SJ93] **She, Z.S., Jackson, E.**: *Constrained Euler system for Navier Stokes turbulence*, Physical Review Letters, **70**, 1255–1258, 1993.

[Sm67] **Smale, S.**: *Differentiable dynamical systems*, Bullettin of the American Mathematical Society, **73** , 747–818, 1967.

[SM99] **Scotti, A., Meneveau, C.**: A fractal model for large eddy simulation of turbulent flow, Physica D, **127**, 198–232, 1999.

[So63] **Sobolev, S.L.**: *Applications of Functional analysis in Mathematical Physics*, Translations of the American Mathematical Society, vol 7, 1963, Providence.

[St93] **Stein, E.**: *Harmonic analysis*, Princeton University press, 1993.

[Ta35] **Taylor, G.I.**: , Proceedings of the Royal Society of London, **A151**, 421, 1935.

[VW93] **van de Water, W.**: *Experimental study of scaling in fully developed turbulence*, in "Turbulence in spatially extended systems", ed. R. Benzi, C. Basdevant, S. Ciliberto, Nova Science Publishers, Commack (NY), 1993.

[Wa90] **Waleffe, F.**: *The nature of triad interactions in homogeneous turbulence*, Physics of fluids, **4A**, 350–363, 1990.

[Wi87] **Wisdom, J.**: *Chaotic behavior in the solar system*, Proceedings of the Royal Society of London, **A 413**, 109–129, 1987.

[Wi89] **Wittaker, E.T.**: *A treatise on the analytic dynamics of particles & rigid bodies*, Cambridge University Pres, 1989, (reprint).

[WL98] **Woitkowsky, M.P., Liverani, C.**: *Conformally Symplectic Dynamics and Symmetry of the Lyapunov Spectrum*, Communications in Mathematical Physics, **194**, 47–60, 1998.

[Yo82] **Young, L.S.**: *Dimension, entropy and Liapunov exponents*, Ergodic theory and dynamical systems, **2**, 109, 1982.

[YO86] **Yakhot, V., Orszag, S.A.**: *Renormalization group analysis of turbulence. I. Basic theory*, Journal of Scientific Computing, **1**, 3–51, 1986.

Name index

Subject index

Citations index