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 $t \geq 0$ 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

$$\begin{aligned}
A(p) &= \lim_{T \to \infty} A_T(p) = \\
&= \lim_{T \to \infty} \frac{1}{2} \sum_{\underline{n}, \underline{n}' \in Z^\ell} \Phi_{\underline{n}} \overline{\Phi}_{\underline{n}'} \, \frac{1}{T} \frac{e^{i(\underline{\omega} \cdot \underline{n} - p)T} - 1}{i(\underline{\omega} \cdot \underline{n} - p)} \frac{e^{-i(\underline{\omega} \cdot \underline{n}' - p)T} - 1}{-i(\underline{\omega} \cdot \underline{n}' - p)} = \\
&= \lim_{T \to \infty} 2 \sum_{\underline{n}, \underline{n}' \in Z^\ell} \Phi_{\underline{n}} \overline{\Phi}_{\underline{n}'} \, e^{i \underline{\omega} \cdot (\underline{n} - \underline{n}') \frac{T}{2}} \frac{1}{T} \frac{\sin(\underline{\omega} \cdot \underline{n} - p) \frac{T}{2}}{i(\underline{\omega} \cdot \underline{n} - p)} \frac{\sin(\underline{\omega} \cdot \underline{n}' - p) \frac{T}{2}}{-i(\underline{\omega} \cdot \underline{n}' - p)} = \\
&= \pi \sum_{\underline{n}} |\Phi_{\underline{n}}|^2 \, \delta(p - \underline{\omega} \cdot \underline{n})
\end{aligned} \tag{5.1.6}$$

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 $a_j\ge1$ 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

$$\underline{v}_k=\begin{pmatrix}a_0&1\\1&0\end{pmatrix}\begin{pmatrix}a_1&1\\1&0\end{pmatrix}\cdots\begin{pmatrix}a_k&1\\1&0\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}$$

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 $q_k\ge2^{(k-1)/2}$ for $k\ge0$ and $p_k\ge2^{(k-1)/2}$ 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 $s\ge 0$, 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 $(q_jq_{j-1})^{-1}>|\alpha_{j-1}-r|>|\alpha_{j-1}-\alpha|=|p_{j-1}q-q_{j-1}p|/qq_{j-1}\ge 1/qq_{j-1}$, because $|p_{j-1}q-q_{j-1}p|\ge 1$. 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:* : $q|\alpha-r|>q|\alpha-\alpha_{j+1}|=q|pq_{j+1}-qp_{j+1}|/qq_{j+1}\ge 1/q_{j+1}\ge q_j|\alpha_j-r|$).

[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 $\eta_0(q) = \varepsilon_n = |q_n r - p_n|$ for $q_n \le q < q_{n+1}$.

[5.1.18]: Let n even: the point $q_n r \bmod 1$ 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 $q_{n-1}r + aq_n r$ 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* $n\, r \bmod 1$, $n = 1, 0, \ldots, , T$ is:

$$\begin{aligned} q_n \le& T < q_n + q_{n-1} && \varepsilon_{n-1} \\ q_n + q_{n-1} \le& T < 2q_n + q_{n-1} && \varepsilon_{n-1} - \varepsilon_n \\ &\ldots && \\ (a_{n+1}-1)q_n \le& T < a_{n+1}q_n + q_{n-1} \equiv q_{n+1} && \varepsilon_{n-1} - (a_{n+1}-1)\varepsilon_n \end{aligned}$$

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 $a_j \equiv 1$. 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: $|q_n r - p_n| > 1/Cq_n$ 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: $A_n r_n^2 + B_n r_n + C_n = 0$. 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 $|\hat\chi(\underline{w})| \le \Gamma_\alpha^\ell |\underline{w}|^{-\alpha}$ 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 $a_1 \ge 1$ 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 $a_2 \ge 1$ of iterations is reached (a check that $a_2 \ge 1$ 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

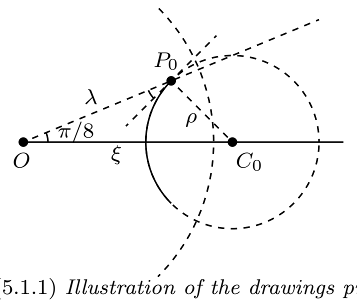
$$z' = \frac{a}{b}\frac{(zb + bc/a)}{zb + d} = \frac{a}{b} - \frac{(ad - bc)/b^2}{z + d/b} \qquad (*)$$

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 $z' = \frac{az+c}{\bar{c}z+\bar{a}}$, 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).



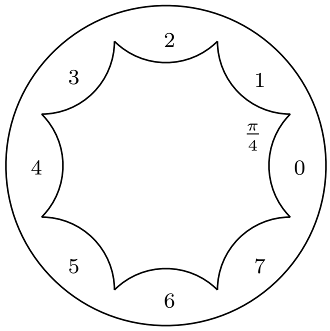


Fig. (5.1.1) *Illustration of the drawings proposed in* [5.1.34] and [5.1.35].

[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].

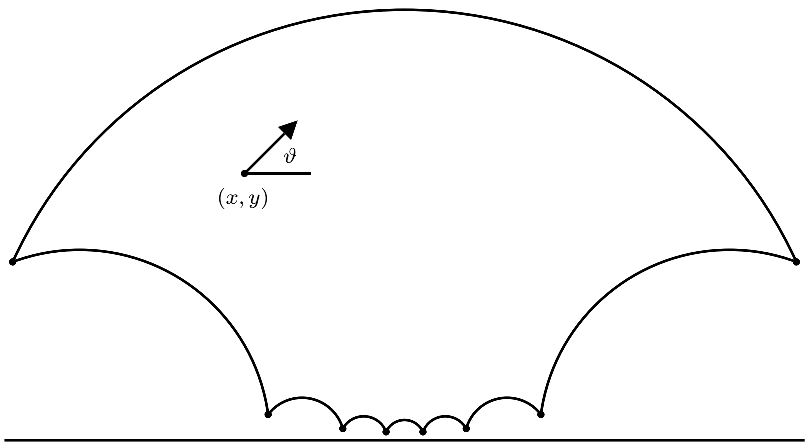


FIg. (5.1.2): *A fundamental domain in the Lobatchevsky plane: it is the image of the domain in the front cover. The (x,y,ϑ) denotes a light ray at (x,y) and direction ϑ.*

[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, $n \geq 0$ integer, contain "all informations concerning the property $\mathcal{P}$" that can be found in the transformations S_t, with $t \geq 0$ 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: $k \to f(S^k x)$. Define:

1 Definition *(power spectrum for discrete time evolutions):*
(1) We define the "discrete autocorrelation function" *of the observable f on the motion $x \to S^n x$ to be the average $\Omega(k, x)$ of the product $f(S^{h+k}x)$ times $f(S^h x)$ 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 $x \to S^n x$ 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 $n \to S^n x$ 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 $x \to S^n x$ is, setting aside problems (often nontrivial) of exchange of limits, the limit

$$\lim_{N\to\infty} \frac{1}{2N} \left| \sum_{k=0}^{N-1} f(S^k x)\, e^{ipk} \right|^2 \tag{5.2.2}$$

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 $f(S^k x)$ 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 $k \to f(S^k x)$ is quasi periodic, *i.e.* if $f(S^k x) = \varphi(k\underline{\gamma})$ 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 $x = (\varphi_1, \varphi_2), Sx = (\varphi'_1, \varphi'_2)$.

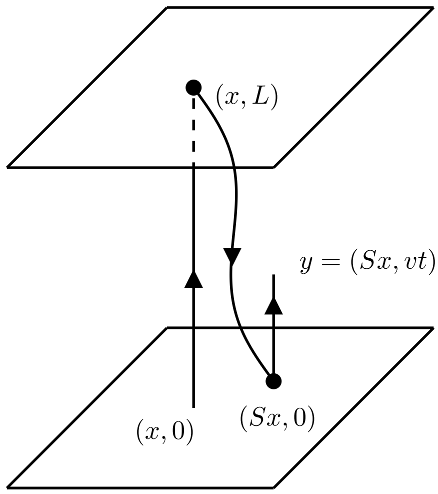


Fig.(5.2.1) *A continuous dynamical system with an embedded "Arnold's cat". The curved line is just a visual aid to the identification of (x, L) with $(Sx, 0)$.*

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 $x_k = Sx_{k-1}$, $k \ge 1$). 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 $x_k = Sx_{k-1}$, $k \ge 1$ 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*

$$\big|\frac{1}{N}\sum_{i=1}^{N} f_k(x_i) - \int_Q \rho(x)\, f_k(x)\, dx\big| < \varepsilon \quad \text{for each} \quad k = 1, \ldots, n \qquad (5.2.3)$$

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)): $|n_1 v_{\pm,1} + n_2 v_{\pm,2}| \ge D/(|n_1| + |n_2|)$ 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 $|f_{C^k\underline{n}}| \le \frac{|\underline{n}|^3 F}{(\lambda_+^k D)^3}$, and also $|f_{C^k\underline{n}}| < F'$ if F' is the maximum of $|f|$.
It follows that $|f_{\underline{n}} f_{C^k\underline{n}}| \le \frac{F}{|\underline{n}|^3} F'^{1-\varepsilon}(|\underline{n}|^3 F D^{-3}\lambda_+^{-3k})^\varepsilon)$ 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

$$C_{F,G}(k)\equiv\int_M\mu(du)\,F(S^ku)G(u)\xrightarrow[k\to\infty]{}\big(\int_M\mu(du)\,F(u)\big)\big(\int_M\mu(du)\,G(u)\big)\tag{5.3.4}$$

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 neverthelsess there is a non empty set of zero measure in phase space consisting of points for which the collision with U_ε is recurrent.

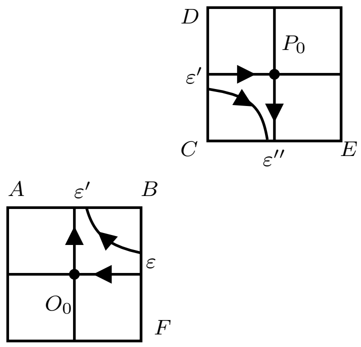


Fig. (5.3.1) *Illustration of the construction in* [5.3.6].

[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 $b \neq a$ 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 $E_n = E + n\,r \,\mathrm{mod}\,1$, $n \in Z$, 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* $n \geq 0$

$$\begin{aligned}|S^n(x+dx) - S^n(x)| \equiv& |\partial S^n(x) \cdot dx| < Ce^{-\lambda n}|dx|, \quad \text{if} \quad dx \in R^s(x)\\ |S^n(x+dx) - S^n(x)| \equiv& |\partial S^n(x) \cdot dx| > Ce^{\lambda n}|dx|, \quad \text{if} \quad dx \in R^i(x)\end{aligned} \tag{5.4.1}$$

(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 $Sx_{k-1}=x_k$ 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

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \log \frac{|\partial S^n d\xi_1|}{|d\xi_1|} = \lambda_1(x),\\
&\lim_{n\to\infty} \frac{1}{n} \log \frac{|\partial S^n d\xi_1 \wedge d\xi_2|}{|d\xi_1 \wedge d\xi_2|} = \lambda_1(x) + \lambda_2(x), \qquad (5.4.6)\\
&\lim_{n\to\infty} \frac{1}{n} \log \frac{|\partial S^n d\xi_1 \wedge d\xi_2 \wedge d\xi_3|}{|d\xi_1 \wedge d\xi_2 \wedge d\xi_3|} = \lambda_1(x) + \lambda_2(x) + \lambda_3(x), \quad \ldots
\end{aligned}$$

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 $d\xi_1 \notin V_2$, the second if $d\xi_1 \notin V_2$ and $d\xi_2 \notin V_3$, *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 $d(S^ky, S^kx) \le C(x)e^{-\lambda_j k}$, $k \ge 0$.

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 $d(S^nx, N) > 0$: 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 $d(S^nx, N) > C(x)n^{-\gamma'}$. (*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.* $m^{-1}\sum_{j=0}^{m-1} f(S^jx) \ge 0$ 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: $\overline{f}(x) = \lim_{n\to\infty} n^{-1}\sum_{j=0}^{n-1} f(S^jx)$. (*Idea:* Let $f_{sup}(x) = \limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(S^jx)$ and likewise define $f_{inf}(x) = \liminf_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(S^jx)$; 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

$$\limsup_{n\to\infty} n^{-1}\log||T_n|| \le 0, \qquad \lim_{n\to\infty} n^{-1}\log||T^{n\wedge q}|| = \lambda_q \qquad q = 1,2,\ldots,\ell \qquad (*)$$

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 $\lim \frac{1}{k}\log t_k^{(q)} = \mu^{(i)}$. 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 $\frac{1}{k}\log t_k^{(p)} \xrightarrow[k\to+\infty]{} \mu^{(i)}$.

[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 $\delta_1 > 0$,

$$\begin{aligned}
|(u,u')| \le& \max_{all\ u'',||u''||=1} e^{-(n+1)\mu^{(r')}} |(u, \sum_{r(i)\ge r'} e^{(n+1)\mu^{(r')}} P_i u'')| = \\
=& e^{-(n+1)\mu^{(r')}} \max_{all\ u'',||u''||=1} |(\sum_{r(i)\ge r} e^{(n+1)\mu^{(r')}} P_i u, u'')| \le \\
\le& e^{-(n+1)\mu^{(r')}} ||\sum_{r(i)\ge r} e^{(n+1)\mu^{(r')}} P_i u|| \le \\
\le& e^{-(n+1)\mu^{(r')}+(n+1)\delta_1} ||\sum_{r(i)\ge r} t_{n+1}^{(i)} P_i u|| \le \\
\le& e^{-(n+1)\mu^{(r')}+(n+1)\delta_1} ||(T^{(n+1)*}T^{(n+1)})^{1/2}u||
\end{aligned}$$

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 $(u_\alpha, u'_\beta) = U_{\alpha\beta} \equiv \overline{U_{\beta\alpha}^{-1}}$ 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 $|z|\ne 1$, and let $\nu(z)=2$ if $z>0$ and $\nu(z)=1$ if $z<0$; define

$$x'=(z',\varphi_1'\varphi_2')=\begin{cases} f(z)\\ \varphi_1+\nu(z)\varphi_2 \quad \bmod 2\pi\\ \nu(z)\varphi_1+(\nu(z)^2+1)\varphi_2 \quad \bmod 2\pi\end{cases}$$

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

$$\tilde{\mathcal{M}}(ze^{-in\omega}) = \begin{pmatrix} \frac{\lambda}{2} + \frac{\lambda}{2} z^2 e^{-2in\omega} - E\, z\, e^{-in\omega} & ze^{-i\omega n} \\ -ze^{-i\omega n} & 0 \end{pmatrix} \qquad \textit{and}$$
$$\tilde{\mathcal{M}}_N(z) = \tilde{\mathcal{M}}(ze^{-i(N-1)\omega}) \ldots \tilde{\mathcal{M}}(z)$$

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.)

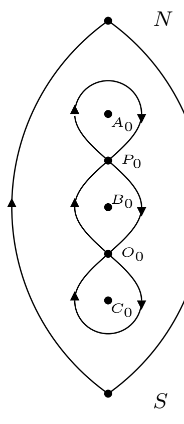


Fig. (5.4.1) Illustration of the one of the four quarters of the sphere in the construction of a non normal attracting set, *c.f.r.* problem [5.4.28].

[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 mod 2π, 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

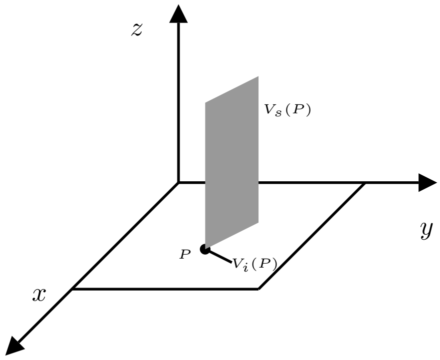


Fig. (5.5.1): The grey surface is a piece of the stable plane $V_s(P)$ of P while the segment is a piece of the unstable plane $V_i(P)$.

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

$W^{\delta,s}_x$ passing through x and having at each of its points y in A the plane $V_s(y)$ as tangent plane. Therefore $W^{\delta,s}_x$ 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 $W^{\delta,s}_x$ and $V_s(x)$ are identical up to second order in the distance from x.

Likewise one can define a portion of regular surface $W^{\delta,i}_x$ having everywhere $V_i(y)$ as tangent plane at each of its points y, and $W^{\delta,i}_x$ 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 $\bmod 2\pi$) and they cannot any longer be regarded as graphs.

(2) The surfaces $W^{\delta,s}_x$ and $W^{\delta,i}_x$ are such that if $y \in W^{\delta,s}_x$ 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 $y \in W^{\delta,i}_x$ 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 $W^{\delta,s}_x, W^{\delta,i}_x$ 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 $W^{\delta(x),s}_x$ and $W^{\delta(x),i}_x$ 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 $n \to -\infty$.

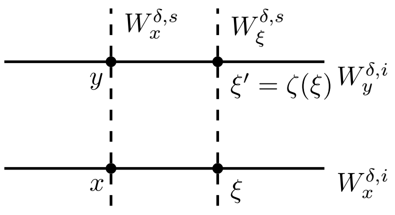


Fig. (5.5.2): The correspondence ζ between two nearby portions $W^{\delta,i}_x$ and $W^{\delta,i}_y$ of unstable manifolds established by the stable manifolds $W^{\delta,s}_\bullet$ that intersect them. The points x, y and ξ, ξ' are two pairs of corresponding points. The segment $\delta^o_x = x\xi$ is 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

$$\begin{aligned}&\frac{\text{mass}(\delta^0_x)}{\text{mass}(\delta^0_y)} \stackrel{def}{=} \lim_{N'\to\infty} \frac{\delta^0_x \prod_{j=0}^{N'} e^{-\lambda_i(S^{-j}x)}}{\delta^0_y \prod_{j=0}^{N'} e^{-\lambda_i(S^{-j}y)}} \equiv \\ &\equiv \lim_{\substack{N\to\infty \\ N'\to\infty}} \frac{\prod_{j=-N'}^{N} e^{-\lambda_i(S^jx)}\delta^N_{S^Nx}}{\prod_{j=-N'}^{N} e^{-\lambda_i(S^jy)}\delta^N_{S^Ny}} \equiv \prod_{j=-\infty}^{\infty} \frac{e^{-\lambda_i(S^jx)}}{e^{-\lambda_i(S^jy)}}\end{aligned} \tag{5.5.4}$$

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 $\delta^N_{S^Nx}$ is the segment of unstable manifold image of δ^0_x under the action of S^N, and a similar notation is used for $\delta^N_{S^Ny}$.

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 δ^0_x δ^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 $W^{\delta,i}_x, W^{\delta,s}_x$ on an attractive set A. Let the set Δ be a "parallelogram" consisting in the union of the surfaces $W^{\delta,i}_\xi$ that pass through the points $\xi \in W^{\delta/2,s}_x$ (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*

$$\int_\Delta f(y)\mu(dy) = \int_{W^s_x \cap A} \nu(d\xi) \int_{W^i_\xi} f(\xi,\sigma)\rho_\xi(\sigma)d\sigma \tag{5.5.5}$$

where ν is a suitable probability distribution concentrated on $W^{\delta/2,s}_x \cap A$, $d\sigma$ is the surface element on $W^{\delta,i}_\xi$, a point $y \in \Delta$ located on the surface element $d\sigma \in W^{\delta,i}_\xi$ is denoted (ξ,σ) and $\rho_\xi(\sigma)$ is a suitable function called the density on the unstable manifold.

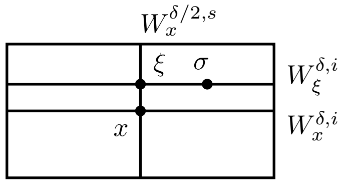


Fig.(5.5.3): *The rectangle represents Δ and the figure illustrates (5.5.5). The horizontal segments represent portions of unstable manifolds and the vertical ones portions of stable manifolds.*

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

$$\frac{\rho_\xi(\sigma)}{\rho_{\xi'}(\sigma)} = \prod_{j=-\infty}^{\infty}\frac{e^{-\lambda_i(S^{-j}x)}}{e^{-\lambda_i(S^{-j}y)}}, \quad \text{and} \quad \frac{\rho_\xi(\sigma)}{\rho_\xi(\sigma')} = \prod_{j=0}^{\infty}\frac{e^{-\lambda_i(S^{-j}x)}}{e^{-\lambda_i(S^{-j}x')}} \tag{5.5.6}$$

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* $\operatorname{per}_n(A)$ *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

$$\langle f\rangle_n = \frac{\sum_{x\in \operatorname{per}_n(A)} \Lambda_i(x)^{-1} f(x)}{\sum_{x\in \operatorname{per}_n(A)} \Lambda_i(x)^{-1}} \tag{5.5.7}$$

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 $\delta(y) \geq \frac{1}{2}\delta(x)$.
(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 $d(S^n x, S^n y) \leq C(x)e^{-\lambda n}$ for $n \geq 0$, likewise if all the nonnegative exponents are $> \lambda > 0$ then for each $y \in W_x^{\delta(x),i}$ it is $d(S^{-n}x, S^{-n}y) \leq C(x)e^{-\lambda n}$ for $n \geq 0$, 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 $d(S^n x, S^n y) \xrightarrow[n\to+\infty]{} 0$: 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.)

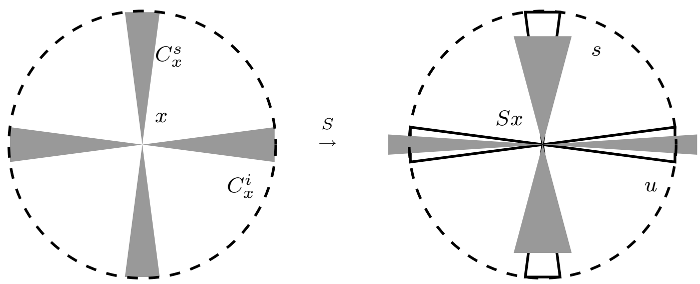


Fig. (5.5.4): *Symbolic representation of the cone property of the map S for the stable (vertical) manifold and for the unstable (horizontal) manifold.*

[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 $S^kx\in P_{\sigma_k}$. 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^kx 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 $P_0 = [0, 0.1]$, $P_1 = [0.1, 0.2]$, …, $P_9 = [0.9, 1]$. 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

$$\nu(\hat\sigma_0, \ldots, \hat\sigma_p) = \lim_{N \to +\infty} N^{-1} \Big(\text{number of values of } h \text{ such that } \sigma_h = \hat\sigma_0, \ldots \sigma_{h+p} = \hat\sigma_p \text{ with } h \le N\Big) \tag{5.6.1}$$

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

$$E\begin{matrix} 0 & 1 & \ldots & p \\ \hat\sigma_0 & \hat\sigma_1 & \ldots & \hat\sigma_p \end{matrix} \equiv \bigcap_{j=0}^{p} S^{-j} P_{\hat\sigma_j} \tag{5.6.2}$$

we see that $E\,{}^{0\;1\;\cdots\;p}_{\hat\sigma_0\,\hat\sigma_1\,\ldots\,\hat\sigma_p}$ is the set of points which visit the sets $P_{\hat\sigma_0},\ldots P_{\hat\sigma_p}$ at times $0,\ldots,p$, so that we should have

$$\nu(\hat\sigma_0,\hat\sigma_1,\ldots,\hat\sigma_k)=\mu\big(E\,{}^{0\;\;1\;\ldots\;p}_{\hat\sigma_0\,\hat\sigma_1\,\ldots\,\hat\sigma_p}\big) \tag{5.6.3}$$

(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

$$\mu_{\underline{\sigma}}\big(C\,{}^{0\;\;1\;\ldots\;p}_{\hat\sigma_0\,\hat\sigma_1\,\ldots\,\hat\sigma_p}\big)=\nu(\hat\sigma_0,\ldots,\hat\sigma_p) \tag{5.6.4}$$

where $C\,{}^{0\;1\;\cdots\;p}_{\hat\sigma_0\,\hat\sigma_1\,\ldots\,\hat\sigma_p}$ is a *cylinder* in $\mathcal{S}$, *i.e.* it is the set of sequences of $\mathcal{S}$ such that $\sigma_0=\hat\sigma_0,\ldots,\sigma_p=\hat\sigma_p$; or, more formally,

$$C\,{}^{0\;\;1\;\ldots\;p}_{\hat\sigma_0\,\hat\sigma_1\,\ldots\,\hat\sigma_p}=\bigcap_{j=0}^{p}\tau^{-j}C\,{}^{0}_{\hat\sigma_j} \tag{5.6.5}$$

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 $C\,{}^{-p\;\cdots\;0\;\;1\;\cdots\;p}_{\hat\sigma_{-p}\,\ldots\,\hat\sigma_0\,\hat\sigma_1\,\ldots\,\hat\sigma_p}$ are needed to determine μ: such measures will be naturally defined by their probabilities $\mu_{\underline{\sigma}}(C\,{}^{0\;\;\cdots\;p+1\;p+2\;\cdots\;2p}_{\hat\sigma_{-p}\,\cdots\,\hat\sigma_0\;\;\hat\sigma_1\;\cdots\;\hat\sigma_p})$ 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 $\mathcal{N}_{\varepsilon,p} \le n^p$). We define (when the limit exists)*

$$s(\underline{\sigma}) = \lim_{\varepsilon \to 0^+} \lim_{p \to \infty} \frac{1}{p} \log \mathcal{N}_{\varepsilon,p} \le \log n \tag{5.6.8}$$

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 $s \ge s(\mathcal{P})$.*

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

$$\mu(C^{\,0\ \ 1\ \ldots\ p}_{\,a_0\, a_1\, \ldots\, a_p}) = \prod_{i=0}^{p} p_{a_i} \qquad (5.6.11)$$

It is an immediate consequence of the theorem that

$$s(\mu) = -\sum_i p_i \log p_i \qquad (5.6.12)$$

(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

$$\mu(C^{\,0\ \ 1\ \ldots\ p}_{\,a_0\, a_1\, \ldots\, a_p}) = \pi_{a_0} \prod_{i=0}^{p-1} p_{a_i a_{i+1}} \qquad (5.6.13)$$

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:

$$s(\mu, S) = -\sum_{i=1}^n \sum_{j=1}^n \pi_i\, p_{ij} \log p_{ij} \qquad (5.6.14)$$

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:

$$\begin{aligned} X &= \sum_{i\in\mathcal{C}_2(k)} \mu(\cap_{j=0}^{k-1} S^{-j} P_{\sigma_j}) \equiv \sum_{i\in\mathcal{C}_2(k)} p_i = \\ &= \sum_{i\in\mathcal{C}_2(k)} \big(\mu(\cap_{j=0}^{k-1} S^{-j} P_{\sigma_j})\big)^{\frac{d-1}{d}+\frac{1}{d}} \le \\ &\le \sum_{i\in\mathcal{C}_2(k)} \Big(\big(\Gamma|\partial \cap_{j=0}^{k-1} S^{-j} P_{\sigma_j}|\big)^{\frac{d}{d-1}}\Big)^{\frac{d-1}{d}} e^{-\eta k/d} \lambda^{-k} \le C' e^{-\eta k/d}). \end{aligned}$$

[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

$$\begin{aligned} &-\sum_i p_i \log p_i = -\sum_{i\in\mathcal{C}_1(k)} p_i \log p_i - \sum_{i\in\mathcal{C}_2(k)} p_i \log p_i \le \\ &\le (d \log \lambda + \eta) k + X\big(\sum_{i\in\mathcal{C}_2(k)} \tfrac{-p_i}{X} \log \tfrac{p_i}{X}\big) + \log X \le \\ &\le (d \log \lambda + \eta) k + X\, k\, \log(n+1) + X\, \log X \end{aligned}$$

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 $P_0 = [0, a_0]$,$P_1 = [a_0, a_1]$,..., $P_{n-1} = [a_{n-2}, 1]$.

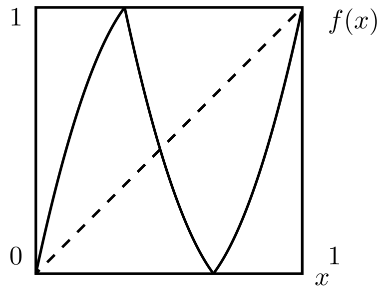


Fig. (5.7.1): *A continuous expansive map of the interval* $[0, 1]$.

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).

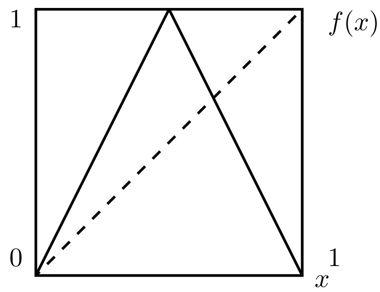


Fig. (5.7.2): *The graph of the "tent map".*

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 $\underline{v}_\pm$).

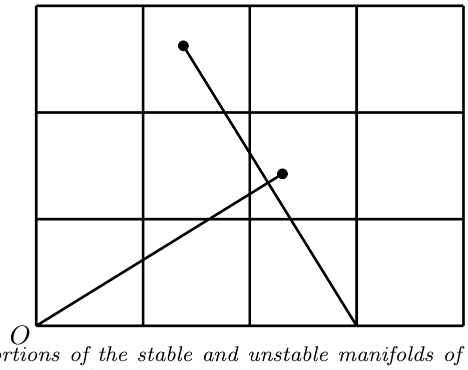


Fig. (5.7.3): *Portions of the stable and unstable manifolds of the Arnold's cat map drawn after representing T^2 as a lattice of copies of the square $[0, 2\pi]^2$ with opposite sides identified.*

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).

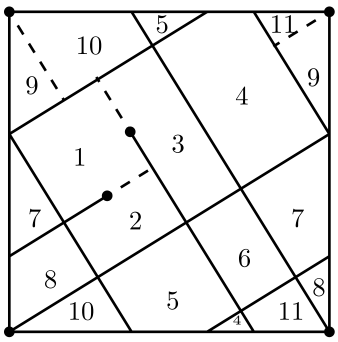


Fig. (5.7.4): *The continuation parts are marked as dashed and come out of the marked points; note that the four vertices of the square are in reality the same point (i.e. the origin) because of the periodic conditions that must be imagined at the boundaries.*

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 $S^2 P_\sigma \cap P_{\sigma''} \neq \emptyset$, *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).

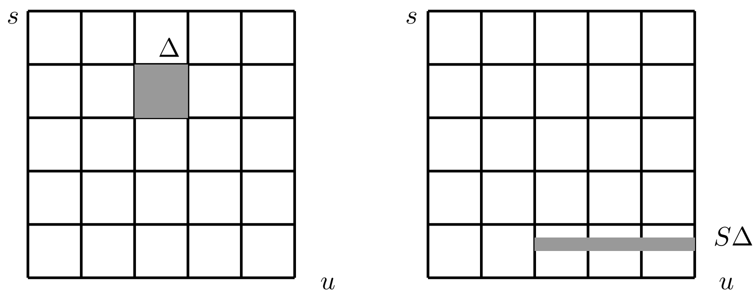


Fig. (5.7.5): *The figures illustrate, symbolically as squares, a few elements of a markovian pavement. An element Δ of it is transformed by S into $S\Delta$ in such a way that the part of the boundary that contracts ends up exactly on a boundary of some elements of $\mathcal{P}$.*

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 $S^k x$ 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 $S^jx\in[a_{\sigma_j},a_{\sigma_j+1}]$. 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 $S^kx=a_i$ 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 $\int_0^1(L^k\,1)(x)\,dx\equiv1$, for $k\ge1$. (*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

$$B^{-1} < \frac{(\mathcal{L}^n 1)(\underline{\sigma})}{(\mathcal{L}^n 1)(\underline{\sigma}')} < B$$

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] $\nu_0(\mathcal{L}^k 1) = 1$ 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

$$|(\mathcal{L}^n 1)(\underline{\sigma}) - (\mathcal{L}^n 1)(\underline{\sigma}')| < D d(\underline{\sigma}, \underline{\sigma}')^{\lambda}$$

(*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 $n \to \mathcal{L}^n 1$ 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 $\mathcal{L}^k f > 0$ and, if $||f||$ denotes the maximum of f and $\nu = h\nu_0$ is the measure defined in problem [5.7.8], $||\mathcal{L}^k f|| \ge B^{-1}\nu(\mathcal{L}^k f) = B^{-1}\nu(f)$. (*Idea:* Proceed as in [5.7.5] and show that $\mathcal{L}^p f(\underline{\sigma}) \ge B^{-1}\mathcal{L}^p(\underline{\sigma}')$ for $p \ge k$ and hence integrate both sides with respect to $\nu(d\underline{\sigma})$ using that $\nu(\mathcal{L}^k 1) = 1$, 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 $k, n \ge 0$. (*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 $n \ge k$, as consequences of the inequality in [5.7.9]:

$$\begin{aligned}
&\nu(|\mathcal{L}^n g_k|) = \nu(|\mathcal{L}^k \mathcal{L}^{n-k} g_k|) = \nu(|\mathcal{L}^k \Big((\mathcal{L}^{n-k} g_k)_+ - (\mathcal{L}^{n-k} g_k)_-\Big)|) = \\
&\nu\Big(\Big|\mathcal{L}^k \Big((\mathcal{L}^{n-k} g_k)_+ - \frac{1}{B}\nu((\mathcal{L}^{n-k} g_k)_+) - (\mathcal{L}^{n-k} g_k)_- \frac{1}{B}\nu((\mathcal{L}^{n-k} g_k)_-)\Big)\Big|\Big) \le \\
&\nu(\mathcal{L}^k \Big((\mathcal{L}^{n-k} g_k)_+ - \frac{1}{B}\nu((\mathcal{L}^{n-k} g_k)_+) + (\mathcal{L}^{n-k} g_k)_- - \frac{1}{B}\nu((\mathcal{L}^{n-k} g_k)_-)\Big)) = \\
&= (1 - B^{-1})\nu(|\mathcal{L}^{n-k} g_k|)
\end{aligned}$$

(*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 $(\mathcal{L}^{n-k} g_k)_\pm - \frac{1}{B}\nu((\mathcal{L}^{n-k} g_k)_\pm) \ge 0$.)

[5.7.12]: Evince from [5.7.11] that $\nu(|\mathcal{L}^n g_k|) \le (1 - B^{-1})^{n/k} \nu(|g_k|)$ 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 $d(S_\varepsilon^n x, S_\varepsilon^n x') \le Ce^{-\lambda' n}$ with $0 < \lambda' < \lambda$ and C suitable. Likewise if x and x' have histories eventually equal in the past it is $d(S_\varepsilon^n x, S_\varepsilon^n x') \le Ce^{-\lambda' n}$. 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):

$$\begin{aligned}
&\lim_{p\to\infty} -p^{-1}\sum_{\hat\sigma_0,\ldots,\hat\sigma_p} \mu(\cap_{i=0}^{\infty}S^{-j}P_{\hat\sigma_j})\log\mu(\cap_{i=0}^{\infty}S^{-j}P_{\hat\sigma_j}) =\\
&=\lim_{p\to\infty}\Big[+p^{-1}\log C - p^{-1}\sum_{\sigma_0,\ldots,\sigma_p} C^{-1}e^{-\sum_{k=0}^p \lambda_e(S^k\underline{\sigma})}\,(-\sum_{k=0}^{p}\lambda_e(S^k\underline{\sigma}))\Big]=\\
&=\int \lambda_e(x)\mu(dx)
\end{aligned}$$

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.

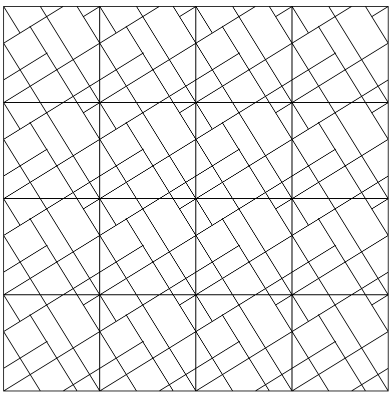


Fig. (5.7.6) *Markovian pavement of* §5.7