Extension of Onsager's reciprocity to large fields and the chaotic hypothesis

Giovanni Gallavotti
Fisica, Universitá di Roma, "La Sapienza", 00185 Roma, Italia

We show that the "fluctuation theorem", a consequence of the *Chaotic Hypothesis* of [1] (formulated again below), can be interpreted as extending to arbitrary forcing fields Green-Kubo's formulae, hence Onsager's reciprocity, in a class of reversible nonequilibrium statistical mechanical systems.

PACS numbers: 47.52.+j, 05.45.+b, 05.70.Ln, 47.70.-n

Interest on nonequilibrium Statistical Mechanics and Fluid Turbulence has recently produced a wealth of results, see [2] to quote only a few. I concentrate on results on thermostatted systems with a *reversible* dissipation mechanism. Physical relevance of such kind of thermostats is questioned (their reversibility sounding a contradictory or, at best, exotic feature): in [2] a connection with old debates and new problems was perceived clearly and to some extent discussed. But it is important that the mentioned works do contain first rate data, relations among them and deep theoretical suggestions that hint at, and call for, a theory. It is also clear that any such theory has to deal with, and say something non trivial about, the *well recognized problem* of "what is the analogue of the Gibbs' ensembles out of equilibrium?" Hence I do not enter here into the debate on the direct physical interest of reversible thermostats, [3].

In the early 70's Ruelle, [4], proposed an answer ("ensemble" = "SRB distribution", see below) and, since, many have been looking for consequences of that idea, a bit too advanced at the time to be widely appreciated. The key work [5] provided a puzzling experimental result and an indication for its theory based on Ruelle's principle. This led [1] to "predict", as a consequence of a form of Ruelle's principle ("chaotic hypothesis"), the result of [5], with *no free parameters* available.

Although the results required a *reversible* dissipation mechanism their generality seemed promising. One should keep in mind that we look for a non equilibrium analogue of Boltzmann's heat theorem; he derived out of clear general principles, but not necessarily very clearly related to Physics at the time, that $(dU + pdV)/T$ is exact: a property without free parameters "always valid". The fluctuation theorem seemed a step in this direction.

As a consequence I wanted to test whether the chaotic hypothesis is consistent (at least for the considered wide class of systems) with other properties that are known or universally believed to be correct in generic non equilibrium systems. Not many of them exist: perhaps only the fluctuation dissipations theorems (*i.e.* Onsager's reciprocity and Green–Kubo's formulae) which concern derivatives of thermodynamic quantities in terms of forcing fields *evaluated at zero forcing*. The purpose of this paper is to show that the fluctuation theorem of [1], valid with small or large forcing reduces to Onsager's relations and Green–Kubo's formulae at zero forcing: i.e. it is a general law with no free parameters, valid out of equilibrium with no conditions and reducing to known results near equilibrium. Hence it is a result of the type that one looks for in the attempt to establish a theory of nonequilibrium. That the chaotic hypothesis implied Onsager's reciprocity was already noted in [6] with no reference to the fluctuation theorem; but the experimental results of [7] suggested that there could have been a much less technical derivation directly *from the fluctuation theorem.* This is done in the present paper by showing that the latter general theorem, a property of *large* fluctuations, becomes a trivial identity at zero fields but, if before letting the fields to 0 one divides by appropriate powers of the forcings and at the same time one uses the central limit theorem for *small* fluctuations, one gets Green–Kubo's relations (hence reciprocity). We, therefore, conclude that the fluctuation theorem can be regarded as an extension to non zero fields of Onsager's reciprocity theorem valid only at 0 forcing.

A typical system studied here will be N point particles subject to *(a)* mutual and external conservative forces with potential $V(\vec{q}_1,\ldots,\vec{q}_N)$, *(b)* external (non conservative) forces, *forcing agents*, $\{\vec{F}_j\}$, $j=1,\ldots,N$, whose strength is measured by parameters $\{G_i\}$, $i=1,\ldots,s$, and *(c)* also to forces $\{\vec{\varphi}_j\}$, $j=1,\ldots N$, generating constraints that provides a model for the thermostatting mechanism that keeps the energy of the system from growing indefinitely (because of the continuing action of the forcing agents). An observable $O(\{\vec{q},\dot{\vec{q}}\})$ evolves under the time evolution S_t solving the equations of motion:

$$m\ddot{\vec{q}}_j = -\partial_{\vec{q}_j}V(\vec{q}_j) + \vec{F}_j(\{G\}) + \vec{\varphi}_j \qquad (1)$$

($m =$ particles mass; $\vec{\varphi}_j$ "thermostatting" forces assuring approach to a (non equilibrium) stationary state). Time evolution of O on the motion starting at $x = (\vec{q},\dot{\vec{q}})$ is a function $t \to O(S_t x)$; *motion statistics* is the stationary probability distribution μ_+ on *phase space* $\mathcal{C}$ such that:

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T O(S_t x)\,dt = \int_{\mathcal{C}} O(y)\mu_+(dy) \stackrel{def}{=} \langle O\rangle_+ \qquad (2)$$

for all data $x \in \mathcal{C}$ except a set of zero measure with respect to the volume $\overline{\mu}_0$ on $\mathcal{C}$. The distribution μ_+ is assumed to exist: a property called *zero-th law*, [1, 6].

The thermostatting mechanism will be described by force laws $\vec{\varphi}_j$ enforcing the constraint that kinetic energy (or total energy) of the particles, or of subgroups of the particles, remains constant, [8]. It is convenient also to imagine that the constraints keep the total kinetic energy bounded (hence phase space is bounded). Constraint forces $\{\vec{\varphi}_j\}$ will be supposed such that the system is *reversible*: this means that there will be a map i, defined on phase space, anticommuting with time evolution: *i.e.* $S_t\, i \equiv i\, S_{-t}$. Examples of such thermostats are in *Nosé–Hoover*'s class, see for instance [1, 2].

Reversibility is a key assumption, [1, 9].

In [1, 7] the above systems, at least when *chaotic i.e.* when showing at least one positive Lyapunov exponent, are supposed to verify following hypothesis:

Chaotic hypothesis: A reversible many particle system in a stationary state can be regarded as a transitive Anosov system for the purpose of computing the macroscopic properties.

This means that the attractor is assumed hyperbolic in a strict mathematical sense ("Anosov" is a technical statements about existence and mild regularity of stable and unstable manifolds at each attractor point); *transitive* means that the stable and unstable manifolds of each attractor point are dense on the attractor, see [4].

In [7] a broader formulation of the hypothesis is given that applies also to cases in which the attractor is smaller than the whole phase space (which may happen at really large forcing and fixed number of particles): but in this paper we adhere to the original formulation, [1], to simplify the analysis.

In the quoted references it is argued that the chaotic hypothesis should be considered in the same way as the *ergodic hypothesis* in equilibrium statistical mechanics. It is assumed as correct even in cases in which it cannot be mathematically strictly valid: but this can be done only for the purpose of deriving statistical properties of a few relevant observables. An analogue of this procedure is the derivation of the second law from ergodicity (*i.e.* from the microcanonical ensemble): the law ("Boltzmann's heat theorem") is derived supposing ergodicity and it is assumed valid even when the ergodic hypothesis is obviously false (*e.g.* for the free gas in a box).

It might be surprising that *there are non trivial consequences of the hypothesis*, besides the existence of the distribution μ_+ in (2) which in this context is called the *SRB distribution*, and that they *can be tested*, [5, 7]. The main one is the *fluctuation theorem*, see below.

In [1] the divergence of the r.h.s. of Eq. (1) is a quantity $-\sigma(x)$ defined on phase space that has been identified with the *entropy production rate*.

The *chaotic hypothesis*, (CH), implies a *fluctuation theorem* or (FT) which, [1], is a property of the fluctuations of the entropy production rate. Namely if we denote $\langle\sigma\rangle_+ = \int_{\mathcal{C}} \sigma(y)\mu_+(dy)$ the time average, over an infinite time interval by Eq. (2), then the *dimensionless* finite time average $p = p(x)$:

$$\frac{1}{\tau}\int_{-\tau/2}^{\tau/2} \sigma(S_t x)\, dt \overset{def}{=} \langle\sigma\rangle_+\, p \tag{3}$$

has a statistical distribution $\pi_\tau(p)$ with respect to the stationary state distribution μ_+ such that:

$$\frac{1}{\tau\langle\sigma\rangle_+\, p}\log\frac{\pi_\tau(p)}{\pi_\tau(-p)} \to_{\tau\to\infty} 1 \tag{4}$$

provided (of course) $\langle\sigma\rangle_+ > 0$. Following [1] a reversible system for which $\langle\sigma\rangle_+ > 0$ will be called *dissipative*. Ruelle's H-theorem states that $\langle\sigma\rangle_+ \geq 0$ and, at least if there are no nontrivially 0 Lyapunov exponents (*e.g.* if (CH) is assumed), $\langle\sigma\rangle_+ > 0$ unless the stationary distribution μ_+ has the form $\rho(x)\overline{\mu}_0(dx)$, [9].

Hence we shall suppose that the system is dissipative when the forcing $\vec{G}$ does not vanish and, without real loss of generality, that $\sigma(ix) = -\sigma(x)$ and that $\sigma(x) \equiv 0$ when the external forcing vanishes, writing:

$$\sigma(x) = \sum_{i=1}^{s} G_i J_i^0(x) + O(G^2) \tag{5}$$

Assuming fast decay of σ–σ correlations (to be expected if the (CH) is accepted, [13]) then by a result of Sinai [13] the entropy verifies a *limit theorem*; *i.e.*:

$$\lim_{\tau\to\infty}\frac{1}{\tau}\log\pi_\tau(p) = -\zeta(p) \tag{6}$$

where $\zeta(p)$ is analytic for p in the interval $[-p^*, p^*]$ within which it can vary (model dependent)[10, 11]. The function $\zeta(p)$ can be conveniently computed because its transform $\lambda(\beta) = \lim_{\tau\to\infty}\frac{1}{\tau}\log\int e^{\beta\tau\,(p-1)\langle\sigma\rangle_+}\pi_\tau(p)dp$ can be expressed by a *cumulant expansion.* Once $\lambda(\beta)$ is "known" then $\zeta(p)$ is recovered via a Legendre transform; $\zeta(p) = \max_\beta\big(\beta\langle\sigma\rangle_+(p-1) - \lambda(\beta)\big)$, [10, 11].

By using the cumulant expansion for $\lambda(\beta)$ we find that $\lambda(\beta) = \frac{1}{2!}\beta^2 C_2 + \frac{1}{3!}\beta^3 C_3 + \ldots$ where the coefficients C_j are $\int_{-\infty}^{\infty}\langle\sigma(S_{t_1}\cdot)\sigma(S_{t_2}\cdot)\ldots\sigma(S_{t_{j-1}}\cdot)\sigma(\cdot)\rangle_+^T\, dt_1\ldots$ if $\langle\ldots\rangle_+^T$ denote the cumulants of the variables $\sigma(x)$.

In our case the cumulants of order j have size $O(G^j)$, by Eq. (5), so that:

$$\zeta(p) = \frac{\langle\sigma\rangle_+^2}{2C_2}(p-1)^2 + O((p-1)^3G^3) \tag{7}$$

(note the first term in r.h.s. giving the central limit theorem). Eq. (7), together with the (FT) (4), yields at fixed p the key relations:

$$\langle\sigma\rangle_+ = \frac{1}{2}C_2 + O(G^3) \tag{8}$$

We define, [10]: *current* $J_i(x) = \partial_{G_i}\sigma(x)$, *transport coefficients* $L_{ij} = \partial_{G_j}\langle J_i(x)\rangle_+|_{G=0}$ and we study L_{ij}.

To derive Green–Kubo formulae, (GK), we first look at the r.h.s. of the first of Eq. (8) discarding $O(G^3)$: the r.h.s becomes quadratic in G with coefficients:

$$\frac{1}{2}\int_{-\infty}^{\infty} dt\,\big(\langle J_i^0(S_t\cdot)J_j^0(\cdot)\rangle_+ - \langle J_i^0\rangle_+\langle J_j^0\rangle_+\big)\big|_{G=0} \qquad (9)$$

On the other hand the expansion of $\langle\sigma\rangle_+$ in the l.h.s. of Eq. (8) to second order in G gives:

$$\langle\sigma\rangle_+ = \frac{1}{2}\sum_{ij}\big(\partial_{G_i}\partial_{G_j}\langle\sigma\rangle_+\big)\big|_{G=0}G_iG_j \qquad (10)$$

because the first order term vanishes (by Eq. (5), or (8)). The r.h.s. of (10) is the sum of $\frac{1}{2}G_iG_j$ times $\partial_{G_i}\partial_{G_j}\int\sigma(x)\mu_+(dx)$ which equals the sum of the following three terms: the first is $\int\partial_{G_i}\partial_{G_j}\sigma(x)\mu_+(dx)$, the second is $\int\partial_{G_i}\sigma(x)\partial_{G_j}\mu_+(dx) + (i\leftrightarrow j)$ and the third is $\int\sigma(x)\partial_{G_i}\partial_{G_j}\mu_+(dx))$, all evaluated at $G=0$. The first addend is 0 (by time reversal), the third addend is also 0 (as $\sigma=0$ at $G=0$). Hence:

$$\partial_{G_i}\partial_{G_j}\langle\sigma\rangle_+|_{G=0} = \Big(\partial_{G_j}\langle J_i^0\rangle_+ + \partial_{G_i}\langle J_j^0\rangle_+\Big)|_{G=0} \quad (11)$$

and it is easy to check, again by using time reversal, that:

$$\partial_{G_j}\langle J_i^0\rangle_+|_{G=0} \,=\, \partial_{G_j}\langle J_i\rangle_+|_{G=0} \,=\, L_{ij} \qquad (12)$$

Thus equating r.h.s and l.h.s. of Eq. (8), as expressed respectively by Eq. (9) and (11) we express the matrix $\frac{L_{ij}+L_{ji}}{2}$ getting (GK) at least if $i=j$: a relation sometimes called a "fluctuation dissipation theorem", [12].

We want to show that the above ideas *also* suffice to prove Onsager reciprocity, (OR), *i.e.* $L_{ij} = L_{ji}$. The main remark is that we can *extend* (FT) theorem to give properties of *joint* distribution of the average of σ, (3), and of the corresponding μ_+–average of $G_j\partial_{G_j}\sigma$. In fact defining *dimensionless* j-current $q = q(x)$ as:

$$\frac{1}{\tau}\int_{-\tau/2}^{\tau/2} G_j\partial_{G_j}\sigma(S_tx)dt \stackrel{def}{=} G_j\langle\partial_{G_j}\sigma\rangle_+\, q \qquad (13)$$

where the factor G_j is there only to keep σ and $G_j\partial_{G_j}\sigma$ with the same dimensions.

Then if $\pi_\tau(p,q)$ is the joint probability of p,q the *same* proof of the (FT) in [1] yields also:

$$\lim_{\tau\to\infty}\frac{1}{\tau\langle\sigma\rangle_+ p}\log\frac{\pi_\tau(p,q)}{\pi_\tau(-p,-q)} = 1 \qquad (14)$$

and the limit theorem in (6) is extended, [13], to:

$$\lim_{\tau\to\infty}\frac{1}{\tau}\log\pi_\tau(p,q) = -\zeta(p,q) \qquad (15)$$

We can compute $\zeta(p,q)$ in the same way as $\zeta(p)$ by considering first the transform $\lambda(\beta_1,\beta_2)$:

$$\lim_{\tau\to\infty}\tfrac{1}{\tau}\log\int e^{\tau(\beta_1\,(p-1)\langle\sigma\rangle_+ + \beta_2\,(q-1)\langle G_j\partial_{G_j}\sigma\rangle_+)}\pi_\tau(p,q)dpdq \qquad (16)$$

and then the Legendre trasform:

$$\zeta(p,q) = \max_{\beta_1,\beta_2}\big(\beta_1\,(p-1)\langle\sigma\rangle_+ + \beta_2\,(q-1)\langle G_j\partial_{G_j}\sigma\rangle_+ - \\ -\lambda(\beta_1,\beta_2)\big) \qquad (17)$$

The function $\lambda(\vec\beta)$, $\vec\beta = (\beta_1,\beta_2)$, is evaluated by the cumulant expansion, as above, and one finds:

$$\lambda(\vec\beta) = \frac{1}{2}\,(\vec\beta, C\,\vec\beta) + O(G^3) \qquad (18)$$

where C is the 2×2 matrix of the second order cumulants. The coefficient C_{11} is given by C_2 appearing in (7); C_{22} is given by the same expression with σ replaced by $G_j\partial_{G_j}\sigma$ while C_{12} is the mixed cumulant:

$$\int_{-\infty}^{\infty}\big(\langle\sigma(S_t\cdot)\,G_j\partial_{G_j}\sigma(\cdot)\rangle_+ - \langle\sigma(S_t\cdot)\rangle_+\,\langle G_j\partial_{G_j}\sigma(\cdot)\rangle_+\big)dt \qquad (19)$$

Hence if $\vec w = \begin{pmatrix}(p-1)\langle\sigma\rangle_+\\ (q-1)\langle G_j\partial_{G_j}\sigma\rangle_+\end{pmatrix}$ we get:

$$\zeta(p,q) = \frac{1}{2}\,\big(C^{-1}\vec w,\vec w\big) + O(G^3) \qquad (20)$$

completely analogous to (7). But the (FT) in (14), implies that $\zeta(p,q)-\zeta(-p,-q)$ is q *independent*: this means, as it is immediate to check:

$$-(C^{-1})_{22}\langle G_j\partial_{G_j}\sigma\rangle_+ - (C^{-1})_{21}\langle\sigma\rangle_+ = 0 + O(G^3) \quad (21)$$

which because of (8), and of $(C^{-1})_{22} = C_{11}/\det C$, becomes the analogue of (8):

$$\langle G_j\partial_{G_j}\sigma\rangle_+ = \frac{1}{2}C_{12} + O(G^3) \qquad (22)$$

Then, proceeding as in the derivation of (9) through (12) (*i.e.* expanding both sides of (22) *to first order* in the G_i's and using (19)) we get that $\partial_{G_i}\langle\partial_{G_j}\sigma\rangle_+$ is given by the integral in (9). This means that $L_{ij} = L_{ji}$ and the (GK) follow together with the (OR).

Thus Eq. (8),(22) and the ensuing (GK), and (OR), are a consequence of (FT), (4), and of its (obvious) extension, (14), in the limit $G\to 0$, when combined with the expansion (7) for entropy fluctuations. Those theorems and the fast decay of the $\sigma\sigma$ correlations, [13], are all natural consequences of (CH) for reversible statistical systems, which is the starting point of our considerations. Reversibility is here assumed *both in equilibrium and in non equilibrium*: this is a feature of gaussian thermostat models, [1], but by no means of all models, [5, 8].

Of course while the (OR) and (GK) only hold around equilibrium, *i.e.* they are properties of G–derivatives evaluated at $G = 0$, and the expansion for $\lambda(\beta)$ is a general consequence of the correlation decay, the (FT) also holds far from equilibrium, *i.e.* for large G and can be considered a generalization of the (OR) and (GK).

Evidence for (8),(22) arose in [7] in an effort to interpret results of various numerical experiments and an apparent incompatibility of the *a priori* known non gaussian nature of the distribution $\pi_\tau(p)$ and "gaussian looking" empirical distributions. In [7] the situation arising at really large fields, when the attractor is strictly smaller than the whole phase space, is also discussed.

The above ideas as well as attempts to give a more fundamental role to gaussian thermostats gave rise, recently, to several papers [3].

Acknowledgements: I profited from many discussions and hints from F. Bonetto and P. Garrido. I am particularly indebted and grateful to E.G.D. Cohen for important comments, criticism, suggestions and much needed encouragement, and to G. Gentile for pointing out an error in the first version and for many comments. Partial support from Rockefeller U., CNR-GNFM, ESI and the EU program: "Stability and Universality in Classical Mechanics", # ERBCHRXCT940460.

[1] Gallavotti, G., Cohen, E.G.D.: Phys. Rev. Lett., **74**, 2694–2697, 1995 and J. Stat. Phys., **80**, 931–970, 1995.

[2] Hoolian, B.D., Hoover, W.G., Posch, H.A.: Phys.Rev. Lett. **59**, 10, 1987; Hoover, W.G., Posch, H.A.: Phys. Rev **E49**, 1913, 1994; Dellago, Ch., Posch., H.A., Hoover, W.G.: Phys. Rev. **E53**, 1485, 1995; Evans, D.J., Cohen E.G.D, Morriss, G.P.: Phys. Rev.: **142A**, 5990, 1990.

[3] Gallavotti, G.: Phys. Lett. A, **223**, 91–95, 1996 and Physica D, **105**, 163–184, 1997.

[4] Ruelle, D.: *Chaotic motions and strange attractors*, Lezioni Lincee, notes by S. Isola, Cambridge U. Press, 1989; Ruelle, D.: Ann. of N. Y. Acad. of Sci., **357**, 1–9, 1980; Eckmann, J.P., Ruelle, D.: Rev. Mod. Phys. **57**, 617–656, 1993.

[5] Evans, D.J.,Cohen, E.G.D., Morriss, G.P.: Phys. Rev. Lett., **71**, 2401–2404, 1993.

[6] Gallavotti, G.: J. Stat. Phys., **84**, 899–926, 1996.

[7] Bonetto, F., Gallavotti, G., Garrido, P.: Physica D, **105**, 226–252, 1997.

[8] Examples of such thermostats are the gaussian thermostats and a few concrete cases can be found in [1]. Gaussian thermostats and reversibility importance and possibility of linking them with the dynamical systems theory of SRB distributions, [13], has been stressed in the fundamental paper [5] whose experimental results and theoretical hints led to formulating of the (CH) as a reinterpretation of Ruelle's principle, [9].

[9] Ruelle, D.: J. Stat. Phys., **86**, 1-25, 1996. This theorem holds under much more general assumptions: for any dynamical system which is mildly smooth and with bounded phase space. Intuitively it says that, if phase space is bounded, the volume cannot *expand on the average* (*i.e.* "forever"); and if there is an invariant distribution that is absolutely continuous (*i.e.* it is given by a suitable density on phase space) then the volume cannot *contract on the average* (*i.e.* "forever"). Finally if the volume does not contract in the average then there is an absolutely continuous distribution. Even the first statement requires a detailed discussion. One should not forget that in dissipative systems stationary states are concentrated on zero volume regions so that volume there could be well expanding in the average with no contradiction with the intuition suggested above. The point is that if evolution is volume expanding on the attractor then "provided the dynamics is mildly regular" (e.g. if the divergence of the equations of motion is differentiable) the expansivity on the attractor would extend to expansivity in its vicinity. This would be a contradiction at least if the dynamics near the attractor is "very close" to that on the attractor (and a sufficient condition, proposed in the referred paper, is that there should be no 0 Lyapunov exponent, beyond the trivial one). Strictly speaking the theorem has been proved in the above reference only for maps (with the divergence replaced by the logarithm of the jacobian), but I think that its proof extends to flows, see also comment [13] below.

[10] Gallavotti, G.: Math. Phys. Electronic J., MPEJ, **1**, 1–12, 1995, (http:// mpej. unige. ch). Strictly speaking this paper deals with discrete time systems; the corresponding continuous time case is discussed in [11].

[11] Gentile, G.: Forum Mathematicum, **10**, 89–118, 1998.

[12] By fluctuation theorem we mean expressing transport coefficients L_{ij} as in (9): this is more precisely the (FD) of the first kind at 0 frequency, following the terminology of R. Dorfman, *Transport coefficients*, (entry for the Italian Encyclopedia of Physical Sciences, vol. VI), Roma 1996, eq. (3.7). It coincides with Green–Kubo's formula.

[13] Sinai, Y.G.: Russian Mathematical Surveys, **27**, 21–69, 1972. Also: *Lectures in ergodic theory*, Lecture notes in Mathematics, Princeton, 1977. For discrete time Anosov systems correlation decay of any smooth observable is exponential. Here we consider continuous time hence, strictly speaking, fast decay is an assumption: note however that in [1, 7] the hypothesis is made for discrete dynamical systems obtained from continuous time systems of (1) by timing observations with characteristic events (*e.g.* collisions): a procedure also called *Poincaré section.* We use continuous time because it is simpler conceptually, apart from the latter problem (discussed and essentially solved in [11]).

Published on *Physical Review Letters*, **77**, 4334-4337, 1996.