Remark on the (non)convergence of ensemble densities in dynamical systems

S. Goldstein\textsuperscript{a)} and J. L. Lebowitz\textsuperscript{b)}

Departments of Mathematics and Physics, Rutgers University, Piscataway, New Jersey 08854-8019

Y. Sina\textsuperscript{c)}

Department of Mathematics, Princeton University, Princeton, New Jersey 08544

(Received 11 March 1998; accepted for publication 6 April 1998)

We consider a dynamical system with state space \( M \), a smooth, compact subset of some \( \mathbb{R}^n \), and evolution given by \( T_t \), \( x_t = T_t x \), \( x \in M \); \( T_t \) is invertible and the time \( t \) may be discrete, \( t \in \mathbb{Z} \), \( T_t = T^t \), or continuous, \( t \in \mathbb{R} \). Here we show that starting with a continuous positive initial probability density \( \rho(x,0) > 0 \), with respect to \( dx \), the smooth volume measure induced on \( M \) by Lebesgue measure on \( \mathbb{R}^n \), the expectation value of \( \log \rho(x,t) \), with respect to any stationary (i.e., time invariant) measure \( \nu(dx) \), is linear in \( t \), \( \nu(\log \rho(x,t)) = \nu(\log \rho(x,0)) + Kt \). \( K \) depends only on \( \nu \) and vanishes when \( \nu \) is absolutely continuous with respect to \( dx \).

\[ \mu_\nu(f) = \int f(T_t x) \rho(x,0) dx = \int f(x) \rho(x,0) dx \]  

for any bounded measurable \( f(x) \). The weak convergence of \( \rho(x,t) \) to \( \bar{\rho} \), expressed by Eq. (3), is clearly compatible with the fact that when \( T_t \) preserves \( dx \), the Gibbs entropy \( S_\mu = -\int \rho(x,t) \log \rho(x,t) dx \), and indeed any \( \tilde{F} = \int F(\rho(x,t)) dx \), is constant in time.

Unfortunately, it is sometimes thought that this constancy of \( S_\mu \) for Hamiltonian evolutions is a manifestation of the conflict between microscopic reversibility and the second law of thermodynamics, and that the resolution of this conflict requires at least an acceptance of weak convergence as the mathematical expression of the approach to equilibrium characteristic of macroscopic irreversibility, and perhaps even necessitates changes in the microscopic physical laws, cf. Ref. 3(b). This concern and its proposed resolution are based on a misunderstanding of the origin of the observed time asymmetry of macroscopic physical systems, which really concerns not probability densities but the behavior of individual systems whose microstates \( x_t = T_t x \) are points in a very high-dimensional phase space \( M \). In fact, the second law refers not to \( S_\mu \) but to an entropy defined for individual macroscopic systems, whose observed irreversible behavior is due first and foremost to the large discrepancy between the

\textsuperscript{a)}Electronic mail: oldstein@math.rutgers.edu
\textsuperscript{b)}Electronic mail: lebowitz@math.rutgers.edu
\textsuperscript{c)}Electronic mail: sinai@math.princeton.edu

The long-time behavior of ensemble densities for dynamical systems, the analysis of which was initiated by Gibbs, is widely linked with the origins of thermodynamic irreversibility. While regarding the linkage as often misguided, we note here some simple, and perhaps surprising, features of this behavior. In particular we find, with great generality, an exactly linear time dependence for a natural modified entropy functional.
scale of macroscopic observables (which behave irreversibly) and microscopic scales and to the nature of “typical” initial conditions for the microstate $x$ of the system, cf. Ref. 3.

These conceptual issues are, however, not the main concern of this brief note, even though it was motivated by the paper of R. Fox in this issue$^4$ in which such problems are discussed for the baker’s transformation. In that paper Fox notes the constancy of $F$ when $F = \log \rho$ for this transformation. Here we are concerned with what happens to functions of $\rho(x,t)$ when $T_t$ does not preserve $dx$ and $\nu$ may not be absolutely continuous.

Let $\nu$ be a stationary probability measure, and let $\mu_t^{(1)}$ and $\mu_t^{(2)}$ be two measures on $M$, evolving according to the dynamics. If $\mu_t^{(1)}$ is absolutely continuous with respect to $\mu_t^{(2)}$, i.e., $\mu_t^{(1)}(dx) = g(x,t)\mu_t^{(2)}(dx)$, then it follows directly from Eq. (2) that

$$g(x,t) = g(T_{-t}x,0).$$

(4)

Suppose that $g(x,0)$ is continuous in $x$. Then, given a function of $g$, $f(g)$, integrable with respect to $\nu$, we have $\nu(f(g(x,t))) = \nu(f(g(x,0)))$ for all $t$. Assume now that $\mu_t^{(1)}$ and $\mu_t^{(2)}$ are absolutely continuous with respect to $dx$, with continuous positive densities $\rho_1(x,t)$ and $\rho_2(x,t)$. Then $g(x,t) = \rho_1(x,t)/\rho_2(x,t)$ and

$$\nu(f(g)) = \int_M f(\rho_1(x,t)/\rho_2(x,t)) \nu(dx) = \text{const}. \tag{5}$$

Setting $f(g) = \log g$ yields

$$\nu(\log \rho_1(x,t)) - \nu(\log \rho_2(x,t)) = C$$

(6)

independent of $t$. Put now $\rho_2(x,t) = \rho(x,t)$ and $\rho_1(x,t) = \rho(x,t+\tau)$. Equation (6) then becomes for all $\tau$

$$\nu(\log \rho(x,t+\tau)) - \nu(\log \rho(x,t)) = K(\tau). \tag{7}$$

Noting that $K(\tau_1 + \tau_2) = K(\tau_1) + K(\tau_2)$ we obtain a rather surprising result,

$$\nu(\log \rho(x,\tau)) = \nu(\log \rho(x,0)) + K \tau, \tag{8}$$

with $K$ independent of $\tau$. In other words, the average of the log of the density with respect to the stationary measure $\nu$ is linear in the time. On the other hand it follows from Eq. (6) that the growth rate of $\nu(\log \rho(x,t))$ does not depend on $\rho$. Hence $K$ depends only on the dynamics $T_t$ and the stationary probability measure $\nu$. Consequently, we can compute $K$ by taking for our initial (unnormalized) density $\rho(x,0) = 1$. We then get

$$K = \nu \left. \frac{dJ}{dt} \right|_{t=0}, \tag{9}$$

where $J(x,t)$ is the Jacobian of the transformation $T_{-t}$, for continuous time and

$$K = \nu(\log J(x)), \tag{10}$$

where $J(x) = J(x,1)$, for discrete time. If $\nu$ is absolutely continuous with respect to $dx$, i.e., $\nu(dx) = \tilde{\rho}(x)dx$, then putting $\rho_2(x,t) = \tilde{\rho}(x)$ and $\rho_1 = \rho$ in Eq. (6) we see that

$$\int_M [\log \rho(x,t)] \tilde{\rho}(x) dx \text{ is independent of } t, \text{ i.e., } K \text{ vanishes for such a } \nu.$$ 

In the case of a continuous time evolution given by a (smooth) vector field, $\dot{x} = \mathbf{v}(x)$, the right-hand side of Eq. (9) is just $\nu(-\nabla \cdot \mathbf{v})$. Equations (8) and (9) can then also be obtained directly for a smooth, positive $\rho(x,0)$ by starting with the continuity equation

$$\frac{\partial \rho(x,t)}{\partial t} = -\nabla \cdot (\rho \mathbf{v}(x)). \tag{11}$$

We then find

$$K = \frac{d}{dt} \int_M \log \rho(x,t) \nu(dx)$$

$$= - \int_M \rho^{-1} \nabla \cdot (\rho \mathbf{v}) \nu(dx)$$

$$= - \int_M (\nabla \cdot \mathbf{v} + (\nabla \log \rho) \cdot \mathbf{v}) \nu(dx). \tag{12}$$

On the other hand, the time derivative of $\mu_t(\phi)$ is, for any smooth $\phi(x)$, given by

$$\frac{d}{dt} \mu_t(\phi) = - \mu_t(\mathbf{v} \cdot \nabla \phi). \tag{13}$$

Hence, by the stationarity of $\nu$, $\nu(\mathbf{v} \cdot \nabla \phi) = 0$ and so the second term in the square brackets in Eq. (12) vanishes, yielding explicitly

$$K = -\nu(\nabla \cdot \mathbf{v}). \tag{14}$$

Equations (7) and (13) are to be compared with what happens to the rate of change of the Gibbs entropy $S_\mu$, for $\mu_t(dx) = \rho(x,t)dx$. A straightforward computation gives

$$\frac{d}{dt} S_\mu = - \frac{d}{dt} \int_M \rho(x,t) \log \rho(x,t) dx$$

$$= \int_M (\nabla \cdot \mathbf{v}) \rho(x,t) dx = \mu_t(\nabla \cdot \mathbf{v}). \tag{15}$$

$\dot{S}_\mu$ has been of much interest recently in connection with “thermostated” nonequilibrium systems. Under suitable conditions on $T_t$, it can be shown that $\mu_t(dx) = \rho(x,t)dx$ for $\mathbf{mu}(t)$ is a Sinai, Ruelle, Bowen (SRB) measure. In such cases

$$- \frac{d}{dt} S_\mu \to - \nu(\nabla \cdot \mathbf{v}) \tag{16}$$

with $\nu(\nabla \cdot \mathbf{v}) \leq 0$. The equality holds if and only if $\nu$ is absolutely continuous with respect to $dx$, i.e., $\nu(\cdot) = \tilde{\rho}(x)dx$. On the other hand when $T_t$ is “time reversible” in the sense that there exists a transformation $R$ on $M$, preserving $dx$, such that $R^2 = I$ and $RT_tx = T_{-t}Rx$, then

$$K = - \nu(\nabla \cdot \mathbf{v}) = \nu(\mathbf{v} \cdot \nabla) = - K_\perp. \tag{17}$$

Thus, writing $S_\perp(t) = - \nu(\log \rho(x,t))$ we have in this case that
\[ S_{\mu}(t) \sim S_{\sigma}(t) \quad \text{for} \quad t \to \pm \infty \]

with

\[ S_{\sigma}(t) = S_{\sigma}(0) = K_{\sigma}t. \]

As an illustrative example consider a flow on a circle, with \( v(x) = -\sin x + \omega, \) where \( x \in [-\pi, \pi] \) with periodic boundary conditions and \( \omega \) is a constant. This example corresponds to a particle moving in the plane with velocity \( \mathbf{u} \) under the action of an electric field \( \mathbf{E} \) and a magnetic field \( \mathbf{h} \) perpendicular to the plane. The speed \(|\mathbf{u}|\) is kept equal to one by a Gaussian thermostat; 2, 5 \( \chi \) is the angle between the velocity \( \mathbf{u} \) and \( \mathbf{E} \) and \( \omega \sim h/|E| \). (This flow is time reversible, with \( R \) given by reflection through \( \pi/2 \), the minimum of \( v \).) For \(|\omega| < 1 \), \( \nu(x) \) are delta functions at \( x_{\pm} = \arcsin \omega \) with \( |x_{+}| < |x_{-}| \). We clearly have \( K_{\pm} = \cos x_{\pm} = \sqrt{1 - \omega^2} = -K_{-} > 0 \). For \(|\omega| > 1 \) there is a unique stationary state, \( \nu(dx) = \bar{\rho}(x)dx \), with \( \bar{\rho}(x) \) proportional to \( 1/|v(x)| \) and \( K = 0 \) on general grounds as well as by explicit computation. At \(|\omega| = 1 \), \( x_{+} = x_{-} \), \( \nu_{+} = \nu_{-} = \nu \) with \( K = 0 \) so \( K \) is continuous in \( \omega \).

Another observation which follows from Eq. (5) is that for an absolutely continuous \( \nu \), with density \( \bar{\rho}(x) \),

\[ B_{\rho} = \int_{M} \left| \frac{\rho(x,t)}{\bar{\rho}(x)} - 1 \right| \bar{\rho}(x)dx \tag{18} \]

is independent of \( t \). For \( \rho = 1 \) Eq. (18) is just the \( L_{1} \) distance between \( \mu \) and \( \nu \), since \( M \) is compact, \( \int_{M}dx = |M| < \infty \), Eq. (17) also implies, by the Schwartz inequality, that \( \int_{M} |\rho(x,t) - \bar{\rho}(x)|^2 dx \geq |M|^{-1} B_{\bar{\rho}}^2 > 0 \) unless \( \rho = \bar{\rho} \), and a similar statement is true of the higher norms. Thus there can be no convergence to zero of the \( L_{2} \) and higher norms of \( \rho(x,t) - \bar{\rho}(x) \).

We conclude by noting that the long time behavior of \( \rho(x,t) \) was discussed in Ref. 6 for hyperbolic maps. It was explained there that conditional probability densities induced by \( \mu_{t} \) on the unstable manifolds converge, as \( t \to \infty \), pointwise with their derivatives to the corresponding densities given by \( \nu \). Along stable directions, however, the densities \( \rho(x,t) \) are extremely irregular, as might be suggested by the preservation of the integrals discussed above.

ACKNOWLEDGMENTS

S.G. was supported in part by NSF Grant No. DMS 95-04556. J.L.L. was supported in part by NSF Grant No. DMR 95-23266.