Variational formulation and ergodic invariants

A. Carreira, M.O. Hongler and R. Vilela Mendes
Centro de Física da Matéria Condensada, Av. Gama Pinto 2, P-1699 Lisbon Codex, Portugal

Received 26 September 1990; revised manuscript received 14 March 1991; accepted for publication 15 March 1991
Communicated by A.R. Bishop

A family of ergodic invariants for discrete time dynamical systems is constructed from a variational formulation of the dynamics. Some examples are presented for the case of one-dimensional piecewise linear maps.

1. Introduction

In the ergodic theory of dynamical systems the central objects to be studied are the measures invariant under time evolution. Various parameters are associated to the invariant measures, namely the Lyapunov exponents, the entropy, the information dimension, Ruelle's rotation number. As emphasized by several authors [1] the characterization of the dynamics is in general not exhausted by the ergodic parameters listed above. Other ergodic parameters may play a useful role in the characterization of the probabilistic properties of dynamical systems. The following simple argument [2] shows why this must be so. Any given ergodic parameter is defined by an infinite time limit. This quantity will fluctuate and in general the fluctuations will not be Gaussian. The quantity describing the distribution of the fluctuations is again an ergodic invariant and the same reasoning applies in turn to its fluctuations, etc.

The possible existence of an infinite number of independent ergodic invariants raises the question of its construction and classification. Constructions related to the reasoning above have been used to characterize the range of possible fluctuations around the Lyapunov exponent and to obtain "generalized Lyapunov exponents" which seem appropriate to characterize intermittent behavior [3,4].

In this paper we use a somewhat different approach to the construction of a family of ergodic invariants for a dynamical system. Rather than relating the new invariants to the fluctuations of a previously known quantity, we try to analyse the problem anew, attempting to extract the invariants from a general characterization of the dynamics. One of the aims of this approach would be to obtain a sort of complete set of ergodic invariants. This goal we do not claim to have achieved, in particular because the notion of completeness is not easy to characterize.

Our basic tool is a variational principle for general maps [5] which describes the dynamics of the system by the critical points of an action functional. The functional may be constructed both for conservative and non-conservative maps. The action functional being critical if and only if the trajectories are determined by the dynamical system, it should contain all the information about the system. Thus suggests that whatever invariant dynamical information may be obtained from the system it may also be obtained from the action. Of particular interest are the analytical properties of the action in the neighborhood of the critical points. It is then natural to look for a family of invariants in the coordinate invariant features of the Hessian, i.e. in the properties of its eigenvalue distribution.

In section 2 we review briefly the variational formulation for dissipative maps proposed in ref. [5] and in
section 3 a family of ergodic invariants is constructed. Finally in section 4 several examples are worked out.

2. Variational formulation for dissipative maps

Let \((M, f)\) be a differentiable discrete time dynamical system defined by the mapping \(f: M \rightarrow M\), \(M\) being an open set of \(\mathbb{R}^d\). Now we define the space \(Y_N^x\) of ordered \((N+1)\)-tuples with initial condition \(x_0\),

\[Y_N^x = \{ x = (x_0, x_1, \ldots, x_N); \ x_0 \text{ fixed}, \ x_1 \ldots x_N \in M \}. \tag{2.1}\]

An orbit segment of \(f\) with initial condition \(x_0\) is an element of \(Y_N^x\) for which \(x_i = f(x_{i-1})\).

In \(Y_N^x\) we define a topology by the metric

\[d(x, y) = \sup |x_i - y_i|. \]

The dual \(Y_N^x^*\) is the space of continuous functionals \(Y_N^x \rightarrow \mathbb{R}\). A differentiable functional has a stationary (critical) point if \(\frac{\partial F}{\partial x_i} = 0\), \(i = 1, \ldots, N\; \alpha = 1, \ldots, d\).

The strong inverse variational problem is:

Given a dynamical system \((M, f)\) find a family of functionals \(A_N \in Y_N^x \rightarrow \mathbb{R}\) that are stationary if and only if \(x \in Y_N^x\) \((\forall x_0, \forall N)\) is an orbit segment of \(f\).

Notice that it is not required that \(\frac{\partial A_N}{\partial x_i} = 0\) coincides with the equations of motion. Actually when this requirement is imposed on \(A_N\) the problem has in general no solution unless \((M, f)\) is a conservative system.

By contrast the inverse variational problem as formulated above has the general solution \[A_N = \sum_{\alpha=1}^{d} \sum_{k \neq j \geq 0} [x_i^\alpha - f^\alpha(x_{i-1})]G_{kj}[x_j^\alpha - f^\alpha(x_{j-1})], \quad i = 1 \quad j \geq 0 \]

with

\[G_{kj} = c_k \delta_{k,j} \quad \tag{2.2}\]

or

\[G_{kj} = c_k \delta_{k,j+1} \quad \text{(N even)} \].

Both choices are solutions to the strong inverse variational problem. The restriction to \(N\) even in (2.4) comes from the "if and only if" condition. The nature of the critical points is in general very different for the choices (2.3) and (2.4). Of particular interest for the results of the next sections is the particular case \(G_{kj} = \delta_{k,j}\).

3. Ergodic invariants

Here we consider the functional

\[A_N = \sum_{\alpha=1}^{d} \sum_{k \neq j \geq 0} [x_i^\alpha - f^\alpha(x_{i-1})][x_j^\alpha - f^\alpha(x_{j-1})]. \tag{3.1}\]

Since there is a one-to-one correspondence between critical points of \(A_N\) and orbits of the dynamical system, it is natural to assume that the essential dynamical information on the system is contained in the analytical structure of \(A_N\) in the neighborhood of the critical points. The first non-trivial information is in the Hessian \(H_N\) of the critical points, namely in its eigenvalue distribution,
\[ \frac{1}{2} H_N = \frac{1}{2} \frac{\partial^2 A_N}{\partial x^\alpha \partial x^\beta} = \delta_{\alpha \beta} \delta_{j,k} - (1 - \delta_{j,N}) \frac{\partial f^\alpha(x_k)}{\partial x^\beta} - (1 - \delta_{j,N}) \delta_{j,k-1} \frac{\partial f^\beta(x_j)}{\partial x^\alpha} + \delta_{j,k} (1 - \delta_{j,N}) \frac{\partial f^\beta(x_j)}{\partial x^\alpha} \frac{\partial f^\beta(x_j)}{\partial x^\beta} \].

(3.2)

**Lemma.** The eigenvalues of the Hessian \( H_N \) of the functional \( A_N \) at the critical points are all positive.

From (3.2) one obtains for the second variation of \( A_N \) on the orbits (critical points)

\[ \frac{1}{2} \delta^2 A_N = \sum_{\alpha,j} \delta x^\alpha_j - (1 - \delta_{j-1,N}) \sum_{\beta} \delta x^\beta_{j-1} \frac{\partial f^\alpha(x_{j-1})}{\partial x^\beta_{j-1}} \left( \delta x^\alpha_j - (1 - \delta_{j-1,N}) \sum_{\beta} \delta x^\beta_{j-1} \frac{\partial f^\alpha(x_{j-1})}{\partial x^\beta_{j-1}} \right). \]

(3.3)

Because \( \delta^2 A_N \geq 0 \), the eigenvalues cannot be negative. On the other hand if \( \delta^2 A_N = 0 \) each term in (3.3) must vanish, i.e.

\[ \left( \delta x^\alpha_j - (1 - \delta_{j-1,N}) \sum_{\beta} \delta x^\beta_{j-1} \frac{\partial f^\alpha(x_{j-1})}{\partial x^\beta_{j-1}} \right) = 0 \quad \forall \alpha, j. \]

For \( j = 1 \) one obtains \( \delta x^\alpha_0 \delta x^\alpha_0 = 0 \) (recall that \( x_0 \) is the initial condition, not a functional variable). Therefore \( \delta^2 A_N = 0 \) implies \( \delta x^\alpha_j = 0 \forall \alpha, j \), i.e. there are no zero eigenvalues.

The next result shows that the momenta of the eigenvalue distribution are ergodic invariants.

**Theorem.** Let \((M, \nu)\) be a measure space, \( \nu \) a probability measure and \( f \) a measurable mapping in \( M \) leaving \( \nu \) invariant. Let \( \mu_p^{(N)} \) be the \( p \)-momentum of the eigenvalue distribution of the Hessian \( H_N \) associated to \( f \) by eq. (3.2).

\[ \mu_1^{(N)} = \frac{1}{N} \sum_{i=1}^{N} \lambda_i, \quad (3.4a) \]

\[ \mu_p^{(N)} = \frac{1}{N} \sum_{i=1}^{N} (\lambda_i - \bar{\lambda})^p, \quad p = 2, 3, \ldots. \]

(3.4b)

Then, provided the \( \mu_p^{(N)} \) are of class \( L^1(\nu) \), the limit \( \lim_{N \to \infty} N^{-1} \mu_p^{(N)} = \mu_p \) exists. (In (3.4b) we have used the notation \( \bar{\lambda} = \mu_1 \).)

The quantities \( \mu_p^{(N)} \) are related to the traces of powers of the Hessian

\[ \mu_1^{(N)} = \frac{1}{N} \text{Tr} H_N. \]

(3.5a)

\[ \mu_p^{(N)} = \frac{1}{N} \sum_{k=0}^{p} \binom{p}{k} (-1)^k \text{Tr} H_N^{p-k} \bar{\lambda}^k. \]

(3.5b)

From (3.2) it follows that each diagonal element of \( H_N \) is a function involving only coordinates defined at \( q \) different times,

\[ (H_N)_{\alpha,\alpha} = (H_N)_{\alpha,\alpha}(x_{i-\epsilon}, \ldots, x_i, \ldots, x_{i+\epsilon}) . \]
with $\epsilon + \epsilon' = q - 1$ and $|\epsilon - \epsilon'| \leq 1$.

Instead of $H^q$, consider the matrix $H^{q}$ as a block of dimension $N$ imbedded in a larger matrix of dimension $N+q$ in such a way that all diagonal elements of $H^q$ have the generic form and are not truncated by the effects of proximity of the matrix boundaries. Then the sequence of functions $\text{Tr} H^q \equiv q - 1$ and $|\epsilon - \epsilon'| \leq 1$.

Instead of $\epsilon + \epsilon' = q - 1$ and $|\epsilon - \epsilon'| \leq 1$. Consider the matrix $H^q$ as a block of dimension $N$ imbedded in a larger matrix of dimension $N+q$ in such a way that all diagonal elements of $H^q$ have the generic form and are not truncated by the effects of proximity of the matrix boundaries. Then the sequence of functions $\text{Tr} H^q$ satisfies the additivity condition

\[ \text{Tr} H^q_{N+q} = \text{Tr} H^q(x) + \text{Tr} H^q(f(M)x) . \]

Because the eigenvalues are all positive $\int \text{Tr} H^q(x) \, d\nu > 0$.

For a probability space with a measure $\nu$ and a measure-preserving transformation $f$, Kingman's sub-additive ergodic theorem [6,7] states that, if a sequence $(g_n)$ of $L^1$-functions satisfies $g_{n+k} \leq g_n + g_k f^n$ almost everywhere (a.e.) and $\exists M \geq 0$ such that $\int g_n \, d\nu \geq -M$, then $\lim n^{-1} g_n$ exists almost everywhere and

\[ \int \lim \frac{1}{n} g_n \, d\nu = \lim \frac{1}{n} \int g_n \, d\nu \]

for every invariant measurable set $A$. All conditions of the sub-additive ergodic theorem are fulfilled for $\text{Tr} H^q$ and therefore $\lim_{N \to \infty} N^{-1} \text{Tr} H^q(x)$ exists $\nu$-almost-everywhere. Furthermore for every measurable invariant set $A$

\[ \int \lim_{N \to \infty} \frac{1}{N} \text{Tr} H^q(x) \, d\nu = \lim_{N \to \infty} \frac{1}{N} \int \text{Tr} H^q(x) \, d\nu . \]

Then the same statements hold for $\lim_{N \to \infty} N^{-1} \text{Tr} H^q(x)$ because the sums in the traces differ at most in $q$ terms.

Existence of these limits implies the existence of the $\mu_p$'s.

Explicit expressions for the invariants $\mu_p$ as functions of the orbit coordinates are easily obtained from (3.2) and (3.5). For example for one-dimensional mappings $x_{n+1} = f(x_n)$ one obtains

\[ \frac{1}{2} \mu_1 = 1 + \lim_{N \to \infty} \frac{1}{N} \sum_i \left[ f'(x_i) \right]^2 , \quad (3.6a) \]
\[ \frac{1}{2} \mu_2 = \lim_{N \to \infty} \frac{1}{N} \sum_i \left[ f'(x_i) \right]^4 + 2 \mu_1 - \frac{1}{2} \mu_1 - 3 , \quad (3.6b) \]
\[ \frac{1}{2} \mu_3 = \lim_{N \to \infty} \frac{1}{N} \sum_i \left[ f'(x_i) \right]^6 + 3 \lim_{N \to \infty} \frac{1}{N} \sum_i \left[ f'(x_i) \right]^2 \left[ f'(x_i) \right] + \frac{1}{2} \mu_2 - \frac{1}{2} \mu_1 + \frac{3}{2} \mu_2 - \frac{1}{2} \mu_1 \mu_2 - \frac{1}{2} . \quad (3.6c) \]

4. Hessian spectrum and ergodic invariants. Examples

From (3.2) and (3.5) the invariants $\mu_p$ are obtained as explicit functions of the orbit coordinates and may in all cases be computed numerically by iterating the maps and taking limiting averages. In simple cases they may be obtained directly from the spectrum of the Hessian. Furthermore, the fact that the moments of the eigenvalue distribution are ergodic invariants, implies that the spectrum itself may be taken as ergodic characterization of the dynamical system. This point of view is emphasized in our second example where some features like the spectral gaps are found to be related to dynamical properties. Even the (numerically computed) integrated density of eigenvalues may provide useful characterizations of the dynamics (see ref. [5]).

The spectrum is found by solving the characteristic equation
\[ S_\lambda(\lambda) = \text{Det} \left[ \frac{1}{2} H_\lambda - \lambda \right] = 0. \]  

(4.1)

For one-dimensional maps \( x_{n+1} = f(x_n) \), \( H_\lambda \) is a tridiagonal matrix. The \( S_\lambda(\lambda) \) form a Sturm sequence of polynomials [8] with the recurrence relation

\[ S_0(\lambda) = 1, \quad S_1(\lambda) = 1 + \left[ f'(x_1) \right]^2 - \lambda, \quad \ldots, \]

(4.2a)

\[ S_{n+1}(\lambda) = 1 + \left[ f'(x_{n+1}) \right]^2 - \lambda, S_n(\lambda) = [f'(x_n)]^2 S_{n-1}(\lambda). \]

(4.2b)

In particular, if the motion is chaotic, this will be a random Sturm sequence.

We now discuss a few examples chosen from the class of piecewise linear maps, for which analytical results may be obtained.

Consider the following one-dimensional maps.

\[ x_{n+1} = 2ax_n, \quad x \in \left[ 0, \frac{1}{2a} \right], \]

(4.3)

\[ x_{n+1} = 2a(1-x_n), \quad x \in \left[ \frac{1}{2}, 1 \right], \]

(4.4)

They are sketched in fig. 1.

The maps (4.3) and (4.4) lead to identical Sturm sequences:

\[ S_0(2) = 1, \quad S_1(2) = 1 + 4a^2 - \lambda, \quad \ldots, \]

(4.5)

\[ S_{n+1}(2) = (1 + 4a^2 - \lambda)S_n(2) - 4a^2 S_{n-1}(2). \]

Using the results of ref. [9], it can easily be shown that

\[ S(\lambda) = (-1)^n \lambda \sin \arccos \left( \frac{\lambda - 4a^2 - 1}{4a} \right) \left\{ \sin \arccos \left( \frac{\lambda - 4a^2 - 1}{4a} \right) \right\}^{-1}. \]

(4.6)

(4.7)

From (4.6) one obtains for the Hessian spectrum

\[ \left| \lambda \right|_N = 1 + 4a^2 + 4a \cos \left( \frac{\pi k}{N+1} \right), \quad k = 1, \ldots, N. \]

The momenta of the eigenvalue distribution are obtained explicitly,

\[ \frac{1}{2} \mu_1^{(N)} = 1 + 4a^2, \quad \left( \frac{1}{2} \right)^n \mu_1^{(N)} = (4a)^n \frac{1}{N} \sum_{k=1}^{N} \cos \left( \frac{\pi k}{N+1} \right). \]

(4.8a)

In the limit \( N \to \infty \) one obtains for the invariants

\[ \frac{1}{2} \mu_1 = 1 + 4a^2, \]

\[ \left( \frac{1}{2} \right)^p \mu_p = \frac{(4a)^p}{\pi} \int_0^{\pi} (\cos u)^p du = (4a)^p \frac{(p-1)!!}{p!!}, \quad (p \text{ even}). \]

(4.8b)

\[ = 0 \quad (p \text{ odd}). \]

(4.8c)

Notice that, from the point of view of the invariants derived from the Hessian, the dynamical systems defined by eqs. (4.3) and (4.4) are entirely equivalent. This reflects the similarity of their measure properties. From ref. [10] it is indeed known that they share a lot of common properties. They both have a unique invariant measure and the dynamics induced by the Perron–Frobenius operator is asymptotically periodic.
Let us now turn our attention to the asymmetric roof map \cite{11} and a closely related monotonic piecewise linear system. These dynamical systems are defined by the equations ($a>0.5$)

\begin{align}
    x_{n+1} &= 2ax_n, \quad x \in [0, 1/2a], \\
    x_{n+1} &= \frac{2a(1-x_n)}{2a-1}, \quad x \in ]1/2a, 1], \\
\end{align}

(4.9)

and

\begin{align}
    x_{n+1} &= 2ax_n, \quad x \in [0, 1/2a], \\
    x_{n+1} &= \frac{2ax_n-1}{2a-1}, \quad x \in ]1/2a, 1].
\end{align}

(4.10)

These maps are sketched in fig. 2.

Eqs. (4.9) and (4.10) present a richer behavior than the previous maps (4.3), (4.4). The Sturm sequence (4.2) involves again identical coefficients for both systems; however, the recurrence depends now on the trajectories. Let us discuss some properties of the spectrum of the Hessian. Consider the eigenvalue problem

\begin{equation}
    \frac{1}{2} H_N u = \lambda u, \quad u = (u_1, \ldots, u_N),
\end{equation}

(4.11)

where $H_N$ is the Hessian matrix of eq. (3.2). Defining

\begin{equation}
    V_{l+1} = \begin{pmatrix} u_l \\ u_{l+1} \end{pmatrix}, \quad l = 1, \ldots, N-1,
\end{equation}

(4.12)

the eigenvalue problem (4.11) may be rewritten in the equivalent form

\begin{equation}
    V_{l+1} = T_l V_l, \quad l = 1, \ldots, N-1,
\end{equation}

(4.13)

where the transfer matrix $T_l$ for a one-dimensional map is

\begin{equation}
    T_l = \begin{pmatrix} 0 & 1 \\ f'(x_{l-1}) & f'(x_l) \end{pmatrix}, \quad f'(x) = \frac{1+f'(x)^2-\lambda}{f'(x)}.
\end{equation}

(4.14)

For the examples (4.9) and (4.10) we observe that the transfer matrix $T_l$ takes along the trajectory one of the following four values,

Fig. 1. (a) Mapping of eq. (4.3). (b) Mapping of eq. (4.4).

Fig. 2. (a) Asymmetric roof map, eq. (4.9). (b) Monotonic map, eq. (4.10).
The spectral properties of the Hessian may now be discussed using the same techniques that are used to study the vibrations of disordered lattices [12]. One might think in terms of a lattice with four different atom species leading to the four different transfer matrices \( (4.14a)-(4.14d) \). However, the analogy is not complete because, in a sequence generated by eq. (4.13), \( T^{(1)} \) and \( T^{(2)} \) can only be followed by \( T^{(1)} \) or \( T^{(3)} \) and \( T^{(3)} \) and \( T^{(4)} \) by \( T^{(2)} \) or \( T^{(4)} \). In particular repetition of \( T^{(2)} \) and \( T^{(3)} \) is not possible. Products of the matrices \( T^{(2)} \) and \( T^{(3)} \) repeat themselves when a trajectory oscillates between the two regions (L=left and R=right, see fig. 2) of the partition of the interval. Hence, to analyse the spectral gaps, rather than the transfer matrices \( T^{(2)} \) and \( T^{(3)} \), we consider the products

\[
T^{(2)}_+ T^{(3)}_+ = \left\{ \begin{array}{c}
\pm \frac{\beta}{\alpha} \\
\frac{1 + \beta^2 - \lambda}{\alpha} \\
\frac{1 + \alpha^2 - \lambda}{\alpha} 
\end{array} \right\} 
\]

(4.15a)

\[
T^{(3)}_+ T^{(2)}_+ = \left\{ \begin{array}{c}
\pm \frac{\alpha}{\beta} \\
\frac{1 + \alpha^2 - \lambda}{\alpha} \\
\frac{1 + \beta^2 - \lambda}{\beta} 
\end{array} \right\} 
\]

(4.15b)

where \( \alpha = 2a(2a-1)^{-1} \) and \( \beta = 2a \).

Using eqs. (4.14a,d) and (4.15a,b) we can now discuss the properties of the spectrum of \( H_N \) as for a tetraatomic disordered lattice. This analogy leads to the calculation of the spectral gaps which, according to the phase–angle representation [12], occur when the eigenvalues of the transfer matrices \( (4.14a,b) \) and \( (4.15a,b) \) are real. This implies the following allowed regions for the Hessian eigenvalues.

\[
\lambda \in \left[ 1 + 4a^2 - 4a; 1 + 4a^2 + 4a \right],
\]

(4.16a)

\[
\lambda \in \left[ 1 + \frac{4a^2}{(2a-1)^2} - \frac{4a}{2a-1} ; 1 + \frac{4a^2}{(2a-1)^2} + \frac{4a}{2a-1} \right],
\]

(4.16b)
Fig. 3. Spectral regions associated to the three types of transfer matrices and spectral gaps. \( \epsilon = (a - 1)/a \). The \( \lambda \)-axis is a symmetry axis for the graph.

\[
\lambda \in [1; 1 + \frac{1}{2}(\alpha^2 + \beta^2 - \sqrt{(\alpha^2 + \beta^2)^2 - 16\alpha\beta})] \quad \text{or} \quad [1 + \frac{1}{2}(\alpha^2 + \beta^2 + \sqrt{(\alpha^2 + \beta^2)^2 - 16\alpha\beta}); 1 + \alpha^2 + \beta^2].
\] (4.16c)

Eqs. (4.14a,d) imply the relations (4.16a,b) while the last relation (4.16c) follows both from (4.15a) and (4.15b). The allowed regions for the spectrum of the Hessian are the same for the maps (4.9) and (4.10). As in the first example, this is related to the similarity of their invariant measure properties. In fig. 3 we have marked the boundaries of the spectral regions associated to each one of the relations (4.16).

It is known [10,11] that both maps (4.9) and (4.10) \((a > \frac{1}{2})\) have the Lebesgue measure as their invariant measure. Therefore a typical trajectory will contain sequences of transfer matrices of the three types (4.14a), (4.14d) and (4.15). Invoking the Saxon–Hutner theorem [12] we conclude that spectral gaps exist in the intersection of the complements of the domains defined by (4.16). These are the shaded regions in fig. 3.

The density of eigenvalues in the complement of the spectral gaps is however expected to depend on the relative frequency of occurrence of sequences of each type. For example, for \( a \) in the interval \([\frac{1}{2}, \frac{1}{2} + \epsilon]\), where \( \epsilon \) is a small positive quantity, one has intermittent behavior with long sequences of \( T^{(1)} \) type dominating the dynamics. Then one expects relation (4.16a) to characterize the region of higher density of eigenvalues, with only a few eigenvalues lying in the complement of this region (excluding the gaps). This also suggests that the spectral distribution and gaps of the Hessian \( H_N \) might provide a rigorous characterization of intermittency.
References