Journal of Statistical Physics, Vol. 45, Nos. 1/2, 1986

A Note on Sinaï and Bunimovich's Markov Partition for Billiards

Yves-Emmanuel Lévy[1]

Received March 6, 1986; revision received June 16, 1986

I present a construction of Markov partitions related to a statistical description of a class of hyperbolic dynamical systems of $\mathbb{R}^2$ with singularities, through a general algorithm. This construction, correct, in contrast to previous attempts, applies in particular to billiards problems, yielding an easy-to-handle Markov partition; as an application, the incorrectness of a lemma of Bunimovich and Sinai (which was known to need improvement) is made clear.

KEY WORDS: Billiards; Markov partition; Markov process; dynamical system with singularity; hyperbolic mapping.

Many physical systems can be modeled by discrete-time dynamical systems. In particular, the physicist is interested in the properties of some nice invariant measure μ, the Bowen–Ruelle measure, which is, in ergodic cases, the unique distribution asymptotically reached by the set of the iterates of an arbitrary initial condition—except maybe for a set of initial conditions of zero probability. The Bowen–Ruelle measure can also be thought of as the probability measure reached as the limit of the evolution of any initially smooth density.

Little can be said in general about dynamical systems. Among them, one class can and has been described in an extensive way, the class of smooth (C^2) hyperbolic systems; this paper is a contribution to the understanding of hyperbolic systems with singularities, a larger class. Note that I restrict consideration in this paper to mappings of $\mathbb{R}^2$. I recall that hyperbolicity is the existence of a constant $\lambda > 1$ and, for almost every point x of the phase space, of two directions $e_u(x)$ and $e_s(x)$ (unstable and

[1] Centre de Physique Théorique (Laboratoire du CNRS LP 014), Ecole Polytechnique, F-91128 Palaiseau, France.

0022-4715/86/1000-0063$05.00/0

822/45/1-2-5

stable) along which the tangent mappings DT^n are expanding (resp. contracting) by a factor at least λ^n, for any $n \in \mathbb{N}$. When this makes sense, the two manifolds tangent, respectively, to these two vector fields are called unstable and stable manifolds.

The methods of symbolic dynamics have been shown to be a natural tool of great efficiency in the study of dynamical systems. In particular, they allow one to give a statistical description of some hyperbolic dynamical systems; it is this statistical description, which allows the use of the thermodynamic formalism, on which I focus here.

Symbolic dynamics makes use of a (finite or countable) partition of the phase space to code any point into an infinite chain of symbols, namely the sequence of the elements of the partition to which the successive (future and past) iterates of the point belong; the original mapping maps into the shift on these sequences. Various properties of dynamical systems can be studied via the analogous properties of the associated symbolic systems.[3] Among the possible partitions yielding such a correspondence, one class is of special interest for hyperbolic systems: the Markov partitions, which lead as a process on the infinite sequences of symbols to a special class known as topological Markov chains. These processes are defined by a transition matrix, the elements of which are 0's and 1's, forbidding or allowing the corresponding transition. As a property of partitions, the Markov property reads as some fitting condition to the foliation induced by the hyperbolicity, which is given explicitly below. The objects and concepts related to hyperbolicity take a very simple form in terms of these topological Markov chains.

The results which may be derived via the symbolic formalism are of two types: topological results (topological transitivity and mixing, topological entropy), and metric ones. Symbolic coding by Markov partitions has been used to construct and investigate the properties of an invariant measure via its conditional probabilities along the unstable and stable manifolds.[4,5] It also allows[6] the description of the Bowen–Ruelle measure as an equilibrium state for some potential related to the Jacobian of the map.

The billiards on $\mathbb{R}^2$ with a periodic configuration of scatterers (known as Sinaï billiards) lead to an interesting example of hyperbolic systems with singularities, where the singularity set corresponds to the trajectories tangent to some scatterer. For this system, an invariant measure, absolutely continuous with respect to the Lebesgue measure, is explicitly known. Bunimovich and Sinaï[1,2] use the statistical description of this measure to derive the ergodic and mixing properties of the system from the probabilistic properties of the finite Markov chains.

The present paper intends to clarify one question raised by the proofs in Ref. 1.[2] The point is that the partition the authors construct there does not have the Markov property, which is necessary to their description; I present here, in a slightly more pedagogical way, a correct construction which seems to be what the authors actually had in mind. My construction goes through a general algorithm, which does not apply only to billiards, and which makes the Markov partition an easy to use tool, with easy to check properties. Applied to Sinaï and Bunimovich's billiard problem, it allows one to make clear that a lemma in Ref. 1 is incorrect; the lemma relies on the finiteness of some class of elements of the Markov partition used in the canonical isomorphism, and we will see that there is no reason to assume this finiteness. This lemma provides some bound of the correlation from the past of the conditional transition probabilities; this estimate (Ref. 1, Lemma 6.6) is a key point necessary to derive the ergodic and mixing properties of the Bowen–Ruelle measure; the authors, aware of the problem raised by the proof of the lemma, have now made an (as yet unpublished) improvement of the proof.

In this paper, I deal with some mapping T of some invariant set $X \subset \mathbb{R}^2$ such that $\mu(X) = 1$, where μ is an invariant measure for T. The mapping T is hyperbolic and twice continuously differentiable except on a one-dimensional regular singularity curve S. Note that, in the case of a mapping with singularity, the stable and unstable manifolds present singularities as they cross S, or (some of) the images of S by the mapping. The mapping studied in Refs. 1 and 2 (which is such a T) is the Poincaré map of the periodic billiard dynamics associated with the successive reflections on scatterers; the two variables are the speed angle and an abscissa along the scatterer. Another example can be found in Ref. 7, where the authors study some particular hyperbolic piecewise linear (thus, singular) mappings which exhibit a strange attractor.

Before giving the construction of the Markov partition in our case, I recall the definition of Markov partitions. I call an unstable (stable) fiber any measurable subset of finite length of some regular component of the unstable (stable) manifold of some point. A subset C of X will be called a parallelogram if for any $x \in C$ there are unique unstable and stable fibers $c_u(x)$, $c_s(x) \subset C$ that allow us to describe the parallelogram in the following sense:

$$C = \bigcup_{y \in c_u(x)} \bigcup_{z \in c_s(x)} [c_s(y) \cap c_u(z)]$$

and such that, for y, $z \in C$, $c_s(y) \cap c_u(z)$ is a unique point.

[2] Note that throughout the present paper I have taken for simplicity $m = 1$ and set $S = S_{-1} \cup S_1$.

Note that, given a partition into parallelograms C, the corresponding decompositions of the space into unstable (resp. stable) fibers $c_u(x)$ $[c_s(x)]$ are unique. We can now state the Markov property: A partition η is Markovian when its elements are parallelograms and, moreover,

$$c_u(Tx) \subset Tc_u(x) \quad \text{and} \quad Tc_s(x) \subset c_s(Tx) \qquad \text{for any} \quad x \in X$$

where c_u (c_s) is the partition into unstable (stable) fibers induced by η.

In the regular case, the induced fibers $c_u(x)$ $[c_s(x)]$ are connected components of unstable (stable) manifolds; the elements of the partition are interiors of plain parallelograms, they are connected and finite in number. The existence of a singularity curve S makes the partition more intricate. Note in particular that a regular component of an unstable (stable) fiber cannot cross any T^nS, $n>0$ $(n<0)$. An element of the Markov partition must be a parallelogram; since the singularity curve S can neither cross the borders of a piece not be itself a border (since S is part neither of an unstable nor of a stable manifold), the elements of the partition have to be infinite in number; each element is disconnected and displays a Cantor-like structure in the two directions, in order to avoid S and all its images by the map.

I give now the construction of a Markov partition η for some hyperbolic systems in $\mathbb{R}^2$ with singularity. η will be a partition of X (e.g., the unit square in Ref. 1 or the strange attractor in Ref. 7). Let us suppose a finite pre-Markov partition η_0 has been constructed, which fulfills the hypotheses Ref. 1, Theorem 5. η_0 is a finite partition, whose elements are parallelograms with pieces of stable and unstable manifolds as borders except for a set of elements whose union covers a neighborhood D_0 of S and which are triangles, parallelograms, or pentagons and may admit S as a border. Moreover, except for this set of elements, η_0 has the Markov property.

In D_0, the refinement of η_0 by its images $T^{\pm 1}\eta_0$ produces some parallelograms, and leaves a smaller neighborhood D_1 of S of "bad elements." The iteration of this process yields a covering of the neighborhood of S by an infinite sequence of parallelograms, the nearer the smaller. The same construction must be performed around each image T^nS for n integer, and in a coherent way in order to get the Markov property; the neighborhoods of the T^nS in which the same hierarchical procedure is applied must be thought of as thin stripes approximately oriented in the unstable (stable) direction for $n>0 (n<0)$, which are cut out of previously connected parallelograms.

Here I introduce some convenient conventions. Let (q, p), for p and q in $\mathbb{N}$, be the partition $T^{-q}\eta_0 \vee T^{-q+1}\eta_0 \vee \cdots \vee T^{p-1}\eta_0 \vee T^p\eta_0$: note that

(p, q) is a refinement of η_0. Let $T^{\pm n}A$, for any set A, denote the union $T^n A \cup T^{-n}A$. For any partition α, let α denote the mapping that associates to x the element $\alpha(x)$ of α containing x (when this makes sense).

Let D_n be the union of the elements of (n, n) that are adjacent to S. I will assume that, as in Refs. 1 and 7, an exponentially decaying bound on $\mu(D_n)$ has been obtained. Let $E_n = D_n \backslash D_{n+1}$: E_n is the union of the parallelograms drawn by $(n+1, n+1)$ in D_n. Let $\Delta_p = \bigcup_{0 \leqslant k \leqslant p+1} T^{\pm k} D_p$; since $\mu(\limsup_p \Delta_p) = 0$, we can, by the Borel–Cantelli lemma, index almost every point with the function $P(x) = \text{Max}\{p \mid x \in \Delta_p\}$. Note that $P(x) = p$ implies not only $x \in \bigcup_{0 \leqslant k \leqslant p+1} T^{\pm k} D_p$, but also $x \in \bigcup_{0 \leqslant k \leqslant p+1} T^{\pm k} E_p$. Then, suppose that some x, with $P(x) = p$, belongs to D_p. It is natural to define $\eta(x) = (p+1, p+1)(x)$; thus, if x with $P(x) = p$ belongs to $T^n D_p$, $|n| < p+1$, it is natural to define $\eta(x) = (p+1-n, p+1+n)(x)$, which, I recall, is the image by T^n of some element of $(p+1, p+1)$.

I come now to the definition of η. Let $N(x) = \text{Min}\{n \leqslant 0 \mid x \in T^{\pm n} D_p\}$; for some $x \in X$, let $P(x) = p$ and $N(x) = n$ (p, thus n, is μ-almost-surely finite); then I define η by setting

$$x \in T^n E_p \backslash T^{-n} E_p \Rightarrow \eta(x) = (p+1-n,\ p+1+n)(x)$$

$$x \in T^n E_p \cap T^{-n} E_p \Rightarrow \eta(x) = (p+1+n,\ p+1+n)(x)$$

$$x \in T^{-n} E_p \backslash T^n E_p \Rightarrow \eta(x) = (p+1+n,\ p+1-n)(x)$$

The rephrasing of this definition in term of domains of equality of η with the (q, p)'s that follows is a useful tool; in particular, it allows a tedious but automatic check of the Markov property.

Let $\Delta_{n,p} = (\bigcup_{0 \leqslant k \leqslant n-1} T^{\pm k} D_p) \cup (\bigcup_{n \leqslant k \leqslant p+1} T^{\pm k} D_{p+1}) \cup (\bigcup_{k > p+1} T^{\pm k} D_{k-1})$. We have

$$\eta = (p+1-n, p+1+n) \qquad \text{on } (T^n E_p \backslash T^{-n} E_p) \backslash \Delta_{n,p}$$

$$\eta = (p+1-n, p+1+n) \qquad \text{on } (T^n E_p \cap T^{-n} E_p) \backslash \Delta_{n,p}$$

$$\eta = (p+1-n, p+1+n) \qquad \text{on } (T^{-n} E_p \backslash T^n E_p) \backslash \Delta_{n,p}$$

Note that a piece of the partition is a Cantor set, obtained as the complement of an infinite set of (unstable and stable) strips—that is, $\Delta_{n,p}$—cut out of a regular, connected parallelogram—that is, the element of $(p+1 \pm n, p+1 \pm n)$.

The authors of Ref. 1, in short, construct independently two partitions η_+ and η_-, which are adapted to the covering of the neighborhoods of, respectively, the $T^n S$ and the $T^{-n} S$ ($n \in \mathbb{N}$); η is then defined by $\eta = \eta_+ \vee \eta_-$. They define, more or less explicitly, indices $P_+(x)$, $N_+(x)$

allowing them to describe η_+, and similarly for η_-. The absence of the Markovian property comes, in this description, from the jumps experienced, along a trajectory, by P_+ and P_- when N_+ or N_- becomes zero.

I come now to the statement about the partition on which Ref. 1, Lemma 6.6 relies fundamentally. In the proof of this lemma it is assumed that the image by T of any element of η is exactly an element of η, except for a finite set of elements. Obviously, this exceptional set should contain all the elements of η obtained from parallelograms of (q, p) with $q=0$ or $p=0$, since η is a refinement of η_0; but the known properties of folding and expansion of the images of S prove that these elements are infinite in number.

Although the partition I construct does not allow one to get the desired results of Ref. 2, it is an efficient tool for the study of this type of dynamical system. This Markov partition is the generalization of the finite partitions used by Anosov to deal with C^2 hyperbolic mappings; in particular, an analogous partition was used in Ref. 7 to deal with the Lozi map, and to extend general results on entropy, Lyapunov exponents, and Hausdorff dimension of attractors to that mapping.

ACKNOWLEDGMENTS

The interest of Markov partitions to dynamical systems was first pointed out to me by J. Lascoux and P. Collet; I thank them and F. Ledrappier for very clarifying discussions.

REFERENCES

1. L. A. Bunimovich and Ya. G. Sinaï, Markov partitions for dispersed billiards, *Commun. Math. Phys.* **78**:247 (1980).
2. L. A. Bunimovich and Ya. G. Sinaï, Statistical properties of the Lorentz gas with a periodic configuration of scatterers, *Commun. Math. Phys.* **78**:479 (1981).
3. V. M. Alekseev and M. V. Yakobson, Symbolic dynamics and hyperbolic dynamic systems, *Phys. Rep.* **75**:287 (1981).
4. Ya. G. Sinaï, Markov partitions and *C*-diffeomorphisms, *Funct. Anal. Appl.* **2**:61 (1968).
5. Ya. G. Sinaï, Construction of Markov partitions, *Funct. Anal. Appl.* **2**:245 (1968).
6. Ya. G. Sinaï, Gibbs measures in ergodic theory, *Russ. Math. Surv.* **166**:21 (1972).
7. P. Collet and Y. E. Lévy, Ergodic properties of the Lozi mappings, *Commun. Math. Phys.* **93**:461 (1984).