Estimating a generating partition from observed time series: Symbolic shadowing

Yoshito Hirata,* Kevin Judd,† and Devin Kilminster‡
Centre for Applied Dynamics and Optimization, School of Mathematics and Statistics, The University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia

(Received 18 July 2003; revised manuscript received 10 February 2004; published 30 July 2004)

We propose a deterministic algorithm for approximating a generating partition from a time series using tessellations. Using data generated by Hénon and Ikeda maps, we demonstrate that the proposed method produces partitions that uniquely encode all the periodic points up to some order, and provide good estimates of the metric and topological entropies. The algorithm gives useful results even with a short noisy time series.

DOI: 10.1103/PhysRevE.70.016215 PACS number(s): 05.45.Tp, 05.45.Vx

I. INTRODUCTION

The theory of symbolic dynamics is a powerful tool for the investigation of discrete time dynamical systems. The main idea is that of a *partition*, that is, a finite collection $\mathcal{A}$ of disjoint subsets whose union is the state space M. By identifying each $A\in\mathcal{A}$ with a unique *symbol*, we have a sequence of symbols that correspond to each trajectory of the original system—the sequence is produced as the evolving state visits different regions of the partition. This idea is at its most powerful when the partition is chosen to be a *generating partition*, that is, when the assignment of symbol sequences to trajectories is unique, up to a set of measure zero. Generating partitions have been found to be desirable in applications such as chaotic communication [1], and in problems of parameter estimation [2].

As might be expected, generating partitions have often proven difficult to find. When the dynamical system is defined by a known map, there are methods by which to find generating partitions using "primary tangencies" [3] and also by consideration of unstable periodic orbits [4–6]. In cases where the dynamical system is not fully known, for example, in the practically important case in which only a time series of observations is available, there appears to be no previously published methods for obtaining a generating partition. However, recently Kennel and Buhl [7] proposed a method for finding a generating partition from a time series which uses "symbolic false nearest neighbors" to "localize" the region specified by a finite block of symbols.

We propose a method for estimating a generating partition from a time series. It is significantly different from that of Kennel and Buhl in that we approximate the generating partition by tessellations of state space and use certain fundamental properties of generating partitions observed by Eckmann and Ruelle [8]. In Sec. II, we define necessary notions and outline the conceptual basis of our algorithm. In Sec. III, we state our algorithm. In Sec. IV, we apply the proposed method to time series data generated by Hénon and Ikeda maps, show that the estimated partitions are close to those found in previous work [3,4], and show that they yield good estimates of the metric and topological entropies. The partition obtained from the algorithm may depend on how it is initialized, so in Sec. V, we examine the effects of initialization and discuss how to deal with them. In Sec. VI, we evaluate the proposed method and compare it with previous methods [4,7].

II. CONCEPTUAL BACKGROUND

We begin with several paragraphs that establish notation for some special partitions, then state a fundamental result concerning these partitions and generating partitions. Motivated by this fundamental result we develop the ideas that underpin and justify our methodology.

In all that follows, we shall assume that the state space M of the dynamical system in question is a bounded subset of d-dimensional Euclidean space with the usual two-norm $\|x\|$ for $x\in M$, and that the dynamics $f:M\to M$ are continuous. We also assume, for simplicity, that f can be inverted. (The following argument should be easily generalized for a noninvertible case.) In this case, an *initial condition* $x_0\in M$ gives rise to a unique *trajectory* $\cdots,x_{-1},x_0,x_1,\cdots\in M$, satisfying $x_t=f(x_{t-1})$. We suppose that we are given N consecutive points of a trajectory, and our problem is to estimate a generating partition from *this time series*.

A partition $\mathcal{A}$ of M is a finite collection of disjoint sets whose union is M. Given this partition, we define the function $\phi:M\to\mathcal{A}$, where $\phi(x)=A$ when $x\in A$. We call $\phi(x)$ the *symbol* assigned to x. We use a partition rather than a topological partition because we can define ϕ for all the points in M; Eckmann and Ruelle [8] likewise used a partition to define a generating partition. Corresponding to the trajectory $\cdots,x_{-1},x_0,x_1,\cdots$ is the *symbol sequence* $\cdots X_{-1}.X_0X_1\cdots$, where $X_t=\phi(x_t)$. We insert a dot in the symbol sequence to provide a place maker: it indicates the present, or *center*, X_0 of the symbol sequence, where the present, or center, occurs immediately after the dot.

As the initial point determines the symbol sequence, we can consider a map $\Phi:M\to\mathcal{A}^{\mathbb{Z}}$, which assigns to x_0 the symbol sequence $\Phi(x_0)=\cdots X_{-1}.X_0X_1\cdots$.

Let $\sigma:\mathcal{A}^{\mathbb{Z}}\to\mathcal{A}^{\mathbb{Z}}$ be the map such that $\sigma(\cdots X_{-1}.X_0X_1\cdots)=\cdots X_{-1}X_0.X_1\cdots$. We call σ the *shift map*, and a subset of $\mathcal{A}^{\mathbb{Z}}$ closed under σ a *shift space*. Observe that $\Phi(f(x_0))$

*Electronic address: yoshito@maths.uwa.edu.au
†Electronic address: kevin@maths.uwa.edu.au
‡Electronic address: devin@maths.uwa.edu.au

$=\cdots X_{-1}X_0.X_1\cdots=\sigma(\cdots X_{-1}.X_0X_1\cdots)=\sigma(\Phi(x_0))$, that is, the following diagram commutes:

$$\begin{array}{ccc} x & \overset{f}{\longrightarrow} & f(x) \\ \Phi\downarrow & & \Phi\downarrow \\ \Phi(x) & \overset{\sigma}{\longrightarrow} & \sigma(\Phi(x))=\Phi(f(x)). \end{array}$$

We say that $\mathcal{A}$ is a *generating partition* if Φ is one to one, up to a set of measure zero, on M and a subset of $\mathcal{A}^{\mathbb{Z}}$ that is a shift space. (See, for example, Lind and Marcus [Ref. [9], pp. 5,6] for a rigorous definition of such a shift space using "forbidden blocks.") The shift space in question is the image $\Phi(M)$, and on this shift space the map Φ has an inverse Φ^{-1} and we have $f(x)=\Phi^{-1}(\sigma(\Phi(x)))$. Therefore, when the partition is a generating partition, the original dynamics and the symbol dynamics are conjugate.

Consider a finite subsequence of consecutive symbols, which we will call a *substring*. For convenience, we denote the substring $X_{t_1}X_{t_1+1}\cdots X_{t_2}$ by $X_{[t_1,t_2]}$. For $m,n>0$, let $\Phi_{[-m,n]}:M\to\mathcal{A}^{m+n+1}$ be the map such that $\Phi_{[-m,n]}(x_0)=X_{[-m,n]}$. The map $\Phi_{[-m,n]}(x)$ gives the finite substring of $\Phi(x)$ which starts at $t=-m$ and ends at $t=n$. Observe that $\Phi_{[-m,n]}(M)$ is the set of all possible substrings of length $(m+n+1)$. More important for what follows, observe that the preimage $\Phi^{-1}_{[-m,n]}(X_{[-m,n]})$ is the set of points $x\in M$ that have the same substring $X_{[-m,n]}$.

The preimage $\Phi^{-1}_{[-m,n]}$ is defined on finite symbol sequences, however, it is useful to extend the definition of $\Phi^{-1}_{[-m,n]}$ to infinite sequences as follows. Let $\mathcal{X}$ denote an admissible infinite symbol sequence, $\cdots X_{-1}.X_0X_1\cdots$, in shift space $\Phi(M)$. Define $\Phi^{-1}_{[-m,n]}(\mathcal{X})=\Phi^{-1}_{[-m,n]}(X_{[-m,n]})$, that is, to find $\Phi^{-1}_{[-m,n]}(\mathcal{X})$, first extract the finite subsequence $X_{[-m,n]}$, then find the preimage $\Phi^{-1}_{[-m,n]}(X_{[-m,n]})$.

We are now in a position to state the key motivation for our algorithm. For a set $E\in M$, let $\mathrm{diam}(E)=\sup\{\|x-y\|:x,y\in E\}$. Call the maps $\Phi_{[-m,n]}$ *localizing* if $\sup_{\mathcal{X}}\mathrm{diam}(\Phi^{-1}_{[-k,k]}(\mathcal{X}))\to 0$ as $k\to\infty$. Eckmann and Ruelle [8] stated that if the $\Phi_{[-m,n]}$ are localizing, then this is a sufficient condition for $\mathcal{A}$ being a generating partition. Hence, a possible guide to finding a generating partition is to find a partition such that longer substrings $X_{[-m,n]}$ should specify smaller regions of points with the same substring.

When the $\Phi_{[-m,n]}$ are localizing, the set of points $\Phi^{-1}_{[-m,n]}(\mathcal{X})$ tends to be small for large m and n. Hence this set may be accurately located by a single point. For each substring $S\in\Phi_{[-m,n]}(M)$, assign a point $r_S\in M$ to be called the *representative* of $\Phi^{-1}_{[-m,n]}(S)$, typically we choose a point near the center of the set. It follows (from theorem 1 in the Appendix) that if $\Phi_{[-m,n]}$ are localizing, then the representatives can be chosen so that

$$\sup_{x\in M}\|x-r_{\Phi_{[-k,k]}(x)}\|\to 0,\quad k\to\infty. \tag{1}$$

Conversely, if $\sup_{x\in M}\|x-r_{\Phi_{[-k,k]}(x)}\|\to 0$ as $k\to\infty$, then $\Phi_{[-m,n]}$ are localizing (theorem 2). It is also true that if $\sup_{x\in M}\|x-r_{\Phi_{[-k,k]}(x)}\|\to 0$ as $k\to\infty$, then the partition is generating on M (theorem 3).

A set of representatives, for fixed m and n, can also be used to specify a partition by tessellating the state space using the representatives. For each substring $S\in\Phi_{[-m,n]}(M)$, we define its *tile* T_S to be the set of points in M that have the representative r_S as their nearest neighbor, that is,

$$T_S=\{x\in M:\|x-r_S\|\leqslant\|x-r_{S'}\|,\ \forall\, S'\in\Phi_{[-m,n]}(M)\}. \tag{2}$$

Then for each $A\in\mathcal{A}$ we collect all tiles T_S, for $S=S_{-m}\cdots S_n\in\Phi_{[-m,n]}(M)$, satisfying $S_0=A$, to form a set,

$$B_{A,[-m,n]}=\cup\{T_S:S\in\Phi_{[-m,n]}(M),S_0=A\}. \tag{3}$$

This set is a good approximation of A. In fact, when $\Phi_{[-m,n]}$ is localizing,

$$B_{A,[-k,k]}\to A,\quad k\to\infty, \tag{4}$$

where the limit is in the sense that if $B_i\to A$, then $\sup_{x\in A}\{\inf\|x-y\|:y\in B_i\}\to 0$ and $\sup_{y\in B_i}\{\inf\|x-y\|:x\in A\}\to 0$ (corollary 1).

Clearly, a method will find a generating partition if it finds $\Phi_{[-m,n]}$ that are localizing. Our aim is to do this approximately given only time series data $x_1,x_2,\cdots,x_N$. Equations (1) and (4) essentially imply conditions that can do this without explicitly constructing the $\Phi_{[-m,n]}$, that is, we need only construct suitable sets of representatives,

$$R_{[-m,n]}=\{r_S:S\in\Phi_{[-m,n]}(M)\}. \tag{5}$$

Equation (1) implies we require that $\max_i\|x_i-r_{\Phi_{[-m,n]}(x_i)}\|$ is small for large m and n. Therefore, we claim that the following minimization yields good estimates of the generating partition:

$$\min_{R_{[-m,n]},\mathcal{A}}\sum_{t=m+1}^{N-n}\|x_t-r_{\Phi_{[-m,n]}(x_t)}\|^2, \tag{6}$$

where the minimization is over the representatives and the partition. However, this minimization is awkward, because it requires explicit specification of the partition. However, it is sufficient to just label x_t by suitable symbols X_t, that is, we need to optimize over $X_1,X_2,\cdots,X_N$, given the representatives. Hence, the problem reduces to the minimization,

$$\min_{R_{[-m,n]},X_{[1,N]}}\sum_{t=m+1}^{N-n}\|x_t-r_{X_{[t-m,t+n]}}\|^2. \tag{7}$$

We call $\sum_{t=m+1}^{N-n}\|x-r_{X_{[t-m,t+n]}}\|^2$ the *discrepancy*, since it is a measure of how well each time series point is approximated by its corresponding representative. It is shown in theorem 4 that the partition is a generating partition if the dynamics are ergodic and the discrepancy divided by N goes to 0 as $N\to\infty$ and $m=n=k\to\infty$. As the discrepancy is non-negative, we try to minimize the discrepancy to estimate a generating partition.

One might refer to this method for estimating a generating partition as *symbolic shadowing*, because the original data

$\{x_t\}$ are eventually shadowed by the time series $\{r_{X_{[t-m,t+n]}}\}$. Specifically, there exists $\delta > 0$ such that any time series $\{x_t\}_{t=m+1}^{N}$ has a symbol sequence $X_{[1,N]}$ satisfying $\|x_t - r_{X_{[t-m,t+n]}}\| < \delta$ for $t = m+1, m+2, \cdots, N-n$. This is similar to the δ trace in work by Bowen [10]. For further information on shadowing, see work by Pilyugin [11] and by Palmer [12]. The idea of shadowing is applied to state estimation [13–15], noise reduction [16–19], and finding unstable periodic orbits [20–22].

To find an approximate solution to Eq. (7), we use an iterative algorithm that repeats the following two minimizations. First we fix a set of representatives and minimize the discrepancy over the symbol sequence:

$$\min_{X_{[1,N]}} \sum_{t=m+1}^{N-n} \|x_t - r_{X_{[t-m,t+n]}}\|^2. \tag{8}$$

Next we fix the symbol sequence and minimize the discrepancy over the set of representatives:

$$\min_{R_{[-m,n]}} \sum_{t=m+1}^{N-n} \|x_t - r_{X_{[t-m,t+n]}}\|^2. \tag{9}$$

This algorithm is similar to the Linde-Buzo-Gray (LBG) algorithm [23] in information theory [24]. In a way similar to that of Gray *et al.* [25], the feasibility of this algorithm can be shown [26].

Equation (9) is solved exactly, in a way similar to the least squared method. For each substring S, the solution is given by the least squares method:

$$r_S = \frac{1}{|\{i : X_{[i-m,i+n]} = S\}|} \sum_{\{i: X_{[i-m,i+n]}=S\}} x_i. \tag{10}$$

Optimizing Eq. (8) is difficult because it is a combinatorial optimization. However, we may find a *reasonable* approximation using the property of Eq. (4): We find for x_t the closest representative, whose substring is $S_{-m}\cdots S_n$. Then we assign $X_t = S_0$. This is expected to work because a longer substring specifies a smaller region for each x_t, and the closer symbols are to the center of the substring, the more significant they are expected be in locating x_t.

To begin the iteration of Eqs. (8) and (9) requires an initial set of representatives, and there are many ways that this can be done. One could try random initial representatives, or perhaps successive partitioning of the data into regions containing equal numbers of data points, then use the means of these as initial representatives. However, it is better to incorporate some basic dynamical information, for example, we know that unstable periodic points must have unique periodic symbol sequences [4,5]. Hence, a good initial set of representatives can be obtained by finding, and appropriately labeling, the lowest order unstable periodic points. This also has the advantage of providing a lower bound on the number of partitions, that is, the number of unique symbols required.

III. ALGORITHM

Our iterative algorithm can be stated as follows.

(1) We prepare an initial partition. First find unstable periodic points from a time series. Assign to each unstable periodic point a substring of length l of type $S_{-m}S_{-m+1}\cdots S_{-1}.S_0\cdots S_n$ ($m = \lfloor l/2 \rfloor$ and $n = \lfloor (l-1)/2 \rfloor$) over alphabet $\mathcal{A}$ so that the unstable periodic points are encoded uniquely. Let each unstable periodic point be the representative $r_{S_{[-m,n]}}$ of the substring $S_{[-m,n]} \in \mathcal{A}^l$.

(2) For each observed point x_t, find its closest representative $r_{S_{-m}S_{-m+1}\cdots S_{-1}.S_0\cdots S_n}$. Then make X_t to be S_0.

(3) Classify x_t depending on its substring $X_{[t-m,t+n]}$: Let $C_S = \{x_t : X_{[t-m,t+n]} = S, m+1 \leq t \leq N-n\}$. Set C_S is a set of points whose currently allocated substring is S.

(4) For each substring $S \in \mathcal{A}^l$, update its representative:

$$r_S = \sum_{y \in C_S} \frac{y}{|C_S|}. \tag{11}$$

(5) Return to step (2) until the set of representatives and the symbol sequence no longer change, or they cycle. (See comments below.)

(6) Increase the length of the substrings by $l \leftarrow l+1$, $m \leftarrow \lfloor l/2 \rfloor$ and $n \leftarrow \lfloor (l-1)/2 \rfloor$. Return to step (3) until a stopping criterion is achieved. (See comments below.)

There could be several possible stopping criteria. For example, one may stop the algorithm when the length of substrings reaches a certain length. Another idea is to use the *discrepancy*. One may stop the algorithm when $\sum_{i=m+1}^{N-n} \|x_i - R_{X_{[i-m,i+n]}}\|^2$ or $\max_i \|x_i - R_{X_{[i-m,i+n]}}\|^2$ becomes smaller than a certain prescribed value. An optimal stopping criterion is unresolved. In this paper, we stop when $\sum_{i=m+1}^{N-n} \|x_i - R_{X_{[i-m,i+n]}}\|^2/(N-m-n)$ is sufficiently small.

In the majority of tested cases, a stationary state is achieved in step (5), however, we have observed the algorithm to alternate between states in step (5), and cannot rule out other longer periodic behavior. This is a side effect of using tessellation to approximate the partition. We have always observed that the differences between alternating states were small.

IV. EXAMPLES

Since only a time series is required in order to apply the proposed method, we do not have to know the map of the dynamical system. Therefore, it can be used if one reconstructs state space by forming an embedding. However, in this paper, we use data sets generated from known models so that we can compare estimated partitions with those obtained in the literature.

The first example is the Hénon map [27]:

$$\begin{pmatrix} u_{t+1} \\ v_{t+1} \end{pmatrix} = \begin{pmatrix} 1 - a u_t^2 + b v_t \\ u_t \end{pmatrix}, \tag{12}$$

where $(a,b) = (1.4, 0.3)$. The following calculations take $x_t = (u_t, v_t)$, and use a time series of 10 000 data points.

In this example, we decided to stop the algorithm when $\sum_{i=m+1}^{N-n} \|x_i - R_{X_{[i-m,i+n]}}\|^2/(N-m-n) < (0.05)^2$. The algorithm

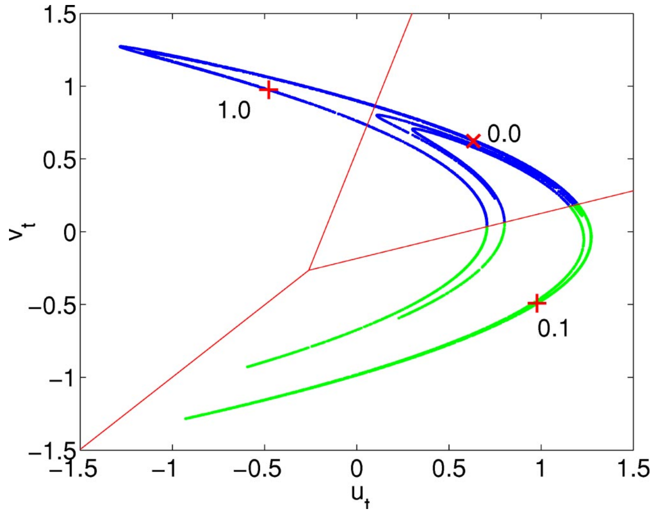

FIG. 1. (Color online) Initial partition $\Phi_{[-1,0]}^{-1}(S)$ for the Hénon map. The symbol × shows the fixed point, and + the periodic points of period 2 which are obtained from the time series using the method of Auerbach *et al.* [28]. Blue and green points indicate the points which were initially labeled by symbols 0 and 1, respectively.

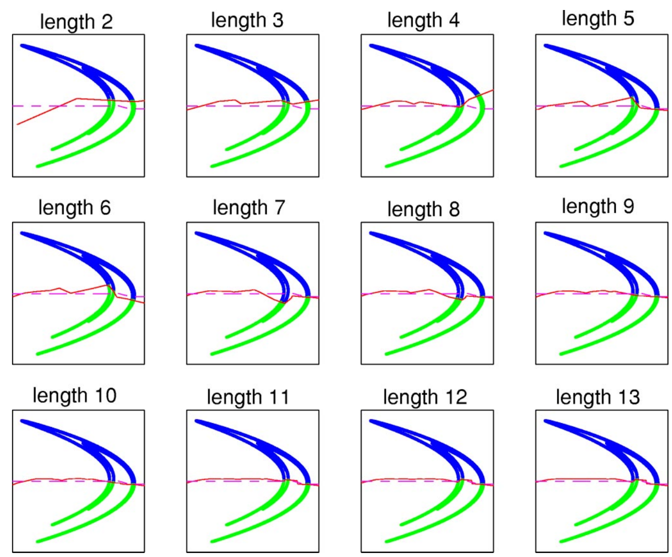

FIG. 2. (Color online) Refining the estimate for a generating partition during iteration, using the Hénon map. Blue and green points correspond to symbols 0 and 1, respectively. For each graph, the red solid line shows the partition line obtained using the proposed algorithm, and the magenta dashed line shows the one conjectured by Grassberger and Kantz [3].

was initialized using the lowest order unstable periodic points, or a fixed point and a period-2 orbit, which were found using the method of Auerbach *et al.* [28]. (There are more sophisticated methods for detecting the unstable periodic points from a time series, such as those given in Refs. [29,30]. However, in this case, we did not need to use them because given a time series of length 10 000 the method of Auerbach *et al.* [28] was able to detect periodic points up to order 2.) The period-1 point was assigned the initial substring 0.0 and the period-2 points arbitrarily assigned 0.1 and 1.0, as shown in Fig. 1. Figure 1 also shows the initial partition of sets $\Phi_{[-1,0]}(M)$.

During the iteration, the partition was changed, as shown in Fig. 2. We observed that convergence was not "monotonic," in the sense that the partitions are not nested, however, the approximate partition gradually became close to that conjectured by Grassberger and Kantz [3]. When the algorithm stopped, we obtained the partition shown in Fig 3. The final estimated partition used 883 representatives and the length of the substrings was 13.

To analyze the rate of convergence, we measured two quantities at each iteration: the *mean error* defined as $\sum_{t=m+1}^{N-n}\|x_t - R_{X_{[t-m,t+n]}}\|^2/(N-m-n)$ and the *maximum error* defined as $\max_t\|x_t - R_{X_{[t-m,t+n]}}\|^2$. In Fig. 4 we plot these quantities for the length of substring l. Both of the errors decreased, but not always monotonically.

We compare in Fig. 5 the symbol sequence of each length obtained with that generated from the same time series using the generating partition conjectured by Grassberger and Kantz [3], who found homoclinic tangencies and connected some of them to construct the partition. As the length of the substrings grew, the symbol sequence obtained came close to that generated using the partition of Grassberger and Kantz [3]. For substrings of length 13, it was found that of 10 000 symbols only 12 symbols were different between the two symbol sequences.

We also tested this partition by comparing the uniqueness of the symbol sequences assigned to unstable periodic points, and by comparing the values of the metric entropy and the topological entropy.

We obtained the unstable periodic points for the Hénon map using the method of Biham and Wenzel [31], and found that this partition encodes unstable periodic points uniquely up to period 17.

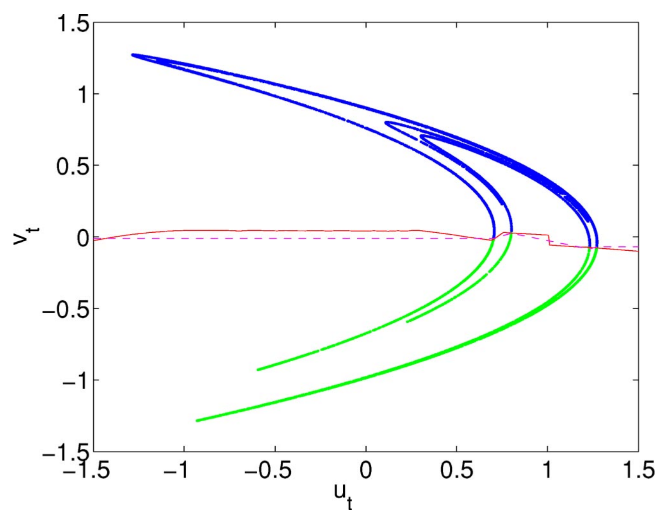

FIG. 3. (Color online) Partition $\mathcal{A}$ estimated using 10 000 data points generated from the Hénon map. Blue and green points are points in the time series corresponding to symbols 0 and 1, respectively. The red solid line shows the partition line obtained using the proposed method, and the magenta dashed line shows the generating partition conjectured by Grassberger and Kantz [3].

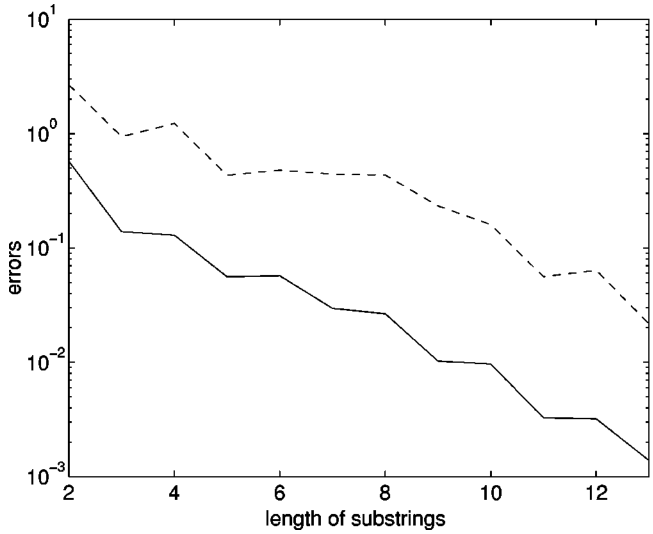

FIG. 4. Change in error during iteration of the algorithm, the Hénon map. The solid and broken lines show the mean and maximum errors, respectively.

Under the assumption that Pesin's identity holds, the metric entropy is equal to the sum of positive Lyapunov exponents [8], which is, in this case, 0.6048 using 2 for the logarithmic base [3]. (In what follows, we always use 2 for the logarithmic base when evaluating Lyapunov exponents, metric entropy, and topological entropy.) For the topological entropy, the most accurate value in the literature is 0.670 75 with a root mean square error of 0.000 04 [32]. Using the methods of Kennel and Mees [33] and of Hirata and of Mees [34], we estimated, from the symbol sequence obtained, metric entropy of 0.6223 and topological entropy of 0.6746, respectively, both of which agree well with the literature values.

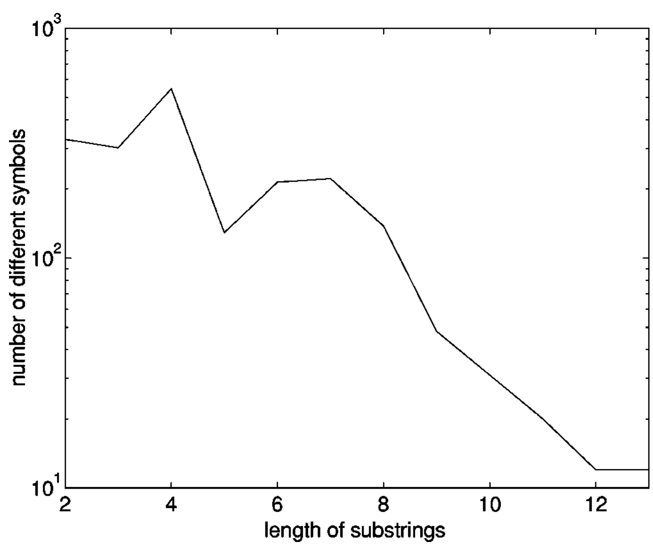

FIG. 5. Comparison of the symbol sequence obtained with that generated using the generating partition conjectured by Grassberger and Kantz [3], the Hénon map. For each length of substring, we counted the number of symbols in symbol sequence which did not agree with those assigned by the conjectured generating partition. As the length of the substrings grew, the number of "wrong" symbols, out of 10 000, decreased gradually, indicating that the symbol sequence obtained was close to the symbol sequence generated using the partition of Grassberger and Kantz.

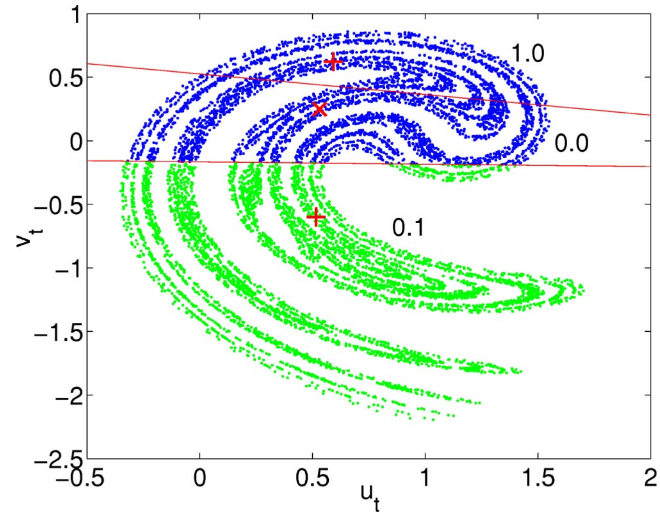

FIG. 6. (Color online) Initial partition for the Ikeda map. The initial substrings were assigned to unstable periodic points up to period 2 as shown. Blue and green points correspond to those with $\Phi_0(x)=0$ and 1, respectively.

The second example is the Ikeda map [35]:

$$\begin{pmatrix} u_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} 1 + a(u_t \cos\,\theta - v_t \sin\,\theta) \\ a(u_t \sin\,\theta + v_t \cos\,\theta) \end{pmatrix}, \tag{13}$$

where $\theta=0.4-b/(1+u_t^2+v_t^2)$, $a=6.0$, and $b=0.9$. We generated time series data of length 10 000.

First the algorithm was initialized by finding the low order unstable periodic points from the time series. We found unstable periodic orbits of periods 1 and 2 using the method of Auerbach *et al.* [28] and assigned substrings as shown in Fig. 6. We stopped the algorithm when the mean error became less than $(0.05)^2$, and obtained partition $\mathcal{A}$ shown in Fig. 7, which looks similar to that in Ref. [4]. It had the substrings of length 11 and 1386 representatives. The change in error is shown in Fig. 8. The mean error decreased monotonically, whereas the maximum error decreased gradually, but not monotonically.

We tested this partition using three indices: unique coding for unstable periodic points, the metric entropy, and the topological entropy.

We calculated the unstable periodic points using the method of Davidchack and Lai [36], and confirmed that this partition encodes the unstable periodic points uniquely up to period 8.

The metric entropy can be compared with the sum of the positive Lyapunov exponents under the assumption that Pesin's identity holds [8]. The numerical value for the positive Lyapunov exponent of the Ikeda map is known to be 0.726 [37].

For the topological entropy, we obtained the theoretical value using the numbers of unstable periodic points listed in Ref. [36]. Let $N(p)$ be the number of fixed points for the

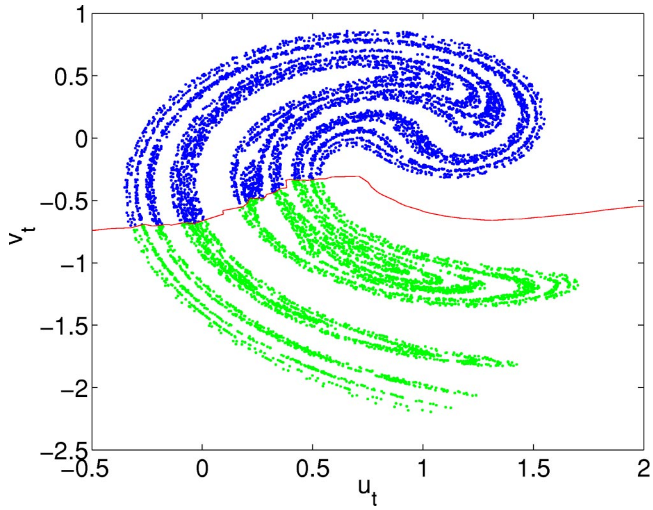

FIG. 7. (Color online) Partition estimated using 10 000 data points generated from the Ikeda map. The blue and green points are points in the time series encoded using symbols 0 and 1, respectively.

p-time map. The topological entropy is approximated by $1/p \log N(p)$. Averaging over those values obtained for periods between 14 and 22, one would get the estimate of topological entropy 0.8685 with its root mean square 0.0009.

The methods of Kennel and Mees [33] and of Hirata and Mees [34] were applied for the symbol sequence obtained from the data, and gave metric entropy of 0.7578 and topological entropy of 0.8752, respectively. These estimates have errors of 0.03 and 0.01, respectively.

We also tested the proposed algorithm using 50 000 points data of the Ikeda map, whose first 10 000 points were the same as the previous data. We applied the algorithm with the stopping criterion attaining substrings of length 15. We confirmed that the estimated partition could encode the unstable periodic points uniquely up to period 12, and yielded a metric entropy of 0.7450 and a topological entropy of 0.8748. These improvements demonstrate that a longer time series and longer substrings give a more accurate estimate of a generating partition.

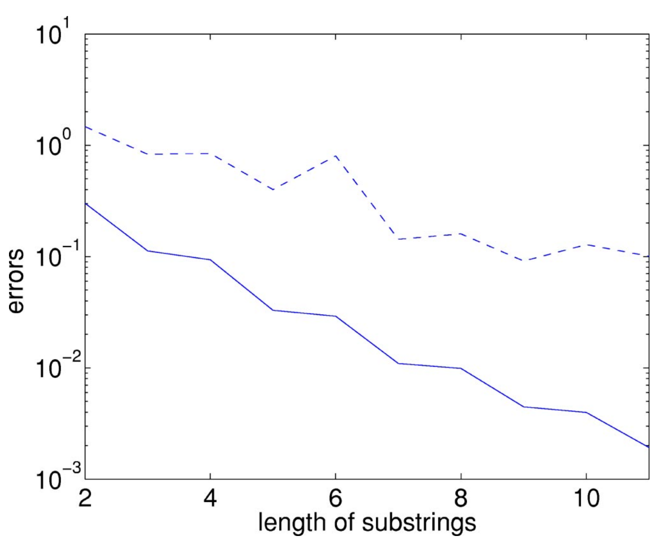

FIG. 8. Change in error during iterations for the Ikeda map. The solid and dashed lines show the mean and maximum errors, respectively.

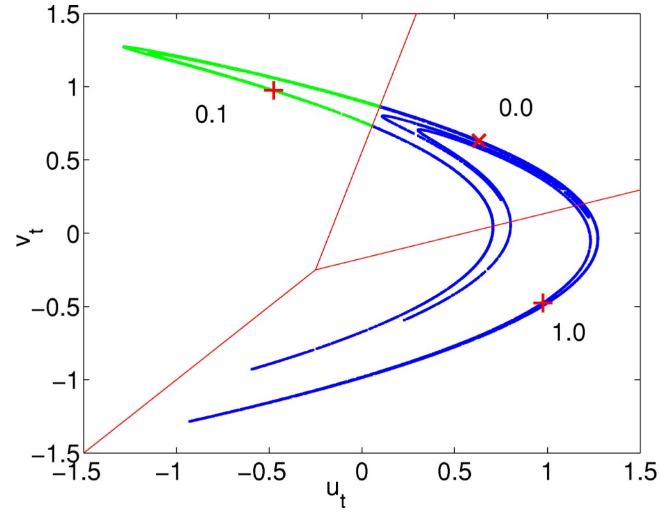

FIG. 9. (Color online) Assigning the initial condition in the opposite way from Fig. 1.

V. CONSEQUENCE OF BAD INITIAL REPRESENTATIVES

Unfortunately, sometimes the algorithm fails due to poor choice of labeling of the initial representatives. To make the dependence clear, we again used the data generated from the Hénon map, but this time, we assigned the initial condition in the opposite way for period-2 points from that in Sec. IV (Fig. 1), as shown in Fig. 9, and applied the same procedure with the same parameters. For comparison, we stopped the algorithm when it had attained substrings of length 13. We obtained the partition shown in Fig. 10.

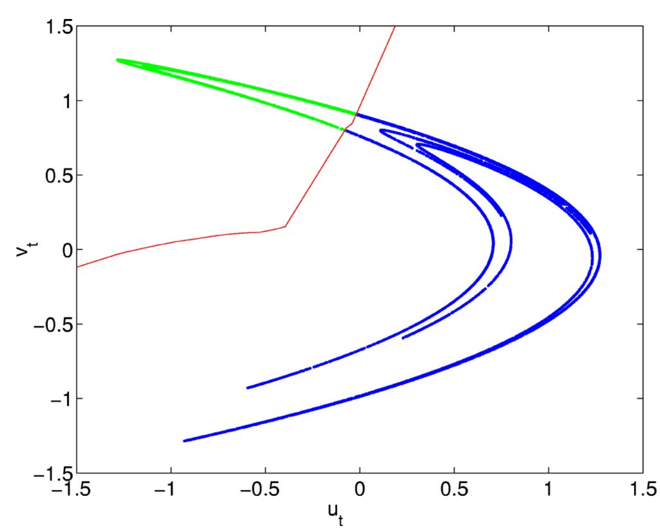

FIG. 10. (Color online) Partition obtained using a different initial partition. Blue and green points correspond to symbols 0 and 1, respectively. We exchanged the initial substrings for unstable periodic points of period 2 and attempted to find a generating partition.

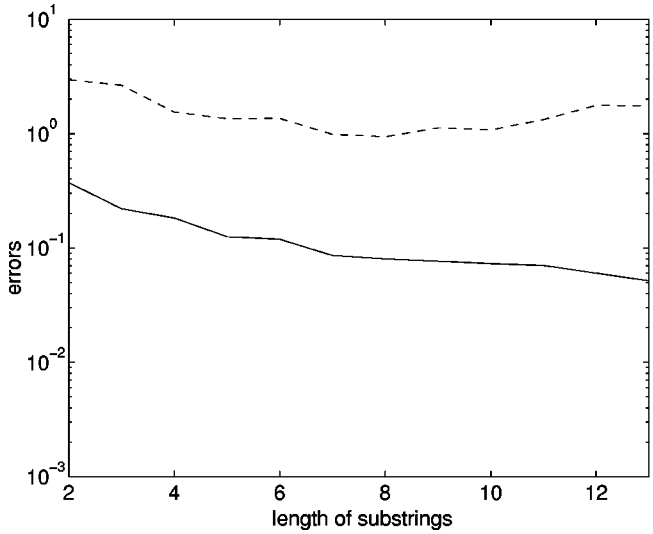

FIG. 11. Change in error when we assign the initial labels for the representatives as shown in Fig. 9. The solid and broken lines show the mean and maximum errors, respectively.

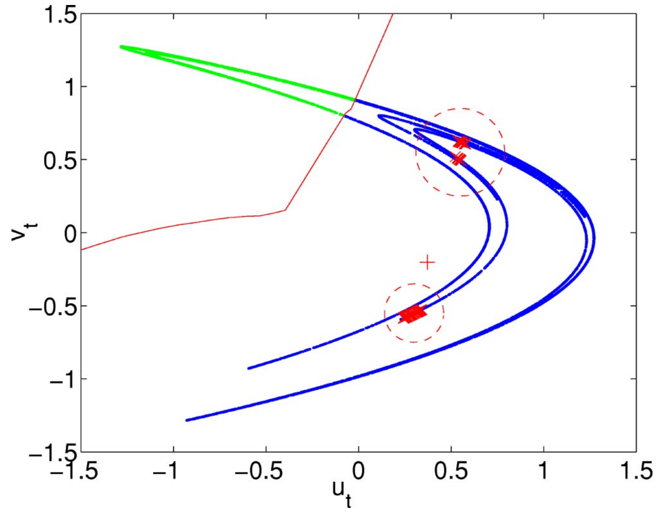

FIG. 13. (Color online) Example of representatives which are badly placed. In this example, the data of the Hénon map, the algorithm was started with the initial partition in Fig. 9. Marked by × are the points with substring 010 100.000 0101. These points are in the three disconnected regions. Averaging over these points gave +, which is their representative.

In the case of the Ikeda map, giving the initial symbols for period-2 points in the opposite way did not make much difference in the results except for the fact that the regions for symbols 0 and 1 were swapped.

A. Fixing a failure of the initial partition

As the example of the Hénon map shows, the success of the algorithm always relies on good choice and labeling of the initial representatives.

Failure of the algorithm can be detected easily as discussed in Sec. V B. To fix the problem, we first try different labels for the representatives; there is always only a small number of them for any given choice of representatives. If relabeling does not solve the problem, then this would suggest that the number of representatives is insufficient or the number of symbols is insufficient. Hence, we may have to increase the number of representatives, for example, by increasing the length of initial substring l, or we may have to increase the number of symbols.

B. Detecting a failure of the initial partition

Fortunately, an ill-prepared initial partition can be easily identified. We propose two ways by which to check the algorithm's progress. (When the algorithm succeeds, we do not observe any of the following.)

One criterion is to check the decrease in errors. When we chose the poor initial partition for the Hénon map, the mean error did not decrease satisfactorily (Fig. 11), because the maximum error returned to the initial level. This suggests that the maximum error can be an indicator.

A second criterion is the distribution of the representatives. When the algorithm fails, we tend to find some of the representatives are out of the attractor. Before explaining the reason why, we first compare the distributions. In the first Hénon map setting, all the representatives were found to be on the attractor [Fig. 12(a)]. But in the opposite setting, some of the representatives were found to be off the attractor [Fig. 12(b)]. This phenomenon can be explained as follows. In Fig. 12(b), we observe that one of the representatives that corresponds to a substring 010 100.000 0101 is well outside the attractor. Figure 13 shows the all data points whose substring is 010 100.000 0101. We can see that these points are found in three different regions of the attractor that are not contiguous. This is counter to requirements of Eq. (1). These points, with the same substring, are not well localized or well

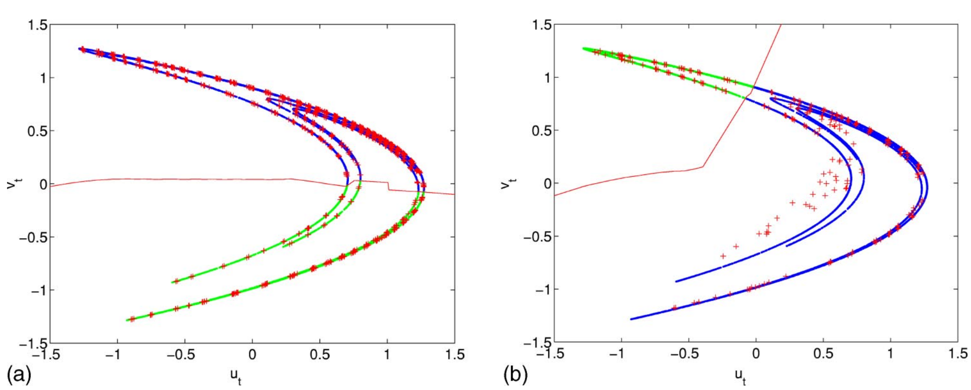

FIG. 12. (Color online) Distribution of representatives. (a) In the original setting (Fig. 1), all representatives are located on the attractor. But in (b), the opposite setting (Fig. 9), some representatives are located out of the attractor. This shows that the partition obtained from the opposite setting is bad.

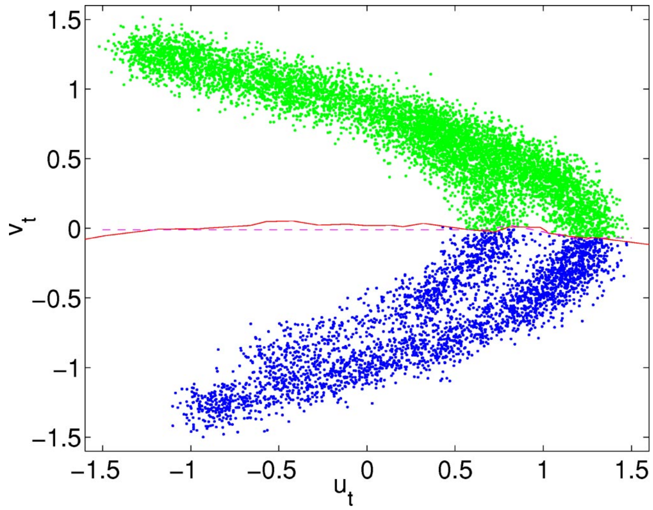

FIG. 14. (Color online) Estimating a generating partition using 10 000 point data of the Hénon map, contaminated by 10% Gaussian noise. Blue and green dots show points with symbols 0 and 1, respectively. The red solid line shows the partition line obtained using the proposed method and, the magenta dashed line shows the generating partition conjectured by Grassberger and Kantz [3].

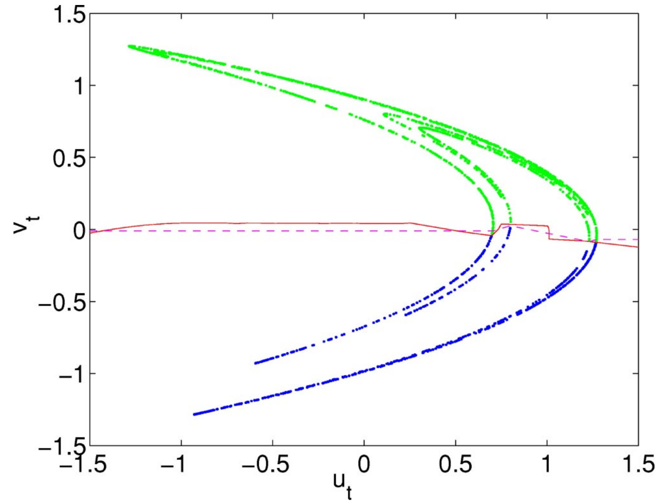

FIG. 15. (Color online) Estimating a generating partition using 2000 point clean data of the Hénon map. Blue and green dots show points with symbols of 0 and 1, respectively. The red solid line shows the partition line obtained using the proposed method and the magenta dashed line shows the generating partition conjectured by Grassberger and Kantz [3].

represented by one point. Consequently, the measures of error do not decrease, because these points prevent the errors going to zero.

VI. EVALUATION

There are two other methods for estimating a generating partition. When a system is given, a commonly used technique is to encode the unstable periodic points up to certain order uniquely [4,5]. Recently, Kennel and Buhl [7] proposed another algorithm for estimating a generating partition from a time series. The proposed method has a rigorous justification, given in the Appendix, but the method of Kennel and Buhl [7] does not.

The proposed algorithm gives useful results for noisy time series. We tested the noisy case by adding 10% Gaussian noise to the 10 000 points data of the Hénon map. Instead of detecting the unstable periodic points, we initialized the algorithm in the following way. First we split the time series at the median of v_t. Second we encoded each point with the symbol 0 if its v_t is smaller than the median, and with symbol 1 otherwise. In this way, we obtained the initial symbol sequence. Third we started the algorithm from step (3) by classifying points of the time series using their substrings of length 2. We stopped the algorithm when it converged with substrings of length 13.

The partition obtained from the noisy time series is shown in Fig. 14, which we note is close to the partition conjectured by Grassberger and Kantz [3]. Despite the noise, only 22 points out of 10 000 points are labeled differently than the partition conjectured by Grassberger and Kantz. It was confirmed that this partition encodes the unstable periodic points uniquely up to order 15. Using the symbol sequence obtained, we estimated 0.626 for the metric entropy and 0.677 for the topological entropy. This agrees well with their theoretical values 0.6048 [3] and 0.670 75 [32], respectively. The results in this noisy case are comparable to the results of Kennel and Buhl [7]. The proposed method is much better than identifying and labeling unstable periodic points from the noisy time series, because it is also hard to ensure that one detects all the unstable periodic points up to a certain order from a noisy time series.

The proposed algorithm gave useful results with a short time series. We took the first 2000 points of the clean data of the Hénon map and applied the algorithm in the same way we did for the noisy case. The partition obtained is shown in Fig. 15. Out of 2000 symbols, this partition has 12 symbols different from the partition conjectured by Grassberger and Kantz [3]. This partition encodes the unstable periodic points uniquely up to period 14. Using this partition, we estimated 0.642 for the metric entropy and 0.684 for the topological entropy, both of which are close to their theoretical values 0.6048 [3] and 0.670 75 [32], respectively.

The proposed algorithm needs to assign fewer parameters in advance than that of Kennel and Buhl [7].

The proposed algorithm is deterministic, that is, given the same initial condition, it gives the same solution. However, the algorithm of Kennel and Buhl [7] uses "differential evolution" [38], a genetic-type algorithm, which is stochastic optimization. Therefore, each application of the algorithm can generate a different answer. Differential evolution runs fast in general because it can reduce the number evaluations of the cost functions [38].

The proposed algorithm runs fast because the cost function is simple to solve and we try to avoid evaluating the cost function as much as possible. The proposed algorithm needs to evaluate the cost function only when one considers stopping the algorithm. We tested the speed of the proposed algorithm using a computer with CPU Pentinum III 1 GHz and 256 MB memory. The program was written and run using MATLAB 6.5. When we used the 10 000 point data of the

Hénon map, the algorithm took 564 s from step (2) to the end. When we used the first 2000 points in the above test, the algorithm took 63 s. The noise did not affect the performance much. When we add 10% Gaussian noise to the 10 000 point data in the above test, the algorithm took 589 s. If we do not require in step (5) that the symbol sequence converges completely, then computation is faster.

VII. CONCLUSION

We proposed a method for estimating a generating partition from observed time series data generated from an invertible map. The partition is approximated by tessellating state space with representatives, that is, points in state space, each of which has a distinct substring of a certain length. Using our scheme, we stated the problem of finding a generating partition as finding the minimum discrepancy between a series of points in the data and one specified by a symbol sequence and representatives. By solving this minimization problem approximately using an iterative algorithm, we found an estimate for a generating partition.

We applied our method for time series of Hénon and Ikeda maps. Estimates of generating partitions obtained from 10 000 data points uniquely encoded all periodic points of order less than 18 for the Hénon map and 9 for the Ikeda map, respectively. They also gave reasonable values for the metric and topological entropies.

Details of the proof for the proposed method and implementation for short and noisy time series are discussed elsewhere [26].

ACKNOWLEDGMENTS

We are indebted to Dr. Leonard A. Smith for his helpful comments on the discrepancy between points in time series data and their representatives. We also acknowledge Dr. Des Hill for checking the draft. Y.H. would like to thank Dr. Gary Froyland for his encouraging comments, Dr. Matthew B. Kennel and the Institute for Nonlinear Science at the University of California, San Diego, for discussions and hospitality, and Professor Alistair I. Mees, Prediction Company and School of Mathematics and Statistics, the University of Western Australia, for financial support.

APPENDIX

Here we briefly describe proof of the theorems. Let μ be the invariant measure of the system.

Theorem 1. If $\mathcal{A}$ is localizing, there exists a function $r_{[-m,n]}(x)=r_{\Phi_{[-m,n]}(x)}$ such that $\lim_{k\to\infty}\sup_{x\in M}\|r_{[-k,k]}(x)-x\|=0$.

Proof. Let $\mathcal{X}$ be a symbol sequence that corresponds to x. Define $r^*_{[-m,n]}(x)$ by

$$\frac{1}{\mu(\Phi^{-1}_{[-m,n]}(\mathcal{X}))}\int_{\Phi^{-1}_{[-m,n]}(\mathcal{X})} y\,d\mu(y). \tag{A1}$$

For the lth coordinate, we have the following inequalities:

$$\begin{aligned}|x^l - r^*_{[-m,n]}(x)^l| &= \left|\frac{1}{\mu(\Phi^{-1}_{[-m,n]}(\mathcal{X}))}\int_{\Phi^{-1}_{[-m,n]}(\mathcal{X})}(x^l-y^l)d\mu(y)\right| \\ &\le \frac{1}{\mu(\Phi^{-1}_{[-m,n]}(\mathcal{X}))}\int_{\Phi^{-1}_{[-m,n]}(\mathcal{X})}|x^l-y^l|d\mu(y) \\ &\le \frac{1}{\mu(\Phi^{-1}_{[-m,n]}(X))}\int_{\Phi^{-1}_{[-m,n]}(\mathcal{X})}\|x-y\|d\mu(y) \\ &\le \frac{\int_{\Phi^{-1}_{[-m,n]}(\mathcal{X})}\operatorname{diam}(\Phi^{-1}_{[-m,n]}(\mathcal{X}))d\mu(y)}{\mu(\Phi^{-1}_{[-m,n]}(\mathcal{X}))} \\ &\le \frac{\mu(\Phi^{-1}_{[-m,n]}(\mathcal{X}))\operatorname{diam}(\Phi^{-1}_{[-m,n]}(\mathcal{X}))}{\mu(\Phi^{-1}_{[-m,n]}(\mathcal{X}))} \\ &\le \operatorname{diam}(\Phi^{-1}_{[-m,n]}(\mathcal{X})),\end{aligned}$$

resulting in

$$|x^l - r^*_{[-m,n]}(x)^l| \le \sup_{\mathcal{X}}\operatorname{diam}(\Phi^{-1}_{[-m,n]}(\mathcal{X})). \tag{A2}$$

Therefore, if $\lim_{k\to\infty}\sup_{\mathcal{X}}\operatorname{diam}(\Phi^{-1}_{[-k,k]}(\mathcal{X}))=0$, then $\lim_{k\to\infty}|x^l-r^*_{[-k,k]}(x)^l|=0$. When $|x^l-r^*_{[-m,n]}(x)^l|<\epsilon<1$, we have $|x^l-r^*_{[-m,n]}(x)^l|^2<|x^l-r^*_{[-m,n]}(x)^l|$. Therefore, it follows that

$$\begin{aligned}\|x - r^*_{[-m,n]}(x)\|^2 = \sum_l |x^l - r^*_{[-m,n]}(x)^l|^2 &< \sum_l |x^l - r^*_{[-m,n]}(x)^l| \\ &\le \sum_l \sup_{\mathcal{X}}\operatorname{diam}(\Phi^{-1}_{[-m,n]}(\mathcal{X})) \\ &\le d\sup_{\mathcal{X}}\operatorname{diam}(\Phi^{-1}_{[-m,n]}(\mathcal{X})).\end{aligned}$$

Taking the supremum over M, we have $\sup_{x\in M}\|x-r^*_{[-m,n]}(x)\|^2\le d\sup_{\mathcal{X}}\operatorname{diam}(\Phi^{-1}_{[-m,n]}(\mathcal{X}))$. Hence, if $\lim_{k\to\infty}\sup_{\mathcal{X}}\operatorname{diam}(\Phi^{-1}_{[-k,k]}(\mathcal{X}))=0$, then $\lim_{k\to\infty}\sup_{x\in M}\|x-r^*_{[-k,k]}(x)\|^2=0$. We can conclude that is $r_{[-m,n]}(x)$ satisfies the desired property.

The converse of theorem 1 also holds, that is, the following theorem.

Theorem 2. If $\lim_{k\to\infty}\sup_{x\in M}\|r_{[-k,k]}(x)-x\|=0$, then $\mathcal{A}$ is localizing.

Proof. Let $\mathcal{X}$ be a symbol sequence for $x\in M$. Suppose there exists $y,z\in M$ such that $\Phi_{[-m,n]}(x)=\Phi_{[-m,n]}(y)=\Phi_{[-m,n]}(z)$.

$$\begin{aligned}\|y-z\| &\le \|y-r_{\Phi_{[-m,n]}(y)}\| + \|z-r_{\Phi_{[-m,n]}(y)}\| \\ &= \|y-r_{\Phi_{[-m,n]}(y)}\| + \|z-r_{\Phi_{[-m,n]}(z)}\| \\ &\le 2\sup_{w\in M}\|w-r_{\Phi_{[-m,n]}(w)}\|.\end{aligned}$$

By taking the supremum over $y,z\in A$, we have

$$\operatorname{diam}\Phi^{-1}_{[-m,n]}(\mathcal{X}) = \sup_{y,z\in\Phi^{-1}_{[-m,n]}(\mathcal{X})} \|y-z\| \leq 2\sup_{w\in M}\|w - r_{\Phi_{[-m,n]}(w)}\|.$$

Taking the supremum over all the admissible symbol sequences, we have

$$\sup_{\mathcal{X}} \operatorname{diam}\,\Phi^{-1}_{[-m,n]}(\mathcal{X}) \leq 2\sup_{w\in M}\|w - r_{\Phi_{[-m,n]}(w)}\|. \quad \text{(A3)}$$

Because $\sup_{w\in M}\|w-r_{\Phi_{[-k,k]}(w)}\|\to 0$ as $k\to\infty$, it follows that

$$\lim_{k\to\infty}\sup_{\mathcal{X}} \operatorname{diam}\Phi^{-1}_{[-k,k]}(\mathcal{X}) = 0, \quad \text{(A4)}$$

indicating that $\mathcal{A}$ is localizing.

Consequently, when we choose the representatives using Eq. (A1), condition $\lim_{k\to\infty}\sup\|r_{[-k,k]}(x)-x\|=0$ is equivalent to localizing.

Lemma 1. Let M_0 be a subset of M. If $\lim_{k\to\infty}\|x-r_{[-k,k]}(x)\|=0$ for every $x\in M_0$, then $\mathcal{A}$ assigns a unique symbol sequence to each point on M_0.

Proof. Take any two distinct points $x,y\in M_0$. Then there exists $\epsilon>0$ such that $\|x-y\|>\epsilon>0$. The triangle inequalities make the following inequalities true:

$$\begin{aligned}
&\|r_{[-k,k]}(x) - r_{[-k,k]}(y)\| \\
&\quad = \|(x-y) + \{(r_{[-k,k]}(x) - x) - (r_{[-k,k]}(y) - y)\}\| \\
&\quad \geq \|x-y\| - \|(r_{[-k,k]}(x) - x) - (r_{[-k,k]}(y) - y)\| \\
&\quad \geq \|x-y\| - \|r_{[-k,k]}(x) - x\| - \|r_{[-k,k]}(y) - y\|.
\end{aligned}$$

Then for $\epsilon>0$, there exists k_1 such that if $k\geq k_1$, then $\|x-r_{[-k,k]}(x)\|<\frac{1}{3}\epsilon$. Similarly, there exists k_2 such that if $k\geq k_2$, then $\|y-r_{[-k,k]}(y)\|<\frac{1}{3}\epsilon$. Let $k_0=\max\{k_1,k_2\}$. If $k\geq k_0$, then it follows that $\|r_{[-k,k]}(x)-r_{[-k,k]}(y)\|\geq\frac{1}{3}\epsilon>0$. As the distance between $r_{[-k,k]}(x)$ and $r_{[-k,k]}(y)$ is positive, two points $r_{[-k,k]}(x)$ and $r_{[-k,k]}(y)$ are different from each other, indicating that x and y are assigned different substrings. As any two points in M_0 have different substrings, a point in M_0 has its unique symbol sequence.

Therom 3. If $\lim_{k\to\infty}\sup_{x\in M}\|x-r_{[-k,k]}(x)\|=0$, then $\mathcal{A}$ is generating on M.

Proof. If $\lim_{k\to\infty}\sup_{x\in M}\|x-r_{[-k,k]}(x)\|=0$, then $\lim_{k\to\infty}\|x-r_{[-k,k]}(x)\|=0$ for every $x\in M$. Using lemma 1, the partition $\mathcal{A}$ assigns a unique symbol sequence to each point in M, meaning that $\mathcal{A}$ is a generating partition on M.

When f is ergodic on M, then

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N-m-n}\sum_{t=m+1}^{N-n}\|x_t - r_{[-m,n]}(x_t)\|^2 \\
&\quad = \int_M \|x - r_{[-m,n]}(x)\|^2 d\mu(x).
\end{aligned}$$

Denote this quantity by $H_{\mathcal{A},[-m,n]}$.

Theorem 4. If f is ergodic on M and $\lim_{k\to\infty}H_{\mathcal{A},[-k,k]}=0$, then the partition $\mathcal{A}$ is generating on M.

Proof. The following chain of equalities holds:

$$\begin{aligned}
&\lim_{k\to\infty}\int \|x - r_{[-k,k]}(x)\|^2 d\mu(x) \\
&\quad = \int \lim_{k\to\infty}\|x - r_{[-k,k]}(x)\|^2 d\mu(x) \\
&\quad = \int_{\{x:\,\lim_{k\to\infty}\|x-r_{[-k,k]}(x)\|=0\}} \lim_{k\to\infty}\|x - r_{[-k,k]}(x)\|^2 d\mu(x) \\
&\qquad + \int_{\{x:\,\lim_{k\to\infty}\|x-r_{[-k,k]}(x)\|>0\}} \lim_{k\to\infty}\|x - r_{[-k,k]}(x)\|^2 d\mu(x) \\
&\quad = \int_{\{x:\,\lim_{k\to\infty}\|x-r_{[-k,k]}(x)\|>0\}} \lim_{k\to\infty}\|x - r_{[-k,k]}(x)\|^2 d\mu(x) = 0,
\end{aligned}$$

where the first equality holds because as M is bounded, the inside $\|x-r_{[-k,k]}(x)\|^2$ of the integral is bounded, and we can swap the order of the integral and the limit.

To have the last equality valid, we need

$$\mu(\{x:\lim_{k\to\infty}\|x - r_{[-k,k]}(x)\| > 0\}) = 0, \quad \text{(A5)}$$

meaning that

$$\mu(\{x:\lim_{k\to\infty}\|x - r_{[-k,k]}(x)\| = 0\}) = 1. \quad \text{(A6)}$$

Therefore, from lemma 1, there exists a subset $M_0\subset M$ of full measure such that $\mathcal{A}$ assigns a unique symbol sequence to each point on M_0, indicating that $\mathcal{A}$ is generating on M.

This theorem means that if a partition has the property that the average distance between a point and its representative converges to 0, then any two points can be distinguished from each other with probability 1. Hence $\lim_{k\to\infty}H_{\mathcal{A},[-k,k]}=0$ gives a generating partition.

Theorem 5. If $\lim_{k\to\infty}\sup_{x\in M}\|x-r^*_{[-k,k]}(x)\|=0$, then $B_{A,[-k,k]}\to A$ in the sense that $\sup_{x\in A}\inf_{y\in B_{A,[-k,k]}}\|x-y\|\to 0$ and $\sup_{x\in B_{A,[-k,k]}}\inf_{y\in A}\|x-y\|\to 0$.

Proof. We first prove $\sup_{x\in A}\inf_{y\in B_{A,[-k,k]}}\|x-y\|\to 0$. We divide A into two sets: $A\cap B_{A,[-k,k]}$ and $A\cap B^c_{A,[-k,k]}$. For $x\in A\cap B_{A,[-k,k]}$, we have $\inf_{y\in B_{A,[-k,k]}}\|x-y\|=0$. For $x\in A\cap B^c_{A,[-k,k]}$, we have $r^*_{[-k,k]}(x)\in B_{\mathcal{A},[-k,k]}$ because $r^*_{[-k,k]}(x)$ is located at the center of the tile included in $B_{\mathcal{A},[-k,k]}$. Therefore, for $x\in A\cap B^c_{A,[-k,k]}$, the following inequality stands:

$$\inf_{y\in B_{A,[-k,k]}}\|x-y\| \leq \|x - r^*_{[-k,k]}(x)\|. \quad \text{(A7)}$$

Hence,

$$\sup_{x\in A}\inf_{y\in B_{A,[-k,k]}}\|x-y\| \leq \sup_{x\in A}\|x - r^*_{[-k,k]}(x)\| \to 0, \quad \text{(A8)}$$

resulting in the desired relation.

For the second part: $\sup_{y\in B_{A,[-k,k]}}\inf_{x\in A}\|x-y\|\to 0$, divide $B_{A,[-k,k]}$ into two sets: $B_{A,[-k,k]}\cap A$ and $B_{A,[-k,k]}\cap A^c$. For $y\in B_{A,[-k,k]}\cap A$, there exists $x\in A$ such that $\|x-y\|=0$. For $y\in B_{A,[-k,k]}\cap A^c$, let $c_{[-m,n]}(y)$ be a map of M into $\{r^*_{[-m,n]}(M)\}$, giving the closest representative to y in metric $\|\cdot\|$. From the triangle inequality, we also have $\|x-y\| \leq \|x-c_{[-k,k]}(y)\| + \|c_{[-k,k]}(y)-y\|$ for every $x \in A$. Because $c_{[-k,k]}(y)$ is the closest representative to y, $\|c_{[-k,k]}(y)-y\| \leq \|z-y\|$ for $z \in \{r^*_{[-k,k]}(M)\}$. Especially $\|c_{[-k,k]}(y)-y\| \leq \|r^*_{[-k,k]}(y)-y\|$. Hence, the following inequality is true for $x \in A$ and $y \in B_{A,[-k,k]}$:

$$\inf_{x\in A}\|x-y\| \leq \|x-y\| \leq \|x-c_{[-k,k]}(y)\| + \|c_{[-k,k]}(y)-y\| \leq \|x-c_{[-k,k]}(y)\| + \|r^*_{[-k,k]}(y)-y\|.$$

Because $y \in B_{A,[-k,k]}$, the representative $c_{[-k,k]}(y)$ is the one whose tile is included in $B_{A,[-k,k]}$. As $x \in A$ is arbitrary, we can choose x whose representative is $c_{[-k,k]}(y)$. Then, the first term has the upper bound,

$$\|x-c_{[-k,k]}(y)\| = \|x-r^*_{[-k,k]}(x)\| \leq \sup_{x\in A}\|x-r^*_{[-k,k]}(x)\|.$$

By taking the supremum over $B_{A,[-k,k]}$, we have

$$\begin{aligned}\sup_{y\in B_{A,[-k,k]}}\inf_{x\in A}\|x-y\| &\leq \sup_{y\in B_{A,[-k,k]}}\{\sup_{x\in A}\|x-r^*_{[-k,k]}(x)\| + \|r^*_{[-k,k]}(y)-y\|\}\\ &\leq \sup_{x\in A}\|x-r^*_{[-k,k]}(x)\| + \sup_{y\in B_{A,[-k,k]}}\|r^*_{[-k,k]}(y)-y\|\\ &\leq 2\sup_{x\in M}\|x-r^*_{[-k,k]}(x)\| \to 0.\end{aligned}$$

Therefore, the statement holds.

Corollary 1. If $\Phi_{\mathcal{A}}$ is localizing, then $B_{\mathcal{A},[-k,k]} \to A$ for each $A \in \mathcal{A}$.

Proof. Using theorem 1, the localizing property means that $\lim_{k\to\infty}\sup_{x\in M}\|r^*_{[-k,k]}(x)-x\| = 0$. From Theorem 5, the statement should follow.

[1] S. Hayes, C. Grebogi, and E. Ott, Phys. Rev. Lett. **70**, 3031 (1993).
[2] X. Z. Tang, E. R. Tracy, A. D. Boozer, A. de Brauw, and R. Brown, Phys. Rev. E **51**, 3871 (1995).
[3] P. Grassberger and H. Kantz, Phys. Lett. **113A**, 235 (1985).
[4] R. L. Davidchack, Y.-C. Lai, E. M. Bollt, and M. Dhamala, Phys. Rev. E **61**, 1353 (2000).
[5] J. Plumecoq and M. Lefranc, Physica D **144**, 231 (2000).
[6] J. Plumecoq and M. Lefranc, Physica D **144**, 259 (2000).
[7] M. B. Kennel and M. Buhl, Phys. Rev. Lett. **91**, 084102 (2003).
[8] J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).
[9] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding* (Cambridge University Press, Cambridge, UK, 1995).
[10] R. Bowen, J. Diff. Eqns. **18**, 333 (1975).
[11] S. Y. Pilyugin, *Shadowing in Dynamical Systems* (Springer, Berlin, 1999).
[12] K. Palmer, *Shadowing in Dynamical Systems: Theory and Applications* (Kluwer Academic, Dordrecht, The Netherlands, 2000).
[13] K. Judd and L. Smith, Physica D **151**, 125 (2001).
[14] K. Judd, L. Smith, and A. Weisheimer, Physica D **190**, 153 (2004).
[15] K. Judd, Physica D **183**, 273 (2003).
[16] S. M. Hammel, Phys. Lett. A **148**, 421 (1990).
[17] M. Davies, Int. J. Bifurcation Chaos Appl. Sci. Eng. **3**, 113 (1992).
[18] M. Davies, Physica D **79**, 174 (1994).
[19] D. Ridout and K. Judd, Physica D **165**, 26 (2002).
[20] B. A. Coomes, H. Koçak, and K. J. Palmer, Contemp. Math. **172**, 115 (1994).
[21] B. A. Coomes, H. Koçak, and K. J. Palmer, Numer. Algorithms **14**, 55 (1997).
[22] B. A. Coomes, H. Koçak, and K. J. Palmer, Aust. Math. Soc. Gaz. **24**, 183 (1997).
[23] Y. Linde, A. Buzo, and R. M. Gray, IEEE Trans. Commun. **COM-28**, 84 (1980).
[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1990).
[25] R. M. Gray, J. C. Kieffer, and Y. Linde, Inf. Control. **45**, 178 (1980).
[26] Y. Hirata, Ph.D. dissertation, School of Mathematics and Statistics, The University of Western Australia, Crawley, Australia, 2003.
[27] M. Hénon, Commun. Math. Phys. **50**, 69 (1976).
[28] D. Auerbach, P. Cvitanović, J.-P. Eckmann, G. Gunaratne, and I. Procaccia, Phys. Rev. Lett. **58**, 2387 (1987).
[29] P. So, E. Ott, S. J. Schiff, D. T. Kaplan, T. Sauer, and C. Grebogi, Phys. Rev. Lett. **76**, 4705 (1996).
[30] P. So, E. Ott, T. Sauer, B. J. Gluckman, C. Grebogi, and S. J. Schiff, Phys. Rev. E **55**, 5398 (1997).
[31] O. Biham and W. Wenzel, Phys. Rev. Lett. **63**, 819 (1989).
[32] J. Jacobs, E. Ott, and B. R. Hunt, Phys. Rev. E **57**, 6577 (1998).
[33] M. B. Kennel and A. I. Mees, Phys. Rev. E **66**, 056209 (2002).
[34] Y. Hirata and A. I. Mees, Phys. Rev. E **67**, 026205 (2003).
[35] K. Ikeda, Opt. Commun. **30**, 257 (1979).
[36] R. L. Davidchack and Y.-C. Lai, Phys. Rev. E **60**, 6172 (1999).
[37] R. Brown, P. Bryant, and H. D. Abarbanel, Phys. Rev. A **43**, 2787 (1991).
[38] R. Storn and K. Price, J. Global Optim. **11**, 341 (1997).