Title

Estimating good discrete partitions from observed data: Symbolic false nearest neighbors

Journal

Physical Review Letters, 91(8)

ISSN

0031-9007

Authors

Kennel, Matthew B

Buhl, Michael

Publication Date

2003-08-01

Peer reviewed

Estimating Good Discrete Partitions from Observed Data: Symbolic False Nearest Neighbors

Matthew B. Kennel* and Michael Buhl†

Institute For Nonlinear Science, University of California, San Diego, La Jolla, California 92093-0402, USA

(Received 1 October 2002; revised manuscript received 14 May 2003; published 21 August 2003)

A symbolic analysis of observed time series requires a discrete partition of a continuous state space containing the dynamics. A particular kind of partition, called "generating," preserves all deterministic dynamical information in the symbolic representation, but such partitions are not obvious beyond one dimension. Existing methods to find them require significant knowledge of the dynamical evolution operator. We introduce a statistic and algorithm to refine empirical partitions for symbolic state reconstruction. This method optimizes an essential property of a generating partition, avoiding topological degeneracies, by minimizing the number of "symbolic false nearest neighbors." It requires only the observed time series and is sensible even in the presence of noise when no truly generating partition is possible.

DOI: 10.1103/PhysRevLett.91.084102 PACS numbers: 05.45.Tp

State space reconstruction, in particular the time-delay embedding [1], is a universally popular key representation tool that opened experimentally observed time series to analysis as dynamical systems. The state is represented as sequences of vectors in some finite-dimensional vector space with continuous coordinates, very often Euclidean. A key property central to its popularity is that, given sufficient coordinates, *almost all reconstructions are generically satisfactory*. The reconstructed dynamics are equivalent to the underlying dynamics.

Instead of a continuous vector space, what about a discrete, low-precision representation—sequences of *symbols*? A discrete representation opens up the many powerful techniques of information and communication theory in addition to the connection between discrete mathematics and dynamical systems via the theoretical study of symbolic dynamics. Experimentally, symbolic techniques have attracted significant interest in a number of areas [2], but choosing the initial representation is still a problem. The obligatory discretization requires a *partition*: a coloring of the state space [3], $\mathbf{x} \in R^d$, into nonoverlapping regions and associated symbols so that any $\mathbf{x}$ is assigned a single symbol s from a finite alphabet, representable as the integers $0, 1, \ldots, A-1$. A partition $\mathcal{P}$ defines a discretization of the observed data sequence $\mathbf{x}_i$ $(i \in 1 \ldots N)$ into a symbolic sequence s_i. Which discretizations retain invariants of the original dynamics in the sequence of symbols? Unfortunately, the situation is unlike time-delay embedding: *simple partitions do not generically reconstruct the dynamics*. Are the *ad hoc* partitions often used still satisfactory? Unfortunately they are often not so. Bollt *et al.* [4] examined the degradation in the symbolic dynamics which results from the frequently used empirical "histogram partition." A suboptimal partition induces improper projections or degeneracies, where a single observed symbolic orbit may correspond to more than one topologically distinct state space orbit. This results in finding the wrong topological entropy, because some distinct transitions are improperly merged. Mathematically, the ideal is a so-called "generating partition" (GP), where symbolic orbits uniquely identify one continuous space orbit, and thus the symbolic dynamics is equivalent to the continuous space dynamics. Unfortunately there is no satisfactory general theory saying how to find a GP, except for one dimensional maps ($d = 1$), where one partitions at the critical points (minima, maxima, or discontinuities). Additionally, there is significant work on a class of chaotic communication schemes, where one targets specific orbits with small perturbations [5]. These optimally make use of a GP so that a digital transmission can be controlled to a unique orbit. Implementations so far have used only nearly $d = 1$ dynamics, perhaps because GPs were not known for higher dimensional oscillators.

For symbolic data analysis and experimental chaotic communication, a method to approximate good partitions from realistic observed data *alone*—not knowing the equations of motion—is urgently needed, and is the problem we attack. Davidchack *et al.* [6] recently described a partition algorithm which successively colors unstable periodic orbits (UPOs) to ensure unique codings (all UPOs have unique codes under a GP). This is practical if one already knows the dynamics, as the necessary high-order UPOs are very difficult to obtain from observed data alone. Reference [7] gives a similar method.

Any partition $\mathcal{P}$ of a stationary dynamical system yields a symbol stream which has a ($\mathcal{P}$-dependent) Shannon entropy rate h. The supremum of h—in the limit over all increasingly fine (high-alphabet) partitions—gives the dynamically invariant Kolmogorov-Sinai (KS) entropy rate h_{KS}. More strikingly, a GP also achieves this supremum with a finite, and, one hopes, small alphabet and simple partition. This suggests maximizing a statistical *estimator* of h, $\hat{h}$, over candidate partitions [8]. This apparently sensible algorithm is flawed as demonstrated by the following counterexample. Consider a partition of the state space with a fine box size ϵ where each region is *randomly* assigned a symbol from the alphabet. For

sufficiently small ϵ, any finite length symbol sequence will be indistinguishable from a memoryless information source with the maximal rate $h = \log A$, as each datum could have encountered a different partition element with a new random symbol. With the maximum possible entropy, this partition would be selected over competitors but is clearly wrong as the symbol stream says nothing about the original dynamics. Beyond this extreme case, there are other practical issues. First, estimation of h (asymptotic rate, not block entropy) is not trivial to do well. Second, with observational noise larger alphabets will inevitably appear to give significantly higher entropies even if they are not actually much better at encoding the dynamics. As the true entropy rate of the system is not already known (often a quantity one wants to estimate *given* a good partition), there is no absolute statistical target which confirms whether the proposed partition is at all close or far from the ideal. Selecting partitions with entropy estimates seems to fail in our practice.

In place of entropy, we assert our criterion for a good partition: short sequences of consecutive symbols should localize the corresponding continuous state space point as well as possible. A useful partition maintains the benefits of a low-precision symbolic representation with minimum distortion of the original state space dynamics. Our idea is to form a geometrical embedding of the symbolic sequence under the candidate partition and evaluate, and minimize, a statistic which quantifies the apparent errors in localizing state space points.

We embed the symbol sequence into the unit square [9]:

$$\mathbf{y}_i = \left(\sum_{k=1}^{k_{\max}} s_{i-(k-1)}/A^k, \sum_{k=1}^{k_{\max}} s_{i+k}/A^k \right) \tag{1}$$

($k_{\max}$ is chosen so $A^{-k_{\max}}$ is as small as the computational precision). For $A = 2$, the first coordinate of $\mathbf{y}_i$ is the binary fraction whose digits start at s_i and go backwards in time, the second is with the sequence going forward from s_{i+1}. The distribution on $\mathbf{y}$ is like a $\mathcal{P}$-dependent symbolic version of the invariant measure.

Given $\mathbf{x}_i$ and a partition $\mathcal{P}$, the symbolic embedding (1) yields a parallel series $\mathbf{y}_i$, defining points on some map $\mathbf{y} = \phi_{\mathcal{P}}(\mathbf{x})$. We want $\phi_{\mathcal{P}}$ to be *injective*, i.e., $\phi_{\mathcal{P}}(\mathbf{x}) = \phi_{\mathcal{P}}(\mathbf{x}')$ implies $\mathbf{x} = \mathbf{x}'$. With finite data, we desire that if $\|\phi_{\mathcal{P}}(\mathbf{x}) - \phi_{\mathcal{P}}(\mathbf{x}')\|$ is small, so is $\|\mathbf{x} - \mathbf{x}'\|$. By construction, sufficiently near points in $\mathbf{x}$ have close symbolic sequences in their most significant digits. In a good partition, additionally, nearby points in $\mathbf{y}$ remain close when mapped back into the $\mathbf{x}$ space. By contrast, bad partitions induce topological degeneracies where similar symbolic words map back to globally distant regions of state space, the problem described in [4].

We quantify how well any candidate partition achieves our ideal. We find the *nearest neighbor*, in Euclidean distance, to each point $\mathbf{y}_i$. Conventional k-d tree algorithms [10] efficiently provide the index of the nearest neighbor to any point in a data set, denoted $\mathcal{N}[i] = \text{argmin}_{k \neq i} \|\mathbf{y}_k - \mathbf{y}_i\|$. We find the distance of those same two points back in $\mathbf{x}$ space, $D_i = \|\mathbf{x}_{\mathcal{N}[i]} - \mathbf{x}_i\|$. We normalize the set of D_i by a monotonic transformation: given any D, find its rank $R \in [0, 1]$ in the cumulative distribution of *random* two-point distances $\|\mathbf{x}_\alpha - \mathbf{x}_\beta\|$. Large R means that localizing well in symbol space did not localize well in the original state space.

Better partitions give fewer *symbolic false nearest neighbors*, i.e., a smaller J_{sfnn}, defined to be the proportion of the R_i greater than some threshold η. This resembles the false neighbors statistic for time-delay embeddings [11]: both count large-deviation "mistakes" in a related space which result from topological misembedding in the tested space. Appropriate values for η which define a macroscopic deviation are $\eta \approx 0.01$–0.3, depending on the noise in $\mathbf{x}_i$. An alternative statistic is K_{sfnn}, defined to be the arithmetic average of the largest γ percentile of the set of R_i. Using J_{sfnn}, η may need tuning depending on the noise scale and dynamical system, but the effect of varying γ is lower. On the downside, K_{sfnn} does not necessarily converge to near zero for the optimal partition. We typically find good results with $\gamma \approx 0.01$–0.05, that is, averaging the largest 1%–5% of the R_i. The thresholds should be set so that the best $\mathcal{P}$ gives as few R_i as possible above the size of the effective noise in $\mathbf{x}$, η in particular one might view as the largest reasonably expected noise plus nearest neighbor distance. The interpretation of γ is less clear but the method is less sensitive to γ.

For concrete calculations, we represent $\mathcal{P}$ with a small number of free parameters. Inspired by [6], we define partitions with respect to a mA sized set of radial-basis "influence functions," $f_k(\mathbf{x}) = \alpha_k / \|\mathbf{x} - \mathbf{z}_k\|^2$, each with a preassigned associated symbol S_k. For each $\mathbf{x}$, one $f_l(\mathbf{x})$ will generically result in a larger value than all other $f_k(\mathbf{x})$, $k \neq l$. Then $\mathcal{P}(\mathbf{x}) = S_l$. The set of α and $\mathbf{z}$ are the free optimization variables, constraining $\alpha \geq 0$. The $\mathbf{z}$ are initialized to random selections from $\mathbf{x}_i$, the α to random variates in $[0, 1)$. There are m functions assigned to each of the A symbols of the alphabet. In [6] the $\mathbf{z}_k$ were fixed on the UPOs and their symbols varied; here, the $\mathbf{z}_k$ and α_k vary freely but the S_k remain fixed. We minimize J_{sfnn} or K_{sfnn} over the $mA(d + 1)$ free parameters using "differential evolution" [12], a genetic algorithm suitable for continuous free parameters.

Figure 1 shows the final $\mathcal{P}$ on 2000 data points from the Ikeda map [13]. It shows the best result (lowest J_{sfnn}) out of six restarts changing only the random seed governing the initial conditions (the results were not much worse on the other runs, however). Our result is very close to the partition knowing the dynamics.

In a stationary information source, the number of distinct length-p code words will scale, for asymptotic p, as $N_p \propto e^{h_T p}$ where h_T is the *topological entropy*, a dynamical invariant. We validate $\mathcal{P}$ with an estimate of the deficiency between h_T implied by $\mathcal{P}$ and the correct h_T: $\delta h_T = p_{\max}^{-1} \sum_{p=1}^{p_{\max}} p^{-1} \log(N_p / \tilde{N}_p)$. N_p is the number of distinct period-p UPOs (computed knowing the equations

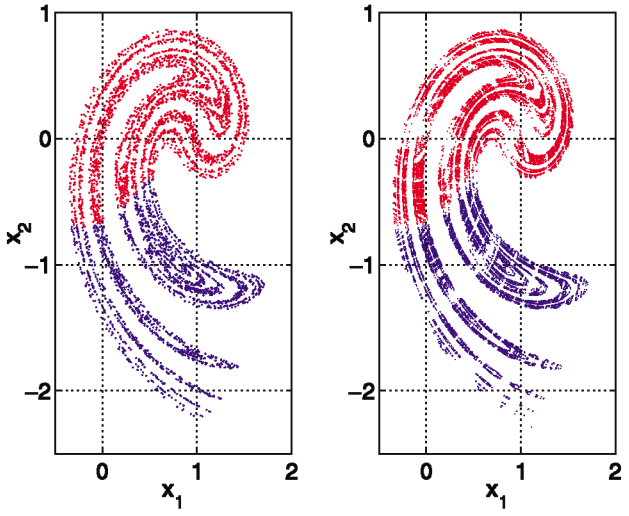

FIG. 1 (color online). Left: partition estimated by J_{sfnn} optimization on 2000 data points from the Ikeda map. Right: partition calculated with foreknowledge of UPOs, numerically extracted from the equation of motion. The partition we estimate from observed data alone is quite close to a presumably correct partition, calculated from the method of [6]. The measure on the two figures is not the same: the left-hand figure is a sample of the natural measure, whereas the right-hand figures shows UPOs up to period 16. They avoid regions of homoclinic tangencies, contributing to the blank spaces.

of motion), $\tilde{N}_p$ the number of such UPOs with unique p-symbol codes under some $\mathcal{P}$. A GP gives $\delta h_T = 0$, with $\delta h_T \to 0$ for better (less UPO-degenerate) partitions. Figure 2 shows δh_T on each new best partition found during the optimization. The optimization target, J_{sfnn}, decreases monotonically by construction; though δh_T does not decrease perfectly monotonically, the trend toward zero is evident. This gives evidence that minimizing J_{sfnn} also refines approximations to GPs.

Figures 3–5 demonstrate applications of the algorithm. Figure 3 shows the effect of noise on a system where the GP is analytically known. The Lozi map (see analysis in [9]) is similar to the Henon map but replaces the quadratic nonlinearity with a piecewise linear one. We find a partition which is close to the noise-free GP even when the data have been contaminated by significant amounts of additive noise. Figures 4 and 5 show estimated partitions

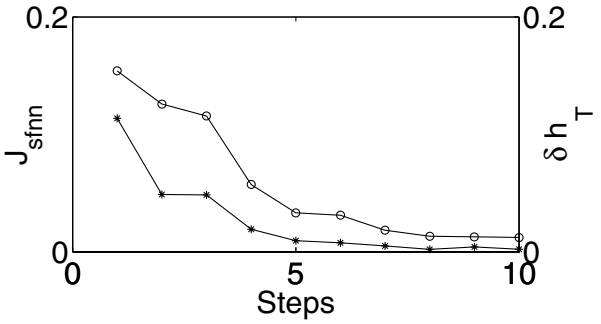

FIG. 2. For each new best partition: minimization target J_{sfnn} (circles, left-hand scale), estimated deficiency in topological entropy δh_T (asterisks, right-hand scale). Minimizing J_{sfnn} minimizes δh_T, maximizing topological entropy.

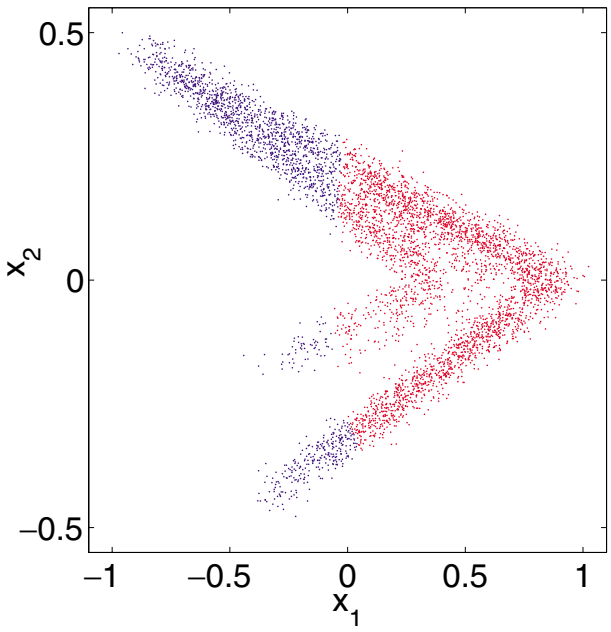

FIG. 3 (color online). Minimizing K_{sfnn} with $\gamma = 0.01$: estimated partition for a time series of 5000 data points from the Lozi map with 10% additive by amplitude Gaussian noise. Either the x_1 or x_2 axes are GPs for the noiseless map. Here despite the noise the algorithm finds a partition close to what would be a GP.

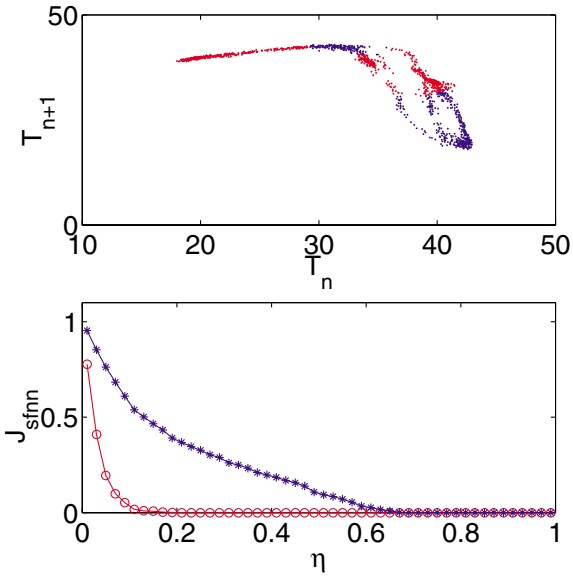

FIG. 4 (color online). Top: estimated $\mathcal{P}$ for interbubble time intervals [16] (arbitrary units), minimizing K_{sfnn}. Note that partitions may have the same symbol in multiple regions. Bottom: $J_{\mathrm{sfnn}}(\eta)$ vs η for the optimized partition (circles) and for a naive equiprobable histogram partition with the same alphabet (asterisks). For the optimized partition there are very few large distance errors, e.g., J_{sfnn} observed above $\eta = 0.1$.

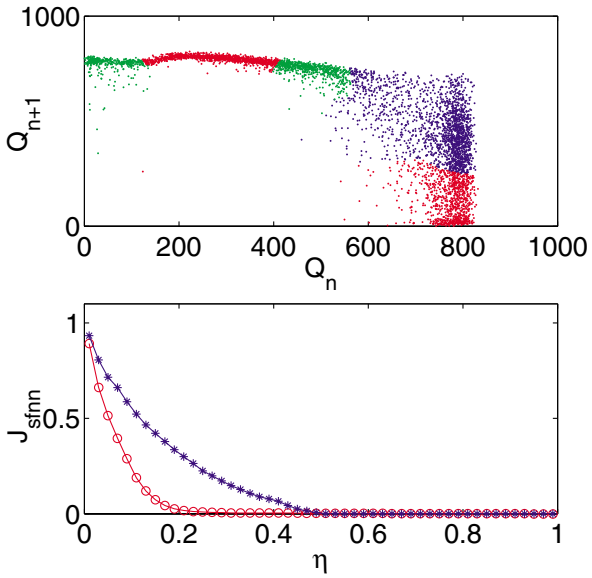

FIG. 5 (color online). Same as Fig. 4 but with combustion engine heat release [17] time series (energy, arbitrary units), and $A = 3$. The noise level is higher thus there remain more moderately sized distances, even with a larger alphabet which usually results in better localization.

on experimental data sets where no analytical form of the equations (much less partitions) are known. Because of noise, dynamical or observational, a certain level of divergence D for symbolic nearest neighbors is inevitable, but minimizing large deviations is still a useful goal. There are very few rank distances with $R \geq 0.2$ or 0.3 compared to a basic histogram partition.

It must be kept in mind that GPs are not necessarily unique. The optimization algorithm may find distinct partitions, all of which are reasonably satisfactory. (At a minimum any iterate of a GP is also a GP). There are coexisting solutions, which is why the optimization problem requires a global search. We conjecture this is one reason that understanding the geometrical diversity of generating partitions has been difficult (e.g., [14]). Well-localizing partitions for nonchaotic periodic or quasiperiodic data can also be estimated, but there may be many equally good solutions without additional constraints. In summary, we demonstrate a practical algorithm to estimate good symbolic partitions from time series of observed state vectors. It addresses the topological degeneracy problems demonstrated in [4]. It does not require computation of local derivatives to estimate homoclinic tangencies nor accurate enumeration of high-order UPOs, making it feasible for realistic data sets [15].

*Electronic address: mkennel@ucsd.edu
†Electronic address: mbuhl@ucsd.edu

[1] N. H. Packard *et al.*, Phys. Rev. Lett. **45**, 712 (1980); T. Sauer, J. A. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579–616 (1992).
[2] Among many others in the literature, P. Graben, J. D. Saddy, M. Schlesewsky, and J. Kurths, Phys. Rev. E **62**, 5518–5541 (2001); M. Lehrman and A. B. Rechester, Phys. Rev. Lett. **87**, 164501 (2001); Z. Hongxuan and Z. Yisheng, Med. Eng. Phys. **23**, 523–528 (2001); J. Godelle and C. Letellier, Phys. Rev. E **62**, 7973–7981 (2000); R. Steuer, W. Ebeling, D. F. Russell, S. Bahar, A. Neiman, and F. Moss, Phys. Rev. E **64**, 061911 (2001).An applied review is found in C. S. Daw, C. E. A. Finney, and E. R. Tracy, Rev. Sci. Instrum. **74**, 915–930 (2003).
[3] We assume that a state space of sufficient dimension has been reconstructed already, and dynamics on it are as a discrete map—for flows a Poincaré section is taken.
[4] E. M. Bollt, T. Stanford, Y-C. Lai, and K. Zyczowski, Phys. Rev. Lett. **85**, 3524 (2000); E. M. Bollt, T. Stanford, Y-C. Lai, and K. Zyczowski, Physica (Amsterdam) **154D**, 259–286 (2001).
[5] S. Hayes *et al.*, Phys. Rev. Lett. **73**, 1781 (1994); E. Bollt, Y-C. Lai, and C. Grebogi, Phys. Rev. Lett. **79**, 3787 (1997).Y-C. Lai, E. Bollt, and C. Grebogi, Phys. Lett. A **255**, 75–81 (1999).
[6] R. L. Davidchack, Y-C. Lai, E. M. Bollt, and M. Dhamala, Phys. Rev. E **61**, 1353–1356 (2000).
[7] J. Plumecoq and M. Lefranc, Physica (Amsterdam) **144D**, 231–258 (2000).
[8] J. P. Crutchfield and N. H. Packard, Int. J. Theor. Phys. **21**, 433 (1982); R. Steuer, L. Molgedey, W. Ebeling, and M. A. Jimenez-Montano, Eur. Phys. J. B **19**, 265–269 (2001).
[9] P. Cvitanovic, G. H. Gunaratne, and I. Procaccia, Phys. Rev. A **38**, 1503 (1988).
[10] J. H. Friedman, J. L. Bentley, and R. A. Finkel, ACM Trans. Math. Software **3**, 209–226 (1977); R. F. Sproull, Algorithmica **6**, 579–589 (1991).
[11] M. B. Kennel and H. D. I. Abarbanel, Phys. Rev. E **66**, 026209 (2002); M. B. Kennel, R. Brown, and H. D. I. Abarbanel, Phys. Rev. A, **45**, 3403–3411 (1992).
[12] R. Storn and K. Price, J. Global Optim. **11**, 341–359 (1997).
[13] Dynamical equations and parameters are as in [6].
[14] L. Jaeger and H. Kantz, J. Phys. A **30**, L567 (1997).
[15] See AIP Document No. E-PRLTAO-91-082332 for a complete FORTRAN 95 source code, implementation notes, and additional examples (e.g., 1D and quasiperiodic maps). A direct link to this document may be found in the online article's HTML reference section. The document may also be reached via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html) or from ftp.aip.org in the directory /epaps/. See the EPAPS homepage for more information.
[16] K. Nguyen, C. S. Daw, P. Chakka, M. Cheng, D. D. Bruns, C E. A. Finney, and M. B. Kennel, Chem. Eng. J. (Lausanne) **64**, 191–197 (1996).
[17] R. M. Wagner, J. A. Drallmeir, and C. S. Daw, Int. J. Eng. Res. **1**, 301–320 (2000).