Detecting metastable states of dynamical systems by recurrence-based symbolic dynamics

Peter Beim Graben, Axel Hutt

Detecting recurrence domains of dynamical systems
by symbolic dynamics

Peter beim Graben[1,2,3,*] and Axel Hutt[3]

[1]*Dept. of German Language and Linguistics,*
Humboldt-Universität zu Berlin, Germany
[2]*Bernstein Center for Computational Neuroscience Berlin, Germany*
[3]*Cortex Project, INRIA Nancy Grand Est, France*

(Dated: March 18, 2013)

Abstract

We propose an algorithm for the detection of recurrence domains of complex dynamical systems from time series. Our approach exploits the characteristic checkerboard texture of recurrence domains exhibited in recurrence plots (RP). In phase space, RPs yield intersecting balls around sampling points that could be merged into cells of a phase space partition. We construct this partition by a rewriting grammar applied to the symbolic dynamics of time indices. A maximum entropy principle defines the optimal size of intersecting balls. The final application to high-dimensional brain signals yields an optimal symbolic recurrence plot revealing functional components of the signal.

States of complex dynamical systems often dwell for relatively long time in a particular domain of their phase spaces before the trajectory moves into another region. This is the case for metastability [1] and several kinds of instability such as saddles that are connected by heteroclinic trajectories [2, 3] or, e.g., the "wings" of the Lorenz attractor that are centered around its unstable foci [4]. According to Poincaré's famous *recurrence theorem* [5], we could refer to such behavioral regimes as to *recurrence domains* of a dynamical system. The detection of recurrence domains has become increasingly important in recent time in several applications such as spin glasses [1], molecular configurations [6], in the geosciences [7] and in the neurosciences [2, 8–10].

For the identification of recurrence domains from time series, their characteristic slow time scales have been separated from the fast dynamics of phase space trajectories by several clustering algorithms [1, 2, 6, 8, 11, 12]. One method, sometimes called *Perron clustering* [6], starts with an *ad hoc* partitioning of the system's phase space that leads to an approximate Markov chain description [1, 6, 8, 11, 12]. Applying spectral clustering methods to the resulting transition matrix yields the time scales of the process, while their corresponding (left-)eigenvectors allow the unification of cells into a partition of metastable states [8, 12]. Another approach by Hutt and Riedel [2] utilizes the slowing-down of the system's trajectory within saddle sets by means of phase space clustering.

Several of such methods are numerically rather time-consuming. For instance, Markov chain modeling requires an *ad hoc* partitioning of the complete system's phase space into equally populated cells, from which transition probabilities must be estimated by counting measures. Subsequent spectral clustering methods perform various matrix multiplications and clustering techniques. All these algorithms are numerically rather expensive as illustrated in [8].

In this Letter, we propose a parsimonious algorithm for detecting recurrence domains from measured or simulated time series. Our starting point is Eckmann et al.'s [13] recurrence plot (RP) method for visualizing Poincaré's recurrences. The proposed method is numerically less time-consuming and advantageous especially for high-dimensional data since it simply exploits the recurrence structure of the system's dynamics.

When $x_j \in \mathbb{R}^d$ is the system's state at (discretized) time j in phase space $\mathbb{R}^d$ of dimension d, the element

$$R_{ij} = \Theta(\varepsilon - ||x_j - x_i||) \tag{1}$$

of the *recurrence matrix* $\boldsymbol{R} = (R_{ij})$ is one if x_j is contained in a "ball" $B_\varepsilon(x_i)$ of radius $\varepsilon > 0$ centered at state $x_i \in \mathbb{R}^d$ and zero otherwise [13, 14], as mediated by the Heaviside step function Θ. Eckmann et al. [13] have already pointed out that RPs display recurrence domains as a characteristic "checkerboard texture". We illustrate this in Fig. 1 with the paradigmatic Lorenz attractor [4].
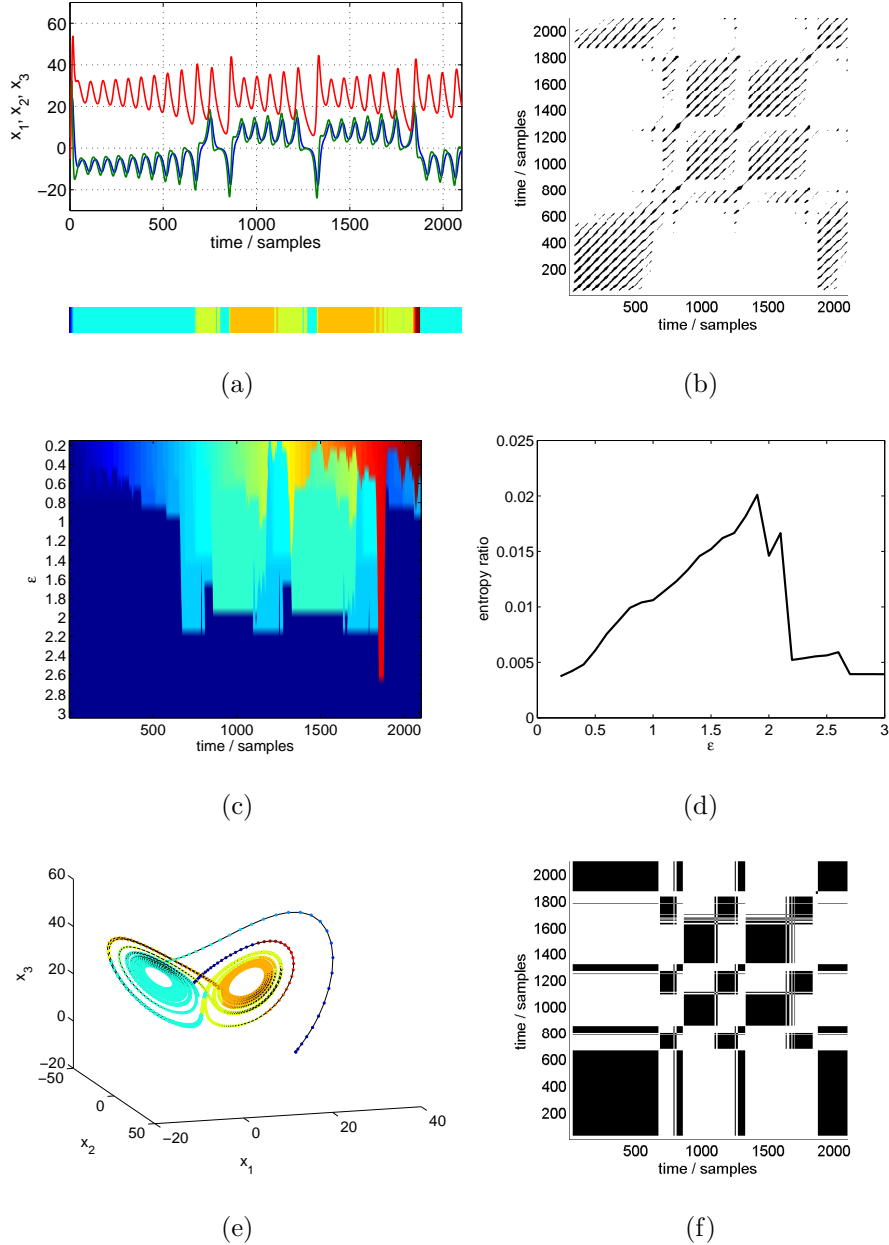
Figure 1: (Color online) Recurrence-based symbolic dynamics of the Lorenz attractor [4]. (a) Time series $\boldsymbol{x}_t$ (upper panel) and optimal encoding s' (color bar beneath). (b) ε-Recurrence plot [Eq. (1)] for $\varepsilon = 5.0$ and Euclidian norm; black pixels denote $R_{ij} = 1$, white ones $R_{ij} = 0$. (c) Symbolic dynamics s' for range $\varepsilon \in [0.2, 3]$. (d) Dependence of entropy ratio [Eq. (5)] from ε. (e) Phase space partition into recurrence domains for optimal encoding $\varepsilon^* = 1.9$. (f) Symbolic recurrence plot [Eq. (2)] of optimal encoding.

The upper panel of Fig. 1(a) displays the x_1, x_2 (blue and green) and x_3 (red) time series of the Lorenz attractor starting with initial condition $\boldsymbol{x} = [20, 5, -5]^T$, with integration

4

interval $[0, 20]$ and sampling $\Delta t = 0.0095$. The two wings of the attractor [shown in Fig. 1(e)] clearly correspond to positive, respectively, negative x_1, x_2. The recurrence plot in Fig. 1(b) exhibits the typical texture of diagonal line patterns that are characteristic for oscillatory dynamics. These oscillators correspond to the attractor's wings. Going along the line of identity (LOI) reveals transient transitions between about four recurrence domains, while the checkerboard texture of these diagonal line patters indicates that there are indeed only two recurrence domains being involved, namely the wings, that are repeatedly explored by the system's trajectory.

For uniform ε, the recurrence matrices obtained from Eq. (1) are reflexive, $R_{ii} = 1$ (the LOI), and symmetric, $R_{ij} = R_{ji}$, but in general not transitive, i.e. $R_{ij} = 1$ and $R_{jk} = 1$ do not necessarily imply $R_{ik} = 1$. In order to cope with this disadvantage, Donner et al. and later Faure and Lesne [15, 16] suggested to compute the recurrence matrix from words in a symbolic dynamics [17] through

$$R_{ij}^+ = \delta_{w_i w_j} \, , \tag{2}$$

where w_i, w_j are words of length m at times i and j in a symbolic sequence $s = a_1 a_2 \ldots a_n$. Here, $\delta_{ab} = 1$ if $a = b$ and zero otherwise, denotes the Kronecker matrix. Symbolic RPs given by Eq. (2) are also transitive, because symbolic dynamics results from a partition of the system's phase space into equivalence classes from an equivalence relation.

In contrast to [15, 16] we here construct a phase space partition and thereby its resulting symbolic dynamics from the ε-RP (1). For that aim we first observe that $R_{ij} = 1$ if two ε-balls $B_\varepsilon(x_i)$ and $B_\varepsilon(x_j)$ intersect: $B_\varepsilon(x_i) \cap B_\varepsilon(x_j) \neq \emptyset$. We could therefore start with an initial partition of the phase space into a family of ε-balls around the sampling points x_i and its set complement and then merge all intersecting balls together. The result is a partition of phase space into disjoint sets.

In order to achieve this construction we consider the ε-RP $\boldsymbol{R}$ [Eq. (1)] as a *grammatical rewriting system* over the time indices of a given trajectory x_t [18]. Thus, we first map a trajectory x_t to the sequence of successive time indices, regarded as symbols: $x_t \rightarrow s_t = t$. Then we define a formal grammar of rewriting rules: if $i > j$ and $R_{ij} = 1$ create a rule $i \rightarrow j$. To enforce transitivity for $i > j > k$, $R_{ij} = 1$ and $R_{ik} = 1$, we first eliminate the redundancy by rewriting only $i \rightarrow k$ and then create an additional rule $j \rightarrow k$. Finally, we apply this grammar to the initial sequence of time indices $s_t = t$ in order to replace large indices by smaller ones, thus exploiting the recurrence structure of the data. The result is

a transformed symbolic sequence s'_t, whose symbolic RP $\boldsymbol{R}^+$ [Eq. (2)] [15, 16] becomes also transitive.

Let us illustrate the procedure by means of a simple example. Assume, we have a series of only five data points $(x_1, x_2, \ldots, x_5)$ that gave rise to the recurrence matrix

$$\boldsymbol{R} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

The algorithm starts in the 5th row, detecting a recurrence $R_{52} = 1$. Since $5 > 2$, we create a rewriting rule $5 \rightarrow 2$. Because the next recurrence in row 5 is trivial, the algorithm continues with row 4, where $R_{41} = R_{42} = 1$. Now, two rules $4 \rightarrow 1$ and $4 \rightarrow 2$ could be generated. However, the latter is redundant. Therefore the algorithm only records the rule $4 \rightarrow 1$. Moreover, transitivity is taken into account by an additional rule $2 \rightarrow 1$. Next, row 3 does not contribute to the algorithm and rows 2 and 1 can be neglected due to the symmetry. Recursively applying this grammar to the symbolically encoded time series $s = \ll 12345 \gg$ yields $s' = \ll 11311 \gg$, i.e. a system with two recurrence domains $\ll 1 \gg$ and $\ll 3 \gg$.

In order to validate our construction, we employ the method to the Lorenz attractor as shown in Fig. 1(c). Here, each row is the symbolically encoded time series s' from Fig. 1(a) using a color code. For small values of ε (top rows) there are almost no intersecting ε-balls such that each ball is represented by a separate color from the light spectrum. Increasing ε towards the bottom rows yields more and more intersections, eventually leading to one big cluster of merged ε-balls for $\varepsilon > 2.6$. For intermediate values of ε essentially two recurrence domains emerge that are connected by transients.

Interestingly, Fig. 1(c) also reveals that our recurrence-based symbolic dynamics is rather robust against variations of the ball size ε which is reflected by the vertical band structure of the symbolic sequences.

Guided by the principle of maximal entropy, we assume that the system spends equal portions of time in its recurrence domains and derive a utility function of the symbolic encoding from the entropy of the symbol distribution

$$H(\varepsilon) = -\sum_{k}^{M(\varepsilon)} p_k \log p_k \,, \tag{4}$$

6

where p_k is the relative frequency of symbol k and $M(\varepsilon)$ the cardinality of the symbolic repertoire obtained for ball size ε. The entropy ratio

$$h(\varepsilon) = \frac{H(\varepsilon)}{M(\varepsilon)} \tag{5}$$

is then a good estimator for a given encoding because small values of ε lead to an almost uniform distribution of rare symbols that is punished by the large alphabet. By contrast, large values of ε give rise to a trivial partition with small entropy. Thus, the quantity $h(\varepsilon)$ will assume a global maximum for an optimal value

$$\varepsilon^* = \arg \max_{\varepsilon} h(\varepsilon) \tag{6}$$

reflecting a uniform distribution of a small number of recurrence domains.

We plot the dependence of $h(\varepsilon)$ for the Lorenz system in Fig. 1(d) and choose the optimal ball size $\varepsilon^* = 1.9$ for the symbolic dynamics in the color bar of Fig. 1(a). One can easily recognize that one wing is uniquely represented by the turquois symbol, while the other one is represented by orange and light green symbols. The distribution of these symbols in phase space is shown in Fig. 1(e) using the same color palette for the samples $\boldsymbol{x}_t$. Here, one wing is completely captured by the union of turquois ε-balls, while the other one needs two partitions cells, indicated in orange and light green which is due to a gap in the second wing in our numerics.

Finally, Fig. 1(f) depicts the symbolic RP Eq. (2) where the characteristic checkerboard texture of the Lorenz attractor's recurrence domains is significantly enhanced.

In order to also present a proof-of-concept for our method applied to real-world data, we reanalyze event-related electroencephalographic (EEG) data from a language processing experiment [19] in Fig. 2 since Hutt and Riedel [2] have argued that components in the event-related brain potential (ERP) can be regarded as saddle sets and therefore as recurrence domains in the EEG.
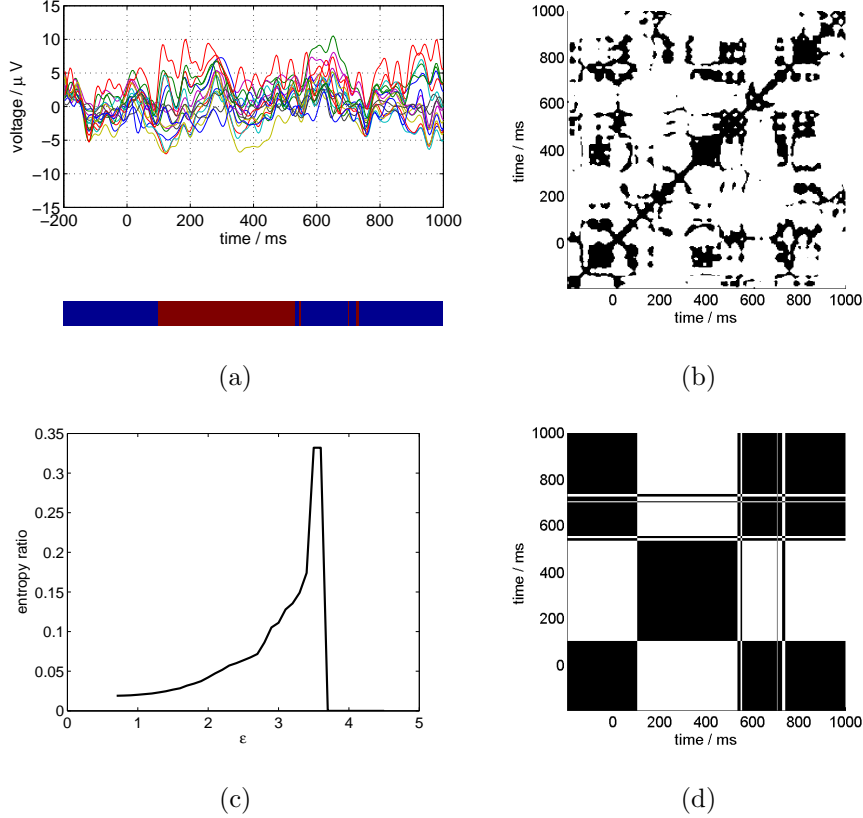
Figure 2: (Color online) Recurrence-based symbolic dynamics of a 17-dimensional ERP data set of a single subject [19]. (a) Single subject ERP time series (upper panel) and symbolic encoding s' with $\varepsilon^* = 3.5\,\mu$V (color bar beneath) for $d = 17$ scalp channels. (b) ε-Recurrence plot [Eq. (1)] for $\varepsilon = 8.0\,\mu$V. (c) Utility function $h(\varepsilon)$ [Eq. (5)]. (d) Symbolic recurrence plot [Eq. (2)] for optimal encoding $\varepsilon^* = 3.5\,\mu$V.

Figure 2(a) displays the averaged ERP time series of a single subject encountering a linguistic processing problem in German [19]. In Fig. 2(b) we present the conventional RP [Eq. (1)] for $\varepsilon = 8.0\,\mu$V. Figure 2(c) shows the utility function $h(\varepsilon)$ from Eq. (5). The optimal encoding s' is obtained for $\varepsilon^* = 3.5\,\mu$V which is depicted as the color bar in Fig. 2(a). The symbolic dynamics exhibits three interesting properties: *i*) the pre-stimulus interval is represented by one distinguished recurrence domain; *ii*) the time interval for lexical access around 400 ms post stimulus is represented by another recurrence domain; *iii*) the time window of syntactic reanalysis around 600 ms is represented by the first recurrence domain again. This results from the optimization constraint to obtain uniformly distributed recurrence domains. Also the symbolic RP in 2(d) nicely reveals the existence of two substantial

8

recurrence domains.

In this Letter we proposed a parsimonious algorithm for the detection of recurrence domains in complex dynamical systems. In contrast to techniques based on Markov chains, which require an *ad hoc* partitioning of the system's phase space and the estimation of transition probabilities, our approach exploits the recurrence structure of the system's dynamics thereby partitioning the phase space into unions of intersecting ε-balls along the actual trajectory.

The proposed method could have a number of interesting applications in many different fields, such as molecular dynamics [6], geo- [7], and neurosciences [3, 8–10] for the identification of recurrence domains. Moreover, it could also be useful for the analysis of complex networks for solving graph partition and related problems by taking the transitive closure of the graph's adjacency matrix. Finally we concede that further research is required to obtain appropriate utility functions for real-world problems that violate the uniformity assumption for recurrence domains in order to detect e.g. saddle sets in heteroclinic dynamics [3, 10] or to identify functional ERP components [2].

––––––––––

* Electronic address: `peter.beim.graben@hu-berlin.de`

[1] H. Larralde and F. Leyvraz, Physical Review Letters **94**, 160201 (2005).

[2] A. Hutt and H. Riedel, Physica D **177**, 203 (2003).

[3] M. I. Rabinovich, R. Huerta, P. Varona, and V. S. Afraimovich, PLoS Computational Biology **4**, e1000072 (2008).

[4] E. N. Lorenz, Journal of the Atmospheric Sciences **20**, 130 (1963).

[5] H. Poincaré, Acta Mathematica **13**, 1 (1890).

[6] P. Deuflhard and M. Weber, Linear Algebra and its Applications **398**, 161 (2005).

[7] G. Froyland, K. Padberg, M. H. England, and A. M. Treguier, Physical Review Letters **98**, 224503 (2007).

[8] C. Allefeld, H. Atmanspacher, and J. Wackermann, Chaos **19**, 015102 (2009).

[9] K. J. Friston, NeuroImage **5**, 164 (1997).

[10] M. I. Rabinovich, R. Huerta, and G. Laurent, Science **321**, 48 (2008).

[11] G. Froyland, Physica D **200**, 205 (2005).

[12] B. Gaveau and L. S. Schulman, Physical Reviews E **73**, 036124 (2006).

[13] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, Europhysics Letters **4**, 973 (1987).

[14] N. Marwan and J. Kurths, Physics Letters **A 336**, 349 (2005).

[15] R. Donner, U. Hinrichs, and B. Scholz-Reiter, The European Physical Journal Special Topics **164**, 85 (2008).

[16] P. Faure and A. Lesne, International Journal of Bifurcation and Chaos **20**, 1731 (2010).

[17] B.-L. Hao, *Elementary Symbolic Dynamics and Chaos in Dissipative Systems* (World Scientific, Singapore, 1989).

[18] J. E. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages, and Computation* (Addison–Wesley, Menlo Park, California, 1979).

[19] P. beim Graben, D. Saddy, M. Schlesewsky, and J. Kurths, Physical Reviews E **62**, 5518 (2000).