Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica D 200 (2005) 205–219
PHYSICA D
www.elsevier.com/locate/physd

Statistically optimal almost-invariant sets

Gary Froyland*

School of Mathematics, The University of New South Wales, Sydney, NSW 2052, Australia

Received 6 June 2003; accepted 4 November 2004

Communicated by E. Kostelich

Abstract

Chaotic dynamical systems are often transitive, although this transitivity is sometimes very weak. It is of interest to divide the phase space into large regions, between which there is relatively little communication of trajectories. We present fast, simple algorithms to find such divisions. The present work builds on the results of Froyland and Dellnitz [G. Froyland, M. Dellnitz, Detecting and locating near-optimal almost-invariant sets and cycles, SIAM J. Sci. Comput. 24 (6) (2003) 1839–1863], focussing on a statistical description of transitivity that takes into account the fact that trajectories tend to visit different regions of phase space with different frequencies. The new work takes advantage of theoretical results from the theory of reversible Markov chains. A new adaptive algorithm is put forward to efficiently deal with situations where the boundaries of the weakly communicating regions are complicated. This algorithm is illustrated with the standard map. Relevant convergence results are proven.

PACS: 02.50.Ga; 02.70.Hm; 05.45.Ac; 89.75.Fb

Keywords: Almost-invariant; Meta-stable; Macrostructure; Transitivity

1. Introduction

Let $T : X \to X$ be continuous, with $X \subset \mathbb{R}^d$ compact. The mapping T defines a dynamical system on X. We assume that the domain X contains an attracting[1] set $M = \bigcap_{i=0}^{\infty} T^i(X)$, where $M \subseteq X$ and equality is allowed. If T defines a "chaotic" dynamical system, then T is *transitive* on M, i.e., trajectories may move from any given relatively open set in M into any other in a finite amount of time. In some cases, this transitivity can be very weak, in the sense that even though it is *possible* for trajectories to move from any part of phase space to any other, certain transitions

* Tel.: +61 2 9385 7050; fax: +61 2 9385 7123.
E-mail address: froyland@maths.unsw.edu.au (G. Froyland).

[1] It is also possible to treat saddles and unstable sets in a completely analogous way, but for simplicity we restrict the exposition to attracting sets.

0167-2780/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.physd.2004.11.008

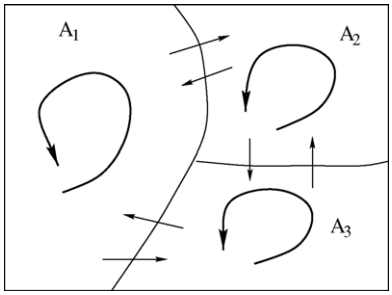

Fig. 1. Schematic representation of three almost-invariant sets.

between different regions are very *unlikely*. In this weakly transitive situation, there are *regions of phase space in which trajectories tend to stay for a very long time before entering another region.* We will call these regions *almost-invariant sets*; our goal is to detect their existence and to locate their positions in phase space.

Quantifying almost-invariance of individual sets. Given a particular dynamical system, we would like to find a collection of "optimal" almost-invariant sets. In order to set up and solve a suitable optimisation problem we require a quantification of the notion of almost-invariance. In the sequel, we will assume that our transitive dynamical system has a unique probability measure μ satisfying

$$\mu(A) = \lim_{N\to\infty} \frac{\#\{T^i x \in A : 0 \le i \le N-1\}}{N} \tag{1}$$

for Lebesgue almost all $x \in X$. The probability measure μ describes the distribution of Lebesgue almost-all long trajectories in phase space; it is commonly known as the *physical-invariant measure* or *natural-invariant measure* for T. The measure $\mu(A)$ of a set A is high if trajectories visit A frequently; if trajectories never visit A, then $\mu(A) = 0$. For a set $A \subset M$, the ratio

$$\rho_\mu(A) = \mathrm{Prob}_\mu\{Tx \in A | x \in A\} = \frac{\mu(A \cap T^{-1}A)}{\mu(A)}$$

is the probability that a point in A will remain in A after one iteration of T. In Fig. 1 we show a schematic representation of three almost-invariant sets A_1, A_2, A_3. The thick arrows represent high probability transitions and the thin arrows represent transitions with a low probability of occurring. The values $\rho_\mu(A_1)$, $\rho_\mu(A_2)$, $\rho_\mu(A_3)$ will all be close to unity.

We introduce the following criteria and problem definition.

Criteria 1.

(A) The candidate almost-invariant sets $A_1, \ldots, A_q$ have roughly the same measure; to achieve this, we impose the condition that $\mu(A_k) > s$ for $k = 1, \ldots, q$, where $0 < s \le 1/q$ is given.
(B) The ratios

$$\rho_\mu(A_k) = \frac{\mu(A_k \cap T^{-1}A_k)}{\mu(A_k)}$$

are large for each $k = 1, \ldots, q$.

Problem 1. Find a measurable partition $\{\hat{A}_1, \ldots, \hat{A}_q\}$ of M such that

$$\rho_\mu(\hat{A}_1, \ldots, \hat{A}_q) = \max \left\{ \frac{1}{q} \sum_{k=1}^{q} \rho_\mu(A_k) : \{A_1, \ldots, A_q\} \text{ measurably partitions } M \text{ and satisfies Criterion 1A} \right\} \tag{2}$$

A solution $\{\hat{A}_1, \ldots, \hat{A}_q\}$ of this problem will satisfy both Criteria 1A and B, and we say that it describes an *optimal almost-invariant partition of M into q sets*.

Remark 1. Clearly, satisfying Criterion 1B alone is not particularly useful; it may be true that once an orbit enters a set A_k, it has a high probability of remaining in A_k, but if $\mu(A_k)$ is very small, then there is a very slim chance of the orbit entering A_k in the first place. By imposing the additional condition that $\mu(A_k) > s$, we ensure that every set in the partition is physically meaningful.

The separate problem of determining a suitable *number q* of sets to include in the partition is treated in [7,12].

Early ideas on the notion of almost-invariance are contained in [7]. The current work builds on the results of [12] which focussed on the problem of satisfying Criteria 1A and B with the physical measure μ replaced by Lebesgue measure m. The methods presented here are similar to those in [12], but have been modified to provide better performance for this statistical description of almost-invariance. These new methods provide the ability to obtain rigorous bounds on how large the value $\rho_\mu(A_k)$ may be, and satisfy relevant optimality criteria. In addition, we introduce a new adaptive approach that efficiently determines the boundaries between the almost-invariant sets.

Related work includes [16,9,13] which consider meta-stable sets in the context of time reversible conservative dynamics of molecular systems. A recent preprint [8] applies ideas from graph theory (notably "congestion") to identify almost-invariant sets in dynamical systems of the generality considered here.

2. First steps towards a solution: discretisation of phase space

As Problem 1 stands, we are maximising over an uncountably infinite domain; there are no restrictions on the shape or size of the sets $A_1, \ldots, A_q$, apart from the conditions that each A_k is measurable and does not differ in μ-measure too greatly from any of the other sets. We will make the problem tractable by reducing the class of sets in which we search for a maximum. The first step in a practical solution of Problem 1 is to discretise the phase space by covering M with many small boxes; $M \subset \bigcup_{j=1}^{n} B_j$. This covering process can be accomplished efficiently through the use of multilevel methods; see [7].

Optimising over a finite collection. For a box collection $\{B_1, \ldots, B_n\}$ with $M \subset \bigcup_{j=1}^{n} B_j$, define

$$\mathcal{C}_n = \left\{ A \subset X : A = \bigcup_{j \in \mathcal{I}} B_j, \mathcal{I} \subset \{1, \ldots, n\} \right\}.$$

We will restrict our choices of almost-invariant sets to elements of $\mathcal{C}_n$.

Definition 1. We will call a collection of coverings $\{S_n\}$, where $S_n = \bigcup_{i=1}^{n} B_i$, of M *tight* if $\mu(S_n \Delta M) \to 0$ as $n \to \infty$. For brevity, we often call individual coverings tight if they are members of a tight collection.

Problem 2. Find a collection $\{\hat{A}_1, \ldots, \hat{A}_q\}$ with $\hat{A}_k \in \mathcal{C}_n$, $1 \le k \le q$, such that

$$\rho_\mu(\hat{A}_1, \ldots, \hat{A}_q) = \max_{A_1, \ldots, A_q \in \mathcal{C}_n} \left\{ \frac{1}{q} \sum_{k=1}^{q} \rho_\mu(A_k) : \{A_1, \ldots, A_q\}, \text{ partitions a tight covering of } M \text{ and satisfies Criterion 1A} \right\}$$

In practice, we will usually not know the physical measure μ a priori, and will have to numerically estimate it. This estimate μ_n will be obtained in a natural way from our box covering of M; the details will be explained in Section 3. For the moment, we state a convergence result that says that we lose nothing by restricting ourselves to sets comprised of unions of boxes in the limit of the box diameters going to zero.

An approximation result. Setting

$$\rho_\mu^{\max} = \sup\left\{\sum_{k=1}^{q}\rho_\mu(A_k) : \{A_1,\ldots,A_q\} \text{ is a measurable partition of } M \text{ that satisfies Criterion 1A}\right\}, \tag{3}$$

we have the following approximation result.

Theorem 1 ([12]). *Let $\mathcal{C}_n$, $n \geq 1$, denote a sequence of partitions of a tight collection $\{S_n\}$ of coverings of X, with $\max_{B\in\mathcal{C}_n}\operatorname{diam}(B) \to 0$ as $n \to \infty$. If $\mu_n \to \mu$ strongly, then*

$$\max_{A_1^n,\ldots,A_q^n\in\mathcal{C}_n}\{\rho_{\mu_n}(A_1^n,\ldots,A_q^n) : \{A_1^n,\ldots,A_q^n\} \text{ is a partition of } S_n \text{ and } \mu_n(A_k) > s \text{ for } k=1,\ldots,q\} \to \rho_\mu^{\max}. \tag{4}$$

In the event that the condition $\mu_n(A_k^n) > s$ for $k = 1,\ldots,q$ cannot be met, $\rho_{\mu_n}(A_1^n,\ldots,A_q^n)$ is understood to be zero.

Restricting the class of sets over which we search for optimal almost-invariant partitions does not immediately help a great deal, because it is computationally infeasible to consider all possible combinations of sets in $\mathcal{C}_n$. If there are n boxes covering M, there are q^n different ways we can form q sets as unions of these boxes, and therefore q^n candidate partitions to be evaluated in (3). We therefore turn to a global approach, considering the dynamics of T on the box collection to be a very large, but finite, Markov chain.

More detail on the material covered in this section may be found in [7,12].

3. Markov chains and almost-invariant sets

We create a large n-state Markov chain where each box $B \in \{B_1,\ldots,B_n\}$ corresponds to a state in the Markov chain. Define the $n \times n$ transition matrix P as

$$P_{ij} = \frac{m(B_i \cap T^{-1}B_j)}{m(B_i)},$$

where m denotes normalised Lebesgue measure on $\bigcup_{j=1}^n B_j$. The (i,j)th entry of P is the probability that a randomly selected point $x \in B_i$ has its image in B_j. In this sense, the dynamics of the Markov chain may be thought of as the dynamics of T, plus a small amount of dynamic noise.

Easy approximation of μ. The left eigenvector p of P provides an estimate μ_n of the physical-invariant measure by setting $\mu_n(B_i) = p_i$ for all $i = 1,\ldots,n$. The measure μ_n is extended to all Borel measurable sets of M via

$$\mu_n(A) = \sum_{i=1}^{n}\frac{m(A\cap B_i)}{m(B_i)}p_i. \tag{5}$$

As the box diameters go to zero ($n \to \infty$), it may be proven in some situations that $\mu_n \to \mu$ strongly; see [7] for maps perturbed by small amounts of random noise, and [10,11] for purely deterministic situations.

Simple computation of ρ_{μ_n}. It is straightforward to compute $\rho_{\mu_n}(A)$ if $A \in \mathcal{C}_n$. Since $A \in \mathcal{C}_n$, it is of the form $A = \bigcup_{i\in\mathcal{I}} B_i$, for some set of box indices $\mathcal{I} \subset \{1,\ldots,n\}$, and it is relatively easy to see the following lemma.

Lemma 1 ([12, Proposition 6.4]).

$$\rho_{\mu_n}(A) = \frac{\sum_{i,j\in\mathcal{I}} p_i P_{ij}}{\sum_{i\in\mathcal{I}} p_i}. \tag{6}$$

3.1. Time reversal and almost-invariance

The transition matrix P induces a related transition matrix R via a time reversal transformation. In this section we consider the properties of R and their relation to almost-invariant sets.

Invariance of $\rho_{\mu_n}(A)$ under time reversal. The Markov chain defined by the transition matrix P is not reversible[2] in general. This is because

$$p_i P_{ij} \neq p_j P_{ji},$$

or in terms of probabilities of set transitions,

$$\mu_n(B_i \cap T^{-1}B_j) \neq \mu_n(B_j \cap T^{-1}B_i).$$

The interpretation of the latter expression is that the probability of a point $x \in M$ being in box B_i at time t and in box B_j at time $t+1$ is not equal to the probability of a point $x \in M$ being in box B_j at time t and in box B_i at time $t+1$. If it were, the system would be reversible and would behave identically under the reversal of time.

An important observation is that Criteria 1A and 1B are invariant under time reversal.[3] Indeed, the approximate invariance ratios ρ_{μ_n} are also invariant under time reversal as we now demonstrate. Given a partition $\{\mathcal{I}_1, \ldots, \mathcal{I}_q\}$ of the box indices $\{1, \ldots, n\}$ (i.e., $\mathcal{I}_k \subset \{1, \ldots, n\}$, $\bigcup_{k=1}^{q} \mathcal{I}_k = \{1, \ldots, n\}$ and $\mathcal{I}_k \cap \mathcal{I}_\ell = \emptyset$, $k \neq \ell$), the invariance ratio

$$\rho_{\mu_n}(A_k) = \frac{\sum_{i,j\in\mathcal{I}_k} p_i P_{ij}}{\sum_{i\in\mathcal{I}_k} p_i},$$

remains unchanged under time-reversal. That is, if we define

$$\hat{\rho}_{\mu_n}(A_k) = \frac{\sum_{i,j\in\mathcal{I}_k} p_i \hat{P}_{ij}}{\sum_{i\in\mathcal{I}_k} p_i}$$

then it is clear that $\hat{\rho}_{\mu_n}(A_k) = \rho_{\mu_n}(A_k)$. Since the almost-invariance ratios of the individual sets $A_1, \ldots, A_q$ remain unchanged under time-reversal, so does the total cost function:

$$\rho_{\mu_n}(A_1, \ldots, A_q) = \frac{1}{q}\sum_{k=1}^{q} \rho_{\mu_n}(A_k).$$

Because our cost function is unchanged, without loss, we may replace P with $R = (P + \hat{P})/2$ to make our Markov chain reversible. In the remaining sections, R will denote this reversibilisation.

[2] Markov chains are called *reversible* if $P = \hat{P}$, where $\hat{P}_{ij} = p_j P_{ji}/p_i$ is the transition matrix governing the time-reversed system.

[3] It is trivial that Criterion 1A is invariant under time reversal; Criterion 1B is invariant under time reversal for invertible T, as one has $\mu(A \cap TA) = \mu(T^{-1}A \cap A)$ by T-invariance of μ.

Optimality properties of eigenvectors of reversible Markov chains. The right eigenvectors of reversible transition matrices satisfy minimality properties that are closely connected with the Criteria 1A and 1B. In particular, the right eigenvector $\hat{v}^{(2)}$ of R corresponding to the second largest eigenvalue λ_2 satisfies the following minimisation property

$$\min_{v \text{ is nonconstant}} \left\{ \frac{\sum_{i,j} p_i R_{ij}(v_i - v_j)^2}{\sum_i p_i v_i^2 - (\sum_i p_i v_i)^2} \right\} \tag{7}$$

is realised at precisely $v = \hat{v}^{(2)}$, with the value $2(1-\lambda_2)$. This result is known as Rayleigh's theorem (see, e.g. [5, p. 205]).

How is this related to almost-invariance? Consider dividing the n states of the Markov chain governed by R into two disjoint subsets $\mathcal{I}_1$ and $\mathcal{I}_2$, and imagine that in addition to the constraint that "v is nonconstant", we also insist that v takes only the values ± 1, depending on which subset state i is in. If $i \in \mathcal{I}_1$, we set $v_i = 1$, while if $i \in \mathcal{I}_2$, we set $v_i = -1$. Substituting these values into (7), we obtain the minimisation problem

$$\min_{\substack{\mathcal{I}_1 \cup \mathcal{I}_2 = \{1,\ldots,n\} \\ \mathcal{I}_1 \cap \mathcal{I}_2 = \emptyset}} \frac{4(1 - \sum_{i,j\in\mathcal{I}_1} p_i R_{ij} - \sum_{i,j\in\mathcal{I}_2} p_i R_{ij})}{1 - (\sum_{i\in\mathcal{I}_1} p_i - \sum_{i\in\mathcal{I}_2} p_i)^2}. \tag{8}$$

The numerator will be small if most transitions between i and j occur when *both* i and j are in either $\mathcal{I}_1$ or $\mathcal{I}_2$. The denominator will be large (and therefore the quotient smaller) when the combined weights of states in $\mathcal{I}_1$ and $\mathcal{I}_2$ are approximately equal. Thus a selection of $\mathcal{I}_1$ and $\mathcal{I}_2$ to minimise (8) will produce a good bisection on two counts: first, *the total weight of both collections of states is approximately equal*, and second, *most transitions are within either $\mathcal{I}_1$ or $\mathcal{I}_2$*.

The expression (8) can be rewritten in terms of boxes and measures as

$$\min_{\substack{A_1, A_2 \in \mathcal{C}_n \\ A_1 \cup A_2 = X, A_1 \cap A_2 = \emptyset}} \frac{4(1 - \mu_n(A_1 \cap T^{-1}A_1) - \mu_n(A_2 \cap T^{-1}A_2))}{1 - (\mu_n(A_1) - \mu_n(A_2))^2}. \tag{9}$$

Thus (9) has small values when

(i) the measures of A_1 and A_2 are similar and
(ii) the bulk of the measure in A_1 (resp. A_2) remains in A_1 (resp. A_2) under one iteration,

These two conditions correspond exactly to Criteria 1A and 1B.

In view of these remarks, we intend to relax the $v_i = \pm 1$ condition implicit in (8), and use the eigenvector v that minimises (7) to provide a good partition of the phase space. How we use v is explained in the next section. The eigenvector v contains global information on the dynamics of the Markov chain and the fact that eigenvectors are relatively cheap to numerically compute makes this spectral approach to finding almost-invariant partitions very attractive.

Theoretical bounds for $\rho_{\mu_n}(A)$. An additional advantage of using the reversible transition matrix R for our computations is that there exist theoretical bounds[4] for $\max_{A\in\mathcal{C}_n} \rho_{\mu_n}(A)$ when $\mu_n(A) \le 1/2$ in terms of the second eigenvalue λ_2:

$$1 - \sqrt{2(1-\lambda_2)} \le \max_{A\in\mathcal{C}_n} \rho_{\mu_n}(A) \le \frac{1+\lambda_2}{2}. \tag{10}$$

[4] These results are easily obtained from [3, Theorem 11.3; 5, Theorem 6.4.3].

Since $\rho_{\mu_n}(A_1, \ldots, A_q)$ is a convex combination of the individual ratios $\rho_{\mu_n}(A_k)$, $k = 1, \ldots, q$, we immediately obtain the upper bound:

$$\rho_{\mu_n}(A_1, \ldots, A_n) \leq \frac{1 + \lambda_2}{2}. \tag{11}$$

This bound gives one a guide as to the best possible invariance ratio, when searching for a maximum of $\rho_{\mu_n}(A_1, \ldots, A_q)$.

4. Using eigenvectors of R to produce almost-invariant partitions

Two almost-invariant sets. Let us be more specific about how we intend to use the second eigenvector $\hat{v}^{(2)}$ to divide the phase space M into two almost-invariant sets. The eigenvector $\hat{v}^{(2)}$ provides an ordering of the boxes $\{B_1, \ldots, B_n\}$ covering M. Let $\{I(1), \ldots, I(n)\}$ be a permutation of $\{1, \ldots, n\}$ with the property that $\hat{v}^{(2)}_{I(i)} \leq \hat{v}^{(2)}_{I(i+1)}$ for $i = 1, \ldots, n-1$. This ordering of the elements of $\hat{v}^{(2)}$ into ascending order provides an ordering of the boxes: $B_{I(1)}, \ldots, B_{I(n)}$. Those elements of $\hat{v}^{(2)}$ that are near to -1 we "map" to -1 and those that are near to 1 we "map" to 1. This "mapping" is achieved by

$$\hat{v}^{(2)}_i \leq c \Rightarrow \hat{v}^{(2)} \mapsto -1, \tag{12}$$

$$\hat{v}^{(2)}_i > c \Rightarrow \hat{v}^{(2)} \mapsto 1 \tag{13}$$

for a suitable choice of $\min_{1\leq i\leq n} \hat{v}^{(2)}_i \leq c \leq \max_{1\leq i\leq n} \hat{v}^{(2)}_i$. This mapping will define the two sets A_1 and A_2 via

$$A_1 = \bigcup_{i\in\mathcal{I}_1} B_i, \qquad \text{where } \mathcal{I}_1 = \{i : v^{(2)}_i \leq c\}, \tag{14}$$

$$A_2 = \bigcup_{i\in\mathcal{I}_2} B_i, \qquad \text{where } \mathcal{I}_2 = \{i : v^{(2)}_i > c\}. \tag{15}$$

Methods for finding a good choice of c are described in [12]; they include (i) an exhaustive search along the vector $\hat{v}^{(2)}$ and (ii) clustering the values $\hat{v}^{(2)}_i$ and choosing c to divide the two clusters found.

q almost-invariant sets. When we search for $q > 2$ almost-invariant sets, we have the option of using more information from further eigenvectors $\hat{v}^{(3)}, \ldots, \hat{v}^{(q)}$. Each of these eigenvectors provides an ordering of the boxes in a completely analogous way to $\hat{v}^{(2)}$. By combining the information from the various orderings, we select our q almost-invariant sets. In practice, we use clustering algorithms to identify q distinct clusters in the set

$$V_\ell = \{(\hat{v}^{(2)}_i, \ldots, \hat{v}^{(\ell)}_i) \in \mathbb{R}^{\ell-1} : i = 1, \ldots, n\}$$

with $\ell = \lceil \log_2 q \rceil + 1$. Index sets are determined by

$$\mathcal{I}_k = \{i \in \{1, \ldots, n\} : i \in \text{ Cluster \#}k\}, \quad k = 1, \ldots, q,$$

and the almost-invariant sets themselves by

$$A_k = \bigcup_{i\in\mathcal{I}_k} B_i$$

(see [12] for details). The use of multiple eigenvectors to represent points in higher-dimensional Euclidean space appears in the context of minimal graph cuts [6,2] and implicitly in conformation identification [9].

5. The basic algorithm and numerical results

We begin this section by describing the computational steps required to perform the analysis described above. The main steps are similar to the algorithm presented in [12]; the crucial practical differences are the use of the reversibilised transition matrix R instead of the Laplacian matrix L, and the use of weighted fuzzy clustering in place of the balancing techniques described in [12]. Beyond this, the use of R provides many theoretical advantages over the methods proposed in [12]. The relevant convergence result for the Basic Algorithm is Theorem 1. The numerical results section compares the Basic Algorithm with results obtained using methods in [12].

5.1. The basic algorithm

In order to arrive at a partition of M into q almost-invariant sets made up of small boxes, we have to proceed through the following six main steps.

5.1.1. Main computational steps

1. Computation of box covering $\{B_1, \ldots, B_n\}$.
2. Computation of transition matrix P.
3. Computation of the reversibilised transition matrix R.
4. Computation of the large eigenvalues of P to determine a value of q.
5. Computation of the large eigenvalues of R and their corresponding eigenvectors.
6. Selecting almost-invariant sets by identifying clusters in V_ℓ.

Steps 1 and 2 are described in [7]. Step 3 is straightforward, and Steps 4 and 5 may be accomplished using standard numerical packages such as MATLAB. Step 6 is achieved using a variant of fuzzy clustering, where points in V_ℓ are weighted according to their corresponding μ-mass; this variant is described in the following section. Once the box covering and transition matrix R have been constructed, the Basic Algorithm for separating the box collection in the covering into q sets $A_1, \ldots, A_q$ is as follows.

Algorithm 1 (Basic Algorithm).

1. Compute the eigenvectors of R, $\hat{v}^{(2)}, \ldots, \hat{v}^{(\ell)}$, that correspond to the ℓ largest eigenvalues (not including the eigenvalue 1). Normalise each eigenvector to have an l^2-norm of 1.
2. Identify q clusters in the data set $V_\ell = \{(\hat{v}_i^{(2)}, \ldots, \hat{v}_i^{(\ell)}) \in \mathbb{R}^{\ell-1} : i = 1, \ldots, n\} \subset \mathbb{R}^{\ell-1}$ using the weighted fuzzy clustering technique described in Section 5.2.
3. Denote $\mathcal{I}_k = \{i \in \{1, \ldots, n\} : i \in \text{Cluster}\,\#k\},\ k = 1, \ldots, q$.
4. Return $A_k = \bigcup_{i \in \mathcal{I}_k} B_i, k = 1, \ldots, q$.

We now describe a modification of the standard fuzzy clustering algorithm [4] to take into account weights assigned to each point in the set being clustered.

5.2. Weighted fuzzy clustering

Define $\mathbf{x} = [\hat{v}^{(2)}, \ldots, \hat{v}^{(\ell)}] \in \mathbb{R}^{n \times (\ell-1)}$; this is an $(\ell - 1)$-dimensional embedding of the data contained in the ℓ largest eigenvectors of R (not counting the eigenvector corresponding to the eigenvalue 1). The standard fuzzy

clustering algorithm (see e.g. [4]) searches for clusters in the data set $\mathbf{x}$, considered as n points in $\mathbb{R}^{\ell-1}$. The standard cost function for fuzzy clustering (see Eq. (2) [4]) is

$$J(U,c) = \sum_{i=1}^{n}\sum_{k=1}^{q} U_{ki}^2 \|\mathbf{x}_{i,\cdot} - c_k\|^2, \tag{16}$$

where $c_k \in \mathbb{R}^{\ell-1}$ is the kth cluster centre and U_{ki} is the probability that point i (namely $\mathbf{x}_{i,\cdot}$) is contained in cluster k. We modify this function by adding a factor w_i^2 to the ith data point to weight its importance according to the invariant measure of the ith box B_i. Now, boxes with large weights will tend to be forced into tighter, smaller clusters, and boxes with smaller weights are allowed to be placed into large, sparsely distributed clusters. In this way, we overcome the problem of single clusters containing all the high weight boxes, and cluster according to a balanced mix of distance and weight. The modified cost function is

$$J(U,c) = \sum_{i=1}^{n}\sum_{k=1}^{q} U_{ki}^2 w_i^2 \|\mathbf{x}_{i,\cdot} - c_k\|^2. \tag{17}$$

The modified "distance" is (cf. [4, (3b)])

$$d_{ki} = w_i^2 \|\mathbf{x}_{i,\cdot} - c_k\|^2,$$

and the modified update rule for the cluster centres is (cf. [4, (3a)])

$$c_k = \frac{\sum_{i=1}^{n} U_{ki}^2 w_i^2 \mathbf{x}_{i,\cdot}}{\sum_{i=1}^{n} U_{ki}^2 w_i^2}. \tag{18}$$

The update rule for the probabilities U_{ki} remains unchanged:

$$U_{ki} = \left(\sum_{j=1}^{q} d_{ki}/d_{ji}\right)^{-1}. \tag{19}$$

We have found that weighted fuzzy clustering works very well, yielding sets $A_1, \ldots, A_q$ with similar measure.

5.3. Numerical results

We compare new techniques presented in this paper with the methods of [12]. We return to the Lorenz systems of ODEs:

$$\dot{x} = \sigma(y-x), \qquad \dot{y} = \rho x - y - xz, \qquad \dot{z} = xy - \beta z \tag{20}$$

with standard parameters $\sigma = 10$, $\rho = 28$ and $\beta = 8/3$. As in [12], we cover the Lorenz attractor with 5025 boxes. The transition matrix R has second eigenvalue $\lambda_2 = 0.9804$, leading to bounds of

$$0.8022 \le \max_{A\in\mathcal{C}_n} \rho_{\mu_n}(A) \le 0.9902.$$

Table 1 compares the results of Basic Algorithm 1 to the results obtained in [12] using a related approach. The results displayed in Table 1 show that in this example, Basic Algorithm 1 is comparable to Algorithm 4 in [12], when Algorithm 4 is combined with the balancing procedure described in [12]. Basic Algorithm 1 is clearly superior to the methods of [12] when balancing is not used. Advantages of Basic Algorithm 1 over Algorithm 4 [12] with

Table 1
Data for three almost-invariant sets of the Lorenz system: a comparison of the methods of [12] vs. Basic Algorithm 1

Method	$\mu_n(A_1)/\mu_n(A_2)/\mu_n(A_3)$	ρ_{μ_n}
Algorithm 4 [12] ($\ell = 1$)	0.0588/0.8823/0.0590	0.8888
Algorithm 4 [12] ($\ell = 2$)	0.0308/0.9384/0.0308	0.8847
Algorithm 4 [12] ($\ell = 1$, with balancing)	0.2539/0.4922/0.2539	0.8911
Algorithm 4 [12] ($\ell = 2$, with balancing)	0.1495/0.7011/0.1495	0.9022
Basic Algorithm ($\ell = 1$)	0.2255/0.5491/0.2255	0.8994
Basic Algorithm ($\ell = 2$)	0.2068/0.5865/0.2068	0.9017

balancing are (i) the availability of theoretical bounds on how well one can expect to do, (ii) the existence of a theoretical minimisation principle for the eigenvectors of R and (iii) numerical stability.[5]

6. Hybrid adaptive algorithms, convergence and numerical results

This section contains our main theoretical result. To reduce the number of boxes necessary to provide a good approximation of almost-invariant sets, we propose an adaptive strategy whereby the interiors of each almost-invariant set are covered by large boxes, while the boundaries or interfaces between sets are covered by smaller boxes to more accurately resolve the individual sets. The main steps of the algorithm are as follows:

I. *Order* the boxes using the eigenvectors of R.
II. *Assign* each box to a large almost-invariant set by searching for clusters within the eigenvector values.
III. *Identify* those boxes near the boundaries of the almost-invariant sets for further subdivision.

Steps I and II have already been outlined; they are Steps 1 and 2–4, respectively, of Algorithm 1. Step III is new and we now describe an efficient method of identifying the boundaries of the almost-invariant sets.

6.1. Boundary refinement

Suppose that we are presented with a covering of M consisting of n boxes $\{B_1, \ldots, B_n\}$, i.e., $M \subset \bigcup_{i=1}^{n} B_i$ with the B_i's pairwise disjoint. The boxes in the collection $\{B_1, \ldots, B_n\}$ may be of different sizes and contain different amounts of invariant mass μ. We assume that we have performed Steps I and II, and have assigned each of these n boxes to one of q approximate almost-invariant sets $A_1, \ldots, A_q$. Associated with each box B_i is a number $k_i \in \{1, \ldots, q\}$ describing which of the q sets $A_1, \ldots, A_q$ the box B_i belongs to, e.g., $k_5 = 3$ if box B_5 belongs to the set A_3.

For the following discussion, let us fix k and concentrate on a fixed almost-invariant set A_k. Many of the boxes $B_i \subset A_k$ will be mapped entirely inside A_k in one iteration of our map. These boxes are "good" because they make A_k "look" more invariant; if A_k was truly invariant, then *all* boxes making up A_k would be mapped inside A_k at the next iteration. There will of course, also be many boxes in A_k that are either partially or completely mapped *outside* A_k. In view of our definition of ρ_{μ_n}, clearly those boxes which "take a lot of mass out of A_k under one iteration" reduce ρ_{μ_n} the most. The amount of μ_n-mass flowing from box $B_i \subset A_k$ out of A_k in one iteration is given by

$$\text{Outflow}(i) := \mu_n(B_i)\left(1 - \sum_{j \in \mathcal{I}_k} R_{ij}\right).$$

[5] The modified matrix produced by balancing can become ill-conditioned if the density of μ_n is very small in many boxes: a typical situation for chaotic attractors. In contrast, the stochastic matrix R is well behaved even when $\mu(B_i)$ is very small.

Therefore, "bad" boxes are those with large values of Outflow. To identify boxes for further subdivision we propose the following.

Algorithm 2.

1. Calculate

$$\text{MeanOutflow} := \frac{\sum_{i=1}^{n} \text{Outflow}(i)}{n}.$$

2. Subdivide all boxes B_i with $\text{Outflow}(i) > \text{MeanOutflow}$.

An alternative criterion that will subdivide less boxes at each iteration is:

1a. Calculate

$$\text{MeanOutflow} := \frac{\sum_{i:\text{Outflow}(i)>0} \text{Outflow}(i)}{\#\{i : \text{Outflow}(i) > 0\}}.$$

Once boxes with large Outflow have been subdivided, a new computation is performed on the refined box covering, and a new ordering and assignment is carried out through Steps I and II. In terms of Basic Algorithm 1, we expect the subdivision of boxes with large Outflow to be helpful for the following reason. The eigenvectors of the reversibilised transition matrix R provide orderings of the boxes in the phase space M. Once we have these orderings, we select cut points using the weighted fuzzy clustering algorithm. Boxes near these cut points are ordered as "close" by the eigenvectors; there is some uncertainty as to exactly which side of the cut such boxes should be. By subdividing these boxes and re-applying Basic Algorithm 1 to the refined collection, this separation should become more clear, and the boundaries will become better resolved.

6.2. Balancing boundary refinement, box refinement, and measure approximation

Recall that our goal is to (i) find estimates of optimal almost-invariant sets, while (ii) concentrating on resolving the boundaries of these sets. In order to find good estimates of optimal almost-invariant sets, we need to have the box collection $\{B_1, \ldots, B_n\}$ consisting of reasonably small boxes in terms of μ-measure. If $\{B_1, \ldots, B_n\}$ contains overly large boxes in measure, then *all* of the sets in $\mathcal{C}_n$ may be far (in μ-measure) from the optimal almost-invariant sets, and thus even if we choose the best sets from $\mathcal{C}_n$, they will be relatively poor in terms of almost-invariance. Additionally, in order to have a good estimate of μ, we require the boxes to be small (or at least small in mass). We therefore cannot simply subdivide those boxes that we think are near the boundaries of our current best guess; doing so may lead to poor estimates of μ and skew our results. By refining other boxes that are not designated as boundary boxes, we not only improve our estimate of μ, but also fill $\mathcal{C}_n$ with a richer class of boxes. Thus, we need to somehow balance the refinement of boundary boxes with those away from the boundary. A simple and robust solution is to set a threshold ratio that says that any boxes larger in mass than a fixed ratio of the smallest box *must* be subdivided at the next step, even if they are not boundary boxes. We arrive at the following hybrid adaptive algorithm:

Algorithm 3 (Hybrid Adaptive Algorithm).

0. Begin with an initial box covering $\{B_1, \ldots, B_n\}$ of the set M and a fixed measure ratio $r > 1$.
1. Compute the transition matrix P for the dynamical system on the current box covering.

2. Create the reversibilised matrix R and assign each of the boxes in the current covering to one of q almost-invariant sets using Basic Algorithm 1.
3. Compute the Outflow for every box in the current covering, and mark those boxes with greater than average Outflow for subdivision at the next step.
4. Mark those boxes with measure exceeding $r(\min_{1\le i\le n} \mu_n(B_i))$ for subdivision at the next step.
5. Subdivide the marked boxes to create a refined box covering, and return to Step 1.

The ratio r may be chosen according to the problem at hand. In cases where the physical-invariant measure is highly variable or difficult to approximate, lower values of r should be used to obtain reliable estimates. If the physical-invariant measure is relatively easily approximable, a larger value of r may be used to concentrate more heavily on refining the boundaries of the almost-invariant sets.

The adaptive element of Algorithm 3 can also be incorporated into searches for m-almost-invariant sets as described in [12]. One simply replaces μ_n with m in Steps 3 and 4 of Algorithm 3. The greedy combinatorial searches described in [12] can also be modified to take advantage of the adaptive aspect. Simply substitute a combinatorial search (or any other search method of your choice) in place of Step 2 of Algorithm 3 to identify the almost-invariant sets.

6.3. A convergence result for the hybrid algorithm

In this section, we consider dynamical systems governed by a continuous map T perturbed by a small amount of noise. Such situations are clearly important when modelling real-world systems. The evolution of such a dynamical system is governed by a transition function $\mathcal{P} : X \times \mathcal{B}(X) \to [0, 1]$, where the number $\mathcal{P}(x, A)$ is the probability that by iterating the point x, the image is in the set $A \subset X$. For example,

$$\mathcal{P}(x, A) = \mathcal{N}(x) \int_A \exp\left(\frac{-\|Tx - y\|^2}{2\epsilon}\right) \mathrm{d}m(y)$$

is a transition function describing the evolution $x \mapsto Tx + \xi$, where ξ is normally distributed with mean zero and variance ϵ; the factor $\mathcal{N}(x)$ is a normalising factor. In the following, we will suppose that $\mathcal{P}(x, \cdot)$ has a Lipschitz density; this is certainly the case for a deterministic map perturbed by normally distributed noise, and many other physically relevant situations are also covered. One has the following convergence result for Algorithm 3.

Theorem 2. *Suppose that the map $T : X \to X$ is perturbed by noise with a Lipschitz density, and that the resulting stochastic system has a unique invariant measure μ. Let $\mathcal{C}_n^m$ denote the box collection that is obtained after m applications of Algorithm 3 with $r < \infty$. As $m \to \infty$,*

$$\max_{A_1,\ldots,A_q \in \mathcal{C}_n^m} \{\rho_{\mu_n}(A_1, \ldots, A_q) : \{A_1, \ldots, A_q\} \text{ partition a tight covering of } M\} \to \rho_\mu^{\max}. \tag{21}$$

Proof. We restrict our attention to $\Lambda = \operatorname{supp} \mu \subset M$, and consider the box collection $\mathcal{D}_n^m = \{B \in \mathcal{C}_n^m : \operatorname{Int}(B) \cap \operatorname{supp} \mu \neq \emptyset\}$. We will use the sets in $\mathcal{D}_n^m \cap \operatorname{supp} \mu$ to be the building blocks of almost-invariant sets in $\operatorname{supp} \mu$. Let us assume for the moment that under repeated application of Algorithm 3 ($m \to \infty$), the *diameters* of boxes in $\mathcal{D}_n^m$ approach zero; this is proven in Sublemma 1. Proposition 3.7 [14] may then be applied to show that $\mu_n \to \mu$ strongly. Theorem 6.3 [12] then yields the desired result. □

Sublemma 1. *Under repeated application of Algorithm 3 ($m \to \infty$), the diameters of boxes in $\mathcal{D}_n^m$ approach zero.*

Proof. Let us suppose that there is a sequence of boxes $B_{i_m}^{(m)} \in \mathcal{D}_n^m$, $1 \le i_m \le n$ and an $M \ge 1$ such that $B_{i_m}^{(m)} = B_{i_M}^{(M)}$ for all $m \ge M$. The box $B_{i_M}^{(M)}$ is a problem box because after M applications of Algorithm 3, it is not subdivided

Table 2
Data for three almost-invariant sets of the standard map with $a = 8.0$: a comparison of adaptive vs. non-adaptive methods

Method	# Boxes	$\mu_n(A_1)/\mu_n(A_2)/\mu_n(A_3)$	$\rho_{\mu_n}(A_1)/\rho_{\mu_n}(A_2)/\rho_{\mu_n}(A_3)$	λ_2 bound
Algorithm 1	4096	0.1525/0.6821/.0.1654	0.7510/0.9094/0.7610	0.9557
Algorithm 3 ($r = 25$)	3711	0.1255/0.7464/0.1281	0.7621/0.9328/0.7629	0.9456
Algorithm 3 ($r \approx \infty$)	3545	0.1132/0.7736/0.1132	0.7751/0.9448/0.7752	0.9422

any further, and so the diameter of this box does not go to zero. We will now derive a contradiction, proving the sublemma.

The quantity $r(\min_{1\le i\le n} \mu_n(B_i))$ in Step 4 of Algorithm 3 is no greater than r/n. Therefore any boxes in $\mathcal{D}_n^m$ with measure exceeding r/n will surely be subdivided at the next iteration of the algorithm. If such a box $B_{i_M}^{(M)}$ exists, it must have measure smaller than $r/n = r/n(m)$ at the mth stage of the algorithm for all $m \ge M$.

We wish to apply Theorem 6.1 [15] to show that $\mu_{n(m)} \to \mu$ as $m \to \infty$, where weak convergence is meant. The main condition of this theorem is equivalent to $\max_{1\le i\le n} \mu_n(B_i) < \Delta_n$, where $\Delta_n \to 0$ as $n \to \infty$, and $B_i \in \mathcal{C}_n^m$. Note that there are always *some* boxes being subdivided. The subdivision of boxes with high measure clearly decreases the maximal box weight and the subdivision of boxes with high Outflow also leads to a decrease in measure of all boxes via the condition in Step 4 of Algorithm 3. Thus we may directly apply Theorem 6.1 [15] to conclude that $\mu_{n(m)} \to \mu$ weakly as $m \to \infty$.

Set $s_m = \mu_{(n(m))}(B_{i_M}^{(M)})$ for all $m \ge M$, and $s_\infty = \mu(B_{i_M}^{(M)})$. We are worried about the possibility that $s_m \le r/n(m)$ for all $m \ge M$, as this will mean no further subdivision of $B_{i_m}^{(m)}$ beyond $m = M$. If this occurs then $s_m \to 0$ as $m \to \infty$. Since the interior of $B_{i_M}^{(M)}$ intersects $\operatorname{supp}\mu$, given $x \in \operatorname{Int}(B_{i_M}^{(M)}) \cap \operatorname{supp}\mu$, one can choose $\delta > 0$ sufficiently

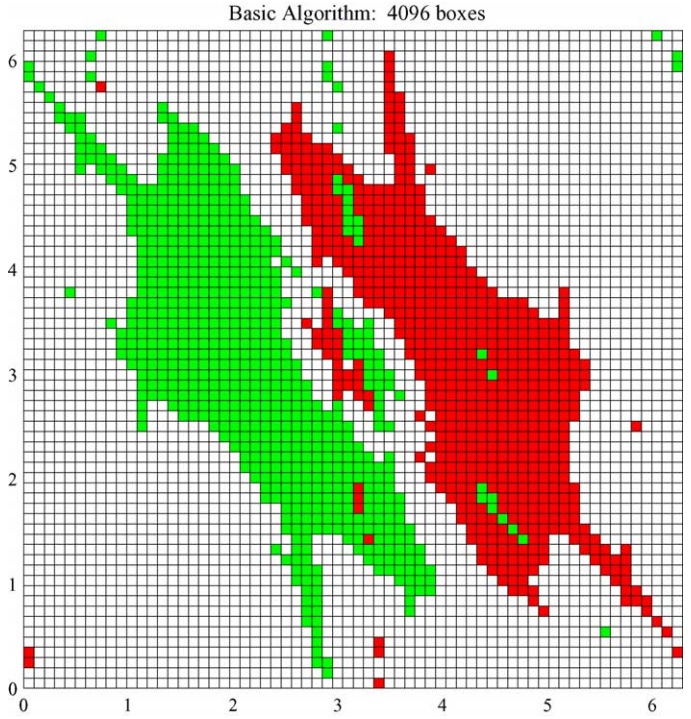

Fig. 2. Three almost-invariant sets for the standard map, using the Basic Algorithm: 4096 boxes shown in $[0, 2\pi) \times [0, 2\pi)$.

small that $O := B_\delta(x) \subset \mathrm{Int}(B_{i_M}^{(M)})$. Moreover, $\mu(O) > 0$ by the definition of support. By weak convergence of μ_n to μ,

$$\liminf_{m\to\infty} s_m = \liminf_{m\to\infty} \mu_{n(m)}(B_{i_M}^{(M)}) = \liminf_{n\to\infty} \mu_n(B_{i_M}^{(M)}) \geq \liminf_{n\to\infty} \mu_n(O) \geq \mu(O) > 0.$$

A contradiction; therefore no such set $B_{i_M}^{(M)}$ exists, and all sets whose interiors intersect supp μ will be subdivided further, decreasing their diameter in the process. □

Sublemma 1 shows that under the assumptions of Proposition 3.7 [14], the weak convergence result of [15] under the "high mass" subdivision algorithm may be easily strengthened to strong convergence of μ_n to μ.

Corollary 1. *Under the assumptions of Proposition 3.7 [14], the "high mass" subdivision algorithm of [15] produces a sequence of approximate invariant measures μ_n that converge to μ strongly under repeated application of the algorithm.*

Proof. This is simply a result of the combination of Sublemma 1 and Proposition 3.7 [14]. □

6.4. Numerical results

We consider a Hamiltonian system preserving Lebesgue measure. Define the so-called standard map $S : \mathbb{T}^2 \to \mathbb{T}^2$ by

$$S(x, y) = (x + y, y + a\sin(x + y)) \qquad \mod 2\pi. \tag{22}$$

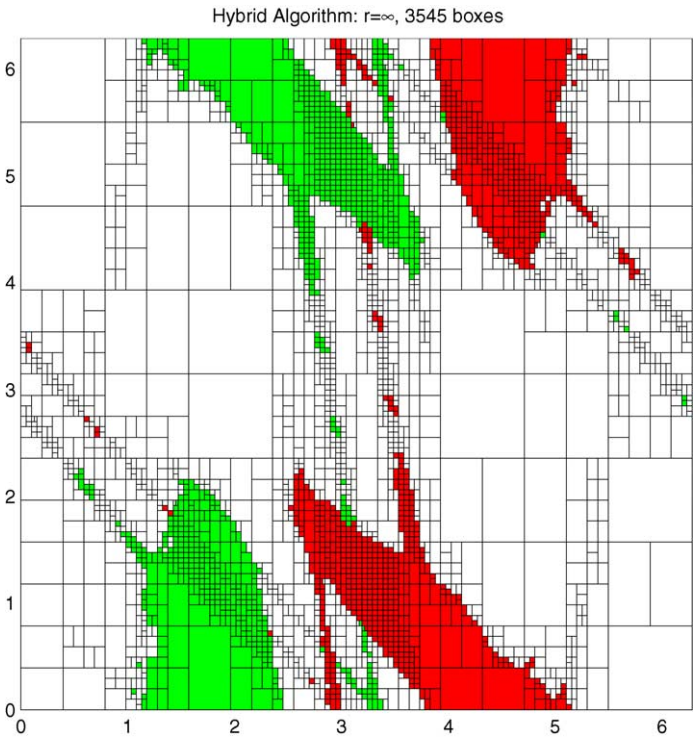

Fig. 3. Three almost-invariant sets for the standard map, using the Hybrid Adaptive Algorithm with $r = \infty$: 3545 boxes shown in $[0, 2\pi) \times [0, 2\pi)$.

We choose the parameter $a = 8.0$; in this parameter region, the standard map has no obvious invariant sets, in contrast the invariant KAM curves existing at lower parameter values (see, e.g. [11, p. 477]). We expect that any almost-invariant sets will have complicated boundaries, and that the Hybrid Adaptive Algorithm 3 may be advantageous. Since the physical measure μ is Lebesgue measure in this example, it is easily approximable, and we choose $r = 25$ in Algorithm 3. We also test Algorithm 3 with $r \approx \infty$ (refining only near boundaries, and assuming the physical measure is always well approximated). These two tests are compared with a uniform box covering containing a similar number of boxes, to reveal the advantages of Algorithm 3 over Algorithm 1. Table 2 displays the results of these tests, searching for three almost-invariant sets using the second and third eigenvectors of R ($\ell = 2$).

It is clear that the Hybrid Adaptive Algorithm 3 is superior to the Basic Algorithm 1 in this example. The bounds of Section 3.1 are shown in the final column of Table 2. In both applications of Algorithm 3, not only has $\rho_{\mu_n}(A_k)$ increased, but the theoretical upper bound has also tightened. Figs. 2 and 3 show the three almost-invariant sets obtained via Algorithms 1 and 3, respectively. One recovers the superior partition of almost-invariant sets shown in Fig. 3 using Algorithm 1 if 65536 uniform boxes are used.

Acknowledgement

Work carried out at the Department of Mathematics and Statistics, The University of Western Australia, Nedlands, WA 6907, Australia, and the Department of Mathematics and Computer Science, University of Paderborn, Paderborn D-33095, Germany, under the DFG grant De 448/5-4.

References

[1] K. Alligood, T. Sauer, J. Yorke, Chaos: An Introduction to Dynamical Systems, Textbooks in Mathematical Sciences, Springer-Verlag, New York, 1997.
[2] C.J. Alpert, A.B. Kahng, S.-Z. Yao, Spectral partitioning with multiple eigenvectors, Discrete Appl. Math. 90 (1999) 3–26.
[3] E. Behrends, Introduction to Markov Chains: With Special Emphasis on Rapid Mixing, Vieweg, Wiesbaden, 2000.
[4] J. Bezdek, R. Hathaway, M. Sabin, W. Tucker, Convergence theory for fuzzy c-means: counterexamples and repairs, IEEE Trans. Syst. Man Cybernetics SMC-17 (5) (1987) 873–877.
[5] P. Brémaud, Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues Number 31 in Texts in Applied Mathematics, Springer, New York, 1999.
[6] P.K. Chan, M.D.F. Schlag, J.Y. Zien, Spectral k-way ratio-cut partitioning and clustering, IEEE Trans Computer-Aided Des. Integrated Circuits Syst. 13 (9) (1994) 1088–1096.
[7] M. Dellnitz, O. Junge, On the approximation of complicated dynamical behavior, SIAM J. Numer. Anal. 36 (2) (1999) 491–515.
[8] M. Dellnitz, R. Preis, Congestion and almost-invariant sets in dynamical systems, Preprint.
[9] P. Deuflhard, W. Huisinga, A. Fischer, C. Schütte, Identification of almost-invariant aggregates in nearly uncoupled Markov chains, Linear Algebra Appl. 315 (2000) 39–59.
[10] G. Froyland, Approximating physical invariant measures of mixing dynamical systems in higher dimensions, Nonlinear Anal. Theory, Meth. Appl. 32 (7) (1998) 831–860.
[11] G. Froyland, Using Ulam's method to calculate entropy and other dynamical invariants, Nonlinearity 12 (1999) 79–101.
[12] G. Froyland, M. Dellnitz, Detecting and locating near-optimal almost-invariant sets and cycles, SIAM J. Sci. Comput. 24 (6) (2003) 1839–1863.
[13] W. Huisinga, Metastability of Markovian systems: a transfer operator approach in application to molecular dynamics, Ph.D. thesis, Free University Berlin, 2001.
[14] O. Junge, Mengenorientierte Methoden zur numerischen Analyse dynamischer Systeme, Ph.D. thesis, University of Paderborn, 2000.
[15] O. Junge, An adaptive subdivision technique for the approximation of attractors and invariant measures: proof of convergence, Dyn. Syst. 16 (3) (2001).
[16] C. Schütte, Conformation dynamics: modelling, theory, algorithm, and application to biomolecules, Habilitation Thesis, Freie Universität Berlin, 1999.