Chaos, Solitons and Fractals 170 (2023) 113329

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

journal homepage: www.elsevier.com/locate/chaos

# A general framework for the rigorous computation of invariant densities and the coarse-fine strategy

S. Galatolo [a], M. Monge [b], I. Nisoli [b,c,d,*], F. Poloni [e]

[a] Università di Pisa, Dipartimento di Matematica, Largo Pontecorvo, 2 56127, Pisa, Italy
[b] Universidade Federal do Rio de Janeiro, Instituto de Matemática, Av. Athos da Silveira Ramos, 149, Edifício do Centro de Tecnologia, Bloco C (Térreo), Cidade Universitária, Brazil
[c] Department of Mathematics, Hokkaido University, N10 W8, Kita-ku, Sapporo 001-0010, Japan
[d] RIES, Hokkaido University, N20 W10, Kita-ku, Sapporo 001-0020, Japan
[e] Università di Pisa, Dipartimento di Informatica, Largo Pontecorvo, 3 56127, Pisa, Italy

ARTICLE INFO

*MSC:*
37M25
37-04
65P99

*Keywords:*
Invariant densities
Finite element reduction
Transfer operator
Coarse-fine strategy
Rigorous error bounds

ABSTRACT

In this paper we present a general, axiomatical framework for the rigorous approximation of invariant densities and other important statistical features of dynamics. We approximate the system through a finite element reduction, by composing the associated transfer operator with a suitable finite dimensional projection (a discretization scheme) as in the well-known Ulam method.

We introduce a general framework based on a list of properties (of the system and of the projection) that need to be verified so that we can take advantage of a so-called "coarse-fine" strategy. This strategy is a novel method in which we exploit information coming from a coarser approximation of the system to get useful information on a finer approximation, speeding up the computation. This coarse-fine strategy allows a precise estimation of invariant densities and also allows to estimate rigorously the speed of mixing of the system by the speed of mixing of a coarse approximation of it, which can easily be estimated by the computer.

The estimates obtained her e are *rigorous*, i.e., they come with exact error bounds that are guaranteed to hold and take into account both the discretization and the approximations induced by finite-precision arithmetic.

We apply this framework to several discretization schemes and examples of invariant density computation from previous works, obtaining a remarkable reduction in computation time.

We have implemented the numerical methods described here in the Julia programming language, and released our implementation publicly as a Julia package.

## 1. Introduction

Several important features of the statistical behavior of a dynamical system are related to the properties of its invariant measures and in particular to the properties of the so called *Physical Invariant Measure*.[1] The knowledge of the invariant measure of interest, gives information on the statistical behavior for the long time evolution of the system. This fact strongly motivates the search for algorithms which are able to compute quantitative information about invariant measures of physical interest, and in particular, algorithms giving an explicit bound on the error which is made in the approximation. The application of such rigorously certified estimates allows to get reliable information on the statistical behavior of the system and perform computer-aided proofs, establishing rigorously proved statements on the statistical behavior of the system (see e.g. [1]).

***Several levels of precision in the estimation of the approximation error***. The problem of approximating some interesting invariant measure of a deterministic or random dynamical system is widely studied in the literature. Some algorithms are proved to converge to the real invariant measure (up to errors in some given metrics) in some classes of systems. Sometimes asymptotical estimates on the rate of

* Corresponding author at: Universidade Federal do Rio de Janeiro, Instituto de Matemática, Av. Athos da Silveira Ramos, 149, Edifício do Centro de Tecnologia, Bloco C (Térreo), Cidade Universitária, Brazil.

*E-mail addresses:* stefano.galatolo@unipi.it (S. Galatolo), maurizio.monge@im.ufrj.br (M. Monge), nisoli@im.ufrj.br, isaia.nisoli@es.hokudai.ac.jp (I. Nisoli), federico.poloni@unipi.it (F. Poloni).

[1] This is a class of invariant measures representing the statistical behavior of large sets of initial conditions and having particular interest in the applications, see [43] for a survey on the subject.

https://doi.org/10.1016/j.chaos.2023.113329
Received 4 February 2023; Accepted 2 March 2023
Available online 14 March 2023

convergence are provided (see e.g. [2–9]); other results and algorithms give an explicit bound on the error (see e.g. [1,10–16]). This is the point of view of the present paper.

We are not only interested to the algorithm but also to a suitable implementation. In fact, implementing such an algorithm in a software which is able to keep track of the various truncations and numerical errors in the computation allows the result of a single computation to be interpreted as a computer-aided proved statement on the behavior of the observed system, and hence it has a mathematical meaning. In the literature the dimension of some nontrivial attractors or repellers was estimated in this way (see e.g. [16–18]), as well as escape rates [19], linear response [20,21], diffusion coefficients [22,23] or the behavior of Lyapunov exponents in models of real phenomena [1,24].

It is worth noting that some negative result are known about the general problem of computing invariant measures up to a small given error. In [25] it is shown that *there are examples of computable*[2] *systems without any computable invariant measure*. This phenomenon shows that there is some subtlety in the general problem of computing invariant measures up to a given error.

***Finite element reductions based on a projection and the present paper***. The techniques used in the literature to establish rigorous bounds on the approximation error are often related to a suitable finite-element reduction of the transfer operator of the system. In this approach the transfer operator is approximated by a finite-rank one. The invariant measures of the system under study can be seen as fixed points of its transfer operator when acting on suitable functional spaces. These fixed points can then be approximated by the fixed points of the finite-dimensional reduction of the operator. Suitable quantitative fixed-point stability results can give a bound of this approximation error.

For this purpose, several approaches have been implemented. The Ulam method is a classical example of such a finite elements reduction, and provides a finite dimensional approximation of the transfer operator with a finite Markov chain obtained by discretizing the space by a cell subdivision; see Section 6 for a precise definition. In this approach, and in other finite-element reductions, the transfer operator is approximated by a finite-dimensional operator defined by the composition of the original operator with suitable projections to a finite-dimensional functional space. In the classical Ulam method, a probability density is approximated by a piecewise constant one and the projection is then a conditional expectation made on the cell subdivision of the whole space. Other approaches use different approximation schemes, as for instance a piecewise linear approximation (see Section 7), piecewise smooth approximations, or even other approximation schemes based on Fourier analysis or Taylor series [15,26], which are suitable for smooth systems. All of these approaches require their own estimates and have advantages for certain classes of systems: for instance, approximation schemes based on the projection to spaces of smooth functions converge faster when used to approximate smooth systems with smooth invariant measures. These approaches can be seen as examples of a general construction in which one defines a finite-dimensional reduction of some operator by composing it with a suitable finite dimensional projection.

In this paper we consider this "projection based" finite-element reduction point view in general, and show that if the finite element reduction method satisfies a certain list of hypotheses, then we can apply a general construction in which the computation of the invariant density up to a small explicit approximation error will work efficiently.

To estimate this approximation error, we will consider the finite element reduction of the system as a small perturbation of the system itself and estimate quantitatively the stability of the invariant measure of a system up this small perturbation. These kinds of estimates are also called quantitative statistical stability estimates. It is known that the quantitative statistical stability of a system is related to the speed of convergence to equilibrium of the system itself: the faster is this speed of convergence, the more the system is statistically stable (see e.g. [27] for a general statement adapted to many convergence rates).[3] This is a delicate point in many papers related to rigorous computations of invariant measures, where the estimate for approximation error involves an estimate for the convergence to equilibrium of the system. Establishing an effective (not only asymptotical) estimate for the convergence to equilibrium of the system is not trivial. This problem is sometimes approached by a-priori estimates which are possible only on restricted families of systems. For example, in circle expanding maps such explicit estimates on the convergence to equilibrium can be done by using Hilbert cones related techniques. In [17] an idea to overcome this difficulty was proposed, and in this paper a construction is shown, in which the *a priori* estimate on the speed of convergence is replaced by some *a posteriori* one which is computed on the finite element reduction of the system. This is a finite dimensional system (and the transfer operator can be represented by a large and sparse matrix) and its speed of convergence to equilibrium can be estimated directly by the computer. This idea allowed [17] to compute with explicit error bounds invariant densities of quite different systems as expanding maps, piecewise-expanding ones without a Markov partition and even non-uniformly expanding ones (examples of Manneville–Pomeau maps), essentially applying the same construction for each one of these systems. An estimate of the convergence rate of a finite-dimensional system, as we need in the "a posteriori" approach, is always possible, but it can be a challenging task when the related matrix is large.

In [19], a method to speed up this computation was proposed and applied to some class of examples. This method exploits the regularization properties of the transfer operator to infer the speed of convergence to equilibrium of a finite-dimensional reduction of the system from a coarser finite-element reduction (hence reducing the dimension of the matrix to be considered when estimating the speed of convergence to equilibrium). We will refer to this kind of approach as a "coarse-fine" approach. In [1], a similar approach was applied to estimate the convergence to equilibrium of high-resolution finite-element reductions (the rank of the reduced operator is of the order of millions) of transfer operators related to a class of random systems which are models of the behavior of the famous Belosouv–Zhabotisky chaotic chemical reaction, proving the existence of a noise-induced phenomenon observed by numerical simulation in 1983 in the article [28].

In the present paper we propose a general systematic formalization of this method, adapting it to different kinds of projection based finite dimensional reductions. We also implemented these ideas in the Julia language [29], in a package called `RigorousInvariantMeasures.jl`, which is part of the JuliaDynamics organization. The package can be installed through the Julia package manager and the source code for the development version can be found at https://github.com/JuliaDynamics/RigorousInvariantMeasures.jl.

Examples of the use of this package can be found in the `examples` directory. Jupyter notebooks detailing its usage were developed for a summer school at Hokkaido University and can be found at https://github.com/orkolorko/HokkaidoSchool; Lectures 1 and 2 are introductory while Lecture 3 and 4 deal with the rigorous approximation of the invariant density for a deterministic dynamical system and a random dynamical system respectively.

[2] Computable, here means that the dynamics can be approximated at any accuracy by an algorithm, see e.g. [25] for precise definition.

[3] In our paper we will consider the transfer operator associated to the system acting on different weaker or stronger spaces with norms $\| \ \|_w, \| \ \|_s$. Here by speed of convergence to equilibrium we mean the rate of convergence to the invariant measure $\mu$ of iterates $L^n\nu$ of regular initial probability measures $\nu$ by the transfer operator $L$. The speed of convergence to equilibrium will be measured as the speed of convergence to 0 of the ratio $\frac{\|L^n\nu-\mu\|_w}{\|\nu\|_s}$. This notion is also related to the speed of mixing of the system.

We will apply this new package to a series of examples already studied in [17], testing systematically the performance of the computations and showing a major speed-up and increase of precision with the new package.

**Structure of the paper and main results.** In Section 2 we describe the properties we require for our general projection based approximation schemes and the kind of operators to which we mean to apply it. We also show the first useful consequences of these properties, as the fact that the if the original transfer operator satisfy a Lasota–Yorke inequality, also the finite element reduction of the transfer operator satisfies it. In Section 3 we show explicit bounds on the approximation errors made on approximating the fixed points of the original operator with the fixed points of the finite dimensional reduction. In Section 4 we show how to improve this bound and related estimates on the convergence to equilibrium by a coarse-fine strategy, in which we discretize the transfer operator at different resolutions, exploiting the regularization properties of the operator and using information from the coarser discretization to understand the behavior of the finer one, greatly improving the efficiency of the computation.

In Sections 6 and 7 we show two examples of approximation schemes, with associated functional analytic setting satisfying the abstract approximation setting defined at 2 : the Ulam scheme and a smoother approximation scheme based on the approximation by piecewise linear functions.

In Section 8 we discuss some algorithmic aspects of the implementation of our ideas, in particular about the construction of the discretized operators and the estimation of norms of powers of discretized operators. Section 9 presents examples and in Section 10 we present some final discussion and remarks.

**Notation 1.1.** In the following, we use $I$ for the identity matrix/operator/function (it is typically clear from the context which one it is), and $e_j$ for the $j$th vector of the canonical basis (i.e., the $j$th column of $I$).

The symbol $v^*$ denotes the conjugate transpose of a vector.

The symbol $\|f\|_{L^p}$ denotes the $L^p$ norm of a function (usually defined on $[0,1]$), whereas the symbol $\|v\|_{\ell^p}$ denotes the $\ell^p$ norm of a vector $v \in \mathbb{R}^n$.

## 2. The abstract setting

In the following we will consider suitable operators between normed vector spaces of functions over a certain compact manifold with boundary $X$; the main example is the transfer operator of a nonsingular dynamical systems on $X$, see Section 5. We will suppose $X$ to be endowed with the normalized Lebesgue measure $m$ as a reference measure. And denote by $\|.\|_{L^1}$ the norm of the associated space $L^1(X,m)$.

**Assumptions on the space 2.1.** Let $(\mathcal{U}_w, \|.\|)$ be a real or complex Banach space of real or complex functions over $X$ containing the indicatrix 1 of the whole space. Let $\mathcal{U}_s \subseteq \mathcal{U}_w$ be a subspace of more regular functions on which a certain seminorm $\|.\|_s$ is defined. Let us suppose that $(\mathcal{U}_s, \|.\|_s + \|.\|)$ is a Banach space which is compactly embedded in $(\mathcal{U}_w, \|.\|)$.

We will suppose that these norms satisfy the following assumptions, there exists positive constants $S_1, S_2 \in \mathbb{R}$ and an element $i \in \mathcal{U}_s^*$ such that:

(1) $\|1\| = 1$,
(2) $\|1\|_s = 0$,
(3) $|i(f)| \le \|f\|$,
(4) $\|.\|_{L^1} \le \|.\|$,
(5) $\|.\| \le S_1 \|.\|_s + S_2 \|.\|_{L^1}$,
(6) i(1) = 1.

**Example 2.2.** Let $\mathcal{U}_s = BV([0,1])$, the space of functions of bounded variation on $[0,1]$, equipped with the seminorm $\|.\|_s = \mathrm{Var}(.)$ and the norm $\|.\| = \|.\|_{L^1}$, with $S_1 = 1, S_2 = 1$ and

$$i(f) = \int_0^1 f \, dm,$$

where $m$ is the Lebesgue measure on $[0,1]$.

**Example 2.3.** Let $\mathcal{U}_s = \mathrm{Lip}([0,1])$, the space of Lipschitz continuous functions on $[0,1]$, equipped with the seminorm $\|.\|_s = \mathrm{Lip}(.)$ and the norms $\|.\| = \|.\|_\infty$, with $S_1 = 1, S_2 = 1$ and

$$i(f) = \int_0^1 f \, dm,$$

where $m$ is the Lebesgue measure on $[0,1]$.

**Assumptions on the operator 2.4.** Let $L$ be an operator acting on $\mathcal{U}_s$ such that

- $i(Lf) = i(f)$,
- $\|L\| < \infty$,

and suppose that there are $A, B, W \in \mathbb{R}$, with $A < 1$ such that for each $n \ge 0$

$$\|Lf\|_1 \le \|f\|_1 \tag{1}$$
$$\|Lf\|_s \le A\|f\|_s + B\|f\|_{L^1} \tag{2}$$
$$\|L^n f\| \le W\|f\| \tag{3}$$

respectively for all functions in $L^1$ and in $\mathcal{U}_s$. We say such an operator satisfies a **one step Lasota–Yorke inequality**.

**Remark 2.5.** The Lasota–Yorke inequality implies that $L$ has a 'regularizing' behavior, up to a certain point. The general form of the Lasota–Yorke inequality is the following: there are $\lambda < 1$, $A', B \ge 0$ s.t. for each $f \in \mathcal{U}_s$ and $n \ge 0$

$$\|L^n f\|_s \le A'\lambda^n \|f\|_s + B\|f\|_{L^1}. \tag{4}$$

An estimate of this kind can be established in many systems having some form of uniform expansiveness, even in the presence of discontinuities or piecewise hyperbolic behavior. We remark that in this case a suitable iterate of $L$ satisfies (2). This is usually sufficient for the computation of invariant densities of a system, as the invariant density of the original system is also invariant for the iterate.

We remark that since $\|f\| \ge \|f\|_{L^1}$, (2) also implies

$$\|Lf\|_s \le A\|f\|_s + B\|f\| \tag{5}$$

for all functions $f \in \mathcal{U}_s$.

A consequence of (2) is a simple regularity estimate on the fixed points of $L$ which will play an important role in our estimation procedure.

**Corollary 2.6.** *If $L$ satisfies a one step Lasota–Yorke inequality and $u$ is a fixed point of $L$:*

$$\|u\|_s \le \frac{B}{1-A}\|u\|_{L^1}. \tag{6}$$

We are interested to compute the invariant density by a suitable finite element reduction of our system. This finite element reduction will be realized by a suitable projection on a finite dimensional space. We now formalize the requirements we ask for this projection.

**Definition 2.7.** Let $n \in \mathbb{N}$, and $P_h$ be a rank $n$ linear operator defined on $\mathcal{U}_s$, with $h := 1/n$.

We say that $P_h$ is a **compatible discretization** if there exists $K$ and $E$ such that:

(1) $P_h = P_h^2$, i.e., $P_h$ is a projection.
(2) $\|P_h f\| \le \|f\|$ for any function $f \in \mathcal{U}_s$.
(3) $\|P_h f\|_s \le \|f\|_s$ for any function $f \in \mathcal{U}_s$.
(4) $\|P_h f - f\| \le Kh \|f\|_s$.
(5) $\|P_h f\|_{L^1} \le \|f\|_{L^1} + Eh \|f\|_s$
(6) $\|P_h f + i(f - P_h f)\|_{L^1} \le \|f\|_{L^1} + Eh \|f\|_s$.

**Remark 2.8.** In general, we could relax Items 4, 5 6, substituting $h$ by $h^\alpha$, with $\alpha > 0$ in the whole paper, or more generally substitute $K \cdot h$ and $E \cdot h$ by functions $K(h)$ and $E(h)$ that go to 0 fast enough as $h$ goes to 0; this is not needed for the projections and functional spaces we study in this paper, but most of the theory adapts to these more general conditions with few differences.

**Remark 2.9.** Condition 6 in Definition 2.7 is used to control the error when the discretization of the operator is not $i$-preserving (see Definition 2.14 and Remark 2.15).

**Definition 2.10.** We will call the finite dimensional space $\mathcal{U}_h := P_h(\mathcal{U}_s)$ the **approximating space**.

The strong norm, the weak norm and the $L^1$ norm induce norms on $\mathcal{U}_h$, that will use the same notation.

**Assumptions on the approximating space 2.11.** We assume that the norms on $\mathcal{U}_h$ satisfy the following inequality; there exist $M$ and $\alpha$ such that for each $f \in \mathcal{U}_h$

$$\|f\|_s \le \frac{1}{h^\alpha} M \|f\|. \tag{7}$$

**Remark 2.12.** Such an inequality is usually false on $\mathcal{U}_s$, but the approximating space $\mathcal{U}_h$ is finite dimensional. If we let $\|f\|_{ns} = \|f\|_s + \|f\|$, this is a norm on $\mathcal{U}_h$ and there exists constants $\Lambda(h), \Theta(h)$, depending on $h$, such that

$$\Lambda(h) \|f\| \le \|f\|_{ns} \le \Omega(h) \|f\|;$$

remark that as $h = 1/n$ goes to 0, $\Omega(h)$ may go to infinity.

The rate at which $\Omega(h)$ goes to infinity depends on the chosen norms and approximation schemes. For the schemes presented in the current paper $\Theta(h) = M/h$; this is not true in general, in other cases, as in Chebyshev and Fourier discretization $\Theta(h)$ may grow faster as shown by Markov–Bernstein inequalities [30]; as an example, suppose $\mathcal{U}_s = C^1([0,1])$, $\|f\|_s = \|f'\|_\infty$, $\|f\| = \|f\|_\infty$ and $P_n$ is the map that associates to $f$ its Chebyshev interpolant of degree $n$.

In this case, by $\|f - P_n f\| \le O(h) \|f\|_s$ but, by Markov–Bernstein

$$\|p'\|_\infty \le n^2 \|p\|_\infty,$$

for all $p$ polynomial of degree at most $n$, i.e., $\Theta(h) = 1/h^2$.

**Remark 2.13.** Another possible generalization is to allow

$$\|f\|_{L^1} \le \tilde{M} \|f\|,$$

instead of fixing $\tilde{M}$ to be equal to 1 as in Assumption 2.1. Again, this is not needed in our current paper, but our methods can be adapted to this case.

**Definition 2.14.** Given a compatible discretization $P_h$ we define the **discretized operator** to be

$$L_h : \mathcal{U}_h \to \mathcal{U}_h, \quad L_h := P_h L P_h,$$

and we define the $i$-**preserving discretized operator** to be

$$Q_h : \mathcal{U}_h \to \mathcal{U}_h, \quad Q_h f := L_h f + 1 \cdot (i(f) - i(L_h f)). \tag{8}$$

**Remark 2.15.** In the two explicit discretizations presented in this paper, $i$ is the integral with respect to the Lebesgue measure. The name $i$-preserving may be interpreted, in these discretizations, as a nickname for integral preserving.

**Remark 2.16.** Depending on the chosen compatible discretization, $L_h$ may preserve $i$. In this case $Q_h$ and $L_h$ are going to denote the same operator.

**Remark 2.17.** From Assumption 2.1 item (2), follows that:

$$\|Q_h f\|_s \le \|L_h f\|_s. \tag{9}$$

From the properties of a compatible discretization follows a straightforward result on the operators $L_h$ and $Q_h$.

**Corollary 2.18.** *Let $P_h$ be a compatible discretization, and suppose that $L$ satisfies a one step Lasota–Yorke inequality (2) with coefficients $A$ and $B$. Then, if $h$ is small enough, a one step Lasota–Yorke inequality holds for $L_h$ and $Q_h$: for all $f \in \mathcal{U}_s$*

$$\|Q_h f\|_s \le \|L_h f\|_s \le (A + EhB) \|f\|_s + B \|f\|_{L^1}. \tag{10}$$

*Moreover, for all $f_h \in \mathcal{U}_h$, we have a stronger one step Lasota–Yorke inequality, since $P_h f_h = f_h$:*

$$\|Q_h f_h\|_s \le \|L_h f_h\|_s \le A \|f_h\|_s + B \|f_h\|_{L^1}. \tag{11}$$

**Proof.** For $f \in \mathcal{U}_s$ and the properties of a compatible discretization we have that

$$\begin{aligned}\|L_h f\|_s &= \|P_h L P_h f\|_s \le \|L P_h f\|_s \le A \|P_h f\|_s + B \|P_h f\|_{L^1} \\ &\le A \|f\|_s + B(Eh \|f\|_s + \|f\|_{L^1});\end{aligned}$$

if $f_h \in \mathcal{U}_h$ we have that $P_h f_h = f_h$, from this follows:

$$\|L_h f_h\|_s = \|P_h L P_h f_h\|_s = \|P_h L f_h\|_s \le A \|f_h\|_s + B \|f_h\|_{L^1}. \quad \square$$

**Corollary 2.19.** *Applying repeatedly the Lasota–Yorke inequality of Corollary 2.18, we get for all $k \in \mathbb{N}$ and $f_h \in \mathcal{U}_h$*

$$\left\|Q_h^k f_h\right\|_s \le A^k \|f_h\|_s + \frac{B}{1-A} \max_i\left(\left\|Q_h^i\right\|_{L^1}\right) \|f_h\|_{L^1}.$$

*If $f \in \mathcal{U}_s$ we have that $Q_h f \in \mathcal{U}_h$, therefore*

$$\left\|Q_h^k f\right\|_s \le A^{k-1}(A + EhB) \|f\|_s + \frac{B}{1-A} \max_i\left(\left\|Q_h^i\right\|_{L^1}\right) \|f\|_{L^1}.$$

**Proof.** If $f_h \in \mathcal{U}_h$ then

$$\begin{aligned}\left\|Q_h^k f_h\right\|_s &\le A \left\|Q_h^{k-1} f_h\right\|_s + B \left\|Q_h^{k-1} f_h\right\|_{L^1} \\ &\le A(A \left\|Q_h^{k-2} f_h\right\|_s + B \left\|Q_h^{k-2} f_h\right\|_{L^1}) + B \left\|Q_h^{k-1} f_h\right\|_{L^1} \\ &\le \cdots \le A^k \|f_h\|_s + B \sum_{j=1}^{k-1} A^{k-1-j} \left\|Q_h^j f_h\right\|_{L^1} \\ &\le A^k \|f_h\|_s + \frac{B}{1-A} \max_i\left(\left\|Q_h^i\right\|_{L^1}\right) \|f_h\|_{L^1}.\end{aligned}$$

If $f \in \mathcal{U}_s$ we have that $Q_h f \in \mathcal{U}_h$, therefore

$$\begin{aligned}\left\|Q_h^k f\right\|_s &= \left\|Q_h^{k-1} Q_h f\right\|_s \\ &\le A^{k-1} \|Q_h f\|_s + B \sum_{j=1}^{k-2} A^{k-1-j} \left\|Q_h^{j+1} f\right\|_{L^1} \\ &\le A^{k-1}(A + EhB) \|f\|_s + B \sum_{j=1}^{k-1} A^{k-1-j} \left\|Q_h^{j+1} f\right\|_{L^1} \\ &\le A^{k-1}(A + EhB) \|f\|_s + \frac{B}{1-A} \max_i\left(\left\|Q_h^i\right\|_{L^1}\right) \|f\|_{L^1}. \quad \square\end{aligned}$$

**Remark 2.20.** Remark that if the Lasota–Yorke inequality (10) is satisfied for a discretization of size $n = 1/h$ then for all discretizations with $\tilde{n} > n$ we have that $1/\tilde{n} = \tilde{h} < h$ and so inequality (10) is satisfied for all finer discretizations and for the original operator $L$. This permits us to prove, in a similar fashion as Corollary 2.19 that we have a uniform iterated Lasota–Yorke inequality. This is the main hypothesis we need to satisfy so that the spectral stability results of [31] holds.

## 3. Fixed point error estimation

In this section, we describe an explicit strategy to derive certified approximations of the fixed point $u$ of $L$ by approximating it with an element $u_h$ of the approximating space.

The following theorems give a slightly improved version of [17, Theorem 3.1] in which we allow for an inexactly computed eigenvector $u_h$ and we take more care about the multiplicative factors $\|P_h\|$.

**Definition 3.1.** Let us consider the generalized "zero average" spaces

$$\mathcal{V}_w^0 := \{v \in \mathcal{V}_w \mid i(v) = 0\},$$

$$\mathcal{V}_s^0 := \{v \in \mathcal{V}_s \mid i(v) = 0\}.$$

When dealing with the discretized operator $Q_h$, we denote by

$$\mathcal{V}_h^0 := \mathcal{V}_s^0 \cap \mathcal{V}_h.$$

**Remark 3.2.** When restricted to $\mathcal{V}_h$, the strong and the weak norm are equivalent, therefore $\mathcal{V}_h^0$ can be equivalently defined as

$$\mathcal{V}_h^0 := \mathcal{V}_w^0 \cap \mathcal{V}_h.$$

**Remark 3.3.** In Examples 2.2, 2.3, the space $\mathcal{V}_w^0$ is the space of average 0 functions, i.e.,

$$\mathcal{V}_w^0 = \mathcal{V}_w \cap \{f \in L^1([0,1]) \mid \int f\,dm = 0\}.$$

Then

$$\mathcal{V}_h^0 = \{u_h \in \mathcal{V}_h \mid h\sum_i u_i\},$$

where $u_i$ are the coordinates of $u_h$.

**Theorem 3.4.** *In the framework of the assumptions on the spaces, operators and discretizations stated in Section 2, Let $L$ be an operator with a fixed point $u$, normalized in a way that $i(u) = 1$, let $Q_h$ be an $i$-preserving discretized operator, and let $u_h \in \mathcal{V}_h$ be any vector such that $\|Q_h u_h - u_h\| \le \varepsilon$, normalized so that $i(u) = i(u_h) = 1$. Let $C_k$, for each $k \in \mathbb{N}$, be a constant such that*

$$\left\|Q_h^k|_{\mathcal{V}_h^0}\right\| \le C_k, \quad k \in \mathbb{N} \tag{12}$$

*and suppose that $\sum_{k=0}^{\infty} C_k < \infty$. Then,*

$$\|u - u_h\| \le \left(\sum_{k=0}^{\infty} C_k\right)(2Kh(1 + \|L\|)\|u\|_s + \varepsilon). \tag{13}$$

Before the proof of the theorem we need to perform some technical estimates.

**Lemma 3.5.** *Let $L : \mathcal{V}_s \to \mathcal{V}_s$ be an $i$-preserving operator, $L_h$ and $Q_h f$ discretized operators as in Definition 2.14 obtained by a compatible discretization as in Definition 2.7. Then,*

$$\|Q_h f - Lf\| \le 2Kh\left(\|L\|\,\|f\|_s + \|Lf\|_s\right).$$

**Proof.** The inequality follows by combining

$$\begin{aligned}\|Q_h f - Lf\| &= \|L_h f - Lf\| + \|1 \cdot (i(Lf) - i(L_h f))\| \\ &\le 2\|L_h f - Lf\|\end{aligned}$$

where we used 2.1, item 1 and 3, and

$$\begin{aligned}\|L_h f - Lf\| &\le \|P_h L(P_h - I)f\| + \|(P_h - I)Lf\| \\ &\le \|L\|\,Kh\,\|f\|_s + Kh\,\|Lf\|_s. \quad \square\end{aligned}$$

**Corollary 3.6.** *If $u \in \mathcal{V}_s$ is a fixed point of the operator $L$, then*

$$\|Q_h u - u\| \le 2Kh(\|L\| + 1)\|u\|_s.$$

**Proof.** By the previous lemma

$$\|Q_h u - u\| = \|Q_h u - Lu\| \le 2Kh\left(\|L\|\,\|u\|_s + \|Lu\|_s\right),$$

observing that $\|Lu\|_s = \|u\|_s$ we have the thesis. $\square$

**Proof of Theorem 3.4.** Let $v = u - u_h \in \mathcal{V}_h^0$. Note that from Corollary 3.6 we get

$$\|Q_h v - v\| \le \|Q_h u - u\| + \|Q_h u_h - u_h\| \le 2Kh(1 + \|L\|)\|u\|_s + \varepsilon.$$

By the triangle inequality,

$$\begin{aligned}\|v\| \le \|Q_h v - v\| + \left\|Q_h^2 v - Q_h v\right\| + \left\|Q_h^3 v - Q_h^2 v\right\| + \cdots + \left\|Q_h^m v - Q_h^{m-1} v\right\| \\ + \left\|Q_h^m v\right\|,\end{aligned}$$

and since $\lim_{m\to\infty}\left\|Q_h^m v\right\| \le \lim_{m\to\infty} C_m\|v\| = 0$ we can take the limit obtaining

$$\|v\| \le \sum_{k=0}^{\infty}\left\|Q_h^k(Q_h v - v)\right\| \le \left(\sum_{k=0}^{\infty} C_k\right)\|Q_h v - v\|. \tag{14}$$

Combining these inequalities we get:

$$\|v\| \le \left(\sum_{k=0}^{\infty} C_k\right)(2Kh(1 + \|L\|)\|u\|_s + \varepsilon). \quad \square$$

While Theorem 3.4 requires an infinite sum, the following lemma shows that it is sufficient to find a value $m$ with $C_m < 1$ to prove the convergence of the series.

**Lemma 3.7.** *Let $Q_h$ be an $i$-preserving discretized operator, and $C_k$ be constants such that $\left\|Q_h^k|_{\mathcal{V}_h^0}\right\| \le C_k$ for each $k = 0, 1, 2, \ldots$. Suppose that $C_m < 1$ for some positive integer $m$. Then,*

1. $\sum_{k=0}^{\infty} C_k \le \frac{1}{1-C_m}(C_0 + C_1 + \cdots + C_{m-1}) < \infty$;
2. *there are real constants $C > 0, \lambda_2 \in (0,1)$ such that $\left\|Q_h^k|_{\mathcal{V}_h^0}\right\| \le C\lambda_2^k$ for each $k$.*

**Proof.** Let $k \in \mathbb{N}$, and use Euclidean division with remainder to write $k = qm + r$. In particular, we have $r \in \{0, 1, \ldots, m-1\}$ and $k < (q+1)m$.

Since $Q_h$ is $i$-preserving, $Q_h(\mathcal{V}_h^0) \subseteq \mathcal{V}_h^0$, hence we can write

$$\left\|Q_h^k|_{\mathcal{V}_h^0}\right\| \le \left\|(Q_h^m|_{\mathcal{V}_h^0})^q(Q_h^r|_{\mathcal{V}_h^0})\right\| \le C_m^q C_r.$$

Then the first estimate follows by summing over all possible $k$

$$\begin{aligned}\sum_{k=0}^{\infty} C_k &\le C_m^0(C_0 + C_1 + \cdots + C_{m-1}) + C_m^1(C_0 + C_1 + \cdots + C_{m-1}) \\ &\quad + C_m^2(C_0 + C_1 + \cdots + C_{m-1}) + \cdots \\ &\le (1 + C_m + C_m^2 + \cdots)(C_0 + C_1 + \cdots + C_{m-1}).\end{aligned}$$

The second estimate follows instead from noting that

$$C_m^q C_r \le (C_m)^{\frac{k}{m}-1}\max(C_0, C_1, \ldots, C_{m-1}),$$

and thus we can take

$$C = \frac{\max(C_0, C_1, \ldots, C_{m-1})}{C_m}, \quad \lambda_2 = (C_m)^{\frac{1}{m}}. \quad \square$$

The first estimate is tighter and is the one that we shall use in numerical computation; the second one is looser but it gives an explicit bound with a geometric series.

**Remark 3.8.** The sequence $\left\|Q_h^k|_{\mathcal{V}_h^0}\right\|$ is related to the speed of convergence to equilibrium of the system mentioned in note 3. Even if these norms are explicitly computable, since $Q_h^k$ is a finite rank operator and can be represented by a matrix, we are not going to compute an enclosure for the norm, but just an upper bound $C_k$, which is enough for our treatment and more practical to compute.

In the case of Markov Transfer operators, this sequence is also related to the convergence of equilibrium of the system, indeed if $\mu$ is invariant for the system and $\nu$ is another probability measure in the strong space we have that $\mu - \nu \in \mathcal{V}_s^0$ and hence the convergence to zero in the weak norm of $Q_h^k(\mu - \nu)$ can be estimated by the sequence $C_k$. Note also that $Q_h^k(\mu - \nu) = \mu - Q_h^k(\nu)$.

**Remark 3.9.** In Theorem 3.4 we have a summability condition on $C_k$. We remark that in the statement and in the proof of the theorem we could exchange the role of $L$ and $Q_h$. If we could prove that $\sum_k \left\| L^k|_{\mathcal{V}_s^0} \right\|$ is summable and find an estimate for each term, this would give us an *a-priori* bound on the approximation error, but in general this a difficult task already for simple maps, as one-dimensional piecewise expanding ones, in the case there is not a Markov partition.

The flexibility of our method lies in the fact that the bound in Theorem 3.4 uses an a-posteriori, computer-assisted estimate which is computed on a finite-dimensional operator $Q_h$: in some sense, we ask the computer to estimate the convergence to equilibrium of the system at a finite resolution. This task is possible even if the dynamics is quite complicated. Of course the complexity increases with the resolution, and to optimize this we have to find a suitable strategy. This is the theme of next section.

### 3.1. The approximation error can be made as small as wanted

Our error estimates are a-posteriori ones: one knows the quality of the approximation only after applying the algorithm. In this section we give an argument showing that if the spaces satisfy Assumptions 2.1, the discretization scheme satisfies Definition 2.7 and the operator satisfies 2.4 we can approximate the stationary density as well as wanted; the argument here mirrors the one in [17] but works under the more general assumptions of this paper.

Suppose $\mathcal{V}_s \subseteq \mathcal{V}_w$ are two vector spaces of Borel signed measures on a certain metric space $X$ endowed with two norms, the strong norm $\| \ \|_s$ on $\mathcal{V}_s$ and the weak norm $\| \ \|$ on $\mathcal{V}_w$, such that $\| \ \|_s \geq \| \ \|$ as before. Let $\overline{\delta} \geq 0$. Let $L_\delta$, $\delta \in [0, \overline{\delta})$ be a family of Markov operators acting on $\mathcal{V}_w$. Denote by $\mathcal{V}_s^0, \mathcal{V}_w^0$ the "zero average" spaces of $\mathcal{V}_s, \mathcal{V}_w$.

**Definition 3.10.** We say that $L : \mathcal{V}_s \to \mathcal{V}_s$ has exponential convergence to equilibrium if there are $\lambda < 0$ and $C \geq 0$ such that for each $n \geq 0$, $f \in \mathcal{V}_s^0$

$$\|L_0^n f\| \leq C e^{\lambda n} \|f\|_s.$$

**Theorem 3.11.** *Let $L_0$ be an linear operator acting on $\mathcal{V}_s, \mathcal{V}_w$, having exponential convergence to equilibrium, and let $L_h = P_h L_0 P_h$ where $P_h$ is a compatible discretization. Let $\bar{h}$ be small and suppose that for all $h \in [0, \bar{h})$*

1. *$L_h$ are Markov operators acting on $\mathcal{V}_w$ and $\mathcal{V}_s$,*
2. *$L_h$ satisfy (5) with constants uniform in $h$,*
3. *$L_h$ satisfy (3) with constant uniform in $h$.*

*Then we can apply Theorem 3.4, finding constant $C_k$ such that when $h$ and $\epsilon$ are small enough, $\|u - u_h\|$ in (13) is as small as wanted.*

Before the proof we need to recall a result which is classical in this setting, and is proved in [32] in the form we will use.

We say that $L_\delta$ is a uniform family of operators if:

**UF1** (Uniform Lasota–Yorke ineq.) There are constants $A, B, \lambda_1 \geq 0$ with $\lambda_1 < 1$ such that $\forall f \in B_s, \forall n \geq 1, \forall \delta \in [0, 1)$ and each operator satisfies a Lasota–Yorke inequality.

$$\|L_\delta^n f\|_s \leq A\lambda_1^n \|f\|_s + B\|f\|_w. \tag{15}$$

**UF2** Suppose that $L_\delta$ approximates $L_0$ when $\delta$ is small in the following sense: there is $C \in \mathbb{R}$ such that $\forall g \in B_s$:

$$\|(L_\delta - L_0)g\|_w \leq \delta C \|g\|_s. \tag{16}$$

**UF3** Suppose that $L_0$ has exponential convergence to equilibrium, with respect to the norms $\| \ \|_w$ and $\| \ \|_s$.

**UF4** (The weak norm is not expanded) There is $M$ such that $\forall \delta, n, g \in B_s \ \ \|L_\delta^n g\|_w \leq M\|g\|_w$.

The following result (see [32], Proposition 45 for the proof) shows that such a uniform family has a uniform rate of contraction of the space $\mathcal{V}_s^0$ and hence a uniform convergence to equilibrium and spectral gap.

**Theorem 3.12** (*Uniform $\mathcal{V}_s^0$ Contraction for the Uniform Family of Operators*). *Let us consider a one parameter family of operators $L_\delta$, $\delta \in [0, 1)$. Suppose that they satisfy UF1, ... UF4, then there are $\lambda_1 < 1$ and $A_2, \delta_2 \geq 0$ such that for each $\delta \leq \delta_2$ and $f \in V_s$*

$$\|L_\delta^k f\|_s \leq A_2 \lambda_1^k \|f\|_s. \tag{17}$$

**Proof of Theorem 3.11.** First we see that we can apply Theorem 3.12 to our family of operators $L_h$. The assumption $UF1$ and $UF4$ are verified due to (2) and (3). The assumption $UF2$ is provided by Lemma 3.5, while $UF3$ is supposed in the assumptions of Theorem 3.11. Applying Theorem 3.12 we get that uniformly on $h$ there are $\lambda_1 < 0, \overline{h}_2 \geq 0, C_1 \geq 0$ such that for each $h \in [0, \overline{h}_2), n \geq 0$, $f \in \mathcal{V}_s$,

$$\|L_h^n f\|_s \leq C_1 e^{\lambda_1 n} \|f\|_s.$$

By (7) we than have that when $f \in U_h^0$

$$\|L_h^n f\|_w \leq \|L_h^n f\|_s \leq h^{-1} C_1 e^{\lambda_1 n} M_1 \|f\|_w.$$

By this we see that a sufficient condition to get $C^n \leq \frac{1}{2}$ is

$$n \geq \lambda^{-1} log(h^{-1} C_1 M_1)$$

by Item 1) of Lemma 3.7 this leads to $\sum_{k=0}^{\infty} C_k \leq 2\ W \lambda^{-1} \log(h^{-1} C_1 M_1)$ and then by (13):

$$\left\|u - u_h\right\| \leq (2\ W \lambda^{-1} \log(h^{-1} C_1 M_1))(2Kh\,(1 + \|L\|)\,\|u\|_s + \varepsilon). \tag{18}$$

Which can be set as small as wanted when $h$ and $\varepsilon$ are small enough. □

## 4. Estimating the convergence to equilibrium with the coarse-fine strategy

This section presents the coarse-fine approach, i.e., a method to use bounds $C_k$ as in (12), estimating the convergence to equilibrium of $Q_h$, to produce analogous bounds $C_k^F$ on the convergence to equilibrium of a finer-resolution approximation $Q_{h_F}$ of $L$, with $h_F < h$. An important ingredient will be the Lasota–Yorke inequality, which is shared by all sufficiently fine compatible discretizations of $L$ (as proved in Corollary 2.18).

A statement of this kind will be given in Corollary 4.6. This corollary will be obtained as a consequence of several intermediate steps, obtaining estimates on the norm of $Q_h^m - Q_{h_F}^m$.

The first ingredient is an iterated version of the Lasota–Yorke inequality for a discretized operator (Corollary 2.18). The approach is somewhat similar to the one used in [19] to rigorously estimate decay of correlation.

**Theorem 4.1.** *Let $L$ be an operator that satisfies Assumption 2.4, $P_h$ be a compatible discretization. Then, for each $k \in \mathbb{N}$ we have the inequality*

$$\begin{bmatrix} \left\| Q_h^k f \right\|_s \\ \left\| Q_h^k f \right\|_{L^1} \end{bmatrix} \leq \left( \begin{bmatrix} 1 & 0 \\ Eh & 1 \end{bmatrix} \begin{bmatrix} A & B \\ 0 & 1 \end{bmatrix} \right)^k \begin{bmatrix} \|f\|_s \\ \|f\|_{L^1} \end{bmatrix}, \quad f \in \mathcal{V}_h,$$

*where $\leq$ is intended to be componentwise.*

**Proof.** Note that $P_h f = f$ since $f \in \mathcal{U}_h$. We have

$$\|Q_h f\|_{L^1} \le \|Lf\|_{L^1} + Eh\,\|Lf\|_s \le \|f\|_{L^1} + Eh\,\|Lf\|_s .$$

Hence

$$\begin{bmatrix} \|Q_h f\|_s \\ \|Q_h f\|_{L^1} \end{bmatrix} \le \begin{bmatrix} 1 & 0 \\ Eh & 1 \end{bmatrix} \begin{bmatrix} \|Lf\|_s \\ \|f\|_{L^1} \end{bmatrix} \le \begin{bmatrix} 1 & 0 \\ Eh & 1 \end{bmatrix} \begin{bmatrix} A & B \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \|f\|_s \\ \|f\|_{L^1} \end{bmatrix}.$$

The rest follows by induction. □

**Corollary 4.2.** *Let $M$ be as in (7). Then,*

$$\left\|Q_h^k f\right\|_s \le R_{k,h,1}\,\|f\|, \quad \left\|Q_h^k f\right\|_{L^1} \le R_{k,h,2}\,\|f\|, \quad f \in \mathcal{U}_h, \tag{19}$$

*where*

$$\begin{bmatrix} R_{k,h,1} \\ R_{k,h,2} \end{bmatrix} := \left( \begin{bmatrix} 1 & 0 \\ Eh & 1 \end{bmatrix} \begin{bmatrix} A & B \\ 0 & 1 \end{bmatrix} \right)^k \begin{bmatrix} \frac{1}{h^\alpha} M \\ 1 \end{bmatrix}. \tag{20}$$

**Corollary 4.3.** *Let $S_1, S_2$ be constants such that $\|f\| \le S_1 \|f\|_s + S_2 \|f\|_{L^1}$. Then,*

$$\left\|Q_h^k\right\| \le S_1 R_{k,h,1} + S_2 R_{k,h,2}. \tag{21}$$

**Remark 4.4.** If $E = 0$, as in the case of the Ulam projection (Section 6), these bounds reduce to

$$\begin{aligned} \mathrm{Var}((Q_h^U)^k f) &\le A^k \,\mathrm{Var}(f) + (1 + A + A^2 + \cdots + A^{k-1}) B\,\|f\|_{L^1} \\ &\le A^k \,\mathrm{Var}(f) + \frac{B}{1-A}\,\|f\|_{L^1}, \end{aligned}$$

which is a classical iterated form of the Lasota–Yorke inequality [17, 31].

For a general projection, instead, $E \neq 0$ and the matrix

$$\mathcal{A}_h = \begin{bmatrix} 1 & 0 \\ Eh & 1 \end{bmatrix} \begin{bmatrix} A & B \\ 0 & 1 \end{bmatrix}$$

has an eigenvalue strictly larger than 1, hence $R_{k,h,1}$ and $R_{k,h,2}$ diverge and $\left\|Q_h^k\right\|$ is not bounded uniformly in $k$. Nevertheless, $\mathcal{A}_h$ is an $O(h)$ perturbation of the power-bounded matrix $\mathcal{A}_0 = \left[\begin{smallmatrix} A & B \\ 0 & 1 \end{smallmatrix}\right]$, so these estimates can be shown to be useful when $k \ll 1/h$.

We can now prove a result that shows that discretizations of the same operator with different grid sizes are 'close' (in a suitable sense). Let us consider two discretizations of the same Perron operator $L$, with $n$ and $n_F$ elements respectively (and grid sizes $h = 1/n$, $h_F = 1/n_F$) respectively. Note that if $n_F$ is a multiple of $n$, then for both $P_h^U$ and $P_h^L$ the finer grid is a refinement of the coarse grid, and $P_h P_{h_F} = P_{h_F} P_h = P_h$.

**Theorem 4.5.** *Let $Q_h, Q_{h_F}$ be two (i-preserving) discretizations of the same Perron operator $L$, obtained with projections such that $P_h P_{h_F} = P_{h_F} P_h = P_h$. Then, for each $f \in \mathcal{U}_{h_F}^0$ we have*

$$\left\|(Q_h^m - Q_{h_F}^m) f\right\| \le 2Kh \sum_{k=0}^{m-1} C_{m-1-k} \left( \left\|Q_{h_F}\right\| \left\|Q_{h_F}^k f\right\|_s + \left\|Q_{h_F}^{k+1} f\right\|_s \right).$$

*where* $\left\|Q_h^k|_{\mathcal{U}_h^0}\right\| \le C_k$.

**Proof.** The key insight is noticing that $L_{h_F} = P_{h_F} Q_h P_{h_F}$, so we can regard $Q_h$ as a further discretization of the operator $Q_{h_F}$, rather than a discretization of $L$. In particular, we can apply Lemma 3.5 with $Q_{h_F}$ in place of $L$. The rest follows once again from a telescopic sum argument.

$$\begin{aligned} \left\|(Q_h^m - Q_{h_F}^m) f\right\| &\le \sum_{k=0}^{m-1} \left\| Q_h^{m-1-k} (Q_h - Q_{h_F}) Q_{h_F}^k f \right\| \\ &\le \sum_{k=0}^{m-1} C_{m-1-k} 2Kh \left( \left\|Q_{h_F}\right\| \left\|Q_{h_F}^k f\right\|_s + \left\|Q_{h_F}^{k+1} f\right\|_s \right). \end{aligned}$$ □

**Corollary 4.6.** *We have*

$$\left\|Q_{h_F}^m|_{\mathcal{U}_{h_F}^0}\right\| \le C_m + 2Kh \sum_{k=0}^{m-1} C_{m-1-k} \left( \left\|Q_{h_F}\right\| R_{k,h_F,1} + R_{k+1,h_F,1} \right). \tag{22}$$

**Proof.** From Theorem 4.5 we have that for all $f \in \mathcal{U}_{h_F}^0$

$$\begin{aligned} \left\|Q_{h_F}^m f\right\| &\le \left\|Q_h^m f\right\| + \left\|(Q_h^m - Q_{h_F}^m) f\right\| \\ &\le C_m \|f\| + \sum_{k=0}^{m-1} C_{m-1-k} 2Kh \left( \left\|Q_{h_F}\right\| \left\|Q_{h_F}^k f\right\|_s + \left\|Q_{h_F}^{k+1} f\right\|_s \right) \end{aligned}$$

Observe that by Corollary 4.2 we have that

$$\left\|Q_{h_F}^k f\right\|_s \le R_{k,h_F,1}\,\|f\|,$$

therefore

$$\left\|Q_{h_F}^m f\right\| \le C_m \|f\| + \sum_{k=0}^{m-1} C_{m-1-k} 2Kh \left( \left\|Q_{h_F}\right\| R_{k,h_F,1} + R_{k+1,h_F,1} \right) \|f\|.$$ □

This estimate requires only the explicit computation of $\left\|Q_{h_F}\right\|$ and of the norms $C_k$ computed on a matrix of size $n < n_F$. Hence its computational cost is $O(n^2\ m + n_F)$, which can be much smaller than $O(n_F^2\ m)$.

**Remark 4.7.** When used alone, this process to derive coefficients $C_m^F$ on a finer grid from coefficients $C_m$ on a coarser grid never gives a practical advantage when used in (13). Indeed, ignoring some moderate factors and summands, we are replacing the estimate

$$\|u - u_h\| \sim h \sum_{k=0}^{\infty} C_k$$

from Theorem 3.4 with

$$\begin{aligned} \left\|u - u_{h_F}\right\| &\sim h_F \sum_{m=0}^{\infty} C_m^F \sim h_F \sum_{m=0}^{\infty} h \sum_{k=0}^{m-1} C_{m-1-k} \frac{1}{h_F} A^k \\ &= h \sum_{m=0}^{\infty} \sum_{j=0}^{m-1} C_j A^{m-1-j} = h \sum_{j=0}^{\infty} \sum_{m=j+1}^{\infty} C_j A^{m-1-j} \\ &= \frac{h}{1-A} \sum_{j=0}^{\infty} C_j, \end{aligned}$$

from Theorem 4.5; and this estimate is worse by a factor $\frac{1}{1-A}$. This rough computation suggests that one is always better off using the bound in Theorem 3.4 on $Q_h$ directly, forgoing $Q_{h_F}$ entirely.

However, another key ingredient is that we have other sources of *a priori* bounds on $C_k^F$ which are more effective for small $k$ and improve this estimate significantly. These different bounds are described in detail in Section 8.5.

## 5. Applying the general strategy to the transfer operators of non-singular maps

The main application of the abstract approximation scheme we present is the approximation of invariant densities for expanding and piecewise expanding dynamical systems on the unit interval $[0, 1]$.

Let $T$ be a measurable map $T : [0,1] \to [0,1]$, we say $T$ is **nonsingular** if $m(T^{-1}(A)) = 0$ if and only if $m(A)$ is equal to 0 for all measurable subsets $A$.

Given a measurable map, the action of the dynamical system extends to the space of probability measures through the push-forward operator associated to the map $T$, usually denoted as $T^*$, which associates to a probability measure $\mu$ the unique measure $T^*\mu$ such that $(T^*\mu)(A) = \mu(T^{-1}A)$ for all measurable set $A$. If $T$ is nonsingular, the space of absolutely continuous measures is preserved by $T^*$; this induces an operator $L : L^1[0,1] \to L^1[0,1]$ on the space of densities, called the **Perron–Frobenius operator** associated to the dynamical

system. It is well known that $L$ in this case is a weak contraction in $L^1$; for each $f \in L^1[0,1]$,

$$\|Lf\|_{L^1} \leq \|f\|_{L^1}.$$

In the case where the map is piecewise expanding we have that the associated Perron–Frobenius operator satisfies a Lasota–Yorke inequality. The following is a classical result, see [33] or [17][Theorem 5.2] for a proof.

**Lemma 5.1** (Var $-L^1$ *Lasota–Yorke Inequality*). *Let $T : [0,1] \to [0,1]$ and suppose there exists a finite partition $\{P_k\}_{k=1}^b$ of $[0,1]$ such that*

(1) $T_k = T|_{P_k}$ *is* $C^2$,
(2) $|T'(x)| > 2$ *for all* $x \in [0,1]$
(3) *the distortion* $|T''(x)/T'(x)^2|$ *is uniformly bounded by a constant* $D$,

*then (2) is satisfied with*

$$A = \sup_x \left|\frac{2}{T'(x)}\right| \quad B = \sup_k \frac{2}{|P_k|} + D, \tag{23}$$

*Maintaining hypothesis (1) and (3), relaxing hypothesis (2) to $|T'(x)| > 1$ for all $x \in [0,1]$ and with the addition that for all $k$ $f(P_k) = [0,1]$, then (2) is satisfied with*

$$A = \sup_x \left|\frac{1}{T'(x)}\right| \quad B = D. \tag{24}$$

In this context it is also well known (see e.g. [34]) that the transfer operator associated to a piecewise expanding map $T$, provided that $T$ is topologically mixing has a unique invariant probability density having bounded variation.

### 5.1. Recalling the needed constants

In the following we will use the basic facts recalled above for the approximation of invariant densities of examples of piecewise expanding maps. We will do this following our general strategy, for different discretizations and using different spaces. We recall that to apply our approximation strategy we have to provide the following bounds:

- the coefficients $A, B$ of a Lasota–Yorke inequality

$$\|Lf\|_s \leq A\|f\|_s + B\|f\|_{L^1}, \quad A < 1,$$

- the constant of the discretization error $K$,
- the "$i$-injection" constant $E$,
- the discretized "strong-weak" constant $M$,
- the "weak-strong+auxiliary" constants $S_1$ and $S_2$,
- a bound on $\|L\|$.

In the next sections we will compute all these constants for the Ulam approximation and for the piecewise linear approximation studied in [17] showing how the application of the coarse-fine strategy brings a substantial improvement in the computing speed and in the precision.

## 6. The Ulam projection

The first projection that we consider is the so-called Ulam projection on the torus. Subdivide $[0,1)$ into $n$ intervals $I_j = [(j-1)h, jh)$, $j = 1,\dots,n$, with the same width $h = 1/n$, and define

$$(P_h^U f)(x) = \frac{1}{h}\int_{I_j} f(y)dy, \quad \text{if } x \in I_j, \quad j = 1,\dots,n.$$

i.e., $P_h^U f$ is the piecewise constant function that is equal on each interval $I_j$ to the integral average of $f$ on $I_j$. Its image $\mathcal{U}_h$ is the space of piecewise constant functions on this grid. A natural basis for $\mathcal{U}_h$ is the one composed of the characteristic functions of the intervals $I_1, I_2, \dots, I_n$. In this basis, the coordinates of a function $f \in \mathcal{U}_h$ are $f_j = f((j-1)h)$ for $j = 1,2,\dots,n$, and

$$\|f\|_{L^1} = \frac{1}{n}\sum_{j=1}^n |f_i| = \frac{1}{n}\left\|\begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix}\right\|_{\ell^1}. \tag{25}$$

Moreover, the matrix associated to $L_h^U$ has elements

$$(L_h^U)_{ij} = \frac{\left|T^{-1}(I_i) \cap I_j\right|}{\left|I_j\right|}. \tag{26}$$

This discretization admits a simple interpretation, first suggested by Ulam in [35, pag.73-75][4]: $(L_h^U)_{ij}$ is the probability that a random point in $I_j$ (under the scaled Lebesgue measure) is mapped by $T$ into the interval $I_i$. Hence $L_h^U$ is the transition matrix of a Markov chain which approximates (in a suitable sense) the dynamic of the map $T$.

**Remark 6.1.** When discretizing the transfer operator of a piecewise expanding map, the matrix $L_h^U$ we obtain, with elements in (26) is sparse. Indeed, we can decompose

$$(L_h^U)_{ij} = \sum_{i=1}^k \frac{\left|T_k^{-1}(I_i) \cap I_j\right|}{\left|I_j\right|}, \tag{27}$$

and by Lagrange's theorem,

$$\left|T_k^{-1}(I_i)\right| \leq \frac{h}{\inf|T'|} < \frac{h}{2},$$

hence $L_h^U$ has at most 2 m nonzero elements in each row.

In this section we will find all the needed constants for the Ulam projection; in the Ulam case, we use the following norms.

**Norms for the Ulam discretization 6.2.** The norms involved in the Ulam approximation scheme are

- the strong seminorm is $\|.\|_s := \mathrm{Var}(.)$,
- the weak norm is $\|.\| := \|.\|_{L^1}$,
- $i(f) = \int f dm$, where $m$ is the Lebesgue measure on $[0,1]$.

The function $i(f)$ is represented in the above basis by the row vector $i^* = \frac{1}{n}[1,1,\dots,1]$.

### 6.1. Establishing the necessary bounds

In this subsection we estimate the necessary constants for our approximation procedure. Most of the estimates are trivial or well known, and are proved for a matter of completeness.

**Lemma 6.3.** *Let $P_h^U$ be the Ulam discretization on $n$-elements. Then:*

(1) $\|P_h f - f\| \leq \frac{h}{2}\|f\|_s$, *therefore* $K = 1/2$,
(2) $\|P_h f\|_{L^1} = \|f\|_{L^1}$ *and* $\int_X (f - P_h f)dx = 0$, *therefore* $E = 0$,
(3) *if* $f_h \in \mathcal{U}_h$ *we have that* $\|f_h\|_s \leq 2\frac{\|f_h\|}{h}$, *therefore* $M = 2$,
(4) $\|f\| = \|f\|_{L^1}$ *therefore* $S_1 = 0, S_2 = 1$.

**Proof.** We refer to [1] for a proof of (1). Since $P_h$ is a positive operator, we have that $\|P_h f\|_{L^1} \leq \|P_h 1\|_{L^1} = \|1\|_{L^1} = 1$; moreover

$$\int_0^1 f - P_h f dx = \sum_{i=0}^n \int_{I_i}\left(f(x) - \frac{1}{h}\int_{I_i} f(y)dy \cdot \chi_{I_i}(x)\right)dx = 0,$$

[4] For the interested reader, it can be found at https://archive.org/details/collectionofmath0000ulam/page/73

therefore $E = 0$, item (2).

If $f_h \in \mathcal{U}_h$, we have that $f_h = \sum_{i=0}^{n} f_i \chi_{I_i}$ and

$$\mathrm{Var}\, f_h = \sum_{i=0}^{n-1} |f_{i+1} - f_i| \le 2 \sum_{i=0}^{m} |f_i| \le 2h \|f\|_{L^1},$$

therefore $M = 2$, item (3).

Item (4) follows from the fact that $\|f\| = \|f\|_{L^1}$. □

### 6.2. Spectral picture for $L_h^U$

Note that the Ulam projection is, by its definition, $i$-preserving, i.e., $i(P_h^U f) = i(f)$. In particular, this implies that $L_h^U = Q_h^U$.

We have $i^* Q_h^U = i^*$, hence $Q_h^U$ is a stochastic matrix, which is also irreducible and a-periodic by the mixing hypothesis. By the Perron–Frobenius theorem, its largest eigenvalue is $\lambda_1 = 1$, and the associated eigenvector $u_h$ has strictly positive entries; moreover, the second largest eigenvalue is $\lambda_2 < 1$. In particular, $1 = \left\|Q_h^U\right\|_{L^1} = \left\|(Q_h^U)^k\right\|_{L^1}$ for all $k \in \mathbb{N}$, while $\left\|(Q_h^U)^k|_{\mathcal{U}_h^0}\right\| = O(\lambda_2^k)$, where

$$\mathcal{U}_h^0 := \{g \in \mathcal{U}_h : i^* g = 0\}. \tag{28}$$

## 7. The piecewise linear projection

In this section we will find all the needed constants for the piecewise linear projection on $[0, 1)$.

The piecewise linear projection is defined as follows. Divide $[0, 1]$ into $n$ equal intervals, delimited by equispaced nodes $\{a_i = \frac{i}{n}\}_{i=0}^{n}$. Let $\phi_i(x)$ be the piecewise linear function

$$\phi_i(x) = \begin{cases} n(x - a_{i-1}) & x \in [a_{i-1}, a_i] \\ -n(x - a_i) & x \in [a_i, a_{i+1}] \\ 0 & x \in [a_{i-1}, a_{i+1}]^c \end{cases}$$

and define

$$(P_h^L f)(x) = \sum_{i=0}^{n} f(a_i)\phi_i(x)$$

i.e., $P_h^L f$ is the piecewise linear function that which interpolates $f(a_i)$ on the given nodes. The image $\mathcal{U}_h$ of $P_h^L$ is the space of piecewise linear functions on this grid. A natural basis for this space is $(\phi_i(x))_{i=1,\dots,n}$. Given a function $f \in \mathcal{U}_h$, its coordinates in this basis are $f_j = f(a_{j-1})$ for $j = 1, 2, \dots, n$, and

$$\|f\|_{L^\infty} = \max_{i=1,\dots,n} |f_i| = \left\| \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} \right\|_{\ell^\infty}.$$

The matrix associated to $L_h$ has elements

$$(L_h)_{ij} = \sum_{x \in T^{-1}(a_i)} \frac{\phi_j(x)}{|T'(x)|}. \tag{29}$$

**Norms for the piecewise linear discretization 7.1.** The norms involved in the piecewise linear approximation scheme are

- the strong norm $\|.\|_s := \|.\|_{Lip}$,
- the weak norm $\|.\| := \|.\|_\infty$,
- $i(f) = \int f dm$, where $m$ is the Lebesgue measure on $[0, 1]$.

The function $i(f)$ is represented in the above basis by the row vector $i^* = \frac{1}{n}[1, 1, \dots, 1]$.

### 7.1. Expanding maps and the Lasota–Yorke inequality

In this case we need to prove that the operator $L$ preserves a stronger norm; this is proved in the next theorem.

**Theorem 7.2** (Lip $-L^1$ *Lasota–Yorke Inequality*). *Let $T$ be in $C^2(S^1)$, with $|T'(x)| > 1$ and $|T''/(T')^2| < D$. Then, an inequality (2) holds with*

$$A = \sup_x \frac{(2D + 1)}{|T'(x)|} \qquad B = D(D + 1)$$

**Proof.** Since $T \in C^2(S^1)$, $|T'(x)| > 1$ there exists (at least) one fixed point of $T$; we can label this fixed point as 0 and see $T$ as a map satisfying (24); in specific, there exists a partition $\{P_k\}_{k=1}^{b}$ such that $T(P_k) = [0, 1]$ for all $k$; we denote by $T_k := T|_{P_k}$. Please remark that $\lambda$ and $D$ are defined in the proof of (24).

$$\begin{aligned} |Lf(x) - Lf(y)| &= \left| \sum_k \frac{f(T_k^{-1}x)}{T_k'(T_k^{-1}x)} - \frac{f(T_k^{-1}y)}{T_k'(T_k^{-1}y)} \right| \\ &\le \left| \sum_k \frac{f(T_k^{-1}x) - f(T_k^{-1}y)}{T_k'(T_k^{-1}x)} \right| + \left| \sum_k \frac{f(T_k^{-1}y)}{T_k'(T_k^{-1}y)} - \frac{f(T_k^{-1}y)}{T_k'(T_k^{-1}x)} \right| \\ &\le \mathrm{Lip(f)}\lambda|x - y| \sum_k \frac{1}{T_k'(T_k^{-1}x)} + \left| \sum_k \frac{f(T_k^{-1}y)}{T_k'(T_k^{-1}y)} \left(1 - \frac{T_k'(T_k^{-1}y)}{T_k'(T_k^{-1}x)}\right) \right|. \end{aligned}$$

and

$$\begin{aligned} \left|1 - \frac{T_k'(T_k^{-1}y)}{T_k'(T_k^{-1}x)}\right| &= \left|\frac{T_k'(T_k^{-1}x) - T_k'(T_k^{-1}y)}{T_k'(T_k^{-1}x)}\right| \\ &\le \lambda \mathrm{Lip}(T_k')|T_k^{-1}x - T_k^{-1}y| \le \lambda^2 \mathrm{Lip}(T_k')|x - y|. \end{aligned}$$

Hence

$$\mathrm{Lip}(Lf) \le \lambda \|L1\|_\infty \mathrm{Lip}(f) + D\|Lf\|_\infty. \tag{30}$$

Now we use (24) to estimate

$$\begin{aligned} \|Lf\|_\infty &\le \mathrm{Var}(Lf) + \|Lf\|_{L^1} \le \lambda\, \mathrm{Var}(f) + D\|f\|_{L^1} \\ &\quad + \|f\|_{L^1} \le \lambda\, \mathrm{Lip}(f) + (D + 1)\|f\|_{L^1}, \end{aligned}$$

and, in particular, for the constant function 1, $\|L1\|_\infty \le D + 1$. Plugging these two bounds into (30) we get the thesis. □

**Remark 7.3.** Under the same hypotheses, the matrix $L_h$ with elements in (29) is sparse. Indeed, for each $x \in [0, 1)$ at most two of the functions $\phi_j(x)$ are nonzero, hence $L_h$ has at most 2 m nonzero elements in each row.

**Remark 7.4.** We need $(2D + 1)\lambda < 1$ for this to be a valid Lasota–Yorke inequality. If this property does not hold, then we can replace $T$ with one of its iterates $T^k$. Clearly $T^k$ has the same invariant measure as $T$; moreover, the values of $\lambda$ and $D$ are replaced by $\lambda^k$ and $D(1 + \lambda + \cdots + \lambda^{k-1}) < \frac{D}{1-\lambda}$. In particular, for sufficiently large $k$ one has $\lambda^k(2\frac{D}{1-\lambda} + 1) < 1$, hence this strategy works. Note, though, that $T^k$ has $kb$ monotonic branches instead of $b$, hence the associated matrix is less sparse and the whole method is more computationally expensive.

### 7.2. Establishing the necessary bounds

**Lemma 7.5.** *Let $P_h^L$ be the piecewise linear discretization on $n$-elements. Then:*

1. $\|P_h f - f\| \le \frac{h}{2}\|f\|_s$, *therefore* $K = 1/2$,
2. $\left\|P_h f - 1 \cdot \int_X (f - P_h f)dx\right\|_{L^1} \le \|f\|_{L^1} + \mathrm{Lip}(f)h/2$, *therefore* $E = 1/2$,
3. *if* $f_h \in \mathcal{U}_h$ *we have that* $\|f_h\|_s \le 2\frac{\|f\|}{h}$, *therefore* $M = 2$,
4. $\|f\| \le \|f\|_{L^1} + \|f\|_s$ *therefore* $S_1 = 1, S_2 = 1$,
5. $\|L\| \le D + 1$.

**Proof.** To prove item (1) we study $f|_{[a_i,a_{i+1}]}$ and suppose that $f(a_i) = 0, f(a_{i+1}) = 0$; in this case, $\|f|_{[a_i,a_{i+1}]}\|_\infty \le \mathrm{Lip}(f)\cdot(x-a_i)$ and $\|f|_{[a_i,a_{i+1}]}\|_\infty \le \mathrm{Lip}(f)\cdot(a_{i+1}-x)$. This means that the graph of $|f(x)|$ lies under the graphs of the linear functions $y_1(x) = \mathrm{Lip}(f)\cdot(a_{i+1}-x)$ and $y_2(x) = \mathrm{Lip}(f)\cdot(x-a_i)$ which intersect in $(a_{i+1}-a_i)/2$, therefore

$$\left\|f|_{[a_i,a_{i+1}]}\right\|_{L^\infty} \le \mathrm{Lip}(f)\frac{h}{2}$$

and, since it is true for all $i$ we have

$$\left\|f - P_h^L f\right\|_{L^\infty} \le \mathrm{Lip}(f)\frac{h}{2}.$$

This implies that

$$\left|\int_0^1 f - P_h^L f\,dx\right| \le \int_0^1 |f - P_h^L f|dx \le \mathrm{Lip}(f)\frac{h}{2}.$$

From this follows

$$\left\|P_h f - 1\cdot\int_X (f-P_hf)dx\right\|_{L^1} \le \|P_hf\|_{L^1} + \left\|1\cdot\int_X (f-P_hf)dx\right\|_{L^1}$$
$$\le \|f\|_{L^1} + \mathrm{Lip}(f)\frac{h}{2}.$$

and that

$$\left\|P_h^L f\right\|_{L^1} = \int_0^1 |P_h^L f(x)| - |f(x)| + |f(x)|dx$$
$$\le \int_0^1 |f(x)| + \|P_h^L f(x)| - |f(x)\|dx$$
$$\le \|f\|_{L^1} + \int_0^1 |P_h^L f(x) - f(x)|dx \le \|f\|_{L^1} + \mathrm{Lip}(f)\frac{h}{2},$$

therefore $E = 1/2$.

We prove now Item (3). If $f_h \in \mathcal{U}_h$ we have that $f_h(x) = \sum_{i=0}^n a_i\cdot\phi_i(x)$, where the $\phi(x)$ are piecewise linear. Therefore

$$\mathrm{Lip}(f) = \max_i \frac{|a_{i+1}-a_i|}{h} \le 2\frac{\max_i|a_i|}{h} = 2\frac{\|f\|}{h}$$

Item (4) follows from the fact that for $x \ne \tilde{x}$

$$|f(\tilde{x})| = |f(x) + \frac{f(\tilde{x})-f(x)}{\tilde{x}-x}(\tilde{x}-x)| \le |f(x)| + \mathrm{Lip}(f)\cdot|x-\tilde{x}|.$$

Suppose now that $|f|$ attains its maximum in $\tilde{x}$, and integrate:

$$\int_0^1 \|f\|_\infty dx \le \int_0^1 |f(x)| + \mathrm{Lip}(f)\cdot|x-\tilde{x}|dx \le \|f\|_{L^1} + \mathrm{Lip}(f).$$

Item (5) follows from the fact that $\|L\|_{L^\infty} = \|L1\|_{L^\infty}$, since $L$ is a positive operator, and $\|L1\|_{L^\infty} \le D+1$ as in the proof of Theorem 7.2. □

### 7.3. Spectral picture for $L_h^L$ and $Q_h^L$

Compared with the Ulam projection, the spectral picture is more blurry for $L_h^L$ and $Q_h^L$. The matrix $L_h^L$ is still a non-negative matrix, but since $P_h^L$ is not $i$-preserving its first eigenvector $\lambda_1$ is not in general equal to 1.

The row vector $i^*$ is a left eigenvector of $Q_h^L$ with eigenvalue equal to 1, however, $Q_h^L$ is not a non-negative matrix, so we do not have all the results implied by the Perron–Frobenius theory of Markov chains; in particular, $\left\|Q_h^L\right\|_{L^1} > 1$ in general (and our experimental results suggest that even the limit $\lim_{h\to 0}\left\|Q_h^L\right\|_{L^1} = 1$ does *not* hold).

Nevertheless, the results by Keller and Liverani (see Corollary 2.19) ensure that $\lambda_2$ is smaller than 1 for sufficiently small values of $h$.

## 8. Practical computation

In this section we present the results that permit us to efficiently compute the objects and the constants involved in our treatment. There are three main points in the algorithm:

- Computing an interval matrix $\mathbf{L} \ni L_h$ that encloses $L_h$;
- Computing a fixed point vector for $\mathbf{L}$;
- Computing norm estimates $C_k \ge \left\|Q_h^k|_{\mathcal{U}_h^0}\right\|$ for $k = 0,1,2,\dots,m$, and reaching a $m$ such that $C_m < 1$.

We will address them one by one in the next sections.

### 8.1. Assembling the sparse matrices

Recall that the projection $P_h$ permits us to build a discretization of $L$, the projected operator $L_h := P_hLP_h$.

Then, $L_h : \mathcal{U}_h \to \mathcal{U}_h$ can be represented by a square matrix in a suitable basis of $\mathcal{U}_h$.

We describe here a strategy to compute the matrix associated to $L_h$ for the case of the Ulam and piecewise linear projections on the torus $[0,1)$. With some abuse of notation, we will denote with the same symbol both the operator (acting on functions on $[0,1]$) and the matrix that represents it. We assume that the dynamic $T$ is composed of $b$ continuous and monotonic branches $T_1, T_2, \dots, T_b$, whose domains form a partition of $[0,1)$.

The partition underlying the projection (which is typically equispaced) can be described by an increasing sequence $\mathcal{Y} : 0 = y_0 < y_1 < y_2 < \cdots < y_{n-1} < y_n = 1$ that partitions $[0,1)$ of $T$ into $\bigcup_{j=i}^n I_j$, with $I_j = [y_{j-1}, y_j)$. We assume that the co-domain $[0,1)$ of $T : [0,1) \to [0,1)$ is partitioned according to this sequence $\mathcal{Y}$; then, its domain $[0,1)$ is decomposed into $nb$ intervals $T_k^{-1}(I_j)$, some of them possibly empty; their endpoints are an increasing sequence $\mathcal{X} : 0 = x_0 < x_1 < \cdots < x_N = 1$ that defines a partition of the domain $[0,1)$ of $T$. We say that this sequence $\mathcal{X}$ is the *pull-back* of the sequence $\mathcal{Y}$, and we denote it by $\mathcal{X} = T^{-1}(\mathcal{Y})$. An example is shown in Fig. 1. The endpoints $x_i$ are either preimages $T_k^{-1}(y_j)$ for some $k$ and $j$, or endpoints of the domain of each branch; clearly we have $N \le nb$, but some intervals may be missing if the map is not full-branch; for instance, in the example in Fig. 1 the interval $T^{-1}(I_1)$ is empty, and hence there are 7 intervals instead of $8 = 4\cdot 2$ in the partition $\mathcal{X}$.

Interval arithmetic methods such as the interval Newton method [36] can be used to compute tight inclusion intervals $\mathbf{x}_i$ for each element $x_i$ of the pull-back partition, given explicit formulas to compute each branch of the map $T_k$. Once the $\mathbf{x}_i$ are available, inclusions $\mathbf{L}_{ij}$ for the matrix elements in either (26) or (29) are easy to compute.

The computation of the $\mathbf{x}_i$ can be performed automatically; we sketch how the method works for the second branch $T_2$ of the dynamic in Fig. 1. One starts from the endpoints $(a,b)$ of $\mathrm{dom}(T_2)$. By checking how $T_2(a), T_2(b)$ compare with the elements of the sequence $\mathcal{Y}$, one can determine that $\mathbf{x}_3 = a$, $\mathbf{x}_7 = b$, and that three unknown values $x_4 = T_2^{-1}(y_1), x_5 = T_2^{-1}(y_2), x_6 = T_2^{-1}(y_3)$ need to be computed. We can use a bisection strategy to reduce the number of iterations needed in the interval Newton method, as follows. We first compute $\mathbf{x}_5$ by applying the interval Newton method to find a zero of $x \mapsto T_2(x) - y_2$, using the whole domain $\mathrm{hull}(\mathbf{x}_3, \mathbf{x}_7)$ as a starting interval. Once $\mathbf{x}_5$ has been computed, we obtain $\mathbf{x}_4$ by applying the interval Newton method to find a zero of $x \mapsto T_2(x) - y_1$ using the tighter interval $\mathrm{hull}(\mathbf{x}_3, \mathbf{x}_5)$ as a starting point instead of the whole domain, and similarly we use $\mathrm{hull}(\mathbf{x}_5, \mathbf{x}_7)$ as a starting interval in the interval Newton method to compute $\mathbf{x}_6$.

**Remark 8.1.** Since each branch of $T$ is expanding, the preimage problem is well-conditioned, and we expect to be able to compute enclosures with radius $\mathrm{rad}(\mathbf{x}_i)$ of the same order of magnitude as the machine precision used.

**Remark 8.2.** This description in terms of pull-backs of partitions has the additional benefit that pull-backs of composed maps are particularly easy to compute, since $(S\circ T)^{-1}(\mathcal{Y}) = T^{-1}(S^{-1})(\mathcal{Y})$.

An explicit algorithm to compute a sparse interval matrix $\mathbf{L} \ni L_h$ is sketched in Algorithm 1. It has complexity $O(nb)$, since the sets $S_\ell$ have dimension $O(1)$. The algorithm returns the sparse matrix in coordinate list format, i.e., a list $\mathcal{S}$ of triples $(i,j,\mathbf{c})$ such that $\mathbf{L}_{ij} = \sum_{(i,j,\mathbf{c})\in\mathcal{S}} \mathbf{c}$. Note that the list $\mathcal{S}$ will in general contain multiple entries with the same $i$ and $j$.

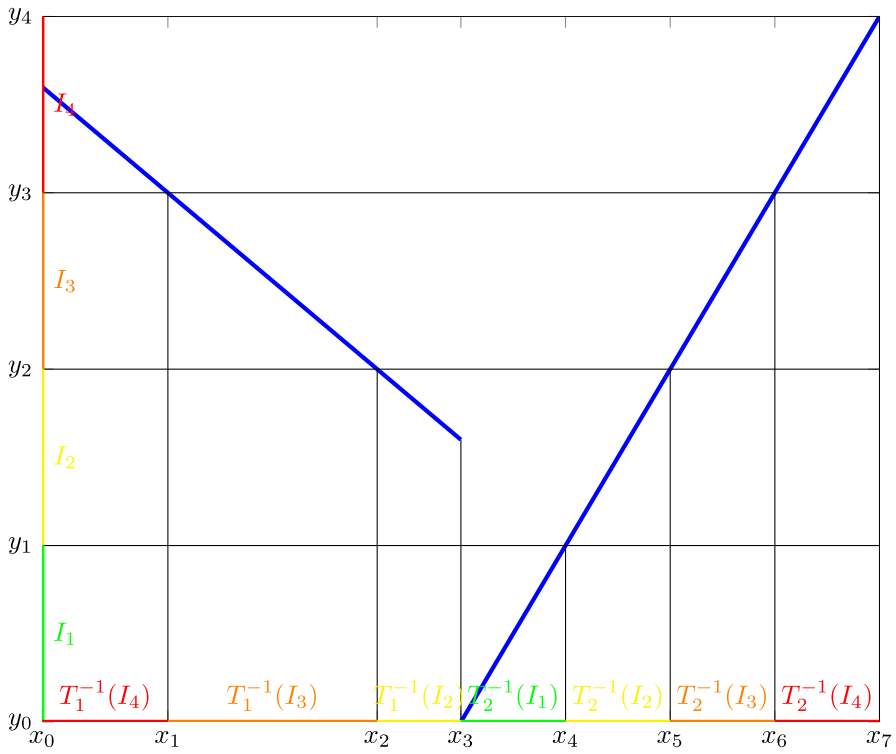


**Fig. 1.** An example of "pull-back": the pull-back along the map $T$ (drawn in blue) of the sequence $(y_j)$ on the $y$-axis is the sequence $(x_i)$ on the $x$-axis.

**Algorithm 1** Assembling a sparse interval matrix $\mathbf{L} \ni L_h$ (for the Ulam or piecewise linear discretization)

**function** ASSEMBLE_LH($T, n$)
**Require:** $T$, partition $\mathcal{Y}$ (typically an equispaced one)
**Ensure:** A list $S$ of triples $(i, j, \mathbf{c})$
  compute enclosures $(\mathbf{x}_\ell)_{\ell=1}^N$ for the pull-back $\mathcal{X} = T^{-1}(\mathcal{Y})$,
  **for** $\ell = 1, 2, \dots, N$ **do**
    determine the set $S_\ell = \{j : (\mathbf{x}_{\ell-1}, \mathbf{x}_\ell) \cap I_j \neq \emptyset\}$ or $S_\ell = \{j : \phi_j(\mathbf{x}_\ell) \neq 0\}$ via a binary search on $\mathcal{Y}$;
    **for** all $j \in S_\ell$ **do**
      push $(i, j, \mathbf{c})$ into $S$, where $\mathbf{c}$ is a summand of (27) or (29),
      and $i$ is the index such that $T((x_{\ell-1}, x_\ell)) \subseteq (y_{i-1}, y_i)$;
    **end for**
  **end for**
**end function**

### 8.2. Numerically approximating the fixed point

We compute numerically an approximate fixed point $\tilde{u}_h$ of the operator $Q_h$ by using the restarted Arnoldi method [37, Section 10.5] to return its eigenvector with eigenvalue (approximately) 1. While $L_h$ is a sparse matrix, $Q_h$ is not, in general. However, we can compute its (approximate) action on a vector $v$ using

$$Q_h v \approx \operatorname{mid}(\mathbf{L})v + ei^*(v - \operatorname{mid}(\mathbf{L})v).$$

For the Ulam discretization, $Q_h = L_h$ and we can drop the second summand.

In general, the computed eigenvector will not satisfy the equality $i^*\tilde{u}_h = 1$ exactly. The following corollary of Theorem 3.4 allows us to estimate the distance between the computed $\tilde{u}_h$ and the exact fixed point of the operator.

**Corollary 8.3.** *Under the hypothesis of Theorem 3.4 and Lemma 3.7, let $\tilde{u}_h$ be a vector such that $\|Q_h\tilde{u}_h - \tilde{u}_h\| \le \varepsilon_1$ and $|i^*\tilde{u}_h - 1| \le \varepsilon_2 < 1$. Then,*

$$\|u - \tilde{u}_h\| \le \frac{C_0 + C_1 + \cdots + C_{m-1}}{1 - C_m}(2Kh(1 + \|L\|)\|u\|_s + \frac{\varepsilon_1}{1 - \varepsilon_2}) + \frac{\varepsilon_2}{1 - \varepsilon_2}\|\tilde{u}_h\|.$$

**Proof.** Set $u_h = \tilde{u}_h / i^*\tilde{u}_h$. Then,

$$\|Q_h u_h - u_h\| = \frac{\|Q_h\tilde{u}_h - \tilde{u}_h\|}{i^*\tilde{u}_h} \le \frac{\varepsilon_1}{1 - \varepsilon_2},$$
$$\|u_h - \tilde{u}_h\| \le \left\|\frac{1 - i^*\tilde{u}_h}{i^*\tilde{u}_h}\tilde{u}_h\right\| \le \frac{\varepsilon_2}{1 - \varepsilon_2}\|\tilde{u}_h\|.$$

Combining these two bounds with (13) and the first point of Lemma 3.7 gives the desired result. □

### 8.3. Bounding norms of powers computationally

In this section, we describe a computational procedure to obtain rigorous bounds of the form $\left\|Q_h^k|_{U_h^0}\right\| \le C_k$ in practice on a computer. We start by recalling one important notation convention we stated in Notation 1.1

**Notation 8.4.** The symbol $\|f\|_{L^p}$ denotes the $L^p$ norm of a function (usually defined on $[0, 1]$), whereas the symbol $\|v\|_{\ell^p}$ denotes the $\ell^p$ norm of a vector $v \in \mathbb{R}^n$.

In the Ulam projection, since the 'continuous' norms $\|\cdot\|_{L^1}$ and the 'discrete' norm $\|\cdot\|_{\ell^1}$ differ only by a constant (see (25)), we have $\left\|Q_h^k|_{U_h^0}\right\|_{L^1} = \left\|Q_h^k|_{U_h^0}\right\|_{\ell^1}$, and similarly for $L^\infty$ and $\ell^\infty$ in the piecewise linear projection. Hence we can replace these norms with matrix norms for which there are classical formulas

$$\|M\|_{\ell^1} = \max_i \sum_j \left|M_{ij}\right|, \quad \|M\|_{\ell^\infty} = \max_j \sum_i \left|M_{ij}\right|. \tag{31}$$

However, even after reducing to a discrete setting, computing matrix norms restricted to a certain subspace $\mathcal{U}_h^0$ is not a textbook problem. The following bound allows one to solve it.

**Lemma 8.5.** *Let*

$$U = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ -1 & 0 & \cdots & 0 \\ 0 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 \end{bmatrix} = \begin{bmatrix} e^* \\ -I \end{bmatrix} \in \mathbb{R}^{n\times(n-1)},$$

*and* $\mathcal{U}_h^0 = \ker([1,1,\dots,1]^*) = \operatorname{Im} U$. *Then, for each* $M \in \mathbb{R}^{n\times n}$ *and each* $\ell^p$ *norm one has* $\left\| M|_{\mathcal{U}_h^0} \right\| \le \|MU\|$.

**Proof.** We have for each $z \in \mathbb{R}^{n-1}$

$$\|Uz\| = \left\| \begin{bmatrix} z_1 + z_2 + \cdots + z_{n-1} \\ -z_1 \\ -z_2 \\ \vdots \\ -z_{n-1} \end{bmatrix} \right\| \ge \|z\|.$$

Moreover,

$$\left\| M|_{\mathcal{U}_h^0} \right\| = \sup_{x\in\mathcal{U}_h^0\setminus\{0\}} \frac{\|Mx\|}{\|x\|} = \sup_{z\in\mathbb{R}^{n-1}\setminus\{0\}} \frac{\|MUz\|}{\|Uz\|} \le \sup_{z\in\mathbb{R}^{n-1}\setminus\{0\}} \frac{\|MUz\|}{\|z\|} = \|MU\|. \quad \square$$

**Remark 8.6.** Note that $\|Uz\|_{\ell^1} \le 2\|z\|_{\ell^1}$ and $\|Uz\|_{\ell^\infty} \le (n-1)\|z\|_{\ell^\infty}$, so this bound is off by at most a factor 2 in the $\ell^1$ norm and by at most a factor $n-1$ in the $\ell^\infty$ norm.

**Remark 8.7.** For a generic projection, an analogous procedure can be devised. Let $U \in \mathbb{R}^{n\times(n-1)}$ be a matrix whose columns are a basis of $\mathcal{U}_h^0 = \ker i^*$, and suppose that $\|Uz\| \ge \alpha\|z\|$. Then, by the same reasoning, we have that

$$\left\| M|_{\mathcal{U}_h^0} \right\| = \sup_{x\in\mathcal{U}_h^0\setminus\{0\}} \frac{\|Mx\|}{\|x\|} = \sup_{z\in\mathbb{R}^{n-1}\setminus\{0\}} \frac{\|MUz\|}{\|Uz\|} \le \sup_{z\in\mathbb{R}^{n-1}\setminus\{0\}} \frac{\|MUz\|}{\alpha\|z\|} = \frac{1}{\alpha}\|MU\|.$$

An estimate for $\alpha$ can be obtained automatically for any norm $\|.\|$ for which we know explicit constants $c, C$ such that

$$c\|z\|_{\ell^2} \le \|z\| \le C\|z\|_{\ell^2},$$

using a rigorous estimate for

$$\sigma_{\min} = \min(\|Uv\|_{\ell^2}/\|v\|_{\ell^2})$$

obtained from the SVD decomposition of $U$, using techniques to rigorously certify eigenvalues as in [38].

Therefore

$$\|Uz\| \ge c\|Uz\|_{\ell^2} \ge c\eta\|z\|_{\ell^2} \ge c\sigma_{\min}C\|z\|.$$

**Remark 8.8.** In the case of a more general weak norm $\|\cdot\|$, we can reduce the problem to the computation of the $\ell^1$ and $\ell^\infty$ norms of the operator. To do so, we need three estimates

$$\|v\| \le W_1\|v\|_{\ell^1} + W_2\|v\|_{\ell^\infty}$$

and

$$\|v\|_{\ell^1} \le \alpha_1\|v\|, \quad \|v\|_{\ell^\infty} \le \alpha_\infty\|v\|,$$

which imply

$$\|P\| \le \frac{W_1}{\alpha_1}\|P\|_{\ell^1} + \frac{W_2}{\alpha_2}\|P\|_{\ell^\infty}.$$

**Remark 8.9.** There is some linear algebra literature on fast estimation of matrix norms, for instance [39], but unfortunately we cannot use it here. Indeed, these estimators return only a guaranteed *lower* bound $C \le \|M\|$. Providing a lower bound is a simpler problem, since it is sufficient to show that $\|Mx\| \ge C$ for a suitable norm-1 vector $x$; giving a rigorous upper bound, instead, requires proving that $\|Mx\| \le C$ for *all* norm-1 vectors.

### 8.4. Handling machine arithmetic errors when bounding norms

In principle, one can obtain a rigorous estimate for $\left\| Q_h|_{\mathcal{U}_h^0} \right\|$ from the results in the previous section by computing $\|\mathbf{L}U\|$ using interval arithmetic; however, matrix–vector products in interval arithmetic may be slow (as was the case for our computational environment), so we describe here an alternative procedure in which the matrix–vector products are computed using floating-point arithmetic: we replace $L_h$ with the floating-point matrix $M = \operatorname{mid}(\mathbf{L})$, and keep track of the error directly, in a sort of normwise ball arithmetic, bounding the error with $\delta = \|\operatorname{rad}(\mathbf{L})\|$. We work out the required bounds in this section, for both the $\ell^1$ norm (used in the Ulam projection) and the $\ell^\infty$ norm (used in the piecewise linear projection). We first need to bound the computational error produced by products with $M$.

**Lemma 8.10.** *Given* $M \in \mathbb{R}^{n\times n}$ *and* $v \in \mathbb{R}^n$, *let* $\tilde{w} = \mathsf{fl}(Mv)$ *be the vector obtained by evaluating the product* $w = Mv$ *in an inexact floating-point arithmetic system with machine precision* $\mathfrak{u}$. *Then, for both norms* $\|\cdot\|_{\ell^1}$ *and* $\|\cdot\|_{\ell^\infty}$, *it holds that*

$$\|\tilde{w} - w\| \le \gamma_z\|M\|\|v\|,$$

*where* $\gamma_z := \frac{z\mathfrak{u}}{1-z\mathfrak{u}}$, *and* $z$ *is the maximum number of nonzero entries in a row of* $M$.

**Proof.** This result follows from [40, Section 3.5], after noting that for a sparse matrix we can replace $\gamma_n$ with $\gamma_z$, since each sum has at most $z$ terms (as already argued in [17]). $\square$

The main results used to bound the total error are the following. The simplest case is that of an $i$-preserving projection, for which $Q_h = L_h$.

**Lemma 8.11.** *Let* $\tilde{v}_0 = v_0 \in \mathbb{R}^n$ *be a given fixed vector, and let* $M \in \mathbb{R}^{n\times n}$ *be a matrix such that* $\|L_h - M\| \le \delta$. *For each* $k = 1, 2, \dots$, *let* $\tilde{v}_{k+1} := \mathsf{fl}(M\tilde{v}_k)$ *be the vector obtained by evaluating the product* $Mv_k$ *in floating-point arithmetic, and let the sequence* $\epsilon_k$ *be defined recursively as*

$$\epsilon_0 = 0, \quad \epsilon_{k+1} = \gamma_z\|M\|\|\tilde{v}_k\| + \delta\|\tilde{v}_k\| + \|L_h\|\epsilon_k. \tag{32}$$

*Then,*

$$\left\| \tilde{v}_k - (L_h)^k v_0 \right\|_{\ell^1} \le \epsilon_k, \quad k = 0, 1, 2, \dots.$$

**Proof.** Arguing by induction, we have

$$\left\| \tilde{v}_{k+1} - L_h^{k+1}v_0 \right\| \le \|\tilde{v}_{k+1} - M\tilde{v}_k\| + \|M\tilde{v}_k - L_h\tilde{v}_k\| + \left\| L_h(\tilde{v}_k - L_h^k v_0) \right\| \le \gamma_z\|M\|\|\tilde{v}_k\| + \delta\|\tilde{v}_k\| + \|L_h\|\epsilon_k. \quad \square$$

Note that for the Ulam projection $\left\| L_h^U \right\|_{\ell^1} = 1$, so we can remove that factor.

If the projection is not $i$-preserving, the corresponding estimate for $Q_h$ is slightly more involved, because we have to keep track of the second summand in $Q_h = L_h + ei^*(I - L_h)$. Let us introduce the matrix $N = I - ei^*$, so that

$$Q_h v = NL_h v + ei^* v,$$

and the second summand vanishes if $v \in \mathcal{U}_h^0$. This suggests that we can approximate the action of $Q_h$ with that of $NL_h$.

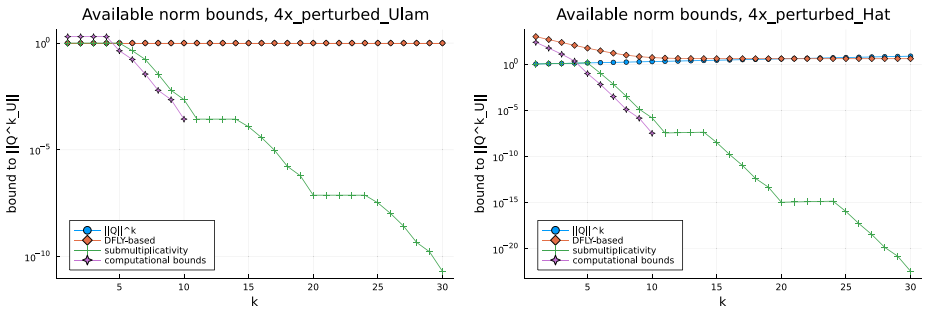


**Fig. 2.** Comparison of the norm bounds obtained from various sources for the Ulam (left) and piecewise linear (right) discretization of $T(x) = 4x + 0.01\sin(8\pi x) \mod 1$ with $n = 1024$; they are obtained with $k_{\max} = 10$ computational norm bounds and do not rely on coarser grids.

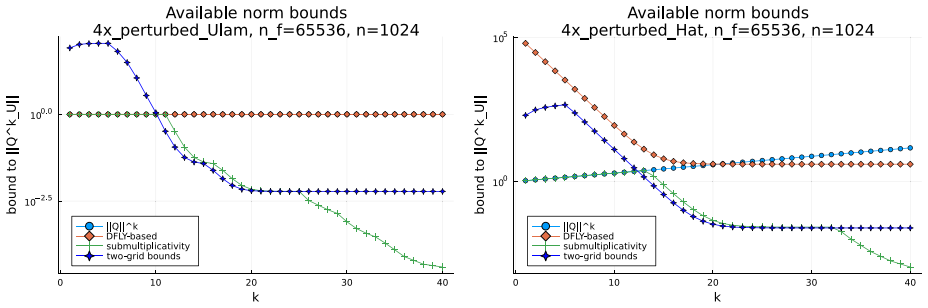


**Fig. 3.** Comparison of the norm bounds obtained from various sources for the Ulam (left) and piecewise linear (right) discretization of $T(x) = 4x + 0.01\sin(8\pi x) \mod 1$ with $n_F = 65536$; they are obtained from $k_{\max} = 10$ computational norm bounds on the coarse grid with $n = 1024$.

**Lemma 8.12.** *Let $\tilde{v}_0 = v_0 \in \mathcal{U}_h^0$ be a given fixed vector, and let $M \in \mathbb{R}^{n\times n}$ be a matrix such that $\|L_h - M\| \le \delta$. For each $k = 1, 2, \dots$, let $\tilde{w}_{k+1} = \mathsf{fl}(M\tilde{v}_k)$ and*

$$\tilde{v}_{k+1} = \mathsf{fl}(N\tilde{w}_{k+1}) = \mathsf{fl}\left(\tilde{w}_{k+1} - e\left(\sum_{i=1}^{n} f_i(\tilde{w}_{k+1})_i\right)\right) \tag{33}$$

*be the vectors obtained by approximating $Q_h\tilde{v}_k$ in floating-point arithmetic, and let the sequence $\epsilon_k$ be defined recursively as*

$$\begin{aligned}\epsilon_0 &= 0,\\ \epsilon_{k+1} &= \gamma_{n+2}\left\|\begin{bmatrix} I & -ei^*\end{bmatrix}\right\|\left(\|w_{k+1}\| + \|ei^*\|\,\|\tilde{w}_k\|\right)\\ &\quad + \|N\|\left(\gamma_z\|M\| + \delta\right)\|\tilde{v}_k\| + \|Q_h\|\,\epsilon_k.\end{aligned} \tag{34}$$

*Then,*

$$\left\|\tilde{v}_k - Q_h^k v_0\right\|_{\ell^\infty} \le \epsilon_k, \quad k = 0, 1, 2, \dots.$$

**Proof.** Standard forward error analysis of the formula (33) gives

$$\|\tilde{v}_{k+1} - N\tilde{w}_{k+1}\| \le \left\|\begin{bmatrix} I & -ei^*\end{bmatrix}\right\|\gamma_{n+2}\|\tilde{w}_{k+1}\|.$$

This bound is essentially the same that would follow from applying Lemma 8.10 to the product $\tilde{v}_{k+1} = \begin{bmatrix} I & -ei^*\end{bmatrix}\tilde{w}_k$, only with $\gamma_{n+2}$ instead of $\gamma_{n+1}$ because forming the products in $ei^*$ could in principle introduce another relative error of the magnitude of the machine precision. (Note that this additional error term can be omitted in the case of the piecewise linear discretization, since $e$ is the vector of all ones and products with its entries are exact.)

Moreover,

$$\begin{aligned}\|(NL_h - Q_h)\tilde{v}_k\| &= \|ei^*\tilde{v}_k\|\\ &= \|ei^*(\tilde{v}_k - N\tilde{w}_k)\|\\ &\le \|ei^*\|\,\|\tilde{v}_k - N\tilde{w}_k\|\\ &\le \|ei^*\|\left\|\begin{bmatrix} I & -ei^*\end{bmatrix}\right\|\gamma_{n+2}\|\tilde{w}_k\|.\end{aligned}$$

Once we have established these bounds, we can conclude with the triangle inequality:

$$\begin{aligned}\left\|\tilde{v}_{k+1} - Q_h^{k+1}v_0\right\| &\le \|\tilde{v}_{k+1} - N\tilde{w}_{k+1}\| + \|N(\tilde{w}_{k+1} - M\tilde{v}_k)\|\\ &\quad + \|N(M - L_h)\tilde{v}_k\|\\ &\quad + \|(NL_h - Q_h)\tilde{v}_k\| + \left\|Q_h(\tilde{v}_k - Q_h^k v_0)\right\|\\ &\le \left\|\begin{bmatrix} I & -ei^*\end{bmatrix}\right\|\gamma_{n+2}\|w_{k+1}\| + \|N\|\gamma_z\|M\|\,\|v_k\|\\ &\quad + \|N\|\,\delta\,\|v_k\|\\ &\quad + \|ei^*\|\left\|\begin{bmatrix} I & -ei^*\end{bmatrix}\right\|\gamma_{n+2}\|w_k\| + \|Q_h\|\,\epsilon_k. \quad \square\end{aligned}$$

In the case of the piecewise linear projection, $\|ei^*\|_{\ell^\infty} = 1$ and $\left\|\begin{bmatrix} I & -ei^*\end{bmatrix}\right\|_{\ell^\infty} = \|N\|_{\ell^\infty} = 2$.

All the norms appearing in these lemmas can be replaced with computable bounds from above. To obtain a bound for $Q_h$, we can use

$$\|Q_h\| = \|M + (L_h - M) + e(i^* - i^*L_h)\| \le \|M\| + \delta + \|e\|\,\|i^* - i^*L_h\|.$$

A bound $\|i^* - i^*L_h\| \le \|i^* - i^*\mathbf{L}\|$ can be computed with a single vector–matrix product performed in interval arithmetic. In practice this approach performed quite well in our examples, since $\|i^* - i^*\mathbf{L}\|$ is quite small for all the experiments described in Section 9.

A full algorithm, for both the $\ell^1$ and $\ell^\infty$ norms, is sketched in Algorithm 2. If $M$ has at most $z$ nonzeros in each row, this computation requires $O(n^2 z k_{\max})$ arithmetic operations.

**Remark 8.13.** It follows from (34) that $\epsilon_{k+1} \ge \|Q_h\|^k\epsilon_1$, i.e., in the non-$i$-preserving case the bounds grow by at least a factor $\|Q_h\|$ at each iteration. A more careful analysis could be made to replace some terms $\|Q_h\|^k$ with $\|Q_h^k\|$; we have implemented that and combined it with the bounds (21), but in the end we observed no practical advantage, since the bounds produced by (21) are much worse than $\|Q_h\|^k$ for moderate values of $k$, see Figs. 2 and 3.

### 8.5. Aggregating norm bounds from various sources

Bounds on the form $\left\|Q_h^k|_{\mathcal{U}_h^0}\right\| \le C_k$ come from various sources, some *a priori*, some requiring explicit computation:

(1) $\left\|Q_h^k|_{\mathcal{U}_h^0}\right\| \le \|Q_h^k\| \le \|Q_h\|^k \le (\|\mathbf{L}\| + \|e\|\,\|i^* - i^*\mathbf{L}\|)^k$, from basic norm properties. For the Ulam discretization, $\|Q_h\| = 1$, hence

**Algorithm 2** Algorithms to estimate norms of powers

**function** NORM1_OF_POWERS($M$, $k_{\max}$)
**Require:** $M = \mathrm{mid}(\mathbf{L})$, $\delta = \|\mathrm{rad}(\mathbf{L})\|$
**Ensure:** bounds $C_k \geq \left\|(Q_h)^k|_{\mathcal{U}_h^0}\right\|_{\ell^1}$ for $k = 1, 2, \dots, k_{\max}$
  **for** $k = 1, \dots, k_{\max}$ **do**
    $C_k \leftarrow 1$
  **end for**
  **for** $j = 1, 2, \dots, n-1$ **do**
    $v \leftarrow e_1 - e_{j+1}$;
    **for** $k = 1, \dots, k_{\max}$ **do**
      $v \leftarrow Mv$ ▷ Rounding to nearest
      $v \leftarrow v - ei^* v$ ▷ Skipped if $i$-preserving
      $C_k \leftarrow \max(C_k, \|v\|_{\ell^1} + \epsilon_k)$ ▷ Rounding up; $\epsilon_k$ as in (32) or (34)
    **end for**
  **end for**
**end function**
**function** NORMINF_OF_POWERS($M$, $k_{\max}$)
  ▷ compute bounds $C_k \geq \left\|(Q_h)^k|_{\mathcal{U}_h^0}\right\|_{\ell^\infty}$ for $k = 1, 2, \dots, k_{\max}$
  **for** $k = 1, \dots, k_{\max}$, $i = 1, \dots, n$ **do**
    $S_{ik} \leftarrow 0$
  **end for**
  **for** $j = 1, 2, \dots, n-1$ **do**
    $v \leftarrow e_1 - e_{j+1}$;
    **for** $k = 1, \dots, k_{\max}$ **do**
      $v \leftarrow Mv$ ▷ Rounding to nearest
      $v \leftarrow v - ei^* v$ ▷ Skipped if $i$-preserving
      $S_{ik} \leftarrow S_{ik} + |v_i| + \epsilon_k$ ▷ Rounding up; $\epsilon_k$ as in (32) or (34)
    **end for**
  **end for**
  **for** $k = 1, \dots, m$ **do**
    $C_k \leftarrow \max_i S_{ik}$
  **end for**
**end function**

this bound is the constant 1. This norm is fast to compute, and effective for low values of $k$, but it will never get below 1, as $\|Q_h\| \geq 1$.

(2) $\left\|Q_h^k|_{\mathcal{U}_h^0}\right\| \leq \left\|Q_h^k\right\| \leq S_1 R_{k,h,1} + S_2 R_{k,h,2}$, from (21). For the Ulam discretization, $E = 0$, hence this bound is once again the constant 1. This a-priori bound requires only the Lasota–Yorke inequality constants, but it is typically equal of worse than the other alternatives.

(3) $\left\|Q_h^k|_{\mathcal{U}_h^0}\right\| \leq \min_{0<i<k}(C_i C_{k-i})$, which comes from the sub-multiplicativity of norms and the fact that $Q(\mathcal{U}_h^0) \subseteq \mathcal{U}_h^0$. This estimate is based on the bounds $C_i$ obtained for $i < k$ with the other methods, but once those are available it is cheap to compute and effective. It becomes useful only after bounds smaller than 1 have already been obtained for at least some $i < k$.

(4a) computational estimates obtained with Algorithm 2. These bounds can be poor for small values of $k$, but they are our only resource to get non-trivial bounds smaller than 1 in the first place. As their cost scales with $O(n^2)$, these can be computed effectively only for discretizations with moderate $n$.

(4b) estimates obtained from a coarser grid using (22). These bounds are an effective replacement of those in Item 4a when $n$ is large. Exactly like the bounds in Item 4a, these are typically poor for small values of $k$, but they are the key ingredient to achieve bounds smaller than 1 in the coarse-fine strategy.

For each $k$, our upper bound $C_k$ is the minimum of the bounds coming from items (1)–(4a) (or (4b)). It is essential to use multiple sources of bounds: the bound in item (1) is effective for small values of $k$; the bound in item (4) is the only one that can go below 1, and the bound in (3) can be used to combine the other ones and extend them to larger values of $k$. We plot in Figs. 2 and 3 the norm bounds obtained from all these sources, on two representative examples with the Ulam and piecewise linear discretizations.

We note that the two-grid strategy is not guaranteed to succeed and yield a bound $C_m^F < 1$ for some $m$: in particular, when $n$ is too small (and $h$ too large), the second term in the right-hand side of (22) is greater than 1 even for large values of $m$. An example is shown in Fig. 4.

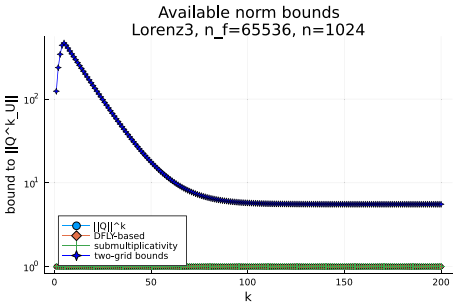


**Fig. 4.** Norm bounds obtained from various sources for the Ulam discretization of the third iterate of the Lorenz map (38) with $n_F = 65536$; they are obtained from $k_{\max} = 10$ computational norm bounds on the coarse grid with $n = 1024$.

### 8.6. The algorithms

Putting everything together, we can formulate the following algorithms. To compute a *one-grid* bound for a dynamic using a discretization with $n$ equal intervals, we

(1) (DFLY coefficients) Compute the coefficients $A, B$ of the Lasota–Yorke inequality (2). This computation requires finding rigorous bounds on the $T'$ and the distorsion $T''/(T')^2$ on each branch of the dynamic, via interval optimization. Its cost does not depend on the discretization size $n$.

(2) (matrix assembly) Construct an interval sparse matrix $\mathbf{L} \ni L_h$ with Algorithm 1. Its cost is $O(bn)$.

(3) (eigenvalue computation) Compute an approximated eigenvector $Q_h \tilde{u}_h \approx \tilde{u}_h$ using the restarted Arnoldi method in machine arithmetic. Also compute rigorous bounds $\varepsilon_1 \geq \|Q_h \tilde{u}_h - \tilde{u}_h\|$ and $\varepsilon_2 \geq |i^* \tilde{u}_h - 1|$ which will be needed in 8.3.

(4) (norms of powers) Compute norm bounds $\left\|Q_h^k|_{\mathcal{U}_h^0}\right\| \leq C_k$ for $k = 1, 2, \dots, k_{\max}$, using Algorithm 2 to obtain some first computational bounds and the techniques in Section 8.5 to refine them. The value of $k_{\max}$ chosen must be sufficient to obtain $C_{k_{\max}} < 1$; if this inequality does not hold, we can repeat the computation with a larger value of $k_{\max}$. If we choose to multiply by 2 the value of $k_{\max}$ at each restart, then the cost of this step is $O(n^2 z k_{\max})$, with $z \sim b$ and $k_{\max} \sim \log n$ (by the arguments in Section 3.1). Assuming a constant number of iterations suffices, its cost is $O(nz)$.

(5) (error estimation) Using interval arithmetic or directed rounding to get rigorous bounds, compute the bound for $\|u - \tilde{u}_h\|$ in 8.3. The cost for this step is merely $O(k_{\max})$, since $\varepsilon_1$ and $\varepsilon_2$ have already been computed.

The computational cost of this algorithm scales as $O(n^2 \log n)$, seriously limiting its usefulness when large values of $n$ are required. To reduce the cost, we can compute instead a *two-grid* bound as follows, using a coarse grid with $n_C$ equal intervals and a fine grid with $n_F$ equal intervals.

(1) (DFLY coefficients) Compute the coefficients $A, B$ of the Lasota–Yorke inequality (2), as above.

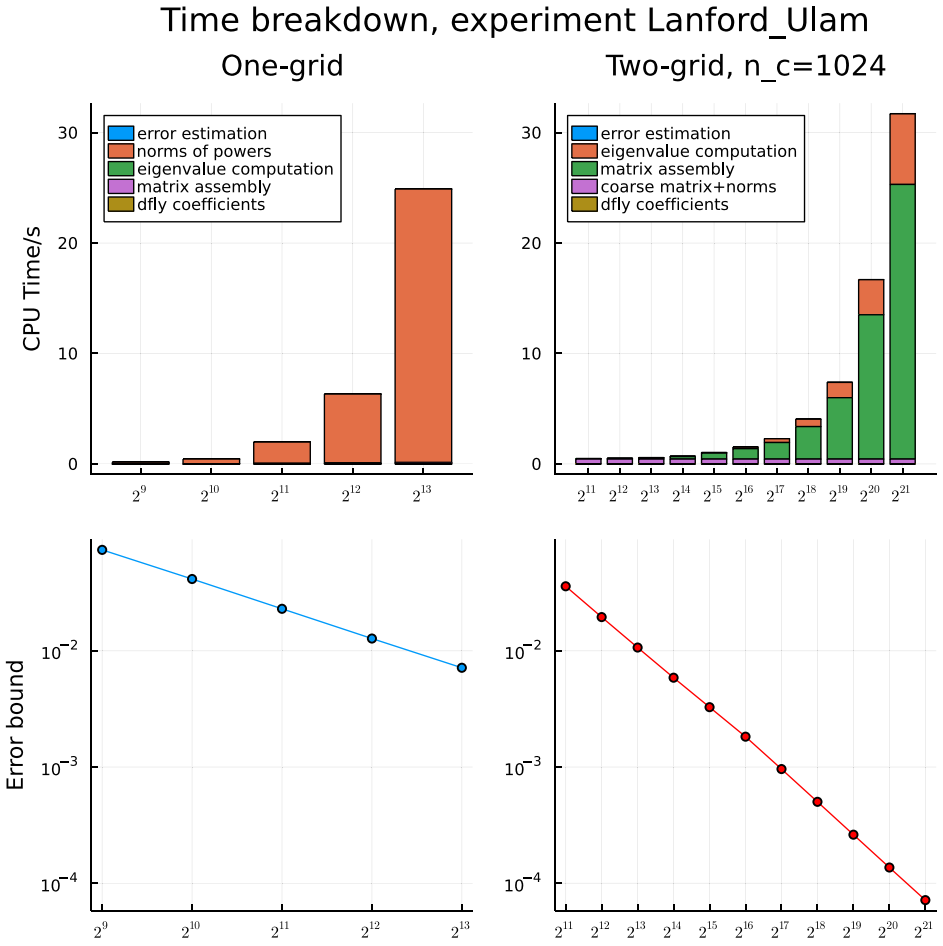


**Fig. 5.** Error bounds proved and CPU time breakdown along the five steps of the algorithms in Section 8.6 for the Lanford map (35).

(2) (coarse matrix+norms) Perform steps 2 and 4 of the previous algorithm with $n = n_C$. The cost is $O(n_C^2 b \log n_C)$ as argued above.
(3) (matrix assembly) Construct $\mathbf{L} \ni L_{h_F}$ with Algorithm 1. Its cost is $O(bn_F)$.
(4) (eigenvalue computation) Compute an approximated eigenvector $Q_{h_F}\tilde{u}_{h_F} \approx \tilde{u}_{h_F}$, as well as $\varepsilon_1$ and $\varepsilon_2$ as above. This step costs $O(n_F z)$.
(5) (error estimation) Compute norm bounds $C_{k,F}$ using the techniques in Section 8.5, with Step 4b instead of 4a, and use them to compute a bound for $\left\|u - \tilde{u}_{h_F}\right\|$ using 8.3 (with $n = n_F$). This step costs $O(k_{\max})$.

The total cost depends quadratically on $n_C$, but only linearly on $n_F$. We shall see that this algorithm outperforms the one-grid strategy for suitable values of $n_C$ and $n_F$.

## 9. Numerical experiments

The proposed algorithm has been implemented in the Julia language for both the Ulam (Section 6) and piecewise linear projection (Section 7). Our code is available on https://github.com/JuliaDynamics/RigorousInvariantMeasures.jl. The following numerical experiments have been performed with Julia 1.7.1 on an Imac i7-4790K 4.00 GHz.

### 9.1. The Lanford map

As a first experiment, we compute the invariant measure of

$$T : [0,1] \to [0,1], \quad T(x) = 2x + \frac{1}{2}x(1-x) \mod 1 \tag{35}$$

with the Ulam projection. We tested both the one-grid described above, with various powers of 2 as the values of $n$, and the two-grid bound, with $n_C = 1024$ and various powers of 2 as the values of $n_F$. We display in Fig. 5 the rigorous error bounds on $\|u - \tilde{u}_h\|_{L^1}$ that have been proved, and a breakdown of how the CPU time is divided between the steps of each algorithm described in Section 8.6. Bounds on the same quantity have been computed in [17], but working on the iterate $T^2$ in place of $T$ was necessary there, because the inequality [17, Theorem 5.2] there is weaker than Lemma 5.1 here. The major innovation in this work is the two-grid strategy, which allows to prove bounds as small as $10^{-4}$ in less than one minute of CPU time. With the two-grid strategy (on the right), larger dimensions can be used, and the majority of time is spent assembling the matrix $\mathbf{L}_{h_F}$ and computing its fixed point vector.

A detailed analysis of the tradeoff between error bound and CPU time obtained with various choices of $n, n_C, n_F$ is shown in Fig. 6. One can see from this plot that the error scales approximately as $t^{-1/2}$ with the one-grid strategy, and approximately as $t^{-1}$ with the two-grid strategy, as predicted by our complexity estimates. After an initial period to amortize the power norm computation, all sufficiently large choices of $n$ have similar asymptotic efficiency; this suggests that to improve the precision of an estimate it is better to keep $n$ constant and increase the value of $n_F$.

Using the technique above we were able to compute an enclosure for the Lyapunov exponent of the Lanford map

$$\int \log(|T'|)\, df \in [0.657657, 0.657667]$$

where the diameter of enclosure is $9.45 \cdot 10^{-6}$. This estimate was produced with $n_C = 2^{11}$ and $n_F = 2^{25}$ in 1476 s; most of this time was spent assembling the matrix $\mathbf{L}_{h_F}$.

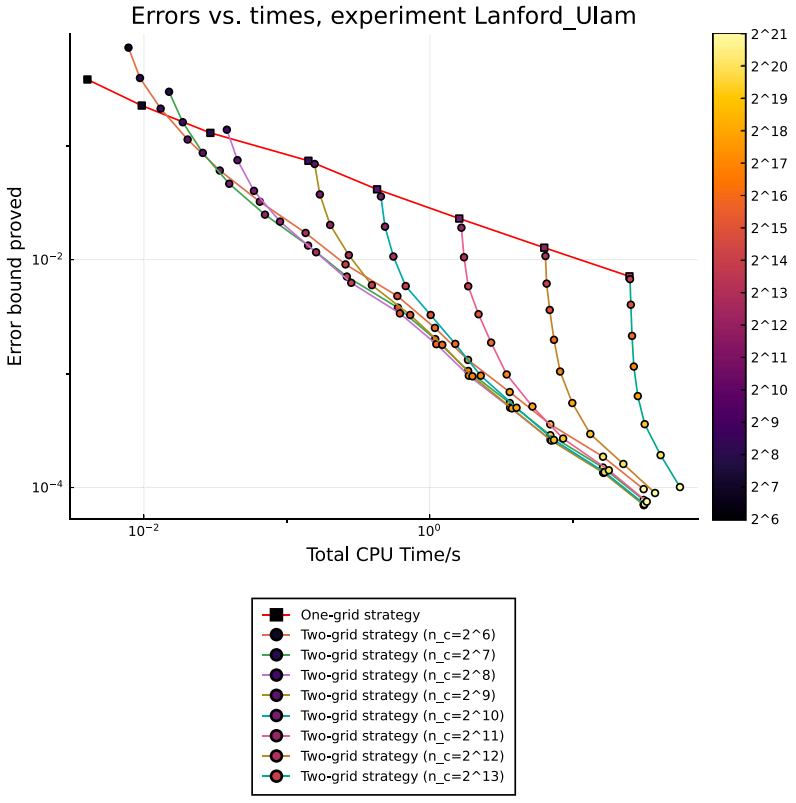


**Fig. 6.** Error bound vs. time for various choices or $n$ and $n_F$, for the Lanford map (35). The marker color represents the value of $n$ or $n_F$.

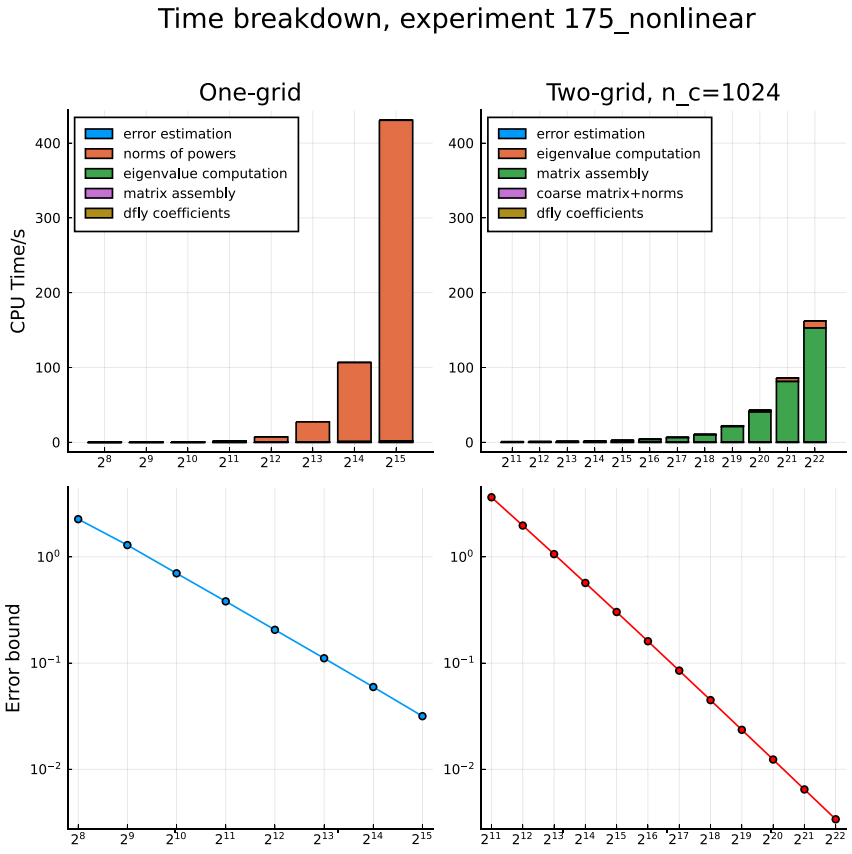


**Fig. 7.** Error bounds proved and CPU time breakdown for the non-Markov map (36).

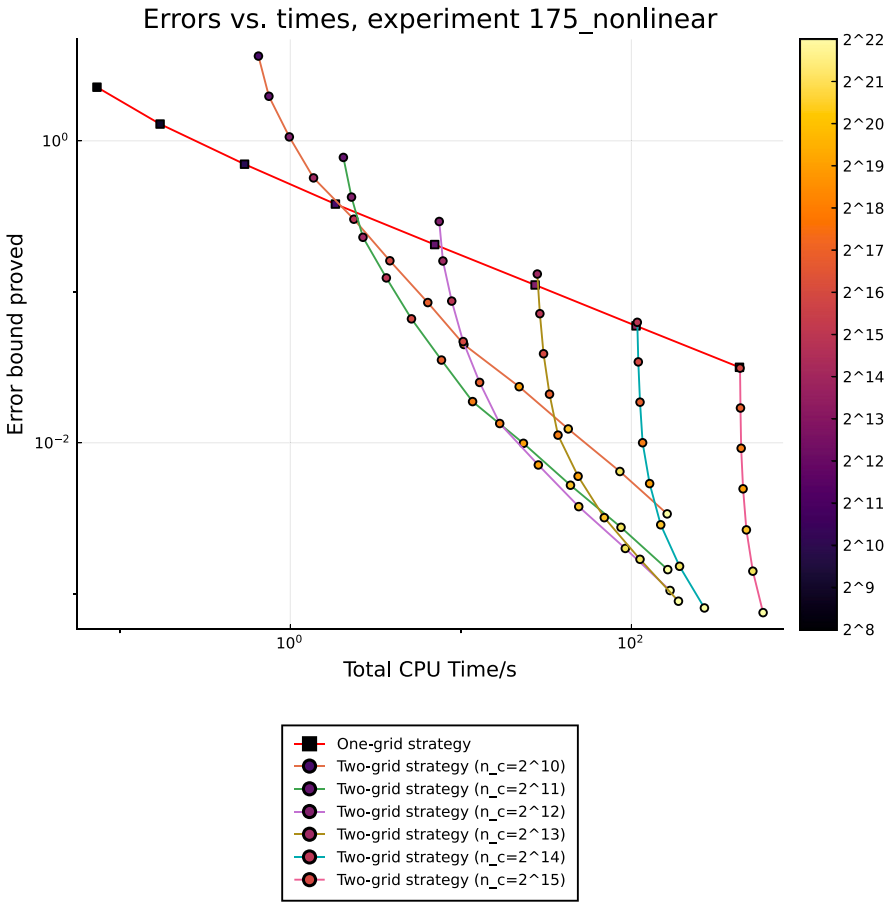


**Fig. 8.** Error bound vs. time for various choices of $n$ and $n_F$ on the non-Markov map (36).

### 9.2. A non-linear non-Markov map

We consider the following nonlinear modification of $\frac{17}{5}x \mod 1$:

$$T(x) = \begin{cases} \frac{17}{5}x & 0 \le x \le \frac{17}{5}, \\ \frac{34}{25}(x-\frac{5}{17})^2 + 3(x-\frac{5}{17}), & \frac{5}{17} < x \le \frac{10}{17}, \\ \frac{34}{25}(x-\frac{10}{17})^2 + 3(x-\frac{10}{17}), & \frac{10}{17} < x \le \frac{15}{17}, \\ \frac{17}{5}(x-\frac{15}{17}) & \frac{15}{17} < x \le 1, \end{cases} \tag{36}$$

again with the Ulam projection. This is another of the dynamics considered in [17], this time without modification. We display the same information in Figs. 7 and 8.

This experiment is more challenging, especially since $n_C = 2^{10}$ is required to reach a bound $C_{k_{\max}} < 1$ with the two-grid strategy, but the same features appear in the plots, highlighting in particular the massive improvements provided by the two-grid strategy.

Using the technique above we were able to compute an enclosure for the Lyapunov exponent of this map

$$\int \log(|T'|)\, df \in [1.21933, 1.22016]$$

where the diameter of enclosure is 0.00082. The computation time to obtain such an approximation was 2110 s.

### 9.3. A Markov perturbation of $4x \mod 1$

The next example we consider is

$$T(x) = 4x + 0.01\sin(8\pi x) \mod 1. \tag{37}$$

In this experiment, we use the piecewise linear discretization to provide a bound to $\|u - \tilde{u}_h\|_{L^\infty}$ in the $L^\infty$ norm, again replicating an example in [17]. The results are reported in Figs. 9 and 10.

Despite the different projection, the workload and results are very similar. Note that assembling the matrix $\mathbf{L}$ is more expensive than in the other examples; this is not related to the different projection, but it is due to the fact that providing certified enclosures for trigonometric functions is computationally expensive.

Using the technique above we were able to compute an enclosure for the Lyapunov exponent of this map

$$\int \log(|T'|)\, df \in [1.38530, 1.38531]$$

where the diameter of enclosure is $3.3 \cdot 10^{-6}$. The computation time to obtain such an approximation was 3016 s.

### 9.4. One-dimensional Lorenz map

The Lorenz system is a famous example of a 3-dimensional vector flows that, presents a strange attractor. We refer to [41] for a historical introduction to the geometric model of the Lorenz system and a careful presentation of its construction; the example we present in this subsection is the one-dimensional map associated to the stable foliation of the geometric Lorenz system studied in [42].

This map is

$$T(x) = \begin{cases} \theta\left|x-\frac{1}{2}\right|^\alpha & 0 \le x < \frac{1}{2}, \\ 1-\theta\left|x-\frac{1}{2}\right|^\alpha & \frac{1}{2} < x \le 1, \end{cases} \tag{38}$$

with $\alpha = 51/64$ and $\theta = 109/64$. Note that the derivative of this map goes to $\infty$ as we approach $1/2$, so the one-step Lasota–Yorke inequality

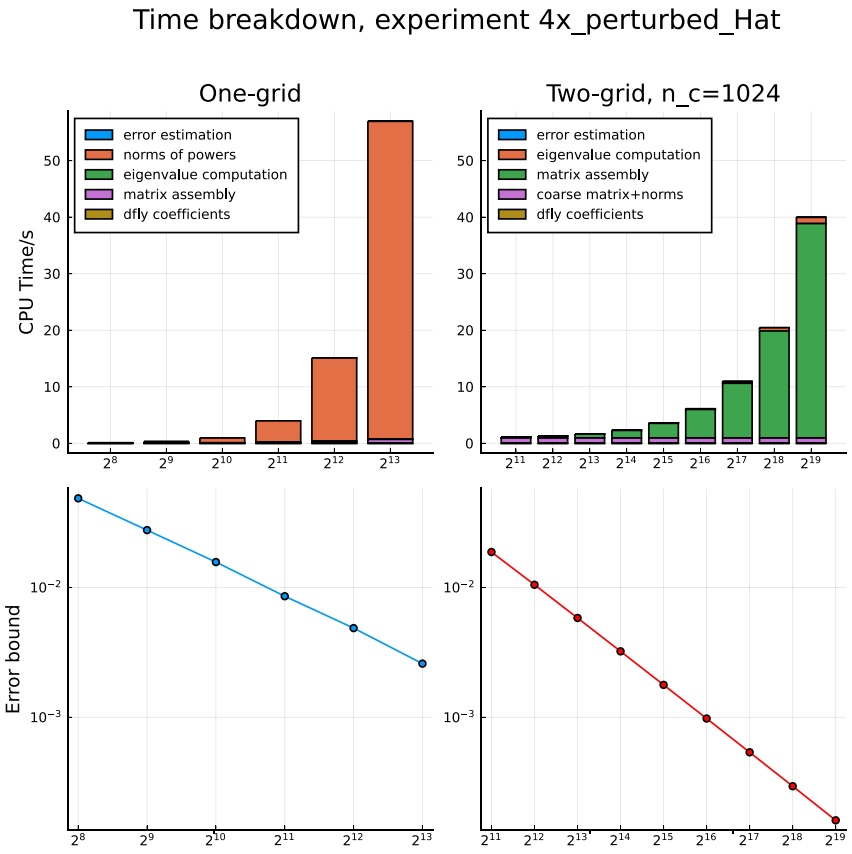


**Fig. 9.** Error bounds proved and CPU time breakdown for the "$4x$ perturbed" map (37).

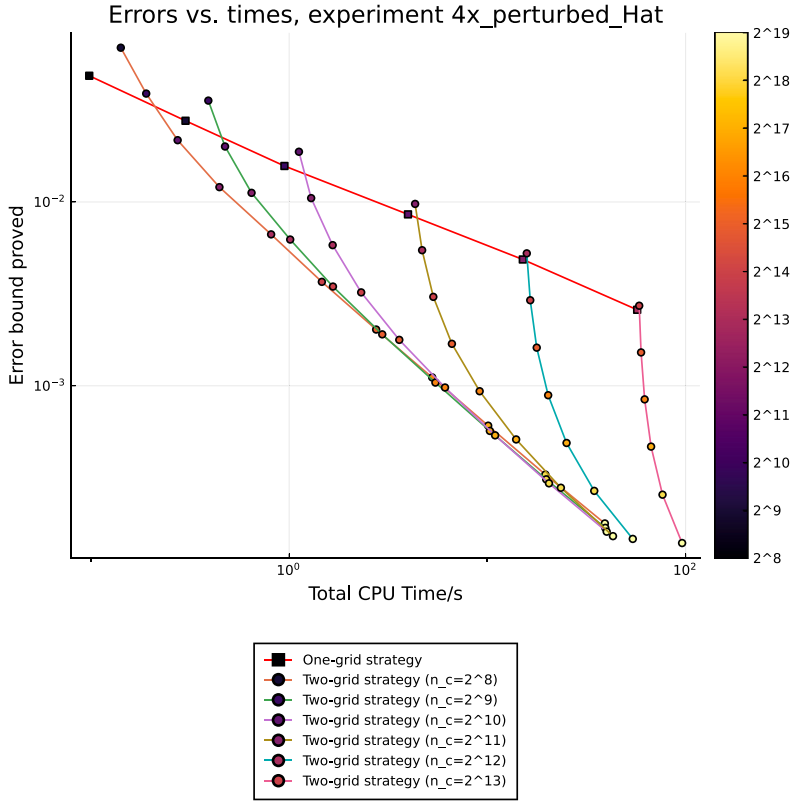


**Fig. 10.** Error bound vs. time for various choices of $n$ and $n_F$ on the "$4x$ perturbed" map (37).

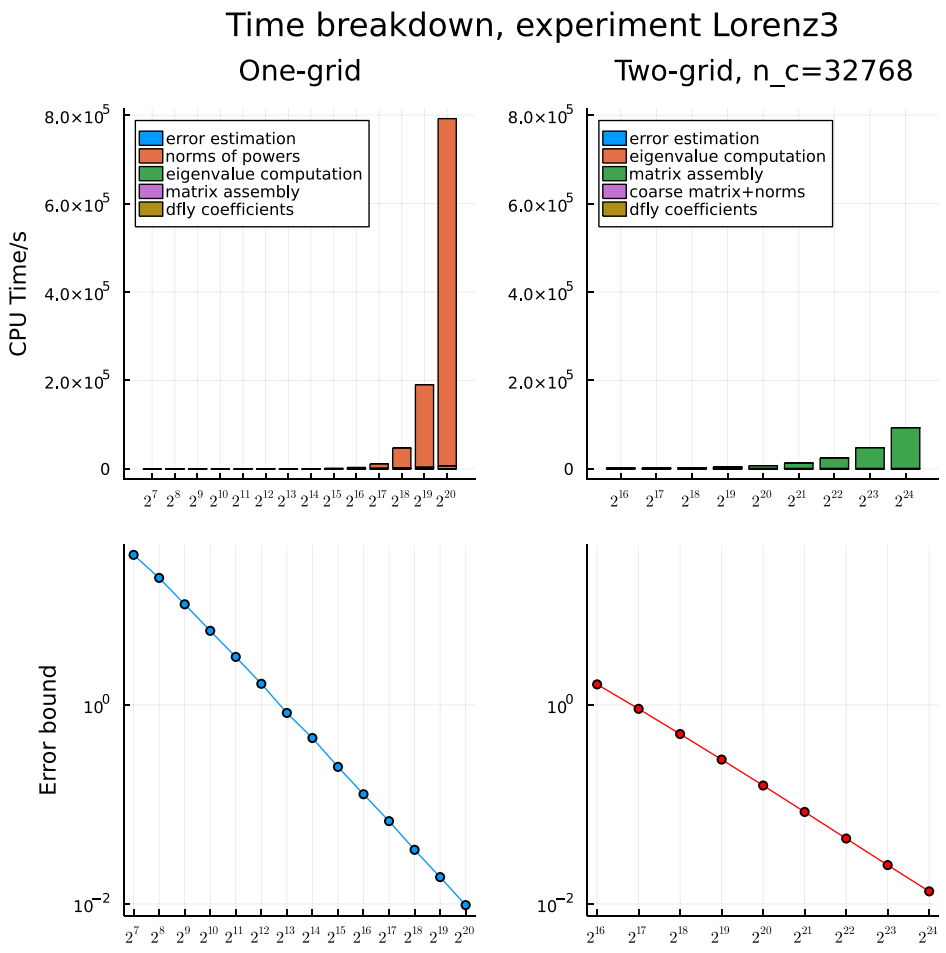


**Fig. 11.** Error bounds proved and CPU time breakdown for the third iterate of the Lorenz map (38).

which we have been using in the other examples does not hold; by direct computation, one sees that $T''/(T')^2$ behaves as $|x-1/2|^{-\alpha}$ near $1/2$, and hence it is unbounded.

**Lemma 9.1.** *Let $T : [0,1] \to [0,1]$ and suppose there exists a finite partition $\{P_k\}_{k=1}^{b}$ of $[0,1]$ such that*

(1) *$T_k = T|_{P_k}$ is $C^2$,*
(2) *$|T'(x)| > 2$ for all $x \in [0,1]$.*

*Let $I_l = \{x \mid |T''/(T')^2| \geq l\}$ and suppose there exists an $l$ such that*

$$A = \frac{1}{2}\int_{I_l}\left|\frac{T''}{(T')^2}\right| dm + \frac{2}{\inf(|T'|)} < 1,$$

*then*

$$\operatorname{Var} Lf \leq A \operatorname{Var}(f) + \left(\max_k \frac{2}{|P_k|} + l\right)\|f\|_{L^1}.$$

To prove a one-step Lasota–Yorke inequality for our example, we applied this lemma to the third iterate of the map $T$. The coefficients in the obtained inequality are large ($A \approx 0.922, B \approx 48.43$) and quite expensive to compute (about one minute).

The results obtained are presented in Figs. 11 and 12. One can see that the error bounds are generally worse than those obtained with the previous maps; in particular, we need to use a larger value of the coarse discretization size $n$. Indeed, the two-grid strategy fails to produce useful bounds when used with $n = 1024$: due to the large value of $B$, with this choice of $h$ the formula (22) produces only bounds for $Q^m_{h_F}|\mathcal{U}^0_{h_F}$ that are larger than 1, hence the convergence of the series appearing in (13) cannot be proved with Lemma 3.7 and the method fails. Nevertheless, larger values of $n$ and $n_F$ yields valid bounds for the error, as shown in Fig. 12; the two-grid strategy eventually surpasses the efficiency of the one-grid bounds, and for instance it is faster by an order of magnitude when one seeks to prove an error bound of $10^{-2}$.

Using the technique above with $n_C = 2^{17}$, $n_F = 2^{24}$, we can compute an enclosure for the Lyapunov exponent of this map

$$\int \log(|T'|)\, df \in [0.580676, 0.786467]$$

where the diameter of enclosure is 0.2058. The computation time to obtain such an approximation was 101 828 s.

### 9.5. Limitations of machine arithmetic

In several computations involved in our algorithm, floating point arithmetic gives a lower bound on the attainable precision:

- the diameter of the interval entries of the interval matrix representing the discretized operator is generically bounded below by machine precision,
- machine floating point arithmetic is going to be the main source of the error stemming from the computation of the residual $\|Pu_h - u_h\|$.

A possible strategy to overcome machine arithmetic limitations could be to first compute a coarse approximation in machine arithmetic, allowing us to estimate mixing rates $C_k$, and then compute a finer approximation in higher precision floating point arithmetic, i.e., a "low-precision coarse – high-precision fine" scheme.

While this corresponds to a small modification of the code, no experiments have been done in this direction.

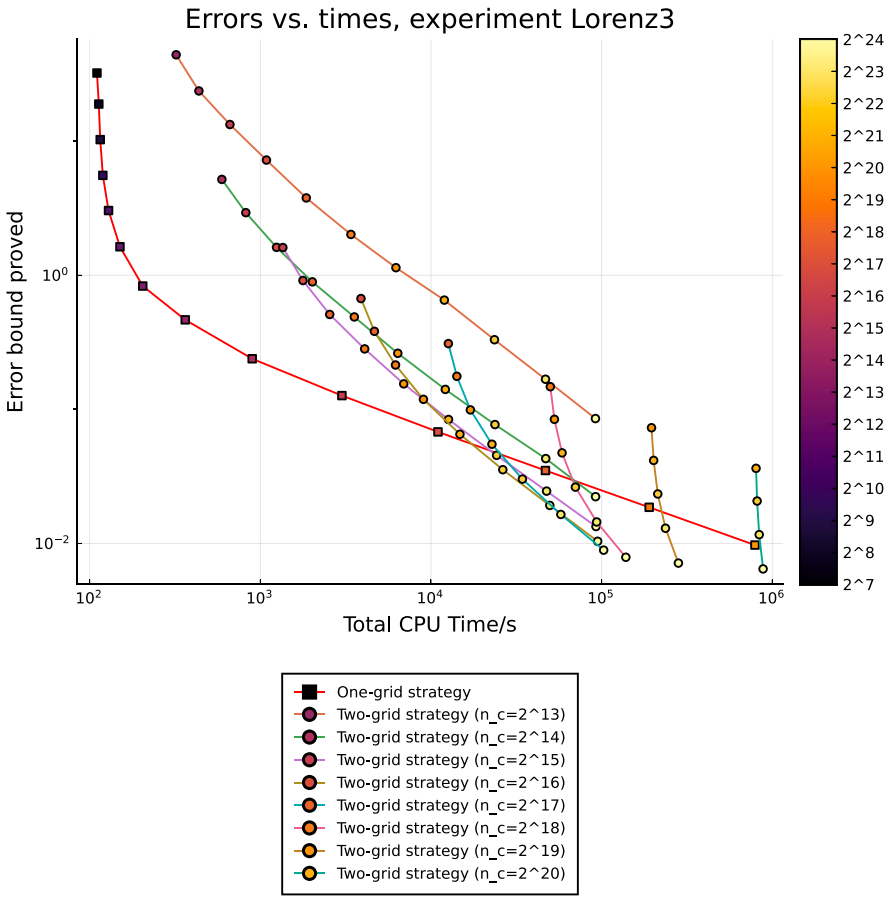


**Fig. 12.** Error bound vs. time for various choices of $n$ and $n_F$ on the third iterate of the Lorenz map (38).

Another, much more serious problem arising from machine precision is the numerical error arising in our norm estimates. If the discretized operator is not sparse, it may be impossible to prove that one of its iterates contracts $\mathcal{U}_0$, due to the estimates we need to put in place to guarantee an upper bound of the norm, see Section 8.4. This can also be solved by using higher precision floating point numbers, but the computational overhead would be difficult to manage.

## 10. Final remarks and considerations

In this paper we introduced a general framework for the approximation of invariant measures. We gave a finite set of inequalities that, once proved, give rise to an algorithm for the approximation, once we can prove computationally the existence of an $m$ such that $\left\|Q_h^m|_{\mathcal{U}_0}\right\| \leq C_m < 1$.

On the computational side, the major contribution of this paper is the new "coarse-fine" framework based on two discretizations with grids of different sizes; this framework greatly reduces the computational burden of the estimation algorithms introduced in [17]. The experiments in [17] relied on computational norm estimation with Algorithm 2, which requires $O(mn^2)$ floating point operations to obtain estimates $C_k$ for $k \leq m$. Typically, $n \approx 10^5$ to $10^6$ is needed to get a meaningful estimate, so this computation was doable, but extremely slow. Here, we give a strategy to combine bounds from various sources in Section 8.5, including in particular those coming from the coarse-fine strategy (22). This improvement gives a major reduction in the computational time: while the results in [17] were obtained on a supercomputing cluster, we can replicate them in a few minutes on a common laptop computer.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The code package is open and available on GitHub

## Acknowledgments

SG is partially supported by the research project PRIN 2017S35 EHN_004 "Regular and stochastic behavior in dynamical systems" of the Italian Ministry of Education and Research. IN is partially supported by CNPq, CAPES (through the programs PROEX and the CAPES-STINT project "Contemporary topics in non uniformly hyperbolic dynamics"). The author is currently under "Afastamento do país para qualificação profissional, apresentação de trabalhos técnico-cientìficos e colaboração institucional do pessoal docente e técnico-administrativo" from UFRJ and is currently a Specially Appointed Associate Professor at Hokkaido University, and would like to thank Prof. Yuzuru Sato for the hospitality and scientific discussions. The author would like to thank Prof. Hiroki Sumi and Kyoto university for their hospitality during the final revision of this article. This work was supported by the Research Institute for Mathematical Sciences, an International Joint Usage/Research Center located in Kyoto University. FP is partially supported by INDAM, by project PRA_2020_61 "complex network analysis: from theory to applications" of the university of Pisa, and by the National Centre for HPC, Big Data and Quantum Computing/Spoke 6, Multiscale Modeling and Engineering Applications.

## References


[43] Young Lai-Sang. What are SRB measures, and which dynamical systems have them? J Stat Phys 2002;108(5):733–54.
[1] Galatolo Stefano, Monge Maurizio, Nisoli Isaia. Existence of noise induced order, a computer aided proof. Nonlinearity 2020;33(9):4237–76.
[2] Ding Jiu, Du Qiang, Li Tien-Yien. High order approximation of the Frobenius-Perron operator. Appl Math Comput 1993;53:151–71.
[3] Ding Jiu, Zhou Aihui. The projection method for computing multidimensional absolutely continuous invariant measures. J Stat Phys 1994;77(3):899–908.
[4] Dellnitz Michael, Junge Oliver. Chapter 5 - Set oriented numerical methods for dynamical systems. In: Fiedler Bernold, editor. Handbook of dynamical systems. Handbook of dynamical systems, vol. 2, Elsevier Science; 2002, p. 221–64.
[5] Dellnitz Michael, Junge Oliver. On the approximation of complicated dynamical behavior. SIAM J Numer Anal 1999;36(2):491–515.
[6] Bose Christopher, Murray Rua. The exact rate of approximation in Ulam's method. Discrete Contin Dyn Syst 2001;7(1):219–35.
[7] Murray Rua. Ulam's method for some non-uniformly expanding maps. Discrete Contin Dyn Syst 2010;26(3):1007–18.
[8] Froyland Gary. Extracting dynamical behavior via Markov models. In: Nonlinear dynamics and statistics. Boston, MA: Birkhäuser Boston; 2001, p. 281–321.
[9] Crimmins Harry, Froyland Gary. Fourier approximation of the statistical properties of Anosov maps on tori. Nonlinearity 2020;33(11):6244–96.
[10] Keane Michael, Murray Rua, Young Lai-Sang. Computing invariant measures for expanding circle maps. Nonlinearity 1998;11(1):27–46.
[11] Bahsoun Wael, Bose Christopher. Invariant densities and escape rates: Rigorous and computable approximations in the L infinity-norm. Nonlinear Anal TMA 2011;74(13):4481–95.
[12] Liverani Carlangelo. Rigorous numerical investigation of the statistical properties of piecewise expanding maps. A feasibility study. Nonlinearity 2001;14(3):463–90.
[13] Pollicott Mark, Jenkinson Oliver. Computing invariant densities and metric entropy. Comm Math Phys 2000;211(3):687–703.
[14] Ippei Obayashi. Computer-assisted verification method for invariant densities and rates of decay of correlations. SIAM J Appl Dyn Syst 2011;10(2):788–816.
[15] Wormell Caroline. Spectral Galerkin methods for transfer operators in uniformly expanding dynamics. Numer Math 2019;142(2):421–63.
[16] Galatolo Stefano, Monge Maurizio, Nisoli Isaia. Rigorous approximation of stationary measures and convergence to equilibrium for iterated function systems. J Phys A 2016;49:274001.
[17] Galatolo Stefano, Nisoli Isaia. An elementary approach to rigorous approximation of invariant measures. SIAM J Appl Dyn Syst 2014;13(2):958–85.
[18] Jenkinson O, Pollicott M. Rigorous effective bounds on the Hausdorff dimension of continued fraction Cantor sets: A hundred decimal digits for the dimension of E2. Adv Math 2018;325:87–115.
[19] Galatolo Stefano, Nisoli Isaia, Saussol Benoît. An elementary way to rigorously estimate convergence to equilibrium and escape rates. J Comput Dyn 2015;2(1):51–64.
[20] Bahsoun Wael, Galatolo Stefano, Nisoli Isaia, Niu Xiaolong. A rigorous computational approach to linear response. Nonlinearity 2018;31(3):1073–109.
[21] Pollicott M, Vytnova P. Linear response and periodic points. Nonlinearity 2016;29(10):3047–66.
[22] Bahsoun Wael, Galatolo Stefano, Nisoli Isaia, Niu Xiaolong. Rigorous approximation of diffusion coefficients for expanding maps. J Stat Phys 2016;163(6):1486–503.
[23] Jenkinson O, Pollicott M, Vytnova P. Rigorous computation of diffusion coefficients for expanding maps. J Stat Phys 2018;170(2):221–53.
[24] Marangio L, Sedro J, Galatolo S, Garbo A Di, Ghil Michael. Arnold maps with noise: Differentiability and non-monotonicity of the rotation number. J Stat Phys 2019.
[25] Galatolo Stefano, Hoyrup Mathieu, Rojas Cristóbal. Dynamics and abstract computability: Computing invariant measures. Discrete Contin Dyn Syst 2011;29(1):193–212.
[26] Bandtlow Oscar F, Slipantschuk Julia. Lagrange approximation of transfer operators associated with holomorphic data. 2020.
[27] Galatolo Stefano. Quantitative statistical stability, speed of convergence to equilibrium and partially hyperbolic skew products. J Ecole Polytech — Math 2018;5:377–405.
[28] Matsumoto K, Tsuda I. Noise-induced order. J Stat Phys 1983;31(1):87–106.
[29] Bezanson Jeff, Edelman Alan, Karpinski Stefan, Shah Viral B. Julia: A fresh approach to numerical computing. SIAM Rev 2017;59(1):65–98.
[30] Milovanović Gradimir V. Extremal problems and inequalities of Markov-Bernstein type for polynomials. In: Rassias Themistocles M, Srivastava Hari M, editors. Analytic and geometric inequalities and applications. Dordrecht: Springer Netherlands; 1999, p. 245–64.
[31] Keller Gerhard, Liverani Carlangelo. Stability of the spectrum for transfer operators. Ann Sc Norm Super Pisa - Cl Sci 1999;Ser. 4, 28(1):141–52.
[32] Galatolo Stefano. Statistical properties of dynamics. Introduction to the functional analytic approach. arXiv:1510.02615.
[33] Lasota A, Yorke James A. On the existence of invariant measures for piecewise monotonic transformations. Trans Amer Math Soc 1973;186:481–8.
[34] Viana Marcelo. Stochastic dynamics of deterministic systems. Lecture notes XXI braz. math. colloq., Rio de Janeiro: IMPA; 1997.
[35] Ulam SM. A collection of mathematical problems. Interscience tracts in pure and applied mathematics, no. 8, New York-London: Interscience Publishers; 1960, p. xiii+150.
[36] Tucker Warwick. Validated numerics: A short introduction to rigorous computations. Princeton University Press; 2011.
[37] Golub Gene H, van Loan Charles F. Matrix computations. 4th ed.. JHU Press; 2013.
[38] Miyajima Shinya. Verified bounds for all the singular values of matrix. Japan J Ind Appl Math 2014;31:513–39.
[39] Higham Nicholas J. Estimating the matrix $p$-norm. Numer Math 1992;62(1):539–55.
[40] Higham Nicholas J. Accuracy and stability of numerical algorithms. 2nd ed.. Society for Industrial and Applied Mathematics; 2002.
[41] Araújo Vítor, Pacifico Maria José. Three-dimensional flows. Ergebnisse der mathematik und ihrer grenzgebiete. 3. Folge / A series of modern surveys in mathematics, vol. 53, Berlin, Heidelberg: Springer; 2010.
[42] Galatolo Stefano, Nisoli Isaia. Rigorous computation of invariant measures and fractal dimension for maps with contracting fibers: 2D Lorenz-like maps. Ergodic Theory Dynam Systems 2016;36(6):1865–91.