# Ulam's scheme revisited: digital modeling of chaotic attractors via micro-perturbations

Gábor Domokos[1] and Domokos Szász[2]
[1] Department of Strength of Materials,
Budapest University of Technology and Economics,
H-1521 Budapest, Hungary.
[2] Institute of Mathematics,
Budapest University of Technology and Economics,
H-1521 Budapest, Hungary.

June 7, 2002

**Abstract**

We consider discretizations $f_N$ of expanding maps $f : I \to I$ in the strict sense: i.e. we assume that the only information available on the map is a finite set of integers. Using this definition for computability, we show that by adding a random perturbation of order $1/N$, the invariant measure corresponding to $f$ can be approximated and we can also give estimates of the error term. We prove that the randomized discrete scheme is equivalent to Ulam's scheme applied to the polygonal approximation of $f$, thus providing a new interpretation of Ulam's scheme. We also compare the efficiency of the randomized iterative scheme to the direct solution of the $N \times N$ linear system.

## 1 Introduction

Computer simulations of chaotic dynamical systems necessarily distort the qualitative behaviour due to roundoff errors. In particular, the invariant measure (if it exists) cannot be preserved: any computer simulation generates only a finite set of orbits with finite period. Our main aim is to understand how and how efficiently one can reconstruct the original invariant measure based on the finite information provided by the computer.

Our conceptual approach relies upon the following earlier methods:
**1.** Shortly after Rényi [14] discovered the existence of invariant measures for a wide class of maps, Ulam [19] already addressed the problem of their computability. His approach, called *'Ulam's scheme'*, reduces the infinite-dimensional problem to a finite dimensional one by averaging the density function between meshpoints. Nevertheless, the direct, exact computation of this discrete problem is still impossible, since the coefficients depend on the *exact* behaviour of the original map between meshpoints.
**2.** Liverani recognized this difficulty ([13]) and proposed to compute the Ulam scheme corresponding to a piecewise linear approximation of the original map; we will refer to this approach as *'Liverani's scheme'*. He proved that the density function associated with this scheme still converges to the original density function.

In Liverani's work the piecewise linear approximation is constructed by evaluating the function on a (uniform) mesh and connecting the meshpoints. He did not consider the *roundoff errors*

occuring during the evaluation of the function values, thus *he interpreted the independent variable as an integer (rational) number, while the function value was represented by a real number.* In this paper we adopt the point of view that the only thing *given* in any computer code is a finite set of (binary) integers, and therefore there is no reason to neglect the effect of the roundoff errors in the function value once we considered the discrete character of the independent variable. By adopting this view, *we are able to stick consistently and precisely to the actual framework provided by digital computer arithmetic.*

Our first goal is to utilize Liverani's results to prove convergence for the Ulam scheme associated with the 'true' digital image of the original function. This image will be constructed on an $N \times M$ mesh, $N$ and $M$ describing the arithmetic precision in the $x$ and $y$ directions, respectively (thus the roundoff errors will be of order $1/N$, $1/M$). One of our main results — in fact, Theorem 3 from subsection 4.4 — is that $(N/M) \to 0$ is sufficient (and most likely also needed) to achieve convergence, which indicates that Liverani's scheme provides a *qualitatively correct* model of well-constructed real computer simulations. What is more, this agreement is at the same time also *quantitatively precise* as far as an error estimate of the same Theorem for finite $N, M$ is concerned. By considering the more precise model, we provide a justification for Liverani's approach, and simultaneously we provide an algorithm which can be implemented on any computer. (We note that although convergence requires $(M/N) \to 0$, the $N = M$ case is of considerable mathematical interest since it guarantees the Markov property of the piecewise linear approximation and we will investigate this case separately in subsection 3.3).

One might expect that the suggested, more precise interpretation of roundoff errors will result in a less elegant mathematical construction. This expectation is partially fulfilled, since the proof of convergence, while relying heavily on Liverani's strong results, is at some points technically more involved. On the other hand, it turns out that the 'true' digital scheme, including all roundoff errors, provides an interesting and absolutely new interpretation (cf. subsection 3.2): the probability transition matrix associated to the Ulam scheme defined for the 'true' digital image is identical with the same matrix for a Markov chain, which is constructed as a special ('micro'-) *random perturbation* of the discrete map. This interesting connection enables us to compute the invariant density associated with the Ulam scheme with considerably less effort. The second, equally important goal of the paper is to introduce this perturbation method and show its connection to Ulam's scheme.

Let us explain the idea behind our micro-perturbation. It is well-known ([10]) that the quality of the simulation can be radically improved by adding random noise to the system. Earlier research (see [10] and references there) concentrated on methods where the order $\mathcal{O}(\epsilon)$ of the random perturbation exceeded the order of roundoff errors $\mathcal{O}(1/N)$, $1/N$ denoting the smallest number differing from zero in the computer arithmetic.

Although the random perturbation improves the model in the statistical sense, it necessarily distorts ('smears out') the original map, hence it is desirable to reduce it to the necessary minimum. The special random perturbation associated with the Ulam scheme is on the scale $\mathcal{O}(1/N)$ of the roundoff errors. As opposed to previous approaches, our random perturbation is sensitive to the local behaviour of the map; the key observation is that the amount of information lost in the discretization is -in first approximation- proportional to the derivative of the function, so the range of the added random noise should have the same order as the lost information.This 'micro-perturbation' introduced in subsection 3.1 minimizes the distortion of the original map, in fact, this is a perturbation with *minimal range* capable of reproducing the discretized version of the invariant density function. Since we approach the problem from the point of view of real computations, we not only prove convergence but also provide rigorous, essentially effective error bounds for finite $N, M$.

In a recent study ([13]) Liverani listed some fundamental questions about the computability of ergodic maps. In this paper we interpret 'computability' in the strict sense of doing operations with integers. In subsection 5.2 we give non-rigorous estimates for our method to make possible its comparison with the computation effort of different schemes. We prove our results for expanding maps on the unit interval and formulate conjectures on some generalizations.

The main thrust of the paper is to obtain new mathematical insight by utilizing strong results from earlier work combined with our ideas and theorems, also, our goal is to obtain an 'optimal' practical tool to be used in simulations. This tool connects two, apparently distinct approaches: Ulam's scheme and the random perturbation approach.

Section 2 contains the basic assumptions and definitions, it introduces the Perron-Frobenius operator and the associated Ulam scheme. Section 3 describes the micro-perturbation, and *as a main result it shows its identity to the Ulam scheme.*The last subsection formulates results concerning the $N = M$ quadratic grid. Section 4 describes *the second main result: a rigorous error bound for the density function obtained via the micro-perturbation.*The proof utilizes several strong results from the literature. Section 5 describes the diadic map as a rather transparent example and provides non-rigorous comparison to the direct solution of the linear system. Section 6 summarizes the results and points out possible generalizations and conjectures.

The paper is structured in such a way that the readers primarily interested in the application of the new algorithm can concentrate on subsections 2.1, 2.2, 3.1 and on sections 5 and 6.

# 2 Assumptions and definitions

## 2.1 Assumptions about the continuous map $f$

We investigate maps of the unit interval $f : [0,1] \to [0,1]$ with independent variable $x \in [0,1]$. Assuming certain properties (like piecewise smooth, expanding, topologically mixing, Markov) implies the existence of a unique, absolutely continuous invariant measure $\mu(x)$ with corresponding density function $\rho(x)$. We will introduce the polygonal approximation $\bar{f}_N(x)$, which, for sufficiently high $N$, inherits several of the the mentioned properties and generates the density function $\bar{\rho}_N(x)$. These fundamental concepts are defined below in more detail.

We assume about the map $f : [0,1] \to [0,1]$ that it is

**(PS)** piecewise $C^2$-smooth, i. e. $\exists a_0 = 0 < a_1 < \ldots < a_l = 1$ such that $f$ is $C^2$-smooth in every $I_i = (a_i, a_{i+1})$ (and its extension to each $I_i^{cl}$ is also $C^2$; $i = 0, 1, \ldots, l-1$).

**(E)** expanding, i. e. $\exists \alpha \in (0,1)$, such that $|f'(x)| \geq 1/\alpha$ for every $x \in \cup_{i=0}^{l-1} I_i$.

**Note:** In almost all derivations the assumption $\alpha < 1$ will be satisfactory, except for Theorem 3, where we will need $\alpha < 1/2$, because in the proof we utilize Lemma 4.1 of Liverani [13]. This latter assumption of course holds for a sufficiently high iterate of a map satisfying **(E).** .

Under these conditions, the map $f$ is known to have at most a finite number of absolutely continuous $f$-invariant measures $\mu$ (for details, weaker conditions, further references see [1], [13], [20], [22]:we note that since [13] is a recent comprehensive survey, in many cases we do not cite the original sources, rather we refer to the survey paper.) For simplicity of notation, the map is also assumed to be piecewise increasing, but this requirement is absolutely not essential.

To ensure the uniqueness of the measure $\mu(x)$ we suppose

**(TM)** topological mixing, i. e. there exists an interval $I^* \subset [0,1]$ such that $f(I^*) = I^*$, and every orbit $\{f^n(x) : n \in \mathbf{Z}_+\}, x \in (0,1)$ eventually enters $I^*$, and finally $f|I^*$ is topologically mixing meaning that for every interval $J \subset I^*$ one can find an index $n$ such that $f^n(J) = I^*$.

Under this condition the unique density will be denoted by $\rho(x)$.

Finally, only in part of our discussion (see subsection 3.3) the Markov property of maps will also be used:

**(M)** $f$ is Markov, i. e. for every $i = 0, 1, \ldots, l-1$ $\exists$ $\beta_i \subset \{0, 1, \ldots, l-1\}$ such that $f(I_i) \subset \cup_{j \in \beta_i} I_j^{cl}$. (In our case, an equivalent formulation of Markovity is just: $f(+0), f(a_1 \pm 0), \ldots, f(a_{l-1} \pm 0), f(1-0) \in \{a_0, a_1, \ldots, a_l\}$.)

## 2.2 Assumptions about the discretized map $f_N$ and the polygon $\bar{f}_N$

The discretized map can be interpreted on the $N \times M$ lattice, where $N$ is the integer describing the arithmetic precision of the computer in the independent variable $x$ (the smallest value of $x$ different from zero is $1/N$), $M$ is the integer describing the arithmetic precision of the function values $f(x)$. It might appear natural to assume $N = M$, however, this could severely handicap the simulation. On a quadratic grid the derivative of a polygonal line can only have integer values and thus it will differ by $\mathcal{O}(1)$ quantities from $f'(x)$, no matter how dense the grid. In other words, the approximating polygon is a 'zigzag' line, with large relative angles. In order to obtain convergence in the first derivative, we need

$$\lim_{N \to \infty} \frac{N}{M} = 0, \tag{1}$$

which can be guaranteed by defining $M$ as a super-linear function of $N$, e.g. $M = N \log N$. For simplicity, and observing computer arithmetic (see below) we will assume $K = M/N$ =integer throughout the paper.

If we want to realize the above described $N \times M$ grid on a digital computer, careful attention must be payed to the actual arithmetic of the CPU. Current processors, using the IEEE Standard 754-1985 [15] represent 'real'-type variables in logarithmic form as binary fractions, consequently the mesh on $[0, 1]$ is far from being uniform. Nevertheless, on any interval $[2^k, 2^{k+1}]$ the mesh is uniform, so one has to transform the map onto such an interval [17]. The different mesh densities $N$ and $M$ can be realized by defining $x$ as a single precision variable and $f(x)$ as a double precision one, resulting in $M = 2^{2k}, N = 2^k, M = N^2$. Both the uniformization of the mesh and the construction of two different mesh densities can be also performed by the direct truncation of the variables after prescribed number $(\log_2(N), \log_2(M))$ of binary digits, which is equivalent to using fixed-point variables instead of floating-point ones (cf. [17]). As mentioned, due to the binary representation we always have $M = 2^k, N = 2^l$, however, we do not use this fact.

The discretization will depend on the two mesh densities $M$ and $N$, however, in the forthcoming discussion we regard $M$ as a fixed function of $N$ (satisfying (1)) and use only the single subscript $N$. After these preliminaries, we define the discretized map as

$$f_N(\frac{i}{N}) = \frac{1}{M} \left[ M f(\frac{i}{N}) \right], i = 0, 1, 2, \ldots, N, \tag{2}$$

where [ ] denotes the integer part of a real number. We also define the difference operator on $f_N$ as

$$df_N(\frac{i}{N}) = N \left( f_N(\frac{i+1}{N}) - f_N(\frac{i}{N}) \right), i = 0, 1, 2, \ldots N-1. \tag{3}$$

Since $f$ is piecewise increasing, the discrete derivative should be positive. Nevertheless, close to each discontinuity of $f(x)$, $df_N(i/N)$ will be typically non-positive for one value of $i$. To eliminate this discrepancy, we introduce the corrected discrete derivative $Df_N$:

$$Df_N(\frac{i}{N}) = \begin{cases} df_N(\frac{i}{N}) & if \quad df_N(\frac{i}{N}) > 0 \\ df_N(\frac{i-1}{N}) & otherwise. \end{cases} \tag{4}$$

As mentioned before, the discrete map produces a finite set of orbits with finite periods. The information about $f(x)$, carried by the point-set $f_N(i/N)$ is rather limited in many other respects as well: $f(x)$ may take arbitrary values between the meshpoints $i/N$ (observing the expanding property (E)). One could interpolate different functions $\bar{f}_N(x)$ between the points $f_N(i/N)$; the simplest assumption would be 0th order: $\bar{f}_N(x) = f_N(i/N), i/N \le x < (i+1)/N$, however, this piecewise constant function would not inherit the expanding property (E) of $f(x)$. The simplest interpolating function which does satisfy this condition (for sufficiently large $N$) is the piecewise linear interpolation, yielding the polygonal line

$$\begin{array}{rcl} \bar{f}_N(x) & = & \left(f_N(\frac{i}{N}) + Df(\frac{i}{N})(x - \frac{i}{N})\right) \mod 1, \\ x & \in & I_i, i = 0, 1, 2, \ldots N-1, \end{array} \tag{5}$$

where $a \mod 1 = a - [a]$ and $I_i$ denotes the interval $\frac{i}{N} \le x < \frac{i+1}{N}$.

The latter truncation is necessary, since otherwise, due to the corrected discrete derivative $Df_N$, at the discontinuities of $f(x)$ the polygon could exit the unit square.

We can imagine the unit square subdivided into $N$ vertical and $M$ horizontal stripes. As the polygon $\bar{f}_N(x)$ traverses the $i$th vertical stripe, simultaneously it traverses

$$s_i = Df_N(i/N)\frac{M}{N} \tag{6}$$

horizontal stripes. We will utilize this approach in subsection 3.2 when we tie the Ulam scheme to the random perturbaton.

As we will show later, for sufficiently high $N$, $\bar{f}_N(x)$ inherits the expanding (E) and topological mixing (TM) property of $f(x)$, resulting the existence of the unique density $\bar{\rho}_N(x)$.

### 2.3 The Perron-Frobenius operator $P_f$

The fundamental tool for the rigorous study of piecewise expanding maps is the so-called transfer operator (also called the Perron-Frobenius or the Ruelle-Perron-Frobenius operator) defined as follows

$$P\phi(x) = P_f\phi(x) = \sum_{fy=x} \frac{\phi(y)}{|f'(y)|} \qquad (0 \le x \le 1)$$

The spectral properties of this operator over the Banach space $\mathcal{B}_{\mathcal{BV}}$ of functions of bounded variation supplied by the norm

$$\|\phi\|_{BV} = \mathrm{Var}(\phi) + |\phi|_1 \tag{7}$$

are well-understood. Below we present in a nutshell the most important properties of the transfer operator, necessary for our discussion.

It is suitable to work in two Banach spaces of functions: $\mathcal{B}_{BV}$ and $\mathcal{L}_1$. Of course, $\mathcal{B}_{\mathcal{BV}} \subset \mathcal{L}_1$ and $\| \ \|_1 \le \| \ \|_{BV}$ (cf. (7)), and moreover, these spaces are adapted, meaning that every $\mathcal{B}_{BV}$-bounded set is precompact in $\mathcal{L}_1$, i.e. for every set of functions who's $\| \ \|_{\mathcal{BV}}$ norms are bounded, one can select a subsequence which converges in the $\| \ \|_1$ norm (cf. [2]). It is appropriate to consider transfer operators $P : \mathcal{B}_{BV} \to \mathcal{B}_{BV}$ with the norm

$$|||P||| = \sup\{|Ph|_1 \Big| h \in \mathcal{B}_{\mathcal{BV}}, \|h\|_{\mathcal{BV}} \le 1\} \tag{8}$$

The theory is based on the so called Doeblin-Fortet or Lasota-Yorke inequality valid for maps satisfying (PS) and (E): there exist $C_1, C_2 > 0$ such that for every $n \in \mathbf{N}$ and for every $h \in \mathcal{B}_{BV}$

$$\|P^n h\|_{BV} \leq C_1 \alpha^n \|h\|_{BV} + C_2 \|h\|_1, \tag{9}$$

see [9], [13] and [20]. We note that under our conditons the essential spectral radius $R_{\text{ess}}$ of the transfer operator $P$ is not larger than $\alpha$. For obtaining the constants $C_1$ and $C_2$ in a simple form, Liverani [13] assumes $\alpha < 1/2$ and obtains $C_1 = 2, C_2 = 1$.(cf. [9] and [13], footnote 7.)

For us the following properties will be important:

Property 1. (PS) and (E) imply that apart from a finite number of eigenvalues lying on the unit circle $|z| = 1$ the spectrum of $P$ lies in a disc of radius $\alpha < 1$. In addition, there exists at least one absolutely continuous $f$-invariant density $\rho(x)$.

Property 2. If, moreover, (TM) is also satisfied, then the only eigenvalue of $P$ on the unit circle is 1, it is simple and its eigenfunction is the only absolutely continuous $f$-invariant density $\rho(x)$.

### 2.4 Ulam's scheme as an operator: $U_{f,N}$

As early as 1960, S. Ulam already addressed the computational aspects of dynamical systems. In doing so, he introduced a discretization scheme for the calculations of maps, continuous in space. Later several authors took up the investigation of this scheme, most notably they were interested in the quality of this approximation, in particular, in its speed of convergence.

For simplicity, we will only recall Ulam's scheme in the case of expanding maps of the interval.

Mathematicians prefer to consider operators in suitable Banach spaces, so for definiteness, we use the Banach space $\mathcal{B}_{BV}$ of functions of bounded variation defined by the norm $\| \ \|_{BV}$ (cf.(7)). The idea is to discretize functions in space by using the averaging operator $\mathcal{A}_N : \mathcal{B}_{BV} \to \mathcal{B}_{BV}$ defined by

$$\mathcal{A}_N \Phi(x) = N \int_{\frac{i}{N}}^{\frac{i+1}{N}} \Phi(y) dy \text{ whenever } \quad x \in I_i.$$

Then Ulam's scheme is the operator $U_{f,N} : \mathcal{B}_{BV} \to \mathcal{B}_{BV}$ defined via

$$U_{f,N} \Phi = \mathcal{A}_N P_f \mathcal{A}_N \Phi$$

The polygonal approximation $\bar{f}_N(x)$ (introduced in (2)) is piecewise linear on each subinterval $I_i : \frac{i}{N} \leq x < \frac{i+1}{N}$. As pointed out by Liverani, in this case $U_{\bar{f}_N,N}$ is completely described by the probability transition matrix

$$P_{ij} = \frac{m(\bar{f}_N^{-1} I_j \cap I_i)}{m(I_i)}, \tag{10}$$

where $m$ is the Lebesgue measure. $P$ is also called the *area overlaps Markov chain* because $P_{ij}$ is the proportion of Lebesgue area in $I_i$ which is mapped into $I_j$ under one iteration of $\bar{f}_N$. Clearly, we have $\sum_{j=1}^N P_{ij} = 1$ for each $i$ (cf. [7], [13]). As Liverani points out, the Ulam operator has also one eigenfunction associated with the single eigenvalue on the unit circle, we will denote this invariant density for $U_{\bar{f}_N,N}$ by $\bar{\rho}_{N,U}$. This density can be computed by solving the linear system for $\rho_j$:

$$\sum_{i=1}^{N} P_{ij} \rho_i = \rho_j \tag{11}$$

# 3 The micro-perturbation $f_N^*$ and its interpretation as an Ulam scheme $U_{\bar{f}_N,N}$

In this section we will introduce a random micro-perturbation $f_N^*$ associated with $f_N$ and show that it is, in essence, identical to the Ulam operator $U_{\bar{f}_N,N}$, so the corresponding invariant densities $\rho_N^*$ and $\bar{\rho}_{N,U}$ are identical as well. We discuss the special case $N = M$ where Ulam's operator becomes identical with the Perron-Frobenius operator and for this case we will prove the identity of $\rho_N^*$ and $\bar{\rho}_N$.

## 3.1 Definition of the micro-perturbation $f_N^*$

In the further discussion we assume (PS), (E), (M) and (TM). The key observation motivating the suggested random perturbation is that the amount of information lost during discretization is -in first approximation- proportional to the discrete derivative $df$: the larger $df$, the more intermediate lattice points are omitted while the independent variable moves by $1/N$. The added random perturbation will be represented by the random variables $\epsilon_i^t$, the superscript referring to the discrete time of the iteration, the subscript to the discrete variable of the map; the random variables $\epsilon_i^t$ are independent in both respect. Based on these considerations the randomized discrete map $f_N^*$ is the following:

$$f_N^*(\frac{i}{N}) = \left(f_N(\frac{i}{N}) + \epsilon_i^t\right) \bmod 1, \tag{12}$$

where

$$P_i = P(\epsilon_i^t = 0) = P(\epsilon_i^t = \frac{1}{M}) = \ldots = P(\epsilon_i^t = \frac{1}{N} Df_N(\frac{i}{N}) - \frac{1}{M}) = \frac{N}{MDf_N(\frac{i}{N})}. \tag{13}$$

Returning to the idea of stripes, introduced in subsection 2.2, we can interpret $P_i$ as the probability of hitting one horizontal stripe if we are in the $i$th vertical stripe.

## 3.2 The micro-perturbation interpreted as an Ulam scheme

Our goal is to prove

**Theorem 1** $\rho_N^*(i/N) = \bar{\rho}_{N,U}(x)$ *if* $x \in I_i$.

Proof of Theorem 1: Let us return to the probability transition matrix $P_{ij}$ (10) associated with the Ulam scheme $U_{\bar{f}_N,N}$, and determine the elements of this matrix. We identified the number $s_i$ of horizontal stripes traversed by $\bar{f}_N(x)$ for $x \in I_i$ in (6). Now we introduce $s_{ij}$ as the number of stripes traversed by $\bar{f}_N$ for $x \in I_i, \bar{f}_N(x) \in I_j$. The length of the horizontal projection of this segment can be calculated by simple proportionality as

$$m(\bar{f}_N^{-1} I_j \cap I_i) = \frac{s_{ij}}{s_i} \frac{1}{N} = \frac{s_{ij}}{MDf_N(i/N)}, \tag{14}$$

and, using (10) and $m(I_i) = 1/N$, the overlap ratio can be expressed as

$$P_{ij} = s_{ij} \frac{N}{MDf_N(i/N)}. \tag{15}$$

Now we can construct a probability transition matrix $T$ for the randomized scheme $f_N^*$: the element $T_{ij}$ will denote the probability $P(f_N^*(i/N) \in I_j)$. As we noted after equation (13), the

probability of hitting one horizontal stripe is $P_i$, since there are $s_{ij}$ horizontal stripes involved, we have, based on (13):

$$T_{ij} = s_{ij} P_i = s_{ij} \frac{N}{MDf_N(i/N)}, \tag{16}$$

so we have clearly

$$P_{ij} = T_{ij}, i, j \in \{1, 2, \dots N\}. \tag{17}$$

Since the two matrices coincide, the eigenfunctions coincide as well. Q.e.d.

### 3.3 Exact results for the $N \times N$ mesh

If our goal is to obtain $\lim_{N\to\infty} \bar{\rho}_{N,U} \to \rho$ then (as we will see later) it is essential that (1) is observed. On the other hand, the $M = N$ special case offers interesting mathematical aspects since the polygon $\bar{f}_N(x)$ constructed on a $N \times N$ quadratic mesh automatically satisfies the conditions of the Markov property **M.** This leads to the following result:

**Theorem 2** *$\bar{\rho}_N(x) = \rho^*_N(\frac{i}{N})$, if $N = M$ and $x \in I_i$.*

The first step is to prove

**Lemma 1** *If $f'(x)$ is constant in each interval $I_i : i = 0, 1, \dots, l-1$, then $\rho(x)$ is also piecewise constant on the same intervals as $f'(x)$.*

**Proof of Lemma 1:**

Instead of the R-P-F operator $P_f$ let us consider its discretized version over $\mathbf{R}^l$

$$P^{(d)}\phi(i) = \sum_{fI_j \supset I_i} \frac{\phi(j)}{|f'(y)|}$$

where $y \in I_j$ is an arbitrary point such that $f(y) \in I_i$; $i = 0, 1, \dots, l-1$. From our conditions it is evident that $P^{(d)}$ is correctly defined. Moreover, analogously to Properties 1. and 2. from section 2, it is a contraction in $\mathbf{R}^l$, whose only eigenvalue on the unit circle is 1 with a unique positive eigenvector

$$P^{(d)}\phi^{(d)} = \phi^{(d)} \tag{18}$$

satisfying the normalization $\sum_{i=0}^{l-1} \phi^{(d)}(i) = 1$. Obviously, the function $\phi(x) = \phi^{(d)}(i)$ if $x \in I_i$ is a solution of the fixpoint equation $P\phi = \phi$, therefore it is the unique $f$-invariant probability density. Q.e.d.

The polygon $\bar{f}_N(x)$ satisfies the conditions of Lemma 1: it inherited the expanding property from $f(x)$ and it has piecewise constant derivative on equal subintervals of length $1/N$ the images of which are unions of such subintervals, so the Markov property is also guaranteed. Hence we have

**Corollary 1** *$\bar{\rho}_N(x)$ is piecewise constant on all sub-intervals $I_i$.*

Since $\bar{\rho}_N(x)$ is piecewise constant, the averaging operator $\mathcal{A}_N$ is identity, so we have $\bar{\rho}_N \equiv \bar{\rho}_{N,U}$, and, via Theorem 1, we arrive at the statement of Theorem 2. Q.e.d.

As we can see, for the $N = M$ case the Ulam operator $U_{\bar{f}_N,N}$ becomes identical with the Perron-Frobenius operator $P_{\bar{f}_N}$.

# 4 The convergence result: bound for $\|\bar{\rho}_{N,U} - \rho\|_1$

In this section we investigate the $N \to \infty$ limit. However, we not only want to prove convergence for the density function $\bar{\rho}_{N,U}$, our goal is to give a rigorous bound on the error $\|\bar{\rho}_{N,U} - \rho\|_1$. These densities are the eigenfunctions belonging to the eigenvalue 1 of the corresponding operators, so our first goal is to estimate the distance of the densities assuming that the distance $|||U_{\bar{f}_N} - P_f|||$ of the operators in the norm (8) is available; this will be performed in subsection 4.1. In the next step, following Liverani's ideas, we decompose the estimate of the operators' distance into two difference norms and estimate $|||U_{\bar{f}_N} - P_{\bar{f}_N}|||$ and $|||P_{\bar{f}_N} - P_f|||$ independently; this is the topic of subsection 4.2. The latter estimate requires the distance $d(f, \bar{f}_N)$, which we supply in subsection 4.3. Finally, the main result on the error bound will be formulated in subsection 4.4.

Most of these estimates follow from a series of strong results obtained in [13], [7], [8] and [9] which we will summarize concisely, supplemented with our own considerations, in particular, the distance estimate in subsection 4.3.

## 4.1 The distance $\|\bar{\rho}_{N,U} - \rho\|_1$ of the densities estimated based on the distance $|||U_{\bar{f}_N} - P_f|||$ of the corresponding operators

**Proposition 1** *Consider pairs of maps $f$ and $g$ satisfying (PS) and (E) and fix $f$. [9], [13]. Then we have*

1. *Let $r \in (\alpha, 1)$ and $\eta = \frac{\log \frac{r}{\alpha}}{\log \frac{1}{\alpha}} > 0$. If $\lambda$ is an isolated eigenvalue of $P_f$ with $|\lambda| > r$ and $\delta$ satisfies $B_\delta(\lambda) \cap \sigma_\alpha(P_f) = \{\lambda\}$, then $\exists$ a constant $K_1 = K_1(\delta, r)$ such that*

$$|||\Pi_f^{\lambda,\delta} - \Pi_g^{\lambda,\delta}||| \leq K_1 |||P_f - P_g|||^\eta$$

   *Here $B_\delta(\lambda) = \{z \in \mathbf{C} \big| |z - \lambda| \leq \delta\}$ and, in general,*

$$\Pi_f^{\lambda,\delta} = \frac{1}{2\pi i} \int_{\partial B_\delta(\lambda)} \frac{1}{z - P_f} dz$$

   *denotes the projection of $P_f$ onto the part of its spectrum lying in $B_\delta(\lambda)$.*

2. *If $d(f, g)$ and $\delta$ are small enough, then*

$$Rank\ \Pi_f^{\lambda,\delta} = Rank\ \Pi_g^{\lambda,\delta}$$

3. *For $\lambda = 1$, 2 is true and 1 is also valid if we change the right hand side to $K_1 |||P_f - P_g||| \ \Big|\log |||P_f - P_g|||\Big|$.*

From statement 3 it actually follows that

$$\|\bar{\rho}_{N,U} - \rho\|_1 = K_1 u |\log u| \tag{19}$$

where

$$u = |||U_{\bar{f}_N} - P_f|||. \tag{20}$$

and $K_1$ is explicitly given in [2],[7],[8]. Our next goal is to identify $u$.

### 4.2 The distance $|||U_{\bar{f}_N} - P_f|||$ between the operators

Following Liverani [13], will decompose the distance into two parts:

$$|||U_{\bar{f}_N} - P_f||| \leq |||U_{\bar{f}_N} - P_{\bar{f}_N}||| + |||P_{\bar{f}_N} - P_f||| \tag{21}$$

The first term on the left-hand side can be identified by using Lemma 4.1 of Liverani ([13]), yielding after simple calculations using the constants given in subsection 2.3 and considering $\alpha \leq 1/2$ (c.f. the note in subsection 2.1, after the definition of the expanding property):

$$|||U_{\bar{f}_N} - P_{\bar{f}_N}||| \leq \frac{1}{N}(1+2\alpha)(1+\frac{1}{2}\alpha) \leq \frac{2.5}{N}. \tag{22}$$

We identify the second term on the left-hand side of (21) by using

**Proposition 2** *Consider pairs of maps $f$ and $g$ satisfying (PS) and (E) and fix $f$ [9], [13]. Then we have*

$$|||P_f - P_g||| \ \leq 12d(f,g)$$

*where*

$$d(f,g) = \inf \quad \Big\{\kappa > 0 \mid \exists A \subset I, \exists \omega : I \to I \ \ s.th.\ m(A) > 1-\kappa,\ \omega \ \ is\ a\ diffeo,$$
$$f|_A = g \circ \omega|_A \ and \ \sup_{x\in I}|\omega(x) - x| < \kappa,\ \sup_{x\in I}|1/\omega'(x) - 1| < \kappa \ \Big\}.$$

By setting

$$d_0 = d(f, \bar{f}_N), \tag{23}$$

Proposition 2 yields

$$|||P_f - P_{\bar{f}_N}||| \ \leq 12d(f,\bar{f}_N) = 12d_0. \tag{24}$$

Now we can express the original distance as

$$u = |||U_{\bar{f}_N} - P_f||| \leq 12d_0 + \frac{2.5}{N}, \tag{25}$$

and the only remaining task is to compute an upper bound for $d_0$; this will be done in the next subsection.

### 4.3 The distance $d_0$ between the maps $f$ and $\bar{f}_N$

Our goal is to find the distance $d(f,\bar{f}_N)$ defined (23), so we have to show the existence of a diffeomorphism $\omega : A \to A$, where $A \subset I, 1 - m(A) < \kappa$, such that

$$f|_A \circ \omega = \bar{f}_N|_A, \tag{26}$$

with $|\omega(x) - x| < \kappa$, $|1/\omega' - 1| < \kappa$ and we have to provide an explicit estimate $\kappa$ in terms of the discretization parameters $N, M$ and other properties of $f(x)$.

The set $\bar{A}_N$ is defined as the union of all sub-intevals $I_i : [i/N, (i+1)/N]$ on which $f$ has no discontinuities, so we have

$$m(\bar{A}) = 1 - \kappa,\ \kappa = \frac{J}{N}, \tag{27}$$

where $J$ denotes the number of discontinuities of $f(x)$. Let

$$L = \sup |f''(x)|, x \in \bar{A},$$

so the Taylor series of $f(x)$ at $x = x_0$ can be written as

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \varphi_1 \frac{L}{N^2}, |\varphi_1| \leq 1. \tag{28}$$

(In the forthcoming formulas we will use similar constants $|\varphi_j| \leq 1$.) Analogously, the derivative can be expressed as

$$f'(x) = f'(x_0) + \varphi_2 \frac{2L}{N}. \tag{29}$$

Let $x_0 = i/N,\ f'(x_0) = b > 1,\ \bar{f}'_N(x_0) = a > 1,\ X = x - x_0$;in the sub-interval $X \in [0, 1/N]$ we can express these functions as

$$\begin{array}{rcl} \bar{f}_N(X) & = & aX + C \\ f(X) & = & bX + \varphi_3 \frac{1}{M} + C + \varphi_1 \frac{L}{N^2}, \end{array} \tag{30}$$

where $\varphi_3 = M(f(x_0) - \bar{f}_N(x_0))$, $C = \bar{f}_N(x_0)$. Based on (26) and (30) we can express the diffeomorphism as

$$\omega(X) = \bar{f}_N^{-1}(f(X)) = \frac{b}{a}X + \varphi_4 \frac{L}{N^2} + \varphi_5 \frac{1}{M} \tag{31}$$

where $\varphi_4 = \varphi_1/a$, $\varphi_5 = \varphi_3/a$. We are interested in

$$\omega(X) - X = (\frac{b}{a} - 1)X + \varphi_4 \frac{L}{N^2} + \varphi_5 \frac{1}{M}, \tag{32}$$

where $(b/a - 1)$ can be expressed based on Figure 1 as

$$\frac{b}{a} - 1 = \varphi_8 \frac{L}{N} + \varphi_9 \frac{N}{M}, \tag{33}$$

where $\varphi_8 = \varphi_1\varphi_7,\ \varphi_9 = \varphi_6\varphi_7,\ \varphi_6 = M(f((i+1)/N) - \bar{f}_N((i+1)/N)), \varphi_7 = 1/a$, see Figure 1. Using (32) and (33) we can write

$$\omega(X) - X = \varphi_{11} \frac{2L}{N^2} + \varphi_{12} \frac{2}{M} < 2\left(\frac{1}{M} + \frac{L}{N^2}\right), \tag{34}$$

where $\varphi_{11} = (\varphi_8\varphi_{10} + \varphi_4)/2$, $\varphi_{12} = (\varphi_9\varphi_{10} + \varphi_5)/2$, $\varphi_{10} = NX$. Similar considerations yield

$$|\frac{1}{\omega'(X)} - 1| = |\frac{\bar{f}_N(\omega(X))}{f'(X)}| = \varphi_7\varphi_{13} \frac{3L}{N} + \varphi_7\varphi_6 \frac{N}{M} < \frac{3L}{N} + \frac{N}{M}, \tag{35}$$

where $\varphi_{13} = (2\varphi_2\varphi_1)/3$.

This diffeomorphism $\omega$ is valid in a 'large' sub-interval of each interval $I_i : [i/N, (i+1)/N]$ belonging to $\bar{A}$, at each end a segment of length $1/M$ has to be chopped off; the union of the remaining 'large' sub-intervals will be denoted by $A$, for illustration see Figure 1. The Lebesgue measure of $A$ can be expressed as

$$m(A) = 1 - \kappa,\ \kappa = \frac{J}{N} + \frac{2(N - J)}{M} \tag{36}$$

The three expressions (34), (35) and (36) are simplified (and also minimized in leading order) if we set $M = N^2$. Our next goal is to choose $d(f, \bar{f}_N)$ as the largest of the three estimates, yielding

$$d_0 = d(f, \bar{f}_N) = max\{\frac{2L+2}{N^2}, \frac{3L+1}{N}, \frac{J+2}{N}\} = max\{\frac{3L+1}{N}, \frac{J+2}{N}\} \tag{37}$$

which can be evaluated for any given $J, L, N$.

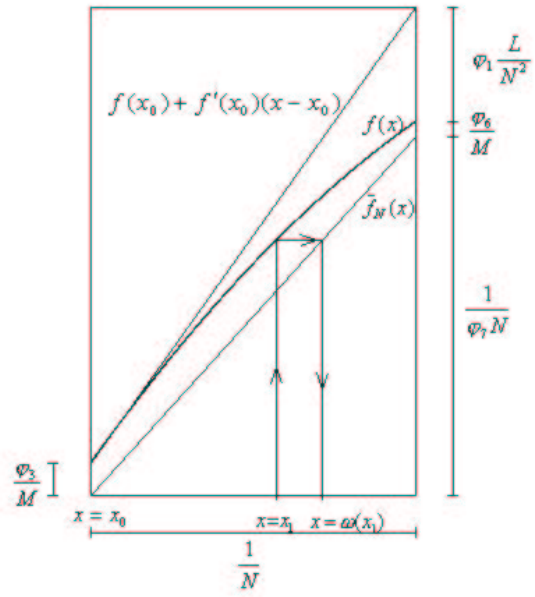


Figure 1: The linearized local behaviour of $f$ and $\bar{f}_N$ and the diffeomorphism $\omega$ on $x \in [i/N, (i+1)/N]$.

### 4.4 The final result for the error bound

Using all the preliminary computations in subsections 4.1, 4.2 and 4.3 we can now formulate our final result on the error bound associated with the computed density function $\bar{\rho}_{N,U}$:

**Theorem 3**

$$\|\bar{\rho}_{N,U} - \rho\|_1 = K_1 u |\log u|$$

*where*

$$u \leq 12 d_0 + \frac{2.5}{N},$$

$$d_0 = max\{\frac{3L+1}{N}, \frac{J+2}{N}\},$$

$$L = \sup |f''(x)|, x \in \bar{A},$$

*and $J$ denotes the number of discontinuities of $f(x), x \in I$.*

The proof of the theorem is contained in section 4, in particular we used equations (19) (25) and (23) to construct the statement.

# 5 Example and comparison to other methods

## 5.1 The diadic map

We illustrate our results on the diadic map

$$f(x) = 2x \bmod 1, \tag{38}$$

which is also known as the roughest qualitative approximation of the Lorenz attractor ([5]). The map (38) obviously satisfies the criteria for the properties (PC),(E) and (TM) and generates on $I$ the uniform invariant density $\rho(x) \equiv 1$. Not only is (38) piecewise linear, its derivative is an integer, so $\forall N : f(x) \equiv \bar{f}_N(x)$, if we assume $M = 2^k N$, and consequently $\forall N : \rho(x) \equiv \bar{\rho}_N(x) \equiv \bar{\rho}_{N,U}(x)$. Due to the very transparent structure of the map, the $\rho^*_N(i/N) \to \bar{\rho}_N(x)$ correspondence can be demonstrated by an elementary argument, without referring to Theorem 2, even for the otherwise problematic $N = M$ case: Since $\forall i : Df_N(i/N) = 2$, any trajectory of length $k$, belonging to $f^*_N$ is a binary tree, the $2^k$ endpoints of which cover the discrete unit interval $\bar{I} = \{0, 1/N, 2/N, \ldots, (N-1)/N, 1\}$ $K_N = [2^k/N]$ times without gaps, in addition they cover once the remaining fraction of length $F = 2^k - NK_N$. Thus the expected number of iterates in each cell differs from the 'uniform' value $2^k/N$ maximally by 1, the density quickly converges to the uniform one as $k \to \infty$. (Note that for the practically interesting case for real-type computer arithmetic $N = 2^{k_0}$, we have convergence already at $k = k_0$, cf [17].)

Up to now we were concerned with the comparison of the continuous map $f$ and the *randomly perturbed* discrete map $f^*_N$ and we proved that under certain conditions the latter can -in the statistical sense- approximate the previous one. One can only hope for weaker results when looking at the *deterministic* discrete dynamics of $f_N$: as mentioned previously, it differs radically from the continuous one, since the discrete system possesses only a finite number of orbits with finite periods. We will use the diadic map to illustrate this behaviour and to compare it with the continuous case.

The structure of the discrete dynamics depends sensitively on $N$: as reported in [4] and [10], different arithmetic precision yielded drastically different global behaviour. There are several options to measure the similarity between the discrete and continuous dynamics. In the cited works this measurement was not rigorous, the authors selected some trajectories based on individual considerations and observed the long-term behaviour. The rigorous measurement could be done on different principles:

(a) The 'strict' principle measures the statistics of the endpoints belonging to a (large) collection of trajectories with fixed length, the initial values of which are confined to an interval of length $< 1/N$. In the continuous case, the endpoints will be spread over $I$ according to the density $\rho(x)$, while in the discrete system, due to the deterministic iteration, each endpoint will be identical, *regardless of $f$ and $N$* .

(b) The 'liberal' principle measures the statistics of one trajectory, the starting point of which is selected at random, statistical event is each step of iteration. In this case the results depend sensitively on $f$ and $N$.

A systematic analysis of the second, 'liberal' similarity measure is carried out in [3], where the diadic map is investigated in detail; we emphasize some interesting features below.

The global behaviour of $f_N$ can be described either by a (degenerate) transition matrix $T$, where each of the transition probabilities is either 0 or 1, or with an oriented graph, which we will call the *symbolic dynamics.* The structure of this graph is very informative: e.g. in case of the diadic map, if $N = M = 2^k$, then the graph is a binary tree, promising the worst possible statistics under the 'liberal' definition, since instead of being spread out, all trajectories converge

to the same point. The 'optimal' statistics, most similar to the uniform density of the continuous system, would be produced if (almost) all cells would be included in a single cycle. Although in real (computer) applications one has very limited choice of $N$, and we have always $N = 2^k$, it is of theoretical interest to ask whether there exists such $N$ to which the 'optimal', cyclic graph belongs. A more detailed number-theoretic study of the problem ([21]) reveals that the answer is positive, if $N$ is a prime of the form $N = 4p + 1$, where $p$ is also prime, $N - 1$ cells will form a single cycle. However, the structure of the graph changes abruptly as $N$ is varied: we illustrate $N = M = 651, 652, 653, 654$ in Figure 2. Observe the cyclic structure for $N = 653 = 4 * 163 + 1$ and the appearance of large number of small cycles for $N = 651$. For more details on the evolution of the graph structure see [3] and [18].

The example of the diadic map illustrates that the deterministic discrete system has little in common with the continuous map, the graph describing the structure of the discrete map is highly sensitive to the applied arithmetic precision. The simple structure of the diadic map allowed us to make exact statements based on number theory. Although such precise formulations are hardly possible in case of other maps, the above mentioned qualitative features are rather general and explain the phenomena reported in [4] and [10].

### 5.2 Comparison to other methods

The perturbation described in (12) can be added to the computer code with very few lines, it provides a useful tool for simulations [17]. In general, it is an open question how this method compares to previously suggested ones (cf. [10],[13]) in terms of computing time and efficiency. In particular, comparison to the direct solution of the linear equation system (11) would be of interest; as Kifer [10] observes, statistical simulations based on random perturbation are generally believed to be faster than the direct solution of the linear equation system. Liverani [12] raised a more specific question: is it possible to define under which conditions the iteration of the randomized scheme (12-13) is more efficient than the direct solution of (11)? This comparison is meaningful, since the iteration of the scheme produces a statistical approximation of $\rho^*$, the linear equation yields $\bar{\rho}_{U,N}$, and Theorem 1 guarantees that the two objects are identical. A rigorous answer to Liverani's question would be difficult not only from the mathematical point of view (estimate on the $\mathcal{L}_1$-distance of the statistical approximation of $\rho^*$, $\bar{\rho}_N$ and $\rho$), it would involve many parameters (e.g. the number of floatig point operations used by the random number generator) so the final result would be hard to interpret. Our goal is *not* to give a rigorous answer, rather, we would like to provide some rough estimates which emphasize only the essential parameters and give an idea for computations.

We will assume that the application of the scheme (12) is equivalent to the random sampling of the variable $\bar{\rho}_N$. (This is, strictly speaking, not true if we regard one iterative step as one statistical event ('liberal' principle of statistical measurement, as defined in subsection 5.1), if we adopt the 'strict' principle and regard the endpoint of a trajectory of fixed (short) length as a statistical event then we are somewhat closer to the assumption.) Under this assumption, using one iterative step as a statistical event (and neglecting the above mentioned computational effort to produce a random number), the computational 'cost' $C_E$ of one statistical event is

$$C_E = F, \tag{39}$$

where $F$ is the number of floating point operations (FLOPs) needed to evaluate $f(x)$. The $\chi^2$ test for goodness of fit [11] requires approximately

$$E \geq \frac{10}{p_{min}} \tag{40}$$

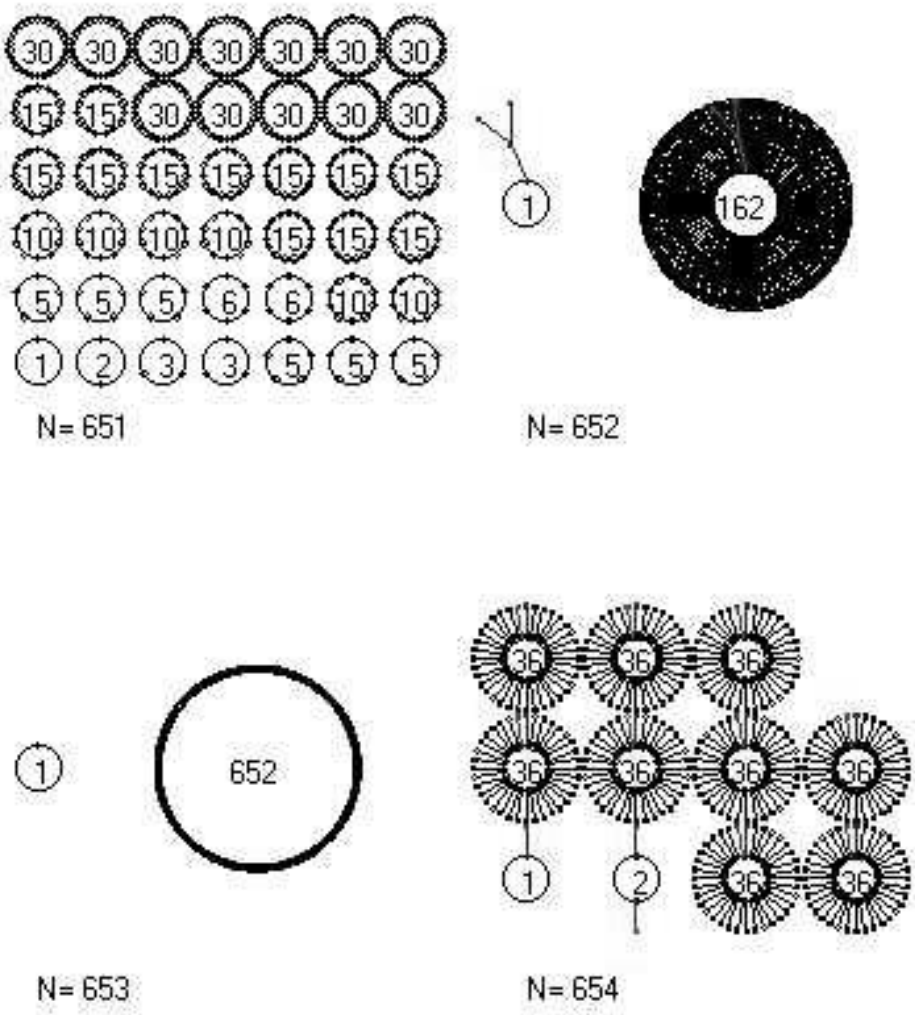


Figure 2: Graphs characterizing the discretized diadic map at $N = M = 651, 652, 653, 654$. Circle indicate cyclic parts of the graph, length of cycle written inside circle. Fixed point corresponds to cycle with length 1. Software by R. Tóth.

statistical events, where $p_{min}$ denotes the smallest probability associated with $\rho^*$. (If $\rho^*$ is uniform, then $p_{min} = 1/N$.) Combining (39) with (40) provides an estimate for the necessary computation effort $C_{stat}$ in FLOPs:

$$C_{stat} = EC_E = \frac{10F}{p_{min}}. \tag{41}$$

Since we are interested in the eigenvector $\rho_i$ in (11), corresponding to the largest eigenvalue $\lambda_1 = 1$, $\rho_i$ can be succesively approximated by computing the powers of $P_{ij}$. This requires roughly

$$C_{lin} = FN + kN^2 \tag{42}$$

FLOPs, the first term corresponds to the computation of the coefficients, the second term to $k$ iterative steps, each involving $N^3$ FLOPs. The convergence of this iteration is controlled by the ratio $\frac{\lambda_i}{\lambda_1}$ of the eigenvalues. For simplicity, we will assume that a minimum of $k = 10$ iterative steps are needed to reduce the error to a tolerable level. For large $N$, this value of $k$ is certainly not sufficient. Comparing the two estimated computation efforts $C_{lin}$ and $C_{stat}$, we can make the following observation:

*If $p_{min} > F/N^2$ then the statistical simulation is faster.*

This condition implies that the solution of the linear system (11) is faster if we apply *very* low arithmetic precision and the function $f(x)$ is *very* difficult to evaluate and/or the sought density function $\bar{\rho}_N$ is *very* far from uniform. Such a situation could arise if $f(x)$ is a Poincarè map and it is evaluated via the integration of a very complicated differential equation and we are only interested in a precision sufficient for graphics: assuming $F = 1000$, $N = 10000$ and $p_{min} = (1/N)/10 = 0.00001$ the two methods are approximately equal. We must not forget that both estimates $C_{stat}$ and $C_{lin}$ were based on a very modest requirement concerning the accuracy. Comparison based on *equal errors* would be much more difficult, however, we believe based on numerical experience that it would yield similar results. Our rough estimates indicate that if we move to higher $N$, the statistical simulation is the only feasible choice.

## 6 Summary and concluding remarks

In this paper we studied the relationship between the statistical properties of piecewise continuous, expanding maps and their discretized versions. The discretization was performed both in the independent variable *and* in the function value, so we investigated an exact model of schemes generated by the application of finite computer arithmetic precision; we believe that this is a new approach. We investigated the following objects:

- $f(x)$, a piecewise continuous, expanding map on the unit interval,
- $f_N(i/N)$, the discretized version of $f(x)$, generated on an $N \times M$, $M > N$ lattice by a computer with maximal precision $1/M$,
- $\bar{f}_N(x)$, the polygonal line defined by the points $f_N(i/N)$,
- $U_{\bar{f}_N,N}$, the Ulam operator associated with $\bar{f}_N$,
- $f_N^*(i/N)$, the randomized version of $f_N(i/N)$,
- $\rho(x)$, the density function generated by $f(x)$,
- $\bar{\rho}_{N,U}(x)$, the density function generated by the Ulam scheme $U_{\bar{f}_N,N}$,

- $\rho_N^*(i/N)$, the discrete density, generated by $f_N^*(i/N)$.

The relationships between these objects are illustrated in scheme (43):

$$\begin{array}{ccccccc} f(x) & & & & & \to & \rho(x) \\ & \nwarrow\nwarrow & & & & & \uparrow\uparrow \\ \downarrow & & \bar{f}_N(x) & \to & U_{\bar{f}_N,N} & \to & \bar{\rho}_{N,U}(x) \\ & \nearrow & & & & & \uparrow \\ f_N(i/N) & \to & f_N^*(i/N) & & & \to & \rho_N^*(i/N) \end{array} \tag{43}$$

In this scheme single arrows indicate that one object defines/generates the other one, double arrows indicate some kind of convergence. Observe that only the three objects in the third row are 'real' in the sense that they appear directly in the computer simulation. The objects in the upper two rows exist only as far as we can obtain information on them from the computer simulation or, possibly, from the physical system. We can observe that the polygonal approximation serves as an intermediate tool for the approximation of the continuous system. As we can see, our main goal was to approximate $\rho(x)$, based on information contained in $f_N(i/N)$, a finite set of integers.

The step from the 'real' objects in the third row to the 'intermediate' objects in the second row is performed in section 3: we proved that the density $\rho_N^*$, generated by the suggested randomization $f_N^*$, is identical with the invariant measure $\bar{\rho}_{U,N}$ of the Ulam scheme associated with the polygonal approximation of the original function. At a given level of discretization this is the most one can ask for.

The next step was to show the relationship between the 'intermediate' second row and the objects of real physical interest in the first row; section 4 was devoted to this task. The proof of Theorem 3, stating the convergence $\bar{\rho}_{N,U} \to \rho$, required strong results from the literature, combined with our considerations. In fact, Theorem 3 claims more than convergence, it establishes an error bound, implying that with the suggested scheme the original invariant density function $\rho(x)$ can be approximated with *arbitrary precision* (for sufficiently large $N$).

It is interesting to observe that for the latter result we required unequal mesh densities $N$ and $M$ on the horizontal ad vertical axis respectively, with $M/N \to 0$ as $N \to \infty$. The computer offers the same maximal mesh density in both directions, so using unequal densities (e.g. by using single and double precision arithmetic) is equivalent to omit certain information, nevertheless, this proves to be the efficient strategy. This is explained by considering that although the omitted data carries additional information on the function values, it blocks the convergence in the derivative. Omitting this data is analogous to 'averaging' the oscillating values of the discrete derivative and provides a smaller error in the approximated density, even for finite $N$.

Scheme 43 can serve as an illustration to Liverani's comment [13]: '...the linear approximation is used to control the nonlinearity, while the Ulam scheme is used to control the lack of the Markov property.' In our approach, the linearization $\bar{f}_N(x)$ served both the purpose of controlling the nonlinearity, and, due to the discretization in both directions, also as an *exact* model of finite computer arithmetic. The role of the Ulam scheme is specially highlighted in subsection 3.3, where we investigate the $N = M$ quadratic grid. This discretization guarantees the Markov property of $\bar{f}_N(x)$, consequently the Ulam scheme 'vanishes' and becomes identical with the Perron-Frobenius operator.

If we have $f = \bar{f}_N$, then of course the randomized map produces the *exact* density $\rho(x)$; this is the situation in case of the diadic map if we apply a mesh with $M = 2^k N$, discussed in subsection 5.1. We also illustrated the extreme sensitivity of the deterministic discrete map to the arithmetic precision $1/N$: in case of the diadic map this could be explained in exact number theoretic terms,

however, the sensitivity itself is characteristic of more complicated maps as well. Our results show that the randomly perturbed discrete map is much more robust: small variation of $N$ will not change the statistical properties.

Subsection 5.2 dealt with the comparison to the direct solution of the linear equation system. Based on non-rigorous estimates and arguments we found that the choice between the methods depends primarily on the computational complexity of $f(x)$ and the applied precision $N$. If the former is very large while the latter is very small, the direct solution of the linear system is preferable, otherwise our randomized iteration scheme proves to be the feasible choice.

Although we only proved our results for expanding maps, we expect this algorithm to be helpful in the computation for non-expanding maps as well, in fact, we expect this algorithm to perform in such cases better compared to other perturbation methods. In case of such maps the statistical behaviour can depend infinitely sensitively on parameters [6], [16]; needless to say that based on finite information one can not hope to capture this sensitivity. On the other hand, stable fixedpoints and their neighbourhood are *not* affected by our random perturbation ($Df_N = 0$), so one can hope to capture stable attractors reliably in a sufficiently fine discretization. In fact, simple inspection of equation (11) reveals that stable fixed points are preserved for sufficiently high $N$. This property is not shared with other random perturbations which are insensitive to the local properties of the map. Based on the above considerations we believe that the suggested method can be a useful tool when one tries to answer *at fixed arithmetic precision* the fundamental questions raised by Liverani ([13]): whether the map is ergodic, mixing, or expanding.

## Acknowledgement

The authors thank Yuri Kifer, Carlangelo Liverani, Imre Szeberényi and Réka Tóth for their helpful suggestions and remarks. Special thanks to Gerhard Keller for his comments on subsection 5.2. The support of OTKA grant T031744 and the Bolyai Fellowship (GD) and OTKA grants T26176 and T32022 (DSz). is gratefully acknowledged.

## References

[1] V. Baladi. Transfer operators and decay of correlations. In *Advanced Series in Nonlinear Dynamics*, volume 16, Singapore, 2000. World Scientific.

[2] M. Blanc and G. Keller. Random perturbations of chaotic dynamical systems: stability of the spectrum. *Nonlinearity*, 11:1351–1364, 1998.

[3] G. Domokos. Digital modelling of chaotic motion. *Studia Sci. Math. Hung.*, 25:323–341, 1990.

[4] G. Haller and G. Stépán. Micro-chaos in digital control. *J. Nonlinear Science*, 6:415–448, 1996.

[5] P.J. Holmes and F.C Moon. Strange attractors and chaos in nonlinear mechanics. *Trans. ASME Ser.E, J. Appl. Mech.*, 50:1021–1032, 1983.

[6] M. Jakobson. Absolutely continuous invariant measures for one-parameter families of one-dimensional maps. *Comm. Math. Phys.*, 81:39–88, 1981.

[7] R. Keane, M. Murray and L.S. Young. Computing invariant measures for expanding circle maps. *Nonlinearity*, 11:27–46, 1998.

[8] G. Keller. Stochastic stability in some chaotic dynamical systems. *Monatsh. Math.*, 94:313–333, 1982.

[9] G. Keller and C. Liverani. Stability of the spectrum for transfer operators. *Ann. Scuola Norm. Sup. Pisa Cl Sci (4)*, 28:141–152, 1999.

[10] Yu. Kifer. Computations in dynamical systems via random perturbations. *Discrete and Continuous Dynamical Systems*, 3:457–476, 1997.

[11] G.A. Korn and T. M. Korn. *Mathematical handbook for scientists and engineers.* McGraw-Hill Book Company, New York, St. Louis, etc., second edition, 1968.

[12] C. Liverani. Oral communication to the authors. 2001.

[13] C. Liverani. Rigorous numerical investigation of the statistical properties of piecewise expanding maps - a feasibility study. *Nonlinearity*, 14:463–490, 2001.

[14] A. Rényi. Representations of real numbers and their ergodic properties. *Acta Math.Acad. Sc. Hung.*, 8.:477–493, 1957.

[15] W. Stallings. *Computer organization and architecture.* IEEE Computer Society, 1985.

[16] H. Thunberg. Unfolding of chaotic unimodal maps and the parameter dependence of natural measures. *Nonlinearity*, 14:323–337, 2001.

[17] L.I. Tóth. Digital simulation of chaotic systems. *Budapest University of Technology and Economics, Scientific Student Competition*, 2002.

[18] R. Tóth. Discrete dynamics of the diadic map. *Budapest University of Technology and Economics, Scientific Student Competition*, 2002.

[19] S. Ulam. *Problems in modern mathematics.* Interscience Publishers, 1960.

[20] M. Viana. Stochastic dynamics of deterministic systems. In *Col. Bras. de Matematica, lecture notes*, 1997.

[21] I.M. Vinogradov. *Foundations of number theory (in Russian).* State Editing House for Theoretical Technical Literature, Moscow, Leningrad, 1949.

[22] L.S. Young. Statistical properties of dynamical systems with some hyperbolicity. *Annals of Mathematics*, 147:585–650, 1998.