Math. Struct. in Comp. Science (2014), vol. 24, iss. 3, e240304, 25 pages.
doi:10.1017/S0960129512000758

Probability, statistics and computation in dynamical systems

STEFANO GALATOLO[†¶], ISAIA NISOLI[‡¶]
and CRISTÓBAL ROJAS[§¶]

[†]*Dipartimento di Matematica Applicata, Università di Pisa,*
via Buonarroti 1, Pisa, Italy
Email: galatolo@dm.unipi.it
[‡]*Universidade Federal do Rio de Janeiro,*
Instituto de Matemática Cidade Universitària – Ilha do Fundão,
Rio de Janeiro 21945-909, Brazil
Email: nisoli@im.ufrj.br
[§]*Departamento de Matemáticas, Universidad Andres Bello,*
República 220, Santiago, Chile
Email: cristobal.rojas@unab.cl

Received 15 November 2011; revised 5 June 2012

We discuss some recent results related to the deduction of a suitable probabilistic model for the description of the statistical features of a given deterministic dynamics. More precisely, we motivate and investigate the computability of invariant measures and some related concepts. We also present some experiments investigating the limits of naive simulations in dynamics.

1. Introduction

This paper is a contribution to the debate raised by this special issue of *Mathematical Structures in Computer Science*. Quoting the Preface to this special issue:

'Kolmogorov's probabilistic syntax has been considered capable of quite satisfactorily hosting and formally organising any specific *factual* probabilistic problem. Furthermore, hardly anybody has seemed to be troubled by the fact that, in any given factual probabilistic situation, in order to calculate predictions, we have to specify numerically the individual probabilities of each event involved in that *particular* factual situation, whereas Kolmogorov's theory of probabilities contains exclusively *general* constraints on a probability measure, which are quite independent of any particular probabilistic situation.' (from the Preface)

In other words, while Kolmogorov's axiomatisation of probabilities is suitable for the mathematical analysis of the concept of probability, it does not tell us how to calculate the individual probabilities that rule the statistics of a specific factual probabilistic situation.

[†] Stefano Galatolo wishes to thank 'Gruppo Nazionale per l'Analisi Matematica, la Probabilitá e le loro Applicazioni'(GNAMPA, INDAM), for financial support.
[‡] Isaia Nisoli wishes to thank the CNPQ (Conselho Nacional de Pesquisas Cientìficas e Tecnologicas CNPq-Brazil), and IM-UFRJ (Instituto de Matematícas of Universidade Federal do Rio do Janeiro).
[§] The research of Cristóbal Rojas was partially supported by project DI-134-12/R (Universidad Andres Bello) and project Fondecyt 11110226 (government of Chile).
[¶] The authors are grateful for the CINECA Award N. HP10C42W9Q, 2011 for the provision of high performance computing resources and support.

Depending on the point of view taken, this problem has several sides, with subtle different interpretations, each of which requires its own approach to be discussed. For example, in some simple cases (such as coin tossing), symmetry considerations might be used to give convincing numerical values of the relevant probabilities. However, in most practical situations, simple symmetry considerations cannot be applied. Instead, a deeper study of the observed system can help in the creation of a suitable probabilistic model. One possibility is to collect statistical data about the factual situation, and then try to extrapolate from that data to the probabilistic model that best fits the data. However, this empirical approach is always reinforced by assumptions made on the underlying phenomenon.

There are other possibilities when the probabilistic situation arises from a physical phenomenon governed by a deterministic law. Indeed, most real situations related to classical physical phenomena can be nicely modelled by dynamical systems. As we shall see, there are several classes of phenomena for which a careful study of the underlying dynamical system can help us develop a good probabilistic model – at least, good for answering some important classes of questions (for example, for forecasting the averages of given observables).

The procedure for creating the probabilistic model from the deterministic law (modelled as a dynamical system) will correspond to the computation of a *physical* invariant measure. We shall see that while this procedure can be implemented in an important class of cases, there are some subtleties involved that might, in some cases, prevent the procedure from being implemented at all.

More precisely, if we model our phenomenon as a dynamical system, the resulting mathematical problems are related to ergodic theory and its computability aspects. In fact, a significant part of the ergodic theory of dynamical systems is motivated by this and similar problems. Later in the paper, we shall explain how ergodic theory offers tools for selecting, for any given dynamical model (arising from a physical phenomenon governed by a deterministic law), a natural probability measure describing its behaviour. This process of taking the deterministic law producing the factual situation and determining a probability distribution that describes it can be seen as a way of relating abstract Kolmogorov probabilities to factual 'probabilistic' situations, provided, of course, we have access to the underlying deterministic law.

The remaining question is whether the numerical values of the selected probability distribution can be calculated, say, by a modern computer. This question can be approached in a rigorous way using computable analysis. We shall present a brief introduction to this theory, together with some recent mathematical results concerning the above-mentioned problem of determining the computability of invariant measures. It will be seen that, again in theory, while there exists several tools, which can by used in many interesting classes of systems to compute the interesting equilibrium probabilities[†], there are also some pathological probabilistic situations for which there is no algorithm

[†] These results are rather abstract, and while they state that some specific tasks can be achieved in principle, the actual implementation of the procedure often involves overcoming several practical complications, such as the computational resources required.

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

to compute any of the equilibrium distributions describing it. On the other hand, if we assume that the underlying system is a little noisy, the situation improves dramatically: not only do all the interesting measures become computable, but they can be computed efficiently (*viz.* in polynomial time).

There are a number of other problems related to the question of determining the computability of the statistical properties in dynamical systems, such as computing the speed of convergence of the above-mentioned time averages, and the computation of points that behave as 'random' (typical) initial conditions. We shall briefly discuss some recent results related to these issues later in the paper.

In contrast to the rigorous computation of the statistical behaviour of a system (the computation of a suitable invariant mesasure), and unlike the empirical/statistical observation mentioned earlier, there is the use of *naive* simulations on a computer (ignoring rounding and other errors), which is still the most-performed method in the applied sciences. It is a fascinating topic because naive simulations seem to work reasonably well in many cases, despite the rounding errors, which when coupled with the typical instability of chaotic systems can, in principle, destroy any similarity to the behaviour of system that is meant to be simulated. We shall discuss some issues related to the limits of this approach later in the paper.

1.1. *Organisation of the paper*

In Section 2, we give a brief introduction to ergodic theory, including a description of how to select, from the deterministic description of the system, the 'physically relevant' probability distributions. In Section 3, we briefly describe how Turing computability can be used to determine rigorously whether continuous objects (such as real functions or probability distributions) can be computed, or not. In Section 4, we review some of the recent results dealing with the question of the computability of relevant invariant measures. In particular, we shall describe:

(i) a general method allowing the computation of the physical measure in many interesting cases;
(ii) an example of a system where no invariant measure can be computed; and
(iii) a result supporting the conjecture that non-computability phenomena are not robust.

In Section 5, we review the above-mentioned results related to the rigorous computation of initial conditions exhibiting *typical* statistical behaviour (rigorous simulations). Finally, in Section 6, we present a few numerical experiments providing some simple illustrations of the problems that might arise when using non-rigorous computations, and the limits of naive simulation (with rounding errors).

2. Dynamical systems and invariant probability distributions

2.1. *Probability enters the dynamical model*

In this section we shall give a brief description of some elementary ergodic theory in discrete time dynamical systems as models of evolving physical systems. For a complete

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

treatment, see, for instance, Walters (1982) and Mañé (1987). A dynamical system consists of a metric space X representing all the possible states the system can ever occupy, and and a map $T : X \to X$ representing the dynamics. In principle, such a model is *deterministic* in the sense that complete knowledge of the state of the system, say $x \in X$, at some initial time, entirely determines the future *trajectory* of the system: $x, T(x), T(T(x)), ...$ However, despite the deterministic character of the system, there are many interesting situations for which it is impossible to predict any particular feature about any specific trajectory. For example, if we are interested in knowing some feature (such as energy, temperature...) of such a system that can be quantified, and thus represented by a function f assigning to each state $x \in X$ the corresponding quantity $f(x)$, then it is impossible to predict (compute) the sequence of values $f(x), f(T(x)), f(T^2(x)), ...$ This is a consequence of the famous sensitivity to initial conditions (chaotic behaviour) and the impossibility of making measurements with infinite precision (approximation): two initial conditions that are very close to each other (so they are indistinguishable by physical measurement) may diverge in time, rendering the true evolution unpredictable in practice.

On the other hand, if we look instead at the averages of the function f along the orbit

$$A_f(x) = \lim_{n\to\infty} \frac{f(x) + f(T(x)) + ... + f(T^{n-1}(x))}{n},$$

they tend to be quite stable and, more importantly, *they do not depend on* x.

Suppose for the moment that the point x is such that the above limit exists for each continuous, compactly supported function f. Then, the map $f \to A_f(x)$ defines a continuous functional $C_c(X) \to \mathbb{R}$. By the Riesz representation theorem, this functional can be represented by a measure μ_x[†]. This measure is *invariant* for the dynamics, $\mu(T^{-1}A) = \mu(A)$ for each measurable set A, and represents the statistical limit distribution of the trajectory of the point x in the space X.

Therefore, by changing our point of view and just looking for the probability distribution related to the system's statistics, we may get some useful information about the system's behaviour.

A common situation is one in which the phase space can be divided into *regions* exhibiting qualitatively different statistical limiting behaviours. Within each region, all the initial conditions give rise to a trajectory that approaches an 'attractor', on which the limiting dynamics take place (and which can be quite complicated). Thus, different initial conditions within the same region may lead in the long term to configurations that have quite different pointwise behaviours, but that are identical in a statistical sense. Any probability distribution supported in the region will also evolve in time, approaching a limiting *invariant* distribution, which is supported in the attractor, and which describes the dynamics of the *equilibrium* situation in statistical terms. Note that, depending on the scale at which we are measuring time steps, the *limiting equilibrium* regime may be approached in experiments after a short period of time. For instance, in some particle

[†] Referring again to the Preface of this special issue, this shows how measures (the Kolmogorov approach to the concept of probability) come into the game very naturally when considering this kind of question, and that they are naturally and unavoidably related to the empirical statistical facts (in the frequentist approach).

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

systems, the collisions occur so rapidly that the equilibrium situation is attained almost instantly with respect to our scale of time.

A major inconvenience with such a mathematical description is that there are often several probability measures that are invariant for a given dynamics (and it is often an infinite dimensional convex set). In other words, the theory says that several limiting equilibrium behaviours may occur, whereas, in practice, we usually observe only one, or a finite set. Our task then is to understand how to select the prevalent equilibrium behaviour we observe in practice, and, moreover, to compute the probability invariant distribution associated with it.

2.2. *Ergodic systems and physical probabilities*

The basic idea is to select the invariant measures that emerge as a limiting statistical behaviour for a *large* set of initial conditions, where the meaning of *large* will be made precise later in this section.

A first result to help in this is the ergodic theorem. We say that a system is *ergodic* when it is undecomposable in the sense that if $T^{-1}(A) = A$ up to a set of zero measure, then A has total or null measure.

Theorem 2.1 (pointwise ergodic theorem). If a system is ergodic, then for any $f \in L^1(X, \mu)$,

$$\lim_{n\to\infty} \frac{S_n^f(x)}{n} = \int f \, \mathrm{d}\mu, \tag{1}$$

for μ almost each x, where

$$S_n^f = f + f \circ T + \ldots + f \circ T^{n-1}.$$

In words, this theorem says that time averages of an observable f computed along μ-typical orbits coincide with phase space averages with respect to μ. That is, the measure μ describes the statistical behaviour of the system when started from almost every (with respect to μ) initial point x. However, as we said earlier, for a given dynamics T, there could be many possible different ergodic measures (think about all the measures supported on periodic orbits), each of which having its own notion of 'almost everywhere', and the question then arises as to whether the measure μ we are looking at describes a significant part of the system. It might perfectly well be that, for instance, the measure μ is concentrated in a negligible part of the system, and thus describes a probabilistic behaviour that is very unlikely to occur, so we almost never observe it.

We shall now explain a possible approach to the task of selecting the measures that are more important for describing the 'real, physical, observed' behaviour of the system. A physical invariant measure is a measure that arises as a time average for a *large* set of initial conditions, where large here is in the sense of the natural, uniform measure that can be put on the configuration space X of our system before starting the dynamics. For example, when X is a manifold, this measure is usually taken to be the Lebesgue measure (equidistribution).

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

We say that a point x *belongs to the basin of attraction* of an invariant measure μ if (1) holds at x for each bounded continuous f. In words, this means that if we assume the initial state of the system to be x, then its probabilistic behaviour is described by μ (in a way similar to the description at the beginning of Section 2).

When X is a manifold (possibly with boundary), *a physical measure*[†] is an invariant measure μ whose basin of attraction has positive Lebesgue measure. This says that if the system starts at some *random* (with respect to equidistribution) initial condition, then there is a 'positive probability' that the statistical behaviour of the system is governed by μ.

Physical measures are often a very small (finite) subset of the set of all invariant measures. It turns out that physical measures also have several other good features, such as a certain stability under perturbations and noise – for more details and a general survey, see Young (2002). Moreover, they are often associated with attractors. The presence of an attractor implies that a whole basin of attraction (with positive Lebesgue measure) is mapped by the dynamics to this attracting set (in some sense, the physical invariant measure in chaotic systems is where the Lebesgue measure is sent after many iterations), which therefore supports a physical measure.

We shall give some concrete examples of physical measures for non-trivial systems in Section 4.

A physical measure is thus a reasonable and natural notion of probability when applied to a physical phenomena modelled by a dynamical system, provided we are interested in the behaviour of averages, or other long term features of the dynamics.

Given this, it becomes important for us to have a way to compute this measure from the description of the elementary rules of the system. Once we have this computation, we can say things like 'the numerical value of the probability that the event A occurs is p', without making further assumptions on the model.

3. Turing-rigorous computability over continuous spaces

In this section we present a very brief introduction to the concepts of computable analysis needed to understand the meaning of the following results. We shall mainly restrict our attention to computability on $\mathbb{R}$, though much more general definitions and results can be given – see Brattka *et al.* (2008) and Weihrauch (2000) or Galatolo *et al.* (2011b) for more on the general theory.

3.1. *Computability over* $\mathbb{R}$

The intuitive notion of an *algorithmic* or *effective procedure* on symbolic objects was formalised in the 1930s in several different and independent ways (by Church, Kleene, Turing, Post, Markov...), the most popular being Turing Machines. All these formalisations turned out to be equivalent: they compute the same functions from $\mathbb{N}$ to $\mathbb{N}$. The class of functions thus defined is now called the class of *computable functions*. The subsets

[†] Physical measures coincide to a large extent with SRB measures – see Young (2002).

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

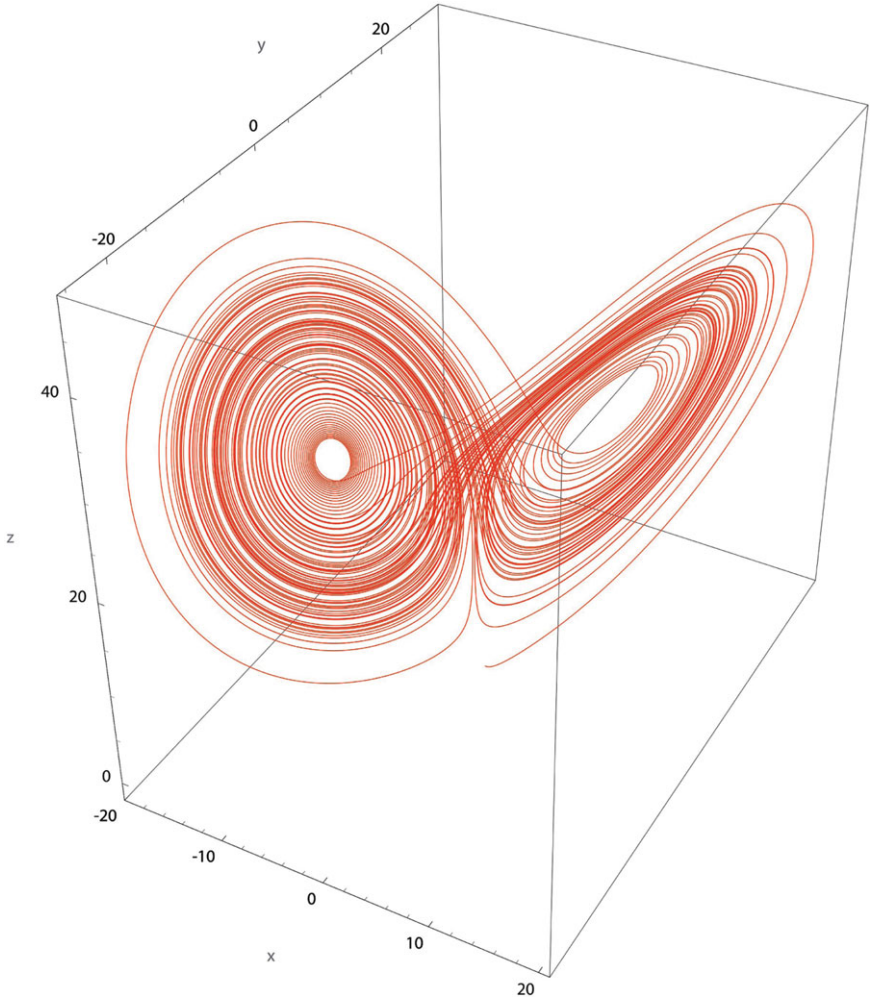

Fig. 1. (Colour online) The Lorenz attractor – the unique physical measure of the system is supported on this attractor

of natural numbers that can be enumerated by a computable function are said to be *recursively enumerable* (r.e.). Computable functions can also operate on other countable sets by identifying them with natural numbers through a convenient fixed numbering. In his seminal paper, Turing introduced the *computable reals* as those $x \in \mathbb{R}$ for which there is a Turing Machine that, upon input n, outputs the first n digits in the binary expansion of x. The following is an equivalent, but more contemporary, definition.

Definition 3.1. A real number x is *computable* if there exists a computable function $\varphi : \mathbb{N} \to \mathbb{Q}$ satisfying

$$|x - \varphi(n)| < 2^n \qquad \forall\ n \in \mathbb{N}.$$

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

The computability of sets is defined in a similar way: specifically, by approximating them by rational intervals. Let $\mathcal{I}_{\mathbb{Q}}$ denote the set of all open rational intervals (which is naturally a numbered set).

Definition 3.2. An open set $A \subset \mathbb{R}$ is *effectively open* if there is a computable function $\varphi : \mathbb{N} \to \mathcal{I}_{\mathbb{Q}}$ such that

$$A = \bigcup_{n \in \mathbb{N}} \varphi(n).$$

Computable real functions are now simply defined as *computably continuous* as follows.

Definition 3.3. A real function $f : \mathbb{R} \to \mathbb{R}$ is *computable* if the pre-images of open rational intervals are all effectively open sets, in a uniform way. That is, there is a computable function $\varphi : \mathcal{I}_{\mathbb{Q}} \times \mathbb{N} \to \mathcal{I}_{\mathbb{Q}}$ such that

$$f^{-1}(I) = \bigcup_{n \in \mathbb{N}} \varphi(I, n).$$

A function f is *computable on* $D \subseteq X$ if, under the above assumptions,

$$f^{-1}(I) \cap D = D \cap \bigcup_{n \in \mathbb{N}} \varphi(I, n).$$

Remark 3.4. It can be shown that a function f is computable if there is a computer program that computes $f(x)$ in the following sense: on input $\epsilon > 0$, the program, during its run, asks the user for approximations of x, and eventually halts and outputs a rational point $q \in \mathbb{R}$ satisfying $|f(x) - q| < \epsilon$. This idea can be formalised using, for example, the notion of *oracle computation*. The resulting notion coincides with that given in the previous definitions.

The following notion will play an important role in proving existence results.

Definition 3.5. A compact set K is *computably compact* if there is an algorithm A_K that, upon input $\varepsilon \in \mathbb{Q}$, outputs a finite list of rational intervals $I_1, ..., I_n$ of diameter at most ε such that

$$K \subset \bigcup_{1 \leqslant i \leqslant n} I_i.$$

That is, the output of $A_K(\varepsilon)$ is a finite ε-cover of K.

Examples of computably compact sets are $[0, 1]$ and $\{x\}$, where x is a computable point. Computable compactness in more general spaces is defined analogously (Galatolo *et al.* 2011a).

Finally, the computability of probability measures is defined by considering measures for which there is an algorithm that can compute integrals of simple functions.

Definition 3.6. A probability measure μ over $[0, 1]$ is *computable* if the function $i \mapsto \int p_i \, d\mu$ is computable, where $\{p_i\}_{i \in \mathbb{N}}$ is an enumeration of all polynomials $p_i : [0, 1] \to \mathbb{R}$ with rational coefficients.

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

It can be shown that a measure is computable in the above sense if and only if it can be approximated up to any precision (in some distance defined on the space of probability measures) by an algorithm producing 'finite' measures (think of clouds of points on the screen) – for more on this, see Hoyrup and Rojas (2009).

4. Computing invariant measures

In this section we review some of the results dealing with the question of whether the relevant invariant measures of a given system can be computed by an algorithm. As might be suspected, there is no general answer to the question and, as we shall see, the computability structure of the space of invariant measures can be intricate. In particular, we shall give:

(i) general methods allowing the computation of the physical measures in many interesting cases;
(ii) an example of a computable system for which NO invariant measure is computable; and
(iii) a result supporting the conjecture that if we suppose the system to be a little noisy, then all physical measures are always efficiently computable, and thus the system becomes statistically predictable.

4.1. *Positive results*

We begin by illustrating a simple idea that allows us to establish the computability of the invariant measure in the most simple case: uniquely ergodic systems. A system is said to be *uniquely ergodic* if it has a unique invariant measure. In particular, this measure must be ergodic and physical (in fact, the trajectory of every initial condition is described by this measure). For simplicity, we shall state the result over $[0, 1]$, but it holds in general spaces as well.

Proposition 4.1. If a computable system $T : [0, 1] \to [0, 1]$ is uniquely ergodic, then its invariant measure is computable.

Proof (idea). The key idea is to observe that, given a measure μ, we can algorithmically detect whether a measure λ is NOT invariant. Indeed, invariant measures must satisfy

$$\int T \, d\lambda = \int T \circ f \, d\lambda$$

for every continuous f. So, by computing the above integrals for a sufficiently rich family of functions, we can detect whether a given measure λ is not invariant. In this way, we can rule out all the non-invariant measures, and by compactness considerations, this allows us to locate the only remaining measure, which is what we are looking for. □

When the system is not uniquely ergodic, the mere detection of non-invariance is not sufficient to locate invariant measures. This is one of the difficulties in extending the above technique to more general cases. To overcome this obstacle, we have to find

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

complementary conditions that allow us to isolate the relevant invariant measure we are looking for, but ones that are also algorithmically detectable.

The so-called *transfer operator* provides a useful tool when trying to find these complementary conditions. A function measurable T between metric spaces naturally induces a function L_T between probability measure spaces. This function L_T is linear, and is called the transfer operator (associated with T). Measures that are invariant for T are fixed points of L_T.

Let us consider the space $PM(X)$ of Borel probability measures over X. The *transfer operator* $L_T : PM(X) \to PM(X)$ is defined as follows. If $\mu \in PM(X)$, then $L_T(\mu) \in PM(X)$ is such that

$$L_T(\mu)[A] = \mu(T^{-1}(A))$$

for each measurable set $A \subseteq X$. This operator describes how the measure is transported by the dynamics. For example, one can check that when δ_x is the delta measure centred in x, then $L_T(\delta_x) = \delta_{T(x)}$.

In the non-rigorous numerical approach[†] (which is commonly used in scientific computing) to the calculations of invariant measures (and many other dynamical quantities), the calculation is very often achieved by computing the fixed points (and some other spectral information) of the transfer operator acting on a suitable function space. The most applied and studied strategy is to find a suitable finite dimensional approximation of L_T (restricted to a suitable function space), which reduces the problem to the computation of the corresponding relevant eigenvectors of a finite matrix.

An example of this is done by discretising the space X by a partition A_i and replacing the system by a (finite state) Markov chain with transition probabilities

$$P_{ij} = \frac{m(A_i \cap T^{-1}A_j)}{m(A_i)}$$

where m is the Lebesgue measure on X (this is called the Ulam method, see, for example, Froyland (2007), Froyland (2001), Liverani 2001 and Dellnitz and Junge (2002)). Then, taking finer and finer partitions, we can, in some cases, show that the finite dimensional model will converge to the real one (and its natural invariant measure to the physical measure of the original system). In some cases there is an estimate of this speed of convergence (see, for example, Froyland (2001) for a discussion), but a rigorous bound on the error (and thus a completely rigorous computation) is known for only a few cases (piecewise expanding or expanding maps – see Liverani (2001) and Galatolo and Nisoli (2012).

In the rigorous framework, we can show that the transfer operator is a computable function, and thus explicit conditions can be given to ensure the computability of some of its fixed points.

[†] Non-rigorous here means calculations that do not take any particular account of rounding and discretisation errors – these include the calculations performed by most of the commonly used numerical software packages. This means that there is no effective estimate of the final distance between what is meant to be computed and what is actually computed.

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

Theorem 4.2 (Galatolo *et al.* 2011a, Theorem 3.2). Let T be computable on $[0,1] \setminus D$. Suppose there is a recursively compact set of probability measures $V \subset PM([0,1])$ such that for every $\mu \in V$, we have $\mu(D) = 0$. Then, every invariant measure isolated in V is computable. Moreover, the theorem is uniform in that there is an algorithm that takes as input finite descriptions of T and V and an ideal ball[†] in $PM([0,1])$ that isolates[‡] an invariant measure μ, and then outputs a finite description of μ.

It is clear that Proposition 4.1 is now a trivial corollary of Theorem 4.2. As we have already said, the main difficulty in the application of Theorem 4.2 is the requirement that the invariant measure we are trying to compute be isolated in V. In general, the space of invariant measures of a given dynamical system could be very large (in particular, an infinite dimensional convex subset of $PM([0,1])$), but there is often some kind of regularity that can be exploited to characterise an interesting invariant measure, and thus isolate it from the rest. Some examples of cases where this method can be applied are given later in the paper.

4.1.1. *Absolutely continuous invariant measures.* The absolute continuity of a measure is a property that can be used as a compact condition to rule out all the non-absolutely continuous measures. In systems possessing a unique such measure, the above method provides the following result.

Proposition 4.3. Let T be computable on $[0,1] \setminus D$ with $dim_H(D) < 1$, and suppose $([0,1], T)$ has a unique absolutely continuous invariant measure μ with bounded density. Then μ is computable (starting from T and a bound for the L^∞ norm of the invariant density).

Similar results hold for maps on manifolds (see Galatolo *et al.* (2011a) again). It is well known that interesting examples of systems having a unique absolutely continuous invariant measure (with bounded density as required) are the topologically transitive *piecewise expanding maps* on the interval or the *expanding maps* on manifolds.

Definition 4.4. A non-singular function $T : ([0,1], m) \to ([0,1], m)$ is said to be piecewise expanding if[§]:

(1) There is a finite set of points $d_1 = 0, d_2, ..., d_n = 1$ such that $T|_{(d_i, d_{i+1})}$ is C^2, and it can be extended to a C^2map on $[d_i, d_{i+1}]$.
(2) There are $c > 0, \lambda > 1$ such that $inf(T^n) \geqslant c\lambda^n$.
(3) T is topologically mixing[¶].

The following result holds for this class of maps.

[†] That is, a basic 'computable' ball, which is a generalisation of the intervals $I_{\mathbb{Q}}$ in Definition 3.2 – see Galatolo *et al.* (2011a).
[‡] The isolating ball is not required if the invariant measure is unique in V.
[§] Note that the definition of piecewise expanding maps frequently used in the current literature is slightly more general than the definition we give here.
[¶] A system is said to be topologically mixing if, given sets A and B, there exists an integer N such that, for all $n > N$, we have $f^n(A) \cap B \neq \varnothing$.

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

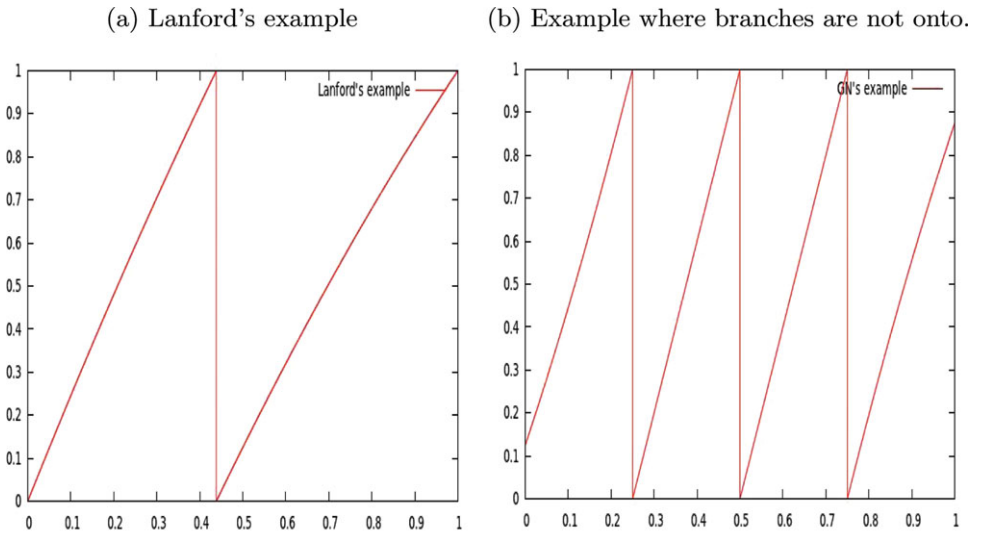

Fig. 2. (Colour online) Two examples of piecewise expanding maps

Proposition 4.5. Suppose a piecewise expanding map T and its first and second derivatives are computable on $[0,1]-\{d_1,...,d_n\}$. Suppose also that its extension to the closed intervals $[d_i, d_{i+1}]$ are computable. Then, the physical measure can be computed starting from a description of the system (the points d_i and the programs computing the map and its derivatives).

4.1.2. *Maximal entropy measures.* In some cases the same method can also be applied to compute the unique invariant measures that maximise entropy. A particular case of great interest is the so-called *quadratic family* over the complex numbers. This corresponds to the family of functions, indexed by $c \in \mathbb{C}$, of the form $z \mapsto z^2 + c$. Each quadratic function f_c has a distinguished invariant set, called the Julia set, which supports the chaotic part of the dynamics. Julia sets are among the mathematical objects that have attracted most attention from the scientific community, possibly because of the beauty of their fractal, self-similar structure. There are countless computer programs for visualising these sets on a computer screen up to any prescribed resolution: that is, one can zoom-in on a given region of the set as deeply as desired, observing more and more details of its structure. The ability to perform such a computation for a set corresponds exactly to the notion of *computable closed set*. Despite the variety of computer programs intended to *compute* the Julia set, Braverman and Yampolsky (2006) showed that there are computable parameters c for which the associated Julia set is NOT computable by any algorithm, so no image of these Julia sets can ever be produced on a computer. This result should be understood as a note of caution when using computer programs that allow for rounding errors.

This means that we cannot accurately visualise the set over which the chaotic dynamics take place, but what about their statistical distribution? This question makes sense since

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

for all $z \neq \infty$ and every continuous test function ψ, we have

$$\frac{1}{2^n} \sum_{w \in (f_c)^{-n}(z)} \psi(w) \underset{n \to \infty}{\longrightarrow} \int \psi d\lambda,$$

where λ is the *Brolin–Lyubich probability measure* (Brolin 1965; Lyubich 1982) supported on the Julia set J_c. This measure is characterised as being the one that maximises entropy. We can thus ask whether the the Brolin–Lyubich measure is computable. Even if $J_c = \mathrm{Supp}(\lambda)$ is not a computable set, the answer does not *a priori* have to be no. In fact, the following result holds.

Theorem 4.6 (Binder *et al.* 2011). The Brolin–Lyubich measure of computable quadratic polynomials is always computable.

This result also follows, essentially, from Theorem 4.2. The difficulty is once again in finding a recursively compact condition that characterises Brolin–Lyubich measure. This measure happens to be singular with respect to Lebesgue, so the absolute continuity argument does not apply. The key property allowing the application of Theorem 4.2 in this case is *balance*: a probability measure μ on $\mathbb{C}$ is said to be *balanced* with respect to a rational function R (of degree at least 2) if for every set $A \subset \mathbb{C}$ on which R is injective, we have

$$\mu(R(A)) = d \cdot \mu(A),$$

where d denotes the degree of R.

Computation of the Brolin–Lyubich measure is then reduced to showing that for probability measures, being balanced is a recursively compact condition. The proof of this fact is not very difficult, but it does require us to overcome some problems, such as the fact that the function $\mu \to \mu(A)$ is not computable – for more details, see (Binder *et al.* (2011)).

4.2. *Negative results and noise*

The method described in the previous section has limitations, some of which are actually impossible to overcome. Indeed, the following result was shown in Galatolo *et al.* (2011a).

Proposition 4.7. There is a computable, continuous map T on the circle having no computable invariant probability measure.

For a description of the system and for applications to reverse mathematics, see Galatolo *et al.* (2011a, Proposition 12). The construction of such a system is somewhat intricate, and the question arises as to whether this is a common phenomenon, or a phenomenon that can be observed in nature, and, in particular, whether the phenomenon persists after the addition of some small noise. In the specific construction given by Proposition 4.7, the function can be made as smooth as desired, but any random perturbation of it would break the particular structure responsible for the non-computability phenomenon.

We are then led to ask whether the addition of a small amount of noise to a system is sufficient to 'destroy' the non-computability phenomenon. The answer, perhaps

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

surprisingly, is yes: while there are dynamical systems for which the invariant measures are non-computable, perturbing such systems makes the invariant measures efficiently computable. Thus, noise, which makes the short-term behaviour of the system harder to predict, may make its long-term statistical behaviour computationally tractable.

In our framework, perturbations are modelled as follows: each iteration f of the system is affected by (a small amount of) random noise. Thus, in the perturbed system T_ε, the state of the system jumps from x to $f(x)$ and then disperses randomly around $f(x)$ with distribution $p^\varepsilon_{f(x)}(\cdot)$. The parameter ε controls the 'magnitude' of the noise, so that $p^\varepsilon_{f(x)}(\cdot) \to f(x)$ as $\varepsilon \to 0$. The following result was proved in Braverman *et al.* (2012).

Theorem 4.8. Let T be a computable system over a compact subset X of $\mathbb{R}^d$ and assume $p^\varepsilon_{f(x)}$ is uniform on the ε-ball around $f(x)$. Then, for almost every $\varepsilon > 0$, the ergodic measures of the perturbed system T_ε are *all computable*.

The assumption of uniformity on the noise is not essential, and it can be relaxed to (computable) smoothness. The result follows from general considerations on the computability and compactness of the relevant spaces. It shows that the non-computability of invariant measures is not robust.

In addition to establishing the result on the *computability* of invariant measures in noisy systems, we can also get upper bounds on the *complexity* of computing these measures. However, these results are beyond the scope of this paper – see Braverman *et al.* (2012) for details.

5. Computing the speed of ergodic convergence and pseudorandom points

As we mentioned earlier, the Birkhoff ergodic theorem (Theorem 2.1) tells us that if the system is ergodic, there is a full measure set of points for which the averages of the values of the observable f along its trajectory (time averages) coincides with the spatial average of the observable f. Similar results can be obtained for the convergence in the L^2 norm, and others. Many more refined results are linked to the speed of convergence of this limit. It is then natural to ask if there is a possibility of computing this speed of convergence.

Some abstract results in Avigad *et al.* (2010) imply that the speed of convergence of such averages in a computable ergodic dynamical system can be estimated algorithmically. On the other hand, it is also shown that there are non-ergodic systems where this kind of estimate is impossible. A very short proof of the result for ergodic systems was given in Galatolo *et al.* (2010).

The results on the computation of invariant measures and the speed of convergence allows us to compute points that are statistically typical for the dynamics. Let μ be a computable probability measure on $[0, 1]$.

Points satisfying the pointwise ergodic theorem (Theorem 2.1) for an observable f are said to be *typical for* f, and points that are typical for each continuous f are said to be *typical for the measure* μ (and for the dynamics).

The set of computable points, being countable, is a very small (invariant) set, compared with the whole space. For this reason, a computable point should only rarely be expected

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

to be typical for the dynamics, as defined above. More precisely, the Birkhoff ergodic theorem (Theorem 2.1) and other almost everywhere theorems that hold for a full measure set, cannot help us decide if there exists a computable point that is typical for the dynamics.

A number of theoretical questions arise naturally from all these facts. Due to the importance of forecasting and simulating a dynamical system's behaviour, these questions also have some practical motivation. In particular, we are led to ask whether there is some computable initial condition realising the typical statistical behaviour, and how to choose such points. Such points could be called *pseudorandom* points, and a result showing their existence and their computability from the description of the system could be seen as a *constructive* version of the pointwise ergodic theorem.

Based on a kind of constructive version of the Borel–Cantelli Lemma (Galatolo *et al.* 2009), and on estimates of the speed of convergence of the ergodic averages mentioned earlier, the following result can be established (see Galatolo *et al.* (2010) for the details).

Theorem 5.1. If $([0,1],\mu,T)$ is a computable ergodic system[†], then there is a uniform sequence x_n of computable points that is dense on the support of μ and such that for each n, we have x_n is typical for μ. Moreover, this sequence can be computed starting from a description of T and μ.

In particular, for the classes of systems where the interesting invariant measure can be computed from the description of the system, we get that the pseudorandom points can, in turn, be computed starting from a description of the system alone.

Corollary 5.2. Each piecewise expanding map, with computable derivatives as in the assumptions of Theorem 4.5, has a sequence of pseudorandom points that is dense in the support of the measure. Moreover, this sequence can be computed starting from the description of the system alone.

These two statements can be seen as *constructive/effective* versions of the ergodic theorem.

6. Dynamics of non-rigorous simulations

In this section we consider in more detail the problem of naive simulations, that is, simulations where we do not take account of the various rounding errors.

We simulate the system by discretising the map and the space to a finite set (the 'floats', the 'doubles', and so on). It is clear that, in principle, the dynamics of the truncated map is different from the original one. In spite of this, naive simulation is widely used, and we can say that, within certain limits, it is quite effective. Of course, there are important limits to this approach, and it is important to understand them. In the next section, we shall present some experiments comparing the statistical behaviour of the truncated map (approximated to a float or double) with the original. We can carry out a rigorous

[†] That is, a system having a computable invariant measure and a dynamics that is computable on an effective G_δ-set (an intersection of a sequence of uniformly effective open sets) – see Galatolo *et al.* (2009) for details.

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

simulation of the behaviour of the orbits of the system using special software packages (**i**RRAM – see Müller (2001)), which are designed for this kind of work. This means that, as in interval arithmetic, we compute a number along with the error in the computation. Then, an iterative step takes place: if the error is too big, we redo the computation with higher precision, and repeat the process until a precise result is obtained.

It is widely believed that naive computer simulations very often produce correct statistical behaviour. The evidence for this is mostly heuristic, with most arguments based on various 'shadowing' results (see, for example, Hasselblatt and Katok (1995, Chapter 18)). In this kind of approach, it can be proved that in a suitable system, every pseudo-trajectory (such as those obtained in simulations with some computation error) is close to a real trajectory of the system.

However, even if we know that what we see in a simulation is near to some real trajectory, we do not know if this real trajectory is typical in some sense[†]. Another limitation of this approach is that shadowing results only hold in systems having some strong hyperbolicity, while many physically interesting systems are not like this.

7. Computing the measure density using the Birkhoff averages

In the first experiment, we shall present one of the possible pathological behaviours of floating point arithmetics and show that the statistical features of the truncated dynamics are quite similar to the original for a certain reasonably long time, but after that time, there is a completely different behaviour.

Consider the map $T : [0,1] \to [0,1]$ given by

$$T(x) = \begin{cases} \frac{1}{8} + 3x + 2x^2 & 0 \leqslant x \leqslant \frac{1}{4} \\ 4(x - \frac{1}{4}) & \frac{1}{4} < x \leqslant \frac{1}{2} \\ 4(x - \frac{1}{4}) & \frac{1}{2} < x \leqslant \frac{3}{4} \\ \frac{7}{8} + 3(x-1) - 2(x-1)^2 & \frac{3}{4} < x \leqslant 1, \end{cases}$$

whose graph is shown in Figure 2b.

This map is a piecewise expanding map of the interval, like those studied in Lasota and Yorke (1973). We know it admits a unique a.c. invariant measure with bounded variation density.

One method for computing an approximation to the density of the invariant measure is to take advantage of Birkhoff's ergodic theorem (Theorem 1). We partition the interval $[0,1]$ into homogeneous smaller intervals and take the first k iterates of a generic orbit (with k big). The ratio of the number of iterates falling in one interval of the partition to

[†] To see why shadowing does not necessarily produce typical statistical behaviour, consider the map $T : [0,1] \to [0,1]$ defined by $T(x) = 2x \pmod 1$, and its discretisation to floating point arithmetics: $\tilde{T} : [0,1] \to [0,1]$. Due to the binary representation of floating point numbers as a finite sequence of 0s and 1s, and because the map $\tilde{T}$ acts as a shift, we get that all the orbits of $\tilde{T}$ are attracted by the point 0, while typical orbits of T are dense in $[0,1]$.
Hence, shadowing holds, and shows real orbits of T converging to 0. So these are real orbits of the system, but are not at all typical.

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

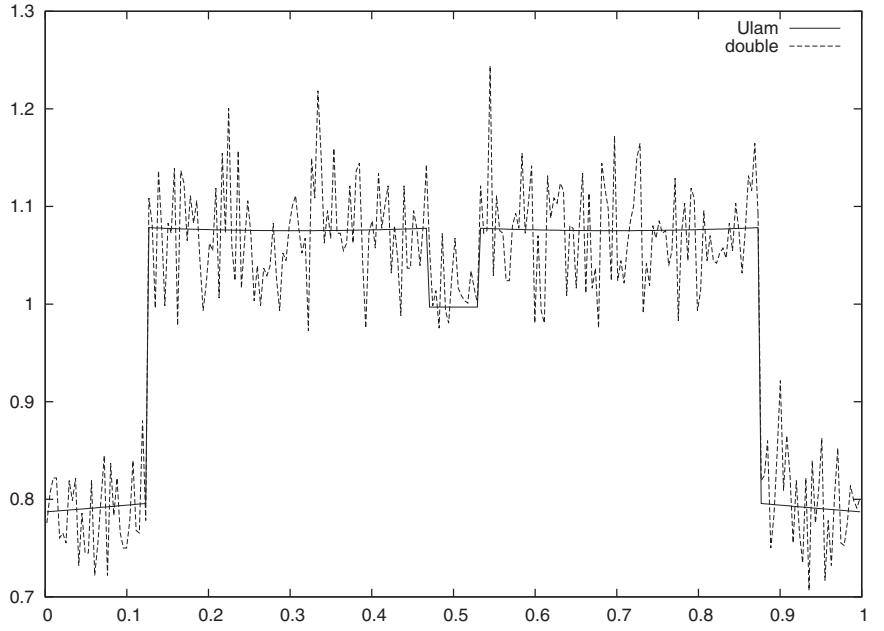

Fig. 3. 100,000 Iterates: Birkhoff averages and Ulam approximation

the number of iterates tends to the measure of the interval with respect to the invariant measure as k goes to infinity.

To see if the implementation of such an algorithm with truncation errors is reliable, we implemented it using the double standard type, and compared the density obtained with that obtained using the Ulam approximation, as in Galatolo and Nisoli (2012).

The Ulam approximation consists of approximating the Perron–Frobenius operator associated with the map T by a finite state Markov chain; we partition the interval $[0, 1]$ into k homogeneous intervals $\{I_i\}_i^k$. The Markov chain is defined by

$$(P)_{ij} = \frac{m(T^{-1}I_j \cap I_i)}{m(I_i)},$$

and the steady state of this Markov chain approximates (in a way made precise in Galatolo and Nisoli (2012)) the invariant measure of the map.

The partitions used for the algorithm and for the Ulam approximation are the same: a homogeneous partition of the interval $[0, 1]$ into 256 intervals.

We took up to 1,000,000 iterates for the point 0.31159286543998735 2675, and drew graphs of the density at different timesteps.

Figure 3 shows the graph for the density computed with 100,000 iterates. The solid line marks the density computed with the Ulam approximation and the dashed line marks the one computed using double precision. Modulo some fluctuations, the density computed using double precision is quite close to the one computed with the Ulam partition.

If we take a larger number of iterates, common sense suggests that the fluctuations will go to 0 as these averages converge, but, at 950,000 iterates, a strange phenomenon

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

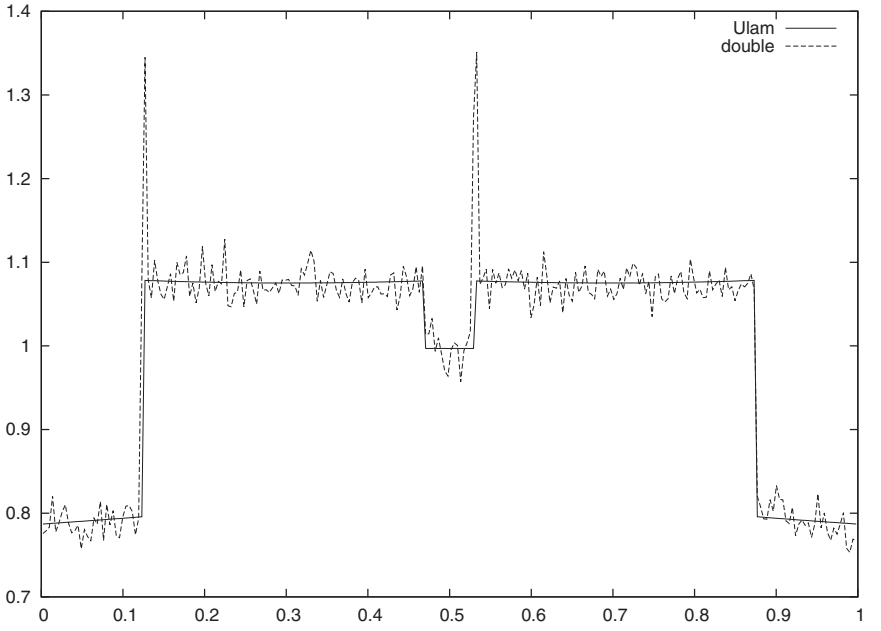

Fig. 4. 950,000 Iterates: Birkhoff averages and Ulam approximation

develops – Figure 4 shows that two peaks appear in the density computed with double precision. It appears as if the orbit computed with double precision is accumulating in these two intervals.

Indeed, taking 1,000,000 iterates, Figure 5 shows that the density computed using doubles is a long way from that computed using the Ulam approximation, and shows some weird behaviour: it appears as if we have entered a periodic orbit contained in those two intervals.

Now, this can be explained by the fact that, since double precision numbers are a finite subset of the real numbers, the function f is represented by a function $\tilde{f}$ from a discrete set into a discrete set, called the *discretised map*.

Since $\tilde{f}$ maps a finite set to itself, if we choose any initial value x_0, we have that eventually, in the sequence of values

$$x_0,\, x_1 = \tilde{f}(x_0),\, x_2 = \tilde{f}(x_1),\, \ldots,\, x_i = \tilde{f}(x_{i-1}), \ldots,$$

one of the values is going to be repeated, that is, $x_i = x_j$ with $i > j$. When this happens, the function repeats the cycle of values $x_i, \ldots, x_{j-1}$. This means that when we deal with functions discretised in floating point arithmetics, all the points are preperiodic.

What happened in our example is that the point, after 947,988 iterations enters a periodic orbit of period 2; we computed this data using Brent's cycle finding algorithm (Brent 1980). Such a behaviour is also observed for other points, for example, 0.91876328768768276874235 1 enters a periodic orbit of period 2 after 13,055,370 iterations.

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

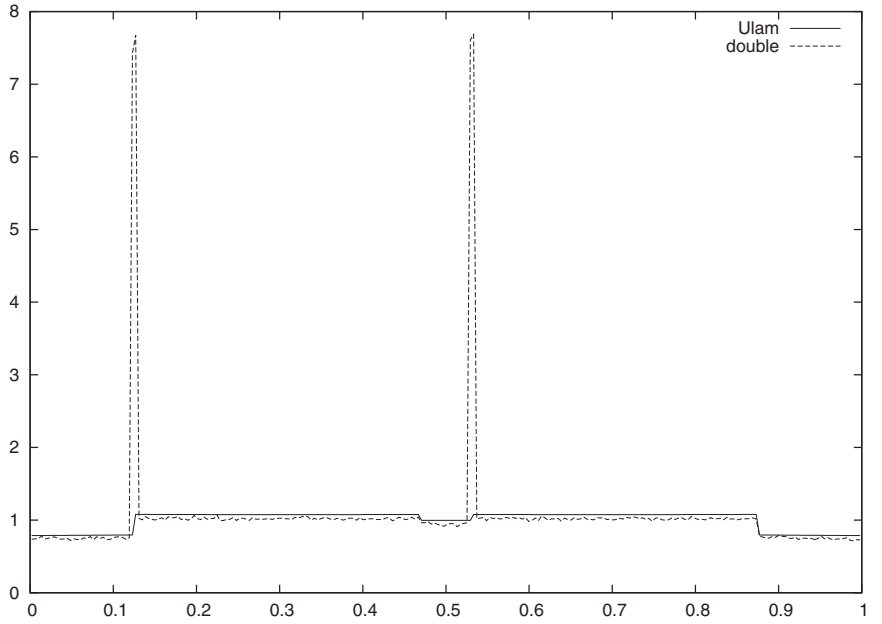

Fig. 5. 1,000,000 Iterates: Birkhoff averages and Ulam approximation

Table 1. *The preperiodic behavior for some point.*

Point	λ	τ
0.0	2	1
0.1	8183624	4816808
0.2	2	3243177
0.3	2	32345141
0.4	2	28
0.5	2	2
0.6	2	28
0.7	2	5067584
0.8	8183624	3612867
0.9	2	14935252

To check whether this behaviour was common, we developed another small experiment in which, given evenly distributed points in the interval $[0, 1]$, we compute the number of iterates needed for them to land on a periodic orbit, denoted by τ, and the period of the periodic orbit they land on, denoted by λ. We show the results of the experiment in Table 1 for ten equidistributed points in $[0, 1]$.

Therefore, a natural question to ask is whether we can find evidence of the convergence of Birkhoff averages for the 'discretised' maps. In the next section we present some further experiments showing how the convergence of this average is anything but trivial.

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

8. Convergence of Birkhoff averages

In this section we look in more detail at the convergence of time averages of a given observable to the space average. We shall see that, even in cases where we do not have the pathology seen above, the convergence of the truncated dynamics is quite approximate compared with the convergence of rigorous iterations.

To study the convergence of Birkhoff averages, we carried out the following numerical experiment:

(1) We computed the ergodic average of an observable using a good approximation of the invariant measure for a piecewise expanding map (by the Ulam method).
(2) We took a large number of homogeneously distributed points in the interval $[0, 1]$.
(3) For each of the points, we computed a large number of iterates, both rigorously and with single float precision.
(4) At chosen timesteps, we compared the Birkhoff average computed along each of these orbits (rigorous and single float) with the ergodic average.
(5) We computed the ratio of points along the rigorous orbit and the single float orbit for which the error on the Birkhoff average was bigger than a given threshold.

To do this, we used the **i**RRAM library, which enables us to compute a real number Turing-rigorously up to a given precision.

Computing an orbit using this library enables us to produce two time series, each containing single precision numbers: the first dataset contains the truncation to floats of all the points in the rigorous orbit and the second contains the orbit of the single precision discretised map. Using these datasets, we computed the Birkhoff averages: since the Birkhoff averages are computed using floating point arithmetic, we can assume that the behaviour with respect to truncation and rounding of the two averages is the same, and that any difference in the behaviour is given purely by the dynamical error.

The Lipschitz observable we chose for our experiment is

$$\phi(x) = \begin{cases} 4x & 0 \leqslant x \leqslant \frac{1}{4} \\ 1 - 4(x - \frac{1}{4}) & \frac{1}{4} < x \leqslant \frac{1}{2} \\ 0 & \frac{1}{2} < x \leqslant 1. \end{cases}$$

We decided to study the behaviour of a known map (one of the maps from Lanford (1998)). The $T : [0, 1] \to [0, 1]$ map we chose is given by

$$T(x) = 2x + 0.5x(x - 1) \bmod 1,$$

whose graph is shown in Figure 2a.

The ergodic average of the observable was computed using an approximation of the invariant measure obtained using the Ulam method on a partition of $[0, 1]$ with $2,097,152$ smaller intervals.

For this map, the ergodic average of the observable (with respect to the approximation of the invariant measure) is 0.2133442063; we took $100,000$ uniformly spaced points and fixed the threshold at 0.15625, and then, at every iterate, we compared the ratio

$$\#\{\text{points exceeding the error bound}\}/\#\{\text{points}\}.$$

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

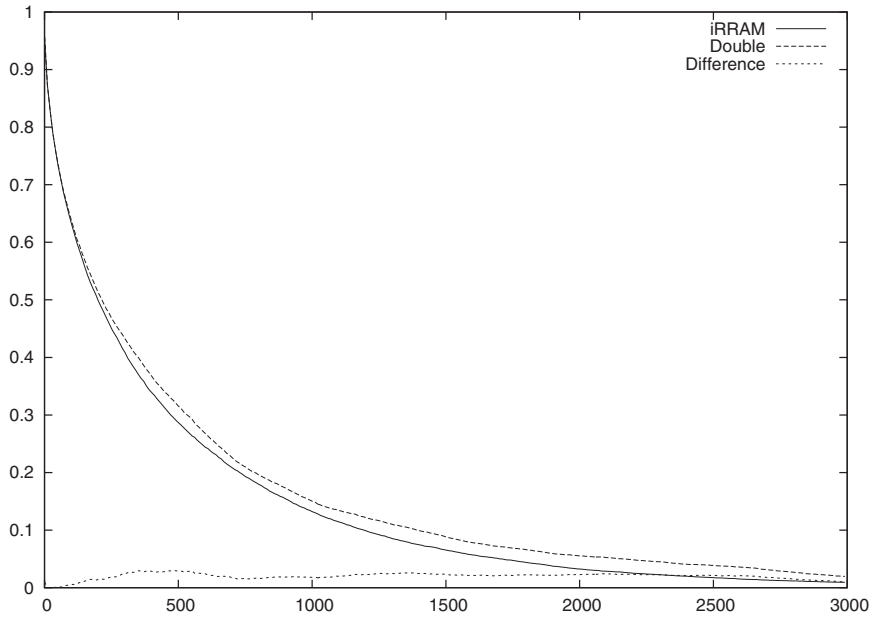

Fig. 6. Probability of exceeding the threshold with:
#Points=100,000, Threshold=0.015625 and Iterates=3,000

The results are given in Figure 6, which shows the 'probability' of exceeding the threshold for the rigorous computation and the discretised orbit, and the difference between the two probabilities.

From the graph, it seems as if the probability of exceeding the error bound decreases exponentially with respect to the number of iterates. The first interesting fact is that the decay rate seems to be smaller for floating point arithmetics. A second is that the difference between the two ratios seems to become constant as the number of iterates increases.

In fact, further experiments show a behaviour that seem to confirm these observations: to show this, we took longer orbits, taking only 2,000 points (Turing-rigorous computations are really expensive) with an error bound of 0.001953. The outcome of this experiment is shown in Figure 7, from which it appears that the number of points for which the error is large, on the floating point orbits, stabilises around 70% after 18,000 iterates.

As a further check, we computed the Birkhoff averages for some points, taking long orbits, and checked whether the absolute difference compared with our estimate of the ergodic average went to 0 – the graph for the point $\sqrt{2}/2$ is given in Figure 8.

If we zoom the tail of the graph, as shown in the inset of Figure 8, we see that, while the error in the Birkhoff average computed on the truncation of the rigorous orbit looks like it is converging to 0, the error in the average computed along the discretised orbit is far from 0, and does not show any convergence.

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

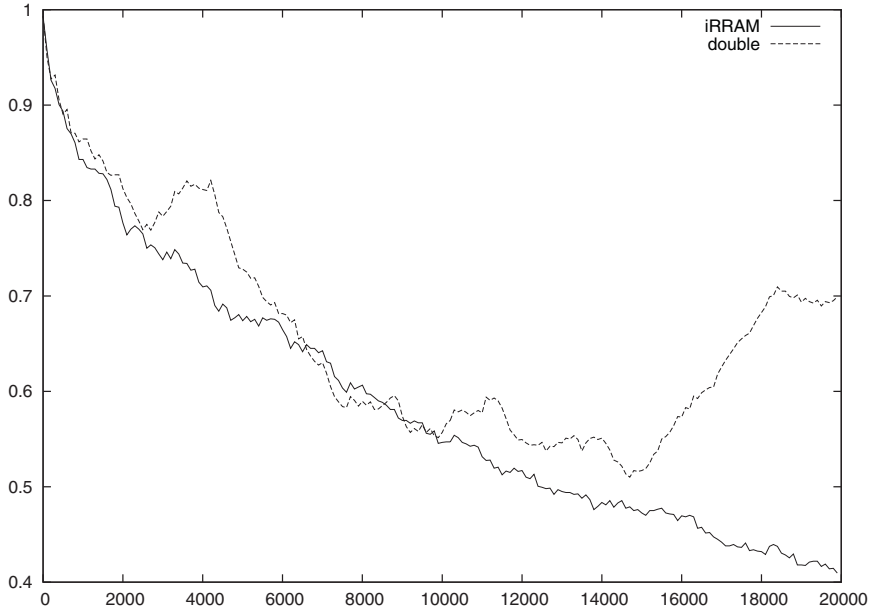

Fig. 7. Probability of exceeding the threshold with:
#Points=2,000, Threshold=0.001953 and Iterates=20,000

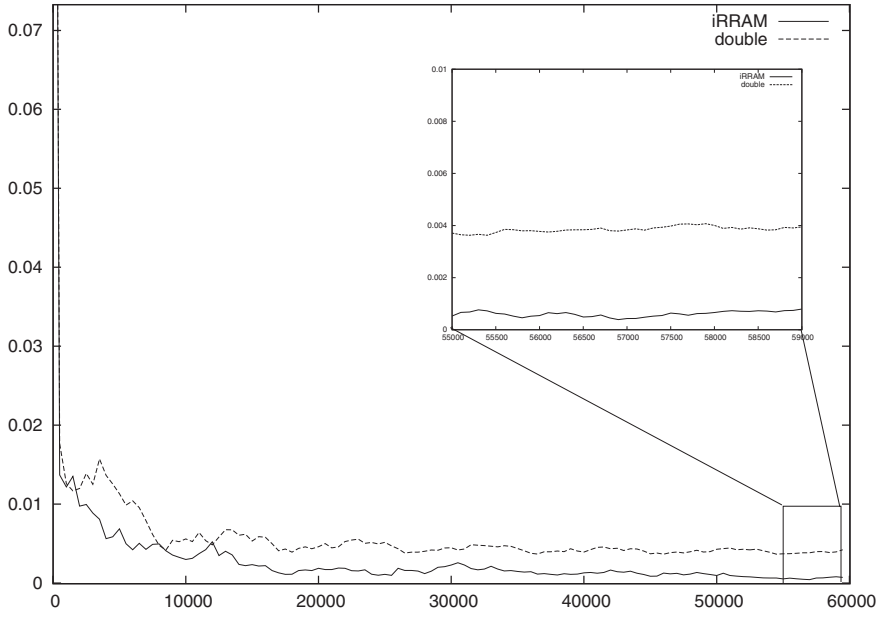

Fig. 8. Absolute difference between the Birkhoff averages and the ergodic average as the number of iterates grows, starting from the point $\sqrt{2}/2$

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

Table 2.

Iterate	**Error for the rigorous orbit**	**Error for the floating point orbit**
59997	0.000943788	0.0041985
59998	0.000940217	0.00420198
59999	0.000936645	0.00420547

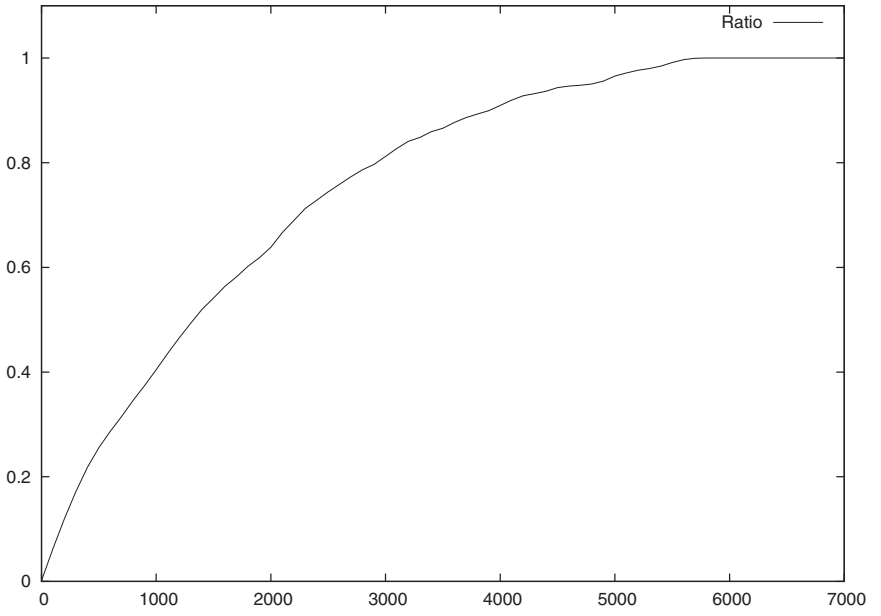

Fig. 9. Ratio between the points that had already reached the periodic orbit and the total number of points as the number of iterates grows

It seems that the dynamical error plays an important role in the convergence of the Birkhoff averages. The behaviour is similar for other points: Table 2 shows the behaviour for the point $\sqrt{3}/2$ when we look at the errors after almost $60,000$ iterates

One idea we have is that this error may be generated by the fact that none of the orbits of the discretised map shows a typical behaviour, even when we consider relatively short times. To investigate this, we took $20,000$ evenly distributed points in $[0,1]$ and, at each iterate, calculated the ratio between the points that had already reached the periodic orbit and the total number of points. As Figure 9 shows, all the points are in a periodic orbit after $6,000$ iterates, which means that, for all of them, we are computing the Birkhoff average along a periodic orbit, and the Birkohff averages on a periodic orbit do not, in general, converge to the ergodic average.

Another interesting fact is that if we look at the length of the periodic orbits where the points land, on average, this length is $14,232.4$ iterates. This means that after $20,000$ iterations, almost all the points we took into consideration have followed a full periodic

https://doi.org/10.1017/S0960129512000758 Published online by Cambridge University Press

cycle, which seems to strengthen our idea that the Birkhoff averages do not converge because we are trapped in periodic orbits.

In our opinion these experiments shows that, in general, using floating point arithmetics to compute Birkhoff averages and invariant measures should not be considered reliable, but not because of truncation and rounding errors, but rather because the dynamics of the discretised map does not mirror the generic dynamic of the real map.

References

Avigad, J., Gerhardy, P. and Towsner, H. (2010) Local stability of ergodic averages. *Transactions of the American Mathematical Society* **362** 261–288.

Binder, I., Braverman, M. and Yampolsky, M. (2006) On computational complexity of Siegel Julia sets. *Communications in Mathematical Physics* **264** 317–334.

Binder, I., Braverman, M. and Yampolsky, M. (2007) Filled Julia sets with empty interior are computable. *Journal FoCM* **7** 405–416.

Binder, I., Braverman, M., Rojas, C. and Yampolsky, M. (2011) Computability of Brolin–Lyubich measure. *Communications in Mathematical Physics* **308 3** 743–771.

Brattka, V., Hertling, P. and Weihrauch, K. (2008) A Tutorial on Computable Analysis. In: Cooper, S. B., Lowe, B. and Sorbi, A. *New Computational Paradigms: Changing Conceptions of What is Computable* 425–491.

Brent, R. (1980) An Improved Monte Carlo Factorization Algorithm. *BIT Numerical Mathematics* **20** 176–184

Braverman, M. (2004) Computational Complexity of Euclidean Sets: Hyperbolic Julia Sets are Poly-Time Computable. In: Brattka, V., Staiger, L. and Weihrauch, K. (eds.) Proceedings of the 6th Workshop on Computability and Complexity in Analysis. *Electronic Notes in Theoretical Computer Science* **120** 17–30.

Braverman, M. (2006) Parabolic Julia Sets are Polynomial Time Computable. *Nonlinearity* **19** (6) 1383–1402.

Braverman, M. and Yampolsky, M. (2006) Non-computable Julia sets. *Journal of the American Mathematical Society* **19** 551–578.

Braverman, M. and Yampolsky, M. (2008a) *Computability of Julia sets*, Algorithms and Computation in Mathematics **23**, Springer-Verlag.

Braverman, M. and Yampolsky, M. (2008b) Computability of Julia sets. *Moscow Mathematics Journal* **8** 185–231.

Braverman, M., Grigo, A. and Rojas, C. (2012) Noise vs computational intractability in dynamics. arXiv:1201.0488.

Brolin, H. (1965) Invariant sets under iteration of rational functions. *Arkiv för matematik* **6** 103–144.

Carleson, L. and Gamelin, T. W. (1993) *Complex Dynamics*, Springer.

Dellnitz, M. and Junge, O. (1999) On the approximation of complicated dynamical behavior. *SIAM Journal on Numerical Analysis* **36** 491–515.

Dellnitz, M. and Junge, O. (2002) Set Oriented Numerical Methods for Dynamical Systems. In: Fiedler, B. (ed.) *Handbook of dynamical systems* **2**, North Holland 221–264.

Froyland, G. (2001) Extracting dynamical behaviour via Markov models. In: Mees, A. (ed.) *Proceedings Nonlinear Dynamics and Statistics: Newton Institute, Cambridge 1998*, Birkhauser 283–324.

Froyland, G. (2007) On Ulam approximation of the isolated spectrum and eigenfunctions of hyperbolic maps. *Discrete and Continuous Dynamical Systems* **17** (3) 203–221.

Galatolo, S. and Nisoli, I. (2012) A simple approach to rigorous approximation of invariant measures. arXiv:1109.2342

Galatolo, S., Hoyrup, M. and Rojas, C. (2009) A constructive Borel–Cantelli lemma. Constructing orbits with required statistical properties. *Theoretical Computer Science* **410** 2207–2222.

Galatolo, S., Hoyrup, M. and Rojas, C. (2010) Computing the speed of convergence of ergodic averages and pseudorandom points in computable dynamical systems. In: Proceedings 7th International Conference on Computability and Complexity in Analysis. arXiv.org/abs/1006.0392v1.

Galatolo, S., Hoyrup, M. and Rojas, C. (2011a) Dynamics and abstract computability: computing invariant measures. *Discrete and Continuous Dynamical Systems* **29** (1) 193–212.

Galatolo, S., Hoyrup, M. and Rojas, C. (2011b) Statistical properties of dynamical systems - simulation and abstract computation. *Chaos, Solitons and Fractals* **45** (1) 1–14.

Hasselblatt, B. and Katok, A. (1995) *Introduction to the Modern Theory of Dynamical Systems*, Encyclopedia of Mathematics and Its Applications **54**, Cambridge University Press.

Hoyrup, M. and Rojas, C. (2009) Computability of probability measures and Martin-Löf randomness over metric spaces. *Information and Computation* **207** 830–847.

Lanford III, O. E. (1998) Informal Remarks on the Orbit Structure of Discrete Approximations to Chaotic Maps. *Experimental Mathematics* **7** 317–324

Lasota, A. and Yorke, J. (1973) On the existence of invariant measures for piecewise monotonic transformations. *Transactions of the American Mathematical Society* **186** 481–488.

Liverani, C. (2001) Rigorous numerical investigations of the statistical properties of piecewise expanding maps – A feasibility study. *Nonlinearity* **14** 463–490.

Lyubich, M. (1982) The measure of maximal entropy of a rational endomorphism of a Riemann sphere. *Funktsional'nyi Analiz i ego prilozheniya* **16** 78–79.

Mañé, R. (1987) *Ergodic theory and differentiable dynamics*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **8**, Springer-Verlag.

Müller, N. (2001) The iRRAM: Exact Arithmetic in C++. In: Blanck, J., Brattka, V. and Hertling, P. (eds.) Computability and Complexity in Analysis. *Springer-Verlag Lecture Notes in Computer Science* **2064** 222–252.

Turing, A. (1936) On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society* (2) **42** 230–265.

Walters, P. (1982) *An introduction to Ergodic Theory*, Springer-Verlag Graduate Texts in Mathematics **79**.

Weihrauch, K. (2000) *Computable Analysis. An Introduction*, Springer-Verlag.

Young, L.-S. (2002) What are SRB measures, and which dynamical systems have them? *Journal of Statistical Physics* **108** 733–754.