

MONOTONIC DYNAMICAL SYSTEMS UNDER SPATIAL DISCRETIZATION

P. DIAMOND, P. KLOEDEN, V. KOZYAKIN, AND A. POKROVSKII

(Communicated by Hal L. Smith)

ABSTRACT. We estimate the probability of replicating the asymptotic behaviour of a dynamical system generated by a monotonic mapping for randomly centered roundoff lattices.

A discrete-time dynamical system on the state space \mathbb{R}^d is generated by the iteration of a mapping $f : \mathbb{R}^d \mapsto \mathbb{R}^d$, that is,

$$\mathbf{x}_{n+1} = f(\mathbf{x}_n), \quad n = 0, 1, 2, \dots,$$

and will often be denoted simply by f in the sequel. The evaluation of the mapping f in a computer involves roundoff and other truncation operations which may have a long-term accumulative effect and make a comparison of the computed and the true dynamics questionable [1], [3], [5]. In this note an estimate is obtained for the probability of replicating the asymptotic behaviour of a dynamical system generated by a monotonic mapping for randomly centered roundoff lattices.

Many important applications in economic and population dynamics involve monotonic dynamical systems. The dynamical behaviour of such systems need not be trivial and indeed can even be chaotic on certain subsets in higher dimensional examples; see [2], [7] Chapter III, [9], [11], [13] and references therein. As will be shown here, this behaviour can be replicated under spatial discretization with high probability, in contrast to many other types of dynamical systems where even asymptotically stable steady states can often be at best approximated by small invariant attracting sets. Moreover, similar ideas and methods to those in this paper are central in Birkhoff – Tarski fixed point analysis as is commonly used, for example, to investigate classical solutions of discontinuous nonlinear systems [4], [11], in particular in the concept of correct modes in such systems [10], [12].

Let Q_h^d denote the h -cube in \mathbb{R}^d centered at the origin, that is,

$$Q_h^d = \{\mathbf{x} = (x^1, \dots, x^d) \in \mathbb{R}^d : -h/2 < x^i \leq h/2, i = 1, \dots, d\},$$

and for each $\mathbf{q} \in Q_h^d$ let $L_{h,\mathbf{q}} = \{\mathbf{q} + h\mathbf{z}, : \mathbf{z} \in \mathbf{Z}^d\}$, where \mathbf{Z}^d is the standard integer lattice in \mathbb{R}^d , be the uniform h -lattice in \mathbb{R}^d centered on \mathbf{q} .

For a given $\mathbf{q} \in Q_h^d$ define by $[\mathbf{x}]_{h,\mathbf{q}} = L_{h,\mathbf{q}} \cap (\mathbf{x} + Q_h^d)$ for $\mathbf{x} \in \mathbb{R}^d$ the roundoff operator from \mathbb{R}^d into $L_{h,\mathbf{q}}$. This can be rewritten in the equivalent coordinate

Received by the editors September 23, 1996 and, in revised form, December 24, 1996.

1991 *Mathematics Subject Classification*. Primary 58F10, 58F12.

Key words and phrases. Roundoff lattice, asymptotic stability, dynamical system.

This research was supported by the Australian Research Council Grant A 8913 2609.

form

$$(1) \quad [\mathbf{x}]_{h,\mathbf{q}}^i = ([x^1 - q^1]_h + q^1, \dots, [x^d - q^d]_h + q^d), \quad i = 1, \dots, d,$$

where $\mathbf{x} = (x^1, \dots, x^d)$, etc, and $[x]_h$ is the usual scalar roundoff operator defined by

$$[x]_h = kh \quad \text{if} \quad (k - \frac{1}{2})h \leq x < (k + \frac{1}{2})h.$$

Let f be a dynamical system in \mathbb{R}^d and denote by $f_{h,\mathbf{q}}$ the mapping of $L_{h,\mathbf{q}}$ into itself defined by

$$(2) \quad f_{h,\mathbf{q}}(\xi) = [f(\xi)]_{h,\mathbf{q}}, \quad \xi \in L_{h,\mathbf{q}}.$$

The mapping $f_{h,\mathbf{q}}$ is called an $L_{h,\mathbf{q}}$ -discretization [14] of f . Given $h > 0$, $\mathbf{q} \in Q_h^d$ and a compact subset Ω of \mathbb{R}^d , we shall say that the triple (h, \mathbf{q}, Ω) is *dynamically robust* if the discretization $f_{h,\mathbf{q}}$ has a single equilibrium $\mathbf{x}_{h,\mathbf{q}} = f_{h,\mathbf{q}}(\mathbf{x}_{h,\mathbf{q}}) \in \Omega \cap L_{h,\mathbf{q}}$ and

$$(3) \quad \lim_{n \rightarrow \infty} |f_{h,\mathbf{q}}^n(\mathbf{x}) - \mathbf{x}_{h,\mathbf{q}}| = 0 \quad \text{for} \quad \mathbf{x} \in L_{h,\mathbf{q}} \cap \Omega.$$

Simple examples show that for arbitrary small h some triples (h, \mathbf{q}, Ω) need not be dynamically robust even if f is a linear exponentially stable system. For instance, the triple $(h, h/2, [-1, 1])$ with $h < 1$ is dynamically robust for the one-dimensional linear system $f(x) = \alpha x$ with $0 \leq \alpha < 1$ if and only if $\alpha = 0$ (see the next section for more details). On the other hand, a triple (h, \mathbf{q}, Ω) can be dynamically robust even if the original system f is not globally asymptotically stable in Ω . This raises the question

Given f and Ω , what is the Lebesgue measure $V(S(f, h, \Omega))$ of the set $S(f, h, \Omega)$ of points $\mathbf{q} \in Q_h^d$ for which the triple (h, \mathbf{q}, Ω) is dynamically robust?

If the original system f has an equilibrium which is (globally) asymptotically stable in Ω , then $h^{-d}V(S(f, h, \Omega))$ is the probability that a discretization with given space step h and a random choice of \mathbf{q} would not “spoil” the phase diagram of the system, whereas if f is not asymptotically stable in Ω it then gives the probability of an “improvement” of the phase diagram. See Blank [5] for the related question of the statistical stabilization of an unstable cycle.

1. MAIN RESULT

We shall consider a semi-ordering $\mathbf{x} \geq \mathbf{y}$ on \mathbb{R}^d defined componentwise by $\mathbf{x} \geq \mathbf{y}$ if $x^i \geq y^i$ for $i = 1, \dots, d$, writing $\mathbf{x} > \mathbf{y}$ if $x^i > y^i$ for $i = 1, \dots, d$, and we shall say that f is *monotonic* on a set $S \subseteq \mathbb{R}^d$ if $f(\mathbf{x}) \geq f(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in S$ with $\mathbf{x} \geq \mathbf{y}$. In addition we shall denote by $V(S)$ the Lebesgue measure of a measurable set $S \subseteq \mathbb{R}^d$, so in particular $V(Q_h^d) = h^d$. Our main result is

Theorem 1. *Let f be monotonic on the set $K(\mathbf{u}_-, \mathbf{u}_+) = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{u}_- \leq \mathbf{x} \leq \mathbf{u}_+\}$ and satisfy the inequalities*

$$(4) \quad f(\mathbf{u}_-) \geq \mathbf{u}_- + \mathbf{h}, \quad f(\mathbf{u}_+) \leq \mathbf{u}_+ - \mathbf{h},$$

where $\mathbf{h} = (h, \dots, h)$. Then

$$(5) \quad \begin{aligned} & h^{-d}V(S(f, h, K(\mathbf{u}_-, \mathbf{u}_+))) \\ & \geq \max \{0, 2 - h^{-d}V(\{\mathbf{x} \in K(\mathbf{u}_-, \mathbf{u}_+) : \mathbf{x} - \mathbf{h}/2 < f(\mathbf{x}) \leq \mathbf{x} + \mathbf{h}/2\})\}. \end{aligned}$$

The proof of Theorem 1 will be given in the next section. The theorem reduces the probability estimate to the problem of estimating the volume of a set in \mathbb{R}^d which is described by an explicit and simple inequality involving only the mapping under consideration. On the other hand, even for a given h there is no simple and straightforward way to describe the set of those q for which the corresponding discretizations satisfy (3). This theorem is useful in the context of numerical simulation. Indeed, for $h = 2^{-N}$ the $L_{h,0}$ -discretization is a natural theoretical model of the implementation of a mapping f in *fixed point format* with N binary digits and radix point in the first position (see, for example, [6], pages 98-100). Such discretizations in fact use only integer arithmetic, which is often an advantage. One of the complications here is the possibility of so-called ghost periodic solutions in the vicinity of an equilibrium of the underlying continuous system. These have been reported even for the simplest linear mappings; see [1], [3] and references therein. Theorem 1 not only refines the theoretical explanation of such effects, but also suggests a way for suppressing it by simply changing the origin of the lattice under consideration, which is often easy to do.

The following corollary to Theorem 1 is a geometrical inequality which follows from (5) and the fact that, by definition, $h^{-d}V(S(f, h, K(\mathbf{u}_-, \mathbf{u}_+))) \leq 1$.

Corollary 1. *For f as in Theorem 1*

$$V(\{\mathbf{x} \in K(\mathbf{u}_-, \mathbf{u}_+) : \mathbf{x} - \mathbf{h}/2 < f(\mathbf{x}) \leq \mathbf{x} + \mathbf{h}/2\}) \geq h^d.$$

To illustrate the above theorem we consider a simple example of an exponentially stable system in \mathbb{R}^1 . Let $f(x) = \alpha x$, where $0 \leq \alpha < 1$ and $u_- < 0 < u_+$. For $0 < h \leq \min\{\alpha u_- - u_-, u_+ - \alpha u_+\}$ it is obvious that

$$\{x \in K(u_-, u_+) : x - h/2 < f(x) \leq x + h/2\} = \left(-\frac{h}{2} \frac{1}{1-\alpha}, \frac{h}{2} \frac{1}{1-\alpha}\right]$$

and hence that

$$V(\{x \in K(u_-, u_+) : x - h/2 < f(x) \leq x + h/2\}) = \frac{h}{1-\alpha}.$$

Inequality (5) in Theorem 1 thus gives

$$(6) \quad h^{-1}V(S(f, h, K(u_-, u_+))) \geq \max\left\{0, 2 - \frac{1}{1-\alpha}\right\}.$$

On the other hand, by explicit calculation

$$S(f, h, K(u_-, u_+)) = \left(-\frac{h}{2}, -\frac{h}{2} \frac{\alpha}{1-\alpha}\right] \cup \left(\frac{h}{2} \frac{\alpha}{1-\alpha}, \frac{h}{2}\right]$$

for $0 \leq \alpha < 1/2$ and $S(f, h, K(u_-, u_+)) = \emptyset$ for $1/2 \leq \alpha < 1$, so

$$V(S(f, h, K(u_-, u_+))) = h \left(1 - \frac{\alpha}{1-\alpha}\right) = h \left(2 - \frac{1}{1-\alpha}\right)$$

for $0 \leq \alpha < 1/2$ and $V(S(f, h, K(u_-, u_+))) = 0$ for $1/2 \leq \alpha < 1$. Hence

$$(7) \quad h^{-1}V(S(f, h, K(u_-, u_+))) = \max\left\{0, 2 - \frac{1}{1-\alpha}\right\}, \quad 0 \leq \alpha < 1.$$

Comparing (6) and (7), we see that the lower bound (5) is sharp in this simple case.

An important advantage of Theorem 1 is that the discretization parameter h does not have to be small. Estimates concerning the case when $h \rightarrow 0$ can be extracted from the theorem in the following way.

Corollary 2. *Suppose that $f(\mathbf{x}_*) = \mathbf{x}_*$ and that the matrix $f'_{\mathbf{x}_*}$ has strictly positive entries and its spectral radius $\sigma(f'_{\mathbf{x}_*}) < 1$. Then for each $\varepsilon > 0$ there exist an element $\mathbf{u}_- = \mathbf{u}_-(\varepsilon) < \mathbf{x}_*$, an element $\mathbf{u}_+ = \mathbf{u}_+(\varepsilon) > \mathbf{x}_*$ and a scalar $h_0 = h_0(\varepsilon) > 0$ such that*

$$(8) \quad \begin{aligned} h^{-d}V(S(f, h, K(\mathbf{u}_-, \mathbf{u}_+))) \\ \geq \max\{0, 2 - (1 + \varepsilon) (\det(I - f'_{\mathbf{x}_*}))^{-1}\}, \quad 0 < h \leq h_0. \end{aligned}$$

The explicit calculation of the determinant $\det(I - f'_{\mathbf{x}_*})$ for higher dimensional systems can be avoided by judicious application of matrix inequalities such as the following (cf. [2], item (i), page 158): if $\sigma(A) \leq 1$ and $A \geq B \geq 0$ componentwise, then $\det(I - A) \leq \det(I - B)$.

2. PROOFS

Proof of Theorem 1. Let $k(h, \mathbf{q})$ denote the cardinality of the set

$$F(h, \mathbf{q}) = L_{h, \mathbf{q}} \cap \bigcap \{\mathbf{x} : f(\mathbf{x}) \in \mathbf{x} + Q_h^d\}.$$

The following assertion is well known [8] and its meaning is geometrically clear.

Lemma 1.

$$\int_{Q_h^d} k(h, \mathbf{q}) d\mathbf{q} = V(\{\mathbf{x} : f(\mathbf{x}) \in \mathbf{x} + Q_h^d\}).$$

We also need the following special and rather straightforward case of the Birkhoff–Tarski Theorem [4].

Lemma 2. *Let g be a monotonic mapping of a finite set $\Omega \subset \mathbb{R}^d$ into itself and let $\mathbf{u} \in \Omega$ be such that either $g(\mathbf{u}) \geq \mathbf{u}$ or $g(\mathbf{u}) \leq \mathbf{u}$. Then the iterations $\mathbf{u}_{n+1} = g(\mathbf{u}_n)$ with $\mathbf{u}_0 = \mathbf{u}$ converge to a fixed point $x_* = g(x_*) \in \Omega$.*

The sequence \mathbf{u}_n is either monotonically increasing or monotonically decreasing with $\mathbf{u}_n \in \Omega$ for all $n \geq 0$. It must settle to a point x_* in Ω after a finite number of steps because Ω is a finite set. The point x_* is obviously a fixed point of g . \square

We apply this lemma to obtain

Proposition 1. $k(h, \mathbf{q}) \geq 1$ for each $\mathbf{q} \in Q_h^d$.

Proof. The elements of the set $F(h, \mathbf{q})$ are precisely the fixed points of the discretization mapping $f_{h, \mathbf{q}}$, which is monotonic in view of the monotonicity of f and (1). In addition, $f_{h, \mathbf{q}}$ maps the finite set $K_{h, \mathbf{q}}(\mathbf{u}_-, \mathbf{u}_+) = L_{h, \mathbf{q}} \cap K(\mathbf{u}_-, \mathbf{u}_+)$ into itself by (4). Finally, there obviously exist points $\mathbf{u} \in K_{h, \mathbf{q}}(\mathbf{u}_-, \mathbf{u}_+)$ with either $f_{h, \mathbf{q}}(\mathbf{u}) \geq \mathbf{u}$ or $f_{h, \mathbf{q}}(\mathbf{u}) \leq \mathbf{u}$, so Lemma 2 can be applied to yield at least one fixed point for the mapping $f_{h, \mathbf{q}}$ in $K_{h, \mathbf{q}}(\mathbf{u}_-, \mathbf{u}_+)$. \square

Moreover, we have

Proposition 2. $\mathbf{q} \in S(f, h, K(\mathbf{u}_-, \mathbf{u}_+))$ if and only if $k(h, \mathbf{q}) = 1$.

Proof. If $k(h, \mathbf{q}) > 1$, then the dynamical system $f_{h, \mathbf{q}}$ has at least two distinct equilibria, and so $\mathbf{q} \notin S(f, h, K(\mathbf{u}_-, \mathbf{u}_+))$.

Suppose now that $k(h, \mathbf{q}) = 1$ and let $\mathbf{x}_{h, \mathbf{q}}$ be the unique fixed point of $f_{h, \mathbf{q}}$ in $K_{h, \mathbf{q}}(\mathbf{u}_-, \mathbf{u}_+)$. By Lemma 2 both of the sequences $f_{h, \mathbf{q}}^n(\mathbf{u}_+)$ and $f_{h, \mathbf{q}}^n(\mathbf{u}_-)$ converge to $x_{h, \mathbf{q}}$, that is,

$$(9) \quad \lim_{n \rightarrow \infty} f_{h, \mathbf{q}}^n(\mathbf{u}_+) = x_{h, \mathbf{q}}, \quad \lim_{n \rightarrow \infty} f_{h, \mathbf{q}}^n(\mathbf{u}_-) = x_{h, \mathbf{q}}.$$

Since $f_{h,\mathbf{q}}$ is monotonic, the two limits in (9) imply the existence of the limits in (3). Hence $\mathbf{q} \in S(f, h, K(\mathbf{u}_-, \mathbf{u}_+))$. \square

Let us now complete the proof of Theorem 1. By Proposition 2 and the equality $\mathbf{x} + Q_h^d = \{x \in \mathbb{R}^d : x - \mathbf{h}/2 < f(x) \leq x + \mathbf{h}/2\}$ we need only to establish that

$$(10) \quad h^{-d}V(\{\mathbf{q} : k(h, \mathbf{q}) = 1\}) \geq 2 - h^{-d}V(\{\mathbf{x} : f(\mathbf{x}) \in \mathbf{x} + Q_h^d\}).$$

Now, by definition, $V(\{\mathbf{q} : k(h, \mathbf{q}) = 1\}) + V(\{\mathbf{q} : k(h, \mathbf{q}) > 1\}) = h^d$ and

$$\int_{Q_h^d} k(h, \mathbf{q}) d\mathbf{q} \geq V(\{\mathbf{q} : k(h, \mathbf{q}) = 1\}) + 2V(\{\mathbf{q} : k(h, \mathbf{q}) > 1\}),$$

so we have

$$\int_{Q_h^d} k(h, \mathbf{q}) d\mathbf{q} \geq 2h^d - V(\{\mathbf{q} : k(h, \mathbf{q}) = 1\}).$$

Finally, by Lemma 1 the last inequality can be rewritten as

$$h^{-d}V(\{\mathbf{q} : k(h, \mathbf{q}) = 1\}) \geq 2 - h^{-d}V(\{\mathbf{x} : f(\mathbf{x}) \in \mathbf{x} + Q_h^d\})$$

which is the same as (10) since $h^{-d}V(\{\mathbf{q} : k(h, \mathbf{q}) = 1\}) \geq 0$. \square

Proof of Corollary 2. Fix an $r > 0$ such that the inequality

$$(11) \quad \inf_{|\mathbf{x} - \mathbf{x}_*| \leq r} f'_\mathbf{x} > 0$$

holds componentwise.

The linear mapping $f'_{\mathbf{x}_*}$ has a unique componentwise strictly positive eigenvector \mathbf{e} with $|\mathbf{e}| = 1$ such that $f'_{\mathbf{x}_*} \mathbf{e} = \sigma(f'_{\mathbf{x}_*}) \mathbf{e}$ due to the Perron–Frobenius theorem ([2], p. 27). Since $\sigma(f'_{\mathbf{x}_*}) < 1$ there exists $l_0 > 0$ such that

$$(12) \quad f(\mathbf{x}_* + l\mathbf{e}) < \mathbf{x}_* + l\mathbf{e}, \quad f(\mathbf{x}_* - l\mathbf{e}) > \mathbf{x}_* - l\mathbf{e} \quad \text{for } 0 < l \leq l_0.$$

Choose also $l_1 < l_0$ such that the estimate

$$(13) \quad V(\{\mathbf{x} \in K(x_* - l_1\mathbf{e}, x_* + l_1\mathbf{e}) : f(\mathbf{x}) \in \mathbf{x} + Q_h^d\}) < (1 + \varepsilon)h^d \det((I - f'_\mathbf{x})^{-1})$$

holds for $h > 0$. By (11) there exists an $l_2 > 0$ such that f is monotonic on the set $K(\mathbf{x}_* - l_2\mathbf{e}, \mathbf{x}_* + l_2\mathbf{e})$. Now denote $l = \min\{l_0, l_1, l_2\}$ and define $\mathbf{u}_- = \mathbf{x}_* - l\mathbf{e}$, $\mathbf{u}_+ = \mathbf{x}_* + l\mathbf{e}$. Finally, choose h_0 to be any positive number for which

$$f(\mathbf{u}_-) > \mathbf{u}_- + \mathbf{h}_0, \quad f(\mathbf{u}_+) < \mathbf{u}_+ - \mathbf{h}_0$$

where $\mathbf{h}_0 = (h_0, \dots, h_0)$. Then due to Theorem 1 we have the estimate

$$\begin{aligned} & h^{-d}V(S(f, h, K(\mathbf{u}_-, \mathbf{u}_+))) \\ & \geq \max\{0, 2 - h^{-d}V(\{\mathbf{x} \in K(\mathbf{u}_-, \mathbf{u}_+) : \mathbf{x} - \mathbf{h}/2 < f(\mathbf{x}) \leq \mathbf{x} + \mathbf{h}/2\})\}. \end{aligned}$$

for $h < h_0$, which implies (8) by (13). \square

REFERENCES

- [1] C. Beck and G. Roepstorff, Effects of Phase Space Discretization on the Long-time Behaviour of Dynamical Systems. *Physica*, **D25** (1987), 173–180. MR **88i**:58152
- [2] A. Berman and R.J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, SIAM Classics in Applied Math. vol. 9, SIAM, Philadelphia, 1994. MR **95e**:15013
- [3] P.M. Binder, Machine iteration of a linear mapping: Local behaviour, *Computers Math. Applic.*, **21** (1991), 133-140. MR **91m**:58084
- [4] G. Birkhoff, *Lattice Theory*, AMS. Colloq. Publ. vol 25., Amer. Math. Soc., Providence (1967). MR **37**:2638
- [5] M. Blank, Pathologies generated by round-off in dynamical systems, *Physica D* **78**, 93-114 (1994). MR **95g**:58115
- [6] J.J.F. Cavanagh, *Digital Computer Arithmetic. Design and Implementation*, McGraw-Hill Book Company, 1984.
- [7] M.W. Hirsch, The dynamical systems approach to differential equations, *Bulletin of AMS*, **11** (1984), 1–64. MR **85m**:58060
- [8] M.G. Kendall and P.A.P. Moran, *Geometrical Probability*, C. Griffin, London (1963). MR **30**:4275
- [9] M.A. Krasnosel'skii, *Positive Solutions of Operator Equations*, Noordhoff, Groningen, 1964. MR **31**:6107
- [10] M.A. Krasnosel'skii and A.V. Pokrovskii, Regular solutions of equations with discontinuous nonlinearities. *Soviet Mathematics*, **17** (1976), 128 – 132. MR **58**:30559
- [11] M.A. Krasnosel'skii and P.P. Zabreiko, *Geometrical Methods of Nonlinear Analysis*, Springer-Verlag, Berlin, 1984. MR **85b**:47057
- [12] A.V. Pokrovskii: Correct solutions of equations with strong nonlinearities. *Soviet Mathematics Doklady*, **29** (1984), 107 – 110. MR **86c**:35056
- [13] H.L. Smith, *Monotone Dynamical Systems*, Amer. Math. Soc., Providence, RI, 1995. MR **96c**:34002
- [14] H.J. Stetter, *Analysis of Discretization Methods for Ordinary Differential Equations*, Springer-Verlag, Berlin (1973). MR **54**:14381

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF QUEENSLAND, BRISBANE 4072, AUSTRALIA
E-mail address: `pmd@maths.uq.edu.au`

FACHBEREICH MATHEMATIK, JOHANN WOLFGANG GOETHE UNIVERSITÄT, D-60054 FRANKFURT AM MAIN, GERMANY
E-mail address: `kloeden@math.uni-frankfurt.de`

INSTITUTE OF INFORMATION TRANSMISSION PROBLEMS, RUSSIAN ACADEMY OF SCIENCE, 19 ERMOLOVOY ST., MOSCOW 101447, RUSSIA
E-mail address: `kozyakin@nov.ippi.ras.ru`

PHYSICS DEPARTMENT, UNIVERSITY COLLEGE, CORK, IRELAND