Discrepancy and discretizations of circle expanding maps I: theory

Pierre-Antoine Guihéneuf, Maurizio Monge

DISCREPANCY AND DISCRETIZATIONS OF CIRCLE EXPANDING MAPS I: THEORY

PIERRE-ANTOINE GUIHÉNEUF AND MAURIZIO MONGE

Abstract. This paper is aimed to study the ergodic short-term behaviour of discretizations of circle expanding maps. More precisely, we prove some asymptotics of the distance between the t-th iterate of Lebesgue measure by the dynamics f and the t-th iterate of the uniform measure on the grid of order N by the discretization on this grid, when t is fixed and the order N goes to infinity. This is done under some explicit genericity hypotheses on the dynamics, and the distance between measures is measured by the mean of a distance we call *discrepancy*. The proof is based on a study of the corresponding linearized problem, where the problem is translated into terms of equirepartition on tori of dimension exponential in t.

A numerical study associated to this work is presented in [GM22].

Contents

1. Introduction

Motivations. In one of the last papers he published [Lan98], Oscar E. Lanford was proposing to study the behaviour of spatial discretizations of expanding maps of the circle in some limiting regime. The question was to decide whether in most of cases, the middle-term ergodic behaviour of such discretizations reflects the actual dynamics of the map.

To fix the notations, let us take $f : \mathbf{S}^1 \to \mathbf{S}^1$ an expanding map (meaning that $f'(x) > 1$ for any $x \in \mathbf{S}^1$) and consider the grid E_N made of N points of $\mathbf{S}^1$ equally spaced. That is, identifying $\mathbf{S}^1 \simeq \mathbf{R}/\mathbf{Z}$ with $[0, 1]$, one sets

$$E_N = \left\{ \frac{i}{N} \;\middle|\; 0 \leq i \leq N-1 \right\}.$$

To each of these grids is associated a projection $P_N : \mathbf{S}^1 \to E_N$ on the nearest point of E_N (for some points one has two choices for the nearest neighbour, one does a choice once for

Date: August 2, 2022.
2010 *Mathematics Subject Classification.* 37M25, 37M05, 37C20, 37C40, 37E10.

all and this choice will not play any role in the sequel). This leads to the definition of the *discretization* of the map f relatively to the grid E_N as

$$\begin{aligned} f_N : E_N &\longrightarrow E_N \\ x &\longmapsto P_N(f(x)). \end{aligned}$$

Remark that if $N = 2^n$, this corresponds to a discretization realizing the rounding of $f(x)$ with n binary digits.

In [Lan98], Lanford asked whether the dynamics of the maps f_N^m looks like the one of f^m in the regime $\log N \ll m \ll \sqrt{N}$. Here is how he justifies these bounds: *"The first $\ll$ allows the computed orbit to deviate macroscopically from the true one over most of its length; the second is in any case usually satisfied in practice and ought to mean that the times considered are short enough so that the effects of the strict finiteness of the space of states are not important. In fact: it might be prudent to replace the second $\ll$ by the stronger condition $\log m \ll \log N$."* His article includes enlightening philosophical thoughts, supported by some numerical experiments.

This research program was tackled by Paul P. Flockermann in his unpublished PhD thesis [Flo02] (under the supervision of Lanford). In this work he obtains partial results towards the limiting behaviour of the "non-injectivity" of the maps f_N – i.e. the quantity $\operatorname{Card}(f_N^k(E_N))/\operatorname{Card}(E_N)$ – when k is fixed and N goes to infinity[1]. These statements are valid under genericity assumptions on the expanding map[2] f: they concern either generic C^r expanding maps, for $1 < r \leq +\infty$, or any real-analytic expanding map different from $x \mapsto 2x$. Lanford and Flockermann were writing an article to complete these partial results (they had an unpublished draft) which unfortunately has never been published.

These results had been obtained independently by Vladimirov in [Vla96] (further works based on this grounding article were published in [DV98, VKD00, DV02a, DV02b]). In this article, the author founds a solid theoretical basis about the discretizations' behaviour: the algebras of quasiperiodic subsets of the lattice, their statistical properties (frequency measurability) under nonresonance conditions, their algebraic properties with respect to discretizations of linear maps, the role of skew products of measure-preserving automorphisms of multidimensional tori in the asymptotic independence and uniform distribution of quantisation errors... This approach reveals more powerful than Flockermann's: in addition to the equidistribution of roundoff errors, Vladimirov gets the fact that the asymptotic rate of injectivity is 0, and some functional central limit theorem, which was published with Vivaldi in [VV03]. Early apparitions of this kind of ideas can be found in the work of Voevodin [Voe67].

One of these results has been re-discovered independently by the first author in [Gui19]: it is proved that the actual limit of the non injectivity rate $\operatorname{Card}(f_N^k(E_N))/\operatorname{Card}(E_N)$ when k is fixed and N goes to infinity is zero. The techniques used in this article are a bit different from the ones of Flockermann and Vladimirov: they involve the notion of "model set", usually used in the study of quasicrystals, and some theorems from basic geometry of numbers, and allow to get similar results in different settings. The approach of the present article is based on these techniques.

[1] More precisely, he treats the corresponding linear case: he gets some local statements.

[2] They become false for the trivial example of $x \mapsto 2x$.

Main results. The aim of this article is to make a contribution in the direction of Lanford's program, by looking at the short-term ergodic behaviour of discretizations: we will compare the actions of the maps f and f_N on respectively Lebesgue measure Leb on $\mathbf{S}^1$ and the uniform measure Leb_N on E_N. This comparison will be made using a distance on measures called here *discrepancy*[3] and denoted Disc, which spans the weak-* topology (see Section 2). Our goal will be to get an asymptotics for the quantity

$$\mathrm{Disc}\left(f_*^k(\mathrm{Leb}), (f_N^k)_*(\mathrm{Leb}_N)\right) \tag{1.1}$$

when k is fixed and N goes to infinity.

As for the previous works already described, we will need genericity assumptions to ensure that there is no phenomenon of resonance between the dynamics and the grid. For example, if $f(x) = 2x \mod 1$ and $N = 2^n$, then the orbit of any point of E_N under f_N eventually falls in the fixed point 0; this is a very specific phenomenon that one wants to avoid to understand what happens "in most cases"[4]. Hence, we will consider *generic properties* on the spaces of C^r expanding maps: a property will be called generic if satisfied on at least a countable intersection of open and dense sets of C^r expanding maps. As these sets of maps are Baire, genericity has some nice natural properties: a generic property is satisfied on a dense set of maps, the fact of satisfying two generic properties is generic, etc. In fact, genericity properties needed to get our results is very weak, so our theorems are also valid under some different genericity assumptions (see [Gui19] for a discussion).

Our main theorem is the following.

Theorem A. *Let $r \geq 1$, f a generic C^r expanding map of the circle $\mathbf{S}^1$, and $k \in \mathbf{N}$. Then*

$$\lim_{N\to+\infty} N^2\,\mathrm{Disc}\left(f_*^k(\mathrm{Leb}), (f_N^k)_*(\mathrm{Leb}_N)\right)^2 = \frac{1}{12} + \frac{1}{12}\sum_{m=0}^{k-1}\left\langle D(f^{k-m}), (L_f^m 1)^2\right\rangle, \tag{1.2}$$

where $\langle\cdot,\cdot\rangle$ stands for the L^2 scalar product, L_f is the RPF transfer operator defined by (1.3), and f^{k-m} is the $(k-m)$-th iterate of f.

This asymptotics tells us at which speed the measures $f_*^k(\mathrm{Leb})$ and $(f_N^k)_*(\mathrm{Leb}_N)$ move apart one from the other. More precisely, from this theorem and an estimation of the terms in Lasota-Yorke inequality (see [EG13] or [Gui15a, Theorem 12.17]), one can easily deduce the following.

Corollary 1.1. *Let $r > 1$, and f a generic C^r expanding map of $\mathbf{S}^1$. Then there exist two constants $1 < c < C$, depending only on $\inf f'$ and the C^{r-1} norm of f', such that for any $K \in \mathbf{N}$, there is $N_0 \in \mathbf{N}$ such that for any $N \geq N_0$ and any $k \leq K$, one has*

$$N^2\,\mathrm{Disc}\left(f_*^k(\mathrm{Leb}), (f_N^k)_*(\mathrm{Leb}_N)\right)^2 \in [c^k, C^k].$$

In other words, for K fixed, if N is large enough, then the behaviour of the discrepancy will be typically exponential for times smaller than K.

[3] Called "the L_p-metrics between distribution functions" in [Rac91].

[4] Following Lanford in [Lan98], it is interesting to note that this hypothesis of genericity, which ensures some uniform repartition properties at a mesoscopic scale (e.g. Proposition 3.7), is also a classical assumption for the problem of deriving fluid mechanics laws from a microscopical deterministic model. See for example [BGSR16].

Let us describe a bit the term $f_*^k(\mathrm{Leb})$. As realized quite a long time ago by physicists, and then in the 70's by mathematicians including David Ruelle, the action of a hyperbolic map on measures can be discribed with the help of the *Ruelle-Perron-Frobenius operator* L_f, defined by (in the case of the circle)

$$L_f\phi : y \mapsto \sum_{f(x)=y} \frac{\phi(x)}{f'(x)}. \tag{1.3}$$

The crucial remark is that if ϕ is the density of some measure μ on $\mathbf{S}^1$, then $L_f\phi$ is the density of the measure $f_*\mu$. A now quite large literature is devoted to link the spectral properties of this operator L_f on suitable Banach spaces and the ergodic behaviour of the dynamics f. In particular, Ruelle proved that if f is a $C^{1+\alpha}$ expanding map of the circle, then the functions $L_f^k(1)$ converge exponentially fast – for the C^α topology – towards the density of a measure called SRB (for Sinaï-Ruelle-Bowen). This measure is moreover the unique absolutely continuous f-invariant probability measure, and the unique physical measure[5] [Via97]. In summary, the measures $f_*^k(\mathrm{Leb})$ will converge "exponentially" towards SRB. Together with Theorem A, this implies that there is some regime $k \ll N$ in which the measures $(f_N^k)_*(\mathrm{Leb}_N)$ converge towards SRB.

Corollary 1.2. *Let $r > 1$, and f a generic C^r expanding map of $\mathbf{S}^1$. Then there exists a constant $C > 1$, depending only on $\inf f'$ and the C^{r-1} norm of f', such that for any $\varepsilon > 0$ and any $k \geq -C \log \varepsilon$, there is $N_0 \in \mathbf{N}$ such that for any $N \geq N_0$, one has*

$$\mathrm{Disc}\left(SRB, (f_N^k)_*(\mathrm{Leb}_N)\right) \leq \varepsilon.$$

This corollary is in fact quite trivial and can be obtained from a direct computation without genericity assumption (see [Gui15a, Theorem 12.19] for an explicit statement). Theorem A specifies this convergence by estimating the discrepancy between continuous and discretized dynamics.

Of course, the discretization procedure studied in this paper is very primitive and inefficient to actually compute SRB measures for circle expanding maps, compared to some other algorithms like the Ulam approximation (see [GHR12] for a survey about this subject, and our second paper [GM22]). Our aim here is to describe to what extent the very naive algorithm for computing SRB measures actually computes an approximation of this measure or not.

Overview of the paper. The proof of Theorem A follows the strategy of [Gui19]: as pointed out by Lanford in [Lan98], it is very fruitful to mimic the proof strategy for the problem consisting in deducing the laws of fluid dynamics from a microscopic model of a gas (hydrodynamic limit), by introducing an intermediate mesoscopic space scale between the microscopic scale $1/N$ of the grid and the macroscopic scale.

So we will start with the study of the corresponding local problem, i.e. the linear case (Sections 3 and 4). First, we will define the corresponding notion of discretization for sequences of linear maps of $\mathbf{R}$. The first step is to link the discrepancy with the roundoff

[5] A measure μ is called *physical* if there is a positive Lebesgue measure set of points $x \in \mathbf{S}^1$ whose Birkhoff sums $\frac{1}{n}\sum_{k=0}^{n-1} \delta_{f^k(x)}$ converge weakly towards μ.

errors made at each iteration (Proposition 3.5). It turns out that under generic conditions, these roundoff errors are statistically uniformly distributed (Proposition 3.2), fact which had previously been obtained by Flockermann in his thesis [Flo02, Theorem 10 page 44]; we will here give two new proofs of this result in Subsections 3.4 and 3.5. As in [Gui19], both are based on ideas from the theory of quasicrystals; the first one follows a direct approach[6], while the second one gives more precise results that allow to get a formula for the cumulated difference (3.3) for time k at some point $x \in \mathbf{Z}$ (which is used to define the discrepancy). This formula involves the value of a piecewise linear functional on a k-dimensional torus at some point depending explicitly on x (Proposition 3.9). This proposition is somehow the heart of the paper, as the first approach fails when one tries to pass to the more complicated framework of the tree linear case.

The set of time-k preimages of a point $x \in \mathbf{S}^1$ by an expanding map f has a structure of complete d-ary tree of height k, where d is the degree of f. So the next step is to study a model of discretizations of linear maps that decorate such a complete d-ary tree (where the coefficients of the linear maps correspond to the derivatives of the expanding map). This is done in Section 4, where we will deduce the behaviour of the discrepancy in this framework from the study conducted in the previous section.

Finally, in Section 5, we will use these results to prove Theorem A. It will be done in two steps. First, we will get a formula involving the derivatives of f along paths of the preimage tree. It will be achieved by combining Thom's transversality theorem for generic maps with the study of the linear case and some suitable application of Taylor's formula. This third tool is elementary but rather technical, and will be obtained by applying a result of [Gui19]. The second step is quite elementary and will allow us to write the formula of Theorem 5.1 in the nicer way of Theorem A.

Numerical experiments. In [GM22] we conduct some numerical experiments relative to the present article. Our aim is twofold: establish the time scale where Theorem A stays valid on some actual examples, and for bigger times try to determine numerically the phenomena underlying the behaviour of (1.1).

It turns out that on the examples we tested, Theorem A stays valid until times k typically logarithmic in N, and that in the regime where $k \geq \log N$ the evolution of (1.1) is not satisfyingly described by some model involving only random perturbations of the dynamics: the fact (involved in the prof of Theorem A) that orbits that merge stay together forever thereafter has a significant impact on the discrepancy (1.1). In [GM22] we propose a model taking this phenomenon into account. We conjecture that this model captures sufficiently well the relevant features of discretizations in the middle-term range to approximate well the evolution of (1.1).

Context. To our knowledge, the first attempt of numerical approximation of physical measures of some Anosov diffeomorphisms dates back to le late 70's, with the works [BCG$^+$79, BCG$^+$78] where the authors study among others the Arnold cat map and some of its perturbations.

The idea of O.E. Lanford consisting in adjusting the length of orbit segments to the discretization order had already been developed in the late 80's by Abraham Boyarsky. In

[6]This direct approach was already presented in the thesis [Gui15a, Chapter 9].

[Boy86], he explains heuristically why one usually finds absolutely continuous measures on simulations. His arguments are based on the tracking of long segments of orbits; the only obstacle for the obtaining of a rigorous proof is the lack of uniformity in Birkhoff 's ergodic theorem[7].

In [GB88], Boyarsky together with Pawel Góra establish the following result, which also relates to the obtaining of absolutely continuous measures from discretizations. *Suppose that f has a unique absolutely continuous invariant measure μ, and that there exists $\alpha > 0$ such that there is a subsequence of f_N admitting a segment of orbit of length bigger than $\alpha \operatorname{Card}(E_N)$. If we denote ν_N the uniform measure on this segment of orbit, then $\nu_N \rightharpoonup \mu$.*

The existence of an orbit segment of length proportional to that of the grid seems to be rarely verified (for example it is not true for a generic circle expanding map, simply because the degree of recurrence is zero). Despite this, this seems to be one of the first theoretical results about discretizations of dynamical systems.

The rigourous study of discretizations of *generic* dynamics has first been proposed by Étienne Ghys in the large audience article [Ghy94]. From this viewpoint, the case of circle homeomorphisms is now quite well understood, due to the work of Tomasz Miernowski [Mie06], whose conclusion is essentially that the discretizations' dynamics resemble the homeomorphism's one (see also the recent preprint [GS21]). The higher dimensional case has been tackled by the first author in his thesis [Gui15a], which includes the case of generic homeomorphisms [Gui15b], C^1 diffeomorphisms [Gui18] and C^r diffeomorphisms and expanding maps[8] [Gui19]. In particular, the article [Gui18] exhibits the following quite unexpected phenomenon[9].

Theorem. *Take a point $x \in \mathbf{S}^1$, and a generic C^1 circle expanding map f. The orbit of x under f_N is finite thus eventually periodic; denote μ_N^x the uniform measure on the limit periodic orbit. Then the sequence $(\mu_N^x)_{N\geq 0}$ accumulates on the whole set of f-invariant probability measures.*

Hence, the ergodic behaviour under the discretization f_N of a (Baire) typical point of the circle does not converge towards the unique physical measure (which exists and is singular, see [CQ01, Qua99]). However, one can expect some more convergent behaviour by averaging over $x \in \mathbf{S}^1$. This leads to the following question.

Question. *For $r > 1$ and a generic C^r expanding map of $\mathbf{S}^1$, do the measures*[10]

$$\lim_{k\to+\infty} \frac{1}{k} \sum_{i=0}^{k-1} (f_N^i)_*(\mathrm{Leb}_N)$$

converge, when the parameter N goes to infinity, towards the SRB measure of f?

This question seems out of reach with only the techniques used in this paper. In the second article of this series [GM22], we show numerical simulations that suggest that the answer to this question may be yes in general.

[7] In fact, the intuition of Boyarsky works for a uniquely ergodic homeomorphism, as proved by Miernowski in [Mie06]. Unfortunately, his result is false in general (see [Gui18]).

[8] As already explained, the proofs of the present article are based on the strategy of this paper.

[9] A similar statement holds for generic measure-preserving C^1-diffeomorphisms.

[10] The convergence of these measures in k is ensured by the finiteness of the map f_N.

Acknowledgements. This project was partially supported by a PEPS/CNRS project and the ANR CODYS. The authors warmly thank Nina Heloin for his careful reading of a first version of this text.

2. Preliminaries: distances on measures

Let μ and ν be two probability measures defined on $\mathbf{S}^1 = \mathbf{R}/\mathbf{Z}$, identified with $[0,1[$. Let F and G be their respective cumulative distribution functions, and $H = F - G$

Definition 2.1. The L^1 Wasserstein distance between μ and ν can be defined by the formula[11]

$$W_1(\mu,\nu) = \min_{c\in\mathbf{R}} \int_0^1 |H-c|,$$

and the minimum is realized by the median of $F-G$, i.e.

$$c_0 = \frac{1}{2}\Big(\sup\big\{c \mid \mathrm{Leb}(H<c) < \mathrm{Leb}(H>c)\big\} + \inf\big\{c \mid \mathrm{Leb}(H<c) > \mathrm{Leb}(H>c)\big\}\Big).$$

Similarly we define another distance on the set of probability measures on $\mathbf{S}^1$, which we call *discrepancy*:

$$\mathrm{Disc}(\mu,\nu) = \left(\min_{c\in\mathbf{R}} \int_0^1 (H-c)^2\right)^{1/2}, \tag{2.1}$$

and the minimum is realized by the mean of $F-G$, i.e. by the number $c_1 = \int_0^1 H$:

$$\mathrm{Disc}(\mu,\nu) = \left(\int_0^1 \left(H(x) - \Big(\int_0^1 H\Big)\right)^2 \mathrm{d}x\right)^{1/2}.$$

Remark that this last expression looks like a variance, and this fact will be useful in the sequel. The proofs of the statements about the numbers c realizing the minima are simple and left to the reader.

Lemma 2.2. *The discrepancy* Disc *is a distance which is invariant under translation. Moreover,*

$$W_1(\mu,\nu) \le \mathrm{Disc}(\mu,\nu)$$

and

$$\mathrm{Disc}(\mu,\nu) \le \sqrt{2} W_1(\mu,\nu)^{1/2}.$$

Thus W_1 and Disc *span the same topology (the weak-*).*

Proof of Lemma 2.2. First we prove that Disc is invariant under translation. More precisely, for any $a \in [0,1]$, we let F_a and G_a be the cumulative distribution functions of μ and ν seen as measures on $[a, a+1]$, and set $H_a = F_a - G_a : [a, a+1] \to \mathbf{R}$. What we want to prove is that for any $a \in [0,1]$, one has

$$\mathrm{Disc}(\mu,\nu)^2 = \int_a^{a+1} \left(H_a(x) - \Big(\int_a^{a+1} H_a\Big)\right)^2 \mathrm{d}x.$$

[11] In [CM95] it is explained why this formula coincides with the classical definition of the Wasserstein distance.

One has $H_a(u) = \int_a^u \mathrm{d}(\mu-\nu)$, thus, using the fact that $\int_0^1 \mathrm{d}(\mu-\nu) = 0$,

$$H_a(u) = \begin{cases} \int_0^u \mathrm{d}(\mu-\nu) - \int_0^a \mathrm{d}(\mu-\nu) = H(u) - H(a) & \text{if } u \leq 1 \\ \int_1^u \mathrm{d}(\mu-\nu) + \int_a^1 \mathrm{d}(\mu-\nu) = H(u-1) - H(a) & \text{if } u \geq 1 \end{cases}$$

Hence, $\int_a^{a+1} H_a = \int_0^1 (H - H(a))$, which implies that

$$H_a(u) - \int_a^{a+1} H_a = H(u \mod 1) - \int_0^1 H,$$

and thus

$$\int_a^{a+1} \left(H_a(x) - \Big(\int_a^{a+1} H_a \Big) \right)^2 \mathrm{d}x = \int_0^1 \left(H(x) - \Big(\int_0^1 H \Big) \right)^2 \mathrm{d}x.$$

We now come to the proof of the inequalities. The first one is simply Cauchy-Schwarz inequality applied to the map $H - c_1$.

For the second one, remark that $H \in [-1,1]$, so that $c_0 \in [-1,1]$ and $|H - c_0| \leq 2$. Hence, $(H-c_0)^2 \leq 2|H-c_0|$ and

$$\mathrm{Disc}(\mu,\nu)^2 \leq \int_0^1 (H-c_0)^2 \leq 2 \int_0^1 |H - c_0| = 2W_1(\mu,\nu)$$

□

3. The simple linear case

Recall that the goal of this section is to treat the corresponding problem of discrepancy between actual and discretized systems, for sequences of linear maps. We first set definitions, which are made to mimic the ones on the circle for the case of linear maps of $\mathbf{R}$.

3.1. Definitions.

We denote $\mathbf{N} = \{0, 1, 2, \dots\}$.

Discretizations of linear maps. Let $(\ell_m)_{m \geq 1}$ be a sequence of homotheties of $\mathbf{R}$, of parameters $\lambda_m > 1$, i.e. $\forall y \in \mathbf{R}$, $\ell_m(y) = \lambda_m y$. We fix $k \in \mathbf{N}^*$.

Definition 3.1. The *discretization* of a linear map $\ell : \mathbf{R} \to \mathbf{R}$ is the map $\widehat{\ell} : \mathbf{Z} \to \mathbf{Z}$ such that for any $x \in \mathbf{Z}$, $\widehat{\ell}(x)$ is the integer closest to $\ell(x)$. More precisely, $\widehat{\ell}(x)$ is the unique integer such that

$$\widehat{\ell}(x) - \ell(x) \in \left] -\frac{1}{2}, \frac{1}{2} \right].$$

We will denote

$$\boldsymbol{\ell}^k = \ell_k \circ \cdots \circ \ell_1, \qquad \widehat{\boldsymbol{\ell}}^k = \widehat{\ell}_k \circ \cdots \circ \widehat{\ell}_1 \qquad \text{and} \qquad \widetilde{\lambda}_m = \prod_{i=m+1}^{k} \lambda_i \tag{3.1}$$

(with the convention that $\widetilde{\lambda}_k = 1$).

Expectation, covariance. The expectation $\mathbb{E}$ and the covariance Var of a map $\mathcal{E} : \mathbf{N} \to \mathbf{R}$ are defined by (whenever the limits make sense)

$$\mathbb{E}[\mathcal{E}] = \lim_{R\to+\infty} \frac{1}{R} \sum_{x=0}^{R-1} \mathcal{E}_x \qquad \text{and} \qquad \mathrm{Var}(\mathcal{E}) = \lim_{R\to+\infty} \frac{1}{R} \sum_{x=0}^{R-1} \big(\mathcal{E}_x - \mathbb{E}[\mathcal{E}]\big)^2. \tag{3.2}$$

Cumulated difference and discrepancy on $\mathbf{R}$. Let μ and ν be two measures on $\mathbf{R}$. Their *cumulated difference* at $y > 0$ is the number

$$c\delta_y(\mu,\nu) = \mu\big(]0,y]\big) + \frac{1}{2}\mu\big(\{0\}\big) - \nu\big(]0,y]\big) - \frac{1}{2}\nu\big(\{0\}\big). \tag{3.3}$$

We define the *(L^2-) discrepancy* Disc as the L^2-average of the cumulated difference $c\delta$ (3.3) (when the limit exists):

$$\mathrm{Disc}(\mu,\nu) = \lim_{R\to+\infty} \mathrm{Disc}_R(\mu,\nu) \qquad \text{where} \qquad \mathrm{Disc}_R(\mu,\nu) = \left(\frac{1}{R}\int_0^R c\delta_y(\mu,\nu)^2\right)^{1/2}.$$

We will be interested in the case where the measures μ and ν are respectively:

- the (correctly normalized) Lebesgue measure $\widetilde{\lambda}_0^{-1}\,\mathrm{Leb}$;
- the uniform measure on the image set $\widehat{\boldsymbol{\ell}}^k(\mathbf{Z})$, that is $\sum_{n\in\mathbf{Z}} \delta_{\widehat{\boldsymbol{\ell}}^k(n)}$

In the sequel we will denote, when no confusion is possible,

$$\begin{aligned} c\delta_y &\overset{\text{def.}}{=} c\delta_y\left(\widetilde{\lambda}_0^{-1}\,\mathrm{Leb},\ \sum_{n\in\mathbf{Z}} \delta_{\widehat{\boldsymbol{\ell}}^k(n)}\right) \\ &= \frac{y}{\widetilde{\lambda}_0} - \mathrm{Card}\left\{x\in\mathbf{N} \mid \widehat{\boldsymbol{\ell}}^k(x) \le y\right\} + \frac{1}{2}, \end{aligned}$$

and the same for the discrepancies Disc_R and Disc.

The half weight given to the singleton $\{0\}$ restores symmetry and ensures that the map $c\delta$ has zero mean (see Remark 3.10).

3.2. **Roundoff errors.** The roundoff error made at the m-th iteration is defined as the difference between the images of $\boldsymbol{\ell}^{m-1}(x)$ by the discretization $\widehat{\ell}_m$ and the initial map ℓ_m, that is

$$e_x^m = \big(\widehat{\ell}_m - \ell_m\big)\big(\widehat{\boldsymbol{\ell}}^{m-1}(x)\big) \in]-1/2, 1/2].$$

The distribution of the vectors

$$\boldsymbol{\varepsilon}_x^k \overset{\text{def.}}{=} (e_x^1, \dots, e_x^k)$$

when x ranges over $\mathbf{Z}$ is given by the following proposition due to P. P. Flockermann (see the thesis [Flo02], Theorem 10 page 44). We will give two alternative proofs of this proposition, both based on linear algebra (contrary to the original proof of Flockermann).

Proposition 3.2 (Flockermann)**.** *If the family $(\widetilde{\lambda}_m^{-1})_{0\le m\le k}$ is $\mathbf{Q}$-free (which is a generic condition on the λ_i's), then the roundoff error vectors $(\boldsymbol{\varepsilon}_x^k)_{x\in\mathbf{Z}}$ are equidistributed in $]-1/2, 1/2]^k$.*

We postpone the proof of this proposition to Sections 3.4 and 3.5.

From the roundoff errors $\boldsymbol{\varepsilon}_x^k$ it is possible to deduce the global error

$$\mathcal{E}_x^k = \widehat{\boldsymbol{\ell}}^k(x) - \boldsymbol{\ell}^k(x)$$

made after k iterations. Indeed, we have

$$\begin{aligned}\mathcal{E}_x^{k+1} =& \big(\widehat{\ell}_{k+1} - \ell_{k+1}\big)\big(\widehat{\boldsymbol{\ell}}^k(x)\big) + \ell_{k+1}\big(\widehat{\boldsymbol{\ell}}^k(x) - \boldsymbol{\ell}^k(x)\big)\\ =& e_x^{k+1} + \ell_{k+1}(\mathcal{E}_x^k).\end{aligned}$$

From this recurrence relation, we deduce that

$$\mathcal{E}_x^k = \sum_{m=1}^{k} \widetilde{\lambda}_m e_x^m. \tag{3.4}$$

Recall that Proposition 3.2 ensures that when the family $(\widetilde{\lambda}_m^{-1})_{1\le m\le k}$ is $\mathbf{Q}$-free, then the errors e_x^m are independent and identically distributed in $[-1/2, 1/2]$ (because $\boldsymbol{\varepsilon}_x^k$ is equidistributed on the product space $[-1/2, 1/2]^k$). From that we deduce the law of the global error $\mathcal{E}^k$, and in particular its covariance

$$\operatorname{Var}(\mathcal{E}^k) = \sum_{m=1}^{k} \widetilde{\lambda}_m^2 \operatorname{Var}(e^m) = \frac{1}{12}\sum_{m=1}^{k} \widetilde{\lambda}_m^2. \tag{3.5}$$

In particular, if there exists $\alpha > 1$ such that $\lambda_m \ge \alpha$ for every m, then

$$\operatorname{Var}(\mathcal{E}^k) \ge \frac{1}{12}\sum_{m=1}^{k} \alpha^{2m} = \frac{\alpha^2(\alpha^{2k}-1)}{12(\alpha^2-1)}.$$

3.3. **Discrepancy and roundoff errors.** In this subsection, we link the asymptotic behaviours of the discrepancy Disc with that of the roundoff errors. We state all properties before proving them. Note that some of these proofs will use Proposition 3.2, which will be proved later on without using the results of this subsection.

The first lemma says that the mean of the map $c\delta$ is zero.

Lemma 3.3.

$$\mathbb{E}[c\delta_{x+1/2}] = \lim_{R\to+\infty} \frac{1}{R}\int_0^R c\delta_y \,\mathrm{d}y = 0.$$

Remark that we have two different notions of means: one continuous (with an integral) and one discrete (with a sum). Both can be easily related: the following lemma links the discrepancy Disc which is obtained as a continuous average of the cumulated difference $c\delta$, with the variance of the map $c\delta$ taken on half integers.

Lemma 3.4. *Whenever the discrepancy and the variance make sense,*

$$\operatorname{Disc}^2 = \frac{1}{12\widetilde{\lambda}_0^2} + \operatorname{Var}_{x\in\mathbf{Z}}\left(c\delta_{x+\frac{1}{2}}\right). \tag{3.6}$$

Finally, the following proposition deals with the covariance: it links the average discrepancy Disc with the covariance of $\mathcal{E}$.

Proposition 3.5. *Let $k \in \mathbf{N}$, and a family $(\ell_m)_{1\le m\le k}$ of of $\mathbf{R}$ of parameters $(\lambda_m)_{1\le m\le k}$ strictly bigger than 1. Whenever the discrepancy and the variance make sense,*

$$\mathrm{Disc}^2 = \frac{1}{12} + \frac{1}{\widetilde{\lambda}_0^2}\,\mathrm{Var}(\mathcal{E}^k).$$

Note that the factor $1/12$ corresponds to the covariance of the uniform distribution on the interval $[-1/2, 1/2]$.

Combined with Proposition 3.2 (more precisely, Equation (3.5)), this immediately gives the following corollary.

Corollary 3.6. *If the family $(\widetilde{\lambda}_m)_{0\le m\le k}$ is $\mathbf{Q}$-free, then*

$$\mathrm{Disc}^2 = \frac{1}{12\widetilde{\lambda}_0^2}\sum_{m=0}^{k}\widetilde{\lambda}_m^2.$$

Proof of Lemma 3.3. The first equality comes from the fact that the map $c\delta$ is affine with slope $\widetilde{\lambda}_0^{-1}$ in restriction to any interval which contains no integer. We will see a more detailed proof of a very similar fact during the proof of Lemma 3.4.

We are left to prove the second equality. Let

$$\mathcal{E}'_x = \widetilde{\lambda}_0^{-1}\mathcal{E}_x^k.$$

Fix $R > 0$, set $R' = \widetilde{\lambda}_0 R$, and denote $x_0 = x_0(R)$ the biggest integer x such that $x+\mathcal{E}'_x \le R$. A linear change of variables leads to

$$\begin{aligned}\frac{1}{R'}\int_0^{R'} c\delta_{y'}\,\mathrm{d}y' &= \frac{1}{R'}\int_0^{R'}\frac{y'}{\widetilde{\lambda}_0} + \frac{1}{2} - \sum_{x=0}^{x_0}\mathbf{1}_{\ell^k(x)+\mathcal{E}_x\le y'}\,\mathrm{d}y'\\ &= \frac{1}{R}\int_0^{R}\left(y + \frac{1}{2} - \sum_{x=0}^{x_0}\mathbf{1}_{x+\mathcal{E}'_x\le y}\right)\mathrm{d}y.\end{aligned}$$

Hence,

$$\begin{aligned}\frac{1}{R'}\int_0^{R'} c\delta_{y'}\,\mathrm{d}y' &= \frac{1}{R}\left(\frac{R^2}{2} + \frac{R}{2} - \sum_{x=0}^{x_0}\int_0^R \mathbf{1}_{y\ge x+\mathcal{E}'_x}\,\mathrm{d}y\right)\\ &= \frac{R}{2} + \frac{1}{2} - \frac{1}{R}\sum_{x=0}^{x_0}(R - x - \mathcal{E}'_x)\\ &= \frac{R}{2} + \frac{1}{2} - (x_0+1) + \frac{x_0(x_0+1)}{2R} + \frac{1}{R}\sum_{x=0}^{x_0}\mathcal{E}'_x\\ &= \frac{(R-x_0)^2 - (R-x_0)}{2R} + \frac{x_0}{R}\frac{1}{x_0}\sum_{x=0}^{x_0}\mathcal{E}'_x\end{aligned}$$

But $|R - x_0|$ is uniformly bounded on R (because $\mathcal{E}'_x$ is uniformly bounded on R), so the first term tends to 0 when R' goes to infinity. The second term, for itself, tends to the mean of $x \mapsto \mathcal{E}'_x$, which is zero by Proposition 3.2. □

Proof of Lemma 3.4. The proof simply consists in remarking that the map $c\delta$ is affine with slope $\widetilde{\lambda}_0^{-1}$ in restriction to any interval that contains no integer. When $R \in \mathbf{N}$, one has

$$\begin{aligned}\frac{1}{R}\int_0^R c\delta_y^2 \,\mathrm{d}y =& \frac{1}{R}\sum_{x=0}^{R-1}\int_{-1/2}^{1/2}\left(c\delta_{x+\frac{1}{2}} + \frac{y}{\widetilde{\lambda}_0}\right)^2 \mathrm{d}y \\ =& \frac{1}{R}\sum_{x=0}^{R-1} c\delta_{x+\frac{1}{2}}^2 + \frac{1}{12\widetilde{\lambda}_0^2}\end{aligned}$$

But by Proposition 3.3], one has $\mathbb{E}\left[c\delta_{k+\frac{1}{2}}\right] = 0$; this gives directly the lemma. □

Proof of Proposition 3.5. We reuse the notations of proof of Lemma 3.3: we set $\mathcal{E}'_x = \widetilde{\lambda}_0^{-1}\mathcal{E}_x$, $R' = \widetilde{\lambda}_0 R$, and denote x_0 the biggest integer x such that $x + \mathcal{E}'_x \le R$.

We will prove that

$$\mathrm{Disc}_R^2\left(\widehat{\ell}^k(\mathbf{Z})\,,\, \widetilde{\lambda}_0^{-1}\,\mathrm{Leb}\right) \underset{R\to+\infty}{\longrightarrow} \frac{1}{12} + \mathrm{Var}(\mathcal{E}').$$

By Equation (3.3), we have

$$\mathrm{Disc}_{R'}^2 = \frac{1}{R'}\int_0^{R'}\left(y'\widetilde{\lambda}_0^{-1} - \sum_{x=0}^{x_0}\mathbf{1}_{\widetilde{\lambda}_0 x + \mathcal{E}_x \le y'} + \frac{1}{2}\right)^2 \mathrm{d}y'$$

A linear change of variables leads to

$$\begin{aligned}\mathrm{Disc}_{R'}^2 &= \frac{1}{R}\int_0^R\left(y + \frac{1}{2} - \sum_{x=0}^{x_0}\mathbf{1}_{x+\mathcal{E}'_x\le y}\right)^2 \mathrm{d}y \\ &= \frac{1}{R}\int_0^R\left(\left(y+\frac{1}{2}\right)^2 - 2\left(y+\frac{1}{2}\right)\sum_{x=0}^{x_0}\mathbf{1}_{x+\mathcal{E}'_x\le y} + \sum_{x,x'=0}^{x_0}\mathbf{1}_{x+\mathcal{E}'_x\le y}\mathbf{1}_{x'+\mathcal{E}'_{x'}\le y}\right)\mathrm{d}y \\ &= \frac{1}{R}\Bigg[\int_0^R\left(y+\frac{1}{2}\right)^2\mathrm{d}y - 2\sum_{x=0}^{x_0}\int_0^R\left(y+\frac{1}{2}\right)\mathbf{1}_{x+\mathcal{E}'_x\le y}\,\mathrm{d}y \\ &\qquad + \sum_{x,x'=0}^{x_0}\int_0^R \mathbf{1}_{x+\mathcal{E}'_x\le y}\mathbf{1}_{x'+\mathcal{E}'_{x'}\le y}\,\mathrm{d}y\Bigg] \\ &= \frac{1}{R}\left[\int_0^R\left(y+\frac{1}{2}\right)^2\mathrm{d}y - 2\sum_{x=0}^{x_0}\int_{x+\mathcal{E}'_x}^R\left(y+\frac{1}{2}\right)\mathrm{d}y + \sum_{x,x'=0}^{x_0}\int_{\max\left(x+\mathcal{E}'_x,x'+\mathcal{E}'_{x'}\right)}^R 1\,\mathrm{d}y\right] \\ &= \frac{1}{R}\Bigg[\frac{\left(R+\frac{1}{2}\right)^3}{3} - \frac{\left(\frac{1}{2}\right)^3}{3} - \sum_{x=0}^{x_0}\left(\left(R+\frac{1}{2}\right)^2 - \left(x+\mathcal{E}'_x+\frac{1}{2}\right)^2\right) \\ &\qquad + \sum_{x,x'=0}^{x_0}\left(R - \max\left(x+\mathcal{E}'_x, x'+\mathcal{E}'_{x'}\right)\right)\Bigg].\end{aligned}$$

But by construction, the map $x \mapsto x + \mathcal{E}'_x$ is increasing, so the last sum can be reindexed:

$$\sum_{x,x'=0}^{x_0} \Big(R - \max\big(x + \mathcal{E}'_x, x' + \mathcal{E}'_{x'}\big)\Big) = \sum_{m=0}^{x_0} \sum_{\substack{\max(x,x')=m \\ x,x' \geq 0}} \big(R - m - \mathcal{E}'_m\big)$$
$$= \sum_{m=0}^{x_0} (2m+1)\big(R - m - \mathcal{E}'_m\big),$$

so one gets:

$$\begin{aligned}
\mathrm{Disc}^2_{R'} =& \frac{1}{R}\left[\frac{R^3}{3} + \frac{R^2}{2} + \frac{R}{4} - \sum_{x=0}^{x_0}\left(R^2 + R + \frac{1}{4} - \left(x^2 + {\mathcal{E}'_x}^2 + \frac{1}{4} + 2x\mathcal{E}'_x + x + \mathcal{E}'_x\right)\right)\right. \\
&\left. + \sum_{x=0}^{x_0}(2x+1)\big(R - (x + \mathcal{E}'_x)\big)\right] \\
=& \frac{1}{R}\left[\frac{R^3}{3} + \frac{R^2}{2} + \frac{R}{4} + \sum_{x=0}^{x_0}\left(- R^2 - x^2 + 2xR + {\mathcal{E}'_x}^2\right)\right] \\
=& \frac{1}{R}\left[\frac{R^3}{3} + \frac{R^2}{2} + \frac{R}{4} - R^2(x_0+1) - \left(\frac{x_0^3}{3} + \frac{x_0^2}{2} + \frac{x_0}{6}\right) + 2R\frac{x_0(x_0+1)}{2} + \sum_{x=0}^{x_0} {\mathcal{E}'_x}^2\right] \\
=& \frac{1}{R}\left[\frac{(R-x_0)^3}{3} - \frac{(R-x_0)^2}{2} + \frac{R-x_0}{6}\right] + \frac{1}{12} + \frac{1}{R}\sum_{x=0}^{x_0} {\mathcal{E}'_x}^2
\end{aligned}$$

As in the proof of Lemma 3.3, we have that $|R - x_0|$ is uniformly bounded in R, so that the first term tends to 0 as R' goes to infinity. As the mean of $x \mapsto \mathcal{E}'_x$ is 0 (see Proposition 3.7), we get finally that

$$\lim_{R' \to +\infty} \mathrm{Disc}^2_{R'} = \frac{1}{12} + \mathrm{Var}(\mathcal{E}').$$

□

3.4. **Uniform distribution of errors: first proof.** This subsection presents the first proof of uniform distribution of errors. It is easier than the one we will see in the next section, and consists in computing projections on k-dimensional tori of vectors depending on the initial condition x.

Fix $k \geq 0$. Recall that we have $\boldsymbol{\varepsilon}^k_x = \big(e^1_x, \cdots, e^k_x\big)$. Moreover, we set $\widehat{\boldsymbol{\ell}}_x = \big(\widehat{\boldsymbol{\ell}}^1(x), \cdots, \widehat{\boldsymbol{\ell}}^k(x)\big)$ the vectors made of the k firsts iterates of x under the discretizations, and denote $u_x = (\lambda_1 x, 0^{k-1}) \in \mathbf{R}^k$, $u_{\mathbf{Z}} = (\lambda_1 \mathbf{Z}, 0^{k-1}) \subset \mathbf{R}^k$, $W^k =]-1/2, 1/2]^k$ and

$$N_{\lambda_1,\cdots,\lambda_k} = \begin{pmatrix} -1 & & & \\ \lambda_2 & -1 & & \\ & \lambda_3 & \ddots & \\ & & \ddots & -1 \\ & & & \lambda_k \quad -1 \end{pmatrix} \in M_k(\mathbf{R}).$$

Finally, we denote pr_{W^k} the projection from $\mathbf{R}^k$ onto the fundamental domain W^k of the quotient space $\mathbf{R}^k/N_{\lambda_1,\cdots,\lambda_k}\mathbf{Z}^k$ (this is indeed a fundamental domain because the matrix $N_{\lambda_1,\cdots,\lambda_k}$ is lower triangular with -1 on the diagonal; remark that in this case the matrix satisfies the conclusion of Hajós theorem [Haj41]).

The following proposition expresses the roundoff error vector $\boldsymbol{\varepsilon}_x^k$ in terms of the projection of the vector u_x on the fundamental domain W^k of $\mathbf{R}^k/N_{\ell_1,\cdots,\ell_k}\mathbf{Z}^k$.

Proposition 3.7.

$$\boldsymbol{\varepsilon}_x^k = \mathrm{pr}_{W^k}(u_x).$$

Thus, when x ranges over $\mathbf{Z}$, the roundoff error vectors $\boldsymbol{\varepsilon}_x^k$ equidistribute on the set $\overline{\mathrm{pr}_{W^k}(u_{\mathbf{Z}})}$.

In particular, as this set is symmetric with respect to 0, the means of each function $x \mapsto e_x^m$, and of $x \mapsto \mathcal{E}_x$, is zero.

Proof of Proposition 3.7. As

$$N_{\lambda_1,\cdots,\lambda_k}\widehat{\boldsymbol{\ell}}_x = \begin{pmatrix} -\widehat{\boldsymbol{\ell}}^1(x) \\ \lambda_2\widehat{\boldsymbol{\ell}}^1(x) - \widehat{\boldsymbol{\ell}}^2(x) \\ \lambda_3\widehat{\boldsymbol{\ell}}^2(x) - \widehat{\boldsymbol{\ell}}^3(x) \\ \vdots \\ \lambda_k\widehat{\boldsymbol{\ell}}^{k-1}(x) - \widehat{\boldsymbol{\ell}}^k(x) \end{pmatrix} = \begin{pmatrix} \lambda_1 x - e_x^1 \\ -e_x^2 \\ -e_x^3 \\ \vdots \\ -e_x^k \end{pmatrix} = u_x - \boldsymbol{\varepsilon}_x^k,$$

the vector u_x can be decomposed into

$$u_x = N_{\lambda_1,\cdots,\lambda_k}\widehat{\boldsymbol{\ell}}_x + \boldsymbol{\varepsilon}_x^k, \tag{3.7}$$

with $\widehat{\boldsymbol{\ell}}_x \in \mathbf{Z}^k$ and $\boldsymbol{\varepsilon}_x^k \in W^k$ (recall that $W^k =]-1/2, 1/2]^k$). As W^k is a fundamental domain of $N_{\lambda_1,\cdots,\lambda_k}\mathbf{Z}^k$, this is a decomposition of u_x into the sum of an element of the lattice $N_{\lambda_1,\cdots,\lambda_k}\mathbf{Z}^k$ and an element of a fundamental domain of this lattice.

The vector u_x being fixed, this condition characterizes completely $\boldsymbol{\varepsilon}_x^k$ and $\widehat{\boldsymbol{\ell}}_x$. In particular, $\boldsymbol{\varepsilon}_x^k$ is equal to the projection of u_x on W^k modulo $N_{\lambda_1,\cdots,\lambda_k}\mathbf{Z}^k$. This implies that the roundoff error vectors $\boldsymbol{\varepsilon}_x^k$ equidistribute on the set $\overline{\mathrm{pr}_{W^k}(u_{\mathbf{Z}})}$ when x ranges over $\mathbf{Z}$. □

Let us explain how this proposition implies Proposition 3.2.

Proof of Proposition 3.2. We begin by remarking that by (3.7), $N^{-1}_{\lambda_1,\cdots,\lambda_k}\boldsymbol{\varepsilon}_x^k$ is equal to the projection of $N^{-1}_{\lambda_1,\cdots,\lambda_k}u_x$ on $N^{-1}_{\lambda_1,\cdots,\lambda_k}W^k$ modulo $\mathbf{Z}^k$ (remark that $N^{-1}_{\lambda_1,\cdots,\lambda_k}W^k$ is a fundamental domain of $\mathbf{Z}^k$). This implies that the sequence of errors $\boldsymbol{\varepsilon}_x^k$ is equidistributed in $\mathbf{R}^k/\mathbf{Z}^k$ if and only if the vectors $N^{-1}_{\lambda_1,\cdots,\lambda_k}u_x$ are equidistributed modulo $\mathbf{Z}^k$ when x ranges over $\mathbf{Z}$. For this purpose, the matrix $N^{-1}_{\lambda_1,\cdots,\lambda_k}$ can be easily computed:

$$N^{-1}_{\lambda_1,\cdots,\lambda_k} = \begin{pmatrix} -1 & & & & \\ -\lambda_2 & -1 & & & \\ -\lambda_3\lambda_2 & -\lambda_3 & \ddots & & \\ \vdots & \vdots & \ddots & -1 & \\ -\lambda_k\cdots\lambda_2 & -\lambda_k\cdots\lambda_3 & \cdots & -\lambda_k & -1 \end{pmatrix},$$

thus

$$N^{-1}_{\lambda_1,\cdots,\lambda_k} u_x = -\begin{pmatrix} \lambda_1 \\ \lambda_2\lambda_1 \\ \lambda_3\lambda_2\lambda_1 \\ \vdots \\ \lambda_k\cdots\lambda_1 \end{pmatrix} x = -\widetilde{\lambda}_0 \begin{pmatrix} \widetilde{\lambda}_1^{-1} \\ \widetilde{\lambda}_2^{-1} \\ \vdots \\ \widetilde{\lambda}_k^{-1} \end{pmatrix} x.$$

As a consequence, by Weyl's criterion, the sequences of errors $\boldsymbol{\varepsilon}^k_x$ is equidistributed in W^k if and only if the family $(\widetilde{\lambda}_m^{-1})_{0\leq m\leq k}$ is $\mathbf{Q}$-free. □

3.5. **Discrepancy: a direct approach.** In the last subsection, we were given $x \in \mathbf{Z}$ and computed the roundoff errors along the positive orbit of x. Now, we adopt a different viewpoint: we are given $n \in \mathbf{Z}$ and want to determine whether n belongs to $\widehat{\boldsymbol{\ell}}^k(\mathbf{Z})$ or not; in the latter case we also want to determine the sequence of roundoff errors in the backward orbit of n. As in the previous section, we will see that these quantities only depend on the projection on some torus of a vector depending only on n. This will allow to compute the cumulated difference $c\delta_n$ from this projection.

Notations. Recall that $W^k =]-1/2, 1/2]^k$. We denote $\Lambda_k = M_{\lambda_1,\cdots,\lambda_k}\mathbf{Z}^{k+1}$, with

$$M_{\lambda_1,\cdots,\lambda_k} = \begin{pmatrix} \lambda_1 & -1 & & & \\ & \lambda_2 & -1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_k & -1 \\ & & & & 1 \end{pmatrix} \in M_{k+1}(\mathbf{R}),$$

and $\widetilde{\Lambda}_k = \widetilde{M}_{\lambda_1,\cdots,\lambda_k}\mathbf{Z}^k$, with

$$\widetilde{M}_{\lambda_1,\cdots,\lambda_k} = \begin{pmatrix} \lambda_1 & -1 & & & \\ & \lambda_2 & -1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_{k-1} & -1 \\ & & & & \lambda_k \end{pmatrix} \in M_k(\mathbf{R}). \tag{3.8}$$

(see Figure 1). Finally, we denote $X_k = \mathbf{R}^k/\widetilde{\Lambda}_k$ the quotient space and pr_{X_k} the projection from $\mathbf{R}^k$ onto X_k. Remark that X_k is a k-dimensional flat torus.

We begin by giving an alternative construction of the image sets $\widehat{\boldsymbol{\ell}}^k(\mathbf{Z})$ in terms of model sets (see [Gui19]). Indeed, denoting p_1 the projection on the k first coordinates and p_2 the projection on the last coordinate

$$\begin{aligned} \widehat{\boldsymbol{\ell}}^k(\mathbf{Z}) &= \{p_2(\lambda) \mid \lambda \in \Lambda_k,\, p_1(\lambda) \in W^k\} \\ &= p_2\Big(\Lambda_k \cap \big(p_1^{-1}(W^k)\big)\Big). \end{aligned} \tag{3.9}$$

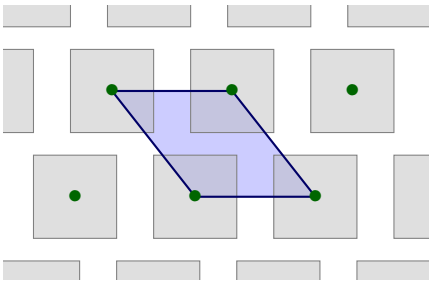

FIGURE 1. The green dots are the points of $\widetilde{\Lambda}_k = \widetilde{M}_{\lambda_1,\cdots,\lambda_k}\mathbf{Z}^k$, the gray squares are $W^k + \widetilde{\Lambda}_k$ and the blue parallelogram is a fundamental domain of $X_k = \mathbf{R}^k/\widetilde{\Lambda}_k$.

Let us explain this construction. Fix a linear map $\ell : \mathbf{R} \to \mathbf{R}$ associated to $\lambda > 1$. An integer $y \in \mathbf{Z}$ belongs to $\widehat{\ell}(\mathbf{Z})$ iff there exists $x \in \mathbf{Z}$ such that $|\ell(x) - y| \leq 1/2$. The last condition can be rephrased as $p_1(v) \in [-1/2, 1/2]$, with

$$v = \begin{pmatrix} \lambda & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

For two linear maps $\ell_1, \ell_2 : \mathbf{R} \to \mathbf{R}$ associated to $\lambda_1, \lambda_2 > 1$, a number $y \in \mathbf{Z}$ belongs to $(\widehat{\ell_2} \circ \widehat{\ell_1})(\mathbf{Z})$ iff there exist $x_1, x_2 \in \mathbf{Z}$ such that $|\ell_2(x_2) - y| \leq 1/2$ and $|\ell_1(x_1) - x_2| \leq 1/2$ (and in this case $(\widehat{\ell_2} \circ \widehat{\ell_1})(x_1) = \widehat{\ell_2}(x_2) = y$). These conditions can be rephrased as $p_1(v) \in [-1/2, 1/2]^2$, with

$$v = \begin{pmatrix} \lambda_1 & -1 & 0 \\ 0 & \lambda_2 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ y \end{pmatrix}.$$

remark that in this case, the roundoff error is given by

$$\boldsymbol{\varepsilon}_x^2 = \begin{pmatrix} x_2 - \lambda_1 x_1 \\ y - \lambda_2 x_2 \end{pmatrix} = -p_1(v),$$

and that $p_2(v) = y$. The same reasoning in arbitrary time k leads to Equation (3.9).

Taking advantage from this viewpoint, we get the following proposition.

Proposition 3.8.

$$y \in \widehat{\boldsymbol{\ell}}^k(\mathbf{Z}) \quad \Longleftrightarrow \quad y \in \mathbf{Z} \text{ and } \mathrm{pr}_{X_k}(0^{k-1}, y) \in \mathrm{pr}_{X_k}(-W^k).$$

In this case, if we denote by $w \in W^k$ the unique point satisfying $\mathrm{pr}_{X_k}(0^{k-1}, y) = \mathrm{pr}_{X_k}(w)$, and $x \in \mathbf{Z}$ the unique integer such that $\widehat{\boldsymbol{\ell}}^k(x) = y$, then the roundoff errors satisfy $\boldsymbol{\varepsilon}_x^k = -w$. As a corollary we get Proposition 3.2.

Proof of Proposition 3.8. By Equation (3.9), we have

$$y \in \widehat{\boldsymbol{\ell}}^k(\mathbf{Z}) \iff y \in \mathbf{Z} \text{ and } \exists v \in \Lambda_k : y = p_2(v),\ p_1(v) \in W^k.$$

But if $y = p_2(v)$, then by the form of the matrix $M_{\lambda_1,\cdots,\lambda_k}$ we can write $v = (\widetilde{v}, 0) + (0^{k-1}, -y, y)$ with $\widetilde{v} \in \widetilde{\Lambda}_k$. Hence,

$$\begin{aligned} y \in \widehat{\boldsymbol{\ell}}^k(\mathbf{Z}) &\iff y \in \mathbf{Z} \text{ and } \exists \widetilde{v} \in \widetilde{\Lambda}_k : (0^{k-1}, -y) + \widetilde{v} \in W^k \\ &\iff y \in \mathbf{Z} \text{ and } (0^{k-1}, y) \in \bigcup_{\widetilde{v} \in \widetilde{\Lambda}_k} \widetilde{v} - W^k. \end{aligned}$$

Thus, $y \in \widehat{\boldsymbol{\ell}}^k(\mathbf{Z})$ if and only if $y \in \mathbf{Z}$ and $\mathrm{pr}_{X_k}(0^{k-1}, y) \in \mathrm{pr}_{X_k}(-W^k)$. Moreover, by construction, $\boldsymbol{\varepsilon}_y^k = -w$.

Then, Proposition 3.2 follows directly from the fact that the points of the form $(0^{k-1}, y)$, with $y \in \mathbf{Z}$, are equidistributed in X_k. To prove this equidistribution, we compute the inverse matrix of $\widetilde{M}_{\lambda_1,\cdots,\lambda_k}$:

$$\widetilde{M}^{-1}_{\lambda_1,\cdots,\lambda_k} = \begin{pmatrix} \lambda_1^{-1} & \lambda_1^{-1}\lambda_2^{-1} & \lambda_1^{-1}\lambda_2^{-1}\lambda_3^{-1} & \cdots & \lambda_1^{-1}\cdots\lambda_k^{-1} \\ & \lambda_2^{-1} & \lambda_2^{-1}\lambda_3^{-1} & \cdots & \lambda_2^{-1}\cdots\lambda_k^{-1} \\ & & \ddots & \ddots & \vdots \\ & & & \lambda_{k-1}^{-1} & \lambda_{k-1}^{-1}\lambda_k^{-1} \\ & & & & \lambda_k^{-1} \end{pmatrix}.$$

Thus, the set of points of the form $(0^{k-1}, y)$ in X_k corresponds to the image of the map

$$\mathbf{Z} \ni y \longmapsto \widetilde{M}^{-1}_{\lambda_1,\cdots,\lambda_k} \begin{pmatrix} 0^{k-1} \\ y \end{pmatrix} = \begin{pmatrix} \widetilde{\lambda}_0^{-1} \\ \widetilde{\lambda}_1^{-1} \\ \vdots \\ \widetilde{\lambda}_{k-1}^{-1} \end{pmatrix} y$$

in the canonical torus $\mathbf{R}^k/\mathbf{Z}^k$. But this map is ergodic when the family $(\widetilde{\lambda}_m^{-1})_{0\le m\le k}$ is $\mathbf{Q}$-free. □

From Proposition 3.8 it is possible to deduce an expression of the difference $c\delta(y)$ in terms of projections on a fundamental domain of X_k, as explained by the following proposition.

Proposition 3.9. *The cumulated difference $c\delta(y)$ only depends on the projection of $(0^{k-1}, y)$ on X_k. Moreover, the induced map $c\delta : X_k \to \mathbf{R}$ is affine when restricted to the fundamental domain*

$$\mathcal{D} = \prod_{i=1}^{k} [1/2 - \lambda_i, 1/2]$$

of X_k, and if $(x_1, \ldots, x_k) \in \mathcal{D}$ is the projection of $(0^{k-1}, y)$ on $\mathcal{D}$ modulo $\widetilde{\Lambda}_k$, we have

$$c\delta(y) = -\frac{1}{2} - \sum_{m=1}^{k} x_m \frac{\widetilde{\lambda}_m}{\widetilde{\lambda}_0}.$$

Remark 3.10. From this proposition one can explain the appearance of normalisation constant in the definition of $c\delta$. Indeed, the mean of the affine map $c\delta$ on $\mathcal{D}$ is equal to its value in the centre of the parallelepiped $\mathcal{D}$:

$$\begin{aligned}\frac{1}{\widetilde{\lambda}_0}\int_{X_k} c\delta(x_1,\dots,x_k)\,\mathrm{d}x_1\cdots\mathrm{d}x_k &= c\delta\left(\frac{1}{2}-\frac{\lambda_1}{2}\,,\ \cdots\ ,\ \frac{1}{2}-\frac{\lambda_k}{2}\right)\\ &= -\frac{1}{2}-\sum_{m=1}^{k}\widetilde{\lambda}_m\frac{1-\lambda_m}{2\widetilde{\lambda}_0}\\ &= -\frac{1}{2\widetilde{\lambda}_0}.\end{aligned}$$

In particular, for $R\in\mathbf{N}$, one has

$$\begin{aligned}\frac{1}{R}\int_0^R c\delta(y)\,\mathrm{d}y &= \frac{1}{R}\sum_{n=0}^{R-1} c\delta(n+1/2)\\ &= \frac{1}{R}\sum_{n=0}^{R-1}\left(c\delta(n)+\frac{1}{2\widetilde{\lambda}_0}\right)\\ &\underset{R\to+\infty}{\longrightarrow} 0,\end{aligned}$$

in other words the map $c\delta$ has zero mean.

Proof of Proposition 3.9. Given $n\in\mathbf{N}$, we want to compute the cumulated difference (3.3):

$$c\delta(n)=\frac{n}{\widetilde{\lambda}_0}-\mathrm{Card}\left\{x\in\mathbf{N}\mid\widehat{\boldsymbol{\ell}}^k(x)\le n\right\}+\frac{1}{2}$$

In other words (as $x\mapsto\widehat{\boldsymbol{\ell}}^k(x)$ is increasing) we search for the biggest $x\in\mathbf{N}$ such that $\widehat{\boldsymbol{\ell}}^k(x)\le n$. Let x be such a number, we have

$$c\delta(n)=\frac{n}{\widetilde{\lambda}_0}-x-\frac{1}{2} \tag{3.10}$$

(remark that this formula allows us to read this discrepancy on the "time 0" set $\mathbf{Z}$ – see Figure 2 –, this is possible by the preservation of order of the maps ℓ_i and $\widehat{\ell}_i$). We are reduced to compute this integer x.

We denote $y=\widehat{\boldsymbol{\ell}}^k(x)$ and $j=n-y\in\mathbf{N}$. In this case, the definition of global error leads to

$$\mathcal{E}_x^k=\widehat{\boldsymbol{\ell}}^k(x)-\boldsymbol{\ell}^k(x)=y-\widetilde{\lambda}_0 x\qquad\Longleftrightarrow\qquad x=\frac{1}{\widetilde{\lambda}_0}\big(y-\mathcal{E}_x^k\big).$$

Thus, applying this to (3.10),

$$c\delta(n)=\frac{n}{\widetilde{\lambda}_0}-\frac{y-\mathcal{E}_x^k}{\widetilde{\lambda}_0}-\frac{1}{2}=\frac{j+\mathcal{E}_x^k}{\widetilde{\lambda}_0}-\frac{1}{2}.$$

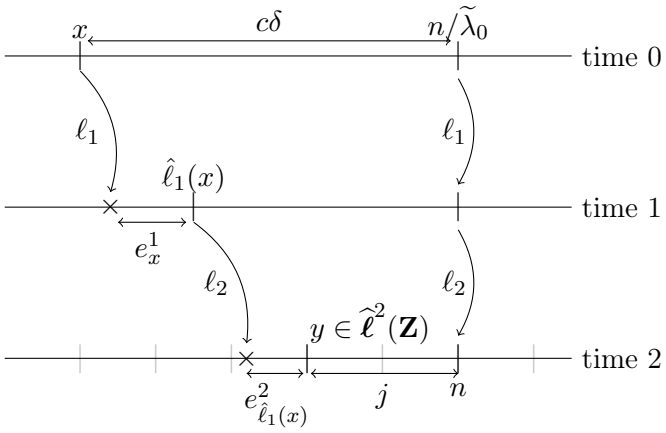

FIGURE 2. Computation of the cumulated difference $c\delta$ in the case $k = 2$. The point y is the biggest point of $\boldsymbol{\ell}^2(\mathbf{Z})$ smaller than n.

But, combining Formula (3.4) page 10 linking the global error $\mathcal{E}_x$ with the roundoff error vector $\boldsymbol{\varepsilon}_x^k$ with the fact that $\boldsymbol{\varepsilon}_x^k = -w$ (Proposition 3.8), one has

$$\mathcal{E}_x^k = -\sum_{m=1}^{k} \widetilde{\lambda}_m w_m,$$

so that

$$c\delta(n) = \frac{1}{\widetilde{\lambda}_0}\Big(j - \sum_{m=1}^{k} \widetilde{\lambda}_m w_m\Big) - \frac{1}{2}$$

with w depending only on the projection of $(0^{k-1}, y) = (0^{k-1}, n-j)$ on X_k.

We have reduced to find, given $n \in \mathbf{Z}$, the smallest $j \in \mathbf{N}$ such that $n - j \in \widehat{\boldsymbol{\ell}}^k(\mathbf{Z})$. First remark that W^k projects injectively (but not surjectively) on the torus $X_k = \mathbf{R}^k/\widetilde{\Lambda}_k$. So we define a partition of X_k into first visit sets W_j in W^k under the iterates of the translation $(0^{k-1}, 1)$ (see Figure 3). More precisely, $v \in X_k$ belongs to W_j iff j is the smallest nonegative integer such that $v + (0^{k-1}, j) \in W^k \bmod \widetilde{\Lambda}_k$.

Thus, if $(0^{k-1}, -n) \in W_j$, then j is the smallest integer such that $(0^{k-1}, -(n-j)) \in W_0 = W^k$. In this case, j is the smallest nonnegative integer such that $n - j \in \widehat{\boldsymbol{\ell}}^k(\mathbf{Z})$, moreover

$$W_i = \mathrm{pr}_{X_k}\left(W^k - (0^{k-1}, i)\right) \setminus \bigcup_{j=0}^{i-1} W_j.$$

For $v \in X_k$, this allows to define

$$c\delta(v) = \frac{1}{\widetilde{\lambda}_0}\Big(j - \sum_{m=1}^{k} \widetilde{\lambda}_m w_m\Big) - \frac{1}{2},$$

where $w = v + (0^{k-1}, j) \in W^k$ (and hence $v \in W_j$).

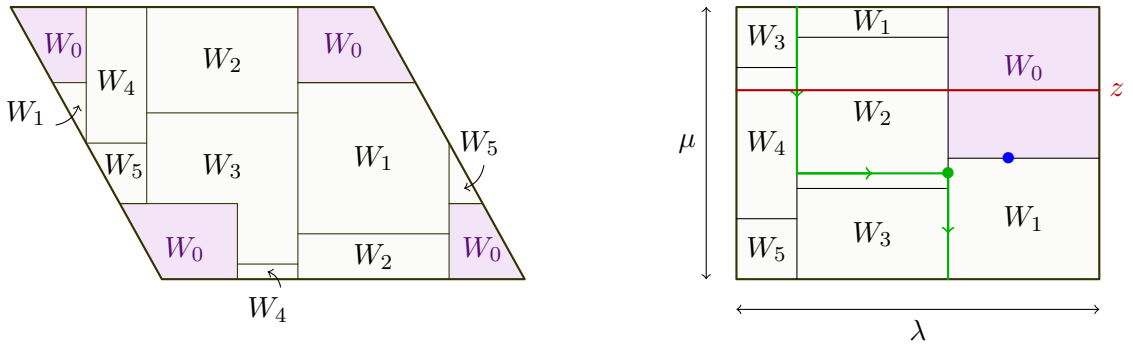

FIGURE 3. The sets W_i for $\lambda = 2.4$ and $\mu = 1.8$ on a canonical fundamental domain (left) and a rectangular fundamental domain $\mathcal{D}$ (right). We recognize a suspension of the rotation $x \mapsto x - 1$ modulo λ.

We just have proved that the map $c\delta$ is piecewise affine on X_k (more precisely, affine in restriction to each set W_i). Let us analyse the partition into the sets W_j a bit further by looking at its projection on the fundamental domain $\mathcal{D} = \prod_{i=1}^{k}[1/2, 1/2 - \lambda_i]$ (see Figure 3). In X_k, the set W_{j+1} is simply obtained by a translation of W_j by $(0, -1)$, removing the intersections with the W_i for $i \leq j$ if necessary. In $\mathcal{D}$, the gluings of opposite faces made to recover X_k correspond to suspensions of rotations $x \mapsto x + e_i$ where e_i is the i-th vector of the canonical basis of $\mathbf{R}^k$. In particular, one sees that all the W_i, except from the $\lceil \lambda_k \rceil$ last, are simply translates of W_0.

It remains to prove that the map $v \mapsto c\delta(v)$ is continuous on $\mathcal{D}$. First remark that it is continuous (because linear) on every set W_j, so we just have to prove the continuity at the boundaries of the W_i's. To do it we reason by recurrence on the dimension.

As a first step, we prove that $c\delta$ is continuous in restriction to the union of two cubes having a face in common orthogonal to the last canonical coordinate vector. In this case, they are two cubes with consecutive indices, say j and $j+1$. Let us take a point v belonging to their common face (the blue point of Figure 3, right). On the one hand, $v \in W_j$, so we can write $v = (w_1, \cdots, w_k)$, where the w_i's are the coordinates of v with respect to the centre of the set W_i. This leads to

$$c\delta(v) = \frac{1}{\widetilde{\lambda}_0}\Big(j - \sum_{m=1}^{k-1} \widetilde{\lambda}_m w_m - \widetilde{\lambda}_k w_k\Big) - \frac{1}{2}$$

But $w_k = -1/2$ and $\widetilde{\lambda}_k = 1$, so

$$c\delta(v) = \frac{1}{\widetilde{\lambda}_0}\Big(j - \sum_{m=1}^{k-1} \widetilde{\lambda}_m w_m + \frac{1}{2}\Big) - \frac{1}{2}.$$

On the other hand, we also have $v \in W_{j+1}$, and so if we write $v = (w'_1, \cdots, w'_k)$ in the coordinates with respect to the centre of W_{j+1},then

$$c\delta(v) = \frac{1}{\widetilde{\lambda}_0}\Big(j + 1 - \sum_{m=1}^{k-1} \widetilde{\lambda}_m w'_m - \widetilde{\lambda}_k w'_k\Big) - \frac{1}{2}.$$

But we have $w_i = w'_i$ for all $i \leq k-1$ and $w'_k = 1/2$, so

$$c\delta(v) = \frac{1}{\widetilde{\lambda}_0}\Big(j+1-\sum_{m=1}^{k-1}\widetilde{\lambda}_m w_m - \frac{1}{2}\Big) - \frac{1}{2}.$$

We deduce that both values of $c\delta(v)$ coincide whether v is seen a an element of W_j or of W_{j+1}.

For the induction's heredity, we use the trivial formula

$$\sum_{m=1}^{k-1}\widetilde{\lambda}_m w_m = \sum_{m=1}^{k-1}\widetilde{\lambda}_m w'_m,$$

where $w_m - w'_m$ is a vector having zero coordinates but the ℓ-th one equal to -1 and the $\ell+1$-th one equal to $\lambda_{\ell+1}$. For each ℓ, using the fact that $c\delta$ is continuous with respect to the $\ell+1$-th coordinate, this tells us that the map $c\delta$ is continuous with respect to the ℓ-th coordinate (geometrically, it consists in following the green path of Figure 3, right).

Remark that one can also prove the continuity by examinating directly what happens on the image spaces $\boldsymbol{\ell}^i(\mathbf{R})$ and performing small translations of the grids.

□

As a corollary, one can get an alternative proof of Proposition 3.5.

Second proof of Proposition 3.5. By Equation (3.6) page 10, one has

$$\mathrm{Disc}^2 = \frac{1}{12\widetilde{\lambda}_0^2} + \mathrm{Var}\left(c\delta\Big(n+\frac{1}{2}\Big)\right).$$

The second term corresponds to the variance of the map $c\delta + 1/(2\widetilde{\lambda}_0)$ on $\mathcal{D}$. As this map is affine with zero mean, and by the form of $\mathcal{D}$, this variance is equal to the sum of the variances of its coordinates, i.e.

$$\begin{aligned}\mathrm{Var}\left(c\delta\Big(n+\frac{1}{2}\Big)\right) &= \sum_{m=1}^{k}\mathrm{Var}_{[1/2-\lambda_m,1/2]}\left(x_m\frac{\widetilde{\lambda}_m}{\widetilde{\lambda}_0}\right)\\ &= \sum_{m=1}^{k}\frac{\lambda_m^2\widetilde{\lambda}_m^2}{12\widetilde{\lambda}_0^2},\end{aligned}$$

thus

$$\mathrm{Disc} = \frac{1}{12\widetilde{\lambda}_0^2}\sum_{m=0}^{k}\widetilde{\lambda}_m^2.$$

□

4. The tree linear case

In this section we adapt the study made in the previous one to stick to the shape of the set of preimages of a point under some expanding map, which has a structure of d-ary tree. We will get get discrepancy estimations for a complete d-ary tree with edges decorated by linear expanding maps.

Fix $r \geq 1$ and $d \geq 2$. We begin by the definition of the set of expanding maps.

Definition 4.1. We denote by $\mathcal{D}^r(\mathbf{S}^1)$ the set of C^r expanding maps of degree d of $\mathbf{S}^1$. More precisely, $\mathcal{D}^r(\mathbf{S}^1)$ is the set of degree d maps $f : \mathbf{S}^1 \to \mathbf{S}^1$ such that the derivative $f^{(\lfloor r \rfloor)}$ is well defined and belongs to $C^{r-\lfloor r \rfloor}(\mathbf{T}^n)$ and such that for every $x \in \mathbf{S}^1$, we have $|f'(x)| > 1$.

The set of preimages of a point $x \in \mathbf{S}^1$ by an expanding map f has a natural structure of complete d-ary tree. We now define the linear setting corresponding to the local behaviour of $f \in \mathcal{D}^r(\mathbf{S}^1)$ using this representation.

Definition 4.2. We set (see also Figure 4)

$$I_k = \bigsqcup_{m=1}^{k} [\![1, d]\!]^m$$

the set of m-tuples of integers of $[\![1, d]\!]$, for $1 \leq m \leq k$.

For $\boldsymbol{i} = (i_1, \cdots, i_m) \in [\![1, d]\!]^m$, we set $\text{length}(\boldsymbol{i}) = m$ and the parent $\wp(\boldsymbol{i}) = (i_1, \cdots, i_{m-1}) \in [\![1, d]\!]^{m-1}$ (with the convention $\wp(i_1) = \emptyset$).

The set I_k is the linear counterpart of the set $\bigsqcup_{m=1}^{k} f^{-m}(\{y\})$. Its cardinal is equal to $d(1 - d^k)/(1 - d)$.

Definition 4.3. Let $k \in \mathbf{N}$. The *complete tree of order* k is the rooted d-ary tree T_k whose vertices are the elements of I_k together with the root $\emptyset$, and whose edges are of the form $(\wp(\boldsymbol{i}), \boldsymbol{i})_{\boldsymbol{i} \in I_k}$ (see Figure 4).

We now consider a family $(\ell_{\boldsymbol{i}})_{\boldsymbol{i} \in I_k}$ of homotheties of $\mathbf{R}$ of parameters $(\lambda_{\boldsymbol{i}})_{\boldsymbol{i} \in I_k} > 1$. In this new case we denote

$$\widehat{\boldsymbol{\ell}}^k\big(\mathbf{Z}^{d^k}\big) = \bigcup_{\boldsymbol{i} \in [\![1,d]\!]^k} \big(\widehat{\ell}_{\wp^{k-1}(\boldsymbol{i})} \circ \cdots \circ \widehat{\ell}_{\boldsymbol{i}}\big)(\mathbf{Z}),$$

$$\widetilde{\lambda}_{\boldsymbol{i}} = \lambda_{\boldsymbol{i}} \lambda_{\wp(\boldsymbol{i})} \cdots \lambda_{\wp^{\text{length}(\boldsymbol{i})-1}(\boldsymbol{i})}, \qquad \widetilde{\lambda}_{tot}^{-1} = \sum_{\boldsymbol{i} \in [\![1,d]\!]^k} \widetilde{\lambda}_{\boldsymbol{i}}^{-1},$$

and for $\boldsymbol{i} \in [\![1, d]\!]^k$ (see also (3.3), we omit the dependance on the measures),

$$c\delta_{\boldsymbol{i}}(y) = \frac{y}{\widetilde{\lambda}_{\boldsymbol{i}}} - \text{Card}\left\{x \in \mathbf{N} \mid \widehat{\ell}_{\wp^{k-1}(\boldsymbol{i})} \circ \cdots \circ \widehat{\ell}_{\boldsymbol{i}}(x) \leq y\right\} + \frac{1}{2} \tag{4.1}$$

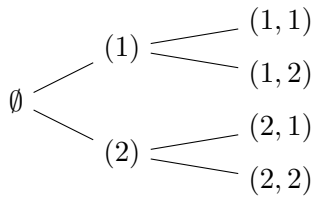

FIGURE 4. The tree T_2 for $d = 2$.

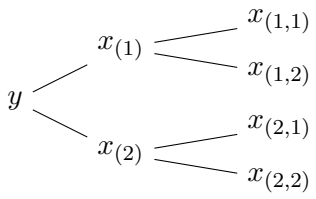

FIGURE 5. The tree associated to the preimages of y, for $k = 2$ and $d = 2$. We have $f(x_{(1,1)}) = f(x_{(1,2)}) = x_{(1)}$, etc.

and, denoting $\mathcal{E}^k_{x,\boldsymbol{i}}$ the cumulative error made by iterating the point x following the path from $\boldsymbol{i}$ during time k.

$$\begin{aligned}\mathrm{Disc}_R^2 =& \frac{1}{R}\int_0^R \left(\widetilde{\lambda}_{tot}^{-1} y - \sum_{\boldsymbol{i}\in[\![1,d]\!]^k} \left(\max\left\{x \mid \widetilde{\lambda}_{\boldsymbol{i}} x + \mathcal{E}^k_{x,\boldsymbol{i}} \le y\right\} - \frac{1}{2}\right)\right)^2 \mathrm{d}y \\ =& \frac{1}{R}\int_0^R \left(\sum_{\boldsymbol{i}\in[\![1,d]\!]^k} c\delta_{\boldsymbol{i}}(y)\right)^2 \mathrm{d}y.\end{aligned}$$

For $\boldsymbol{i},\boldsymbol{i}'\in I_k$ of same length, we denote

$$k_0(\boldsymbol{i},\boldsymbol{i}') = \min\left\{m\in[\![0,k]\!] \mid \wp^m(\boldsymbol{i}) = \wp^m(\boldsymbol{i}')\right\}.$$

We also denote $\delta_{\mathbf{Z}^{dk}} = \sum_{x\in\mathbf{Z}^{dk}} \delta_x$ the uniform measure on $\mathbf{Z}^{dk}$

Proposition 4.4. *Let $k\in\mathbf{N}$, and a family $(\ell_{\boldsymbol{i}})_{\boldsymbol{i}\in I_k}$ of homotheties of $\mathbf{R}$ of parameters $(\lambda_{\boldsymbol{i}})_{\boldsymbol{i}\in I_k}$ strictly bigger than 1. If the family $(\widetilde{\lambda}_{\boldsymbol{i}})_{\boldsymbol{i}\in I_k}$ is $\mathbf{Q}$-free (which is a generic condition), then*

$$\mathrm{Disc}^2\left(\widehat{\boldsymbol{\ell}}^k_*(\delta_{\mathbf{Z}^{dk}}),\ \widetilde{\lambda}_{tot}^{-1}\,\mathrm{Leb}\right) = \frac{1}{12\widetilde{\lambda}_{tot}} + \frac{1}{12}\sum_{\boldsymbol{i},\boldsymbol{i}'\in[\![1,d]\!]^k}\ \sum_{m=k_0(\boldsymbol{i},\boldsymbol{i}')}^{k-1} \frac{\widetilde{\lambda}_{\wp^m(\boldsymbol{i})}\widetilde{\lambda}_{\wp^m(\boldsymbol{i}')}}{\widetilde{\lambda}_{\boldsymbol{i}}\widetilde{\lambda}_{\boldsymbol{i}'}}.$$

Note that for any $m\ge k_0(\boldsymbol{i},\boldsymbol{i}')$, one has $\widetilde{\lambda}_{\wp^m(\boldsymbol{i})} = \widetilde{\lambda}_{\wp^m(\boldsymbol{i}')}$.

To prove this proposition we will need to estimate the expectation of the following correlations:

$$\mathrm{Corr}_{\boldsymbol{i},\boldsymbol{i}'}(y) = c\delta_{\boldsymbol{i}}\left(y+\frac{1}{2}\right)c\delta_{\boldsymbol{i}'}\left(y+\frac{1}{2}\right).$$

Lemma 4.5. *Under the hypotheses of Proposition 4.4, for any $\boldsymbol{i}\neq\boldsymbol{i}'$, $\mathbb{E}\big[\,\mathrm{Corr}_{\boldsymbol{i},\boldsymbol{i}'}\,\big]$ is well defined (that is, the limit exists) and satisfies*

$$\mathbb{E}\big[\,\mathrm{Corr}_{\boldsymbol{i},\boldsymbol{i}'}(y)\big] = \frac{1}{12\widetilde{\lambda}_{\boldsymbol{i}}\widetilde{\lambda}_{\boldsymbol{i}'}}\sum_{m=k_0(\boldsymbol{i},\boldsymbol{i}')}^{k-1}\widetilde{\lambda}_{\wp^m(\boldsymbol{i})}\widetilde{\lambda}_{\wp^m(\boldsymbol{i}')} \ge 0.$$

This lemma is the heart of the proof of Proposition 4.4. It is deduced from the study conducted in the previous section (Proposition 3.9). We first deduce Proposition 4.4 from it.

Proof of Proposition 4.4. By Lemma 3.4, for any $R\in\mathbf{N}$,

$$\mathrm{Disc}_R^2 = \frac{1}{R}\sum_{y=0}^{R-1}\left(\sum_{\boldsymbol{i}\in[\![1,d]\!]^k} c\delta_{\boldsymbol{i}}\left(y+\frac{1}{2}\right)\right)^2 + \frac{1}{12\widetilde{\lambda}_{tot}}.$$

But

$$\left(\sum_{\boldsymbol{i}\in[\![1,d]\!]^k} c\delta_{\boldsymbol{i}}\left(y+\frac{1}{2}\right)\right)^2 = \sum_{\boldsymbol{i}\in[\![1,d]\!]^k} c\delta_{\boldsymbol{i}}\left(y+\frac{1}{2}\right)^2 + \sum_{\boldsymbol{i}\neq\boldsymbol{i}'\in[\![1,d]\!]^k} c\delta_{\boldsymbol{i}}\left(y+\frac{1}{2}\right)c\delta_{\boldsymbol{i}'}\left(y+\frac{1}{2}\right).$$

By Corollary 3.6, the first term has mean

$$\frac{1}{12}\sum_{\boldsymbol{i}\in[\![1,d]\!]^k}\frac{1}{\widetilde{\lambda}_{\boldsymbol{i}}^2}\sum_{m=0}^{k-1}\widetilde{\lambda}_{\wp^m(\boldsymbol{i})}^2,$$

while that of the second one is given by Lemma 4.5. □

Proof of Lemma 4.5. Let $\boldsymbol{i},\boldsymbol{i}'\in[\![1,d]\!]^k$ such that $\boldsymbol{i}\neq\boldsymbol{i}'$. We write $\boldsymbol{i}=(i_k,\dots,i_1)$ and $\boldsymbol{i}'=(i'_k,\dots,i'_1)$ (note the fact that the indices decrease, to correspond to the order of iterations of maps as in the previous part). We also note $\widetilde{\Lambda}_{\boldsymbol{i}}$ the lattice associated to the multipliers $\lambda_{i_1},\dots,\lambda_{i_k}$, see (3.8) page 15 (and the same for $\widetilde{\Lambda}_{\boldsymbol{i}'}$).

By Proposition 3.9, we know that for $y\in\mathbf{Z}$, the cumulated difference $c\delta_{\boldsymbol{i}}(y)$ only depends on the projection of $(0^{k-1},y)$ on the torus $\mathbf{R}^k/\widetilde{\Lambda}_{\boldsymbol{i}}$ and is affine when restricted to the fundamental domain $\mathcal{D}=\prod_{j=1}^k[1/2,1/2-\lambda_{i_j}]$ of $\widetilde{\Lambda}_{\boldsymbol{i}}$: if $(0^{k-1},y)=(x_1,\dots,x_k)\in\mathcal{D}$ mod $\widetilde{\Lambda}_{\boldsymbol{i}}$, we have

$$c\delta_{\boldsymbol{i}}(y+1/2)=-\frac{1}{2}-\sum_{m=1}^{k}x_m\frac{\widetilde{\lambda}_{\wp^{m-1}(\boldsymbol{i})}}{\widetilde{\lambda}_{\boldsymbol{i}}}+\frac{1}{2\widetilde{\lambda}_{\boldsymbol{i}}}.$$

Thus, for any $y\in\mathbf{Z}$ (with transparent notations),

$$\mathrm{Corr}_{\boldsymbol{i},\boldsymbol{i}'}(y)=\widetilde{\lambda}_{\boldsymbol{i}}^{-1}\widetilde{\lambda}_{\boldsymbol{i}'}^{-1}\left(\frac{\widetilde{\lambda}_{\boldsymbol{i}}-1}{2}+\sum_{m=1}^{k}x_m\widetilde{\lambda}_{\wp^{m-1}(\boldsymbol{i})}\right)\left(\frac{\widetilde{\lambda}_{\boldsymbol{i}'}-1}{2}+\sum_{m=1}^{k}x'_m\widetilde{\lambda}_{\wp^{m-1}(\boldsymbol{i}')}\right).$$

By definition of $k_0\overset{\text{def.}}{=}k_0(\boldsymbol{i},\boldsymbol{i}')$, for every $m\in[\![k_0+1,k]\!]$, we have $\widetilde{\lambda}_{\wp^{m-1}(\boldsymbol{i})}=\widetilde{\lambda}_{\wp^{m-1}(\boldsymbol{i}')}$ and $x_m=x'_m$; thus we can denote

$$t_1=\sum_{m=k_0+1}^{k}x_m\widetilde{\lambda}_{\wp^{m-1}(\boldsymbol{i})}=\sum_{m=k_0+1}^{k}x'_m\widetilde{\lambda}_{\wp^{m-1}(\boldsymbol{i}')},$$

$$t_2=\frac{\widetilde{\lambda}_{\boldsymbol{i}}-1}{2}+\sum_{m=1}^{k_0}x_m\widetilde{\lambda}_{\wp^{m-1}(\boldsymbol{i})}\qquad\text{and}\qquad t'_2=\frac{\widetilde{\lambda}_{\boldsymbol{i}'}-1}{2}+\sum_{m=1}^{k_0}x'_m\widetilde{\lambda}_{\wp^{m-1}(\boldsymbol{i}')},$$

so that

$$\widetilde{\lambda}_{\boldsymbol{i}}\widetilde{\lambda}_{\boldsymbol{i}'}\,\mathrm{Corr}_{\boldsymbol{i},\boldsymbol{i}'}(y)=\big(t_1+t_2\big)\big(t_1+t'_2\big).$$

Thus

$$\begin{aligned}\widetilde{\lambda}_{\boldsymbol{i}}\widetilde{\lambda}_{\boldsymbol{i}'}\mathbb{E}\big[\,\mathrm{Corr}_{\boldsymbol{i},\boldsymbol{i}'}(y)\big]\;&=\mathbb{E}[t_1^2]+\mathbb{E}[t_1t'_2]+\mathbb{E}[t_1t_2]+\mathbb{E}[t_2t'_2]\\&=\mathbb{E}[(t_1-\mathbb{E}[t_1])^2]+\mathbb{E}[t_1]^2+\mathbb{E}[t_1t'_2]+\mathbb{E}[t_1t_2]+\mathbb{E}[t_2t'_2].\end{aligned}$$

The hypothesis of independence over $\mathbf{Q}$ of the family $(\widetilde{\lambda}_{\boldsymbol{i}})_{\boldsymbol{i}\in I_k}$ implies that the events t_1, t_2 and t'_2 are "independent". To see it, consider the (big) matrix

$$M_{k+k_0}(\mathbf{R})\ni\widetilde{M}_{\boldsymbol{i},\boldsymbol{i}'}=$$

$$\left(\begin{array}{ccccc|ccc|ccc}
\lambda_{\boldsymbol{i}} & -1 & & & & & & & & & \\
 & \ddots & \ddots & & & & & & & & \\
 & & \lambda_{\wp^{k_0-2}(\boldsymbol{i})} & -1 & & & & & & & \\
 & & & \lambda_{\wp^{k_0-1}(\boldsymbol{i})} & & 0 & & & & -1 & \\
\hline
 & & & & & \lambda_{\boldsymbol{i}'} & -1 & & & & \\
 & & & & & & \ddots & \ddots & & & \\
 & & & & & & & \lambda_{\wp^{k_0-2}(\boldsymbol{i}')} & -1 & & \\
 & & & & & & & & \lambda_{\wp^{k_0-1}(\boldsymbol{i}')} & -1 & \\
\hline
 & & & & & & & & & \lambda_{\wp^{k_0}(\boldsymbol{i})} & \ddots \\
 & & & & & & & & & & \ddots \quad -1 \\
 & & & & & & & & & & \lambda_{\wp^{k-1}(\boldsymbol{i})}
\end{array}\right),$$

remark that the quotient space $\mathbf{R}^{k+k_0}/\widetilde{M}_{\boldsymbol{i},\boldsymbol{i}'}\mathbf{Z}^{k+k_0}$ is obtained from the product space $\mathbf{R}^k/\widetilde{M}_{\lambda_{i_1},\cdots,\lambda_{i_k}}\mathbf{Z}^k \times \mathbf{R}^k/\widetilde{M}_{\lambda_{i'_1},\cdots,\lambda_{i'_k}}\mathbf{Z}^k$ by quotienting by the $k-k_0$ last coordinates, which corresponds to identical multipliers $\lambda_{\wp^j(\boldsymbol{i})} = \lambda_{\wp^j(\boldsymbol{i}')}$, thus to identical projections of $(0^{k-1}, y)$ on the fundamental domains $\mathcal{D}$ and $\mathcal{D}'$. A simple computation of $\widetilde{M}^{-1}_{\boldsymbol{i},\boldsymbol{i}'}$, similar to that of the proof of Proposition 3.8, leads to the fact that when the family $(\widetilde{\lambda}_{\boldsymbol{i}})_{\boldsymbol{i}\in I_k}$ is $\mathbf{Q}$-free, then

$$\begin{aligned} y \mapsto \widetilde{M}^{-1}_{\boldsymbol{i},\boldsymbol{i}'}(0,y) = y\Big(&\widetilde{\lambda}^{-1}_{\boldsymbol{i}},\ \widetilde{\lambda}^{-1}_{\wp(\boldsymbol{i})},\ \cdots,\ \widetilde{\lambda}^{-1}_{\wp^{k_0-1}(\boldsymbol{i})},\\ &\widetilde{\lambda}^{-1}_{\boldsymbol{i}'},\ \widetilde{\lambda}^{-1}_{\wp(\boldsymbol{i}')},\ \cdots,\ \widetilde{\lambda}^{-1}_{\wp^{k_0-1}(\boldsymbol{i}')},\\ &\widetilde{\lambda}^{-1}_{\wp^{k_0}(\boldsymbol{i})},\ \widetilde{\lambda}^{-1}_{\wp^{k_0+1}(\boldsymbol{i})},\ \cdots,\ \widetilde{\lambda}^{-1}_{\wp^{k-1}(\boldsymbol{i})}\Big) \end{aligned}$$

is an ergodic $\mathbf{Z}$-action on the torus $\mathbf{R}^{k+k_0}/\mathbf{Z}^{k+k_0}$. In particular,

$$\mathbb{E}[t_1t_2] = \mathbb{E}[t_1]\mathbb{E}[t_2], \quad \mathbb{E}[t_1t'_2] = \mathbb{E}[t_1]\mathbb{E}[t'_2] \quad \text{and} \quad \mathbb{E}[t_2t'_2] = \mathbb{E}[t_2]\mathbb{E}[t'_2].$$

We deduce that

$$\begin{aligned} \widetilde{\lambda}_{\boldsymbol{i}}\widetilde{\lambda}_{\boldsymbol{i}'}\mathbb{E}\big[\operatorname{Corr}_{\boldsymbol{i},\boldsymbol{i}'}(y)\big] =& \mathbb{E}\big[(t_1-\mathbb{E}[t_1])^2\big] + \mathbb{E}[t_1]^2 + \mathbb{E}[t_1]\mathbb{E}[t'_2] + \mathbb{E}[t_1]\mathbb{E}[t_2] + \mathbb{E}[t_2]\mathbb{E}[t'_2]\\ =& \mathbb{E}\big[(t_1-\mathbb{E}[t_1])^2\big] + \mathbb{E}\big[t_1+t_2\big]\mathbb{E}\big[t_1+t'_2\big]. \end{aligned}$$

But by Remark 3.10, $\mathbb{E}[t_1+t_2]\mathbb{E}[t_1+t'_2] = 0$, so

$$\widetilde{\lambda}_{\boldsymbol{i}}\widetilde{\lambda}_{\boldsymbol{i}'}\mathbb{E}\big[\operatorname{Corr}_{\boldsymbol{i},\boldsymbol{i}'}(y)\big] = \operatorname{Var}(t_1).$$

The computation of the variance of t_1 is obtained by a computation similar to that in second proof of Proposition 3.5 (page 21); we get

$$\begin{aligned} \operatorname{Var}(t_1) =& \sum_{m=k_0+1}^{k} \operatorname{Var}_{x_m\in[1/2-\lambda_{\wp^{m-1}(\boldsymbol{i})},1/2]}\left(x_m\widetilde{\lambda}_{\wp^m(\boldsymbol{i})}\right)\\ =& \frac{1}{12}\sum_{m=k_0+1}^{k}\widetilde{\lambda}^2_{\wp^m(\boldsymbol{i})}\lambda^2_{\wp^{m-1}(\boldsymbol{i})} = \frac{1}{12}\sum_{m=k_0}^{k-1}\widetilde{\lambda}^2_{\wp^m(\boldsymbol{i})}. \end{aligned}$$

Finally,

$$\mathbb{E}\left[\mathrm{Corr}_{\boldsymbol{i},\boldsymbol{i}'}(y)\right] = \frac{1}{12}\sum_{m=k_0}^{k-1}\frac{\widetilde{\lambda}^2_{\wp^m(\boldsymbol{i})}}{\widetilde{\lambda}_{\boldsymbol{i}}\widetilde{\lambda}_{\boldsymbol{i}'}}.$$

□

Remark 4.6. If in the formula we replace all the $\lambda_{\boldsymbol{i}}$'s by d, we get

$$\mathrm{Disc}^2 = \frac{d^{k+1}-1}{12(d-1)}.$$

Without the correlations, it gives

$$\widetilde{\mathrm{Disc}}^{\,2} = \frac{d^{k+1}-d^{-k-1}}{12(d-d^{-1})}.$$

the ratio between both tends to $1+1/d$ when k goes to $+\infty$.

We end this section by a quantitative version of Proposition 4.4. For $E \subset \mathbf{Z}$, we will denote

$$D_R^+(E) = \sup_{x\in\mathbf{R}}\frac{\mathrm{Card}\left(E\cap[x-R,x+R]\right)}{\mathrm{Card}\left(\mathbf{Z}\cap[x-R,x+R]\right)}.$$

Addendum 4.7. *For every $\ell', c \in \mathbf{N}$, there exists a locally finite union of positive codimension submanifolds V_q of $]1,+\infty[^{\mathrm{Card}\, I_k}$ such that for every $\eta'>0$, there exists a radius $R_0>0$ such that if $(\widetilde{\lambda}_{\boldsymbol{i}})_{\boldsymbol{i}\in I_k}$ satisfies $d\big((\widetilde{\lambda}_{\boldsymbol{i}})_{\boldsymbol{i}\in I_k}, V_q\big) > \eta'$ for every q, then for every $R \geq R_0$, and every family $\mathbf{v} = (v_{\boldsymbol{i}})_{\boldsymbol{i}\in I_k}$ of real numbers, we have*[12]

(i) there is a subset of $\mathbf{Z}$ with bounded gaps, made of points which are images of exactly d^k points, each of them having all its roundoff errors close to 0: for every $y\in\mathbf{Z}$, there exists $y'\in\mathbf{Z}$ with $|y-y'|\leq R_0$ and for every $\boldsymbol{i}\in[\![1,d]\!]^k$, a point $x_{\boldsymbol{i}}$ such that

$$\widehat{\boldsymbol{\ell}+\mathbf{v}}^{\,k}\left((x_{\boldsymbol{i}})_{\boldsymbol{i}\in[\![1,d]\!]^k}\right) = \{y'\}$$

and for every $\boldsymbol{i}\in[\![1,d]\!]^k$ and every $m\leq k$, we have

$$|e^m_{x,\boldsymbol{i}}| \leq \frac{1}{\ell'},$$

where $e^m_{x,\boldsymbol{i}}$ is the the error made at the m-th iteration of the point x, applying the discretizations of the maps $\ell_{\wp^j(\boldsymbol{i})}+v_{\wp^j(\boldsymbol{i})}$.

(ii) for any y' like in item (i), the mean of the cumulated difference starting from y' is almost zero (for a set $A\subset\mathbf{R}$, we denote $A-y'$ the translation of A of $-y'$):

$$\left|\frac{1}{R}\int_0^R c\delta_x\left(\widehat{\boldsymbol{\ell}+\mathbf{v}}^{\,k}\left(\mathbf{Z}^{d^k}\right)-y'\,,\,\widetilde{\lambda}_{tot}^{-1}\,\mathrm{Leb}\right)\mathrm{d}x\right| \leq \frac{1}{\ell'}.$$

(iii) for any y' like in item (i), the discrepancy starting from y' is almost the same as the one starting from 0:

$$\left|\mathrm{Disc}^2_R\left(\widehat{\boldsymbol{\ell}+\mathbf{v}}^{\,k}\left(\mathbf{Z}^{d^k}\right)-y'\,,\,\widetilde{\lambda}_{tot}^{-1}\,\mathrm{Leb}\right) - \frac{1}{12\widetilde{\lambda}_{tot}} - \frac{1}{12}\sum_{\boldsymbol{i},\boldsymbol{i}'\in[\![1,d]\!]^k}\sum_{m=k_0(\boldsymbol{i},\boldsymbol{i}')}^{k-1}\frac{\widetilde{\lambda}_{\wp^m(\boldsymbol{i})}\widetilde{\lambda}_{\wp^m(\boldsymbol{i}')}}{\widetilde{\lambda}_{\boldsymbol{i}}\widetilde{\lambda}_{\boldsymbol{i}'}}\right| < \frac{1}{\ell'};$$

[12]The map $\widehat{\ell+v}$ is the discretization of the affine map $\ell+v$.

(iv) there is only a small proportion of the points of the image sets which are obtained by discretizing points close to $\mathbf{Z}+1/2$*: for every* $m \leq k$ *and every* $\boldsymbol{i} \in [\![1,d]\!]^k$*, we have*

$$D_R^+\left\{x \in \big(\ell_{\wp^m(\boldsymbol{i})} + v_{\wp^m(\boldsymbol{i})}\big)(\mathbf{Z}) \;\middle|\; d\big(x, \mathbf{Z}+\frac{1}{2}\big) < \frac{1}{3c\ell'}\right\} < \frac{1}{c\ell'};$$

Sketch of proof of Addendum 4.7. The moral of this addendum is that if a collection of numbers $x_1, \dots, x_k$ is "almost $\mathbf{Q}$-independent", then the rotation of vector $x_1, \dots, x_k$ in $\mathbf{T}^k$ is "ergodic up to ε".

In our particular case, if the collection $(\widetilde{\lambda}_{\boldsymbol{i}})_{\boldsymbol{i}\in I_k}$ does not satisfy any linear dependence relation with small integer coefficients, then for any $\boldsymbol{i}, \boldsymbol{i}'$, the image of the action of $\mathbf{Z}$ by $y \mapsto (0^{k+k_0-1}, y)$ on $\mathbf{R}^{k+k_0}/\widetilde{M}_{\boldsymbol{i},\boldsymbol{i}'}\mathbf{Z}^{k+k_0}$ is "uniformly distributed up to ε". This implies that the events t_1, t_2 and t_2' are "almost independent" and that the variance of t_1 is almost equal to the formula of the proof of Proposition 4.4.

These arguments are formalized by the following improvement of Weyl's criterion:

Lemma 4.8 (Weyl)**.** *Let* dist *be a distance generating the weak-* topology on* $\mathcal{P}$ *the space of Borel probability measures on* $\mathbf{T}^n$*. Then, for every* $\varepsilon > 0$*, there exists a locally finite family of affine hyperplanes* $H_i \subset \mathbf{R}^n$*, such that for every* $\eta > 0$*, there exists* $M_0 \in \mathbf{N}$*, such that for every* $\boldsymbol{\lambda} \in \mathbf{R}^n$ *satisfying* $d(\boldsymbol{\lambda}, H_q) > \eta$ *for every* q*, and for every* $M \geq M_0$*, we have*

$$\operatorname{dist}\left(\frac{1}{M}\sum_{m=0}^{M-1} \bar{\delta}_{m\boldsymbol{\lambda}}\,,\ \operatorname{Leb}_{\mathbf{R}^n/\mathbf{Z}^n}\right) < \varepsilon,$$

where $\bar{\delta}_x$ *is the Dirac measure of the projection of* x *on* $\mathbf{R}^n/\mathbf{Z}^n$*.*

For a proof of this lemma, see [Gui19]. □

5. A formula for the discrepancy of C^r-generic expanding maps

In this section we prove Theorem A, by starting with the more explicit statement Theorem 5.1.

5.1. **First formula.** As a first step towards the proof of Theorem A, one gets a first formula for the discrepancy. Recall that for any $1 \leq r \leq +\infty$, we denote $\mathcal{D}^r(\mathbf{S}^1)$ the set of C^r expanding maps, and that we denote $d \geq 2$ the degree of $f \in \mathcal{D}^r(\mathbf{S}^1)$.

Theorem 5.1. *For any* $1 \leq r \leq +\infty$*, if* f *is a generic element of* $\mathcal{D}^r(\mathbf{S}^1)$*, then for any* $k \in \mathbf{N}$*, one has*

$$\lim_{N\to+\infty} N \operatorname{Disc}\big(f_N^k(E_N), L_f^k(\operatorname{Leb})\big) = \left(\frac{1}{12} + \frac{1}{12}\int_{\mathbf{S}^1} \sum_{x,x'\in f^{-k}(y)} \sum_{m=k_0(x,x')}^{k-1} \frac{1}{(f^m)'(x)(f^m)'(x')}\,\mathrm{d}y\right)^{1/2}. \tag{5.1}$$

Before coming to the proof, let us first make a few comments. First, note that when $f(x) = 2x$, the right part of (5.1) becomes $2^{(k-3)/2}$. It gives an explicit approximation of the asymptotics for generic maps very close to $x \mapsto 2x$.

For $y \in \mathbf{S}^1$, let H_y be the difference between the cumulative distribution functions of respectively:

- The uniform measure on $f_N^k(E_N)$, and
- $L_f^k(\mathrm{Leb})$,

seen as measures on the fundamental domain $[y, y+1]$ of $\mathbf{S}^1$. By the definition of the discrepancy (Equation (2.1)),

$$\mathrm{Disc}\left(f_N^k(E_N), L_f^k(\mathrm{Leb})\right)^2 = \int_0^1 \left(H_y(x) - \left(\int_0^1 H_y\right)\right)^2 \mathrm{d}x \tag{5.2}$$

Lemma 5.2. *There exists a constant $B = B(f,k) > 0$ such that for any y, one has $\|H_y\|_\infty \leq B/N$.*

Proof of Lemma 5.2. The global roundoff error of each point $x \in E_N$ satisfies

$$\left|f_N^k(x) - f^k(x)\right| \leq \frac{A}{N} \qquad \text{where} \qquad A = \frac{\|f'\|_\infty^k}{2(\|f'\|_\infty - 1)}.$$

Thus, for any y, one has $\|H_y\|_\infty \leq 2\frac{A}{N} \cdot \frac{d}{\min_{\mathbf{S}^1} f'}$. □

Hence, the good scale for the discrepancy $\mathrm{Disc}\left(f_N^k(E_N), L_f^k(\mathrm{Leb})\right)$ is at most $1/N$. Theorem 5.1 ensures that this is exactly $1/N$, as it shows that $N\,\mathrm{Disc}\left(f_N^k(E_N), L_f^k(\mathrm{Leb})\right)$ converges towards a positive number when N goes to infinity.

The proof of Theorem 5.1 is mainly based on Proposition 4.4 (more precisely, Addendum 4.7), which treats the linear corresponding case. By applying arguments of [Gui19], one gets the following property:

Proposition 5.3. *Let $r \geq 1$ and f a generic element of $\mathcal{D}^r(\mathbf{T}^n)$. Then for any $k \in \mathbf{N}$, any $\varepsilon > 0$, and any $N \in \mathbf{N}$ large enough, there exists a finite collection $I_p \subset \mathbf{S}^1$ of pairwise disjoint segments such that:*

1) *each segment I_p has length smaller than ε, and the union of the segments I_p has Lebesgue measure bigger than $1-\varepsilon$;*
2) *the left endpoint y_p of each segment I_p is an element of E_N;*
3) *each point y_p is the image of d^k points $x_{\boldsymbol{i},p} \in E_N$ (the maximal possible number) by f_N^k, and the roundoff error vector in time k of each point $x_{\boldsymbol{i},p}$ is ε-small;*
4) *for each p, the discrepancy distribution restricted to the segment I_p and starting from the point y_p is ε-close to the discrepancy distribution associated to the preimage tree of y_p.*

In particular, point (4) ensures that on each segment $[y_p, y_p + R/N] \subset I_p$ such that $R \geq R_0$, the mean of the map $c\delta$ is ε-close to 0, and its variance (which corresponds to the L^2 discrepancy Disc associated to the map f) is ε-close to the discrepancy Disc associated to the preimage tree of y_p.

Proof of Proposition 5.3. We simply apply the arguments of the proof of Theorem 33 of [Gui19], by replacing Lemma 34 of [Gui19] by Addendum 4.7. In particular, this proof tells us that Thom's transversality theorem implies the existence of a family $[y'_p, z_p]$ of segments of length $\gg R_0 N$ (where R_0 is given by Addendum 4.7), with $y'_p, z_p \in 1/N\mathbf{Z}$. For any p, we apply Addendum 4.7 to the point $y = y'_p$, the derivatives given by the preimage tree

starting from y'_p, and the preimage set of y_p as the vector $\mathbf{v}$. This gives us a point $y' = y_p$ (by item (i)), and allows to define the segments $I_p = [y_p, z_p]$.

Point (1) comes from the proof of Theorem 33 of [Gui19] and Points (2) and (3) come from point (i) of Addendum 4.7. For itself, point (4) comes from an application of Taylor formula, the linear formulation of items item (ii) and item (iii) of Addendum 4.7, and the error estimate for nonlinearities of item (iv) of Addendum 4.7 (see the proof of Theorem 33 of [Gui19] for more details). □

Proof of Theorem 5.1. In this proof we denote $H_p = H_{y_p}$, where the y_p are given by Proposition 5.3.

Points (2) and (3) of Proposition 5.3 ensure that for $p \neq p'$, the cumulative distribution functions H_p and $H_{p'}$ (seen as functions of $\mathbf{S}^1$) are close: reasoning as in Lemma 5.2, one gets

$$\|H_p - H_{p'}\|_\infty \leq \varepsilon \frac{B}{N}. \tag{5.3}$$

By Taylor formula, if ε is small enough, in restriction to the interval I_p, the measure $L_f^k(\mathrm{Leb})$ is close to $\widetilde{\lambda}_{tot}^{-1}\,\mathrm{Leb}$, where $\widetilde{\lambda}_{tot}$ denotes the multiplier associated to the preimage tree at y_p. Combined with point (4), this fact implies that the mean of H_p restricted to I_p is small:

$$\left| \frac{1}{|I_p|} \int_{I_p} H_p \right| \leq \frac{\varepsilon}{N}.$$

Hence, fixing $p = 0$, and using the fact that all the H_p's are close, one gets that the mean of NH_0 is small in restriction to the union of the I_p.

Then, using point (1) together with Lemma 5.2, we deduce that the mean of NH_0 is small: there is a constant $C = C(f, k) > 0$ such that

$$\left| \int_{\mathbf{S}^1} H_0 \right| \leq \varepsilon \frac{C}{N}.$$

This fact, together with Equation (5.2), implies that

$$\left| \mathrm{Disc}\left(f_N^k(E_N), L_f^k(\mathrm{Leb})\right)^2 - \int_{\mathbf{S}^1} H_0^2 \right| \leq \frac{\varepsilon}{N^2}.$$

Using Equation (5.3) and Lemma 5.2 again, we deduce that

$$\left| \mathrm{Disc}\left(f_N^k(E_N), L_f^k(\mathrm{Leb})\right)^2 - \sum_p \int_{I_p} H_p^2 \right| \leq \frac{\varepsilon}{N^2}.$$

We now use point (4), which ensures that for any p,

$$\left| N^2 \int_{I_p} H_p^2 - \left(\frac{1}{12} + \frac{1}{12} \int_{I_p} \sum_{x,x' \in f^{-k}(y)} \sum_{m=k_0(x,x')}^{k-1} \frac{1}{(f^m)'(x)(f^m)'(x')}\, \mathrm{d}y \right) \right| \leq \varepsilon |I_p|.$$

Combined with the previous estimation, this gives the theorem. □

5.2. **Proof of theorem A.**

Proof of Theorem A. Recall that in Theorem 5.1 we have, for any x_0 that is an f^r-preimage of y for some r:

$$\widetilde{\lambda}_{x_0} = f'(x_0)\cdot f'(f(x_0))\cdot\dots\cdot f'(f^{r-1}(x_0)) = \frac{\mathrm{d}}{\mathrm{d}x}(f^r(x))|_{x=x_0} = Df^r(x_0),$$

and for each $x, x' \in f^{-k}(y)$ we have $k_0(x,x')$ defined to be the smallest m such that $f^m(x) = f^m(x')$.

Therefore the integral of Theorem 5.1 can also be written as

$$\int_0^1 \sum_{x,x'\in f^{-k}(y)} \sum_{m=k_0(x,x')}^{k-1} \frac{1}{Df^m(x)\cdot Df^m(x')}\,\mathrm{d}y. \tag{5.4}$$

Putting $n+e=k$ (for each pair x, x', being the f-images $n$ot equal for n steps, then $e$qual for e steps), we can split this integral in sums of the form

$$\sum_{\substack{n+e=k\\ n,e\geq 0}} \int_0^1 \sum_{z\in f^{-e}(y)} \sum_{\substack{x,x'\in f^{-n}(z)\\ f^{n-1}(x)\neq f^{n-1}(x')}} \sum_{m=n}^{k-1} \frac{1}{Df^m(x)Df^m(x')}\,\mathrm{d}y.$$

When we fix a certain $0\leq m\leq k-1$, the integral is

$$\int_0^1 \sum_{\substack{n\leq m\\ e=k-n}} \sum_{z\in f^{-e}(y)} \sum_{\substack{x,x'\in f^{-n}(z)\\ f^{n-1}(x)\neq f^{n-1}(x')}} \frac{1}{Df^m(x)Df^m(x')}\,\mathrm{d}y$$

$$= \int_0^1 \sum_{w\in f^{-(k-m)}(y)} \sum_{x,x'\in f^{-m}(w)} \frac{1}{Df^m(x)Df^m(x')}\,\mathrm{d}y$$

as it could be also deduced directly from Equation (5.4) fixing m.

Since $y = f^{k-m}(w)$ and therefore $\mathrm{d}y = Df^{k-m}(w)\,\mathrm{d}w$, dividing the domain into the intervals where f^{k-m} is injective, and then putting again everything together, we can change variable and obtain

$$\int_0^1 \sum_{x,x'\in f^{-m}(w)} \frac{Df^{k-m}(w)}{Df^m(x)Df^m(x')}\,\mathrm{d}w. \tag{5.5}$$

Let us consider now the map $f\times f : [0,1]^2 \to [0,1]^2$, and let $L_{f\times f}$ be its transfer operator. Given an observable $H(w,w') : [0,1] \to \mathbf{R}$, the m-th power of $L_{f\times f}$ can be computed on H as

$$\big(L^m_{f\times f}H\big)(w,w') = \sum_{\substack{f^m(x)=w\\ f^m(x')=w'}} \frac{H(x,x')}{Df^m(x)Df^m(x')},$$

considering that the Jacobian determinant of $(f\times f)^m$ at (x,x') is $Df^m(x)Df^m(x')$. Notice also that if $H(x,x') = h_1(x)\cdot h_2(x')$ we have that

$$\big(L^m_{f\times f}H\big)(w,w') = (L^m_f h_1)(w)\cdot(L^m_f h_2)(w').$$

Therefore, the integral of (5.5) is also the integral on the diagonal

$$\Delta = \{(w, w) : w \in [0, 1]\}$$

of $L^m_{f\times f}H$ for any observable H such that $H(x, x') = Df^{k-m}(w)$ whenever $f^m(x) = w = f^m(x')$. In our case can take for instance

$$H(x, x') = \sqrt{Df^{k-m}(f^m x) \cdot Df^{k-m}(f^m x')},$$

and in the end the integral amounts to the integral $L^m_{f\times f}(H)$ along Δ.

Taking $h(w) = \sqrt{Df^{k-m}(f^m w)}$ we have $H(w, w') = h(w) \cdot h(w')$, therefore

$$\begin{aligned}\int_0^1 \sum_{x,x'\in f^{-m}(w)} \frac{Df^{k-m}(w)}{Df^m(x)Df^m(x')}\,\mathrm{d}w &= \int_0^1 (L^m_{f\times f}H)(w, w)\,\mathrm{d}w\\ &= \int_0^1 (L^m_f h)(w)\,(L^m_f h)(w)\,\mathrm{d}w\\ &= \int_0^1 (L^m_f h)(w)^2\,\mathrm{d}w.\end{aligned}$$

Let $g(x) = \sqrt{Df^{k-m}(x)}$, so that $h(x) = g(f^m(x))$ for short. Let us recall the formula for the transfer operator applied to $g(f^m(x))$:

$$\begin{aligned}(L^m_f h)(w) &= \sum_{x\in f^{-m}(w)} \frac{g(f^m(x))}{Df^m(x)}\\ &= \sum_{x\in f^{-m}(w)} \frac{g(w)}{Df^m(x)}\\ &= g(w) \sum_{x\in f^{-m}(w)} \frac{1}{Df^m(x)}\\ &= g(w)\,(L_f 1)(w).\end{aligned}$$

Therefore, our integral can be written as

$$\begin{aligned}\int_0^1 (L^m_f h)(w)^2\,\mathrm{d}w &= \int_0^1 \left[g(w)\,(L^m_f 1)(w)\right]^2\,\mathrm{d}w\\ &= \int_0^1 Df^{k-m}(w)\,(L^m_f 1)(w)^2\,\mathrm{d}w\\ &= \left\langle Df^{k-m}, (L^m_f 1)^2\right\rangle.\end{aligned}$$

Taking the sum over $m = 0, 1, 2, \ldots, k-1$, we have proved Theorem A. □

Remark 5.4. Stating from Equation (5.5), we can evaluate the case where the sum is restricted to $x = x'$. Changing variable $w = f^m(x)$ have

$$\int_0^1 \sum_{x\in f^{-m}(w)} \frac{Df^{k-m}(w)}{\left[Df^m(x)\right]^2}\,\mathrm{d}w = \int_0^1 \frac{Df^{k-m}(f^m(x))}{\left[Df^m(x)\right]^2} Df^m(x)\,\mathrm{d}x$$

$$= \int_0^1 \frac{Df^{k-m}(f^m(x))}{Df^m(x)} \, \mathrm{d}x$$
$$= \int_0^1 \frac{Df^k(x)}{[Df^m(x)]^2} \, \mathrm{d}x.$$

References

[BCG+78] Ginacarlo Benettin, Mario Casartelli, Luigi Galgani, Antonio Giorgilli, and Jean-Marie Strelcyn, *On the reliability of numerical studies of stochasticity. I. Existence of time averages*, Nuovo Cimento B (11) **44** (1978), no. 1, 183–195. MR 0478237 (57 #17722)

[BCG+79] ———, *On the reliability of numerical studies of stochasticity. II. Identification of time averages*, Nuovo Cimento B (11) **50** (1979), no. 2, 211–232. MR 534103 (81i:58031)

[BGSR16] Thierry Bodineau, Isabelle Gallagher, and Laure Saint-Raymond, *The Brownian motion as the limit of a deterministic system of hard-spheres*, Invent. Math. **203** (2016), no. 2, 493–553. MR 3455156

[Boy86] Abraham Boyarsky, *Computer orbits*, Comput. Math. Appl. Ser. A **12** (1986), no. 10, 1057–1064. MR 862028 (87m:58089)

[CM95] Carlos A. Cabrelli and Ursula M. Molter, *The Kantorovich metric for probability measures on the circle*, J. Comput. Appl. Math. **57** (1995), no. 3, 345–361. MR 1335789

[CQ01] James Campbell and Anthony Quas, *A generic C^1 expanding map has a singular S-R-B measure*, Comm. Math. Phys. **221** (2001), no. 2, 335–349. MR 1845327 (2002d:37038)

[DV98] Phil Diamond and Igor Vladimirov, *Asymptotic independence and uniform distribution of quantization errors for spatially discretized dynamical systems*, Internat. J. Bifur. Chaos Appl. Sci. Engrg. **8** (1998), no. 7, 1479–1490. MR 1661120

[DV02a] ———, *Branching processes and computational collapse of discretized unimodal mappings*, Internat. J. Bifur. Chaos Appl. Sci. Engrg. **12** (2002), no. 12, 2847–2867. MR 1956409

[DV02b] ———, *Set-valued Markov chains and negative semitrajectories of discretized dynamical systems*, J. Nonlinear Sci. **12** (2002), no. 2, 113–141. MR 1894464

[EG13] Peyman Eslami and Pawel Góra, *Stronger Lasota-Yorke inequality for one-dimensional piecewise expanding transformations*, Proc. Amer. Math. Soc. **141** (2013), no. 12, 4249–4260. MR 3105868

[Flo02] Paul Philipp Flockermann, *Discretizations of expanding maps*, Ph.D. thesis, ETH (Zurich), 2002.

[GB88] Paweł Góra and Abraham Boyarsky, *Why computers like Lebesgue measure*, Comput. Math. Appl. **16** (1988), no. 4, 321–329. MR 959419 (89m:58107)

[GHR12] Stefano Galatolo, Mathieu Hoyrup, and Cristóbal Rojas, *Statistical properties of dynamical systems—simulation and abstract computation*, Chaos Solitons Fractals **45** (2012), no. 1, 1–14. MR 2863582 (2012m:37138)

[Ghy94] Étienne Ghys, *Variations autour du théorème de récurrence de Poincaré*, 1994, Republié en 2006 dans "L'héritage scientifique de Henri Poincaré", Belin.

[GM22] Pierre-Antoine Guihéneuf and Maurizio Monge, *Discrepancy and discretizations of circle expanding maps II: simulations*, 2022, arXiv 2206.08000.

[GS21] Stefano Galatolo and Alfonso Sorrentino, *Quantitative statistical stability and linear response for irrational rotations and diffeomorphisms of the circle*, 2021.

[Gui15a] Pierre-Antoine Guihéneuf, *Discrétisations spatiales de systèmes dynamiques génériques*, Ph.D. thesis, Université Paris-Sud, 2015.

[Gui15b] ———, *Dynamical properties of spatial discretizations of a generic homeomorphism*, Ergodic Theory Dynam. Systems **35** (2015), no. 5, 1474–1523. MR 3365731

[Gui18] ———, *Physical measures of discretizations of generic diffeomorphisms*, Ergodic Theory Dynam. Systems **38** (2018), no. 4, 1422–1458. MR 3789171

[Gui19] ———, *Degree of recurrence of generic diffeomorphisms*, Discrete Analysis (2019), no. 1, 1–43.

[Haj41] Georg Hajós, *Über einfache und mehrfache Bedeckung des n-dimensionalen Raumes mit einem Würfelgitter*, Math. Z. **47** (1941), 427–467. MR 0006425 (3,302b)

[Lan98] Oscar E. Lanford, *Informal remarks on the orbit structure of discrete approximations to chaotic maps*, Experiment. Math. **7** (1998), no. 4, 317–324. MR 1678095 (2000b:37035)

[Mie06] Tomasz Miernowski, *Discrétisations des homéomorphismes du cercle*, Ergodic Theory Dynam. Systems **26** (2006), no. 6, 1867–1903. MR 2279269 (2008d:37065)

[Qua99] Anthony Quas, *Most expanding maps have no absolutely continuous invariant measure*, Studia Math. **134** (1999), no. 1, 69–78. MR 1688216 (2000a:37009)

[Rac91] Svetlozar T. Rachev, *Probability metrics and the stability of stochastic models*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, Ltd., Chichester, 1991. MR 1105086

[Via97] Marcelo Viana, *Stochastic dynamics of deterministic systems*, Brazillian Math. Colloquium, IMPA, 1997.

[VKD00] Igor Vladimirov, Nikolai Kuznetsov, and Phil Diamond, *Frequency measurability, algebras of quasiperiodic sets and spatial discretizations of smooth dynamical systems*, Math. Comput. Simulation **52** (2000), no. 3-4, 251–272. MR 1769577

[Vla96] Igor Vladimirov, *Quantized linear systems on integer lattices: a frequency-based approach*, 1996, Parts I,II.

[Voe67] Valentin V. Voevodin, *The asymptotic distribution of round-off errors in linear transformations*, Ž. Vyčisl. Mat i Mat. Fiz. **7** (1967), 965–976. MR 219227

[VV03] Franco Vivaldi and Igor Vladimirov, *Pseudo-randomness of round-off errors in discretized linear maps on the plane*, Internat. J. Bifur. Chaos Appl. Sci. Engrg. **13** (2003), no. 11, 3373–3393. MR 2031150

Pierre-Antoine Guihéneuf: Sorbonne Université and Université de Paris, CNRS, IMJ-PRG, F-75005 Paris, France.

Email address: pierre-antoine.guiheneuf@imj-prg.fr

Maurizio Monge: Instituto de Matemàtica da UFRJ, Av. Athos da Silveira Ramos 149, Centro de Tecnologia, Bloco C Cidade Univesitària, Ilha do Fundão, Caixa Postal 68530 21941-909, Rio de Janeiro, RJ, Brasil

Email address: maurizio.monge@im.ufrj.br