

Generic $\frac{1}{2}$ -discrepancy of $\{n\theta + x\}$

David Ralston

ABSTRACT. We study the pointwise rate of divergence of the sequence of *discrepancy sums*

$$S_i(x) = \sum_{j=0}^{i-1} (\chi_{[0,1/2)} - \chi_{[1/2,1)}) (x + j\theta)$$

for (Lebesgue) generic rotation parameter θ . Almost-sure upper and lower bounds for the rate of divergence are given by a Khinchin-like criterion related to convergence of a certain integral. Concluding remarks address the impossibility of finding a generic asymptotic rate of divergence as well as the related study of how frequently $S_i(x) = 0$.

CONTENTS

1. Introduction	195
2. Existing notation and prior results	197
3. Proof of Theorem 1	200
4. Proof of Theorem 2	202
5. Concluding remarks	204
References	207

1. Introduction

Given an irrational θ and some $x \in [0, 1)$, the $1/2$ -*discrepancy sums* of x are given by

$$S_i(x) = \sum_{j=0}^{i-1} (\chi_{[0,1/2)} - \chi_{[1/2,1)}) (x + j\theta),$$

where all addition is taken modulo one. We are interested in studying the asymptotic growth of $S_i(x)$ for generic θ . However, as $S_i(x)$ are not monotone, it is convenient to define a *maximal range function*:

$$(1) \quad \rho_n(x) = 1 + \max \{S_i(x) : i = 1, \dots, n\} - \min \{S_i(x) : i = 1, \dots, n\}.$$

In [Ra12], it was shown that there is a natural renormalization procedure for studying this sequence; we will briefly reintroduce the necessary notation

Received September 25, 2013.

2010 *Mathematics Subject Classification*. Primary: 11K38, Secondary: 37E20, 37B10.

Key words and phrases. discrepancy, irrational rotation, renormalization, substitution.

in §2, but familiarity with the existing techniques will be helpful. At the center of this renormalization procedure is a map $g : S^1 \rightarrow S^1$, closely related to the Gauss map. While that work was concerned with developing the renormalization procedure in order to construct specific behaviors, we will begin our study of *generic* growth rates by showing first:

Theorem 1. *There is a unique measure μ_g on the circle which is mutually absolutely continuous with respect to Lebesgue measure and is preserved by g . Both Radon–Nikodym derivatives $dx/d\mu_g$ and $d\mu_g/dx$ are essentially bounded. Furthermore, the system $\{S^1, \mu_g, g\}$ is exponentially CF-mixing.*

If $\rho_n(x) \in o(b_n)$ for every $x \in [0, 1)$, then we will simply write $\rho_n \in o(b_n)$. Similarly, we will write $\rho_n \notin o(b_n)$ if for every x we have $\rho_n(x) \notin o(b_n)$. Note that trivially $S_n(x) \in o(b_n)$ for every x if and only if $\rho_n \in o(b_n)$, and similarly for $S_n(x) \notin o(b_n)$.

A function $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ is said to be *regularly varying* if for any $t > 0$ we have $f(tx) \sim f(x)$:

$$f(tx) \in O(f(x)), \quad f(x) \in O(f(tx)).$$

Theorem 2. *Suppose that $f(x)$ is a continuous nondecreasing function which is regularly varying, defined for all $x \geq C$. Define the sequence $\{F(n)\}$ by*

$$(2) \quad F(n) = \int_C^n f(x) dx.$$

Then for almost every θ , either $\rho_n \in o(F(\log n))$ or $\rho_n \notin o(f(\log n))$ according to whether

$$\int_C^\infty f(x) dx < \infty \quad \text{or} \quad \int_C^\infty f(x) dx = \infty.$$

The term *1/2-discrepancy sums* harkens to classical discrepancy theory. The study of discrepancy theory in general tends to be highly probabilistic in flavor, with many results holding in measure and relating to the sequence $\{n_k\theta\}$, where n_k is typically assumed to satisfy strict growth conditions (see, e.g., [AB11] for an overview); our use of $\{k\theta + x\}$ does not change the process through the horizontal shift $+x$, but the slow growth of the multiples of θ used prohibit many useful probabilistic approaches, suggesting a dynamic approach is more appropriate. Our approach therefore is more similar to the original landmark paper in the field [Ke60]. Compared to that result, the results above are similar in that a continued-fraction approach is utilized, but by tailoring the process to a renormalization specific to this particular interval $[0, 1/2]$, we are able to derive *pointwise* behavior, rather than *distributional* behavior. Our bounds in Theorem 2 are not as specific as that of Kesten; by appealing to results in *trimmed ergodic sums*, remarks in the conclusion show that the strong and specific distributional limits obtained by Kesten cannot be achieved pointwise.

2. Existing notation and prior results

All notation is consistent with [Ra12]. We use standard continued fraction notation, and as $\theta \in (0, 1)$ without loss of generality, we omit the integer part and write for $a_i \in \mathbb{Z}^+$

$$\theta = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\ddots}}} = [a_1, a_2, \dots].$$

The partial quotients may be written as $a_i(\theta)$ when θ is not immediately clear from context. The Gauss map acts as the one-sided shift on the sequence of partial quotients for irrational θ in the following manner:

$$\gamma(\theta) = \frac{1}{\theta} - a_1, \quad \gamma([a_1, a_2, \dots]) = [a_2, a_3, \dots].$$

This action extends naturally to rational numbers as well, though this distinction is not important for our purposes. The Gauss map preserves a unique probability measure which is mutually absolutely continuous with respect to Lebesgue measure, and it is exponentially CF-mixing (see, e.g., [AN03] for a definition of this term; we will not use it in any technical way) with respect to this measure. We denote this measure by μ_γ .

The function g acts on irrational $\theta = [a_1, a_2, \dots]$ as follows:

$$g([a_1, a_2, a_3, \dots]) = \begin{cases} [a_2 + 1, a_3, \dots] = 1 - \theta & (a_1 = 1) \\ [1, a_2, a_3, \dots] = \frac{1}{1 + \gamma(\theta)} & (a_1 = 1 \pmod{2}, a_1 \neq 1) \\ [a_3, a_4, \dots] = \gamma^2(\theta) & (a_1 = 0 \pmod{2}). \end{cases}$$

We define

$$E(x) = \max\{n \leq x : n \in \mathbb{Z}, n \equiv 0 \pmod{2}\},$$

and we then let

$$\theta_n = g^n(\theta), \quad \delta_n = 1 - E(a_1(\theta_n))\theta_n.$$

If we assume that $\theta = \theta_0 < 1/2$, then we may label the intervals

$$A = [0, 1/2), \quad B = [1/2, 1 - \theta), \quad C = [1 - \theta, 1).$$

We omit a discussion of standard terminology in substitutions systems (see [Ra12, §2]), except to point out that most techniques are carried out in the compact but disconnected space in which A , B and C are all compact. *This distinction can only cause problems in regards to the endpoints of the intervals, and any orbit can include at most two of the endpoints.* Define the substitutions $\sigma_n = \sigma(\theta_n)$ according to Table 1. For convenience denote

$$\sigma^{(n)} = \sigma_0 \circ \sigma_1 \circ \dots \circ \sigma_{n-1}.$$

Under our labeling of A , B and C , we may without confusion define for finite words Ω in the alphabet $\{A, B, C\}$

$$S(\Omega) = \#A - \#B - \#C,$$

Case	Substitution
$a_1 = 2k, a_3 \neq 1$	$A \rightarrow (A^{k+1}B^{k-1}C)(A^k B^{k-1}C)^{a_2-1}$ $B \rightarrow (A^k B^k C)(A^k B^{k-1}C)^{a_2-1}$ $C \rightarrow (A^k B^k C)(A^k B^{k-1}C)^{a_2}$
$a_1 = 2k, a_3 = 1$	$A \rightarrow (A^k B^k C)(A^k B^{k-1}C)^{a_2}$ $B \rightarrow (A^{k+1}B^{k-1}C)(A^k B^{k-1}C)^{a_2}$ $C \rightarrow (A^{k+1}B^{k-1}C)(A^k B^{k-1}C)^{a_2-1}$
$a_1 = 2k + 1$	$A \rightarrow A^k B^k C$ $B \rightarrow A^{k+1} B^{k-1} C$ $C \rightarrow A$
$a_1 = 1$	$A \rightarrow A$ $B \rightarrow B$ $C \rightarrow C$

TABLE 1. The substitution σ as a function of θ .

and denoting by $\Omega(n)$ the first n letters in Ω (for $0 < n \leq |\Omega|$, the length of Ω), we may similarly define

$$\rho(\Omega) = 1 + \max\{S(\Omega(n)) : n = 1, 2, \dots, |\Omega|\} \\ - \min\{S(\Omega(n)) : n = 1, 2, \dots, |\Omega|\}.$$

By [Ra12, Theorem 1.1], there is a special point $x(\theta)$ for which the sequence of words

$$(3) \quad \Omega_n = \sigma^{(n)}(A)$$

correctly encodes the orbit of $x(\theta)$ under rotation by θ with at most two errors. That is, with two possible exceptions, the sequence of letters in Ω_n correctly specifies in which interval $(x + i\theta) \bmod 1$ lies. Combined with [Ra12, Proposition 5.2], we have

$$(4) \quad \rho(\Omega_n) = \frac{1}{2} \sum_{i=0}^n E(a_1(\theta_n)) + \xi,$$

where $|\xi| \leq 5$. As $\{S_n(x)\}$ must be unbounded for every x [Ke66], the bounded term ξ does not affect the claims of Theorem 2.

The words Ω_n represent the encoding of the first return of $x(\theta)$ to an interval \tilde{I}_n . The intervals \tilde{I}_n are nested and compact (in the associated space in which A , B and C are all compact), and of length

$$|\tilde{I}_n| = \delta_0 \cdot \delta_1 \cdots \delta_{n-1}.$$

The point $x(\theta)$ is the intersection of all \tilde{I}_n ; we point out that for all irrational θ

$$(5) \quad \liminf_{n \rightarrow \infty} \left| \frac{\log(\delta_0 \cdots \delta_{n-1})}{n} \right| \geq \frac{\log 2}{2} > 0,$$

which is to say that the length of \tilde{I}_n always decays at least exponentially fast. This observation is direct from the fact that no two consecutive θ_i may both be larger than $1/2$, and for $\theta_i < 1/2$ we have $\delta_i < 1/2$.

While not explicitly stated in the proof of [Ra12, Theorem 1.1], the following is immediate from the techniques of the proof, especially [Ra12, Proposition 4.1]:

Corollary. *For each $y \in [0, 1)$ and every n , there are two words $\Upsilon_0(y, n)$, $\Upsilon_1(y, n)$, where*

$$\Upsilon_1(y, n) \in \{\sigma^{(n)}(A), \sigma^{(n)}(B), \sigma^{(n)}(C)\},$$

and $\Upsilon_0(y, n)$ is either a proper right factor of one of these words or empty, and the concatenated word

$$(6) \quad \Omega_n(y) = \Upsilon_0(y, n)\Upsilon_1(y, n)$$

correctly encodes the orbit of y , except for at most two errors.

Proof. By [Ra12, Theorem 1.1], the induced transformation on each \tilde{I}_n is rotation by θ_n , and we may therefore label intervals A, B and C within \tilde{I}_n according to this new rotation θ_n ; the specific labeling is outlined in that proof. Let $\Upsilon_0(y, n)$ be the word which encodes the orbit of y (in the disconnected symbol space corresponding to the original transformation) through its return to \tilde{I}_n (if $y \in \tilde{I}_n$, then let $\Upsilon_0(y, n)$ be empty), and let $\Upsilon_1(y, n)$ be $\sigma^{(n)}(*)$, where $*$ represents which of the intervals A, B, C contains this first point in the orbit of y (under the correct labeling of intervals in the induced system on \tilde{I}_n). As the collection

$$\{\sigma^{(n)}(A), \sigma^{(n)}(B), \sigma^{(n)}(C)\}$$

encodes all possible orbits of points in \tilde{I}_n until their return to \tilde{I}_n and the original rotation by θ is minimal, Υ_0 must be a right factor of one of these words. □

It is worth pointing out that $\Omega_n(y)$ correctly encodes the orbit of y , but is not in general a word of the same length as Ω_n .

Define the matrices $M_n = M(\theta_n)$ according to Table 2 (the eigenvalues are readily computable). For convenience, denote

$$M^{(n)} = M_{n-1} \cdot M_{n-2} \cdots M_0,$$

so by [Ra12, Lemma 5.4], we have both that $|\sigma^{(n)}(A)| = |\sigma^{(n)}(B)|$ and

$$(7) \quad M^{(n)} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} |\sigma^{(n)}(A)| \\ |\sigma^{(n)}(C)| \end{bmatrix}.$$

Case	$M(\theta)$	Eigenvalues
$a_1 = 2k, a_3 \neq 1$	$\begin{bmatrix} (a_1 - 1)a_2 + 1 & a_2 \\ (a_1 - 1)a_2 + a_1 & a_2 + 1 \end{bmatrix}$	$ka_2 + 1 \pm \sqrt{ka_2(ka_2 + 2)}$
$a_1 = 2k, a_3 = 1$	$\begin{bmatrix} (a_1 - 1)a_2 + a_1 & a_2 + 1 \\ (a_1 - 1)a_2 + 1 & a_2 \end{bmatrix}$	$ka_2 + 1 \pm \sqrt{ka_2(ka_2 + 2)}$
$a_1 = 2k + 1$	$\begin{bmatrix} a_1 - 1 & 1 \\ 1 & 0 \end{bmatrix}$	$k \pm \sqrt{k^2 + 1}$
$a_1 = 1$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	1

TABLE 2. The matrices $M(\theta)$ used to determine the lengths of $\sigma^{(n)}(A)$, $\sigma^{(n)}(C)$. Note that $2k = E(a_1)$.

3. Proof of Theorem 1

Note that $g^{-1}(1/2, 1) \subset (0, 1/2)$, so any probability measure μ which is preserved by g must have $\mu(1/2, 1) \leq 1/2$: the Gauss measure μ_γ is not preserved by g .

Define the following collection of open intervals (for $n, m, k = 1, 2, \dots$):

$$\left(\frac{1}{2}, 1\right), \quad \left(\frac{1}{2k+2}, \frac{1}{2k+1}\right), \quad \left(\frac{m}{2nm+1}, \frac{m+1}{2n(m+1)+1}\right).$$

The middle intervals are those θ whose continued fraction expansion begins with an odd number (except one); we will refer to the collection of all such intervals as *odd intervals*. The right-most intervals are those θ whose continued fraction expansion begins with the pair $[2n, m, \dots]$, and we will refer to the collection of all such intervals as *even intervals*. The collection of odd and even intervals, together with $(1/2, 1)$ we denote \mathcal{C} , and note that \mathcal{C} covers S^1 except for a countable set of points.

Lemma 3.1. *We have each of the following:*

- (I) \mathcal{C} is a Markov partition for g .
- (II) For $c_1, c_2 \in \mathcal{C}$, if $g(c_1) \cap c_2 \neq \emptyset$, then $c_2 \subset g(c_1)$.
- (III) There is a k such that $(0, 1) \subset g^k(c)$ for each $c \in \mathcal{C}$.
- (IV) The map g is monotone and $1:1$ on each $c \in \mathcal{C}$, and extends to the closure of each c to a C^2 function.
- (V) g is expansive; $\exists k$ and $d > 1$ such that $|(g^k)'(x)| \geq d$ for almost all x .
- (VI) g has the Renyi (or strong distortion) property: $\exists d$ such that for every $c \in \mathcal{C}$

$$\sup_{x \in c} \left(\frac{|g''(x)|}{(g'(x))^2} \right) < d.$$

- (VII) The endpoints of all $c \in \mathcal{C}$ map (via the C^2 extension from item (IV)) to a finite set.

Proof. The restriction of g to each $c \in \mathcal{C}$ is invertible, and we have chosen \mathcal{C} to generate the Borel σ -algebra under g^{-1} ; if $\theta_1 \neq \theta_2$, then there is some minimal index i such that $a_i(\theta_1) \neq a_i(\theta_2)$, from which one sees that there is some k such that either (see for example [Ra12, Eqn. 20])

$$a_1(g^k\theta_1) \neq a_1(g^k\theta_2), \quad \text{or} \quad a_1(g^k\theta_1) = a_1(g^k\theta_2) = 0 \pmod 2, \\ a_2(g^k\theta_1) \neq a_2(g^k\theta_2).$$

So item (I) is shown.

The related items (II) and (III) can both be shown directly using the fact that g maps odd intervals to $(1/2, 1)$, g maps $(1/2, 1)$ to $(0, 1/2)$, and g maps even intervals to $(0, 1)$.

On the even intervals, $g = \gamma^2$ is the square of the Gauss map, and the even intervals are members of the standard Markov partition for γ^2 , which is well-known to have all of these properties. Furthermore, on $(1/2, 1)$ we have $g(\theta) = 1 - \theta < 1/2$. Using the chain rule, then, we need only establish the remaining items for the odd intervals. So, let us consider the odd interval for a fixed k . The reader may verify that for $\theta = [2k + 1, \dots]$ we have

$$\frac{1}{2k + 1} < 1 - 2k\theta < \frac{1}{k + 1}.$$

Using this inequality, one may show:

$$g(\theta) = \frac{1}{1 + \gamma(\theta)} = \frac{\theta}{1 - 2k\theta}, \\ g'(\theta) = \frac{1}{(1 - 2k\theta)^2} \geq (k + 1)^2, \\ g''(\theta) = \frac{4k}{(1 - 2k\theta)^3} \leq 4k(2k + 1)^3.$$

Item (IV) is immediate from the first line (for θ in this odd interval, $\theta < (2k + 1)^{-1}$, so there is no asymptote), and item (V) from the second (recall that $k \geq 1$). Item (VI) follows from the bounds on g' and g'' . \square

The literature surrounding the existence of invariant measures for Markov maps of the interval is vast. We have arranged Lemma 3.1 to match the statement which appears in [BS79]. This theorem gives the existence of a unique probability density μ_g , supported on all of $[0, 1)$, which is invariant under g and continuous with respect to Lebesgue measure, with essentially bounded Radon–Nikodym derivative. Therefore $\{S^1, \mu_g, g\}$ is ergodic. Following the trail of references in that work back to [PV75] actually says more, however, that the system $\{S^1, \mu_g, g\}$ is weakly Bernoulli (and therefore exact). Item (III) gives that g is topologically mixing (any open set contains an open subinterval specified by an initial finite string of partial quotients), so by [Aa97, Corollary 4.7.8], our system is exponentially CF-mixing. That the Radon–Nikodym derivative $d\mu_g/dx$ is bounded away from zero (not just bounded) is given as a remark in the third paragraph of [Aa97, §4.7].

The following Khinchin-like characterization follows immediately via a standard shrinking-target result (or the Borel–Cantelli Lemma) as g is mixing and $\mu_g \sim \mu_\gamma$, the measure preserved by the Gauss map (as both are mutually absolutely continuous with respect to Lebesgue measure):

Corollary 3.1. *Let $\{b_i\}$ be a sequence of positive real numbers for $i = 0, 1, \dots$. Then the inequality*

$$a_1(g^n\theta) > b_n$$

is satisfied almost surely infinitely many times or only finitely many times according to whether the series

$$\sum_{i=0}^{\infty} \frac{1}{b_i}$$

diverges or not.

4. Proof of Theorem 2

For a square real-valued matrix M , let $\|M\|$ be the largest eigenvalue (in modulus), and for a real-valued column vector u , let $\|u\|$ be the largest element (in absolute value).

Lemma 4.1. *With $M(\theta)$ as given by Table 2, we have both $\log \|M(\theta)\|$ and $\log \|M^{-1}(\theta)\|$ in $L^1(X, \mu_g)$.*

Proof. As the two-by-two matrices $M(\theta)$ all have $|\det M(\theta)| = 1$, we have $\|M(\theta)\| = \|M^{-1}(\theta)\|$. As μ_g is mutually absolutely continuous with Lebesgue measure, it therefore suffices to show that

$$\int_0^1 \log \|M(\theta)\| d\theta < \infty.$$

For $\theta \in (1/2, 1)$, we have $\|M(\theta)\| = 1$. If $a_1(\theta) = 2k + 1$ for $k \neq 1$, then

$$\log \|M(\theta)\| = \log |k + \sqrt{k^2 + 1}| < \log(2k + 1).$$

On the other hand, for $a_1(\theta) = 2n$ and $a_2(\theta) = m$ we have

$$\log \|M(\theta)\| = \log |nm + 1 + \sqrt{nm(nm + 2)}| \leq \log(2nm + 2).$$

So we may therefore sum across the even and odd intervals:

$$\begin{aligned} & \int_0^1 \log \|M(\theta)\| d\theta \\ & \leq \sum_{k=1}^{\infty} \int_{\frac{1}{2k+2}}^{\frac{1}{2k+1}} \log(2k + 1) d\theta + \sum_{n,m=1}^{\infty} \int_{\frac{m}{2nm+1}}^{\frac{m+1}{2n(m+1)+1}} \log(2nm + 2) d\theta \\ & = \sum_{k=1}^{\infty} \frac{\log(2k + 1)}{(2k + 1)(2k + 2)} + \sum_{n,m=1}^{\infty} \frac{\log(2nm + 2)}{(2nm + 1)(2n(m + 1) + 1)}, \end{aligned}$$

and the summability of both series is direct. \square

By the Oseledec ergodic theorem, then, for almost every θ we have

$$(8) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(M^{(n)} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) = \lambda$$

for some $\lambda < \infty$.

Proposition 4.1. *For almost every θ , for every $\epsilon > 0$ we have for sufficiently large n (recall (3))*

$$(\lambda - \epsilon)^{n-3} \leq |\Omega_n| \leq (\lambda + \epsilon)^n.$$

Proof. The upper inequality is direct in light of the previous remarks and (7). For the lower inequality we must establish

$$\min\{|\sigma^{(n)}(A)|, |\sigma^{(n)}(C)|\} \geq \max\{|\sigma^{(n-3)}(A)|, |\sigma^{(n-3)}(C)|\}.$$

The proof may be accomplished through an exhaustive case-by-case analysis of different possible forms for the matrices M_{n-1} , M_{n-2} and M_{n-3} . The situation is easiest in the case that $a_1(\theta_{n-1}) = 0 \pmod 2$. For example, suppose that

$$M_{n-1} = \begin{bmatrix} (2k-1)m+1 & m \\ (2k-1)m+2k & m+1 \end{bmatrix}, \quad M^{(n-1)}u = \begin{bmatrix} A \\ B \end{bmatrix}.$$

Then we have

$$M^{(n)}u = \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} ((2k-1)m+1)A + mB \\ ((2k-1)m+2k)A + (m+1)B \end{bmatrix},$$

and we clearly have

$$\min\{C, D\} \geq \max\{A, B\}.$$

The matrices $M(\theta)$ for $a_1(\theta) = 1 \pmod 2$ are less trivial, but the composition of two such matrices is seen to have the desired property. Since two such matrices can only occur separated by an identity matrix, and it is possible that M_{n-1} was the identity matrix, then $(n-3)$ in the exponent is sufficient. \square

Lemma 4.2. $\lambda > 1$.

Proof. See (5). As the length of \tilde{I}_n decays at least exponentially fast (it is not larger than $2^{-n/2}$), the return time of any point in \tilde{I}_n to itself increases exponentially fast. As the entries of $M^{(n)} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ are the two return times of points in \tilde{I}_n to itself (7), we must have $\lambda \geq \sqrt{2}$. \square

Now let $f(x)$ be continuous, nondecreasing and regularly varying, and define $F(t)$ as in the introduction (2). We proceed now under the assumption that $f(x)$ is not integrable. From Corollary 3.1, it follows that for generic θ and infinitely many n , we have $f(n) < (1/2)E(a_1(\theta_n))$. So by (4),

$$\rho(\Omega_n) > f(n).$$

By Proposition 4.1, then, we have infinitely many times N such that (recall (1) for the definition of ρ_N)

$$\rho_N(x(\theta)) \geq f(C \log N),$$

where $C = \log(\lambda + \epsilon)$, so as f was regularly varying, $\rho_n(x(\theta)) \notin o(f(\log n))$.

Similarly, if f is integrable, we eventually have for generic θ that

$$(1/2)E(a_1(\theta_n)) < f(n),$$

so up to a bounded difference which we ignore, we may say that for generic θ we eventually have

$$\frac{1}{2} \sum_{i=0}^{n-1} E(a_1(\theta_i)) < F(n),$$

from which it follows (by considering all $|\Omega_{n-1}| \leq N \leq |\Omega_n|$ and using regularity of f and Proposition 4.1 as before) that $\rho_n(x(\theta)) \in O(F(\log n))$. As we may multiply f by any $\epsilon > 0$ without affecting summability, then, we have $\rho_n(x(\theta)) \in o(F(\log n))$.

So for the specific point $x(\theta)$, the growth of the range of $S_n(x)$ (that is, the sequence $\rho_n(x(\theta))$) grows subject to the desired bounds. It remains to show that the behavior of *any* $\rho_n(y)$ may be accurately considered through the sequence $\rho_n(x(\theta))$.

Note that regardless of the sequence of substitutions σ_i , we have (refer to Table 1 and [Ra12, Proposition 5.1])

$$\rho(\sigma^{(n-1)}(A)) \leq \rho(\sigma^n(*)) \leq \rho(\sigma^n(A)),$$

where $* \in \{A, B, C\}$. Finally, it is clear that for any two words ν_1 and ν_2 , the range of their concatenation is bounded by the sum of their individual ranges:

$$\rho(\nu_1\nu_2) \leq \rho(\nu_1) + \rho(\nu_2).$$

Altogether, then, recall (6), to see that for *any* y , for *every* θ and *any* $n \geq 1$:

$$\rho(\Omega_{n-1}) \leq \rho(\Omega_n(y)) \leq 2\rho(\Omega_n).$$

So, for any $\epsilon > 0$ and sufficiently large n

$$(\lambda - \epsilon)^{n-4} \leq |\Omega_{n-1}| \leq |\Omega_n(y)| \leq 2 \max\{|\sigma^{(n)}(A)|, |\sigma^{(n)}(C)|\} \leq 2(\lambda + \epsilon)^n.$$

All previous arguments using regularity of $\{f(n)\}$ still apply, then, extending the existing arguments to all points for generic θ and completing the proof of Theorem 2.

5. Concluding remarks

We present a pair of simple computations to show the relatively tight control we may generically impose upon $\rho_n(y)$ for any y and for generic θ . A classical application of the Denjoy–Koksma inequality is to show that for θ having bounded partial quotients, for every x we have $\rho_n(x) \in O(\log n)$ (see for example [Co09, §2.1]). It was shown in [Ra12, Theorem 1.4] that in fact

$\rho_n(x) \sim \log(n)$ for such θ . However, by setting $f(x) = x \log x \cdots \log^{(k-1)} x$, the product of the first k iterated logarithms (starting at index zero, and defined for sufficiently large x), we see that for generic θ we have

$$\rho_n \notin o(\log n \cdots \log^{(k)} n).$$

Another classical application of the Denjoy–Koksma inequality is that for generic θ , for every x we have $\rho_n(x) \in o(n^\epsilon)$ for every $\epsilon > 0$ (also in [Co09, §2.1]). To improve upon this bound, consider

$$f(x) = x \log x \cdots \log^{(k-2)} x (\log^{(k-1)} x)^{1+\epsilon}$$

for sufficiently large x . Using the straightforward fact (one may estimate the integral with the corresponding sum over integers) that

$$\int_C^t x \log x \cdots \log^{(k)}(x) \sim t \log t \cdots \log^{(k)} t$$

(setting C sufficiently large so that all terms are defined), we have

$$\begin{aligned} F(\log n) &< \left(\log^{(k-1)}(\log n)\right)^{1+\epsilon} \left(\sum_{i=1}^{\log n} i \log i \cdots \log^{(k-2)}(i)\right) \\ &\sim \log n \cdots \log^{(k-1)}(n) (\log^{(k)}(n))^{1+\epsilon}, \end{aligned}$$

so generically our discrepancy sums grow slower than all such functions.

One might hope that there could be some sequence such that generically $\rho_n \sim b_n$. Such a quest would be a fool’s errand: the sequence $\rho(\Omega_n)$ is given by a partial ergodic sum of the nonintegrable nonnegative function $E(a_1\theta)$, and the growth rate is therefore seen to be similar to the ergodic sums of simply $a_1(\theta)$. While the map g and the Gauss map γ are not identical and do not preserve the same measure, they are both exponentially CF-mixing with respect to mutually absolutely continuous measures.

The partial sums $a_1 + a_2 + \cdots + a_n$ (ergodic sums of the function $a_1(\theta)$ under the action of γ) almost surely do not have a strong law of large numbers (a result known already to Khinchin [Kh35]). However, that example *does* admit a weaker ‘trimmed’ law of large numbers [DV86]: for almost every θ ,

$$\lim_{n \rightarrow \infty} \frac{a_1 + a_1 + \dots + a_n - \max\{a_i : i = 1, \dots, n\}}{n \log n} = \frac{1}{\log 2}.$$

That is to say that the sums grow like $n \log n$ except for rare large partial quotients, corresponding to visits to the cusp in the corresponding flow associated to γ in the modular surface $SL_1(\mathbb{R})/SL_2(\mathbb{Z})$ (Khinchin proved in [Kh35] convergence *in measure* of the ratio $(a_1 + \dots + a_n)/n \log n$). This sort of normalization of trimmed sums was extended to a very general setting in [AN03]; ergodic sums of nonintegrable functions under exponentially CF-mixing maps, for example our own

$$\rho(\Omega_n) = E(a_1(\theta_0)) + \cdots + E(a_1(\theta_{n-1})),$$

may be compared to trimmed sums of regularly varying comparison sequences.

Off the interval $(1/2, 1)$, we have the trivial inequality

$$a_1(\theta) > (1/2)E(a_1(\theta)) > a_1(\theta)/3.$$

As the measure preserved by g is mutually absolutely continuous with respect to the Gauss measure, it follows from [AN03] that the trimmed sums in our system are almost surely not too far removed from the trimmed sums of partial quotients:

$$\frac{1}{2} \sum_{i=0}^{n-1} E(a_1(\theta_i)) - \frac{1}{2} \max\{E(a_1(\theta_i)) : i = 0, 1, \dots, n-1\} \sim n \log n.$$

However, we have as a corollary of Theorem 2 that

$$\limsup_{n \rightarrow \infty} \frac{\rho_n(x)}{\log n \log \log n} = \infty.$$

Finally, we close with a few remarks concerning *visits to zero of the deterministic random walk*. That is, consider the function

$$V_n(x) = \#\{i = 0, 1, \dots, n-1 : S_i(x) = 0\}.$$

The properties of this function have been studied in various contexts. In the event where θ is a quadratic irrational, for example, the sums are particularly well-behaved; there exist constants c_1 and c_2 such that for all n

$$c_1 \left(\frac{n}{\sqrt{\ln n}} \right) \leq \mathbb{E}(V_n) \leq c_2 \left(\frac{n}{\sqrt{\ln n}} \right),$$

a result which appears in [AK82], and significant refinements to this estimate have recently been made [ADDS]. The situation where θ is *not* a quadratic irrational is significantly less amenable to analysis. It follows directly from [CR13, Corollary 4.3], for example, that for θ of the form

$$\theta = [2r_1, s_1, 2r_2, s_2, \dots],$$

with the s_i bounded, we have for all $x \in [0, 1)$ and all $n \in \mathbb{N}$

$$V_{q_{2n}}(x) \leq \tau q_{2n-2},$$

where the constant $\tau = \tau(\theta)$ is independent of x and n , and the q_i refer to the standard continued fraction convergents. As for θ of this form we have

$$q_{2n} = (s_n r_n + 1)q_{2n-2} + s_n q_{2n-3} \geq r_n q_{2n-2},$$

along this subsequence of times we can, for *any* fixed sequence $b_n \rightarrow 0$, easily choose the r_n large enough so that

$$\frac{q_{2n-2}}{q_{2n}} \leq b_n.$$

It follows that

$$\mathbb{E}(V_{q_{2n}}) \leq \max\{V_{q_{2n}}(x) : x \in [0, 1)\} \leq \tau q_{2n-2} \leq \tau(\theta) b_n q_{2n}.$$

While for quadratic irrationals the “appropriate” scaling function is $b_n = 1/\sqrt{\log(q_{2n})}$, for bizarrely constructed θ we may force visits to zero to be very rare events ($b_n \rightarrow 0$ was arbitrary!). The above argument is simply a rewording of the more intricate [CR13, Theorem 4.4] for the particular purpose of counting the visits to zero.

References

- [Aa97] AARONSON, JON. An introduction to infinite ergodic theory. *Mathematical Surveys and Monographs*, 50. *American Mathematical Society, Providence, RI*, 1997. xii+284 pp. ISBN: 0-8218-0494-4. MR1450400 (99d:28025), Zbl 0882.28013.
- [AK82] AARONSON, JON; KEANE, MICHAEL. The visits to zero of some deterministic random walks. *Proc. London Math. Soc.* (3) **44** (1982), no. 3, 535–553. MR656248 (83j:28016), Zbl 0489.60006.
- [AN03] AARONSON, JON; NAKADA, HITOSHI. Trimmed sums for non-negative, mixing stationary processes. *Stochastic Process. Appl.* **104** (2003), no. 2, 173–192. MR1961618 (2004c:60077), Zbl 1075.60511, doi:10.1016/S0304-4149(02)00236-3.
- [AB11] AISTLEITNER, CHRISTOPH; BERKES, ISTVÁN. Probability and metric discrepancy theory. *Stoch. Dyn.* **11** (2011), no. 1, 183–207. MR2771348 (2012d:11163), Zbl 1260.11051, doi:10.1142/S021949371100322X.
- [ADDS] AVILA, A.; DOLGOPYAT, D.; DURIÉV, E.; SARIG, O. The visits to zero of a random walk driven by an irrational rotation. Preprint, 2013. <http://www.wisdom.weizmann.ac.il/~sarigo/CylinderMap6.pdf>.
- [BS79] BOWEN, RUFUS; SERIES, CAROLINE. Markov maps associated with Fuchsian groups. *Inst. Hautes Études Sci. Publ. Math.* **50** (1979), 153–170. MR556585 (81b:58026), Zbl 0439.30033, doi:10.1007/BF02684772.
- [CR13] CHAIKA, JON; RALSTON, DAVID. ω -recurrence in skew products. *Ergodic Theory and Dynamical Systems* **8** (2013), 1–13. arXiv:1109.2999, doi:10.1017/etds.2013.15.
- [Co09] CONZE, JEAN-PIERRE. Recurrence, ergodicity and invariant measures for cocycles over a rotation. *Ergodic theory*, 45–70. *Contemp. Math.* 485, *Amer. Math. Soc., Providence, RI*, 2009. MR2553209 (2010i:37007), Zbl 1183.37005.
- [DV86] DIAMOND, HAROLD G; VAALER, JEFFREY D. Estimates for partial sums of continued fraction partial quotients. *Pacific J. Math.* **122** (1986), no. 1, 73–82. MR825224 (87f:11056), Zbl 0589.10056, doi:10.2140/pjm.1986.122.73.
- [Ke60] KESTEN, HARRY. Uniform distribution mod 1. *Ann. of Math.* (2) **71** (1960), 445–471. MR0113864 (22 #4695), Zbl 0101.03803.
- [Ke66] KESTEN, HARRY. On a conjecture of Erdős and Szűsz related to uniform distribution mod 1. *Acta Arith.* **12** (1966/1967), 193–212. MR0209253 (35 #155), Zbl 0144.28902.
- [Kh35] KHINTCHINE, A. Metrische Kettenbruchprobleme. *Compositio Math.* **1** (1935), 361–382. MR1556899, Zbl 0010.34101, http://www.numdam.org/item?id=CM_1935__1__361_0.
- [PV75] PHILLIPS, E.; VARADHAN, S. Eds. *Ergodic theory. A seminar held at the Courant Institute of Mathematical Sciences, New York University, New York, 1973–1974.* With contributions by S. Varadhan, E. Phillips, S. Alpern, N. Bitzenhofer and R. Adler. *Courant Institute of Mathematical Sciences New York University, New York*, 1975. v+131 pp. MR0486431 (58 #6177), Zbl 0305.28010.
- [Ra12] RALSTON, D. Substitutions and $1/2$ -discrepancy of $\{n\theta + x\}$. *Acta Arith.*, **154** (2012), no. 1, 1–28. MR2943672, Zbl 1272.11092, arXiv:1105.6301.

DEPARTMENT OF MATHEMATICS/CIS, SUNY COLLEGE AT OLD WESTBURY, PO 210,
OLD WESTBURY, NY 11568
ralstond@oldwestbury.edu

This paper is available via <http://nyjm.albany.edu/j/2014/20-11.html>.