Pseudo-randomness of a random Kronecker sequence

Eda Cesaratto, Brigitte Vallée

Pseudorandomness of a Random Kronecker Sequence

Eda Cesaratto[1] and Brigitte Vallée[2]

[1] CONICET and Univ. Nac. de Gral. Sarmiento,
J.M. Gutierrez 1150, 1613 Los Polvorines, Buenos Aires Argentina
ecesarat@ungs.edu.ar
[2] Laboratoire GREYC, CNRS UMR 6072 and Université de Caen,
F-14032 Caen, France
brigitte.vallee@unicaen.fr

Dedicated to Philippe Flajolet (1948–2011)

Abstract. We study two randomness measures for the celebrated Kronecker sequence $\mathcal{S}(\alpha)$ formed by the fractional parts of the multiples of a real α. The first measure is the well-known discrepancy, whereas the other one, the Arnold measure, is less popular. Both describe the behaviour of the truncated sequence $\mathcal{S}_T(\alpha)$ formed with the first T terms, for $T \to \infty$. We perform a probabilistic study of the pseudorandomness of the sequence $\mathcal{S}(\alpha)$ (discrepancy and Arnold measure), and we give estimates of their mean values in two probabilistic settings : the input α may be either a random real or a random rational. The results exhibit strong similarities between the real and rational cases; they also show the influence of the number T of truncated terms, via its relation to the continued fraction expansion of α.

1 Introduction

Measures of Randomness. A measure of randomness on the unit interval $\mathcal{I} := [0, 1]$ tests how a sequence $\mathcal{X} \subset \mathcal{I}$ differs from a "truly random" sequence (see books [11] and [14] for a general discussion on the subject). Such a measure describes the difference between the behaviour of the truncated sequence $\mathcal{X}_T$ formed with the first T terms of the sequence and a "truly random" sequence formed with T elements of $\mathcal{I}$, and explains what happens for $T \to \infty$. Here, we consider the statistical pseudorandomness which is stronger than computational pseudorandomness, widely used in cryptography. Here, we study two measures of statistical pseudorandomness –the discrepancy and the Arnold measure– and we wish to compare them in the particular case of a Kronecker sequence $S(\alpha)$, formed of fractional parts of the multiples of a real α. Such a sequence can be precisely studied since its two measures of randomness are expressed as a function of the continued fraction expansion of the real α.

For a rational of the form $\alpha = u/v$, the sequence $S(\alpha)$ gives rise to an arithmetic progression $k \mapsto k\,u \pmod v$. This is a particular case of the general linear congruential generator (LCG) $x_{k+1} = ax_k + u \pmod v$, obtained here for $a = 1$.

Even if this particular LCG does not belong to the "best" class described by Knuth in [15], it is interesting to study its statistical randomness. The LCG's are widely used for statistical purposes because they are easily implemented and fast (see [15] for a general study). However, they are not adapted to a cryptographic use.

The study of the Arnold measure is just now beginning, with the proposal of Arnold himself in Problem 2003-2 of [1],[2], and the work of Cesaratto, Plagne and Vallée in [7] in the particular case of the Kronecker sequence.

The notion of discrepancy is much more popular, and the case of Kronecker sequences $\mathcal{S}(\alpha)$ is very well studied, with works of Weyl, Hardy, Behnke, Schmidt and Schoissengeier. A summary of the main results can be found in the book [4]. Weyl and Hardy proved that $(1/T)\Delta_T(\alpha)$ tends to zero and asked the question of the speed of convergence to 0. Behnke [5] first observed that the discrepancy $\Delta_T(\alpha)$ is of order $O(\log T)$ if and only if the sequence of quotients which appear in the continued fraction expansion of the real α admits bounded averages. Another class of results shows the influence of the "type" of integer T on the truncated sequence $\mathcal{S}_T(\alpha)$. This is due to the three distance theorem, conjectured by Steinhaus and proved by Surányi [20], Sós [19] and Świerczkowski [21] which states that there are at most three possible distinct distances between geometric consecutive points in the truncated sequence $\mathcal{S}_T(\alpha)$. The length and the number of these distinct distances depend on the relation between the truncation integer T and the continued fraction expansion of the real α. On the one hand, the discrepancy may be small: when the truncation integer T is a continuant of the continued fraction expansion of an irrational α, then the discrepancy satisfies $\Delta_T(\alpha) \leq 3$ ([4], Ch. 5, Sect. 2); moreover, Schoissengeier [23] proved the equality: $\liminf_{T\to\infty} \Delta_T(\alpha) = 1$. On the other hand, there are integers T for which the discrepancy is large: Schmidt [22] proved the existence of an absolute constant $C = (66 \log 4)^{-1}$ for which, for each irrational α, there exists an infinity of integers T so that $\Delta_T(\alpha) \geq C \log T$.

Thus, all the existing works which deal with the discrepancy adopt an "individual" point of view: for which reals α, and for which integers T, the discrepancy of the sequence $\mathcal{S}(\alpha)$ is minimal, maximal?

Our Points of View. We adopt different points of view, which appear to be new:

(a) We compare these two randomness measures (discrepancy and Arnold measure).

(b) We adopt a probabilistic point of view: we choose the "input" α at random, and we wish to study the randomness of a random sequence $\mathcal{S}(\alpha)$. We estimate in particular the mean values of the discrepancy and the Arnold measure, when the number T of terms tends to ∞.

We consider two distinct probabilistic settings: we study the usual case when α is a random real number, but we also focus on the particular rational case, where α is a random rational of the unit interval. This case is never studied in the literature, except in the paper [7]. Even if the behaviour of the sequence $\mathcal{S}_T(u/v)$ is only interesting if $T < v$, we may relate T and v so that they tend both to ∞.

(c) We focus on the special case when the pair (T, α) gives rise to the two distance phenomenon: the computations are easier, but already show very interesting phenomena. We consider two types of integers T: the case of "continuant" type and the case of a general integer which gives rise to a two–distance phenomenon.

Main Results. Our results exhibit three phenomena:

(a) First, the strong parallelism between the behaviour of the two randomness measures (discrepancy and Arnold measures)
(b) Second, the strong similarity between the two probabilistic settings (real and rational cases).
(c) Third, the strong influence of arithmetic properties of the number T of terms, as a function of the continued fraction expansion of the input α. The two distinct types of integers T –the "continuant" type and the general two–distance integer– give rise to distinct phenomena for pseudorandomness of a Kronecker sequence.

We then obtain eight results ($8 = 2^3$). Theorem 1 describes the case when T is an integer of "continuant" type. In this case, the four mean values – discrepancy and Arnold measure, in the real and the rational case– tend to finite limits close to 1. Moreover, for each randomness measure (discrepancy or Arnold measure), the limits in the real case and in the rational case coincide.

Theorem 2 deals with the case when T is a general integer which gives rise to a two–distance phenomenon. For this type of integer T, the mean value of each measure is infinite in the real case. On rationals with a denominator at most N, the mean values are both of order $\log N$, with "similar" constants.

Finally, Theorem 3 studies the case of a general variable which involves the main parameters which appear in the continued fraction expansion. This result may be of independent interest, and also exhibits a strong similarity between the two probabilistic settings (real and rational cases).

This work strongly uses the dynamical analysis methodology, developed by Vallée [3],[25], which combines tools imported from dynamics, such as transfer operators, with various tools of analytic combinatorics: generating functions, Dirichlet series, Perron's formula.

Plan of the Paper. We first introduce the two randomness measures (Section 2). In Section 3, we describe our main results and interpret them in terms of pseudorandomness. Finally, Section 4 provides expressions of these randomness measures as a function of the main parameters which appear in the continued fraction expansion and explains the main steps of the proofs of our results.

2 Notions of Pseudo-randomness

This section describes the two measures of randomness which will be studied, first in the case of a general sequence. Then, it focuses to the particular case of the Kronecker sequences.

2.1 Case of a General Sequence

One considers a sequence $\mathcal{X}$ of the unit interval $\mathcal{I} := [0, 1]$, and, for an integer T, the truncated sequence $\mathcal{X}_T$ formed by the first T elements of the sequence $\mathcal{X}$. After re-ordering the sequence $\mathcal{X}_T$, one obtains an increasing sequence $\mathcal{Y}_T := \{y_i : i \in [1..T]\}$, and the distance $y_{i+1} - y_i$ between consecutive elements is denoted by δ_i, whereas the last distance δ_T is defined as $\delta_T := 1 + y_1 - y_T$.

The main question is: How closely does the truncated sequence $\mathcal{X}_T$ approximate a "truly random" sequence on $\mathcal{I}$? We consider here two main measures.The discrepancy compares the sequence $\mathcal{Y}_T$ to the fixed regular sequence (j/T), whereas the Arnold constant deals with the distances δ_i.

Discrepancy. For a general study of discrepancy, see the two books [14] and [11]. The discrepancy is a measure of how closely the truncated sequence $\mathcal{X}_T$ approximates the uniform distribution on $\mathcal{I}$. We denote by $|\mathcal{Y}|$ the cardinality of a finite set $\mathcal{Y}$, and by $\lambda(\mathcal{J})$ the length of the interval $\mathcal{J} \subset \mathcal{I}$.

A sequence $\mathcal{X}$ of the unit interval $\mathcal{I}$ is called uniformly distributed if,

$$\text{for any interval } \mathcal{J} \subset \mathcal{I}, \qquad \lim_{T\to\infty} \frac{1}{T} |\mathcal{X}_T \cap \mathcal{J}| = \lambda(\mathcal{J}).$$

The discrepancies $D_T(\mathcal{X}), \Delta_T(\mathcal{X})$, given by

$$D_T(\mathcal{X}) := \sup_{\mathcal{J}\subset\mathcal{I}} \left| \frac{1}{T} |\mathcal{X}_T \cap \mathcal{J}| - \lambda(\mathcal{J}) \right|, \qquad \Delta_T(\mathcal{X}) := T D_T(\mathcal{X}), \tag{1}$$

(where the supremum is taken over all the intervals $\mathcal{J} \subset \mathcal{I}$), estimate the speed of convergence towards the uniform distribution. As explained in [18], the discrepancy is expressed with two other sequences, defined by the "signed" distances between the ordered sequence $\mathcal{Y}_T$ and the reference sequence (j/T), namely

$$D_T^+(\mathcal{X}) = \sup_{j\in[1,T]} \gamma_j^+, \;\; D_T^-(\mathcal{X}) = \sup_{j\in[1,T]} \gamma_j^- \;\text{ with }\; \gamma_j^+ := \frac{j}{T} - y_j, \; \gamma_j^- := y_j - \frac{j-1}{T},$$

so that the relation $D_T(\mathcal{X}) = D_T^+(\mathcal{X}) + D_T^-(\mathcal{X})$ holds. In conclusion, the notion of discrepancy is mainly based on the comparison between the ordered sequence $\mathcal{Y}_T$ with the reference sequence (j/T).

Arnold Measure. There exists another measure of randomness, recently introduced by Arnold in [1], [2] and much less studied. Arnold proposed as a measure of randomness of the sequence $\mathcal{X}_T$ the normalized mean-value of the square of the distances δ_i's

$$A_T(\mathcal{X}) = \frac{1}{T} \sum_{i=1}^{T} \left(\frac{\delta_i}{\frac{1}{T}} \right)^2 = T \sum_{i=1}^{T} \delta_i^2 \; .$$

There are three particular values of this constant. When the sequence gives rise to a regular polygon with T vertices, the Arnold constant equals 1 and attains its minimum possible value. More generally, the value of A is close to 1 when the geometric distances δ_i between consecutive elements are close to each other. The maximum value of A_T is obtained in the degenerate case when the sequence $\mathcal{X}_T$ assumes only one value; in this case, one has $A_T = T \cdot 1 = T$. More generally, the

value of A_T is close to T when all the geometric distances between consecutive elements are small except one which is then close to T.

On the other hand, a random choice of T independent uniformly distributed points on the unit torus leads to what Arnold calls the "freedom-liking" value,

$$A_T^* = 2T/(T+1), \qquad A_T^* \to 2 \quad \text{for } T \to \infty.$$

From these observations, it can be inferred that, the value of $A_T(\mathcal{X})$ measures some kind of degree of randomness for the sequence $\mathcal{X}$: if A_T is "much smaller" than A_T^*, this means "mutual repulsion" of points, while if A_T is "much larger" than A_T^*, this means "mutual attraction". On the opposite side, from these two extremal types of non-randomness, the fact that A is "close" to A^* can be considered as a sign of randomness.

2.2 The Particular Case of the Kronecker Sequence

The Kronecker sequence $\mathcal{S}(\alpha)$ associates to $\alpha \in \mathcal{I}$ the fractional parts of the multiples of α,

$$\mathcal{S}(\alpha) := \{\{n\alpha\}; n \in \mathbb{N}\}\,.$$

Here, $\{t\}$ denotes the fractional part of t, namely $\{t\} = t - \lfloor t \rfloor$, where $\lfloor t \rfloor$ denotes the integer part. This sequence satisfies a crucial property which explains its interest: the three distance phenomenon. For any pair (T, α), the truncated sequence $\mathcal{S}_T(\alpha)$ possesses only two or three distinct distances. Both the characterisation of pairs (T, α) for which there exist only two distances, and the values of the distances themselves depend on three main parameters which intervene in the continued fraction expansion of the real α, namely

(a) the quotients m_k,
(b) the denominators q_k of the k-th approximant p_k/q_k of α named continuants,
(c) the distances $\eta_k := |\alpha - (p_k/q_k)|$ between α and its k-th approximant, or more precisely the differences $\theta_k := q_{k-1}\eta_{k-1} = |q_{k-1}\alpha - p_{k-1}|$.

The behaviour of the randomness measures D_T, A_T depends on the "type" of the integer T. We focus on two types of integers T which give rise to the two-distance phenomenon.

(i) the first type when T is of continuant type, i.e T belongs to

$$\mathcal{Q}(\alpha) = \bigcup_{k\geq 0} \mathcal{Q}_k(\alpha) \quad \text{with} \quad \mathcal{Q}_k(\alpha) := \{q_k, q_k + q_{k-1}\},$$

(ii) the case when T is a general two–distance integer, i.e., T belongs to

$$\mathcal{D}(\alpha) = \bigcup_{k\geq 0} \mathcal{D}_k(\alpha) \quad \text{with} \quad \mathcal{D}_k(\alpha) := \{T = m \cdot q_k + q_{k-1};\ 1 \leq m \leq m_{k+1}\}\,.$$

The equality $q_{k+1} = m_{k+1}q_k + q_{k-1}$ entails the inclusion $\mathcal{Q}_{k+1}(\alpha) \subset \mathcal{D}_k(\alpha)$.

In the case where the pair (T, α) gives rise to the two distance phenomenon, the expressions for the Arnold measure and discrepancy (provided later in (3, 4)) are written as a sum of monomials of the form $R_k := m_{k+1}^e q_{k-1}^a q_k^b \theta_k^c \theta_{k+1}^d$. Since the random variables m_k, q_j, θ_ℓ are correlated, it is not easy a priori to study the expectation of such a monomial.

3 Main Results

In this section, we first introduce the two probabilistic models, for the real case and the rational case. Then, we state our main results. Theorems 1 and 2 deal with the discrepancy and the Arnold measure, and Theorem 3 deals with general random variables of the form $q_k^b\,\theta_k^c$.

3.1 Probabilistic Models

There are two different probabilistic models.

Real model. The real α is uniformly chosen in the unit interval $\mathcal{I}$, and the index k tends to ∞. We are interested in the mean values $\mathbb{E}[D_T], \mathbb{E}[A_T]$ for $T \in \mathcal{Q}_k(\alpha)$ or $T \in \mathcal{D}_k(\alpha)$, with $k \to \infty$.

Rational model. Here, we consider the set

$$\Omega = \{(u,v) \in \mathbb{N}^2; \quad 1 \le u < v,\ \gcd(u,v) = 1\},$$

and, for a pair $(u,v) \in \Omega$, the depth $P(u,v)$ denotes the number of steps of the Euclid algorithm on the pair (u,v). We choose here the index k as a function of the depth, and we deal with two main cases: the case where k is a fixed fraction of the depth P of the pair (u,v), namely $k = \lfloor \delta P \rfloor$, for some $\delta \in]0,1[$ fixed, or the more general case when k is a random variable on Ω which is an admissible function of the depth, according to the following definition, already used in [7].

Definition. *A function $F : \mathbb{N} \to \mathbb{N}$ is said to be admissible if there exist two real numbers $a > 0$ and $b < 1$ such that for any integer x, one has $a\,x \le F(x) \le b\,x$.*

A function $K : \Omega \to \mathbb{N}$ is an admissible function of the depth if there exists an admissible function $F : \mathbb{N} \to \mathbb{N}$ for which $K = F \circ P$ where $P : \Omega \to \mathbb{N}$ is the depth function.

For any integer $N > 0$, the subset Ω_N of Ω formed of pairs (u,v) whose denominator v is at most N, is equipped with the uniform probability. We wish to study the asymptotic behaviour of the mean values $\mathbb{E}_N[A_T], \mathbb{E}_N[D_T]$ when α is a random rational of Ω_N, when T belongs to $\mathcal{Q}_k(\alpha)$ or $\mathcal{D}_k(\alpha)$, and k is an admissible function of the depth $P(\alpha)$, and when N tends to ∞.

In Sections 3.2 and 3.3, we show that the two randomness measures share the same behaviour in the real case and in the rational case, for any type of truncation integer T.

3.2 Discrepancy and Arnold Measure for the Continuant Type

Theorem 1 deals with the case when the truncation integer T is of continuant type, i.e. $T = q_k$ or $T = q_k + q_{k-1}$, and proves that the mean values tend to finite values. We then exhibit four constants of interest, and one of them has been already obtained in [7].

Theorem 1. [Discrepancy and Arnold measure for truncation integers of continuant type.] *There are two main cases:*

[Real case.] *When α is a random real of $\mathcal{I}$ and $T \in \mathcal{Q}_k(\alpha)$, the mean values of $\Delta_T(\alpha)$ and $A_T(\alpha)$ are finite, and tend to finite values for $k \to \infty$:*

$$\text{for } T = q_k, \qquad \mathbb{E}[\Delta_T] \sim 1 + \frac{1}{4\log 2} \sim 1.360, \ \mathbb{E}[A_T] \sim \frac{2}{3} + \frac{1}{4\log 2} \sim 1.027,$$

$$\text{for } T = q_k + q_{k-1}, \ \mathbb{E}[\Delta_T] \sim 1 + \frac{1}{2\log 2} \sim 1.721, \ \mathbb{E}[A_T] \sim \frac{2}{3} + \frac{1}{3\log 2} \sim 1.147,$$

with error terms of order $O(\rho^k)$, with $\rho < 1$.

[Rational Case] *When α is a random rational of Ω_N and $T \in \mathcal{Q}_k(\alpha)$, where the index k is an admissible function of the depth $P(\alpha)$, the mean values of $\Delta_T(\alpha)$ and $A_T(\alpha)$, are finite and satisfy, for $N \to \infty$:*

$$\text{for } T = q_k, \qquad \mathbb{E}_N[\Delta_T] \sim 1 + \frac{1}{4\log 2}, \quad \mathbb{E}_N[A_T] \sim \frac{2}{3} + \frac{1}{4\log 2},$$

$$\text{for } T = q_k + q_{k-1}, \qquad \mathbb{E}_N[\Delta_T] \sim 1 + \frac{1}{2\log 2}, \quad \mathbb{E}_N[A_T] \sim \frac{2}{3} + \frac{1}{3\log 2},$$

with error terms of order $O(N^{-\gamma})$, with $\gamma > 0$.

We recall the already known results:

$$\liminf_{T\to\infty} \Delta_T(\alpha) = 1, \qquad \Delta_T(\alpha) \le 3 \quad \text{for } T = q_k(\alpha).$$

The present results show that the asymptotic mean values for the discrepancy, obtained when T is of continuant type, are close to the theoretical infimum. In this case, the Arnold constant is close to 1; following Arnold's interpretation, this is a sign of mutual repulsion of points of the sequence. We conclude from these two facts that a random sequence $\mathcal{S}_T(\alpha)$ is "close" to the sequence (j/T) for T of continuant type.

3.3 Discrepancy and Arnold Measure for a General Two–Distance Integer T

When T is a general two-distance integer, we are interested by the "interpolation curve" which describes the "average" behaviour of the Arnold measure and the discrepancy when the truncation integer T is of the form

$$T = m \cdot q_k + q_{k-1} \text{ with } m = \mu m_{k+1} \text{ and } \mu \in]0, 1[\text{ fixed.}$$

In this case, the integer T does not belong to $\mathcal{Q}_k(\alpha)$, which corresponds to the case $m = 1$ ($\mu = 0$) or $m = m_{k+1}$ ($\mu = 1$). Theorem 2 shows that the mean value is infinite in the real case. On rationals whose denominators are at most N, the mean value is of order $\Theta(\log N)$, and the constant in the dominant term explains the dependence with respect to μ.

Theorem 2. [Discrepancy and Arnold measure for a general two–distance integer] *There are two main cases:*

[Real case.] *When α is a random real of $\mathcal{I}$, when T of the form $T = m \cdot q_k + q_{k-1}$ with $m = \mu m_{k+1}$, $\mu \in]0,1[$ fixed, the mean values of $\Delta_T(\alpha)$ and $A_T(\alpha)$ are infinite.*

[Rational Case] *When α is a random rational of Ω_N, when T is of the form $T = m \cdot q_k + q_{k-1}$ with $m = \mu m_{k+1}$, $\mu \in]0,1[$ fixed, and k an admissible function of depth P, the mean values of $\Delta_T(\alpha)$ and $A_T(\alpha)$ satisfy, for $N \to \infty$,*

$$\mathbb{E}_N[\Delta_T] \sim \frac{\mu(1-\mu)}{2\log 2} \log N, \qquad \mathbb{E}_N[A_T] \sim \frac{\mu(1-\mu)^2}{2\log 2} \log N$$

with error terms of order $O(1/\log N)$.

Our result proves that a truncated Kronecker sequence $\mathcal{S}_T(\alpha)$ does not possess good randomness properties, when its truncation integer is a general two-distance integer. Moreover, in the rational case, our result is more precise, and shows that the mean value of the discrepancy is maximal for a truncation integer T relative to a quotient m close to $(1/2)\, m_{k+1}$ whereas the mean value of the Arnold constant is maximal for a truncation integer T relative to a quotient m close to $(1/3)\, m_{k+1}$. Both asymptotic values are of order $\Theta(\log N)$. And, for most of the admissible truncation integers T, one has $\log T = \Theta(\log N)$. In view of the results of Schmidt and Behnke, and in the case of a general two–distance integer, it would be interesting to determine if the mean values of Δ_T and A_T are of order $\Theta(\log T)$.

3.4 General Study of Random Variables $q_k^b\, \theta_k^c$

It may be of general interest to perform a probabilistic study of the main variables q_k and θ_k, first in a separate way, as it is already done in [16] and [26]. Here, we are interested in a product of the form $q_k^b \theta_k^c$ which involves both variables, which are not independent. Its asymptotic mean values, both in the real and rational case, involves the dominant eigenvalue of the transfer operator $\mathbf{H}_s$ (s a complex parameter) associated to the Euclid dynamical system, defined in (5), when it acts on the space of $\mathcal{C}^1$ functions. This dominant eigenvalue $\lambda(s)$ plays an important role in the following result, which exhibits a strong parallelism between the real and rational cases.

Theorem 3. [Parameters $q_k^b\, \theta_k^c$] *Denote by $\lambda(s)$ the dominant eigenvalue of the transfer operator $\mathbf{H}_s$ defined in (5). There are two main cases:*

[Real case] *(i) For any pair (b,c) with $c > b-1$, the mean value of the product $q_k^b \theta_k^c$ is finite, and satisfies*

$$\mathbb{E}[q_k^b \theta_k^c] = A(b,c)\, \lambda^k\, (1 + (c-b)/2)\, \left[1 + O\left(\rho(b,c)^k\right)\right] \qquad [k \to \infty]\ ,$$

for some positive constants $A(b,c)$, and $\rho(b,c) < 1$. Then, for $c > b$, the mean value tends to ∞, and, for $b - 1 < c < b$, the mean value tends to ∞.

(ii) In the particular case $c = b$, the mean value of the product $q_k^b \theta_k^c$ tends to a constant $A(b,b)$ for $k \to \infty$

(iii) If $c < b - 1$, the mean value $\mathbb{E}[q_k^b \theta_k^c]$ is infinite for any integer k.

[Rational case.] *For any $\delta \in]0,1[$, and any real a, denote by $\sigma(a,\delta)$ the unique real σ solution of the equation*

$$\lambda^{1-\delta}(\sigma) \cdot \lambda^{\delta}(\sigma + a/2) = 1, \qquad \text{with} \quad \sigma(0,\delta) = 1.$$

(i) For any triple (δ, b, c), the mean value of the product $q_k^b \theta_k^c$ on Ω_N, when $k = \lfloor \delta P \rfloor$ is a fraction of the depth P, satisfies

$$\mathbb{E}_N[q_k^b \theta_k^c] \sim A(\delta,b,c)\, N^{2\sigma(c-b,\delta)-2} \left[1 + O(N^{-\gamma(\delta,b,c)})\right] \qquad [N \to \infty],$$

for some positive constants $A(\delta,b,c), \gamma(\delta,b,c)$.

(ii) In the particular case $c = b$, the constant $A(\delta,b,b)$ satisfies $A(\delta,b,b) = A(b,b)$ for any $\delta \in]0,1[$. The mean value of $q_k^b \theta_k^b$ when k is any admissible function of the depth P tends to $A(b,b)$ for $N \to \infty$, the same constant as in the real case.

3.5 Interesting Particular Cases for Random Variables $q_k^b \theta_k^c$

There are three particular cases of interest. The cases $(c = 1, b = 0)$ or $(c = 1, b = -1)$ study the mean value of the k-th approximation of a number α. In the real case, the mean values are of exponential type, with a ratio which involves two possible values of the dominant eigenvalue of the transfer operator $\lambda(3/2)$ or $\lambda(2) \sim 0.1994$. This last value[1] $\lambda(2)$ (discovered in 1994...) plays a central role in the analysis of the Gauss Algorithm [10], and its occurrence in this approximation context was remarked for the first time in [13]. Theorem 3 can also be used as a main step to prove that the random variables $\log q_k$ or $\log \theta_k$ asymptotically follow a gaussian law, both in the real and rational case. (See [16] and [25]).

4 Main Steps for the Proofs

First, with the three distance theorem, we exhibit expressions for the randomness measures in terms of the main parameters m_k, q_k, θ_k of the continued fraction expansion. Then, we describe how to apply the dynamical analysis methodology –a mixing between analysis of algorithms and dynamical systems theory–.

4.1 Euclid Dynamical System and Continued Fractions

The Euclid dynamical system is defined by pair $(\mathcal{I}, V)$ where V is the Gauss map

$$V : \mathcal{I} \to \mathcal{I}, \quad V(x) = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor = \left\{ \frac{1}{x} \right\} \quad \text{for } x \neq 0\ , \quad V(0) = 0,$$

and $\lfloor \cdot \rfloor$ denotes the integer part, and $\{\cdot\}$ denotes the fractional part.

[1] Flajolet called it the Vallée constant...

The restriction of V to the interval $\mathcal{I}_m := [1/(m+1), 1/m]$ is the mapping $V_{[m]} : \mathcal{I}_m \to \mathcal{I}$ defined by $V_{[m]}(x) = (1/x) - m$ whose inverse mapping $h_{[m]} : \mathcal{I} \to \mathcal{I}_m$ is defined by $h_{[m]}(x) = 1/(m+x)$. Denote by $\mathcal{H}$ the set of all inverse mappings.

The trajectory of the real x is the sequence $(x, V(x), V^2(x), \ldots, V^k(x), \ldots)$. It reaches 0 if and only if x is rational. For a rational $x = u/v$, the first index k for which $V^k(x) = 0$ is called the depth of x. This is the number of iterations of the Euclid Algorithm on the pair (u, v), denoted previously by $P(u, v)$. It will be also denoted by $P(u/v)$. The sequence of the digits is defined as

$$(m_1(x), m_2(x), \ldots, m_k(x), \ldots) \quad \text{where} \quad m(x) := \left\lfloor \frac{1}{x} \right\rfloor, \quad m_{k+1}(x) = m(V^k(x)),$$

and x admits a continued fraction expansion (CFE) of the form

$$x = \cfrac{1}{m_1 + \cfrac{1}{m_2 + \cfrac{1}{\ddots + \cfrac{1}{m_k + \cfrac{1}{\ddots}}}}} = [m_1, m_2, \ldots, m_k, \ldots].$$

In any case, a truncation of the continued fraction expansion at depth $k \le P(x)$ produces two continued fraction expansions: the beginning part $[m_1, m_2, \ldots, m_k]$ and the ending part $[m_{k+1}, m_{k+2}, \ldots, m_{k+\ell}, \ldots]$.

The beginning part defines the linear fractional transformation,

$$g_k := h_{[m_1]} \circ h_{[m_2]} \circ \ldots \circ h_{[m_k]} \qquad \text{with} \quad g_k(y) = \frac{p_{k-1}y + p_k}{q_{k-1}y + q_k}$$

together with the k–th approximant of x, namely the rational $p_k/q_k = g_k(0)$. The ending part defines the real $x_k := V^k(x) = [m_{k+1}, m_{k+2}, \ldots]$ via the equality

$$x = g_k(x_k), \qquad \text{or} \quad x_k = \frac{\theta_{k+1}(x)}{\theta_k(x)} \qquad \text{with} \quad \theta_k(x) := |q_{k-1}x - p_{k-1}| \ .$$

Both the continuant q_k and the distance θ_k are expressed with the derivative g'_k,

$$q_k^{-2} = |g'_k(0)|, \qquad \theta_k^2 = |g'_k(x_k)| \ . \tag{2}$$

In the rational case, for $x = u/v$, with coprime integers (u, v), the equality $\theta_k(u/v) = v_k/v$ holds, and involves the sequence v_k of remainders defined by the execution of the Euclid algorithm on the pair (u, v).

4.2 Expressions of Randomness Measures for the Kronecker Sequence

The three distances theorem was conjectured by Steinhaus proved by Surányi [20], Sós [19] and Świerczkowski [21]. Its precise statement is as follows:

Theorem A. [Three distances theorem] *Let α be a real of the unit interval. Consider an integer $T < v$ if α is a rational of the form u/v relative to a pair (u, v) of coprime integers. Then, the truncated Kronecker sequence $\mathcal{S}_T(\alpha) := \{\{k\alpha\}; \quad k \in [0..T-1]\}$ has the three distance property: there are at most three possible values for the distance between geometrically consecutive points.*

(i) Consider the two sequences (q_k) and (θ_k) associated to the real α together with the sequence (m_k) of the quotients, and write the integer $T \geq 0$ under the form

$$T = m \cdot q_k + q_{k-1} + r \qquad \text{with } 1 \leq m \leq m_{k+1} \text{ and } 0 \leq r < q_k .$$

The three possible distances are θ_{k+1}, $\theta_k - m\theta_{k+1}$ or $\theta_k - (m-1)\theta_{k+1}$. Moreover, there are $T - q_k$ such distances equal to θ_{k+1}, r distances equal to $\theta_k - m\theta_{k+1}$ and $q_k - r$ distances equal to $\theta_k - (m-1)\theta_{k+1}$.

(ii) There are only two distances if and only if the integer T is associated to a "remainder" $r = 0$. They are θ_{k+1} and $\theta_k - (m-1)\theta_{k+1}$.

Then, the truncated Kronecker sequence $\mathcal{S}_T(\alpha)$ is a special sequence, where the main "distances" $[\delta_j, \gamma_j^{\pm}]$ can be computed in an explicit way as a function of the three main parameters m_k, q_k, θ_k. This is clear for the distances δ_i which intervene in the Arnold measure (the precise expression can be found in [7]), but this is also true for the distances γ_j^- and γ_j^+ which intervene in the discrepancy at least when the pair (T, α) leads to the two distance situation (the precise expression is in [18]). This is why we focus here in the two distances situation.

Theorem B. *Let α be a real of the interval $\mathcal{I}$. Consider an integer $T < v$ if α is a rational of the form u/v with coprime u, v. Consider the two sequences (q_k) and (θ_k) associated to the real α together with the sequence (m_k) of the quotients, and a two distance integer $T \in \mathcal{D}(\alpha)$ of the form $T = m \cdot q_k + q_{k-1}$ with $1 \leq m \leq m_{k+1}$. The Arnold measure $A_T(\alpha)$ of the sequence $\mathcal{S}(\alpha)$ equals*

$$A_T(\alpha) = (mq_k + q_{k-1}) \left[((m-1)q_k + q_{k-1})\theta_{k+1}^2 + q_k(\theta_k - (m-1)\theta_{k+1})^2 \right] . \quad (3)$$

The discrepancies $\Delta_T(\alpha), D_T(\alpha)$ of the sequence $\mathcal{S}(\alpha)$ satisfy

$$\Delta_T(\alpha) = T \cdot D_T(\alpha) = 1 + (mq_k + q_{k-1} - 1)(\theta_k - m\theta_{k+1}),$$
$$\Delta_T(\alpha) \sim 1 + (mq_k + q_{k-1})(\theta_k - m\theta_{k+1}) . \quad (4)$$

4.3 Various Types of Monomials

The expressions $(3, 4)$ of the discrepancy and Arnold measure are written as a sum of monomials of the form

$$R_k := m_{k+1}^e q_{k-1}^a q_k^b \theta_k^c \theta_{k+1}^d \qquad \text{with } a,\ b,\ c,\ d \text{ and } e \in [0..3].$$

We are interested in two particular cases:

(i) For Theorem 3, one has $e = a = d = 0$. The cost is said of type (T3).

(ii) For Theorems 1 and 2, the costs are homogeneous, and the equalities hold

$$a + b = c + d = 1 \quad \text{[Discrepancy]}, \qquad a + b = c + d = 2 \quad \text{[Arnold measure]} .$$

We then let $f := a + b = c + d$, and the new parameters are a, d, e, f. The monomial is said of type (T1/2).

4.4 Various Strategies for the Analyses

Our general strategy depends on the probabilistic setting.
Real case. We study directly the mean value of the cost R_k, equal to the integral

$$\mathbb{E}[R_k] = \int_{\mathcal{I}} R_k(x)\mathrm{d}x \ .$$

Rational case. When the index k depends on the depth $P(u,v)$ via an admissible function F, this random variable only depends on u/v, and we denote it by $R_{<F>}$ or simply by R. We here perform an indirect study, typical in Analytic Combinatorics, and we introduce the Dirichlet series

$$S_R(s) := \sum_{(u,v)\in\Omega} \frac{R(u,v)}{v^{2s}} = \sum_{n\geq 1} \frac{a_n}{n^{2s}}, \qquad \text{with} \quad a_n := \sum_{(u,n)\in\Omega} R(u,n).$$

Then, the expectation $\mathbb{E}_N[R_{<F>}]$ involves partial sums of coefficients a_n,

$$\mathbb{E}_N[R_{<F>}] = \frac{\Phi(N)}{\Phi_0(N)}, \qquad \text{with} \quad \Phi(N) := \sum_{n\leq N} a_n, \qquad \Phi_0(N) = |\Omega_N|.$$

We then transfer analytic properties of the Dirichlet series into asymptotic properties of the coefficients.

4.5 Generating Operators

We obtain alternative expressions of the integral $\mathbb{E}[R_k]$ or the Dirichlet series $S_R(s)$ as a function of convenient transfer operators, first introduced by Ruelle [17]. The (plain) transfer operator $\mathbf{H}_s$ of the Euclidean dynamical system involves the set $\mathcal{H}$ of the inverse mappings of the mapping V, under the form

$$\mathbf{H}_s[f](x) := \sum_{h\in\mathcal{H}} |h'(x)|^s \cdot f\circ h(x) = \sum_{m\geq 1} \frac{1}{(m+x)^{2s}} \cdot f\left(\frac{1}{m+x}\right). \qquad (5)$$

Due to the expressions given in (2), it can be used to generate continuants q_k, v_k, distances θ_k or digits m_{k+1}. However, for generating products which involve all these variables together, as it is the case in the monomials R_k, it is necessary to deal with the following three extensions described in Figure 1.

$$\begin{aligned}
&\mathbf{H}_{(s,\cdot)}[F](x,y) = \textstyle\sum_{h\in\mathcal{H}} |h'(x)|^s \cdot F(h(x),y),\\
&\mathbf{H}_{(s,\cdot,t)}[F](x,y) = \textstyle\sum_{h\in\mathcal{H}} |h'(x)|^s\, |h'(0)|^t \cdot F(h(x),y),\\
&\mathbf{H}_{(s,t)}[F](x,y) = \textstyle\sum_{h\in\mathcal{H}} |h'(x)|^s |h'(y)|^t \cdot F(h(x),h(y)) \ .
\end{aligned}$$

Fig. 1. The three extensions of the transfer operator

Then, using these various extensions, Figure 2 defines the transfer operator $\mathbf{R}_s^{[k]}$ which is used in each theorem: Theorems 1 and 2 (T1/2), or Theorem 3 (T3).

$$\text{(T3):} \quad \mathbf{H}^k_{(s+c/2,-b/2)}$$

$$\text{(T1/2):} \quad \mathbf{H}_{(s+(d-a)/2,\cdot,-e/2)} \circ \left(\sum_{j=0}^{a} \binom{a}{j} (-1)^{a-j} \mathbf{H}^k_{(s+(f-j)/2,-(f-j)/2)\cdot} \right)$$

Fig. 2. Definition of the transfer operator $\mathbf{R}_s^{[k]}$ used in the study of the monomial $R_k := m_{k+1}^e q_{k-1}^a q_k^b \theta_k^c \theta_{k+1}^d$. In the (T1/2) case, one lets: $f = a + b = c + d$. In the (T3) case, one has $a = d = e = 0$.

Finally, the following proposition holds:

Proposition 1. *The study of the cost* $R_k := m_{k+1}^e q_{k-1}^a q_k^b \theta_k^c \theta_{k+1}^d$ *involves the transfer operator described in Figure 2, together with extensions of transfer operators described in Figure 1.*

$$\textit{In the real case} \qquad \mathbb{E}[R_k] = \textstyle\int_{\mathcal{I}} \mathbf{R}_1^{[k]}[\mathbf{1}](u,0)du$$

$$\textit{In the rational case} \quad S_R(s) = \sum_{p \geq 1} \mathbf{H}^{p-F(p)-1}_{(s,\cdot)} \circ \mathbf{R}_s^{[F(p)]}[\mathbf{1}](u,0) \ .$$

4.6 Main Principles for Dynamical Analysis

We then proceed according to the general dynamical analysis methodology, described for instance in [27]. Our method depends on the probabilistic setting.

In the *real* case, spectral properties of the transfer operator (on the space $\mathcal{C}^1(\mathcal{I})$ lead to asymptotic estimates for the mean value $\mathbb{E}[R_k]$ in terms of dominant eigenvalues and eigenfunctions of transfer operator.

In the *rational* case, with the alternative forms of the Dirichlet series $S_R(s)$ given in Proposition 1, we study the precise behaviour of $S_R(s)$, when s belongs to a vertical strip near $s = 1$ (in terms of analyticity and polynomial growth with respect to $\Im s$), in the same vein as in [12]. With the Perron formula of order two [24] applied to series $S_R(s)$, on a vertical line $\Re s = D > 0$ inside the domain of convergence,

$$\Psi(U) := \sum_{p \leq U} \Phi(p) = \frac{1}{2i\pi} \int_{D-i\infty}^{D+i\infty} S_R(s) \frac{U^{2s+1}}{s(2s+1)} \, \mathrm{d}s,$$

we then obtain estimates on $\Psi(U)$ that we transfer into estimates on $\Phi(p)$, as in [3] and [6]. We then obtain the estimates of Theorems 1, 2, 3.

Conclusions and Open Problems. To the best of our knowledge, this is the first study which adopts a probabilistic point of view on randomness measures for the Kronecker sequence. This study may be extended in three main directions:

(a) The general case where there are three distances, at least in the case of the Arnold sequence. The computations are heavier, but the study is of the same vein as the present study.
(b) Theorem 2 proves that the randomness measures are not good for a general two–distance integer, and the reason is quite simple: this is due to the fact that the quotient m_{k+1} may be large, and it is well-known that the mean value $\mathbb{E}[m_{k+1}]$ is infinite in the real case and of logarithmic order in the rational case (see [27] for the rational case). Then, it would be of great interest to restrict this probabilistic study to "inputs" α for which the sequence of quotients m_k which appear in the continued fraction expansion satisfies one of the following properties :

($b1$) it is bounded – ($b2$) it admits bounded averages.

Previous works of the authors deal with these "constrained" probabilistic models, both in the real case and in the rational case (see [8],[9], and [26]), and use the dynamical analysis methodology. It seems possible to extend these works in order to obtain, in this "restricted" framework, an analog of Theorem 2 which would exhibit finite mean values.
(c) In view of the results of Schmidt and Behnke, and in the study of the real case and a general two–distance integer, it would be interesting to determine if the mean values of Δ_T and A_T are of order $\Theta(\log T)$.

References

1. Arnold, V. I.: Arnold's problems. Springer Phasis (2004)
2. Arnold, V.I.: Topology and statistics of formulae of arithmetics. Russian Math. Surveys 58, 637–664 (2003)
3. Baladi, V., Vallée, B.: Euclidean algorithms are Gaussian. J. Number Theory 110, 331–386 (2005)
4. Beck, J.: Inevitable Randomness in Discrete Mathematics. University Lecture Series, vol. 49. American Mathematical Society, Providence (2009)
5. Behnke, H.: Theorie der Diophantischen Approximationen. Hamb. Abh. 3, 261–318 (1924)
6. Cesaratto, E., Clément, J., Daireaux, B., Lhote, L., Maume-Deschamps, V., Vallée, B.: Regularity of the Euclid Algorithm: application to the analysis of fast gcd Algorithms. Journal of Symbolic Computation 44, 726–767 (2009)
7. Cesaratto, E., Plagne A., Vallée, B.: On the non-randomness of modular arithmetic progressions: a solution to a problem of V. I. Arnold. In: Proceedings of the 4th Colloquium on Mathematics and Computer Science: Algorithms, Trees, Combinatorics and Probability. Discrete Mathematics and Theoretical Computer Science, vol. AG, pp 271-288. DMTCS, Nancy (2006)
8. Cesaratto, E., Vallée, B.: Hausdorff dimension of real numbers with bounded digit averages. Acta Arith. 125, 115–162 (2006)
9. Cesaratto, E., Vallée, B.: Small quotients in Euclidean Algorithms. Ramanujan Journal 24, 183–218 (2011)

10. Daudé, H., Flajolet, P., Vallée, B.: An average-case analysis of the Gaussian algorithm for lattice Reduction. Combinatorics, Probability and Computing 6, 397–433 (1997)
11. Drmota, M., Tichy, R.: Sequences, Discrepancies and Applications. Lecture Notes in Mathematics, vol. 1651. Springer, Berlin (1997)
12. Dolgopyat, D.: On decay of correlations in Anosov flows. Ann. of Math. 147(2), 357–390 (1998)
13. Flajolet, P., Vallée, B.: Continued fraction algorithms, functional operators, and structure constants. Theoretical Computer Science 194(1-2), 1–34 (1998)
14. Kuipers, L., Niederreiter, H.: Uniform distribution of sequences. John Wiley and Sons, New York (1974)
15. Knuth, D.E.: The art of Computer Programming, 3rd edn., vol. 2. Addison Wesley (1998)
16. Lhote, L., Vallée, B.: Gaussian laws for the main parameters of the Euclid Algorithms. Algorithmica 50, 497–554 (2008)
17. Ruelle, D.: Thermodynamic formalism. Addison Wesley (1978)
18. Van Ravenstein, T.: On the discrepancy of the sequence formed from multiples of an irrational number. Bull. Austral. Math. Soc. 31, 329–338 (1985)
19. Sós, V.T.: On the distribution mod 1 of the sequence $n\alpha$. Ann. Univ. Sci. Budapest Eötvös Sect. Math. 1, 127–134 (1958)
20. Surányi, J.: Über die Anordnung der Vielfachen einer reellen Zahl mod 1. Ann. Univ. Sci. Budapest Eötvös Sect. Math. 1, 107–111 (1958)
21. Świerczkowski, S.: On successive settings of an arc on the circumference of a circle. Fund. Math. 46, 187–189 (1959)
22. Schmidt, W.M.: Irregularities of distribution VII. Acta Arith. 21, 45–50 (1972)
23. Schoissengeier, J.: On the discrepancy of $(n\alpha)$. Acta Arith. 44, 241–279 (1984)
24. Tenenbaum, G.: Introduction à la théorie analytique et probabiliste des nombres. Cours Spécialisés 1, SMF (1995)
25. Vallée, B.: Opérateurs de Ruelle-Mayer généralisés et analyse en moyenne des algorithmes de Gauss et d'Euclide. Acta Arith. 81, 101–144 (1997)
26. Vallée, B.: Dynamique des fractions continues à contraintes périodiques. Journal of Number Theory 72, 183–235 (1998)
27. Vallée, B.: Euclidean dynamics. Discrete Contin. Dyn. Syst. 15, 281–352 (2006)