Chapter 4

Pseudorandom generator

In this chapter, we discuss multipurpose pseudorandom generators. Those exclusively for the Monte Carlo integration have been discussed in § 2.5 and will be discussed further in Chapter 5.

4.1 Computationally secure pseudorandom generator

4.1.1 Definitions

Let us introduce the definition of computationally secure pseudorandom generator as well as related notions. Basic ideas can be seen in [2, 49]. For details, see [24, 36].

Definition 4.1

1. This chapter mainly deals with partial recursive functions $f : \{0,1\}^* \to \{0,1\}^*$ of the following form; for each $n \in \mathbb{N}^+$, $f_n := f|_{\{0,1\}^{r(n)}} : \{0,1\}^{r(n)} \to \{0,1\}^{s(n)}$. We then write $f = \{f_n\}_n$. Let M be a Turing machine which computes f. The *time complexity* $T_f(n)$ of f is defined as the maximum number of steps that M needs to compute $f_n(x)$ where x runs over $\{0,1\}^{r(n)}$. This definition applies to functions of several variables as well.

2. A sequence of integers $\{\ell(n)\}_n$ is called a *polynomial parameter* if there exists a constant $c > 0$ such that $\ell(n) = O(n^c)$.

3. $f = \{f_n\}_n$ with $f_n : \{0,1\}^{r(n)} \to \{0,1\}^{s(n)}$ is called a *polynomial time function* if $r(n)$, $s(n)$ and $T_f(n)$ are polynomial parameters. This definition applies to functions of several variables as well.

4. When a random variable Y is distributed uniformly in a finite set B, we write $Y \in_U B$. We assume that Y is independent of all other random variables in the context. The probability measure that governs Y is often written as $\Pr_Y$.

5. $A = \{A_n\}_n$ is called a *random function* if A is of the form $A_n : \{0,1\}^{r(n)} \times \{0,1\}^{s(n)} \to \{0,1\}^{t(n)}$ with inputs $x \in \{0,1\}^{r(n)}$ and $Y \in_U \{0,1\}^{s(n)}$. We often omit the random variable in the notation and say simply "random function $A_n : \{0,1\}^{r(n)} \to \{0,1\}^{t(n)}$". But the time complexity of A is the one for two variable function $A_n(x,y)$. These definitions and notions apply to functions of several variables as well.

In the theory of computational complexity, polynomial time functions are thought to be *feasible* functions, while the other functions are thought to be *infeasible* functions.[†1]

Definition 4.2 A polynomial time function $g = \{g_n\}_n$ is called a *pseudorandom generator* if it is of the form $g_n : \{0,1\}^n \to \{0,1\}^{\ell(n)}$ where $\ell(n) > n$.

Remark 4.3 In Definition 2.5 of Chapter 2, a pseudorandom generator is defined as a single function $g : \{0,1\}^n \to \{0,1\}^L$ where $n < L$. This definition suffices, if we use a pseudorandom generator to solve a particular single problem. But if we use a pseudorandom generator for many purposes, we should define it as a sequence of functions as Definition 4.2, which enables us to choose a suitable $g_n : \{0,1\}^n \to \{0,1\}^{\ell(n)}$ for each individual problem.

The seed which Alice, the player, chooses is regarded as a random variable $Z_n \in_U \{0,1\}^n$. The function g_n stretches it to a pseudorandom number $g_n(Z_n)$, which is a $\{0,1\}^{\ell(n)}$-valued random variable. Of course, $\ell(n) > n$ implies that $g_n(Z_n)$ is not distributed uniformly in $\{0,1\}^{\ell(n)}$.

We next consider functions for tests. Let $A = \{A_n\}_n$, $A_n : \{0,1\}^{\ell(n)} \to \{0,1\}$, be a function or a random function, and set[†2]

$$\delta_{g,A}(n) := \left|\Pr_{Z_{\ell(n)}}(A_n(Z_{\ell(n)}) = 1) - \Pr_{Z_n}(A_n(g_n(Z_n)) = 1)\right|. \tag{4.1}$$

To be a good pseudorandom generator, it is desirable that $\ell(n)$ is much greater than n, the calculation of g_n is quickly done, and that $\delta_{g,A}(n)$ is small enough for many A's. But it is not necessary that $\delta_{g,A}(n)$ is small for *all* A's. We set

$$S_{g,A}(n) := \frac{T_A(n)}{\delta_{g,A}(n)}.$$

Definition 4.4 A pseudorandom generator g is said to be *computationally secure*[†3] if $S_{g,A}(n)$ is not a polynomial parameter for all A's.

If $T_A(n)$ is not a polynomial parameter, neither is $S_{g,A}(n)$. So Definition 4.4 does not care about such A. Thus a computationally secure pseudorandom generator is thought to be one which cannot be rejected by any feasible tests.

4.1.2 Computational security and Monte Carlo method

Pseudorandom numbers are used not only in the Monte Carlo method but also in cryptography. The usage of them in cryptography is as follows. A message, such as document, sound, picture, whatever it may be, is transformed into a finite $\{0,1\}$-sequence x. Generating a pseudorandom number ($\{0,1\}$-sequence) y of the same length as x, we encode x

[†1] This is not always true in problems of practical size. For example, $e^{n/1000}$ is greater than every polynomials in n for sufficiently large n, but for small n, it is less than n^{100}.

[†2] If A is a random function, and if A_n involves a random variable $Y \in_U \{0,1\}^{s(n)}$, the calculation of the probability in (4.1) must take Y into account.

[†3] In computer science, or more precisely, in cryptography, a computationally secure pseudorandom generator is simply called a pseudorandom generator.

by taking the bit-wise XOR (eXclusive OR) of them; i.e., $z := x$ XOR y, which is a coded message. To decode it, do just the same thing; $x = z$ XOR y. In this case, the seed of the pseudorandom number is the common key (or password) for encoding and decoding. If the pseudorandom generator is computationally secure, no one can decode the message z in practice without knowing the key (cf. [24, 36]).

In Definition 4.4, the function A stands for an '*a*dversary' who attacks the cryptosystem using every means available. We admit a random function for A, because adversaries may attack it at random. A computationally secure pseudorandom generator stands up to every feasible attack, and hence it is also called a *cryptographically secure* pseudorandom generator.

Thus the notion of computational security was thought out from a viewpoint other than the Monte Carlo method. But, as a matter of fact, it is useful in the Monte Carlo method, too. Let us explain it.

Suppose that our random variable S is a function of $Z_{\ell(n)} \in_U \{0,1\}^{\ell(n)}$; $S := S(Z_{\ell(n)})$. Note that S should be a function which can be computed in practice. Suppose further that $\ell(n)$ is too large to sample $S(Z_{\ell(n)})$, so we use a pseudorandom number $g_n(Z_n)$, $Z_n \in_U \{0,1\}^n$, instead of $Z_{\ell(n)}$. Namely, we compute $S' := S(g_n(Z_n))$ instead of S. Then we ask if the distribution of S' is close to that of S. Let us compare the distribution functions

$$F(S;t) := \Pr_{Z_{\ell(n)}}(S \le t), \quad F(S';t) := \Pr_{Z_n}(S' \le t), \quad t \in \mathbb{R},$$

of the two. If g is computationally secure, it is assured that $F(S;t)$ and $F(S';t)$ are close to each other. To see this, set a function for test A_n as

$$A_n(x) := \mathbf{1}_{\{S(x)\le t\}}, \quad x \in \{0,1\}^{\ell(n)}.$$

Because S can be computed in practice, the time complexity of A_n is sufficiently small. Then by the definition of computationally security, the difference

$$|F(S;t) - F(S';t)| = \left|\Pr_{Z_{\ell(n)}}(A_n(Z_{\ell(n)}) = 1) - \Pr_{Z_n}(A_n(g_n(Z_n)) = 1)\right|$$

must be very small.

4.1.3 Existence problem

The notion of computationally secure pseudorandom generator is very natural and very simple. However, unfortunately, we do not know if its instance exists.

Let us introduce two computational complexity classes;

$$\mathbf{P} := \left\{ L \subset \{0,1\}^* \;\middle|\; \begin{array}{l} \exists A : \{0,1\}^* \to \{0,1\}, \text{ a polynomial time function, s.t.} \\ \forall x \in \{0,1\}^* \, (x \in L \iff A(x) = 1) \end{array} \right\},$$

$$\mathbf{NP} := \left\{ L \subset \{0,1\}^* \;\middle|\; \begin{array}{l} \exists A : \{0,1\}^* \to \{0,1\}, \text{ a random polynomial time function, s.t.} \\ \forall x \in \{0,1\}^* \, (x \in L \iff \Pr(A(x) = 1) > 0) \end{array} \right\}.$$

$\mathbf{P} \subset \mathbf{NP}$ is obvious, but the inverse inclusion relation is not known. Most of researchers believe that $\mathbf{P} \neq \mathbf{NP}$ holds. This is one of the most important conjectures in the theory of computational complexity.

Concerning pseudorandom generator, we have the following theorem.

Theorem 4.5 *If* **P** = **NP** *holds, there exists no computationally secure pseudorandom generator.*[†4]

Proof. Let $g = \{g_n\}_n$, $g_n : \{0,1\}^n \to \{0,1\}^{\ell(n)}$, be an arbitrary pseudorandom generator. Define $M_n : \{0,1\}^{\ell(n)} \times \{0,1\}^n \to \{0,1\}$ by

$$M_n(y,x) := \begin{cases} 1 & (g_n(x) = y), \\ 0 & (g_n(x) \neq y). \end{cases}$$

Then $M = \{M_n\}_n$ is a polynomial time function. If we set

$$L := \{y \in \{0,1\}^* \mid \exists\, n \in \mathbb{N}^+,\ y \in \{0,1\}^{\ell(n)},\ \exists\, x \in \{0,1\}^n,\ M_n(y,x) = 1\},$$

then $L \in$ **NP**. Since we assume **P** = **NP**, we have $L \in$ **P**. This means that there exists a polynomial time function $A = \{A_n\}_n$, $A_n : \{0,1\}^{\ell(n)} \to \{0,1\}$, such that $y \in L \Leftrightarrow A_n(y) = 1$. Then for $Z_n \in_U \{0,1\}^n$ and $Z_{\ell(n)} \in_U \{0,1\}^{\ell(n)}$, we have

$$\Pr_{Z_n}(A_n(g_n(Z_n)) = 1) = 1, \quad \Pr_{Z_{\ell(n)}}(A_n(Z_{\ell(n)}) = 1) \leq \frac{2^n}{2^{\ell(n)}},$$

which implies $\delta_{g,A}(n) \geq 1 - 2^{n-\ell(n)}$. Since $T_A(n)$ is a polynomial parameter, so is $S_{g,A}(n) = T_A(n)/\delta_{g,A}(n)$. Thus g is not computationally secure. □

There are many candidates for computationally secure pseudorandom generator. To this point, since we do not know if **P** ≠ **NP** holds, we do not know if they are computationally secure. However researchers are optimistic. They think that if some of them are computationally secure, it would be excellent, and if not, there would be some progress in the **P**≠**NP** conjecture. Anyhow, until we come to know that they are not, they may be regarded as secure.

4.1.4 Next-bit-unpredictability

Let us introduce the following property of pseudorandom generator.

Definition 4.6 Let $g = \{g_n\}_n$, $g_n : \{0,1\}^n \to \{0,1\}^{\ell(n)}$, be a pseudorandom generator. Suppose that random variables $Z \in_U \{0,1\}^n$ and $I \in_U \{1,2,\ldots,\ell(n)\}$ are independent under $\Pr_{I,Z}$. For a function or a random function $\tilde{A} = \{\tilde{A}_n\}_n$, $\tilde{A}_n : \{1,\ldots,\ell(n)\} \times \{0,1\}^{\ell(n)} \to \{0,1\}$, define

$$\tilde{\delta}_{g,\tilde{A}}(n) := \Pr_{I,Z}\left(\tilde{A}_n(I, g_n(Z)_{\{1,\ldots,I-1\}}) = g_n(Z)_I\right) - \frac{1}{2}.$$

Here, $g_n(Z)_i$ denotes the i-th bit of $g_n(Z)$, and $g_n(Z)_{\{1,\ldots,i\}} \in \{0,1\}^{\ell(n)}$ is defined by

$$g_n(Z)_{\{1,\ldots,i\}} := (g_n(Z)_1, g_n(Z)_2, \ldots, g_n(Z)_i, \overbrace{0,\ldots,0}^{\ell(n)-i}).$$

Now, g is said to be *next-bit-unpredictable* if

$$\tilde{S}_{g,\tilde{A}}(n) := \left|\frac{T_{\tilde{A}}(n)}{\tilde{\delta}_{g,\tilde{A}}(n)}\right|$$

is not a polynomial parameter for any $\tilde{A}$.

[†4] **P** ≠ **NP** does not imply the existence of computationally secure pseudorandom generator.

Theorem 4.7 *A pseudorandom generator $g = \{g_n\}_n$ is computationally secure, if and only if it is next-bit-unpredictable.*

Proof. *Step 1.* (1) Let g be next-bit-predictable, i.e., let there exist an $\tilde{A}$ such that $T_{\tilde{A}}(n)$ and $\tilde{S}_{g,\tilde{A}}(n)$ are polynomial parameters. Let $I \in_U \{1, \ldots, \ell(n)\}$ and define a random function $A : \{0,1\}^{\ell(n)} \to \{0,1\}$ as

$$A_n(x) := \begin{cases} 1 & (\tilde{A}_n(I, x_{\{1,\ldots,I-1\}}) = x_I) \\ 0 & (\tilde{A}_n(I, x_{\{1,\ldots,I-1\}}) \neq x_I) \end{cases} \qquad x \in \{0,1\}^{\ell(n)}.$$

Then

$$\begin{aligned} \delta_{g,A}(n) &= \left|\Pr_{Z_{\ell(n)}}\left(A_n(Z_{\ell(n)}) = 1\right) - \Pr_{Z_n}\left(A_n(g_n(Z_n)) = 1\right)\right| \\ &= \left|\frac{1}{2} - \Pr_{I,Z_n}\left(\tilde{A}_n(I, g_n(Z_n)_{\{1,\ldots,I-1\}}) = g_n(Z_n)_I\right)\right| \\ &= \left|\tilde{\delta}_{g,\tilde{A}}(n)\right|. \end{aligned}$$

On the other hand, since $T_A(n)$ is a polynomial parameter, so is $S_{g,A}(n)$. Thus g is not computationally secure.

Step 2. Let g be not computationally secure, i.e., let there exist an A such that $T_A(n)$ and $S_{g,A}(n)$ are polynomial parameters. Let $Y \in_U \{0,1\}^{\ell(n)}$ and $W \in_U \{0,1\}$ be independent. For each $i \in \{1, \ldots, \ell(n)\}$ and each $x \in \{0,1\}^{\ell(n)}$, define

$$\tilde{A}_n(i, x) := \begin{cases} Y_i & (A_n(x_1, \ldots, x_{i-1}, Y_i, \ldots, Y_{\ell(n)}) = 1), \\ W & (A_n(x_1, \ldots, x_{i-1}, Y_i, \ldots, Y_{\ell(n)}) = 0). \end{cases}$$

Now, to see that g is next-bit-predictable, let us show that $\tilde{S}_{g,\tilde{A}}(n)$ is a polynomial parameter. In what follows, we write $X := g_n(Z_n)$, $\Pr := \Pr_{Z_n,Y,W}$. We start with the following calculation.

$$\begin{aligned} &\Pr\left(\tilde{A}_n(i, X_{\{1,\ldots,i-1\}}) = X_i\right) - \frac{1}{2} \\ &\quad= \Pr\left(X_i = Y_i,\, A_n(X_1, \ldots, X_{i-1}, Y_i, \ldots, Y_{\ell(n)\}}) = 1\right) \\ &\qquad\quad + \Pr\left(X_i = W,\, A_n(X_1, \ldots, X_{i-1}, Y_i, \ldots, Y_{\ell(n)}) = 0\right) - \frac{1}{2} \\ &\quad= \Pr\left(X_i = Y_i,\, A_n(X_1, \ldots, X_i, Y_{i+1}, \ldots, Y_{\ell(n)}) = 1\right) \\ &\qquad\quad + \frac{1}{2}\Pr\left(A_n(X_1, \ldots, X_{i-1}, Y_i, \ldots, Y_{\ell(n)}) = 0\right) - \frac{1}{2} \\ &\quad= \frac{1}{2}\Pr\left(A_n(X_1, \ldots, X_i, Y_{i+1}, \ldots, Y_{\ell(n)}) = 1\right) \\ &\qquad\quad + \frac{1}{2}\left(1 - \Pr\left(A_n(X_1, \ldots, X_{i-1}, Y_i, \ldots, Y_{\ell(n)}) = 1\right)\right) - \frac{1}{2} \\ &\quad= \frac{1}{2}\Pr\left(A_n(X_1, \ldots, X_i, Y_{i+1}, \ldots, Y_{\ell(n)}) = 1\right) \\ &\qquad\quad - \frac{1}{2}\Pr\left(A_n(X_1, \ldots, X_{i-1}, Y_i, \ldots, Y_{\ell(n)}) = 1\right). \end{aligned}$$

By this, we have

$$\begin{aligned}\tilde{\delta}_{g,\tilde{A}}(n) &= \frac{1}{\ell(n)}\sum_{i=1}^{\ell(n)}\left(\Pr\left(\tilde{A}_n(i, X_{\{1,\dots,i-1\}}) = X_i\right) - \frac{1}{2}\right)\\ &= \frac{1}{2}\cdot\frac{1}{\ell(n)}\sum_{i=1}^{\ell(n)}\left(\Pr\left(A_n(X_1,\dots,X_i,Y_{i+1},\dots,Y_{\ell(n)}) = 1\right)\right.\\ &\qquad\left. -\Pr\left(A_n(X_1,\dots,X_{i-1},Y_i,\dots,Y_{\ell(n)}) = 1\right)\right)\\ &= \frac{1}{2\ell(n)}\left(\Pr(A_n(X) = 1) - \Pr(A_n(Y) = 1)\right).\end{aligned}$$

Therefore we see

$$\left|\tilde{\delta}_{g,\tilde{A}}(n)\right| = \frac{\delta_{g,A}(n)}{2\ell(n)}.$$

Since $T_{\tilde{A}}(n)$ is clearly a polynomial parameter, so is $\tilde{S}_{g,\tilde{A}}(n)$. □

There are many pseudorandom generators used in Monte Carlo methods which are defined by recursive formulas like

$$z_i := f(z_{i-n},\dots,z_{i-1}), \quad i = n, n+1,\dots \tag{4.2}$$

with $(z_0,\dots,z_{n-1})$ being a seed (cf. [17, 27, 29, 47]). Such pseudorandom generators are clearly next-bit-predictable. Therefore, by Theorem 4.7, pseudorandom generators defined by recursive formulas and computationally secure ones are poles apart.

According to Theorem 4.7, to construct a computationally secure pseudorandom generator, we have only to pay attention to its next-bit-unpredictability. This is a guiding principle of designing computationally secure pseudorandom generator. Indeed, under this guiding principle, several candidates for computationally secure pseudorandom generator have been presented. Among them, we introduce the following *B-B-S generator* ([2, 36]).[†5] Let p, q be primes satisfying $p = 3(\mathrm{mod}\ 4)$ and $q = 3(\mathrm{mod}\ 4)$, which are secret. But the product $N = pq$ is open to the public. Let $QR(N)$ denote the set of quadratic residues modulo N. Choosing a seed $x_0 \in QR(N)$, we define a pseudorandom number $\{y_n\}_n$ by

$$x_n := F(x_{n-1}) = x_{n-1}^2 \bmod N, \qquad n = 1, 2, \dots, \tag{4.3}$$

$$y_n := G(x_n) = x_n \bmod 2. \tag{4.4}$$

The inverse function F^{-1} is easy to calculate if p, q are known. But when p, q are unknown and they are very large, F^{-1} is very hard to calculate. Consequently, without knowing the seed x_0, it seems to be almost impossible to predict the next bit y_{m+1} when $\{y_n\}_{n=1}^m$ is given.

Remark 4.8 Such a function F as in (4.3), which is easy to compute but whose inverse is hard to compute, is called a one-way function. In general, assuming the existence of a one-way function F, which is a stronger assumption than $\mathbf{P} \neq \mathbf{NP}$, we can show that there exists a function G, so-called a hard core bit function, such that $\{y_n\}_n$ defined by $y_n := G(F^n(x_0))$, where F^n stands for the n-fold iteration of F, becomes next-bit-unpredictable ([24]).

[†5] In the case of B-B-S generator, since the set of seeds is $QR(N)$, a little modification of the definition in § 4.1.1 is needed.

—— ◇ —— ◇ ——

The security of pseudorandom generator discussed here is considered for random variables which are functions of a finite number of coin tosses. But for applications, it should be considered for general simulatable random variables (§ 1.3). In this context, if we restrict the use of pseudorandom generator to the Monte Carlo integration, the dynamic random Weyl sampling, which we will introduce in § 5.4, can be called a secure pseudorandom generator for general simulatable integrands.

4.2 Pseudorandom generator by means of Weyl transformation

We introduce a pseudorandom generator which is based on probability theory. Any finite dimensional distribution of the pseudorandom number it produces can be explicitly computed, and it converges to the corresponding distribution of the coin tossing process as the size of the seed grows. Of course, we do not know if it is computationally secure, but the difference between $1/2$ and the success probabilities of some special next-bit-predictions converge to 0 exponentially fast as the size of the seed grows (Theorem 4.11).

4.2.1 Definitions

In § 4.2, we discuss a family of $\{0,1\}$-valued stochastic processes $\{Y_n^{(m)}(\bullet;\alpha)\}_{n=0}^{\infty}$, $\alpha \in \mathbb{T}^1$, $m \in \mathbb{N}^+$, on the Lebesgue probability space $(\mathbb{T}^1, \mathcal{B}, \mathbb{P})$ defined by Definition 4.9 below.

Definition 4.9 For each $\alpha \in \mathbb{T}^1$ and $m \in \mathbb{N}^+$, set

$$Y_n^{(m)}(x;\alpha) := \left(\sum_{i=1}^{m} d_i(x+n\alpha)\right) \bmod 2, \quad n = 0, 1 \ldots, \quad x \in \mathbb{T}^1. \tag{4.5}$$

In order to realize (4.5) by computer, we have to approximate real numbers by finite dyadic decimals.

Theorem 4.10 *([53]) For each $\alpha \in \mathbb{T}^1$ and $j, j_1, m \in \mathbb{N}^+$, it holds that*

$$\mathbb{P}\Big(0 \le \exists n \le 2^{j_1} - 1 \text{ s.t. } Y_n^{(m)}(\bullet;\alpha) \ne Y_n^{(m)}(\lfloor\bullet\rfloor_{m+j}; \lfloor\alpha\rfloor_{m+j})\Big) < 2^{-(j-2j_1)}.$$

Proof. Note that

$$\begin{aligned}\Big|(x+n\alpha) - \Big(\lfloor x\rfloor_{m+j} + n\lfloor\alpha\rfloor_{m+j}\Big)\Big| &\le \big|x - \lfloor x\rfloor_{m+j}\big| + n\big|\alpha - \lfloor\alpha\rfloor_{m+j}\big| \\ &< 2^{-m-j} + n2^{-m-j} = (n+1)2^{-m-j}.\end{aligned}$$

By this, we see

$$\mathbb{P}\Big(\lfloor\bullet + n\alpha\rfloor_m \ne \lfloor\lfloor\bullet\rfloor_{m+j} + n\lfloor\alpha\rfloor_{m+j}\rfloor_m\Big) \le \frac{(n+1)2^{-m-j}}{2^{-m}} = (n+1)2^{-j}.$$

Therefore

$$\begin{aligned}
&\mathbb{P}\Big(0 \le \exists n \le 2^{j_1}-1 \text{ s.t. } Y_n^{(m)}(\bullet;\alpha) \ne Y_n^{(m)}(\lfloor\bullet\rfloor_{m+j};\lfloor\alpha\rfloor_{m+j})\Big)\\
&\quad\le \sum_{n=0}^{2^{j_1}-1} \mathbb{P}\Big(\lfloor\bullet + n\alpha\rfloor_m \ne \lfloor\bullet\rfloor_{m+j} + n\lfloor\alpha\rfloor_{m+j}\Big)\\
&\quad< \sum_{n=0}^{2^{j_1}-1} (n+1)2^{-j+1}\\
&\quad= \frac{(2^{j_1}-1)2^{j_1}}{2}\cdot 2^{-j+1} < 2^{-(j-2j_1)}.
\end{aligned}$$

□

According to Theorem 4.10, the discretized process $\{Y_n^{(m)}(\lfloor\bullet\rfloor_{m+j};\lfloor\alpha\rfloor_{m+j})\}_{n=0}^{2^{j_1}-1}$ can be arbitrarily close to the stochastic process (4.5) in distribution by taking j large enough. [†6] Let us regard this discretized stochastic process as a pseudorandom generator;[†7]

$$\{Y_n^{(m)}(\bullet;\lfloor\alpha\rfloor_{m+j})\}_{n=0}^{2^{j_1}-1} : D_{m+j} \cong \{0,1\}^{m+j} \to \{0,1\}^{2^{j_1}}.$$

In case α is irrational, we call this generator the *pseudorandom generator by means of Weyl transformation*.[†8] As we will see below, asymptotic behavior of this generator as $m \to \infty$ is very interesting.

In order to generate a sample of $\{Y_n^{(m)}(\bullet;\lfloor\alpha\rfloor_{m+j})\}_{n=0}^{2^{j_1}-1}$, we define mappings $F_{m+j,\alpha} : D_{m+j} \to D_{m+j}$ and $G_m : D_{m+j} \to \{0,1\}$ by

$$F(\tilde{x}) = F_{m+j,\alpha}(\tilde{x}) := \tilde{x} + \lfloor\alpha\rfloor_{m+j}, \tag{4.6}$$

$$G(\tilde{x}) = G_m(\tilde{x}) := \left(\sum_{i=1}^{m} d_i(\tilde{x})\right) \bmod 2, \tag{4.7}$$

and choose an $\tilde{x}_0 \in D_{m+j} \cong \{0,1\}^{m+j}$ as a seed. Then we have

$$Y_n^{(m)}(\tilde{x}_0;\lfloor\alpha\rfloor_{m+j}) = G(F^n(\tilde{x}_0)), \quad n = 0,1,\dots,2^{j_1}-1. \tag{4.8}$$

Here F^n stands for the n-fold iteration of F. $G(\tilde{x})$ is called the *parity* of the upper m bit of $\tilde{x}$, which can be calculated quickly by computer. A concrete implementation of the pseudorandom generator in C language can be found in § 6.2 or [43].

4.2.2 Hardness of next-bit-prediction

As is seen in (4.6) and (4.7), the pseudorandom generator by means of Weyl transformation $\{Y_n^{(m)}(\bullet;\lfloor\alpha\rfloor_{m+j})\}_{n=0}^{2^j-1}$ has a similar structure as the B-B-S generator (4.3)(4.4). In the

[†6] Here we used a transformation of entropy 0; $\mathbb{T}^1 \ni x \mapsto x+\alpha \in \mathbb{T}^1$. For a chaotic transformation (of positive entropy), we would have no good estimate of approximation as Theorem 4.10.

[†7] We should take the parameters $j, j_1 \in \mathbb{N}^+$ large in accordance with m getting large, but for the sake of simple notation, we do not in this monograph.

[†8] Weyl transformation is a mapping $x \mapsto x+\alpha$ on $\mathbb{T}^1$ where $\alpha \in \mathbb{T}^1$ is irrational. It is also called the *irrational rotation*.

latter case, the next bit seems hard to predict because of the complexity of the inverse function F^{-1}, while in the former case, it seems hard to predict if m is large because of the complexity of the function G. More exactly, as m grows,[†9] the complexity of $G = G_m$ becomes large, and accordingly, next-bit-prediction becomes hard. Let us show below that such phenomena really happen in cases of certain special next-bit-predictions.

To formulate limit theorems, we consider the original process (4.5) instead of the discretized one. For $l \in \mathbb{N}^+$ and $0 \le k_0 < \cdots < k_{l-1}$, let us introduce a function $\tilde{A}$ to predict the value of $Y^{(m)}_{k_{l-1}}(x;\alpha)$ when $\{Y^{(m)}_{k_j}(x;\alpha)\}_{j=0}^{l-2}$ is given. Set

$$F^{(m)}(k_0,\dots,k_{l-1};\alpha) := \mathbb{P}\left(\sum_{j=0}^{l-1} Y^{(m)}_{k_j}(\,\bullet\,;\alpha) = \text{odd}\right). \tag{4.9}$$

As is seen in Theorem 4.13 below, there is a fast algorithm to compute this probability. Using it, we define the following function $\tilde{A} : \{0,1\}^{l-1} \to \{0,1\}$.

$$\tilde{A}(y_{k_0},\dots,y_{k_{l-2}}) := \begin{cases} \mathbf{1}_{\{y_{k_0}+\cdots+y_{k_{l-2}}=\text{even}\}} & (F^{(m)}(k_0,k_1,\dots,k_{l-1};\alpha) \ge \frac{1}{2}), \\ \mathbf{1}_{\{y_{k_0}+\cdots+y_{k_{l-2}}=\text{odd}\}} & (F^{(m)}(k_0,k_1,\dots,k_{l-1};\alpha) < \frac{1}{2}). \end{cases}$$

Suppose that we predict that $Y^{(m)}_{k_{l-1}}(x;\alpha)$ will be $\tilde{A}(Y^{(m)}_{k_0}(x;\alpha),\dots,Y^{(m)}_{k_{l-2}}(x;\alpha))$. Then the success probability of this prediction is

$$\mathbb{P}\left(\tilde{A}(Y^{(m)}_{k_0}(\bullet;\alpha),\dots,Y^{(m)}_{k_{l-2}}(\bullet;\alpha)) = Y^{(m)}_{k_{l-1}}(\bullet;\alpha)\right) = \left|F^{(m)}(k_0,k_1,\dots,k_{l-1};\alpha) - \frac{1}{2}\right| + \frac{1}{2}. \tag{4.10}$$

Indeed, in the case of $F^{(m)}(k_0,k_1,\dots,k_{l-1};\alpha) \ge 1/2$,

$$\begin{aligned} \mathbb{P}\left(\tilde{A}(Y^{(m)}_{k_0}(\bullet;\alpha),\dots,Y^{(m)}_{k_{l-2}}(\bullet;\alpha)) = Y^{(m)}_{k_{l-1}}(\bullet;\alpha)\right) &= \mathbb{P}\left(Y^{(m)}_{k_0}(\bullet;\alpha)+\cdots+Y^{(m)}_{k_{l-1}}(\bullet;\alpha) = \text{odd}\right) \\ &= F^{(m)}(k_0,k_1,\dots,k_{l-1};\alpha), \end{aligned}$$

and in the case of $F^{(m)}(k_0,k_1,\dots,k_{l-1};\alpha) < 1/2$,

$$\begin{aligned} \mathbb{P}\left(\tilde{A}(Y^{(m)}_{k_0}(\bullet;\alpha),\dots,Y^{(m)}_{k_{l-2}}(\bullet;\alpha)) = Y^{(m)}_{k_{l-1}}(\bullet;\alpha)\right) &= \mathbb{P}\left(Y^{(m)}_{k_0}(\bullet;\alpha)+\cdots+Y^{(m)}_{k_{l-1}}(\bullet;\alpha) = \text{even}\right) \\ &= 1 - F^{(m)}(k_0,k_1,\dots,k_{l-1};\alpha), \end{aligned}$$

thus in both cases, (4.10) holds.

The prediction of the next bit by the function $\tilde{A}$ succeeds with probability $\ge 1/2$. About this success probability (4.10), we have the following theorem.

Theorem 4.11 *For $\mathbb{P}$-a.e. $\alpha \in \mathbb{T}^1$, it holds that for any $l \ge 2$, $0 \le k_0 < \ldots < k_{l-1}$, there exists $0 < \rho < 1$ which does not depend on α such that as $m \to \infty$,*

$$\mathbb{P}\left(\tilde{A}(Y^{(m)}_{k_0}(\bullet;\alpha),\dots,Y^{(m)}_{k_{l-2}}(\bullet;\alpha)) = Y^{(m)}_{k_{l-1}}(\bullet;\alpha)\right) - \frac{1}{2} = \left|F^{(m)}(k_0,k_1,\dots,k_{l-1};\alpha) - \frac{1}{2}\right| = O(\rho^m).$$

Theorem 4.11 shows that the next-bit-prediction by $\tilde{A}$ becomes very hard as the size of the seed grows. It is really remarkable that we can see the hardness of next-bit-predictions analytically, although they are special ones.

In the case of $l = 2$, for any $\rho > \sqrt{(1+\sqrt{17})/8} = 0.80024\ldots$, we can show the assertion of Theorem 4.11 (§ 4.3.5 Theorem 4.36).

[†9] m is an approximate size of the seed.

—— ◇ —— ◇ ——

In view of Theorem 4.11, Theorem 4.36 and the result (Table 4.1) of an experiment about Hypothesis 4.18 below, The author suspects that for almost all irrational α, the pseudorandom generator by means of the Weyl transformation is computationally secure.[†10]

4.2.3 Formula of finite dimensional distributions and disappearance of dependence

There exists an algorithm to compute any finite dimensional distribution of the stochastic process $\{Y_n^{(m)}(\bullet;\alpha)\}_{n=0}^{\infty}$.

Lemma 4.12 *(i) For each $\epsilon_n \in \{0,1\}$, $n = 0,1,\dots,k-1$, it holds that*

$$\begin{aligned}&\mathbb{P}\left(Y_n^{(m)}(\,\bullet\,;\,\alpha) = \epsilon_n, \quad n = 0,\dots,k-1\right)\\ &= 2^{-k}\left(\sum_{l=1}^{k}\sum_{0\le k_0<\cdots<k_{l-1}\le k-1}\prod_{j=0}^{l-1}(1-2\epsilon_{k_j})\left(1-2F^{(m)}(k_0,\dots,k_{l-1};\alpha)\right)+1\right).\end{aligned}$$

(ii) If $l \in \mathbb{N}^+$ is odd, we have $F^{(m)}(k_0,\dots,k_{l-1};\alpha) = 1/2$.
(iii) $F^{(m)}(k_0,\dots,k_{l-1};\alpha) = F^{(m)}(0,k_1-k_0,\dots,k_{l-1}-k_0;\alpha)$. Thus we may assume $k_0 = 0$ to know any finite dimensional distribution.

In what follows, we assume l is even and that $\alpha \in \mathbb{T}^1$ is irrational. Let us introduce an algorithm to compute $F^{(m)}(0,k_1,\dots,k_{l-1};\alpha)$. To do this, we need some new notation. First, for each $j = 1,\dots,l-1$, set[†11]

$$\begin{cases}\alpha_j &:= \langle k_j\alpha\rangle,\\ \alpha_j^{(m)L} &:= \lfloor\alpha_j\rfloor_m,\\ \alpha_j^{(m)U} &:= \lceil\alpha_j\rceil_m,\\ \beta_j^{(m)} &:= 2^m(\alpha_j-\alpha_j^{(m)L}),\end{cases}$$

and

$$\beta_0^{(m)} := 1, \quad \beta_l^{(m)} := 0.$$

We next define a permutation $\sigma(m,\bullet)$ on the set $\{0,1,\dots,l-1,l\}$ so that

$$1 = \beta_{\sigma(m,0)}^{(m)} > \beta_{\sigma(m,1)}^{(m)} > \beta_{\sigma(m,2)}^{(m)} > \cdots > \beta_{\sigma(m,l-1)}^{(m)} > \beta_{\sigma(m,l)}^{(m)} = 0, \tag{4.11}$$

in particular, we have $\sigma(m,0) = 0$ and $\sigma(m,l) = l$. Set

$$\alpha_{\sigma(m,j)}^{(m),s} := \begin{cases}\alpha_{\sigma(m,j)}^{(m)U} & (\,j \le s\,),\\ \alpha_{\sigma(m,j)}^{(m)L} & (\,j > s\,),\end{cases}$$

[†10] Even if this is true, it would be impossible to find a concrete example of such α, so **P≠NP** would not follow.

[†11] $\langle t\rangle$ denotes the fractional part of $t \ge 0$, i.e., $\langle t\rangle = t - \lfloor t\rfloor$.

and

$$\boldsymbol{\alpha}^{(m),s} := (\alpha_1^{(m),s}, \dots, \alpha_{l-1}^{(m),s}), \quad s = 0, 1, \dots, l-1.$$

Finally, set

$$D := \bigcup_{m \in \mathbb{N}^+} D_m.$$

Theorem 4.13 *The following formula holds.*

$$F^{(m)}(0, k_1, \dots, k_{l-1}; \alpha) = \sum_{s=0}^{l-1} \left(\beta^{(m)}_{\sigma(m,s)} - \beta^{(m)}_{\sigma(m,s+1)} \right) B(\boldsymbol{\alpha}^{(m),s}). \tag{4.12}$$

Here $B(\bullet)$ is a real valued function defined on $D^{l-1} = \overbrace{D \times \cdots \times D}^{l-1}$, whose value $B(\boldsymbol{\alpha}^{(m),s})$ is determined by

$$B(\boldsymbol{\alpha}^{(0),s}) = 0, \qquad s = 0, 1, \dots, l-1,$$

and the following recursive formula

$$B(\boldsymbol{\alpha}^{(m),s}) = \begin{cases} \dfrac{1}{2} B(\boldsymbol{\alpha}^{(m-1),s_2}) + \dfrac{1}{2} B(\boldsymbol{\alpha}^{(m-1),s_1+s_2}) & (\, s_1 \text{ is even }), \\ \dfrac{1}{2} \left(1 - B(\boldsymbol{\alpha}^{(m-1),s_2}) \right) + \dfrac{1}{2} \left(1 - B(\boldsymbol{\alpha}^{(m-1),s_1+s_2}) \right) & (\, s_1 \text{ is odd }), \end{cases}$$

where s_1, s_2 are given by

$$s_1 := \sum_{j=1}^{l-1} d_m(\alpha_j^{(m),s}), \qquad s_2 := \sum_{j=1}^{s} d_m(\alpha_{\sigma(m,j)}). \tag{4.13}$$

From Theorem 4.11 and Lemma 4.12, the following dependence disappearing theorem[†12] follows.

Theorem 4.14 *([37, 51]) For $\mathbb{P}$-a.e. $\alpha \in \mathbb{T}^1$, each finite dimensional distribution of the stochastic process $\{Y_n^{(m)}(\bullet\,; \alpha)\}_{n=0}^{\infty}$ converges to the corresponding finite dimensional distribution of the coin tossing process as $m \to \infty$ exponentially fast. More exactly, for $\mathbb{P}$-a.e. $\alpha \in \mathbb{T}^1$, it holds that for any k and any $\epsilon_n \in \{0, 1\}$, $n = 0, 1, \dots, k-1$, there exists $0 < \rho < 1$ which does not depend on α such that*

$$\left| \mathbb{P}\left(Y_n^{(m)}(\bullet\,; \alpha) = \epsilon_n, \quad n = 0, \dots, k-1 \right) - 2^{-k} \right| = O(\rho^m), \quad m \to \infty.$$

Here is another theorem.

Theorem 4.15 *([52]) For any irrational $\alpha \in \mathbb{T}^1$, each finite dimensional distribution of $\{Y_n^{(m)}(\bullet\,; \alpha)\}_{n=0}^{\infty}$ converges to the corresponding finite dimensional distribution of the coin tossing process as $m \to \infty$. More exactly, any k, $\epsilon_n \in \{0, 1\}$, $n = 0, 1, \dots, k-1$,*

$$\lim_{m \to \infty} \mathbb{P}\left(Y_n^{(m)}(\bullet\,; \alpha) = \epsilon_n, \quad n = 0, \dots, k-1 \right) = 2^{-k}.$$

[†12] Yasutomi proved, in his papers [50, 51, 52], several extended versions of Theorem 4.14. In this monograph, we use some of his ideas with a little modification to fit the context here.

—— ◇ —— ◇ ——

The author got the idea of the dependence disappearing theorems from Theorem 5.10 in § 5.2.2 below. Such dependence disappearing phenomena probably occur so often in practical numerical calculations. It is very likely to occur that a pseudorandom generator defined by a certain recursive formula such as (4.2) can produce samples of random variable S which look very random, if S is very complicated. This may be a reason why simple pseudorandom generators (cf. [17]) are useful to some extent in practice.

4.2.4 A priori estimate of finite dimensional distributions

By using Theorem 4.13, we can investigate statistical properties of the finite dimensional distributions of $\{Y_n^{(m)}(\,\bullet\,;\alpha)\}_{n=0}^{\infty}$.

Let us consider the two-term correlations. Set

$$\begin{cases} \eta_{n;k}^{(m)}(\,\bullet\,;\alpha) &:= Y_n^{(m)}(\,\bullet\,;\alpha) + Y_{n+k}^{(m)}(\,\bullet\,;\alpha) \pmod 2, \\ S_{N;k}^{(m)}(\,\bullet\,;\alpha) &:= \dfrac{1}{N}\displaystyle\sum_{n=0}^{N-1} \eta_{n;k}^{(m)}(\,\bullet\,;\alpha). \end{cases}$$

If $\{Y_n^{(m)}(\,\bullet\,;\alpha)\}_{n=0}^{\infty}$ were a coin tossing process, the variance of $S_{N;k}^{(m)}(\,\bullet\,;\alpha)$ would be $\sigma_N^2 := 1/(4N)$. For each $k \in \mathbb{N}^+$, in order to test the hypothesis

$$\mathbf{E}\left[S_{N;k}^{(m)}(\,\bullet\,;\alpha)\right] \equiv F^{(m)}(0,k;\alpha) = \frac{1}{2}, \tag{4.14}$$

we compute the probability that

$$\left|S_{N;k}^{(m)}(\,\bullet\,;\alpha) - \frac{1}{2}\right| < 2\sigma_N = \frac{1}{\sqrt{N}}. \tag{4.15}$$

If $\{Y_n^{(m)}(\,\bullet\,;\alpha)\}_{n=0}^{\infty}$ were a coin tossing process, by the central limit theorem, the probability of the event (4.15) would be about 0.95.

Theorem 4.16 [†13] *Let*

$$N^{(m)}(k;\alpha) := \frac{1}{16\left(F^{(m)}(0,k;\alpha) - \frac{1}{2}\right)^2}. \tag{4.16}$$

Then for $m \gg 1$, the probability of the event (4.15) is about 0.92 *(or more), if $N = N^{(m)}(k;\alpha)$ (or $N < N^{(m)}(k;\alpha)$).*

Proof.[†14] If $m \gg 1$, the process $\{Y_n^{(m)}(\,\bullet\,;\alpha)\}_{n=0}^{\infty}$ is very close to coin tossing process, and hence the variance of $S_{N;k}^{(m)}(\,\bullet\,;\alpha)$ is almost equal to σ_N^2. If N is large enough, the distribution of $S_{N;k}^{(m)}(\,\bullet\,;\alpha)$ is close to $\mathcal{N}(1/2 + a, \sigma_N^2)$ by the central limit theorem, where

[†13] This is not exactly a theorem because it is a little bit vague.

[†14] This is not exactly a proof because it is a little bit vague.

$a = F^{(m)}(0,k;\alpha) - 1/2$. Now suppose $N = N^{(m)}(k;\alpha) = 1/(16a^2)$, then since $|a| = \sqrt{N}/4 = \sigma_N/2$, we have

$$\left|S^{(m)}_{N;k} - \frac{1}{2}\right| < 2\sigma_N \iff \begin{cases} -\dfrac{5\sigma_N}{2} < S^{(m)}_{N;k} - \left(\dfrac{1}{2} + a\right) < \dfrac{3\sigma_N}{2} & (a > 0), \\ -\dfrac{3\sigma_N}{2} < S^{(m)}_{N;k} - \left(\dfrac{1}{2} + a\right) < \dfrac{5\sigma_N}{2} & (a < 0). \end{cases}$$

Therefore, in both cases, a simple change of variables shows that the probability of the above event is approximately

$$\int_{-5/2}^{3/2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \;=\; 0.926983. \qquad \square$$

According to Theorem 4.16, if $N \le N^{(m)}(k;\alpha)$, we can expect that the pseudorandom number $\{Y^{(m)}_n(\,\bullet\,;\alpha)\}_{n=0}^N$ will be accepted by the statistical test of hypothesis (4.14) with significance level 8%. Theorem 4.16 assumes $m \gg 1$, but, in fact, for not so large m, the estimate of Theorem 4.16 is valid. See the following example.

Example 4.17 Applying Theorem 4.13 to the case where $\alpha = (\sqrt{5}-1)/2$, $m = 40$ and $k = 305$, we have $F^{(40)}(0,305;\alpha) = 0.5029834$. On the other hand,

$$\frac{1}{16\left(F^{(40)}(0,305;\alpha) - \frac{1}{2}\right)^2} = \frac{1}{16 \times (0.0029834)^2} = 7021.94 \approx 7022.$$

To show Theorem 4.16 is useful in this case, we computed the following probability numerically.

$$\mathbb{P}\left(\left|S^{(40)}_{7022;305}(\bullet;\alpha) - \frac{1}{2}\right| \;<\; \frac{1}{\sqrt{7022}}\right). \tag{4.17}$$

To do this, we generated the sequence

$$\{Y^{(40)}_n(0;\lfloor\alpha\rfloor_{150})\}_{n=1}^{7022\times 10^6+305}$$

by a computer, and we counted

$$p_i := \frac{1}{7022}\#\{7022(i-1)+1 \le j \le 7022\, i \,|\, \eta^{(40)}_{j;305}(0;\lfloor\alpha\rfloor_{150}) = 1\,\},$$

for $i = 1, 2, \ldots, 10^6$. Then we had

$$\text{the mean of } \left\{p_i - \frac{1}{2}\right\}_{i=1}^{10^6} \;=\; 10^{-6}\sum_{i=1}^{10^6}\left(p_i - \frac{1}{2}\right) \;=\; 0.002983535,$$

$$\text{the variance of } \{p_i\}_{i=1}^{10^6} \;=\; 10^{-6}\sum_{i=1}^{10^6}(p_i - 0.502983535)^2 \;=\; 0.0000370605.$$

The mean is close to the theoretical value 0.0029834. The variance is larger by 4% than the theoretical value $1/(4\times 7022) = 0.0000356024$ of coin tossing process. The number of i's which satisfy

$$\left|p_i - \frac{1}{2}\right| \;<\; \frac{1}{\sqrt{7022}}$$

is 921514, which means that the probability (4.17) is approximately 92.15%.

We continue to adopt the golden ratio as the irrational number α for Weyl transformation;

$$\alpha = \frac{\sqrt{5}-1}{2}.$$

Keeping Theorem 4.16 in mind, for $K \in \mathbb{N}^+$, we set

$$a^{(m)}(K) := \max_{1\le k\le K}\left|F^{(m)}(0,k;\alpha)-\frac{1}{2}\right|, \qquad N_c^{(m)}(K) := \frac{1}{16\left(a^{(m)}(K)\right)^2}. \tag{4.18}$$

We call $N_c^{(m)}(K)$ the critical sample number.[†15] We computed the quantities of (4.18) with $K = 10,000$, whose results are shown in the left half of Table 4.1. The number written in ( ) to the right of the value $a^{(m)}(10000)$ is the number k which achieves the maximum of $\left|F^{(m)}(0,k;\alpha)-\frac{1}{2}\right|$.[†16]

Table 4.1: Two-term and multi-term correlations

m	$a^{(m)}(10000)$	(k)	$N_c^{(m)}(10000)$	$b^{(m)}(19)$	$k_1,\dots$
10	0.4860680	(5473)	2.6×10^{-1}	0.1187876	18
20	0.1084934	(1449)	5.3×10^{0}	0.0088276	4, 5, 13, 14, 18
30	0.0435756	(305)	3.3×10^{1}	0.0009169	18
40	0.0029834	(305)	7.0×10^{3}	0.0000769	9
50	0.0001943	(610)	1.7×10^{6}	1.5×10^{-5}	18
60	0.0000136	(8484)	3.4×10^{8}	6.4×10^{-7}	18
70	1.2×10^{-6}	(7264)	4.1×10^{10}	5.9×10^{-8}	1
80	2.0×10^{-7}	(7697)	1.6×10^{12}	7.7×10^{-9}	18
90	8.5×10^{-9}	(165)	8.7×10^{14}	2.1×10^{-9}	16
100	2.8×10^{-9}	(5201)	8.1×10^{15}	3.0×10^{-10}	1

Next, let us estimate general finite dimensional distributions up to K dimension; we estimate the following for even numbers l.

$$F^{(m)}(0,k_1,\dots,k_{l-1};\alpha), \qquad 1\le k_1<\cdots<k_{l-1}\le K.$$

To compute all of them is computationally hard even for a rather small K. But we have a little hope. The right half of Table 4.1 shows the computation result of them for $K = 19$. The left column indicates

$$b^{(m)}(19) := \max_{1\le k_1<\dots<k_{l-1}\le 19}\left|F^{(m)}(0,k_1,\dots,k_{l-1};\alpha)-\frac{1}{2}\right|,$$

and the right one indicates what combination of $k_1,\dots$ the maximum value is achieved. The result of the right half of Table 4.1 convinces us that the following hypothesis should hold.[†17]

[†15] The critical sample number defined in [37] is 4 times as large as $N_c^{(m)}(K)$.

[†16] Since α is irrational, we approximated it by two finite dyadic decimals $\lfloor\alpha\rfloor_{150}$ and $\lfloor\alpha\rfloor_{300}$. For both of them, we got the same table of results (Table 4.1).

[†17] More exactly, for $K = 19$, the equality (4.19) holds for all $37 \le m \le 100$.

Hypothesis 4.18 For each $K \in \mathbb{N}^+$, if $m \gg 1$, it holds that

$$\max_{1\le k_1<\cdots<k_{l-1}\le K}\left|F^{(m)}(0,k_1,\ldots,k_{l-1};\alpha)-\frac{1}{2}\right| \;=\; \max_{1\le k\le K}\left|F^{(m)}(0,k;\alpha)-\frac{1}{2}\right|. \tag{4.19}$$

We have no proof, yet. If Hypothesis 4.18 is correct, we have to estimate only maximum of the two-term correlations.

4.3 Proofs of theorems

In this section, we will prove Lemma 4.12, Theorem 4.13, Theorem 4.15, and Theorem 4.11 in this order.[†18] Proofs will be given not for the stochastic process $\{Y_n^{(m)}\}_{n=0}^{\infty}$, but for an equivalent $\{-1,1\}$-valued process $\{X_n^{(m)}\}_{n=0}^{\infty}$ defined by (4.20) below. $\{Y_n^{(m)}\}_{n=0}^{\infty}$ is better for implementation by computer, while $\{X_n^{(m)}\}_{n=0}^{\infty}$ is better for mathematical analysis.

Let $\{r_i\}_{i=1}^{\infty}$ denote the *Rademacher functions*, i.e.,

$$r_i(x) \;:=\; 1-2d_i(x), \quad x\in\mathbb{T}^1, \quad i\in\mathbb{N}^+.$$

For irrational $\alpha\in\mathbb{T}^1$ and $m\in\mathbb{N}^+$, we define

$$X_n^{(m)}(x;\alpha) \;:=\; \prod_{i=1}^{m} r_i(x+n\alpha), \qquad n\in\mathbb{N}. \tag{4.20}$$

The relation between $\{X_n^{(m)}\}_{n=0}^{\infty}$ and $\{Y_n^{(m)}\}_{n=0}^{\infty}$ is

$$X_n^{(m)}(x;\alpha) \;=\; 1-2Y_n^{(m)}(x;\alpha), \quad \text{or} \quad Y_n^{(m)}(x;\alpha) \;=\; \frac{1}{2}\Big(1-X_n^{(m)}(x;\alpha)\Big).$$

Note the following property; for any k, $h\in\mathbb{N}^+$ and any $\boldsymbol{\epsilon}\in\{-1,1\}^k$, it holds that

$$\mathbb{P}\Big((X_0^{(m)}(\,\bullet\,;\alpha),\ldots,X_{k-1}^{(m)}(\,\bullet\,;\alpha))=\boldsymbol{\epsilon}\Big) \;=\; \mathbb{P}\Big((X_h^{(m)}(\,\bullet\,;\alpha),\ldots,X_{k-1+h}^{(m)}(\,\bullet\,;\alpha))=\boldsymbol{\epsilon}\Big). \tag{4.21}$$

This property is called (strong) stationarity. The proof of (4.21) is readily derived from the translation invariance of the Lebesgue measure — i.e., lengths of intervals do not change by shift.

4.3.1 Proof of Lemma 4.12

Here is a lemma in terms of $\{X_n^{(m)}\}_{n=0}^{\infty}$ which is equivalent to Lemma 4.12.

Lemma 4.12′ *(i) Any finite dimensional distribution of $\{X_n^{(m)}\}_{n=0}^{\infty}$ is derived from the following quantities.*

$$E^{(m)}(k_0,\ldots,k_{l-1};\alpha) \;:=\; \mathbf{E}\left[\prod_{j=0}^{l-1} X_{k_j}^{(m)}(\,\bullet\,;\alpha)\right], \qquad 0\le k_0<\ldots<k_{l-1}, \quad l\in\mathbb{N}^+.$$

[†18] Detailed proofs will be given. The reader may skip this section at the first reading.

In fact, for any $\epsilon_n \in \{-1, 1\}$, $n = 0, 1, \dots, k-1$, *we have*

$$\mathbb{P}\left(X_n^{(m)}(\bullet;\alpha) = \epsilon_n, \quad n = 0, \dots, k-1\right) = 2^{-k}\left(\sum_{l=1}^{k} \sum_{0\le k_0<\dots<k_{l-1}\le k-1} \prod_{j=0}^{l-1} \epsilon_{k_j} E^{(m)}(k_0, \dots, k_{l-1};\alpha) + 1\right). \tag{4.22}$$

(ii) If $l \in \mathbb{N}^+$ *is odd, then* $E^{(m)}(k_0, \dots, k_{l-1};\alpha) = 0$.
(iii) $E^{(m)}(k_0, \dots, k_{l-1};\alpha) = E^{(m)}(0, k_1 - k_0, \dots, k_{l-1} - k_0;\alpha)$.

Proof. (i) Note the following equality.

$$\sum_{l=1}^{k} \sum_{0\le k_0<\dots<k_{l-1}\le k-1} \prod_{j=0}^{l-1} \left(\epsilon_{k_j} X_{k_j}^{(m)}(x;\alpha)\right) = \prod_{n=0}^{k-1} \left(1 + \epsilon_n X_n^{(m)}(x;\alpha)\right) - 1. \tag{4.23}$$

The mean of the left hand side is equal to

$$\sum_{l=1}^{k} \sum_{0\le k_0<\dots<k_{l-1}\le k-1} \prod_{j=0}^{l-1} \epsilon_{k_j} E^{(m)}(k_0, \dots, k_{l-1};\alpha). \tag{4.24}$$

On the other hand, the mean of the right hand side is equal to

$$\mathbf{E}\left[\prod_{n=0}^{k-1} \left(1 + \epsilon_n X_n^{(m)}(\bullet;\alpha)\right)\right] - 1. \tag{4.25}$$

Now, the stuff inside $\mathbf{E}[\,\bullet\,]$ of (4.25) is 2^k, if $X_n^{(m)}(x;\alpha) = \epsilon_n$ holds for each $n = 0, \dots, k-1$, and it is 0 otherwise. Therefor (4.25) is reduced to

$$2^k \mathbb{P}\left(X_n^{(m)}(\bullet;\alpha) = \epsilon_n, \quad n = 0, \dots, k-1\right) - 1. \tag{4.26}$$

Since (4.24) and (4.26) are equal, (4.22) follows.

(ii) Since $r_1(x + \frac{1}{2}) = -r_1(x)$, $r_i(x + \frac{1}{2}) = r_i(x)$, $i \ge 2$, we readily see

$$X_k^{(m)}\left(x + \frac{1}{2};\alpha\right) = -X_k^{(m)}(x;\alpha), \qquad x \in \mathbb{T}^1. \tag{4.27}$$

Then if l is odd,

$$X_0^{(m)}(x;\alpha) \times \cdots \times X_{k_{l-1}}^{(m)}(x;\alpha) = -1$$

and

$$X_0^{(m)}\left(x + \frac{1}{2};\alpha\right) \times \cdots \times X_{k_{l-1}}^{(m)}\left(x + \frac{1}{2};\alpha\right) = 1$$

are equivalent, and hence their probabilities coincide. But the probability of the latter is equal to the probability of

$$X_0^{(m)}(x;\alpha) \times \cdots \times X_{k_{l-1}}^{(m)}(x;\alpha) = 1,$$

by the shift invariance of the Lebesgue measure, so all of these probabilities must be $1/2$. Form this, the assertion (ii) follows. (iii) is obvious by the stationarity (4.21). □

4.3.2 Proof of Theorem 4.13

Here is a theorem in terms of $\{X_n^{(m)}\}_{n=0}^{\infty}$ which is equivalent to Theorem 4.13.

Theorem 4.13′

$$E^{(m)}(0, k_1, \dots, k_{l-1}; \alpha) \;=\; \sum_{s=0}^{l-1} \left(\beta^{(m)}_{\sigma(m,s)} - \beta^{(m)}_{\sigma(m,s+1)}\right) A(\boldsymbol{\alpha}^{(m),s}). \tag{4.28}$$

Here $A(\bullet)$ is a real valued function defined on $D^{l-1} = \overbrace{D \times \cdots \times D}^{l-1}$, whose value $A(\boldsymbol{\alpha}^{(m),s})$ is determined by

$$A(\boldsymbol{\alpha}^{(0),s}) \;=\; 1, \qquad s = 0, 1, \dots, l-1,$$

and a recursive formula

$$A(\boldsymbol{\alpha}^{(m),s}) \;=\; \frac{(-1)^{s_1}}{2}\left(A(\boldsymbol{\alpha}^{(m-1),s_2}) + A(\boldsymbol{\alpha}^{(m-1),s_1+s_2})\right),$$

where s_1 and s_2 have been defined by (4.13).

We will prove Theorem 4.13′. In what follows, we assume that l is an even number. By definition, we have

$$E^{(m)}(0, k_1, \dots, k_{l-1}; \alpha) \;=\; \mathbf{E}\left[\prod_{i=1}^{m} r_i(\bullet) r_i(\bullet + k_1\alpha) \times \cdots \times r_i(\bullet + k_{l-1}\alpha)\right].$$

Keeping this in mind, we define a function $A^{(m)}$ for each $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{l-1}) \in \mathbb{T}^{l-1}$ by

$$A^{(m)}(\boldsymbol{\alpha}) \;:=\; \mathbf{E}\left[\prod_{i=1}^{m} r_i(\bullet) r_i(\bullet + \alpha_1) \times \cdots \times r_i(\bullet + \alpha_{l-1})\right]. \tag{4.29}$$

Lemma 4.19 *For each $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{l-1}) \in (D_m)^{l-1}$, it holds that*

$$\forall m' \geq m, \quad A^{(m')}(\boldsymbol{\alpha}) \;=\; A^{(m)}(\boldsymbol{\alpha}).$$

Proof. Write $A^{(m')}(\boldsymbol{\alpha})$ in the following way.

$$\begin{aligned} A^{(m')}(\boldsymbol{\alpha}) \;&=\; \mathbf{E}\Bigg[\prod_{i=1}^{m} r_i(\bullet) r_i(\bullet + \alpha_1) \times \cdots \times r_i(\bullet + \alpha_{l-1}) \\ &\qquad \times \prod_{i=m+1}^{m'} r_i(\bullet) r_i(\bullet + \alpha_1) \times \cdots \times r_i(\bullet + \alpha_{l-1})\Bigg]. \end{aligned}$$

If $\boldsymbol{\alpha} \in (D_m)^{l-1}$, for $i > m$, we have

$$r_i(x) = r_i(x + \alpha_j), \quad j = 1, \dots, l-1.$$

Then, since l is even, the latter half product is reduced to 1;

$$\prod_{i=m+1}^{m'} r_i(x) r_i(x + \alpha_1) \times \cdots \times r_i(x + \alpha_{l-1}) \;=\; \prod_{i=m+1}^{m'} r_i(x)^l = 1,$$

This shows $A^{(m')}(\boldsymbol{\alpha}) = A^{(m)}(\boldsymbol{\alpha})$. □

Definition 4.20 For each $\boldsymbol{\alpha} \in D^{l-1}$, we define

$$A(\boldsymbol{\alpha}) := \lim_{m\to\infty} A^{(m)}(\boldsymbol{\alpha}).$$

Definition 4.20 is justified by Lemma 4.19. The function A appeared in Theorem 4.13′ is nothing but the one defined by Definition 4.20. Lemma 4.21 below shows that the value of A can be obtained by a recursive formula. To state Lemma 4.21, we need some additional definitions; for each $\boldsymbol{\alpha} \in \mathbb{T}^{l-1}$, we set

$$\begin{aligned} \boldsymbol{\alpha}^{(m)L} &:= (\alpha_1^{(m)L}, \dots, \alpha_{l-1}^{(m)L}), \quad \alpha_j^{(m)L} := \lfloor \alpha_j \rfloor_m, \\ \boldsymbol{\alpha}^{(m)U} &:= (\alpha_1^{(m)U}, \dots, \alpha_{l-1}^{(m)U}), \quad \alpha_j^{(m)U} := \lceil \alpha_j \rceil_m. \end{aligned}$$

Then it is obvious that $\boldsymbol{\alpha}^{(m)L}, \boldsymbol{\alpha}^{(m)U} \in (D_m)^{l-1}$.

Lemma 4.21 *(i)* $A(\overbrace{0, \dots, 0}^{l}) = 1$.
(ii) For each $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{l-1}) \in (D_m)^{l-1}$, *set* $j_0 := \sum_{j=1}^{l-1} d_m(\alpha_j)$. *Then we have*

$$A(\boldsymbol{\alpha}) = \frac{(-1)^{j_0}}{2} \left(A(\boldsymbol{\alpha}^{(m-1)U}) + A(\boldsymbol{\alpha}^{(m-1)L}) \right). \tag{4.30}$$

Proof. (i) Since l is even, we see

$$A(\overbrace{0, \dots, 0}^{l}) = \mathbf{E}\left[\prod_{i=1}^{m} \overbrace{r_i(\bullet) \times \cdots \times r_i(\bullet)}^{l} \right] = 1.$$

(ii) We can show that

$$\prod_{i=1}^{m} r_i(x + \alpha_j) = \begin{cases} \displaystyle\prod_{i=1}^{m-1} r_i\left(x + \alpha_j + 2^{-m}\right) & (\, d_m(\alpha_j) = 1,\ d_m(x) = 1\,), \\ \displaystyle -\prod_{i=1}^{m-1} r_i\left(x + \alpha_j - 2^{-m}\right) & (\, d_m(\alpha_j) = 1,\ d_m(x) = 0\,), \\ \displaystyle -\prod_{i=1}^{m-1} r_i\left(x + \alpha_j\right) & (\, d_m(\alpha_j) = 0,\ d_m(x) = 1\,), \\ \displaystyle\prod_{i=1}^{m-1} r_i\left(x + \alpha_j\right) & (\, d_m(\alpha_j) = 0,\ d_m(x) = 0\,). \end{cases} \tag{4.31}$$

Indeed, if $d_m(\alpha_j) = 0$, then $r_m(x + \alpha_j) = r_m(x)$ and hence

$$\prod_{i=1}^{m} r_i(x + \alpha_j) = \prod_{i=1}^{m-1} r_i(x + \alpha_j) \times r_m(x).$$

From this, we see the third and the fourth cases.

Assume next that $d_m(\alpha_j) = 1$. Then we have $r_m(x + \alpha_j) = -r_m(x)$. Assume further that $d_m(x) = 1$. In this case, we have $d_m(x + \alpha_j) = 0$, and hence for each $i = 1, \dots, m-1$,

$d_i(x+\alpha_j) = d_i(x+\alpha_j+2^{-m})$ holds, namely, $r_i(x+\alpha_j) = r_i(x+\alpha_j+2^{-m})$. Consequently, we see

$$\prod_{i=1}^{m} r_i(x+\alpha_j) \;=\; \prod_{i=1}^{m-1} r_i(x+\alpha_j)\times r_m(x+\alpha_j) \;=\; \prod_{i=1}^{m-1} r_i\left(x+\alpha_j+2^{-m}\right),$$

which shows the first case.

Finally, let us consider the second case, where $d_m(\alpha_j)=1$ and $d_m(x)=0$. This time, since $d_m(x+\alpha_j)=1$, for $i=1,\dots,m-1$, we have $d_i(x+\alpha_j) = d_i(x+\alpha_j-2^{-m})$, i.e., $r_i(x+\alpha_j) = r_i(x+\alpha_j-2^{-m})$. Therefore

$$\prod_{i=1}^{m} r_i(x+\alpha_j) \;=\; \prod_{i=1}^{m-1} r_i(x+\alpha_j)\times r_m(x+\alpha_j) \;=\; -\prod_{i=1}^{m-1} r_i\left(x+\alpha_j-2^{-m}\right),$$

which shows the second case. Thus (4.31) is proved.

In order to make notation simple, let us assume the following situation.

$$d_m(\alpha_j) \;=\; \begin{cases} 1 & (1\le j\le j_0),\\ 0 & (j_0+1\le j\le l-1). \end{cases}$$

Note that $j_0 = \sum_{j=1}^{l-1} d_m(\alpha_j)$. Then

$$\begin{aligned} A(\boldsymbol{\alpha}) &= \mathbf{E}\left[\prod_{i=1}^{m}\left(r_i(\bullet)\prod_{j=1}^{j_0} r_i(\bullet+\alpha_j)\prod_{j=j_0+1}^{l-1} r_i(\bullet+\alpha_j)\right)\right]\\ &= \mathbf{E}\left[-\prod_{i=1}^{m-1} r_i(\bullet)\prod_{j=1}^{j_0}\prod_{i=1}^{m-1} r_i\left(\bullet+\alpha_j+2^{-m}\right)\prod_{j=j_0+1}^{l-1}\left(-\prod_{i=1}^{m-1} r_i(\bullet+\alpha_j)\right);\, d_m(\bullet)=1\right]\\ &\quad +\mathbf{E}\left[\prod_{i=1}^{m-1} r_i(\bullet)\prod_{j=1}^{j_0}\left(-\prod_{i=1}^{m-1} r_i\left(\bullet+\alpha_j-2^{-m}\right)\right)\prod_{j=j_0+1}^{l-1}\prod_{i=1}^{m-1} r_i(\bullet+\alpha_j);\, d_m(\bullet)=0\right]. \end{aligned}$$

Each integrand is independent of the given event $\{d_m(x)=\epsilon\}$ ($\epsilon=0$ or 1), so we see

$$\begin{aligned} A(\boldsymbol{\alpha}) &= \frac{1}{2}\mathbf{E}\left[\prod_{i=1}^{m-1} r_i(\bullet)\prod_{j=1}^{l-1}\prod_{i=1}^{m-1} r_i(\bullet+\alpha_j^{(m-1)U})\times(-1)^{l-j_0}\right]\\ &\quad +\frac{1}{2}\mathbf{E}\left[\prod_{i=1}^{m-1} r_i(\bullet)\prod_{j=1}^{l-1}\prod_{i=1}^{m-1} r_i(\bullet+\alpha_j^{(m-1)L})\times(-1)^{j_0}\right]. \end{aligned}$$

Now, if j_0 is even, so is $l-j_0$, and hence

$$A(\boldsymbol{\alpha}) = \frac{1}{2}A(\boldsymbol{\alpha}^{(m-1)U}) + \frac{1}{2}A(\boldsymbol{\alpha}^{(m-1)L}),$$

If j_0 is odd, so is $l-j_0$, and hence

$$A(\boldsymbol{\alpha}) = -\frac{1}{2}A(\boldsymbol{\alpha}^{(m-1)U}) - \frac{1}{2}A(\boldsymbol{\alpha}^{(m-1)L}).$$

Thus the proof is complete. □

Proof of Theorem 4.13′

Step 1. Set

$$C_j := \bigcup_{s=1}^{2^m} \left[\frac{s}{2^m} - \frac{\beta^{(m)}_{\sigma(m,j)}}{2^m}, \frac{s}{2^m} - \frac{\beta^{(m)}_{\sigma(m,j+1)}}{2^m} \right), \quad j = 0, 1, \ldots, l-1,$$

then for $i = 1, \ldots, m$, we have

$$x \in C_j \implies d_i(x + \alpha_{\sigma(m,p)}) = \begin{cases} d_i\left(x + \alpha^{(m)U}_{\sigma(m,p)}\right) & (1 \le p \le j), \\ d_i\left(x + \alpha^{(m)L}_{\sigma(m,p)}\right) & (j+1 \le p \le l-1), \end{cases}$$

and hence it holds that

$$\begin{aligned} &E^{(m)}(0, k_1, \ldots, k_{l-1}; \alpha) \\ &\quad = \sum_{j=0}^{l-1} \mathbf{E}\left[\prod_{i=1}^{m} \left(r_i(\bullet) \prod_{p=1}^{j} r_i(\bullet + \alpha^{(m)U}_{\sigma(m,p)}) \prod_{p=j+1}^{l-1} r_i(\bullet + \alpha^{(m)L}_{\sigma(m,p)}) \right) ; C_j \right]. \end{aligned}$$

Since the integrand is independent of the event C_j,

$$\begin{aligned} &E^{(m)}(0, k_1, \ldots, k_{l-1}; \alpha) \\ &\quad = \sum_{j=0}^{l-1} \mathbb{P}(C_j) \mathbf{E}\left[\prod_{i=1}^{m} \left(r_i(\bullet) \prod_{p=1}^{j} r_i(\bullet + \alpha^{(m)U}_{\sigma(m,p)}) \prod_{p=j+1}^{l-1} r_i(\bullet + \alpha^{(m)L}_{\sigma(m,p)}) \right) \right] \\ &\quad = \sum_{j=0}^{l-1} \mathbb{P}(C_j) A(\boldsymbol{\alpha}^{(m),j}). \end{aligned}$$

Now, the first half part of Theorem 4.13′ follows from the fact $\mathbb{P}(C_j) = \beta^{(m)}_{\sigma(m,j)} - \beta^{(m)}_{\sigma(m,j+1)}$.

Step 2. The second half part of Theorem 4.13′ can be shown by applying Lemma 4.21. To do this, we will show

$$\left(\boldsymbol{\alpha}^{(m),s}\right)^{(m-1)U} = \boldsymbol{\alpha}^{(m-1),s_1+s_2}, \tag{4.32}$$

$$\left(\boldsymbol{\alpha}^{(m),s}\right)^{(m-1)L} = \boldsymbol{\alpha}^{(m-1),s_2}. \tag{4.33}$$

Step 2.1. Let us prove (4.33) first. To do this, we show

$$\#\left\{ j \,\middle|\, \left(\alpha^{(m),s}_j\right)^{(m-1)L} = \alpha^{(m-1)U}_j \right\} = s_2. \tag{4.34}$$

Confirm the following four implications.

$$\begin{aligned} \alpha^{(m),s}_j \neq \alpha^{(m)U}_j &\implies \alpha^{(m),s}_j = \alpha^{(m)L}_j \\ &\implies \left(\alpha^{(m),s}_j\right)^{(m-1)L} = \left(\alpha^{(m)L}_j\right)^{(m-1)L} = \alpha^{(m-1)L}_j \end{aligned}$$

$$\Longrightarrow \quad \left(\alpha_j^{(m),s}\right)^{(m-1)L} \neq \alpha_j^{(m-1)U}, \tag{4.35}$$

$$\alpha_j^{(m),s} = \alpha_j^{(m)U} \quad \Longrightarrow \quad \left(\alpha_j^{(m),s}\right)^{(m-1)U} = \left(\alpha_j^{(m)U}\right)^{(m-1)U} = \alpha_j^{(m-1)U}, \tag{4.36}$$

$$d_m(\alpha_j^{(m),s}) = 1 \quad \Longleftrightarrow \quad \left(\alpha_j^{(m),s}\right)^{(m-1)L} \neq \left(\alpha_j^{(m),s}\right)^{(m-1)U} \tag{4.37}$$

$$\Longrightarrow \quad \left(\alpha_j^{(m),s}\right)^{(m-1)L} \neq \alpha_j^{(m-1)U}, \tag{4.38}$$

$$d_m(\alpha_j^{(m),s}) = 0 \quad \Longleftrightarrow \quad \left(\alpha_j^{(m),s}\right)^{(m-1)L} = \left(\alpha_j^{(m),s}\right)^{(m-1)U}. \tag{4.39}$$

The contrapositives of (4.35) and (4.38) imply

$$\left(\alpha_j^{(m),s}\right)^{(m-1)L} = \alpha_j^{(m-1)U} \Longrightarrow \begin{cases} \alpha_j^{(m),s} = \alpha_j^{(m)U} \\ \text{and} \\ d_m(\alpha_j^{(m),s}) = 0. \end{cases}$$

The converse is also valid because of (4.36) and (4.39). Thus we have

$$\left(\alpha_j^{(m),s}\right)^{(m-1)L} = \alpha_j^{(m-1)U} \Longleftrightarrow \begin{cases} \alpha_j^{(m),s} = \alpha_j^{(m)U} \\ \text{and} \\ d_m(\alpha_j^{(m),s}) = 0. \end{cases}$$

Consequently,

$$\begin{aligned} J_0 &:= \left\{ j \,\middle|\, \left(\alpha_j^{(m),s}\right)^{(m-1)L} = \alpha_j^{(m-1)U} \right\} \\ &= \left\{ j \,\middle|\, \alpha_j^{(m),s} = \alpha_j^{(m)U},\ d_m(\alpha_j^{(m),s}) = 0 \right\} \\ &= \left\{ \sigma(m,j) \,\middle|\, \alpha_{\sigma(m,j)}^{(m),s} = \alpha_{\sigma(m,j)}^{(m)U},\ d_m(\alpha_{\sigma(m,j)}^{(m),s}) = 0 \right\} \\ &= \left\{ \sigma(m,j) \,\middle|\, 1 \le j \le s,\ d_m(\alpha_{\sigma(m,j)}^{(m),s}) = d_m(\alpha_{\sigma(m,j)}^{(m)U}) = 0 \right\} \\ &= \left\{ \sigma(m,j) \,\middle|\, 1 \le j \le s,\ d_m(\alpha_{\sigma(m,j)}^{(m)L}) = d_m(\alpha_{\sigma(m,j)}) = 1 \right\} =: J_1. \end{aligned}$$

Since (4.13) implies $\#J_1 = s_2$, we see $\#J_0 = s_2$ (4.34).

Step 2.2. By the definition of $\beta_i^{(m)}$, we readily see

$$\beta_i^{(m-1)} = \frac{1}{2}\beta_i^{(m)} + \frac{1}{2}d_m(\alpha_i), \quad 1 \le i \le l-1.$$

Therefore when we sort $\{\beta_i^{(m)}\}_i$ in descending order, those $\beta_i^{(m)}$ whose subscripts i satisfy $d_m(\alpha_i) = 1$ are ranked highly. In particular,

$$\beta_i^{(m-1)} > \beta_j^{(m-1)} \Longleftrightarrow \begin{cases} 1 = d_m(\alpha_i) > d_m(\alpha_j) = 0, \\ \text{or} \\ d_m(\alpha_i) = d_m(\alpha_j) \quad \text{and} \quad \beta_i^{(m)} > \beta_j^{(m)}. \end{cases}$$

Step 2.3. Next, let

$$J_2 := \{\, i \,|\, 1 \le \exists\, j \le s_2 \ \text{ s.t. } \ i = \sigma(m-1, j) \,\}.$$

We want to prove that $J_1 = J_2$. Suppose $i \in J_1$. Then there exists $1 \le j \le s$ such that $i = \sigma(m, j)$ and $d_m(\alpha_i) = 1$. Therefore the definition of s_2 and Step 2.2 imply that $\beta_i^{(m-1)}$ ranks within s_2-th in descending order. This means that $i \in J_2$ and hence that $J_1 \subset J_2$. Since $\#J_1 = \#J_2 = s_2$, we see $J_1 = J_2$.

Step 2.4. $i \in J_2$ implies $\alpha_i^{(m-1),s_2} = \alpha_i^{(m-1)U}$, while $i \notin J_2$ implies $\alpha_i^{(m-1),s_2} = \alpha_i^{(m-1)L}$. Since $J_2 = J_1 = J_0$, if $i \in J_0$ then $\alpha_i^{(m-1),s_2} = \alpha_i^{(m-1)U} = \left(\alpha_i^{(m),s}\right)^{(m-1)L}$, while if $i \notin J_0$ then $\alpha_i^{(m-1),s_2} = \alpha_i^{(m-1)L} = \left(\alpha_i^{(m),s}\right)^{(m-1)L}$. This shows (4.33).

Step 2.5. Now, to prove (4.32), we first show that

$$\#\left\{ j \,\middle|\, \left(\alpha_j^{(m),s}\right)^{(m-1)U} = \alpha_j^{(m-1)U} \right\} = \#\left\{ j \,\middle|\, \left(\alpha_j^{(m),s}\right)^{(m-1)U} \neq \alpha_j^{(m-1)L} \right\} = s_1 + s_2. \tag{4.40}$$

Obviously,

$$\left(\alpha_j^{(m),s}\right)^{(m-1)U} \neq \alpha_j^{(m-1)L} \iff \begin{cases} \left(\alpha_j^{(m),s}\right)^{(m-1)U} \neq \left(\alpha_j^{(m),s}\right)^{(m-1)L} = \alpha_j^{(m-1)L} \\ \text{or} \\ \left(\alpha_j^{(m),s}\right)^{(m-1)U} = \left(\alpha_j^{(m),s}\right)^{(m-1)L} \neq \alpha_j^{(m-1)L} \end{cases}$$

$$\iff \begin{cases} \text{either} \\ \left(\alpha_j^{(m),s}\right)^{(m-1)U} \neq \left(\alpha_j^{(m),s}\right)^{(m-1)L} \\ \text{or} \\ \left(\alpha_j^{(m),s}\right)^{(m-1)L} \neq \alpha_j^{(m-1)L} \end{cases}$$

Namely, we have

$$\left\{ j \,\middle|\, \left(\alpha_j^{(m),s}\right)^{(m-1)U} \neq \alpha_j^{(m-1)L} \right\}$$
$$= \left\{ j \,\middle|\, \left(\alpha_j^{(m),s}\right)^{(m-1)U} \neq \left(\alpha_j^{(m),s}\right)^{(m-1)L} \right\} \cup \left\{ j \,\middle|\, \left(\alpha_j^{(m),s}\right)^{(m-1)L} \neq \alpha_j^{(m-1)L} \right\},$$

which is a disjoint union. Note that (4.37) implies

$$\#\left\{ j \,\middle|\, \left(\alpha_j^{(m),s}\right)^{(m-1)U} \neq \left(\alpha_j^{(m),s}\right)^{(m-1)L} \right\} = \#\left\{ j \,\middle|\, d_m(\alpha_j^{(m),s}) = 1 \right\} = s_1$$

and that (4.34) implies

$$\#\left\{ j \,\middle|\, \left(\alpha_j^{(m),s}\right)^{(m-1)L} \neq \alpha_j^{(m-1)L} \right\} = \#\left\{ j \,\middle|\, \left(\alpha_j^{(m),s}\right)^{(m-1)L} = \alpha_j^{(m-1)U} \right\} = s_2.$$

The above three equations show (4.40).

Step 2.6. Let

$$J_3 := \left\{ j \,\middle|\, d_m(\alpha_j^{(m),s}) = 1 \right\}.$$

From the above arguments, it follows that

$$\left\{ j \,\middle|\, \left(\alpha_j^{(m),s}\right)^{(m-1)U} = \alpha_j^{(m-1)U} \right\} = J_3 \cup J_0 = J_3 \cup J_2,$$

which is a disjoint union. Now, set

$$J_4 := \{\, i \,|\, 1 + s_2 \le \exists\, j \le s_1 + s_2 \ \text{ s.t. } \ i = \sigma(m-1, j) \,\}.$$

Then let us show $J_3 = J_4$. By the definition of $\alpha_j^{(m),s}$, we know that

$$\begin{aligned} J_3 &= \left\{ \sigma(m,j) \,\middle|\, 1 \le j \le s,\ d_m(\alpha_{\sigma(m,j)}) = 0 \right\} \\ &\quad \cup \left\{ \sigma(m,j) \,\middle|\, s+1 \le j \le l-1,\ d_m(\alpha_{\sigma(m,j)}) = 1 \right\} \\ &=: J_5 \cup J_6. \end{aligned}$$

Sorting $\{\beta_i^{(m-1)}\}_i$ in descending order, those $\beta_i^{(m-1)}$ whose subscripts i belong to $J_0 = J_2$ rank within s_2-th. Noting Step 2.2, those $\beta_i^{(m-1)}$ come next whose subscripts i belong to J_6, and those $\beta_i^{(m-1)}$ finally come whose subscripts i belong to J_5. These facts and Step 2.5 imply $J_3 = J_4$.

From all the arguments above, we see that

$$\left\{ j \,\middle|\, \left(\alpha_j^{(m),s}\right)^{(m-1)U} = \alpha_j^{(m-1)U} \right\} = J_2 \cup J_4 = \{\, i \,|\, 1 \le \exists\, j \le s_1 + s_2 \ \text{ s.t. } \ i = \sigma(m-1, j) \,\},$$

which proves (4.32). □

4.3.3 Proof of Theorem 4.15

By Lemma 4.12′, the proof of Theorem 4.15 is reduced to showing the following; *for any even integer $l \in \mathbb{N}^+$ and any $l-1$ integers $k_1 < \cdots < k_{l-1}$, it holds that*

$$\lim_{m\to\infty} E^{(m)}(0, k_1, \ldots, k_{l-1}\,;\, \alpha) = 0$$

for any irrational α.

We first present the idea of the proof. Let us write down the algorithm of Theorem 4.13′ in a concrete case. Suppose that $l = 4$ and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ is given by

$$\begin{cases} \alpha_1 &= 0.011010110\ldots, \\ \alpha_2 &= 0.110011101\ldots, \\ \alpha_3 &= 0.010111011\ldots. \end{cases}$$

All these are expressed in dyadic decimals. If $m = 6$, we see

$$\begin{cases} \alpha_1^{(6)L} = 0.011010, & \alpha_1^{(6)U} = 0.011011, \\ \alpha_2^{(6)L} = 0.110011, & \alpha_2^{(6)U} = 0.110100, \\ \alpha_3^{(6)L} = 0.010111, & \alpha_3^{(6)U} = 0.011000, \end{cases}$$

and

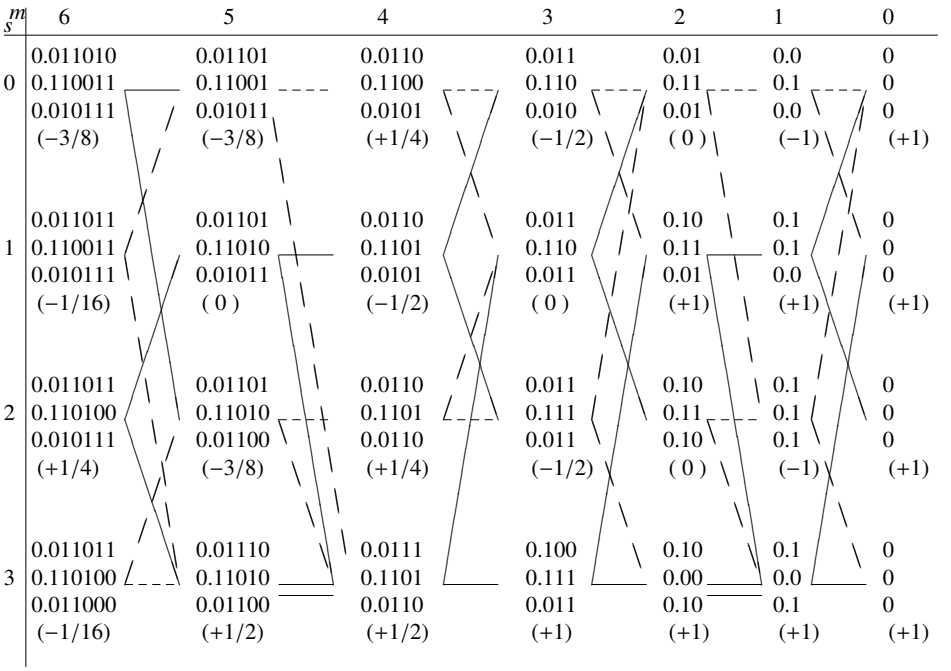
$$1 = \beta_0^{(6)} > \beta_1^{(6)} = 0.110\ldots > \beta_2^{(6)} = 0.101\ldots > \beta_3^{(6)} = 0.011\ldots > \beta_4^{(6)} = 0$$

In this case, we have $\sigma(6, j) = j$, $j = 1, 2, 3$, and[†19]

$$\boldsymbol{\alpha}^{(6),j} = \begin{pmatrix} 0.011010 \\ 0.110011 \\ 0.010111 \end{pmatrix}, \begin{pmatrix} 0.011011 \\ 0.110011 \\ 0.010111 \end{pmatrix}, \begin{pmatrix} 0.011011 \\ 0.110100 \\ 0.010111 \end{pmatrix}, \begin{pmatrix} 0.011011 \\ 0.110100 \\ 0.011000 \end{pmatrix}, \quad j = 0, 1, 2, 3.$$

Now, Figure 4.1 shows a diagram which visualize the algorithm of Theorem 4.13′ and Lemma 4.21. Let us explain it.

Figure 4.1: The diagram of Theorem 4.13′ and Lemma 4.21

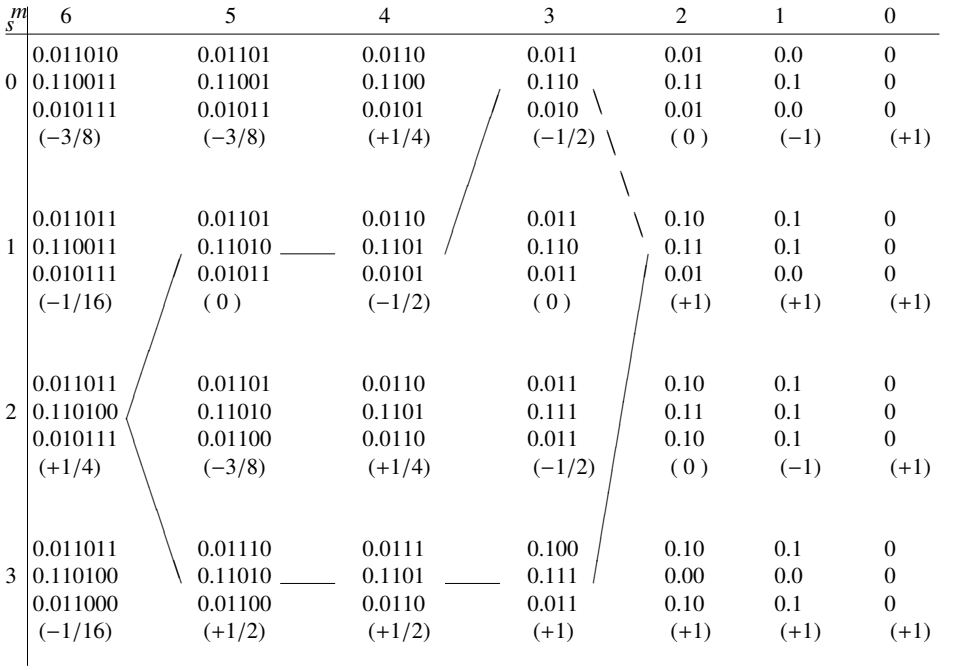


In the diagram, the vector placed at the m-th ($m = 0, \ldots, 6$) column from right and the s-th ($s = 0, \ldots, 3$) row from top represents $\boldsymbol{\alpha}^{(m),s}$. For example, $\boldsymbol{\alpha}^{(6),0}$ is placed at the top-left corner of the diagram. The number in ( ) under each vector $\boldsymbol{\alpha}^{(m),s}$ shows the value of $A(\boldsymbol{\alpha}^{(m),s})$. According to Lemma 4.21, $A(\boldsymbol{\alpha}^{(m),s})$ is computed from its *parents* $A(\boldsymbol{\alpha}^{(m-1),s'})$ and $A(\boldsymbol{\alpha}^{(m-1),s''})$ as (4.30), which is illustrated in the diagram by solid lines if $(-1)^{j_0} = 1$ or broken lines if $(-1)^{j_0} = -1$. For example, the diagram tells that

$$A(\boldsymbol{\alpha}^{(6),2}) = \frac{1}{2}\left(A(\boldsymbol{\alpha}^{(5),1}) + A(\boldsymbol{\alpha}^{(5),3})\right). \tag{4.41}$$

[†19] Vectors are written as column vectors.

Figure 4.2: Two routes that cancel out



In order to prove Theorem 4.15, we will show that $|A(\boldsymbol{\alpha}^{(m),s})| \to 0$ as $m \to \infty$ for each s. We explain the idea using this diagram. Let us, for example, trace the *family line* of $A(\boldsymbol{\alpha}^{(6),2})$ through four generations. The number of all possible routes is $2^4 = 16$. Among them, the two routes shown in Figure 4.2 cancel out. That is, along the lower route of Figure 4.2, $A(\boldsymbol{\alpha}^{(2),1})$ contributes to the calculation of $A(\boldsymbol{\alpha}^{(6),2})$ by $2^{-4}A(\boldsymbol{\alpha}^{(2),1})$, while along the upper route, it contributes by $-2^{-4}A(\boldsymbol{\alpha}^{(2),1})$, and hence the two contributions cancel out. From this, it follows that

$$|A(\boldsymbol{\alpha}^{(6),2})| \le \left(1 - \frac{1}{2^4} \times 2\right) \times \max_{s=0,\dots,3} |A(\boldsymbol{\alpha}^{(2),s})|.$$

The proof of Theorem 4.15 will be done by finding infinitely many such pairs of canceling routes, accordingly iterating the above estimation infinitely many times, and finally by showing $|A(\boldsymbol{\alpha}^{(m),s})| \to 0$ as $m \to \infty$ (Lemma 4.27).

For the proof of Theorem 4.15, we need several lemmas.

Lemma 4.22 *Let* $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{l-1}) \in (\mathbb{T}^1 \setminus D)^{l-1}$ *and* $m \ge 1$.

(i) $\forall j, \quad d_m(\alpha_j^{(m)U}) = d_m(\alpha_j^{(m),l-1}) \ne d_m(\alpha_j^{(m),0}) = d_m(\alpha_j^{(m)L}) = d_m(\alpha_j)$.

(ii) $\left(\boldsymbol{\alpha}^{(m),0}\right)^{(m-1)L} = \boldsymbol{\alpha}^{(m-1),0}$.

(iii) $\left(\boldsymbol{\alpha}^{(m),l-1}\right)^{(m-1)U} = \boldsymbol{\alpha}^{(m-1),l-1}$.

(iv) $\displaystyle\sum_{j=1}^{l-1} d_m(\alpha_j^{(m),0}) \neq \sum_{j=1}^{l-1} d_m(\alpha_j^{(m),l-1}) \pmod 2.$

(v) $\left(\boldsymbol{\alpha}^{(m),l-1}\right)^{(m-1)L} = \left(\boldsymbol{\alpha}^{(m),0}\right)^{(m-1)U}.$

Proof. (i) Obvious from the relations below.

$$\begin{cases} \alpha_j^{(m),l-1} = \alpha_j^{(m)U}, \\ \alpha_j^{(m)U} \neq \alpha_j^{(m)L}, \\ \alpha_j^{(m),0} = \alpha_j^{(m)L}, \qquad d_m(\alpha_j) = d_m(\alpha_j^{(m)L}). \end{cases}$$

(ii) By the definition of s_2 (4.13), $s = 0$ implies $s_2 = 0$, which shows (ii).

(iii) By (i), if $s = l - 1$, we see

$$s_1 = \sum_{j=1}^{l-1} d_m(\alpha_j^{(m),l-1}) = \sum_{j=1}^{l-1} \left(1 - d_m(\alpha_j)\right) = l - 1 - \sum_{j=1}^{l-1} d_m(\alpha_j).$$

On the other hand,

$$s_2 = \sum_{j=1}^{l-1} d_m(\alpha_{\sigma(m,j)}) = \sum_{j=1}^{l-1} d_m(\alpha_j).$$

Thus if $s = l - 1$, then $s_1 + s_2 = l - 1$, which shows (iii).

(iv) By (i), we see

$$\begin{aligned} \sum_{j=1}^{l-1} d_m(\alpha_j^{(m),0}) &= \sum_{j=1}^{l-1} \left(1 - d_m(\alpha_j^{(m),l-1})\right) \\ &= l - 1 - \sum_{j=1}^{l-1} d_m(\alpha_j^{(m),l-1}). \end{aligned} \tag{4.42}$$

Since $l - 1$ is odd, (iv) follows.

(v) Take $p, q \in \mathbb{N}$ so that

$$\begin{cases} \left(\boldsymbol{\alpha}^{(m),l-1}\right)^{(m-1)L} &= \boldsymbol{\alpha}^{(m-1),p}, \\ \left(\boldsymbol{\alpha}^{(m),0}\right)^{(m-1)U} &= \boldsymbol{\alpha}^{(m-1),q}. \end{cases}$$

Then by (4.13) and (i), we see

$$\begin{aligned} p &= \sum_{j=1}^{l-1} d_m(\alpha_{\sigma(m,j)}) = \sum_{j=1}^{l-1} d_m(\alpha_j), \\ q &= \sum_{j=1}^{l-1} d_m(\alpha_j^{(m),0}) = \sum_{j=1}^{l-1} d_m(\alpha_j), \end{aligned}$$

i.e., $p = q$. □

Definition 4.23 In order to make notation simple, we introduce the following mappings.

$$\begin{cases} \mathcal{L}: & (D_m)^{l-1} \ni \boldsymbol{\alpha} \mapsto \boldsymbol{\alpha}^{(m-1)L} & \in (D_{m-1})^{l-1}, \\ \mathcal{U}: & (D_m)^{l-1} \ni \boldsymbol{\alpha} \mapsto \boldsymbol{\alpha}^{(m-1)U} & \in (D_{m-1})^{l-1}. \end{cases}$$

$\mathcal{L}^p$ and $\mathcal{U}^p$ are regarded as mappings $(D_m)^{l-1} \to (D_{m-p})^{l-1}$.

Lemma 4.24 *Let $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{l-1}) \in (\mathbb{T}^1 \setminus D)^{l-1}$ and $r \in \mathbb{N}^+$.*
(i) If $\mathcal{L}^r \boldsymbol{\alpha}^{(m+r),l-1} = \boldsymbol{\alpha}^{(m),0}$, then $\forall s$, $\mathcal{L}^r \boldsymbol{\alpha}^{(m+r),s} = \boldsymbol{\alpha}^{(m),0}$.
(ii) If $\mathcal{U}^r \boldsymbol{\alpha}^{(m+r),0} = \boldsymbol{\alpha}^{(m),l-1}$, then $\forall s$, $\mathcal{U}^r \boldsymbol{\alpha}^{(m+r),s} = \boldsymbol{\alpha}^{(m),l-1}$.
(iii) If

$$\forall j = 1, \dots, l-1, \quad 1 \le \exists p \le r, \quad d_{m+p}(\alpha_j) = 0, \tag{4.43}$$

then $\forall s$, $\mathcal{L}^r \boldsymbol{\alpha}^{(m+r),s} = \boldsymbol{\alpha}^{(m),0}$.
(iv) If

$$\forall j = 1, \dots, l-1, \quad 1 \le \exists p \le r, \quad d_{m+p}(\alpha_j) = 1, \tag{4.44}$$

then $\forall s$, $\mathcal{U}^r \boldsymbol{\alpha}^{(m+r),s} = \boldsymbol{\alpha}^{(m),l-1}$.

Proof. (i) By (4.13), s_2 is an increasing function of s. From this (i) follows.
(ii) By (4.40), $s_1 + s_2$ is an increasing function of s. From this (ii) follows.
(iii) By (4.43) and Lemma 4.22(i), for each j there exists a p such that $d_{m+p}(\alpha_j^{(m+r),l-1}) = 1$. Hence $\alpha_j^{(m+p-1),l-1} > \alpha_j^{(m+p)L}$. By this and Lemma 4.22(ii), we see $\mathcal{L}^r \boldsymbol{\alpha}^{(m+r),l-1} = \boldsymbol{\alpha}^{(m),0}$. Now (i) therefore implies (iii).
(iv) By (4.44), for each j there exists a p such that $\alpha_j^{(m+p-1),0} < \alpha_j^{(m+p)U}$. By this and Lemma 4.22(iii), we see $\mathcal{U}^r \boldsymbol{\alpha}^{(m+r),0} = \boldsymbol{\alpha}^{(m),l-1}$. Now (ii) therefore implies (iv). □

Lemma 4.25 *([52]) let $r := 3k_{l-1}$, let α be any irrational number, and let $\alpha_j := \langle k_j \alpha \rangle$. Then there exist infinitely many m's that satisfy both (4.43) and (4.44).*

Proof. By contradiction. Assume that there exist only a finite number of m's that satisfy both (4.43) and (4.44), namely, assume that there exists an $N \in \mathbb{N}^+$ such that for any $m \ge N$, there is a j_m such that $d_{m+1}(\alpha_{j_m}) = \dots = d_{m+r}(\alpha_{j_m}) = 0$ or $d_{m+1}(\alpha_{j_m}) = \dots = d_{m+r}(\alpha_{j_m}) = 1$. Then we will prove that α is rational by showing $\{d_{N+i}(\alpha)\}_{i=1}^{\infty}$ is periodic.

Step 1. Fix an $m \ge N$. We will show that a finite sequence $\{d_{m+i}(\alpha)\}_{i=1}^{r}$ is periodic with period at most k_{j_m}.

We investigate the dyadic expansion of α by dividing $k_{j_m}\langle 2^m \alpha \rangle$ by k_{j_m}. First, set $R_1 := \lfloor k_{j_m}\langle 2^m \alpha \rangle \rfloor$. Then we have

$$R_1 + \langle 2^m k_{j_m} \alpha \rangle = \lfloor k_{j_m}\langle 2^m \alpha \rangle \rfloor + \langle k_{j_m}\langle 2^m \alpha \rangle \rangle = k_{j_m}\langle 2^m \alpha \rangle.$$

Multiplying both sides by 2,

$$2R_1 + d_{m+1}(k_{j_m}\alpha) + \langle 2^{m+1} k_{j_m} \alpha \rangle = k_{j_m} d_{m+1}(\alpha) + k_{j_m}\langle 2^{m+1} \alpha \rangle.$$

Since

$$\begin{aligned} k_{j_m}\langle 2^{m+1}\alpha \rangle - \langle 2^{m+1} k_{j_m}\alpha \rangle &= \lfloor k_{j_m}\langle 2^{m+1}\alpha \rangle \rfloor + \langle k_{j_m}\langle 2^{m+1}\alpha \rangle \rangle - \langle 2^{m+1} k_{j_m}\alpha \rangle \\ &= \lfloor k_{j_m}\langle 2^{m+1}\alpha \rangle \rfloor, \end{aligned}$$

note that $0 \le k_{j_m}\langle 2^{m+1}\alpha\rangle - \langle 2^{m+1}k_{j_m}\alpha\rangle < k_{j_m}$. Dividing $2R_1 + d_{m+1}(k_{j_m}\alpha)$ by k_{j_m}, we get the quotient Q_1 and the remainder R_2, where

$$\begin{aligned} Q_1 &= d_{m+1}(\alpha), \\ R_2 &= k_{j_m}\langle 2^{m+1}\alpha\rangle - \langle 2^{m+1}k_{j_m}\alpha\rangle, \end{aligned}$$

or

$$R_2 + \langle 2^{m+1}k_{j_m}\alpha\rangle = k_{j_m}\langle 2^{m+1}\alpha\rangle.$$

As above, dividing $2R_2 + d_{m+2}(k_{j_m}\alpha)$ by k_{j_m}, we next get the quotient Q_2 and the remainder R_3, where

$$\begin{aligned} Q_2 &= d_{m+2}(\alpha), \\ R_3 &= k_{j_m}\langle 2^{m+2}\alpha\rangle - \langle 2^{m+2}k_{j_m}\alpha\rangle. \end{aligned}$$

Similarly, we can get (Q_u, R_{u+1}) so that $2R_u + d_{m+u}(k_{j_m}\alpha) = k_{j_m}Q_u + R_{u+1}$, $0 \le R_{u+1} < k_{j_m}$. Now, by the assumption $d_{m+1}(k_{j_m}\alpha) = \ldots = d_{m+r}(k_{j_m}\alpha)$, the sequence $\{(Q_u, R_{u+1})\}_{u=1}^{r}$ depends only on $\{R_u\}_{u=1}^{r}$. Since R_u can take at most k_{j_m} values , this sequence becomes periodic with period at most k_{j_m}. In particular, $Q_u = d_{m+u}(\alpha)$ is also periodic with period at most k_{j_m}.

Step 2. Let $a(0), a(1), \ldots, a(p-1)$ be a sequence with the smallest period $w \le p/2$. Then we will show that if $p' \ge 2w$, the smallest period of any subsequence $a(q), a(q+1), \ldots, a(q+p'-1)$, $0 \le q < q+p' \le p$, is equal to w.

Let the smallest period of $a(q), a(q+1), \ldots, a(q+p'-1)$ be w'. Obviously $w' \le w$. Since, for any $0 \le u \le p-w'-1$, there exist $j \in \mathbb{Z}$ and $0 \le v < w$ such that $u - q = wj + v$, by $v + w' < w + w' \le 2w \le p'$, we have

$$a(u) = a(q + wj + v) = a(q+v) = a(q+v+w') = a(q+wj+v+w') = a(u+w').$$

This means that w' is a period of the original sequence $a(0), a(1), \ldots, a(p-1)$. Because of the minimality of w, we see $w' = w$.

Step 3. Let us show the assertion of the lemma. Let $m \ge N$. Step 1 implies that the sequence $\{d_{m+i}(\alpha)\}_{i=1}^{r}$ is periodic with the smallest period, say $w_m \le k_{j_m}$. Similarly $\{d_{m+i}(\alpha)\}_{i=2}^{r+1}$ is also periodic with the smallest period $w_{m+1} \le k_{j_{m+1}}$. Then Step 2 implies that $\{d_{m+i}(\alpha)\}_{i=2}^{r}$ is periodic with the smallest period w, which must coincide with both w_m and w_{m+1}. Iterating this procedure, we see that $\{d_{N+i}(\alpha)\}_{i=1}^{\infty}$ is periodic with the common period w. □

Let us return to the proof of Theorem 4.15. By Lemma 4.12′, our aim is to show that for any even l and any $1 \le k_1 < \cdots < k_{l-1}$,

$$\lim_{m\to\infty} E^{(m)}(0, k_1, \ldots, k_{l-1}; \alpha) = 0. \tag{4.45}$$

Let α be irrational and and let us again set

$$\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_{l-1}), \qquad \alpha_j := \langle k_j\alpha\rangle. \tag{4.46}$$

Then by Theorem 4.13′, in order to show (4.45), it is sufficient to show that

$$\lim_{m\to\infty} \max_{0\le s\le l-1} \left|A(\boldsymbol{\alpha}^{(m),s})\right| \;=\; 0. \tag{4.47}$$

By Lemma 4.21, we see

$$A(\boldsymbol{\alpha}^{(m),s}) \;=\; \pm\frac{1}{2}\left\{A(\mathcal{U}\boldsymbol{\alpha}^{(m),s}) + A(\mathcal{L}\boldsymbol{\alpha}^{(m),s})\right\}. \tag{4.48}$$

The following lemma follows immediately from (4.48).

Lemma 4.26 $\max_{1\le s\le l-1} |A(\boldsymbol{\alpha}^{(m'),s})| \le \max_{1\le s\le l-1} |A(\boldsymbol{\alpha}^{(m),s})|, \quad m' > m.$

Now, we present a key lemma.

Lemma 4.27 *Let α be irrational, let $r := 3k_{l-1}$, and let $\{m_n\}_{n=0}^{\infty}$ be a sequence of those infinitely many m's in Lemma 4.25 such that $m \ge 2$ and $m_n + r + 2 \le m_{n+1}$. Then we have*

$$\max_{1\le s\le l-1} \left|A(\boldsymbol{\alpha}^{(m_n+r),s})\right| \;\le\; \left(1-\frac{1}{2^{r+1}}\right) \max_{1\le s\le l-1} \left|A(\boldsymbol{\alpha}^{(m_n-2),s})\right|.$$

Proof. Applying (4.48) r times, we get

$$\begin{aligned} A(\boldsymbol{\alpha}^{(m_n+r),s}) \;=\; & \frac{1}{2^r}\epsilon_{\mathcal{U}^r} A(\mathcal{U}^r\boldsymbol{\alpha}^{(m_n+r),s}) + \frac{1}{2^r}\epsilon_{\mathcal{L}\mathcal{U}^{r-1}} A(\mathcal{L}\mathcal{U}^{r-1}\boldsymbol{\alpha}^{(m_n+r),s}) + \\ & \qquad\vdots \\ & + \frac{1}{2^r}\epsilon_{\mathcal{U}\mathcal{L}^{r-1}} A(\mathcal{U}\mathcal{L}^{r-1}\boldsymbol{\alpha}^{(m_n+r),s}) + \frac{1}{2^r}\epsilon_{\mathcal{L}^r} A(\mathcal{L}^r\boldsymbol{\alpha}^{(m_n+r),s}), \end{aligned} \tag{4.49}$$

where $\epsilon_{\mathcal{U}^r}, \dots, \epsilon_{\mathcal{L}^r} = \pm 1$. By Lemma 4.24 and Lemma 4.25,

$$\forall s, \qquad \mathcal{U}^r\boldsymbol{\alpha}^{(m_n+r),s} = \boldsymbol{\alpha}^{(m_n),l-1}, \qquad \mathcal{L}^r\boldsymbol{\alpha}^{(m_n+r),s} = \boldsymbol{\alpha}^{(m_n),0}. \tag{4.50}$$

<u>*Case 1*</u>. Suppose that $\epsilon_{\mathcal{U}^r} = \epsilon_{\mathcal{L}^r}$. By (4.48), for some ϵ, $\epsilon' = \pm 1$, we have

$$\begin{cases} \epsilon_{\mathcal{U}^r} A(\mathcal{U}^r\boldsymbol{\alpha}^{(m_n+r),s}) \;=\; \epsilon_{\mathcal{U}^r}\epsilon\left\{\frac{1}{2}A(\mathcal{U}^{r+1}\boldsymbol{\alpha}^{(m_n+r),s}) + \frac{1}{2}A(\mathcal{L}\mathcal{U}^r\boldsymbol{\alpha}^{(m_n+r),s})\right\}, \\ \epsilon_{\mathcal{L}^r} A(\mathcal{L}^r\boldsymbol{\alpha}^{(m_n+r),s}) \;=\; \epsilon_{\mathcal{L}^r}\epsilon'\left\{\frac{1}{2}A(\mathcal{U}\mathcal{L}^r\boldsymbol{\alpha}^{(m_n+r),s}) + \frac{1}{2}A(\mathcal{L}^{r+1}\boldsymbol{\alpha}^{(m_n+r),s})\right\}. \end{cases} \tag{4.51}$$

Lemma 4.21, Lemma 4.22(iv), and (4.50) imply that $\epsilon \ne \epsilon'$. Then note that from (4.50) and Lemma 4.22(v), it follows that

$$\mathcal{L}\mathcal{U}^r\boldsymbol{\alpha}^{(m_n+r),s} \;=\; \mathcal{U}\mathcal{L}^r\boldsymbol{\alpha}^{(m_n+r),s}. \tag{4.52}$$

Now, using (4.51), we expand (4.49) once again and we get

$$\begin{aligned} A(\boldsymbol{\alpha}^{(m_n+r),s}) \;=\; & \frac{1}{2^{r+1}}\epsilon_{\mathcal{U}^r}\epsilon A(\mathcal{U}^{r+1}\boldsymbol{\alpha}^{(m_n+r),s}) + \frac{1}{2^{r+1}}\epsilon_{\mathcal{U}^r}\epsilon A(\mathcal{L}\mathcal{U}^r\boldsymbol{\alpha}^{(m_n+r),s}) + \\ & \qquad\vdots \\ & + \frac{1}{2^{r+1}}\epsilon_{\mathcal{L}^r}\epsilon' A(\mathcal{U}\mathcal{L}^r\boldsymbol{\alpha}^{(m_n+r),s}) + \frac{1}{2^{r+1}}\epsilon_{\mathcal{L}^r}\epsilon' A(\mathcal{L}^{r+1}\boldsymbol{\alpha}^{(m_n+r),s}). \end{aligned} \tag{4.53}$$

Since $\epsilon_{\mathcal{U}^r}\epsilon \neq \epsilon_{\mathcal{L}^r}\epsilon'$, by (4.52), we see

$$\frac{1}{2^{r+1}}\epsilon_{\mathcal{U}^r}\epsilon A(\mathcal{L}\mathcal{U}^r\boldsymbol{\alpha}^{(m_n+r),s}) + \frac{1}{2^{r+1}}\epsilon_{\mathcal{L}^r}\epsilon' A(\mathcal{U}\mathcal{L}^r\boldsymbol{\alpha}^{(m_n+r),s}) \;=\; 0. \tag{4.54}$$

Thus we get the following estimate.

$$\left|A(\boldsymbol{\alpha}^{(m_n+r),s})\right| \;\leq\; \left(1 - \frac{1}{2^{r+1}} \times 2\right) \max_{0\leq q\leq l-1} \left|A(\boldsymbol{\alpha}^{(m_n-1),q})\right|. \tag{4.55}$$

Case 2. Suppose that $\epsilon_{\mathcal{U}^r} \neq \epsilon_{\mathcal{L}^r}$. This time, we have $\epsilon_{\mathcal{U}^r}\epsilon = \epsilon_{\mathcal{L}^r}\epsilon'$ in (4.53), and hence applying the same method as Case 1 to (4.53), we get the following estimate.

$$\left|A(\boldsymbol{\alpha}^{(m_n+r),s})\right| \;\leq\; \left(1 - \frac{1}{2^{r+2}} \times 2\right) \max_{0\leq q\leq l-1} \left|A(\boldsymbol{\alpha}^{(m_n-2),q})\right|. \tag{4.56}$$

By Lemma 4.26, (4.55) implies (4.56). Thus in both cases, we have the desired estimate (4.56). □

Now, the proof of Theorem 4.15 is easy. By Lemma 4.27, we finally get

$$\begin{aligned}
\max_{0\leq s\leq l-1} \left|A(\boldsymbol{\alpha}^{(m_n+r),s})\right| &\leq \left(1 - \frac{1}{2^{r+1}}\right) \max_{0\leq s\leq l-1} \left|A(\boldsymbol{\alpha}^{(m_{n-1}),s})\right| \\
&\leq \left(1 - \frac{1}{2^{r+1}}\right)^2 \max_{0\leq s\leq l-1} \left|A(\boldsymbol{\alpha}^{(m_{n-2}),s})\right| \\
&\leq \cdots\cdots\cdots \\
&\leq \left(1 - \frac{1}{2^{r+1}}\right)^n \max_{0\leq s\leq l-1} \left|A(\boldsymbol{\alpha}^{(m_0),s})\right| \\
&\leq \left(1 - \frac{1}{2^{r+1}}\right)^n \longrightarrow 0, \qquad n \to \infty.
\end{aligned} \tag{4.57}$$

Thus we see (4.47). □

Remark 4.28 Roughly speaking, for almost all α, the sequence $\{m_n\}_{n=0}^{\infty}$ is almost *arithmetic progression* due to the law of large numbers. This convinces us that the convergence (4.47) should be exponentially fast in n for almost all α (cf. Theorem 4.11′).

Remark 4.29 If $l \geq 4$, the cancellations (4.54) seen in the above proof is a very special ones, and usually there occur many other cancellations in the expansion (4.49). But if $l = 2$, the cancellations seen in the proof are the all that can occur.

4.3.4 Proof of Theorem 4.11

Using ergodic theory, we will prove Theorem 4.11′ below which is equivalent to Theorem 4.11.

Theorem 4.11′ *For $\mathbb{P}$-a.e. $\alpha \in \mathbb{T}^1$, it holds that for any $l \in \mathbb{N}^+$, $0 \leq k_0 < \ldots < k_{l-1}$, there exists $0 < \rho < 1$ which does not depend on α such that*

$$\mathbf{E}\left[\, X_0^{(m)}(\,\bullet\,;\alpha) X_{k_1}^{(m)}(\,\bullet\,;\alpha) \times \cdots \times X_{k_{l-1}}^{(m)}(\,\bullet\,;\alpha)\right] \;=\; o(\rho^m), \quad m \to \infty. \tag{4.58}$$

Formulation by group extension

We will show that the left hand side of (4.58) is a two term correlation of a certain Markov chain, and that its mixing property implies Theorem 4.11′. To this end, we need the following framework.

Note first that by Lemma 4.12′(ii), we have only to show Theorem 4.11′ for each even l. Fix any $0 < k_1 < \cdots < k_{l-1} \in \mathbb{N}^+$. Define a function $f : \mathbb{T}^2 \to \{-1, 1\}$ by

$$f(x, \alpha) := r_1(x)r_1(x + k_1\alpha) \times \cdots \times r_1(x + k_{l-1}\alpha), \qquad (x, \alpha) \in \mathbb{T}^2. \tag{4.59}$$

Here $r_1(x)$ is the first Rademacher function. By using f, we define the *group extension* (or skew product) of the dyadic transformation $\beta : \mathbb{T}^3 \to \mathbb{T}^3$,

$$\beta(x, y, \alpha) := (2x, 2y, 2\alpha), \tag{4.60}$$

as follows.

Definition 4.30 *Let $\Omega := \mathbb{T}^3 \times \{-1, 1\}^2$ and let μ be the uniform probability measure on Ω, i.e.,*

$$\mu := \mathbb{P}^3 \otimes \frac{\delta_{-1} + \delta_1}{2} \otimes \frac{\delta_{-1} + \delta_1}{2}. \tag{4.61}$$

Here $\otimes$ denotes the direct product of probability measures, and δ_i denotes Dirac's δ-measure concentrated at i. Define a transformation $T_f : \Omega \to \Omega$ by [†20]

$$T_f(x, y, \alpha, \epsilon_1, \epsilon_2) := (2x, 2y, 2\alpha, \epsilon_1 f(x, \alpha), \epsilon_2 f(y, \alpha)). \tag{4.62}$$

Obliviously, T_f preserves μ. Define a subset $C \subset \mathbb{T}^3$ by

$$C := \{(x, y, \alpha) \,|\, (x, y, \alpha) \text{ is a discontinuous point of } f(x, \alpha) \text{ or } f(y, \alpha)\}. \tag{4.63}$$

Then C is of probability 0 and $\beta C \subset C$. Let E_j, $j = 1, \ldots, J$, be the connected components of $\mathbb{T}^3 \setminus C$. Define $F_j \subset \Omega$ as

$$\begin{aligned} \{F_j\}_{j=1}^{4J} := \; & \left\{E_j \times \{-1\} \times \{-1\}\right\}_{j=1}^{J} \bigcup \left\{E_j \times \{-1\} \times \{1\}\right\}_{j=1}^{J} \\ & \bigcup \left\{E_j \times \{1\} \times \{-1\}\right\}_{j=1}^{J} \bigcup \left\{E_j \times \{1\} \times \{1\}\right\}_{j=1}^{J}. \end{aligned}$$

Then $\Omega = \bigcup_{j=1}^{4J} F_j$, μ-a.e.

Definition 4.31 *We define a $\{1, 2, 3, \ldots, 4J\}$-valued stochastic process $\{\zeta_m\}_{m=0}^{\infty}$ on (Ω, μ) as follows; $\zeta_m(x, y, \alpha, \epsilon_1, \epsilon_2) := j$ if $T_f^m(x, y, \alpha, \epsilon_1, \epsilon_2) \in F_j$.*

Now, we present a key lemma.

[†20] The method of group extension introduced here is a modification of Yasutomi's idea seen in [50]. There are also a pioneering work by Takanobu[46], in which he proved the strong mixing property of the transformation $T : (x, \alpha, \epsilon) \mapsto (2x, 2\alpha, \epsilon f(x, \alpha))$ on $\mathbb{T}^2 \times \{-1, 1\}$.

Lemma 4.32 *The process $\{\zeta_m\}_{m=0}^{\infty}$ is an irreducible aperiodic stationary Markov chain with transition matrix*[†21]

$$p(i,j) := \mu\left(T_f^{-1}(F_j)\,\middle|\, F_i\right), \quad i,j = 1,\dots,4J.$$

We will prove Lemma 4.32 later. Let $p^m(i,j) := \mu(\zeta_m = j \,|\, \zeta_0 = i)$. Then from Lemma 4.32, the following corollary immediately follows (cf. [1] Theorem 8.9).

Corollary 4.33 *For any $i, j = 1,\dots,4J$, it holds that*

$$p^m(i,j) \longrightarrow \mu(F_j), \quad m \to \infty,$$

and this convergence takes place at an exponential rate in m.

Using Corollary 4.33, Theorem 4.11′ is proved in the following way. First, define four mappings below; for $i = 1, 2$,

$$\begin{array}{ll} \Phi_i : \Omega \to \{-1,1\}, & \Phi_i(x,y,\alpha,\epsilon_1,\epsilon_2) := \epsilon_i, \\ \widetilde{\Phi}_i : \{1,\dots,4J\} \to \{-1,1\}, & \widetilde{\Phi}_i(j) := \Phi_i(F_j) = \epsilon_i \text{ -component of } F_j. \end{array}$$

We then have

$$\begin{aligned} X_0^{(m)}(x;\alpha) \times \cdots \times X_{k_{l-1}}^{(m)}(x;\alpha) &= f(x,\alpha) \times \cdots \times f(2^{m-1}x, 2^{m-1}\alpha) \\ &= \Phi_1(x,y,\alpha,\epsilon_1,\epsilon_2)\Phi_1(T_f^m(x,y,\alpha,\epsilon_1,\epsilon_2)) \\ &= \widetilde{\Phi}_1(\zeta_0(x,y,\alpha,\epsilon_1,\epsilon_2))\widetilde{\Phi}_1(\zeta_m(x,y,\alpha,\epsilon_1,\epsilon_2)). \end{aligned}$$

Note that the right hand side of this does not dependent on $(y,\ \epsilon_1,\ \epsilon_2)$. Hence

$$\mathbf{E}\left[X_0^{(m)}(\,\bullet\,;\alpha) \times \cdots \times X_{k_{l-1}}^{(m)}(\,\bullet\,;\alpha)\right] = \int_{\mathbb{T}^1} dx\, \widetilde{\Phi}_1(\zeta_0(x,y,\alpha,\epsilon_1,\epsilon_2))\widetilde{\Phi}_1(\zeta_m(x,y,\alpha,\epsilon_1,\epsilon_2)).$$

We calculate the following.

$$\begin{aligned} &\int_{\mathbb{T}^1} d\alpha \left(\int_{\mathbb{T}^1} dx\, \widetilde{\Phi}_1(\zeta_0(x,y,\alpha,\epsilon_1,\epsilon_2))\widetilde{\Phi}_1(\zeta_m(x,y,\alpha,\epsilon_1,\epsilon_2)) \right)^2 \\ &\quad = \int_{\mathbb{T}^1} d\alpha \left(\int_{\mathbb{T}^1} dx\, \widetilde{\Phi}_1(\zeta_0(x,y,\alpha,\epsilon_1,\epsilon_2))\widetilde{\Phi}_1(\zeta_m(x,y,\alpha,\epsilon_1,\epsilon_2)) \right. \\ &\qquad \left. \times \int_{\mathbb{T}^1} dy\, \widetilde{\Phi}_2(\zeta_0(x,y,\alpha,\epsilon_1,\epsilon_2))\widetilde{\Phi}_2(\zeta_m(x,y,\alpha,\epsilon_1,\epsilon_2)) \right). \end{aligned} \tag{4.64}$$

If we fix α, then $\Phi_1(\zeta_m(x,y,\alpha,\epsilon_1,\epsilon_2))$ and $\Phi_2(\zeta_m(x,y,\alpha,\epsilon_1,\epsilon_2))$ are random variables with respect to $(x,y,\epsilon_1,\epsilon_2)$ which are independent under the probability measure $\mathbb{P}^2 \otimes (\delta_{-1} +$

[†21] The partition $\{F_j\}_j$ is called a *Markov partition*, and the dynamical system (Ω, T_f) is called a *Markov transformation*. For details about Markov chain, see [1] Section 8.

$\delta_1)/2 \otimes (\delta_{-1} + \delta_1)/2$. Therefore the value of (4.64) becomes

$$\begin{aligned} &= \int_{\mathbb{T}^1} d\alpha \left(\int_{\mathbb{T}^2} dxdy\, \widetilde{\Phi}_1(\zeta_0(x,y,\alpha,\epsilon_1,\epsilon_2))\widetilde{\Phi}_1(\zeta_m(x,y,\alpha,\epsilon_1,\epsilon_2)) \right. \\ &\qquad \left. \times \widetilde{\Phi}_2(\zeta_0(x,y,\alpha,\epsilon_1,\epsilon_2))\widetilde{\Phi}_2(\zeta_m(x,y,\alpha,\epsilon_1,\epsilon_2)) \right) \\ &= \int_{\Omega} d\mu\, \widetilde{\Phi}_3(\zeta_0(x,y,\alpha,\epsilon_1,\epsilon_2))\widetilde{\Phi}_3(\zeta_m(x,y,\alpha,\epsilon_1,\epsilon_2)) \\ &= \sum_{i,j} \widetilde{\Phi}_3(i)\widetilde{\Phi}_3(j)p^m(i,j)\mu(F_i), \end{aligned}$$

where $\widetilde{\Phi}_3 := \widetilde{\Phi}_1 \times \widetilde{\Phi}_2$. Because of Corollary 4.33, as $m \to \infty$, we have

$$\begin{aligned} \sum_{i,j} \widetilde{\Phi}_3(i)\widetilde{\Phi}_3(j)p^m(i,j)\mu(F_i) \;\longrightarrow\; \sum_{i,j} \widetilde{\Phi}_3(i)\widetilde{\Phi}_3(j)\mu(F_j)\mu(F_i) &= \left(\sum_i \widetilde{\Phi}_3(i)\mu(F_i) \right)^2 \\ = \left(\int_\Omega \epsilon_1\epsilon_2\, d\mu \right)^2 &= 0, \end{aligned}$$

which convergence takes place at an exponential rate in m. Hence we have

$$\begin{aligned} &\sum_{m=1}^{\infty} \int_{\mathbb{T}^1} d\alpha \left(\int_{\mathbb{T}^1} dx\, \widetilde{\Phi}_1(\zeta_0(x,y,\alpha,\epsilon_1,\epsilon_2))\widetilde{\Phi}_1(\zeta_m(x,y,\alpha,\epsilon_1,\epsilon_2)) \right)^2 \qquad (4.65) \\ &\qquad = \sum_{m=1}^{\infty} \sum_{i,j} \widetilde{\Phi}_3(i)\widetilde{\Phi}_3(j)p^m(i,j)\mu(F_i) < \infty, \end{aligned}$$

where each term of (4.65) decays exponentially in m. Therefore there exists $0 < \rho_1 < 1$ such that

$$\sum_{m=1}^{\infty} \rho_1^{-m} \int_{\mathbb{T}^1} d\alpha \left(\int_{\mathbb{T}^1} dx\, \widetilde{\Phi}_1(\zeta_0(x,y,\alpha,\epsilon_1,\epsilon_2))\widetilde{\Phi}_1(\zeta_m(x,y,\alpha,\epsilon_1,\epsilon_2)) \right)^2 < \infty. \qquad (4.66)$$

Consequently,

$$\int_{\mathbb{T}^1} d\alpha \sum_{m=1}^{\infty} \left(\rho_1^{-m/2} \int_{\mathbb{T}^1} dx\, \widetilde{\Phi}_1(\zeta_0(x,y,\alpha,\epsilon_1,\epsilon_2))\widetilde{\Phi}_1(\zeta_m(x,y,\alpha,\epsilon_1,\epsilon_2)) \right)^2 < \infty.$$

Finally, we see

$$\rho_1^{-m/2} \int_{\mathbb{T}^1} dx\, \widetilde{\Phi}_1(\zeta_0(x,y,\alpha,\epsilon_1,\epsilon_2))\widetilde{\Phi}_1(\zeta_m(x,y,\alpha,\epsilon_1,\epsilon_2)) \to 0, \quad m \to \infty, \quad \text{a.e.}\alpha. \qquad (4.67)$$

This shows that for almost every α, (4.58) holds. □

Proof of Markov property

Now, let us begin to prove Lemma 4.32. First of all, since T_f preserves μ, the stationarity of the precess $\{\zeta_m\}_{m=0}^{\infty}$ is obvious. By this stationarity, we have

$$\mu(\zeta_m = j \,|\, \zeta_{m-1} = i) = p(i,j), \quad i,j = 1,\dots,4J, \quad m \in \mathbb{N}^+.$$

In the proof below, we use the following abbreviations.

$$T_f^{-m} := (T_f^m)^{-1}, \quad \beta^{-m} := (\beta^m)^{-1}. \tag{4.68}$$

Lemma 4.34 *Let Δ^{-m} denote the partition of Ω consisting of all the connected components $T_f^{-m}F_j$, $j = 1, \dots, 4J$. Then, if $m' < m$, Δ^{-m} is a refinement of $\Delta^{-m'}$, i.e., for any $A \in \Delta^{-m}$ and $A' \in \Delta^{-m'}$, we have either $A \subset A'$ or $A \cap A' = \emptyset$.*

Proof. Let $\tilde{C} := \cup_{\epsilon_1,\epsilon_2=-1,1} C \times \{\epsilon_1\} \times \{\epsilon_2\}$.[†22] Then $T_f\tilde{C} \subset \tilde{C}$. If there are $A \in \Delta^{-m}$, A', $B' \in \Delta^{-m'}$ ($B' \neq A'$) such that $A \cap A' \neq \emptyset$ and $A \cap B' \neq \emptyset$, then $A \cap T_f^{-m'}\tilde{C} \neq \emptyset$. Therefore

$$\emptyset \neq T_f^m(A \cap T_f^{-m'}\tilde{C}) \subset T_f^m A \cap T_f^m T_f^{-m'}\tilde{C} = T_f^m A \cap T_f^{m-m'}\tilde{C} \subset T_f^m A \cap \tilde{C}.$$

But this is impossible, because $T_f^m(A) = F_j$ for some j. □

Let us show the Markov property of $\{\zeta_m\}_{m=0}^\infty$. Assume $m \geq 2$ and $\mu(\zeta_0 = i_0, \dots, \zeta_{m-1} = i_{m-1}) > 0$.

$$\begin{aligned}
&\mu(\zeta_m = j \,|\, \zeta_0 = i_0, \dots, \zeta_{m-1} = i_{m-1}) \\
&\quad = \mu\left(T_f^{-m}F_j \,\middle|\, F_{i_0} \cap T_f^{-1}F_{i_1} \cap \cdots \cap T_f^{-m+1}F_{i_{m-1}}\right) \\
&\quad = \frac{\mu\left(F_{i_0} \cap T_f^{-1}F_{i_1} \cap \cdots \cap T_f^{-m+1}F_{i_{m-1}} \cap T_f^{-m}F_{i_m}\right)}{\mu\left(F_{i_0} \cap T_f^{-1}F_{i_1} \cap \cdots \cap T_f^{-m+1}F_{i_{m-1}}\right)} \\
&\quad = \frac{\mu\left(F_{i_0} \cap T_f^{-1}F_{i_1} \cap \cdots \cap T_f^{-m+1}\left(F_{i_{m-1}} \cap T_f^{-1}F_{i_m}\right)\right)}{\mu\left(F_{i_0} \cap T_f^{-1}F_{i_1} \cap \cdots \cap T_f^{-m+1}F_{i_{m-1}}\right)}.
\end{aligned}$$

The set $T_f^{-m+1}F_{i_{m-1}}$ consists of 8^{m-1} connected components of equal measure. It follows from Lemma 4.34 that some of them, say l connected components, are included in $F := F_{i_0} \cap T_f^{-1}F_{i_1} \cap \cdots \cap T_f^{-m+2}F_{m-2}$, and the other $8^{m-1} - l$ ones are outside of F. The situation is the same for $T_f^{-m+1}\left(F_{i_{m-1}} \cap T_f^{-1}F_{i_m}\right)$ as well. Therefore

$$\begin{aligned}
\mu\left(F \cap T_f^{-m+1}\left(F_{i_{m-1}} \cap T_f^{-1}F_{i_m}\right)\right) &= \frac{l}{8^{m-1}}\mu\left(T_f^{-m+1}\left(F_{i_{m-1}} \cap T_f^{-1}F_{i_m}\right)\right), \\
\mu\left(F \cap T_f^{-m+1}F_{i_{m-1}}\right) &= \frac{l}{8^{m-1}}\mu\left(T_f^{-m+1}F_{i_{m-1}}\right).
\end{aligned}$$

Using μ-invariance of T_f, we have

$$\begin{aligned}
\mu(\zeta_m = j \,|\, \zeta_0 = i_0, \dots, \zeta_{m-1} = i_{m-1}) &= \frac{\mu\left(T_f^{-m+1}\left(F_{i_{m-1}} \cap T_f^{-1}F_{i_m}\right)\right)}{\mu\left(T_f^{-m+1}F_{i_{m-1}}\right)} \\
&= \frac{\mu\left(F_{i_{m-1}} \cap T_f^{-1}F_{i_m}\right)}{\mu(F_{i_{m-1}})} \\
&= \mu(\zeta_m = j \,|\, \zeta_{m-1} = i_{m-1}),
\end{aligned}$$

which proves the Markov property of $\{\zeta_m\}_{m=0}^\infty$. □

[†22] C is the set defined by (4.63).

Proof of ergodicity

The irreducibility follows from the ergodicity of T_f. Let us begin with the following Lemma.

Lemma 4.35 *Let $\phi_i : \mathbb{T}^3 \to \mathbb{C}$, $i = 1, 2$, be measurable functions which satisfy*

$$\phi_1(x, y, \alpha) = \phi_1(2x, 2y, 2\alpha) f(x, \alpha), \qquad a.e., \tag{4.69}$$
$$\phi_2(x, y, \alpha) = \phi_2(2x, 2y, 2\alpha) f(x, \alpha) f(y, \alpha), \qquad a.e. \tag{4.70}$$

Then we have $\phi_1 = \phi_2 = 0$, a.e.

Proof. Let $i = 1, 2$. Since f has no 0's, (4.69) shows that the set of 0's of $\phi_i(x, y, \alpha)$ and the set of 0's of $\phi_i(2x, 2y, 2\alpha)$ coincide, which means it is β-invariant, and hence, of Lebesgue measure 0 or 1 by the ergodicity of β. If its measure is 1, then the proof ends. So suppose that $\phi_i \neq 0$, a.e.. If ϕ_1 and ϕ_2 satisfy (4.69) and (4.70), then the signs of the real parts of ϕ_1 and ϕ_2 also satisfy (4.69) and (4.70). Therefore we may assume that $\phi_i \in \{-1, 1\}$.

We pay attention to the following subset of $\mathbb{T}^3$.

$$A := \left\{ (x, y, \alpha) \;\middle|\; \begin{array}{lll} \frac{1}{2} < x < 1, & \frac{1}{2} < x + k_{l-2}\alpha < 1, & 1 < x + k_{l-1}\alpha < \frac{3}{2}, \\ \frac{1}{2} < y < 1, & \frac{1}{2} < y + k_{l-1}\alpha < 1 & \end{array} \right\}. \tag{4.71}$$

It is easy to see that A is a non-empty domain. If $(x, y, \alpha) \in A$ then

$$\begin{aligned} r_1(x) &= r_1(x + k_1\alpha) = \cdots = r_1(x + k_{l-2}\alpha) = -1, \quad r_1(x + k_{l-1}\alpha) = 1, \\ r_1(y) &= r_1(y + k_1\alpha) = \cdots = r_1(y + k_{l-1}\alpha) = -1, \end{aligned}$$

so, by the fact that l is even and the definition of f (4.59), we see

$$f(x, \alpha) = -1, \quad f(y, \alpha) = 1, \quad (x, y, \alpha) \in A. \tag{4.72}$$

Let us recall the abbreviations (4.68).

$$\beta^{-m} A := \{ (x, y, \alpha) \in \mathbb{T}^3 \,|\, \beta^m(x, y, \alpha) \in A \}.$$

Each connected component of $\beta^{-1}A$ is similar to A itself, in particular, the set

$$B_0^{(-1)} := \left\{ (x, y, \alpha) \;\middle|\; \begin{array}{lll} \frac{3}{4} < x < 1, & \frac{3}{4} < x + k_{l-2}\alpha < 1, & 1 < x + k_{l-1}\alpha < \frac{5}{4}, \\ \frac{3}{4} < y < 1, & \frac{3}{4} < y + k_{l-1}\alpha < 1 & \end{array} \right\}.$$

is a subset of A. So, by (4.72), we have $f(x, \alpha) = -1$, $f(y, \alpha) = 1$ on $B_0^{(-1)}$.

Now, the given equations (4.69) and (4.70) imply that

$$\phi_i(x, y, \alpha)\phi_i(2x, 2y, 2\alpha) = -1, \qquad (x, y, \alpha) \in B_0^{(-1)}. \tag{4.73}$$

If $\phi_i(x, y, \alpha) \equiv 1$, a.e. on A, we would have $\phi_i(2x, 2y, 2\alpha) \equiv 1$, a.e. on $B_0^{(-1)}$, which contradicts (4.73). Therefore $\phi_i \not\equiv 1$ on A. Similarly, $\phi_i \not\equiv -1$ on A. Hence

$$\frac{1}{|A|} \int_A \phi_i(x, y, \alpha) dx dy d\alpha =: a_i \in (-1, 1), \tag{4.74}$$

where $|A|$ stands for the Lebesgue measure of A. Next, Lemma 4.34 implies that, for any m, $f(x,\alpha)$ and $f(y,\alpha)$ are constant on any connected component $B^{(-m)}$ of $\beta^{-m}A$. Therefore, by (4.69) and (4.70), on $B^{(-m)}$,

$$\phi_i(x,y,\alpha)\phi_i(2x,2y,2\alpha) \equiv 1 \quad \text{or} \quad \phi_i(x,y,\alpha)\phi_i(2x,2y,2\alpha) \equiv -1. \tag{4.75}$$

Let us show

$$\frac{1}{\left|B^{(-m)}\right|}\int_{B^{(-m)}} \phi_i(x,y,\alpha)dxdyd\alpha = \pm a_i \tag{4.76}$$

by induction. First, when $m = 1$, by (4.75) and a change of variables $x' = 2x$, $y' = 2y$, $\alpha' = 2\alpha$,

$$\int_{B^{(-1)}} \phi_i(x,y,\alpha)dxdyd\alpha = \pm\int_{B^{(-1)}} \phi_i(2x,2y,2\alpha)dx = \pm\frac{1}{8}\int_A \phi_i(x',y',\alpha')dx'dy'd\alpha'. \tag{4.77}$$

Since $|B^{(-1)}| = \frac{1}{8}|A|$,

$$\frac{1}{\left|B^{(-1)}\right|}\int_{B^{(-1)}} \phi_i(x,y,\alpha)dxdyd\alpha = \pm a_i. \tag{4.78}$$

Next, assume (4.76) up to $m-1$. Then for m, by (4.75), in a similar way as (4.77), we see

$$\int_{B^{(-m)}} \phi_i(x,y,\alpha)dxdyd\alpha = \pm\int_{B^{(-m)}} \phi_i(2x,2y,2\alpha)dxdyd\alpha = \pm\frac{1}{8}\int_{\beta B^{(-m)}} \phi_i(x,y,\alpha)dxdyd\alpha.$$

Here $\beta B^{(-m)}$ is a connected component, say $B^{(-m+1)}$, of $\beta^{-m+1}A$. We also have $|B^{(-m)}| = \frac{1}{8}|B^{(-m+1)}|$, and hence

$$\frac{1}{\left|B^{(-m)}\right|}\int_{B^{(-m)}} \phi_i(x,y,\alpha)dxdyd\alpha = \pm\frac{1}{\left|B^{(-m+1)}\right|}\int_{B^{(-m+1)}} \phi_i(x,y,\alpha)dxdyd\alpha.$$

Thus (4.76) holds for any $m = 1, 2, \ldots$.

Now, the set $\cup_{m=1}^{\infty}\beta^{-m}A$ is dense in $\mathbb{T}^3$, and any cube of edge length $0 < \varepsilon < 1$ includes at least one connected component of $\beta^{-m}A$, if $m \geq \lfloor -\log_2 \varepsilon \rfloor + 2$. Consequently, there exists a $\delta > 0$ such that for any cube $S \subset \mathbb{T}^3$, it holds that

$$-1 + \delta < \frac{1}{|S|}\int_S \phi_i(x,y,\alpha)dxdyd\alpha < 1 - \delta. \tag{4.79}$$

On the other hand, since ϕ_i is a $\{-1,1\}$-valued measurable function, by Lebesgue's density theorem, it holds that, $S(x,y,\alpha;\varepsilon)$ being the cube of edge length $\varepsilon > 0$ and center (x,y,α), we have

$$\lim_{\varepsilon\to 0}\frac{1}{|S(x,y,\alpha;\varepsilon)|}\int_{S(x,y,\alpha;\varepsilon)} \phi_i(x',y',\alpha')dx'dy'd\alpha' = -1 \text{ or } 1, \quad \text{a.e.}(x,y,\alpha) \in \mathbb{T}^3.$$

This contradicts (4.79). Thus $\phi_i \equiv 0$ a.e. □

Let us show the ergodicity of T_f to prove Lemma 4.32. We will show that if a measurable function $\phi : \Omega = \mathbb{T}^3 \times \{-1,1\}^2 \to \mathbb{C}$ is T_f-invariant, i.e.,

$$\phi(x,y,\alpha,\epsilon_1,\epsilon_2) = \phi(2x,2y,2\alpha,\epsilon_1 f(x,\alpha),\epsilon_2 f(y,\alpha)), \quad \mu\text{-a.e.}, \tag{4.80}$$

then $\phi \equiv$constant μ-a.e.

Let $\psi_1(x,y,\alpha) := \sum_{\epsilon_1,\epsilon_2} \phi(x,y,\alpha,\epsilon_1,\epsilon_2)$. Then

$$\begin{aligned}\psi_1(x,y,\alpha) &= \sum_{\epsilon_1,\epsilon_2} \phi(2x,2y,2\alpha,\epsilon_1 f(x,\alpha),\epsilon_2 f(y,\alpha)) \\ &= \sum_{\epsilon_1,\epsilon_2} \phi(2x,2y,2\alpha,\epsilon_1,\epsilon_2) \\ &= \psi_1(2x,2y,2\alpha),\end{aligned}$$

i.e., ψ_1 is invariant under the dyadic transformation β. By the ergodicity of β, we see that $\psi_1 \equiv c =$ constant, a.e. If we take $\phi - c/4$ instead of ϕ, then we have $\psi_1 \equiv 0$, a.e., so we may assume $\psi_1 \equiv 0$, a.e.

Next, let $\psi_2(x,y,\alpha,\epsilon_1) := \sum_{\epsilon_2} \phi(x,y,\alpha,\epsilon_1,\epsilon_2)$. Then

$$\begin{aligned}\psi_2(x,y,\alpha,\epsilon_1) &= \sum_{\epsilon_2} \phi(2x,2y,2\alpha,\epsilon_1 f(x,\alpha),\epsilon_2 f(y,\alpha)) \\ &= \sum_{\epsilon_2} \phi(2x,2y,2\alpha,\epsilon_1 f(x,\alpha),\epsilon_2) \\ &= \psi_2(2x,2y,2\alpha,\epsilon_1 f(x,\alpha)),\end{aligned}$$

i.e.,

$$\begin{aligned}\psi_2(x,y,\alpha,-1) &= \psi_2(2x,2y,2\alpha,-f(x,\alpha)) \\ &= \psi_2(2x,2y,2\alpha,-1)\mathbf{1}_{\{f(x,\alpha)=1\}} + \psi_2(2x,2y,2\alpha,1)\mathbf{1}_{\{f(x,\alpha)=-1\}}.\end{aligned}$$

Since $\psi_2(2x,2y,2\alpha,-1) + \psi_2(2x,2y,2\alpha,1) = \psi_1(x,y,\alpha) \equiv 0$,

$$\begin{aligned}\psi_2(x,y,\alpha,-1) &= \psi_2(2x,2y,2\alpha,-1)\left(\mathbf{1}_{\{f(x,\alpha)=1\}} - \mathbf{1}_{\{f(x,\alpha)=-1\}}\right) \\ &= \psi_2(2x,2y,2\alpha,-1)f(x,\alpha).\end{aligned}$$

By Lemma 4.35, we see $\psi_2(x,y,\alpha,-1) \equiv 0$, a.e., and hence $\psi_2(x,y,\alpha,1) \equiv 0$, a.e. Thus we see $\psi_2(x,y,\alpha,\epsilon_1) \equiv 0$, a.e.$(x,y,\alpha,\epsilon_1)$. Consequently, by definition,

$$\phi(x,y,\alpha,\epsilon_1,-1) + \phi(x,y,\alpha,\epsilon_1,1) = \psi_2 \equiv 0, \quad \text{a.e.},$$

i.e., we can write as

$$\phi(x,y,\alpha,\epsilon_1,\epsilon_2) = \phi(x,y,\alpha,\epsilon_1,1)\epsilon_2.$$

Now exchanging the roles of ϵ_1 and ϵ_2, the same argument leads to

$$\phi(x,y,\alpha,\epsilon_1,\epsilon_2) = \phi(x,y,\alpha,1,\epsilon_2)\epsilon_1, \quad \text{a.e.}$$

From the last two equalities, it immediately follows that

$$\phi(x,y,\alpha,\epsilon_1,\epsilon_2) = \phi(x,y,\alpha,1,1)\epsilon_1\epsilon_2, \quad \text{a.e.}$$

Therefore, ϕ is T_f-invariant (4.80), if and only if

$$\phi(x,y,\alpha,1,1) = \phi(2x,2y,2\alpha,1,1)f(x,\alpha)f(y,\alpha), \quad \text{a.e.}$$

Then Lemma 4.35 implies $\phi \equiv 0$, a.e., thus T_f is ergodic, and hence $\{\zeta_m\}_{m=0}^{\infty}$ is irreducible. □

Proof of aperiodicity

Now we have only to prove the aperiodicity of the process $\{\zeta_m\}_{m=0}^{\infty}$ to prove Lemma 4.32. Since we proved the irreducibility, it is sufficient to show that $\mu(\zeta_0 = \zeta_1) > 0$.

We first set

$$\begin{aligned} F' &:= \left\{ (x,y,\alpha) \;\middle|\; \begin{array}{ll} 0 < x < \frac{1}{2}, & 0 < x + k_{l-1}\alpha < \frac{1}{2}, \\ 0 < y < \frac{1}{2}, & 0 < y + k_{l-1}\alpha < \frac{1}{2} \end{array} \right\}, \\ F &:= F' \times \{1\} \times \{1\} \subset \Omega, \\ F'' &:= \left\{ (x,y,\alpha) \;\middle|\; \begin{array}{ll} 0 < x < \frac{1}{4}, & 0 < x + k_{l-1}\alpha < \frac{1}{4}, \\ 0 < y < \frac{1}{4}, & 0 < y + k_{l-1}\alpha < \frac{1}{4} \end{array} \right\}, \\ H &:= F'' \times \{1\} \times \{1\} \subset \Omega. \end{aligned}$$

Then $F = F_j$ for some $j = 1, \ldots, 4J$, and $H \subset F$. Since $f(x,\alpha) = f(y,\alpha) = 1$ for $(x,y,\alpha) \in F''$, we see

$$\forall (x,y,\alpha,\epsilon_1,\epsilon_2) \in H, \quad \zeta_0(x,y,\alpha,\epsilon_1,\epsilon_2) = \zeta_1(x,y,\alpha,\epsilon_1,\epsilon_2) = j.$$

Since $\mu(H) > 0$, we finally see that $\{\zeta_m\}_{m=0}^{\infty}$ is aperiodic.

Thus we have completed the proof of Lemma 4.32, and hence Theorem 4.11′ (Theorem 4.11). □

4.3.5 Precise estimate of exponential decay of two-term correlation

We will estimate the exponent ρ of the convergence rate appeared in Theorem 4.11′ (Theorem 4.11). For the two-term correlation, we have the following theorem.[†23]

Theorem 4.36 *(cf. [41]) For any* $\rho > \rho_0 := \sqrt{\left(1 + \sqrt{17}\right)/8} = 0.80024\ldots$*, it holds that*

$$\mathbf{E}\left[X_0^{(m)}(\bullet\,;\alpha) X_k^{(m)}(\bullet\,;\alpha)\right] = o(\rho^m), \quad m \to \infty, \quad k \in \mathbb{N}^+, \quad \text{a.e. } \alpha.$$

In fact, more precisely, the following equality holds.

Theorem 4.37

$$\begin{aligned} &\int_{\mathbb{T}^1} d\alpha \left(\mathbf{E}\left[X_0^{(m)}(\bullet\,;\alpha) X_k^{(m)}(\bullet\,;\alpha)\right]\right)^2 \\ &\quad = \left(\frac{1}{2} + \frac{5\sqrt{17}}{102}\right)\left(\frac{1+\sqrt{17}}{8}\right)^m + \left(\frac{1}{2} - \frac{5\sqrt{17}}{102}\right)\left(\frac{1-\sqrt{17}}{8}\right)^m, \quad m, k \in \mathbb{N}^+. \end{aligned}$$

[†23] Takanobu determined ρ for a special four-term correlation. (a private communication)

Theorem 4.37 implies Theorem 4.36, which can be shown just like (4.66) implies (4.67). The constant ρ_0 in Theorem 4.36 is best possible, i.e., it cannot be taken smaller.

Proof of Theorem 4.37. We use a recursion formula to prove the theorem ([41]). First, note that since the Lebesgue measure is invariant under the transformation $\mathbb{T}^1 \ni x \mapsto kx \in \mathbb{T}^1$, it is sufficient to prove it for $k = 1$ only.

Now, for each $m \in \mathbb{N}^+$, set

$$a_m := \int_{\mathbb{T}^1} d\alpha \left(\mathbf{E}\left[X_0^{(m)}(\,\bullet\,;\alpha) X_1^{(m)}(\,\bullet\,;\,\alpha) \right] \right)^2 = \int_{\mathbb{T}^1} d\alpha \left(\int_{\mathbb{T}^1} \prod_{i=1}^{m} r_i(x) r_i(x+\alpha)\, dx \right)^2 .$$

By showing the following two equality, we prove Theorem 4.37.

$$a_1 = a_2 = \frac{1}{3} \tag{4.81}$$

$$a_{m+2} = \frac{1}{4} a_{m+1} + \frac{1}{4} a_m, \quad m \in \mathbb{N}^+ \tag{4.82}$$

Proof of (4.81). Using

$$\int_{\mathbb{T}^1} r_1(x) r_1(x+\alpha) dx = |2 - 4\alpha| - 1, \tag{4.83}$$

we have

$$a_1 = \int_{\mathbb{T}^1} d\alpha \left(\int_{\mathbb{T}^1} r_1(x) r_1(x+\alpha) dx \right)^2 = \int_{\mathbb{T}^1} \left(|2 - 4\alpha| - 1 \right)^2 d\alpha = \frac{1}{3}.$$

Next, note that $r_1(x) r_2(x) = r_1\left(x + \frac{1}{4}\right)$, and we see

$$\begin{aligned} a_2 &= \int_{\mathbb{T}^1} d\alpha \left(\int_{\mathbb{T}^1} r_1(x) r_2(x) r_1(x+\alpha) r_2(x+\alpha) dx \right)^2 \\ &= \int_{\mathbb{T}^1} d\alpha \left(\int_{\mathbb{T}^1} r_1\left(x + \frac{1}{4}\right) r_1\left(x + \alpha + \frac{1}{4}\right) dx \right)^2 \\ &= \int_{\mathbb{T}^1} d\alpha \left(\int_{\mathbb{T}^1} r_1(x) r_1(x+\alpha) dx \right)^2 = a_1. \end{aligned}$$

Proof of (4.82). Imitating (4.29), we define

$$A^{(m)}(\alpha) := \int_{\mathbb{T}^1} \prod_{i=1}^{m} r_i(x) r_i(x+\alpha)\, dx.$$

In this section below, we let $\mathbf{E}$ denote the mean (Lebesgue integral) with respect to α. In particular, we have $a_m = \mathbf{E}\left[A^{(m)}(\alpha)^2 \right]$. Set

$$\begin{cases} \xi_m := \mathbf{E}\left[A(\alpha^{(m)U})^2 + A(\alpha^{(m)L})^2 \right], \\ \eta_m := \mathbf{E}\left[A(\alpha^{(m)U}) A(\alpha^{(m)L}) \right]. \end{cases}$$

Using these quantities, we give a heuristic method to find the recursion formula (4.82).

Lemma 4.38

$$a_m = \frac{1}{3}\xi_m + \frac{1}{3}\eta_m, \tag{4.84}$$

$$\begin{pmatrix}\xi_{m+1}\\ \eta_{m+1}\end{pmatrix} = \begin{pmatrix}\frac{3}{4} & \frac{1}{2}\\ -\frac{1}{4} & -\frac{1}{2}\end{pmatrix}\begin{pmatrix}\xi_m\\ \eta_m\end{pmatrix}. \tag{4.85}$$

Proof. By Theorem 4.13′,

$$A^{(m)}(\alpha) = (1 - 2^m\langle\alpha\rangle_m)\, A(\alpha^{(m)L}) + 2^m\langle\alpha\rangle_m A(\alpha^{(m)U}),$$

where $A(\bullet)$ is the function given in Definition 4.20, and $\langle\alpha\rangle_m := \alpha - \lfloor\alpha\rfloor_m$. From this equality, it follows that

$$\begin{aligned} a_m &= \mathbf{E}\left[(1 - 2^m\langle\alpha\rangle_m)^2\, A(\alpha^{(m)L})^2\right] + \mathbf{E}\left[(2^m\langle\alpha\rangle_m)^2 A(\alpha^{(m)U})^2\right] \\ &\quad +2\mathbf{E}\left[(1 - 2^m\langle\alpha\rangle_m)\, A(\alpha^{(m)L})(2^m\langle\alpha\rangle_m)A(\alpha^{(m)U})\right] \\ &= \mathbf{E}\left[(1 - 2^m\langle\alpha\rangle_m)^2\right]\mathbf{E}\left[A(\alpha^{(m)L})^2\right] + \mathbf{E}\left[(2^m\langle\alpha\rangle_m)^2\right]\mathbf{E}\left[A(\alpha^{(m)U})^2\right] \\ &\quad +2\mathbf{E}\left[(1 - 2^m\langle\alpha\rangle_m)\,(2^m\langle\alpha\rangle_m)\right]\mathbf{E}\left[A(\alpha^{(m)U})A(\alpha^{(m)L})\right], \end{aligned}$$

where we used the independence of $\langle\alpha\rangle_m$ and $\alpha^{(m)L}$ or $\alpha^{(m)U}$. Since $2^m\langle\alpha\rangle_m = \langle 2^m\alpha\rangle$, its distribution is equal to the distribution of α itself, i.e., the uniform distribution, and hence,

$$\begin{aligned} a_m &= \mathbf{E}\left[(1-\alpha)^2\right]\mathbf{E}\left[A(\alpha^{(m)L})^2\right] + \mathbf{E}\left[\alpha^2\right]\mathbf{E}\left[A(\alpha^{(m)U})^2\right] \\ &\quad +2\mathbf{E}\left[(1-\alpha)\alpha\right]\mathbf{E}\left[A(\alpha^{(m)U})A(\alpha^{(m)L})\right] \\ &= \frac{1}{3}\mathbf{E}\left[A(\alpha^{(m)U})^2\right] + \frac{1}{3}\mathbf{E}\left[A(\alpha^{(m)L})^2\right] + \frac{1}{3}\mathbf{E}\left[A(\alpha^{(m)U})A(\alpha^{(m)L})\right], \end{aligned}$$

which shows (4.84). Next, by Lemma 4.21,

$$\begin{aligned} &\mathbf{E}\left[A(\alpha^{(m+1)U})^2 + A(\alpha^{(m+1)L})^2\right] \\ &= \mathbf{E}\left[\frac{1}{4}\left(A(\mathcal{U}\alpha^{(m+1)U}) + A(\mathcal{L}\alpha^{(m+1)U})\right)^2 + \frac{1}{4}\left(A(\mathcal{U}\alpha^{(m+1)L}) + A(\mathcal{L}\alpha^{(m+1)L})\right)^2\right] \\ &= \mathbf{E}\left[\frac{1}{4}\left(A(\alpha^{(m)U}) + A(\alpha^{(m)L})\right)^2 + A(\alpha^{(m)L})^2\,;\, d_{m+1}(\alpha) = 0\right] \\ &\quad + \mathbf{E}\left[A(\alpha^{(m)U})^2 + \frac{1}{4}\left(A(\alpha^{(m)U}) + A(\alpha^{(m)L})\right)^2\,;\, d_{m+1}(\alpha) = 1\right] \\ &= \frac{1}{2}\mathbf{E}\left[\frac{1}{4}\left(A(\alpha^{(m)U}) + A(\alpha^{(m)L})\right)^2 + A(\alpha^{(m)L})^2\right] \\ &\quad + \frac{1}{2}\mathbf{E}\left[A(\alpha^{(m)U})^2 + \frac{1}{4}\left(A(\alpha^{(m)U}) + A(\alpha^{(m)L})\right)^2\right] \\ &= \frac{3}{4}\mathbf{E}\left[A(\alpha^{(m)U})^2 + A(\alpha^{(m)L})^2\right] + \frac{1}{2}\mathbf{E}\left[A(\alpha^{(m)U})A(\alpha^{(m)L})\right]. \end{aligned}$$

And Similarly,

$$\begin{aligned}
&\mathbf{E}\left[A(\alpha^{(m+1)U})A(\alpha^{(m+1)L})\right] \\
&= \mathbf{E}\left[\left(-\frac{1}{2}A(\alpha^{(m)U}) - \frac{1}{2}A(\alpha^{(m)L})\right)A(\alpha^{(m)L})\,;\, d_{m+1}(\alpha) = 0\right] \\
&\quad + \mathbf{E}\left[A(\alpha^{(m)U})\left(-\frac{1}{2}A(\alpha^{(m)U}) - \frac{1}{2}A(\alpha^{(m)L})\right)\,;\, d_{m+1}(\alpha) = 1\right] \\
&= -\frac{1}{4}\mathbf{E}\left[\left(A(\alpha^{(m)U}) + A(\alpha^{(m)L})\right)A(\alpha^{(m)L})\right] \\
&\quad - \frac{1}{4}\mathbf{E}\left[A(\alpha^{(m)U})\left(A(\alpha^{(m)U}) + A(\alpha^{(m)L})\right)\right] \\
&= -\frac{1}{4}\mathbf{E}\left[A(\alpha^{(m)U})^2 + A(\alpha^{(m)L})^2\right] - \frac{1}{2}\mathbf{E}\left[A(\alpha^{(m)U})A(\alpha^{(m)L})\right].
\end{aligned}$$

From these (4.85) follows. □

Now, let us go back to the proof of (4.82). First, by (4.85),

$$\begin{pmatrix}\xi_{m+2}\\ \eta_{m+2}\end{pmatrix} = \begin{pmatrix} \frac{3}{4} & \frac{1}{2} \\ -\frac{1}{4} & -\frac{1}{2}\end{pmatrix}^2\begin{pmatrix}\xi_m\\ \eta_m\end{pmatrix} = \begin{pmatrix} \frac{7}{16} & \frac{1}{8} \\ -\frac{1}{16} & \frac{1}{8}\end{pmatrix}\begin{pmatrix}\xi_m\\ \eta_m\end{pmatrix},$$

and hence

$$\begin{aligned}
a_{m+2} &= \frac{1}{3}\xi_{m+2} + \frac{1}{3}\eta_{m+2} \\
&= \frac{1}{3}\left(\frac{7}{16}\xi_m + \frac{1}{8}\eta_m\right) + \frac{1}{3}\left(-\frac{1}{16}\xi_m + \frac{1}{8}\eta_m\right) \\
&= \frac{1}{8}\xi_m + \frac{1}{12}\eta_m. \qquad (4.86)
\end{aligned}$$

Similarly,

$$a_{m+1} = \frac{1}{3}\xi_{m+1} + \frac{1}{3}\eta_{m+1} = \frac{1}{6}\xi_m. \qquad (4.87)$$

Then to find constants c_1, c_2 such that

$$a_{m+2} = c_1 a_{m+1} + c_2 a_m, \quad m \in \mathbb{N}^+,$$

because of (4.84), (4.86) and (4.87), we have to solve

$$\begin{aligned}
\frac{1}{8}\xi_m + \frac{1}{12}\eta_m &= c_1\frac{1}{6}\xi_m + c_2\left(\frac{1}{3}\xi_m + \frac{1}{3}\eta_m\right) \\
&= \left(\frac{1}{6}c_1 + \frac{1}{3}c_2\right)\xi_m + \frac{1}{3}c_2\eta_m, \quad m \in \mathbb{N}^+.
\end{aligned}$$

Comparing the coefficients of both hand sides, we know that

$$c_1 = c_2 = \frac{1}{4}.$$

This completes the proof of (4.82), and consequently, the proof of Theorem 4.37. □