Journal of Combinatorics
Volume 1, Number 2, 171–232, 2010

Lattice point counting and the probabilistic method

József Beck

Dedicated to the 60th birthday of Joel Spencer

We take a quantitative approach based on probability theory to several number theoretic problems, which all have the common form of counting lattice points in some nice domain. It is well-known that the number of solutions to Pell equations can be counted with a bounded error term. We relax Pell equations to (inhomogeneous) Pell inequalities and study the corresponding question. A naive area principle (=the number of lattice points in and the area of nice domains are close) guides the intuition for the answer, but the intuition is sometimes correct, sometimes not. On the one hand, the intuition fails for continuum many translated copies of the corresponding hyperbolic domain (Theorem 1). On the other hand, the intuition is correct for almost all translated copies (Theorems 2 and 3).

1. From Pell's equation to the Area Principle

1.1.

It is hard to overestimate the importance of the book *Probabilistic Methods in Combinatorics* by the late Paul Erdős and Joel Spencer, published in 1974. This little book is only about 100 pages long, and was published by a minor publisher (*Akadémiai Kiadó, Budapest*); nevertheless, it played an absolutely crucial role in popularizing the so-called "probabilistic method", which since became (arguably) the single most important method in discrete mathematics. (To convince the skeptics, it suffices to have a quick look at the wide range of applications of the method in very different areas from, say, geometry to computer science; see e.g. the later book *The Probabilistic Method* by N. Alon and Spencer.)

This paper is very much in the spirit of the Erdős–Spencer book. It demonstrates the power of the probabilistic method (Markov chains, second moment arguments, etc.) in a field—lattice point counting, (combinatorial) number theory—where these ideas were hardly used before, and the method quickly leads to elegant new results.

Our starting point is the famous Pell's equation. As an illustration, consider for example the special case $x^2 - 2y^2 = \pm 1$. It is well-known that this equation has infinitely many integral solutions; in fact, the set of all integral solutions $(x_k, y_k) \in \mathbb{Z}^2$ forms a cyclic group generated by the least positive solution:

$$x_k + y_k\sqrt{2} = \pm(1+\sqrt{2})^k, \quad k \in \mathbb{Z}.$$

All integral solutions of $x^2 - 2y^2 = 1$ are given by

$$x_k + y_k\sqrt{2} = \pm(1+\sqrt{2})^{2k},$$

and all of $x^2 - 2y^2 = -1$ by

$$x_k + y_k\sqrt{2} = \pm(1+\sqrt{2})^{2k+1}.$$

In particular, all positive integer solutions of $x^2 - 2y^2 = 1$ are given by

$$x_k + y_k\sqrt{2} = (1+\sqrt{2})^{2k} = (3+2\sqrt{2})^k, \quad k = 1, 2, 3, \ldots$$

Taking the algebraic conjugate $x_k - y_k\sqrt{2} = (3-2\sqrt{2})^k$, and adding/subtracting these two equations together, we obtain the explicit formula

(1.1)
$$x_k = \frac{(3+2\sqrt{2})^k + (3-2\sqrt{2})^k}{2} \qquad \text{and} \qquad y_k = (3-2\sqrt{2})^k - (3-2\sqrt{2})^k 2\sqrt{2}.$$

Since $0 < 3-2\sqrt{2} < 1$ (in fact, $0 < 3-2\sqrt{2} < 1/5$), we have

$$x_k = \text{the nearest integer to } \frac{1}{2(3+2\sqrt{2})^k}$$

and

$$y_k = \text{the nearest integer to } \frac{1}{2\sqrt{2}}(3+2\sqrt{2})^k.$$

If k is large, the error is very tiny. For example, the 10th solution of $x^2 - 2y^2 = 1$ in positive integers is the pair

$$x_{10} = 22,619,537 \quad \text{and} \quad y_{10} = 15,994,428.$$

On the other hand,

$$\frac{1}{2}(3+2\sqrt{2})^{10} = 22,619,536.99999998895\ldots$$

and

$$\frac{1}{2\sqrt{2}}(3+2\sqrt{2})^{10} = 15,994,428.000000007815\ldots\ .$$

Let $F(N) = F(\sqrt{2};1;N)$ denote the number of positive integer solutions of the Pell equation $x^2 - 2y^2 = 1$ up to N in the sense $x \ge 1$ and $1 \le y \le N$. (For the simplicity of notation it is convenient to focus on y instead of x.) We have

$$k \le F(N) \Longleftrightarrow (3+2\sqrt{2})^k - \frac{(3-2\sqrt{2})^k}{2\sqrt{2}} \le N,$$

which implies the asymptotic formula

$$F(N) = F(\sqrt{2};1;N) = \frac{\log N}{\log(3+2\sqrt{2})} + O(1). \tag{1.2}$$

Formula (1.2) says that the counting function $F(N) = F(\sqrt{2};1;N)$ has an extremely predictable/deterministic behavior: it is const $\cdot \log N$ plus some negligible *bounded* fluctuation.

Note that (1.2) has some far-reaching generalizations. Let $[\gamma_1,\gamma_2]$ be an arbitrary interval, and let $F(\sqrt{2};[\gamma_1,\gamma_2];N)$ denote the number of positive integer solutions of the Pell inequality $\gamma_1 \le x^2 - 2y^2 \le \gamma_2$, $x \ge 1$ and $1 \le y \le N$. By using the theory of indefinite binary quadratic forms, it is easy to prove the following analog of (1.2):

$$F(\sqrt{2};[\gamma_1,\gamma_2];N) = c_0(\sqrt{2};\gamma_1,\gamma_2)\cdot \log N + O(1), \tag{1.3}$$

where the constant factor $c_0(\sqrt{2};\gamma_1,\gamma_2)$ is independent of N.

What is more, we can switch from $\sqrt{2}$ to any other *quadratic irrational* α: it means that α is a root of a quadratic equation $Ax^2 + Bx + C = 0$ with integral coefficients such that the discriminant $B^2 - 4AC \ge 2$ is not a complete square. An equivalent definition is that $\alpha = (a+\sqrt{d})/b$ where a, b, d are integers with $b \ne 0$ and $d \ge 2$ is not a complete square. Note that the quadratic irrationals are perfectly characterized by their continued fraction: the continued fraction of α is (ultimately) periodic if and only if α is a quadratic irrational. For example,

$$\sqrt{2} = 1 + \cfrac{1}{2+\frac{1}{2+\cdots}} = [1;2,2,2,2,\ldots] = [1;\overline{2}]$$

$$\frac{24-\sqrt{15}}{17} = 1 + \cfrac{1}{5+\frac{1}{2+\cdots}} = [1;5,2,3,2,3,2,3,\ldots] = [1;5,\overline{2,3}].$$

Let's go back to (1.3), that is, to the special case $\alpha = \sqrt{2}$. For example, if $-2 < \gamma_1 \le -1 < 1 \le \gamma_2 < 2$, then

$$c_0(\sqrt{2}; \gamma_1, \gamma_2) = \frac{1}{\log(1+\sqrt{2})} = \frac{2}{\log(3+2\sqrt{2})}, \tag{1.4}$$

if $-1 < \gamma_1 \le 1 \le \gamma_2 < 2$, then

$$c_0(\sqrt{2}; \gamma_1, \gamma_2) = \frac{1}{\log(3+2\sqrt{2})}, \tag{1.5}$$

and finally, if $-1 < \gamma_1 \le \gamma_2 < 1$, then of course

$$c_0(\sqrt{2}; \gamma_1, \gamma_2) = 0. \tag{1.6}$$

1.2. The Naive Area Principle

It is very interesting to compare these well-known asymptotic results about the number of solutions of the Pell equation/inequality to what I like to call the "naive area principle". Perhaps the most natural guiding intuition in *lattice point theory* is the following: if a "nice region" has "large" area, then it should contain a "large" number of lattice points, and the actual number should be "close" to the area.

I refer to this vague intuition as the *Naive Area Principle.* Of course, the heart of the matter is how to define "nice region" precisely. Consider, for example, the infinite open horizontal strip of height one: $0 < y < 1$, $-\infty < x < \infty$; it has infinite area, but it contains no lattice point. I think the reader agrees that the infinite strip is a "nice region", so the Naive Area Principle is clearly violated here.

A less trivial example comes from the Pell inequality

$$-\frac{1}{2} \le x^2 - 2y^2 \le \frac{1}{2}, \tag{1.7}$$

which is a hyperbolic region of infinite area, and contains no lattice point except the origin. I think the reader agrees that the hyperbolic region (1.7) is also "nice", so this is again a violation of the Naive Area Principle.

Next we switch from (1.7) to the general Pell inequality

$$\gamma_1 \le x^2 - 2y^2 \le \gamma_2, \tag{1.8}$$

where $-\infty < \gamma_1 < \gamma_2 < \infty$ are arbitrary real numbers. Of course, the hyperbolic region (1.8) has infinite area; what we want to compute is the area of a finite segment. Consider the finite region

$$H(\sqrt{2};[\gamma_1,\gamma_2];N) = \Big\{(x,y)\in\mathbb{R}^2 :\ \gamma_1 \le x^2-2y^2 \le \gamma_2 \qquad \text{where } x\ge 1 \text{ and } 1\le y\le N\Big\}. \tag{1.9}$$

If N is large compared to γ_1,γ_2, then the finite region $H(\sqrt{2};[\gamma_1,\gamma_2];N)$ looks like a "hyperbolic needle".

It is easy to give a good estimation for the area of the "hyperbolic needle" $H(\sqrt{2};[\gamma_1,\gamma_2];N)$.

Lemma 1.1. *Let $\gamma_1<\gamma_2$, then*

$$\text{area}\Big(H(\sqrt{2};[\gamma_1,\gamma_2];N)\Big) = \frac{\gamma_2-\gamma_1}{2\sqrt{2}}\log N \ +\ O(1), \tag{1.10}$$

where the implicit constant in $O(1)$ is independent of N (but may depend on γ_1 and γ_2).

The proof of (1.10) is based on the familiar factorization

$$x^2-2y^2 = (x+y\sqrt{2})(x-y\sqrt{2}), \tag{1.11}$$

and on the routine computation of the Jacobian of the corresponding substitution (this explains the factor $2\sqrt{2}$ in the denominator in (1.10)). I postpone the details to Section 3.

Now let's return to the Naive Area Principle. Comparing (1.3) with (1.9)–(1.10), it is "reasonable" to expect—in view of the Naive Area Principle—that the counting function $F(\sqrt{2};[\gamma_1,\gamma_2];N)$ is "close" to the area of the hyperbolic needle $H(\sqrt{2};[\gamma_1,\gamma_2];N)$. In other words, it is "reasonable" to expect that

$$c_0(\sqrt{2};\gamma_1,\gamma_2) = \frac{\gamma_2-\gamma_1}{2\sqrt{2}}. \tag{1.12}$$

Unfortunately, the Naive Area Principle is "mostly" violated in the quantitative sense that (1.12) fails for the overwhelming majority of the choices $-\infty<\gamma_1<\gamma_2<\infty$. In fact, the left-hand side and the right-hand side of (1.12) have completely different behavior: the left-hand side of (1.12) has discrete jumps and the right-hand side is a continuous function of γ_1 and γ_2.

For example, as γ_1 and γ_2 run in the interval $-2 < \gamma_1 < \gamma_2 < 2$, the constant factor $c_0(\sqrt{2}; \gamma_1, \gamma_2)$ has only 3 possible values (see (1.4)–(1.6)):

$$0, \quad \frac{1}{\log(3+2\sqrt{2})}, \quad \frac{2}{\log(3+2\sqrt{2})}.$$

This shows—in a quantitative way—how the general Pell inequality (see (1.8))

$$\gamma_1 \leq x^2 - 2y^2 \leq \gamma_2$$

violates the Naive Area Principle.

1.3. Inhomogeneous case—extra large fluctuations

Using the familiar factorization (1.11), we can rewrite the Pell equation $x^2 - 2y^2 = \pm 1$, restricted to positive x, y, as follows:

(1.13)
$$|x^2 - 2y^2| \leq 1 \Longleftrightarrow |y\sqrt{2} - x| \cdot (y\sqrt{2} + x) \leq 1 \Longleftrightarrow \|y\sqrt{2}\| \cdot (y\sqrt{2} + x) \leq 1,$$

where $\|z\|$ denotes, as usual, the distance of a real number z from the nearest integer. Notice that in (1.13) x is the nearest integer to $y\sqrt{2}$ (=an irrational number, namely an integral multiple of $\sqrt{2}$, where the integer $y \geq 1$). Since $y\sqrt{2} \approx x$, (1.13) is "basically" equivalent to the "vague" inequality

$$\|y\sqrt{2}\| \leq \frac{1+o(1)}{2\sqrt{2}y}. \tag{1.14}$$

The vagueness of (1.14) comes from the additive term $o(1)$, which tends to 0 as $y \to \infty$. Formula (1.14) is more like a physicist's notation, but I am sure every mathematician understands it.

An expert in number theory would classify (1.14) as a typical problem in diophantine approximation. Next I give a nutshell summary of diophantine approximation.

The classical problem of the theory of diophantine approximation is to find "good" rational approximations of irrational numbers. More precisely, we want to decide whether an inequality

$$\|n\alpha\| < \frac{1}{n\varphi(n)} \Longleftrightarrow \left\|\alpha - \frac{m}{n}\right\| < \frac{1}{n^2 \cdot \varphi(n)}, \tag{1.15}$$

or in general,

$$\|n\alpha - \beta\| < \frac{1}{n\varphi(n)}, \tag{1.16}$$

where α is a given irrational and β is a given real number, has infinitely many integral solutions in n, and if this is the case, to determine the solutions, or at least determine the asymptotic number of integral solutions. As always, $\|z\|$ denotes the distance of a real z from the nearest integer, and $\varphi(n)$ is a positive increasing function of n.

(1.15) is called a homogeneous, and (1.16) is called an inhomogeneous (diophantine) inequality. For example, in the homogeneous case the best possible result is Hurwitz's well-known theorem: for any irrational α,

$$\|n\alpha\| < \frac{1}{\sqrt{5}n}$$

has infinitely many positive integer solutions.

In the inhomogeneous case we can mention an old result of Kronecker that, for any irrational α and for any real β,

$$\|n\alpha - \beta\| < \frac{3}{n}$$

has infinitely many positive integer solutions. Perhaps the strongest inhomogeneous result is Minkowski's theorem: for any irrational α,

$$\|n\alpha - \beta\| < \frac{1}{4|n|}$$

has infinitely many integer solutions (not necessarily positive), unless $0 < \beta < 1$ is an integral multiple of α modulo one.

The homogeneous case (1.15) has a complete theory based on continued fractions. These are classical results mostly due to Euler and Lagrange. Unfortunately, we know much less about the inhomogeneous case. Very recently I proved some new results in this direction: I basically covered the case where α is an arbitrary quadratic irrational and β is a typical real number. These results form a large part of my recent book Beck [Be010].

Before formulating some main results, first I need to elaborate on the connection between (homogeneous and inhomogeneous) diophantine inequalities such as (1.15)–(1.16) and (homogeneous and inhomogeneous) Pell inequalities.

1.4. Homogeneous and inhomogeneous Pell inequalities

The general form of a quadratic curve on the plane is

$$a_{11}x^2 + a_{12}xy + a_{22}y^2 + a_{13}x + a_{23}y + a_{33} = 0. \tag{1.17}$$

We are interested in the integral solutions $(x, y) \in \mathbb{Z}^2$ of an arbitrary inequality

$$\gamma_1 \le a_{11}x^2 + a_{12}xy + a_{22}y^2 + a_{13}x + a_{23}y \le \gamma_2, \tag{1.18}$$

where $\gamma_1 < \gamma_2$ are given real numbers. Equation (1.18) defines a plane region; the boundary consists of two curves of type (1.17). If the discriminant is negative: $D = a_{12}^2 - 4a_{11}a_{22} < 0$, then (1.18) defines a bounded region where the boundary curves are two ellipses. The case of positive discriminant $D = a_{12}^2 - 4a_{11}a_{22} > 0$ is much more interesting, because then (1.18) defines an unbounded region, where the boundary curves are two hyperbolas. Unboundedness means that we have a chance for infinitely many integral solutions of (1.18).

For simplicity assume that the coefficients a_{11}, a_{12}, a_{22} in (1.18) are integers and $D = a_{12}^2 - 4a_{11}a_{22} > 0$. We can factorize the quadratic part as follows:

$$a_{11}x^2 + a_{12}xy + a_{22}y^2 = a_{11}(x - \alpha y)(x - \alpha' y), \tag{1.19}$$

where

$$\alpha = \frac{-a_{12} + \sqrt{D}}{2a_{11}}, \qquad \alpha' = \frac{-a_{12} - \sqrt{D}}{2a_{11}}. \tag{1.20}$$

Using (1.19) we can rewrite (1.18) in the form

$$\gamma_1 \le (x - \alpha y + \rho_1)(x - \alpha' y + \rho_2) \le \gamma_2, \tag{1.21}$$

where

$$\rho_1 + \rho_2 = \frac{a_{13}}{a_{11}}, \qquad \alpha' \rho_1 + \alpha \rho_2 = -\frac{a_{23}}{a_{11}}$$

(note that γ_1, γ_2 are generic numbers; the pair γ_1, γ_2 in (1.18) is not (necessarily) the same as the pair γ_1, γ_2 in (1.21)).

Without loss of generality we can assume that $|a_{12}| \le a_{11} \le \sqrt{D/3}$ (this is a well-known fact from the Reduction Theory of binary quadratic forms; we omit the proof), and then we have $\alpha > 0 > \alpha'$.

For simplicity assume that the interval $[\gamma_1, \gamma_2]$ is symmetric to 0, i.e., $[\gamma_1, \gamma_2] = [-\gamma, \gamma]$. Also, assume that we are interested in the *positive* integral solutions of (1.21). Since $\alpha > 0 > \alpha'$, for "large" positive x and y the second factor $(x - \alpha' y + \rho_2)$ in (1.21) is also "large" positive, implying that the first factor $(x - \alpha y + \rho_1)$ in (1.21) has to be very small. That is, x has to be the nearest integer to $(y\alpha - \rho_1)$. It follows that the symmetric version of (1.18)

$$(1.22) \qquad -\gamma \leq a_{11}x^2 + a_{12}xy + a_{22}y^2 + a_{13}x + a_{23}y \leq \gamma,$$

where $\gamma > 0$ is a given real number, is equivalent to the diophantine inequality

$$(1.23) \qquad \|y\alpha - \rho_1\| < \frac{c}{y + O(1)} \qquad \text{where } c = \frac{\gamma}{\alpha - \alpha'} = \frac{\gamma a_{11}}{\sqrt{D}}.$$

Let's return to equation (1.18). If the linear part $a_{13}x + a_{23}y$ is missing, i.e., $a_{13} = a_{23} = 0$, then we have a complete theory based on the Pell equation. More precisely,

$$\gamma_1 \leq Q(x,y) \leq \gamma_2 \Longleftrightarrow Q(x,y) = m, \quad \gamma_1 \leq m \leq \gamma_2,\ m \in \mathbb{Z}$$

with $Q(x,y) = a_{11}x^2 + a_{12}xy + a_{22}y^2$, and we have a complete characterization of the integral solutions of $Q(x,y) = m$ for any integer m as follows. For any integer m there is a finite list of "primary solutions", say, (x_j, y_j), $j \in J$ where $|J| < \infty$ and $Q(x,y) = m$ such that, every solution $x = u, y = v$ of $Q(x,y) = m$ can be written in the form

$$u - \alpha v = \pm \left(\frac{u_0 + v_0\sqrt{D}}{2} \right)^n \cdot (x_j - \alpha y_j)$$

for some $j \in J$ and $n \in \mathbb{Z}$, where $x = u_0 > 0, y = v_0 > 0$ is the least positive solution of Pell's equation $x^2 - Dy^2 = 4$. As a byproduct, we obtain that the number of positive integral solutions of

$$\gamma_1 \leq Q(x,y) \leq \gamma_2 \qquad \text{with } 1 \leq x \leq N, 1 \leq y \leq N$$

has the simple asymptotic form $c \log N + O(1)$, where $c = c(a_{11}, a_{12}, a_{22}, \gamma_1, \gamma_2)$ is a constant and the error term $O(1)$ is *uniformly bounded* as $N \to \infty$. (For a more detailed proof; see Lang [La66].)

Exactly the same holds if there is a non-zero linear part $a_{13}x + a_{23}y$ in (1.18), but its effect "cancels out": ρ_1 in (1.21) is an integer.

Finally, if ρ_1 is *not* an integer, then I call (1.21) an *inhomogeneous Pell inequality.* In view of (1.23), an inhomogeneous Pell inequality (1.21) is basically equivalent to an inhomogeneous diophantine inequality

$$\|n\alpha - \beta\| < \frac{c}{n} \tag{1.24}$$

with $c = \gamma a_{11}/\sqrt{D}$, where α is a quadratic irrational defined in (1.20). Inequality (1.24) is a special case of (1.16) where $\varphi(n)$ is a constant.

One of the main results in my book Beck [Be010] is to describe the asymptotic behavior of the number of positive integral solutions of (1.18) for every non-square integer discriminant $D > 0$ and for *almost all* a_{13}, a_{23}. The number of solutions exhibits

(1) extra large fluctuations (proportional to the area(!), see Theorem 1 below),
(2) satisfies an elegant central limit Theorem (see Theorem A later in Section 2),
(3) satisfies a shockingly precise law of the iterated logarithm; see Theorem B later.

For notational simplicity, I formulate the results in the special case of discriminant $D = 8$, which corresponds to the most famous quadratic irrational: $\alpha = \sqrt{2}$.

Since the class number of discriminant $D = 8$ is one, the general form of an inhomogeneous Pell inequality of discriminant $D = 8$ is

$$\gamma_1 \le (x+\beta_1)^2 - 2(y+\beta_2)^2 \le \gamma_2 \tag{1.25}$$

where $\gamma_1 < \gamma_2$ and $\beta_1, \beta_2 \in [0,1)$ are fixed constants. For notational simplicity we restrict ourselves to symmetric intervals $[-\gamma, \gamma]$ in (1.25); note that everything works similarly for general intervals $[\gamma_1, \gamma_2]$.

The factorization

$$(x+\beta_1)^2 - 2(y+\beta_2)^2 = (x+\beta-y\sqrt{2})(x+\beta'+y\sqrt{2}), \tag{1.26}$$

where $\beta = \beta_1 - \beta_2\sqrt{2}$ and $\beta' = \beta_1 + \beta_2\sqrt{2}$, clearly indicates that the asymptotic number of integral solutions of (1.25) heavily depends on the "local" behavior of $n\sqrt{2}$ mod 1. In fact, (1.25) is essentially equivalent to the inhomogeneous diophantine inequality

$$\|n\sqrt{2} - \beta\| < \frac{c}{n} \tag{1.27}$$

with $c = \gamma/2\sqrt{2}$.

To turn the vague term "essentially equivalent" into a precise statement, let $F(\sqrt{2};\beta_1,\beta_2;\gamma;N)$ be the number of integral solutions $(x,y)\in\mathbb{Z}^2$ of (1.25) with $\gamma_2=\gamma$, $\gamma_1=-\gamma$ satisfying $1\le y\le N$ and $x\ge 1$. It means counting lattice points in a long and narrow hyperbola segment. Next let $f(\sqrt{2};\beta;c;N)$ be the number of integral solutions n of (1.27) satisfying $1\le n\le N$, where $\beta=\beta_1-\beta_2\sqrt{2}$. Now essentially equivalent means that, for almost all pairs β_1,β_2, $F(\sqrt{2};\beta_1,\beta_2;\gamma;N)-f(\sqrt{2};\beta;c;N)=O(1)$ as $N\to\infty$, where $c=\gamma/2\sqrt{2}$ (and $\beta=\beta_1-\beta_2\sqrt{2}$). More precisely, we have

Lemma 1.2. *Let $\gamma>0$ and β_2 be arbitrary real numbers. Then for almost all β_1 there exist a finite $0<C(\beta_1,\beta_2,\gamma)<\infty$ such that*

$$\int_0^1 C(\beta_1,\beta_2,\gamma)\,d\beta<\infty \quad \textit{and}$$

$$\left|F(\sqrt{2};\beta_1,\beta_2;\gamma;N)-f(\sqrt{2};\beta;c;N)\right|<C(\beta_1,\beta_2,\gamma)\quad \textit{for all } N\ge 1,$$

where $c=\gamma/2\sqrt{2}$ and $\beta=\beta_1-\beta_2\sqrt{2}$.

I postpone the simple proof to Section 3.

In view of Lemma 1.2 it suffices to study the special case $\beta_2=0$, $\beta_1=\beta$:

$$-\gamma\le(x+\beta)^2-2y^2\le\gamma \tag{1.28}$$

where $\gamma>0$ and $\beta\in[0,1)$ are fixed constants. For simplicity, let $F(\sqrt{2};\beta;\gamma;N)$ denote the number of integral solutions $(x,y)\in\mathbb{Z}^2$ of (1.28) satisfying $1\le y\le N$ and $x\ge 1$.

In the special case $\gamma=1$ and $\beta=0$, (1.28) becomes the simplest Pell equation $x^2-2y^2=\pm1$. The integral solutions (x_k,y_k) form a cyclic group generated by the smallest positive solution $x=y=1$ in the well-known way: $x_k+y_k\sqrt{2}=(1+\sqrt{2})^k$, implying the familiar asymptotic formula

$$F(\sqrt{2};\beta=0;\gamma=1;N)=\frac{\log N}{\log(1+\sqrt{2})}+O(1), \tag{1.29}$$

where $1+\sqrt{2}$ is the fundamental unit of the real quadratic field $\mathbf{Q}(\sqrt{2})$.

In sharp contrast to the bounded fluctuation in the homogeneous case $\beta=0$, the *inhomogeneous* case can exhibit "extra large fluctuations proportional to the area"; see Theorem 1 below. To explain this, first we have to compute the *mean value* of $F(\sqrt{2};\beta;\gamma;N)$ as β runs in the unit interval $0\le\beta<1$.

Lemma 1.3. *We have*

$$\int_0^1 F(\sqrt{2};\beta;\gamma;N)\,d\beta = \frac{\gamma}{\sqrt{2}}\log N + O(1), \tag{1.30}$$

where the implicit constant in $O(1)$ is independent of N (but may depend on γ). Moreover, for an arbitrary subinterval $0 \le a < b \le 1$ we have the limit formula

$$\lim_{N\to\infty} \frac{\frac{1}{b-a}\int_a^b F(\sqrt{2};\beta;\gamma;N)\,d\beta}{\log N} = \frac{\gamma}{\sqrt{2}}. \tag{1.31}$$

Formulas (1.30)–(1.31) express the almost trivial geometric fact that the average number of lattice points contained in all the translated copies of a given region (in our special case: a hyperbola segment) is precisely the area of the region (see Lemma 3.1). I will give a detailed proof of Lemma 1.3 in Section 3.

Next we formulate an "extra large fluctuation" result: Theorem 1. Note that this is hardly more than a simple illustration; we devote another long paper to the many generalizations of Theorem 1 (they require a completely different and much harder proof technique: the Riesz product; see e.g. (1.33)).

Theorem 1. *For $\gamma = 1/2$ there are continuum many "divergence points" $\beta^* \in [0,1)$ in the sense that*

$$\limsup_{n\to\infty} \frac{F(\sqrt{2};\beta^*;\gamma=1/2;n)}{\log n} > \liminf_{n\to\infty} \frac{F(\sqrt{2};\beta^*;\gamma=1/2;n)}{\log n}. \tag{1.32}$$

I postpone the proof of Theorem 1 to Section 3.

Note that the fluctuation const $\cdot \log n$ in (1.32) is as large as possible apart from a constant factor. This follows from Lemma 2.1; see the next section. It is fair to say that Theorem 1 represents a *sophisticated* violation of the Naive Area Principle. The two main results of the paper—Theorems 2 and 3—to be be discussed in the next section, will both support the Naive Area Principle.

I conclude Section 1 with mentioning, without proof, one far-reaching generalization of Theorem 1. It says that Theorem 1 in fact holds for *every* $\gamma > 0$, and we actually have the stronger inequality

$$\limsup_{n\to\infty} \frac{F(\sqrt{2};\beta^*;\gamma;n)}{\log n} > \frac{\gamma}{\sqrt{2}} > \liminf_{n\to\infty} \frac{F(\sqrt{2};\beta^*;\gamma;n)}{\log n}. \tag{1.33}$$

2. Defending the Naive Area Principle

2.1. Determinism vs. randomness

Equations (1.29) and (1.32) display the two extreme cases: (1) bounded fluctuations, and (2) extra large fluctuations proportional to the area. But what fluctuations do we have for a *typical* $0 < \beta < 1$? We show that for a typical β the asymptotic number of solutions $F(\sqrt{2};\beta;\gamma;N)$, as $N \to \infty$, justifies the Naive Area Principle; and going far beyond that, a more thorough look reveals "advanced randomness".

We know from probability theory that the two most important parameters of a random variable are the *expectation* (or mean value) and the *variance.* By (1.30) the *expectation* equals

$$\int_0^1 F(\sqrt{2};\beta;\gamma;N)\,d\beta = \frac{\gamma}{\sqrt{2}}\log N + O(1).$$

Next we explain why it is natural to use exponential scaling here. Note that for any $1 < M < N$, the counting function is "slowly changing" in the following sense:

$$F(\sqrt{2};\beta;\gamma;N) - F(\sqrt{2};\beta;\gamma;M) = O\left(\log(N/M)\right); \tag{2.1}$$

notice that const$\cdot\log(N/M)$ is the corresponding area. The geometric reason behind (2.1) is the exponentially sparse occurence of the lattice points in the corresponding long and narrow tilted hyperbola. The proof of (2.1) is a straightforward application of Lemma 2.1 below.

We have the following corollary of (2.1): If $M = cN$, i.e., n runs in $cN < n < N$ with some constant $0 < c < 1$, then the fluctuation of $F(\sqrt{2};\beta;\gamma;N)$ is a trivial $O(1)$. This negligible constant size change, as n runs in $cN < n < N$ (i.e., (2.1)), explains why it is perfectly natural to switch to the exponential scaling $F(\sqrt{2};\beta;\gamma;e^N)$. In the rest we will often prefer the exponential scaling.

The variance. It comes from the following result: for any $\gamma > 0$ there is a positive effective constant $\sigma = \sigma(\gamma) > 0$ such that

$$\lim_{N\to\infty}\frac{1}{N}\int_0^1\left(F(\sqrt{2};\beta;\gamma;e^N) - \frac{\gamma}{\sqrt{2}}N\right)^2\,d\beta = \sigma^2(\gamma).$$

Note that the proof of this limit formula is based on a combination of Fourier analysis and the arithmetic of the quadratic number field $\mathbf{Q}(\sqrt{2})$; see Beck [Be010] (see also the survey papers [Be98a] and [Be98b]).

The first probabilistic result—nicely fitting the general scheme of "determinism vs. randomness"—is the following (for the proof; see Beck [Be010]).

Theorem A ("central limit theorem")**.** *The renormalized counting function*

$$\frac{F(\sqrt{2};\beta;\gamma;e^N)-\frac{\gamma}{\sqrt{2}}N}{\sigma(\gamma)\sqrt{N}}, \quad 0\le\beta<1$$

has a standard normal limit distribution as $N\to\infty$.

To give at least a vague intuition behind Theorem A, we write

$$G_j(\beta)=F(\sqrt{2};\beta;\gamma;e^j)-F(\sqrt{2};\beta;\gamma;e^{j-1}), \quad j=1,2,\ldots,N.$$

That is, $G_j(\beta)$ is the number of integral solutions $n\in\mathbb{N}$ of (1.28) satisfying $e^{j-1}<n\le e^j$. Note that $G_j(\beta)$ is a bounded function—this follows from the following lemma, and the simple geometric fact that a short hyperbola segment, corresponding to G_j, can be approximated by an inscribed rectangle R_1 of slope $1/\sqrt{2}$ and a circumscribed rectangle R_2 of slope $1/\sqrt{2}$ such that the ratio of the areas is uniformly bounded.

Lemma 2.1. *Every tilted rectangle of slope $1/\sqrt{2}$ and area $1/5$ contains at most one lattice point.*

I postpone the proof of this lemma to the next section. Note that Lemma 2.1 can be easily generalized as follows. The same proof gives that for any quadratic irrational α there is a positive constant $c_0=c_0(\alpha)>0$ such that, every tilted rectangle of slope α and area c_0 contains at most one lattice point.

Let's return to the vague intuition: our key observation is that the bounded function $G_j(\beta)$ resembles the j-th Rademacher function, so the sum

$$F(\sqrt{2};\beta;\gamma;e^N)-\frac{\gamma}{\sqrt{2}}N=\sum_{j=1}^{N}\left(G_j(\beta)-\frac{\gamma}{\sqrt{2}}\right),$$

as a function of $\beta\in[0,1)$, behaves like a sum of N *independent* Bernoulli variables ("N-step random walk")

$$F(\sqrt{2};\beta;\gamma;e^N)-\frac{\gamma}{\sqrt{2}}N\approx\pm1\pm1\pm\cdots\pm1 \quad (N\text{ terms}). \tag{2.2}$$

Our next result—Theorem B—can be interpreted as a variant of Khintchine's famous law of the iterated logarithm in probability theory. We show that the number of solutions $F(\sqrt{2};\beta;\gamma;e^n)$ of (1.28) oscillates between the sharp bounds ($\varepsilon > 0$)

$$\begin{aligned}\frac{\gamma}{\sqrt{2}}n - \sigma\sqrt{n}\sqrt{(2+\varepsilon)\log\log n} &< F(\sqrt{2};\beta;\gamma;e^n) \\ &< \frac{\gamma}{\sqrt{2}}n + \sigma\sqrt{n}\sqrt{(2+\varepsilon)\log\log n}\end{aligned} \tag{2.3}$$

as $n \to \infty$ for *almost all* β. Note that (2.3) fails for $2-\varepsilon$ instead of $2+\varepsilon$ (where $\varepsilon > 0$). Here the main term $\frac{\gamma}{\sqrt{2}}n$ means the "area", so (2.3) can be considered as a highly sophisticated justification of the Naive Area Principle.

(2.3) is particularly interesting in view of the fact that the classical Circle Problem is unsolved (and seems to be hopeless for the available proof techniques). What (2.3) means is that, we can solve a "Hyperbola Problem" instead of the Circle Problem. More precisely, we *can* prove for long and narrow tilted hyperbola segments, what nobody can prove for large concentric circles. Namely, we can show that, for almost all centers (i.e., for almost all values of the translation parameter β), the number of lattice points asymptotically equals the area plus an error, which, even in the worst case scenario, is about the square root of the area. (For circles the corresponding maximum error should be the *square root* of the circumference—at least this is what the conjecture claims.)

Note that the law of the iterated logarithm is one of the most famous results in classical probability theory, and it describes the "maximum fluctuation" in the infinite (one-dimensional) random walk. The term infinite random walk refers to an infinite sequence of random Bernoulli trials, where a trial means tossing a fair coin. Of course, "coin tossing" belongs to the physical world; it is not a mathematical concept. But there is a well-known pure mathematical problem, which is considered equivalent: we can study the digit distribution of a typical real number written in the binary form

$$\beta = \frac{b_1}{2} + \frac{b_2}{2^2} + \frac{b_3}{2^3} + \cdots,$$

where each $b_i = 0$ or 1 (for simplicity assume that $0 < \beta < 1$). The infinite 0–1 sequence

$$b_1 = b_1(\beta), b_2 = b_2(\beta), b_3 = b_3(\beta), \ldots,$$

i.e., the sequence of binary digits of $0 < \beta < 1$, represents an infinite Heads-and-Tails sequence, say, 1 is Heads and 0 is Tails. The sum

$$B_n = B_n(\beta) = b_1 + b_2 + b_3 + \cdots + b_n$$

counts the number of 1's ("Heads") among the first n binary digits of $0 < \beta < 1$. Borel's classical theorem about normal numbers asserts that

$$\frac{B_n(\beta)}{n} \to \frac{1}{2} \quad \text{for almost all } 0 < \beta < 1.$$

Let $S_n = S_n(\beta)$ denote the corresponding error term

$$S_n = S_n(\beta) = 2B_n(\beta) - n = \text{number of Heads} - \text{number of Tails}.$$

That is, $S_n = S_n(\beta)$ represents the number of Heads minus the number of Tails among the first n random trials ("coin tossings").

A well-known theorem of Khintchine [Kh24] asserts that

$$\limsup_n \frac{S_n(\beta)}{\sqrt{2n \log\log n}} = 1 \quad \text{for almost all } 0 < \beta < 1.$$

Notice that Khintchine's theorem is a far-reaching quantitative improvement on Borel's theorem on "normal numbers". The "long form" of Khintchine's theorem says that, for any $\varepsilon > 0$ and for almost all β, we have the following two statements: (1)

$$S_n(\beta) < (1+\varepsilon)\sqrt{2n \log\log n}$$

for all sufficiently large values of n, and (2)

$$S_n(\beta) > (1-\varepsilon)\sqrt{2n \log\log n}$$

holds for infinitely many values of n.

This strikingly elegant and precise result is the simplest form of the law of the iterated logarithm (called the "Khintchine's form").

Let's return to (2.3). The fact that it is an analog of Khintchine's law of the iterated logarithm suggests the vague intuition that the lattice point counting function $F(\sqrt{2}; \beta; \gamma; e^n)$ behaves like a "generalized digit sum" (as β runs in $0 < \beta < 1$).

What we are going to actually formulate below (see Theorem B) are two generalizations/refinements of (2.3). The first generalization is that, for almost all β, (2.3) holds for *all* γ, or in general, for *all* intervals $[\gamma_1, \gamma_2]$. This is a variant of the so-called Cassels's form of the law of the iterated logarithm (see [Ca51]).

The second generalization of (2.3) is the Kolmogorov-Erdős form (see [Er42] and [Fe43]), an ultimate convergence-divergence criterion, which contains the Khintchine's form as a simple corollary.

Theorem B ("law of the iterated logarithm"). *(a) Let $\varepsilon > 0$ be an arbitrarily small but fixed constant. Then for almost all β,*

$$\begin{aligned}\frac{\gamma}{\sqrt{2}}n - \sigma\sqrt{n}\sqrt{(2+\varepsilon)\log\log n} &< F(\sqrt{2};\beta;\gamma;e^n) \\ &< \frac{\gamma}{\sqrt{2}}n + \sigma\sqrt{n}\sqrt{(2+\varepsilon)\log\log n}\end{aligned} \tag{2.4}$$

holds for all $\gamma > 0$ and for all sufficiently large n (i.e., for all $n > n_0(\beta, \gamma)$). Note that (2.4) is sharp in the sense that $2+\varepsilon$ cannot be replaced by $2-\varepsilon$.

(b) Let $\varphi(n)$ be an arbitrary positive increasing function of n. Let $\gamma > 0$ be fixed, then for almost all β,

$$F(\sqrt{2};\beta;\gamma;e^n) > \frac{\gamma}{\sqrt{2}}n + \varphi(n)\sigma\sqrt{n} \tag{2.5}$$

holds for infinitely many n's if and only if the series

$$\sum_{n=1}^{\infty} \frac{\varphi(n)}{n} e^{-\varphi^2(n)/2} \text{ diverges.} \tag{2.6}$$

Exactly the same holds for the other inequality

$$F(\sqrt{2};\beta;\gamma;e^n) < \frac{\gamma}{\sqrt{2}}n - \varphi(n)\sigma\sqrt{n}. \tag{2.6$'$}$$

Remarks. By Lemma 1.2, $f(\sqrt{2};\beta;c;N) = F(\sqrt{2};\beta;\gamma;N) + O(1)$ as $N \to \infty$, where $c = \gamma/2\sqrt{2}$. So Lemma 1.2 implies that Theorems A and B remain true if $F(\sqrt{2};\beta;\gamma;N)$ is replaced with the number of solutions $f(\sqrt{2};\beta;c;N) =$ of the inhomogeneous diophantine inequality (1.27).

In Theorem B(a) there is a dramatic difference between rational β and almost all β. For every rational β the counting function has the form

$$F(\sqrt{2};\beta;\gamma;N) = c(\gamma)\log N + O(1) \quad \text{as } N \to \infty \tag{2.7}$$

for all $\gamma > 0$, and it remains true if $\sqrt{2}$ is replaced by any quadratic irrational. This bounded size fluctuation around the main term $c\log N$ (which is typically *not* the area) jumps up considerably: we have square root (of the area) size fluctuation around the main term (=area), described in (2.4), and this holds for almost all β and all $\gamma > 0$.

Let's return to (2.3): it is a special case of Theorem B(b) with

$$\varphi(n) = ((2 \pm \varepsilon) \log\log n)^{1/2}.$$

Indeed, with this choice of $\varphi(n)$ the series (2.6) is divergent or convergent depending on whether we have $2+\varepsilon$ or $2-\varepsilon$.

We can obtain a much more delicate result with the choice of a large integer $k \geq 4$ and

$$\varphi(n) = \left(2\log_2 n + 3\log_3 n + 2\log_4 n + \cdots + 2\log_{k-1} n + (2\pm\varepsilon)\log_k n\right)^{1/2}.$$

Warning: here, and here only, we use the space-saving notation $\log_2 n = \log\log n$, i.e., it means the iterated logarithm (instead of the usual meaning as the base 2 logarithm), and in general, $\log_k n = \log(\log_{k-1} n)$ denotes the k-times iterated logarithm of n. With this choice of $\varphi(n)$,

$$\sum_{n=1}^{\infty} \frac{\varphi(n)}{n} e^{-\varphi^2(n)/2} \approx \sum_n \frac{1}{n \log n \log_2 n \log_3 n \cdots \log_{k-1} n (\log_k n)^{1\pm\varepsilon/2}},$$

which is divergent or convergent depending on whether we have $2+\varepsilon$ or $2-\varepsilon$.

This example clearly illustrates the remarkable precision of Theorem B(b).

2.2. Formulating Theorems 2 and 3

Next we focus on a simple consequence of Theorem B. Let $c > 0$ be arbitrarily small but fixed, then by Theorem B the inhomogeneous diophantine inequality

$$\|n\sqrt{2} - \beta\| < \frac{c}{n} \tag{2.8}$$

has infinitely many integer solutions $n \geq 1$ for *almost all* β (in the sense of the Lebesgue measure).

Inequality (2.8) corresponds to the hyperbola segment (β is fixed):

$$|y - \beta| < \frac{c}{x}, \quad x \geq 1,$$

which has infinite area.

But we may go further, and consider the smaller region

$$|y-\beta| < \frac{1}{x\log x}, \text{ and the even smaller region } |y-\beta| < \frac{1}{x\log x\log\log x},$$

and so on. They all have infinite area, since

$$\int_e^N \frac{dx}{x\log x} = \log\log N, \qquad \text{and} \qquad \int_{e^e}^N \frac{dx}{x\log x} = \log\log\log N,$$

and the rest all tend to infinity as $N \to \infty$. It is very natural, therefore, to ask the following question.

Question. Consider the inequalities

$$\|n\sqrt{2}-\beta\| < \frac{c}{n\log n} \quad (n \ge 2), \tag{2.9}$$

$$\|n\sqrt{2}-\beta\| < \frac{c}{n\log n\log\log n} \quad (n \ge 3), \tag{2.10}$$

and so on; $0 \le \beta < 1$ is a fixed constant. Is it true that, for *almost all* β (in the sense of the Lebesgue measure), inequality (2.9) (and (2.10), and so on) does have infinitely many integer solutions $n \ge 1$?

Well, the answer is *yes.*

Theorem 2 ("Area Principle for $\sqrt{2}$")**.** *Let $\psi(x)$ be any positive decreasing function of the real variable x with*

$$\sum_n \psi(n) = \infty.$$

Then the inhomogeneous inequality

$$\|n\alpha-\beta\| < \psi(n)$$

has infinitely many integral solutions for almost all $0 \le \beta < 1$ (in the sense of Lebesgue measure).

What is more, there is an interesting generalization of Theorem 2 where $\sqrt{2}$ is replaced by any real α.

To explain this generalization (see Theorem 3 below), I briefly recall the basic question of diophantine approximation: we want to know whether an inequality

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^2}, \quad \text{equivalently,} \quad |q\alpha - p| < \frac{1}{q} \tag{2.11}$$

with integers p, q, or more generally, of an inequality

$$\|q\alpha\| < \psi(q) \tag{2.12}$$

has infinitely many integral solutions in q, and if this is the case, we want to determine the solutions, or at least determine the asymptotic number of integral solutions. (As usual, $\|x\|$ denotes the distance of a real x from the nearest integer, and $\psi(q)$ is a positive increasing function of q.)

It is perfectly natural to study the inhomogeneous analog of (2.12):

$$\|q\alpha - \beta\| < \psi(q) \tag{2.13}$$

where β is an arbitrary fixed real number. Of course, we may assume $0 \leq \beta < 1$.

Is there any connection between the solvability of the homogeneous (2.12) and the inhomogeneous (2.13)? I recall that Theorem 2 is about the Naive Area Principle in the special case $\alpha = \sqrt{2}$. The Naive Area Principle is a vague intuition claiming that a "nice region of infinite area must contain infinitely many lattice points". We know that the Naive Area Principle is false for the hyperbolic region $-1/2 \leq x^2 - 2y^2 \leq 1/2$, which has infinite area and contains only one lattice point (the origin). This Pell inequality is basically equivalent to the diophantine inequality

$$\|q\sqrt{2}\| < \frac{c}{q} \quad \text{with } c \leq 2^{-5/2}, \tag{2.14}$$

and (2.14) has only a finite number of integral solutions in q if the constant $c < 2^{-5/2}$. We can put Theorem 2 in the following form: the failure of the Naive Area Principle for (2.14) is "compensated" by the success of the Naive Area Principle for the inhomogeneous inequality

$$\|q\sqrt{2} - \beta\| < \psi(q), \tag{2.15}$$

for almost all β. That is, (2.15) has infinitely many integral solution q for almost all β, provided $\psi(x)$ is any positive decreasing function of the real variable x with

$$\sum_{n=1}^{\infty} \psi(n) = \infty. \tag{2.16}$$

The next result generalizes the special case $\alpha = \sqrt{2}$ for arbitrary real α.

Theorem 3 ("Area Principle in general"). *Let $\psi(x)$ be any positive decreasing function of the real variable x with*

$$\sum_{n=1}^{\infty} \psi(n) = \infty. \tag{2.17}$$

For any real number α, at least one of the following two cases always holds:

(i) the homogeneous inequality

$$\|q\alpha\| < \psi(q) \tag{2.18}$$

has infinitely many integral solutions,

(ii) the inhomogeneous inequality

$$\|q\alpha - \beta\| < \psi(q) \tag{2.19}$$

has infinitely many integral solutions for almost all $0 \le \beta < 1$ *(in the sense of Lebesgue measure).*

Remarks. (1) Note that divergence condition (2.17) is necessary. Indeed, if

$$\sum_{n=1}^{\infty} \psi(n) < \infty \tag{2.20}$$

then the set of pairs (α, β) for which the inequality

$$\|q\alpha - \beta\| < \psi(q) \tag{2.21}$$

has infinitely many integral solutions q, has 2-dimensional Lebesgue measure zero. This statement immediately follows from the other statement that, for every fixed β, the set of α which satisfy (2.21) for infinitely many q, has Lebesgue measure zero. The second statement has an easy proof as follows: every such α in $0 < \alpha < 1$ is contained in infinitely many intervals of the form

$$\left[\frac{p+\beta}{q} - \frac{\psi(q)}{q}, \frac{p+\beta}{q} + \frac{\psi(q)}{q}\right]$$

with $q \ge N$, $1 \le p \le q$ integers, and the total length of these intervals is less than

$$2 \sum_{q \ge N} \psi(q),$$

which by (2.20) tends to zero as $N \to \infty$.

This means that Theorem 3 is a precise convergence-divergence type result, or we may call it a "zero-one law" (to borrow a well-known concept from probability theory).

(2) Let's return to the inhomogeneous inequality (2.21). If α is rational and β is irrational, then (2.21) has only a finite number of integral solutions for any $\psi(q) \to 0$ as $q \to \infty$. Well, this is trivial. It is far less trivial to find an irrational α and a decreasing function ψ with $\sum_q \psi(q) = \infty$ such that for almost all β (2.21) has only a finite number of integral solutions. We can take any irrational $0 < \alpha < 1$ with "sufficiently large" partial quotients in the following quantitative sense:

$$\alpha = \cfrac{d}{a_1 + \cfrac{1}{a_2 + \cdots}} = [a_1, a_2, a_3, \ldots]$$

where

$$a_k \approx k^{(\log k)^2}, \tag{2.22}$$

and take

$$\psi(q) = \frac{1}{q \log q}. \tag{2.23}$$

Then the denominator q_k of the kth convergent of α is roughly

$$q_k \approx a_1 a_2 \cdots a_k \approx k^{k(\log k)^2}, \tag{2.24}$$

and so

$$\sum_k \frac{d}{\log q_k} \le \text{const} \sum_k \frac{1}{k(\log k)^3} < \infty.$$

I recall the well-known fact

$$\left| \alpha - \frac{p_k}{q_k} \right| < \frac{1}{q_k q_{k+1}}$$

which implies

$$\left| n\alpha - \frac{np_k}{q_k} \right| < \frac{n}{q_k q_{k+1}}. \tag{2.25}$$

If $q_k \le n < q_{k+1}k^{-2}$ and

$$\|n\alpha - \beta\| < \frac{1}{n\log n}$$

then by (2.24)–(2.25)

$$\left\|\beta - \frac{np_k}{q_k}\right\| < \frac{1}{k^2 q_k} + \frac{1}{n\log n} < \frac{2}{k(\log k)^3 q_k}. \tag{2.26}$$

If $q_{k+1}k^{-2} \le n < q_{k+1}$ then define the set

$$A_k = \bigcup_n \left[n\alpha - \frac{1}{n\log n}, n\alpha + \frac{1}{n\log n}\right] \pmod 1 \tag{2.27}$$

where the summation in (2.27) is extended over all n with $q_{k+1}k^{-2} \le n < q_{k+1}$, and motivated by (2.26) define the set

$$B_k = \bigcup_{0\le j<q_k} \left[\frac{j}{q_k} - \frac{2}{k(\log k)^3 q_k}, \frac{j}{q_k} + \frac{2}{k(\log k)^3 q_k}\right] \pmod 1. \tag{2.28}$$

Clearly

$$\sum_k \text{meas}(B_k) \le \sum_k \frac{4}{k(\log k)^3} < \infty, \tag{2.29}$$

where meas stands for the usual Lebesgue measure, and

$$\begin{aligned}\sum_k \text{meas}(A_k) &\le \text{const} \sum_k \frac{\log(k^2)}{k(\log k)^3} \\ &\le \text{const} \sum_k \frac{1}{k(\log k)^2} < \infty. \end{aligned} \tag{2.30}$$

It follows from (2.29)–(2.30) that almost all β are contained only in a finite number of A_k and in a finite number of B_k. In view of (2.26)–(2.28) this implies that, for almost all β, inequality (2.21) has only a finite number of integral solutions (where α and ψ are defined by (2.22)–(2.23)).

For the proofs of Theorems A–B, I refer the reader to my new book Beck [Be010] that will be published soon.

The next section is devoted to the proofs of Theorem 1 and Lemmas 1.1, 1.2, 1.3 and 2.1. Theorems 2 and 3 will be proved in Sections 4 and 5.

Probably the reader is wondering: why do we include a separate proof for Theorem 2 when it is just a special case of Theorem 3 for $\alpha = \sqrt{2}$? Well, the short answer is that we wanted to illustrate a new idea on a simple example. The periodicity of the continued fraction of $\sqrt{2}$ quickly leads us to the concept of *homogeneous Markov chains* (a slight relaxation of independence), and makes it possible to have a shortcut: a direct application of probability theory. Besides the proof of Theorem 2, Markov chains also play a key role in the proofs of Theorems A and B (that I omitted for the lack of space), so the reader can have at least a vague idea how those much longer proofs actually go.

On the other hand, the proof of Theorem 3 does *not* use Markov chains—instead it is based entirely on the theory of continued fraction.

3. Proving Theorem 1 and the lemmas

3.1. Proof of Lemma 1.1

In view of the familiar factorization

$$x^2 - 2y^2 = (x + y\sqrt{2})(x - y\sqrt{2}),$$

it is more convenient to compute the area of the following slight variant of region (1.10): let

$$\begin{aligned} H^*(\sqrt{2};[\gamma_1,\gamma_2];N) = \Big\{(x,y) \in \mathbb{R}^2 :\ & \gamma_1 \le x^2 - 2y^2 \le \gamma_2 \\ & \text{where } 1 \le x + y\sqrt{2} \le 2\sqrt{2}N\Big\}. \end{aligned} \tag{3.1}$$

Consider the substitution

$$u_1 = x + y\sqrt{2}, \qquad u_2 = x - y\sqrt{2}, \tag{3.2′}$$

which is equivalent to

$$x = \frac{u_1 + u_2}{2}, \qquad y = \frac{u_1 - u_2}{2\sqrt{2}}; \tag{3.2″}$$

the corresponding determinant is

$$\frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} 1 & -\sqrt{2} \\ 1 & \sqrt{2} \end{vmatrix} = 2\sqrt{2}.$$

Applying the substitution (3.2′)–(3.2″), we have

$$\begin{aligned}\text{area}(H^*(\sqrt{2};[\gamma_1,\gamma_2];N)) &= \int_{H^*(\sqrt{2};[\gamma_1,\gamma_2];N)} 1\,dxdy\\ &= \frac{1}{2\sqrt{2}}\int_{1\le u_1\le 2\sqrt{2}N}\left(\int_{\gamma_1/u_1\le u_2\le\gamma_2/u_1} 1\,du_2\right)du_1\\ &= \frac{1}{2\sqrt{2}}\int_1^{2\sqrt{2}N}\frac{\gamma_2-\gamma_1}{u_1}du_1 = \frac{\gamma_2-\gamma_1}{2\sqrt{2}}\log N + O(1). \qquad (3.3)\end{aligned}$$

A simple geometric consideration shows that

$$\text{area}\left(H(\sqrt{2};[\gamma_1,\gamma_2];N)\right) = \text{area}\left(H^*(\sqrt{2};[\gamma_1,\gamma_2];N)\right) + O(1),$$

and so (3.3) completes the proof of Lemma 1.1. □

3.2. Proof of Lemma 1.3

First I prove formula (1.30). Consider the hyperbolic needle $H_N(\gamma) = H_N(\sqrt{2};\gamma)$ defined as

(3.4)
$$H_N(\gamma) = \left\{(x,y)\in\mathbb{R}^2:\ -\gamma\le x^2-2y^2\le\gamma \text{ where } 1\le x+y\sqrt{2}\le 2\sqrt{2}N\right\}.$$

Comparing (3.4) with (3.1), we see that

$$H_N(\gamma) = H^*(\sqrt{2};[-\gamma,\gamma];N),$$

so by (3.3) we obtain the area:

$$\text{area}(H_N(\gamma)) = \frac{\gamma}{\sqrt{2}}\log N + O(1). \qquad (3.5)$$

Next we need the following almost trivial result.

Lemma 3.1. *Let $A\subset\mathbb{R}^2$ be a Lebesgue measurable set in the plane with finite measure (that I call the "area"). Then*

$$\int_0^1\int_0^1 |(A+\boldsymbol{x})\cap\mathbb{Z}^2|\,d\boldsymbol{x} = \text{area}(A),$$

where $A+\boldsymbol{x}$ is the translated copy of set A, translated by the vector $\boldsymbol{x}\in\mathbb{R}^2$.

First I derive Lemma 1.3 from Lemma 3.1. By Lemma 3.1,

$$\int_0^1\int_0^1 |(H_N(\gamma)+\mathbf{v})\cap\mathbb{Z}^2|\,d\mathbf{v} = \operatorname{area}(H_N(\gamma)), \tag{3.6}$$

where $A+\mathbf{v}$ denotes the translated copy.

If $\mathbf{v}=(v_1,v_2)\in[0,1)^2$ is chosen in such a way that $v_1-v_2\sqrt{2}\equiv\beta$ (mod 1) is fixed, then clearly

$$\Big|F(\sqrt{2};\beta;\gamma;N)-|(H_N(\gamma)+\mathbf{v})\cap\mathbb{Z}^2|\Big|<c_0(\gamma), \tag{3.7}$$

where $c_0(\gamma)<\infty$ is a constant independent of β and N. Combining (3.6)–(3.7), equation (1.30) follows.

Next we prove (1.31). Let $0\le a<b\le 1$ be fixed, and for any $M\ge 1$ define the parallelogram

$$\mathcal{P}_M=\{\mathbf{v}=(v_1,v_2)\in\mathbb{R}^2:\ a\le v_1-v_2\sqrt{2}\le b,\ 0\le v_1+v_2\sqrt{2}\le M\}. \tag{3.8}$$

If M is large, then $\mathcal{P}_M$ is a long and narrow parallelogram, but we can turn it into a "round" shape by applying an appropriate automorph of the quadratic form x^2-2y^2. The substitution $x_1=x+2y$, $y_1=x+y$ is a fundamental automorph of x^2-2y^2 (indeed, $x_1^2-2y_1^2=(x+2y)^2-2(x+y)^2=-(x^2-2y^2)$), and

$$A^k=\begin{pmatrix}1&2\\1&1\end{pmatrix}^k,\quad k\in\mathbb{Z}$$

give rise to infinitely many automorphs preserving the lattice points and the area. The eigenvectors of $A=\left(\begin{smallmatrix}1&2\\1&1\end{smallmatrix}\right)$ are parallel to the sides of parallelogram $\mathcal{P}_M$, so applying an appropriate power A^k on the long and narrow parallelogram $\mathcal{P}_M$, we obtain a "round" parallelogram $A^k\mathcal{P}_M$ with sides parallel to that of $\mathcal{P}_M$, and

$$\operatorname{area}(A^k\mathcal{P}_M)=\operatorname{area}(\mathcal{P}_M)=\text{const}\cdot M.$$

"Round" means that the diameter of parallelogram $A^k\mathcal{P}_M$ is $O(\sqrt{M})$, so the number of unit squares $[0,1)^2+\mathbf{n}$, $\mathbf{n}\in\mathbb{Z}^2$ intersecting the boundary of $A^k\mathcal{P}_M$ is $O(\sqrt{M})$.

Combining this geometric fact with Lemma 3.1 (see (3.6)), we have

$$\frac{1}{\operatorname{area}(\mathcal{P}_M)}\int_{\mathcal{P}_M}|(H_N(\gamma)+\mathbf{v})\cap\mathbb{Z}^2|\,d\mathbf{v}=\operatorname{area}(H_N(\gamma))\left(1+O(M^{-1/2})\right). \tag{3.9}$$

If $\mathbf{v} = (v_1, v_2) \in [0,1)^2$ is chosen in such a way that $v_1 - v_2\sqrt{2} \equiv \beta \pmod 1$ is fixed, then clearly

$$\left| F(\sqrt{2};\beta;\gamma;N) - |(H_N(\gamma) + \mathbf{v}) \cap \mathbb{Z}^2| \right| < c_0(\gamma, M), \tag{3.10}$$

where $c_0(\gamma, M) < \infty$ is a constant independent of β and N. Combining (3.5), (3.9) and (3.10),

$$\begin{aligned} &\frac{\frac{1}{b-a}\int_a^b F(\sqrt{2};\beta;\gamma;N)\,d\beta}{\log N} \\ &\quad = \left(\frac{\gamma}{\sqrt{2}} + O(1/\log N)\right)\left(1 + O(M^{-1/2})\right) + \frac{c_0(\gamma, M)}{\log N}. \end{aligned} \tag{3.11}$$

Since M can be arbitrarily large, (3.11) implies (1.31). The proof of Lemma 1.3 is complete. □

For the sake of completeness I include a

3.3. Proof of Lemma 3.1

First assume that A is bounded. Let N be a “large” integer. By using the periodicity of $\mathbb{Z}^2$ we have

$$\int_0^N \int_0^N |(A + \mathbf{x}) \cap \mathbb{Z}^2| \; d\mathbf{x} = N^2 \int_0^1 \int_0^1 |(A + \mathbf{x}) \cap \mathbb{Z}^2| \; d\mathbf{x}.$$

On the other hand,

$$\begin{aligned} \int_0^N \int_0^N |(A + \mathbf{x}) \cap \mathbb{Z}^2| \; d\mathbf{x} &= \sum_{\mathbf{n}\in\mathbb{Z}^2} \text{area}\left\{\mathbf{x} \in [0,N]^2 : \; \mathbf{n} \in A + \mathbf{x}\right\} \\ &= \sum_{\mathbf{n}\in\mathbb{Z}^2} \text{area}\left\{(\mathbf{n} - A) \cap [0,N]^2\right\}. \end{aligned}$$

Without loss of generality we can assume that the origin is inside A. Let $d(A)$ denote the diameter of A. Then $(\mathbf{n} - A) \subset [0,N]^2$ if $\mathbf{n} \in [d(A), N - d(A)]^2$, and $(\mathbf{n} - A) \cap [0,N]^2 = \emptyset$ if $\mathbf{n} \notin [-d(A), N + d(A)]^2$. Thus we have

$$\begin{aligned} (N + 2d(A))^2 \cdot \text{area}(A) &\geq \sum_{\mathbf{n}\in\mathbb{Z}^2} \text{area}\left\{(\mathbf{n} - A) \cap [0,N]^2\right\} \\ &\geq (N - 2d(A))^2 \cdot \text{area}(A). \end{aligned}$$

Dividing the last line by N^2, and combining the equations above, Lemma 3.1 follows as N tends to infinity. If A is unbounded, then we approximate A with an increasing sequence $A_1 \subset A_2 \subset A_3 \subset \dots$ of subsets of A such that each A_k is bounded and area$(A \setminus A_k) \to 0$. The last step is to use the continuity of the Lebesgue measure. □

3.4. Proof of Lemma 1.2

For notational simplicity I just prove the special case $\beta_2 = 0$ (the general case is the same). Again the key step is to apply Lemma 3.1. For $1 \le K < L \le \infty$ we define the following four regions:

$$H_{K,L}(\beta;\gamma) = \Big\{(x,y) \in \mathbb{R}^2 :\ -\gamma \le (x+\beta)^2 - 2y^2 \le \gamma \quad \text{where } K \le y \le L,\ x > 0\Big\},$$

$$\widetilde{H}_{K,L}(\beta;\gamma) = \Big\{(x,y) \in \mathbb{R}^2 :\ |x+\beta - y\sqrt{2}| \cdot 2\sqrt{2}y < \gamma \quad \text{where } K \le y \le L,\ x > 0\Big\},$$

$$\widetilde{H}^{+}_{K,L}(\beta;\gamma) = \Big\{(x,y) \in \mathbb{R}^2 :\ |x+\beta - y\sqrt{2}| \cdot (2\sqrt{2}y + 1) < \gamma \quad \text{where } K \le y \le L,\ x > 0\Big\},$$

$$\widetilde{H}^{-}_{K,L}(\beta;\gamma) = \Big\{(x,y) \in \mathbb{R}^2 :\ |x+\beta - y\sqrt{2}| \cdot (2\sqrt{2}y - 1) < \gamma \quad \text{where } K \le y \le L,\ x > 0\Big\}.$$

In view of factorization (1.26), $(x,y) \in H_{K,L}(\beta;\gamma)$ implies that $x + \beta = y\sqrt{2} + o(1)$; in fact, we have the stronger form $x + \beta = y\sqrt{2} + O(1/y)$. Thus there is a threshold $c_1 = c_1(\gamma)$ such that

$$\widetilde{H}^{+}_{K,L}(\beta;\gamma) \subset H_{K,L}(\beta;\gamma) \subset \widetilde{H}^{-}_{K,L}(\beta;\gamma)$$

holds for all $L > K > c_1(\gamma)$. On the other hand, it is trivial that

$$\widetilde{H}^{+}_{K,L}(\beta;\gamma) \subset \widetilde{H}_{K,L}(\beta;\gamma) \subset \widetilde{H}^{-}_{K,L}(\beta;\gamma).$$

Consider now the special case $K = 1$, $L = \infty$, $\beta = 0$, and study the difference set

$$D(\gamma) = \widetilde{H}^{-}_{1,\infty}(0;\gamma) \setminus \widetilde{H}^{+}_{1,\infty}(0;\gamma).$$

We estimate the area of the difference set $D(\gamma)$:

$$\begin{aligned}\operatorname{area}(D(\gamma)) &= O(1)\int_1^\infty \left(\frac{1}{2\sqrt{2}y-1}-\frac{1}{2\sqrt{2}y+1}\right)dy\\ &= O(1)\int_1^\infty \frac{dy}{8y^2-1} = O(1).\end{aligned}$$

Combining this with Lemma 3.1, we have

$$\int_0^1\int_0^1 |(D(\gamma)+\mathbf{v})\cap\mathbb{Z}^2|\,d\mathbf{v} = \operatorname{area}(D(\gamma)) < \infty. \tag{3.12}$$

If $\mathbf{v}=(v_1,v_2)\in[0,1)^2$ is chosen in such a way that $v_1-v_2\sqrt{2}\equiv\beta \pmod 1$ is fixed, then

$$D(\gamma)+\mathbf{v}\supset H_{K,L}(\beta;\gamma)\bigtriangleup \widetilde{H}^{+}_{K,L}(\beta;\gamma), \tag{3.13}$$

where $A\bigtriangleup B=(A\setminus B)\cup(B\setminus A)$ is the symmetric difference of A and B.

Combining (3.12) and (3.13), Lemma 1.2 easily follows. □

3.5. Proof of Lemma 2.1

Consider a rectangle of slope $1/\sqrt{2}$ which contains two lattice points $P=(k,\ell)$ and $Q=(m,n)$; in fact, assume that P,Q are two corner-points of the rectangle. Let PQ-vector$=\mathbf{v}=(m-k,n-\ell)$, and consider the two perpendicular unit vectors

$$\mathbf{e}_1=\left(\frac{\sqrt{2}}{\sqrt{3}},\frac{d}{\sqrt{3}}\right) \quad\text{and}\quad \mathbf{e}_2=\left(\frac{d}{\sqrt{3}},\frac{-\sqrt{2}}{\sqrt{3}}\right).$$

Then the two sides, a and b, of the rectangle can be expressed in terms of the inner products $\mathbf{e}_1\cdot\mathbf{v}$ and $\mathbf{e}_2\cdot\mathbf{v}$:

$$a=|\mathbf{e}_1\cdot\mathbf{v}|=\frac{|p\sqrt{2}+q|}{\sqrt{3}} \quad\text{and}\quad b=|\mathbf{e}_2\cdot\mathbf{v}|=\frac{|p-q\sqrt{2}|}{\sqrt{3}},$$

where $p=m-k$ and $q=n-\ell$. Thus we have

$$\operatorname{area}=ab=\frac{|p\sqrt{2}+q|\cdot|p-q\sqrt{2}|}{3}.$$

Without loss of generality we can assume that $p \geq 0$ and $q \geq 0$. Since $(p,q) \neq (0,0)$, we have

$$|p - q\sqrt{2}| = \frac{|p^2 - 2q^2|}{p + q\sqrt{2}} = \frac{1}{p + q\sqrt{2}},$$

and so

$$\begin{aligned} \text{area} &= \frac{|p\sqrt{2}+q| \cdot |p - q\sqrt{2}|}{3} \geq \frac{1}{3} \cdot \frac{p\sqrt{2}+q}{p+q\sqrt{2}} \\ &\geq \frac{1}{3} \cdot \frac{p/\sqrt{2}+q}{p+q\sqrt{2}} = \frac{d}{3\sqrt{2}} > \frac{1}{5}, \end{aligned}$$

proving Lemma 2.1. □

3.6. Proof of Theorem 1

We show that the set of β's in question ("set of divergence points") contains a Cantor set. This guarantees that the cardinality of the set is continuum.

We make a standard Cantor set construction, i.e., we apply the method of "nested intervals". For notational convenience, we write $F(\sqrt{2};\beta;\gamma;N) = F(\beta;\gamma;N)$. By (1.30),

$$\int_0^1 F(\beta;\gamma;N)\, d\beta = \frac{\gamma}{\sqrt{2}} \log N + O(1),$$

and applying it with $\gamma = 1/4$, we obtain the existence of a $0 < \beta_1 < 1$ and an arbitrarily large integer N_1 such that

$$F(\beta_1;\gamma = 1/4;N_1) > \frac{1}{8} \log N_1.$$

Since $1/4 < 1/2$, there is an interval $I_1 = [a,b]$ with $0 < a < b < 1$ such that $\beta_1 \in I_1$ and

$$F(\beta;\gamma = 1/2;N_1) > \frac{1}{8} \log N_1 \quad \text{for all } \beta \in I_1. \tag{3.14}$$

Next let $\mathbf{n} = (n_1,n_2) \in \mathbb{Z}^2$ be a lattice point such that $\beta_2 = n_1 - n_2\sqrt{2} \in I_1$. Since the equation $|x^2 - 2y^2| \leq 3/4$ does *not* have a non-zero integral solution, trivially

$$F(\beta_2;\gamma = 3/4;N) < \frac{1}{100} \log N \quad \text{for all } N \geq N_2,$$

where $N_2 < \infty$ is a sufficiently large threshold. We can clearly assume that $N_2 > N_1$. Since $3/4 > 1/2$, there is an interval $I_2 = [a, b]$ with some $0 < a < b < 1$ (a and b are generic numbers) such that $\beta_2 \in I_2$ and

$$(3.15) \qquad F(\beta; \gamma = 1/2; N_2) < \frac{1}{100} \log N_2 \quad \text{for all } \beta \in I_2.$$

We can clearly assume that I_2 is a proper subinterval of I_1. Let $I(0) = I_2$, and repeating the second argument, there is another closed subinterval $I(1)$ such that, $I(0) \cup I(1) \subset I_1$, $I(0)$ and $I(1)$ are disjoint, and

$$(3.16) \qquad F(\beta; \gamma = 1/2; N_2^{(1)}) < \frac{1}{100} \log N_2^{(1)} \quad \text{for all } \beta \in I(1).$$

We can clearly assume that $N_2^{(1)} > N_1$.

By (1.31),

$$\frac{1}{|I(0)|} \int_{I(0)} F(\beta; \gamma; N)\, d\beta = (1 + o(1)) \frac{\gamma}{\sqrt{2}} \log N,$$

and applying it with $\gamma = 1/4$, we obtain the existence of a $0 < \beta_3 < 1$ and a large integer N_3 such that

$$F(\beta_3; \gamma = 1/4; N_3) > \frac{1}{8} \log N_3.$$

Since $1/4 < 1/2$, there is an interval $I_3 = [a, b]$ with $0 < a < b < 1$ such that $\beta_3 \in I_3$ and

$$(3.17) \qquad F(\beta; \gamma = 1/2; N_3) > \frac{1}{8} \log N_3 \quad \text{for all } \beta \in I_3.$$

We can clearly assume that I_3 is a proper subinterval of $I(0)$. Write $I(0, 0) = I_3$. Similarly, there is another subinterval $I(0, 1)$ such that, $I(0, 0) \cup I(0, 1) \subset I(0)$, $I(0, 0)$ and $I(0, 1)$ are disjoint, and

$$(3.18) \qquad F(\beta; \gamma = 1/2; N_3^{(1)}) > \frac{1}{8} \log N_3^{(1)} \quad \text{for all } \beta \in I(0, 1).$$

There are similar disjoint subintervals $I(1, 0)$ and $I(1, 1)$ of $I(1)$.

Next let $\mathbf{n} = (n_1, n_2) \in \mathbb{Z}^2$ be a lattice point such that $\beta_4 = n_1 - n_2\sqrt{2} \in I(0, 0)$. Since the equation $|x^2 - 2y^2| \leq 3/4$ does *not* have a non-trivial integral solution,

$$F(\beta_4; \gamma = 3/4; N) < \frac{1}{100} \log N \quad \text{for all } N \geq N_4,$$

where $N_4 < \infty$ is a sufficiently large threshold. We can clearly assume that $N_4 > N_3$. Since $3/4 > 1/2$, there is an interval $I_4 = [a, b]$ with $0 < a < b < 1$ such that $\beta_4 \in I_4$ and

$$F(\beta; \gamma = 1/2; N_4) < \frac{1}{100} \log N_4 \quad \text{for all } \beta \in I_4. \tag{3.19}$$

We can clearly assume that I_4 is a proper subinterval of $I(0, 0)$. Let $I(0, 0, 0) = I_4$, and repeating the last argument, there is another closed subinterval $I(0, 0, 1)$ such that, $I(0, 0, 0) \cup I(0, 0, 1) \subset I(0, 0)$, $I(0, 0, 0)$ and $I(0, 0, 1)$ are disjoint, and

$$F(\beta; \gamma = 1/2; N_4^{(1)}) < \frac{1}{100} \log N_4^{(1)} \quad \text{for all } \beta \in I(0, 0, 1), \tag{3.20}$$

and so on. Repeating this argument, we build an infinite binary tree:

$$I_1 \supset I_{\varepsilon_1} \supset I_{\varepsilon_1,\varepsilon_2} \supset I_{\varepsilon_1,\varepsilon_2,\varepsilon_3} \supset \cdots$$

where $\varepsilon_1 = 0$ or 1, $\varepsilon_2 = 0$ or 1, $\varepsilon_3 = 0$ or 1, and so on.

For an arbitrary infinite 0–1 sequence $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots$, let

$$\beta \in I_1 \cap I_{\varepsilon_1} \cap I_{\varepsilon_1,\varepsilon_2} \cap I_{\varepsilon_1,\varepsilon_2,\varepsilon_3} \cap \cdots ,$$

then by (3.14)–(3.20) there is an infinite sequence $1 < M_1 < M_2 < M_3 < M_4 < \ldots$ of integers such that

$$F(\beta; \gamma = 1/2; M_{2k-1}) > \frac{1}{8} \log M_{2k-1}$$

and

$$F(\beta; \gamma = 1/2; M_{2k}) < \frac{1}{100} \log M_{2k},$$

where $k = 1, 2, 3, \ldots$. This proves Theorem 1. □

4. Proof of Theorem 2: Markov chains

4.1. Quadratic irrational scale and Markov chains

Instead of using the familiar decimal representation of real numbers, we switch to the unusual "$(1 + \sqrt{2})$ scale representation" of the translation constant β. The "$(1 + \sqrt{2})$ scale representation" of real numbers goes as follows. Since $2 < 1+\sqrt{2} < 3$, we can certainly write every real $0 < \beta < 1$ in the form

$$\beta = \frac{b_1}{1+\sqrt{2}} + \frac{b_2}{(1+\sqrt{2})^2} + \frac{b_3}{(1+\sqrt{2})^3} + \cdots \tag{4.1}$$

where $b_i \in \{0,1,2\}$ for every $i \geq 1$, but this representation is clearly not unique. Since

$$2(1+\sqrt{2})^{-1} + (1+\sqrt{2})^{-2} = 2(\sqrt{2}-1) + (3-2\sqrt{2}) = 1, \tag{4.2a}$$

and in general,

$$2(1+\sqrt{2})^{-i} + (1+\sqrt{2})^{-i-1} = (1+\sqrt{2})^{-i+1}, \tag{4.2b}$$

we can guarantee uniqueness by enforcing the following *Extra Rule* in (4.1):

$$b_i = 2 \quad \text{implies} \quad b_{i+1} = 0. \tag{4.3}$$

We also use a somewhat similar representation for any integer n; the novelty is *alternation*. To motivate this "alternating representation", I recall the fact that $(1+\sqrt{2})^j = p_j + q_j\sqrt{2}$, $j = 1,2,3,\ldots$ describes the whole family of positive solutions $1 \leq x = p_j$, $1 \leq y = q_j$ of the Pell equation $x^2 - 2y^2 = \pm 1$. It follows that $(1-\sqrt{2})^j = p_j - q_j\sqrt{2}$, implying

$$q_j\sqrt{2} - p_j = -(1-\sqrt{2})^j = \frac{(-1)^{j+1}}{(1+\sqrt{2})^j}, \tag{4.4}$$

that is, the distance of $q_j\sqrt{2}$ from the nearest integer has an alternating positive-negative behavior as $j = 1,2,3,\ldots$.

This alternating nature of $q_j\sqrt{2} - p_j$ motivates the following "alternating representation of integers": we search for n in the form with k odd

$$n = d_k q_k - d_{k-1}q_{k-1} + d_{k-2}q_{k-2} - d_{k-3}q_{k-3} + d_{k-4}q_{k-4} \mp \ldots \tag{4.5}$$

where $d_j \in \{0,1,2\}$ for all $j \leq k$, and $d_k \neq 0$. Here q_j is the "y" in the jth positive solution of the Pell equation $x^2 - 2y^2 = \pm 1$, that is,

$$q_j = \frac{(1+\sqrt{2})^j - (1-\sqrt{2})^j}{2\sqrt{2}}. \tag{4.6}$$

The reason why (4.1) and (4.5) form a perfect match is explained by the following argument:

$$n\sqrt{2}-\beta=\sum_{j=1}^{k}\Big((-1)^{j+1}d_jq_j\sqrt{2}-b_j(1+\sqrt{2})^{-j}\Big)-\sum_{i>k}b_i(1+\sqrt{2})^{-i},$$

and so by working modulo one we have (see (4.4))

$$n\sqrt{2}-\beta\equiv\sum_{j=1}^{k}\Big((-1)^{j+1}d_j(q_j\sqrt{2}-p_j)-b_j(1+\sqrt{2})^{-j}\Big)-\sum_{i>k}b_i(1+\sqrt{2})^{-i}$$

$$\equiv\sum_{j=1}^{k}\Big((d_j-b_j)(1+\sqrt{2})^{-j}\Big)-\sum_{i>k}b_i(1+\sqrt{2})^{-i}, \tag{4.7}$$

where (4.7) is a (mod 1) equality. Formula (4.7) tells us how to find an integer n such that $\|n\sqrt{2}-\beta\|$ is "very small". Assume that the $(1+\sqrt{2})$ scale representation of β (see (4.1)) has a long block of consecutive 0s: there is an odd k such that

$$b_k\neq 0,\ b_{k+1}=b_{k+2}=b_{k+1}=\cdots=b_{k+\ell}=0, \tag{4.8}$$

where ℓ is "large". Choose an integer n in the form (4.5) such that

$$d_j=b_j\quad\text{for } 1\le j\le k. \tag{4.9}$$

Then by (4.7)

$$\|n\sqrt{2}-\beta\|\le\sum_{i>k+\ell}b_i(1+\sqrt{2})^{-i}$$

$$\le\sum_{i>k+\ell}2(1+\sqrt{2})^{-i}=\frac{\sqrt{2}}{(1+\sqrt{2})^{k+\ell}} \tag{4.10}$$

where ℓ is defined by (4.10); it is the length of the zero-block.

Let's apply (4.9) in (4.5); I claim that the resulting n satisfies the lower bound (recall that k is odd)

$$n\ge q_{k-1}. \tag{4.11}$$

Indeed, since $d_k=b_k\neq 0$, by the Extra Rule we have $d_{k-1}=b_{k-1}\neq 2$, and so

$$n\ge q_k-q_{k-1}-2q_{k-3}-2q_{k-5}-q_{k-7}-\cdots. \tag{4.12}$$

We have

$$\begin{aligned} q_k - q_{k-1} &= q_{k-1} + q_{k-2}, \\ q_{k-2} - 2q_{k-3} &= q_{k-4}, \\ q_{k-4} - 2q_{k-5} &= q_{k-6}, \end{aligned}$$

and so on, and using these inequalities in (4.12) we obtain (4.11).

On the other hand, we have the easy upper bound

$$n \le q_{k+1}. \tag{4.13}$$

Indeed,

$$n \le 2q_k + 2q_{k-2} + 2q_{k-4} + 2q_{k-6} + \cdots , \tag{4.14}$$

and because $2q_i = q_{i+1} - q_{i-1}$, the right-hand side of (4.14) is a telescoping sum, implying (4.13).

By using formula (4.6) in (4.10) we have

$$\begin{aligned} \|n\sqrt{2} - \beta\| &\le \frac{\sqrt{2}}{(1+\sqrt{2})^{k+\ell}} = \frac{2\sqrt{2}}{(1+\sqrt{2})^{k+1}} \frac{1}{2(1+\sqrt{2})^{\ell-1}} \\ &< \frac{1}{q_{k+1}} \cdot \frac{1}{(1+\sqrt{2})^{\ell-1}} \le \frac{1}{n} \cdot \frac{1}{(1+\sqrt{2})^{\ell-1}}, \end{aligned} \tag{4.15}$$

where in the last step we used (4.13), $q_{k-1} \le n \le q_{k+1}$, and ℓ is defined by (4.8) (it is the length of the zero-block).

In view of (4.15) and (4.8) we can say that, the longer the zero-block in (4.8), the better inequality (4.15). This leads to the following question: Why is it true that almost all real numbers $0 < \beta < 1$ contain "long" zero-blocks of the "digits" in the $(1 + \sqrt{2})$ scale representation (see (4.1))?

The digit sequence $b_1 = b_1(\beta)$, $b_2 = b_2(\beta)$, $b_3 = b_3(\beta)$, ... in (4.1) does *not* form independent random variables as β runs in the unit interval $0 < \beta < 1$: the Extra Rule "$b_i = 2$ implies $b_{i+1} = 0$" clearly contradicts (statistical) independence. What we have here is in fact a *homogeneous Markov chain.*

A good way to define a (finite) *Markov chain* is to look at it as an asymmetric random walk on a (finite) directed graph (with loops and multiple edges). For example, our concrete Markov chain $b_1 = b_1(\beta)$, $b_2 = b_2(\beta)$, $b_3 = b_3(\beta)$, $\cdots$ can be visualized as an asymmetric random walk on a directed graph with 3 vertices. The vertices are officially called *states.* Our Markov chain has three states: 0, 1, and 2, representing the three possible values $b_k = b_k(\beta) = 0$ or 1 or 2.

A homogeneous Markov chain has a short term memory in the following sense: conditional upon the present, the future does *not* depend on the past.

The transition matrix $A = (p_{i,j})_{i,j}$ $(0 \le i, j \le 2)$ with the transition probabilities $p_{i,j}$ = "probability to go from state i to state j (in one step)" completely describes a homogeneous Markov chain. In our special case

$$A = \begin{pmatrix} p_{0,0} & p_{0,1} & p_{0,2} \\ p_{1,0} & p_{1,1} & p_{1,2} \\ p_{2,0} & p_{2,1} & p_{2,2} \end{pmatrix} = \begin{pmatrix} \tau & \tau & \tau^2 \\ \tau & \tau & \tau^2 \\ 1 & 0 & 0 \end{pmatrix} \tag{4.16}$$

where $\tau = \sqrt{2} - 1 = (\sqrt{2} + 1)^{-1}$.

The steady-state behavior (or long-term behavior) of the Markov chain is described by the stationary distribution $\mathbf{q} = (q_0, q_1, q_2)$, which is a probability distribution satisfying the fixpoint equation

$$\mathbf{q} = \mathbf{q}A, \quad \text{that is,} \quad q_j = q_0 p_{0,j} + q_1 p_{1,j} + q_2 p_{2,j}, \quad j = 0, 1, 2 \tag{4.17}$$

A simple calculation gives

$$q_0 = \frac{1}{2}, \qquad q_1 = \frac{\sqrt{2}}{4}, \qquad q_2 = \frac{2 - \sqrt{2}}{4} = \tau q_1 \tag{4.18}$$

The stationary (or fixpoint) distribution is also a limit distribution in the following sense (justifying the name "long-term behavior"). The kth power A^k of the transition matrix represents the k-step transition probabilities $p_{i,j}(k)$ = "probability to go from state i to state j in k steps": $A^k = (p_{i,j}(k))_{i,j}$ $(0 \le i, j \le 2)$. The eigenvalues λ_1, λ_2, λ_3 of the transition matrix are $\lambda_1 = 1$, $\lambda_2 = -\tau^2 = 2\sqrt{2} - 3$, $\lambda_3 = 0$, so we can rewrite the transition matrix as

$$A = B^{-1} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} B$$

for some invertible matrix B, and we obtain

$$A^k = B^{-1} \begin{pmatrix} \lambda_1^k & 0 & 0 \\ 0 & \lambda_2^k & 0 \\ 0 & 0 & \lambda_3^k \end{pmatrix} B = B^{-1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & (2\sqrt{2} - 3)^k & 0 \\ 0 & 0 & 0 \end{pmatrix} B \tag{4.19}$$

By (4.19) we have the following simple formula for the k-step transition probabilities:

$$p_{i,j}(k) = q_j + c_{i,j}(2\sqrt{2} - 3)^k, \tag{4.20}$$

where $q_0 = 1/2, q_1 = \sqrt{2}/4, q_2 = (2-\sqrt{2})/4$ is the stationary distribution (see (4.17) and (4.18)), and $c_{i,j}$ are appropriate constants independent of k. Since $|2\sqrt{2}-3| < 1$ (in fact $< 1/5$), (4.20) tells us that the k-step transition probability $p_{i,j}(k)$ converges to q_j, as $k \to \infty$, exponentially fast.

It is worthwhile to know the recipe how to determine the constant factors $c_{i,j}$ in (4.20). Comparing (4.16) to (4.20) with $k = 1$, we have

$$\tau = p_{0,0} = q_0 - c_{0,0}\tau^2 \implies c_{0,0} = \frac{1/2-\tau}{\tau^2},$$

and similarly,

$$\begin{aligned} c_{1,0} &= \frac{1/2-\tau}{\tau^2}, \qquad c_{2,0} = \frac{1}{2\tau^2}, \\ c_{0,1} &= \frac{\sqrt{2}/4-\tau}{\tau^2} = c_{1,1}, \qquad c_{2,1} = \sqrt{2}/4\tau^2, \\ c_{0,2} &= \frac{2-\sqrt{2}}{4\tau^2} - 1 = c_{1,2}, \qquad c_{2,2} = \frac{2-\sqrt{2}}{4\tau^2}. \end{aligned}$$

Even if equation (4.20) is relatively simple, it is rather inconvenient to work with the k-step transition probabilities $p_{i,j}(k)$. Luckily there is a simple way to go back to independence: the trick is to switch to “0”, “1”, “20” (instead of the original values “0”, “1”, “2”); formally,

$$b_1 b_2 b_3 \ldots = B_1 B_2 B_3 \ldots \qquad \text{where } B_i = 0 \text{ or } 1 \text{ or } 20. \tag{4.21}$$

The sequence defined by (4.21)

$$B_1(\beta),\ B_2(\beta),\ B_3(\beta), \ldots \qquad \text{as } 0 < \beta < 1 \tag{4.22}$$

does form independent random variables with common distribution

(4.23)
$$\Pr[B_i = 0] = \Pr[B_i = 1] = \sqrt{2}-1 \text{ and } \Pr[B_i = 20] = (\sqrt{2}-1)^2 = 3-2\sqrt{2},$$

where Pr (i.e., “probability”) means the ordinary one-dimensional Lebesgue measure.

The independence in (4.22) comes from self-similarity, that is to say, it is a corollary of the one-digit periodicity of the continued fraction $\sqrt{2} = [1; 2, 2, 2, \ldots]$.

Notice that (4.23) is just a restatement of (4.2)–(4.3), or using the Markov chain terminology, (4.23) comes from the first two rows of the transition matrix in (4.16) (the third row explains the use of "20" instead of "2").

A long zero-block of b_is is equivalent to a long zero-block of B_is (with a possible loss of "20" at the beginning), so we reduced our number-theoretic problem to a purely probabilistic question—long runs of 0s—for independent trials. The simplest probabilistic model comes from tossing a fair coin repeatedly, and then the analog problem is to study the long runs of Heads. This natural problem is somehow ignored by practically all textbooks of probability theory, so we have to make a detour here to solve this problem.

4.2. Long runs of Heads

Suppose that we toss a fair coin N times, and write down the outcomes; thus we obtain a TH-sequence where T and H stand for Tails and Heads: say $THHTHTHTT\dots THH$. Let $L = L(N)$ denote the length of the longest block of consecutive Heads. This L is a random variable with possible values $0 \le L \le N$. What is the typical size of $L = L(N)$? It is easy to guess that $L = L(N) \approx \log_2 N$ (binary logarithm of N). What is surprising is that $L = L(N)$ is in fact concentrated on a constant number of values $\log_2 N + O(1)$ centered at $\log_2 N$ with probability close to one. The following result gives the complete answer (we don't really need such a delicate result, but the proof is short and very instructive).

Lemma 4.1. *For simplicity assume that N is a power of two (i.e., $\log_2 N$ is an integer); then for any fixed integer d we have*

$$\Pr[L = L(N) = \log_2 N + d] = e^{-2^{-d-2}} - e^{-2^{-d-1}} + o(1), \tag{4.24}$$

where the error term $o(1)$ tends to zero if d is fixed and $N \to \infty$.

Remarks. As far as I know, the study of $L(n)$ goes back (at least) to Erdős and Rényi, but this particular result seems to be unpublished. I learned it from János Komlós (oral communication)—it is probably his (unpublished) theorem, or perhaps it is folklore; I don't know.

The maximum of the exponential expression in (4.24) is attained at $d = -1$, and $d = -2, 0, -3, 1$ give the remaining relatively large values in decreasing order:

$$\begin{aligned} e^{-1/2} - e^{-1} &= 0.2387 \quad \text{for } d = -1 \\ e^{-1} - e^{-2} &= 0.2325 \quad \text{for } d = -2 \\ e^{-1/4} - e^{-1/2} &= 0.1723 \quad \text{for } d = 0 \\ e^{-2} - e^{-4} &= 0.1174 \quad \text{for } d = -3 \\ e^{-1/8} - e^{-1/4} &= 0.1037 \quad \text{for } d = 1. \end{aligned}$$

Notice that the five values $d = -1, -2, 0, -3, 1$ represent more than 85 percent probability, so the longest run of Heads $L = L(N)$ is basically concentrated on a constant number of values $\log_2 N + O(1)$. There is nothing surprising about $\log_2 N$, but the extreme concentration around $\log_2 N + O(1)$, and the elegant limit theorem above is truly surprising.

Since this result is primarily an illustration, the proof below is somewhat sketchy; I leave some details of the calculations to the reader.

Proof of Lemma 4.1. We are going to prove

$$\Pr[L = L(N) \geq \log_2 N + d] = 1 - e^{-2^{-d-1}} + o(1). \tag{4.25}$$

Notice that (4.25) immediately implies (4.24); indeed,

$$\begin{aligned} \Pr[L = L(N) = \log_2 N + d] = \Pr[L = L(N) \\ &\geq \log_2 N + d] - \Pr[L = L(N) \geq \log_2 N + d + 1] \\ &= \left(1 - e^{-2^{-d-1}} + o(1)\right) - \left(1 - e^{-2^{-d-2}} + o(1)\right) \\ &= e^{-2^{-d-2}} - e^{-2^{-d-1}} + o(1). \end{aligned}$$

The proof of (4.25) uses the Inclusion-Exclusion Principle and the trick is to include the "T" at both endpoints of the longest run of Heads. This way the evaluation of the usually very messy Inclusion-Exclusion formula becomes a routine exercise for the Poisson paradigm.

More precisely, if $H \cdots H$ is the longest block of consecutive Heads, then there is a T at both ends (unless the block is already at the end, i.e., it begins at 1 or ends at N), and we consider the T-closed H-block $TH \cdots HT$, or possibly $H \cdots HT$ (if it begins at 1) or $TH \cdots H$ (if it ends at N). The crucial property of the T-closed H-blocks is that they cannot overlap, except that they may share a common T at the end.

Let $E(i, j)$ denote the event that the ith outcome is Tails, the rth outcome is Heads with $i + 1 \leq r \leq i + j$, and the $(i + j + 1)$st outcome is Tails again; this is a typical T-closed H-block $TH\cdots HT$. Also, let $E_{start}(j)$ denote the event that the rth outcome is Heads with $1\le r\le j$, and the $(j+1)$st outcome is Tails; this is a T-closed H-block $H\cdots HT$ that begins at 1. Finally, let $E_{end}(j)$ denote the event that the $(N-j)$th outcome is Tails, and the rth outcome is Heads with $N-j+1\le r\le N$; this is a T-closed H-block $H\cdots HT$ that ends at N. Notice that the event $\{L=L(N)\ge \log_2 N+d\}$ is the union of the events

$$A_1=\bigcup E(i,j) \quad \text{where } 1\le i,\ i+j+1\le N,\ j\ge \log_2 N+d \tag{4.26}$$

and

$$A_2=\bigcup E_{start}(j) \quad \text{where } N\ge j\ge \log_2 N+d \tag{4.27}$$

and

$$A_3=\bigcup E_{end}(j) \quad \text{where } N\ge j\ge \log_2 N+d. \tag{4.28}$$

To compute the probability of a union of events, we have to use the Inclusion-Exclusion formula

(4.29)
$$\Pr\left[\cup_i E_i\right]=\sum_i \Pr[E_i]-\sum_{i_1<i_2}\Pr[E_{i_1}\cap E_{i_2}]+\sum_{i_1<i_2<i_3}\Pr[E_{i_1}\cap E_{i_2}\cap E_{i_3}]\mp\cdots$$

(4.29) is rather hopeless in general (it contains too many terms), but because of symmetry and the non-overlapping of the T-closed H-blocks, for the unions (4.26)–(4.28) this turns out to be a relatively simple calculation.

Note that

(4.30)
$$\Pr[E(i,j)]=2^{-j-2},\qquad \Pr[E_{start}(j)]=2^{-j-1},\qquad \Pr[E_{end}(j)]=2^{-j-1},$$

and so the linear part of (4.29) (see the first sum on the right-hand side) is easy to evaluate: let $n = \log_2 N$, then

$$\begin{aligned}\text{Linear}[A_1] &= \sum_{n+d\le j\le N-2}\ \sum_{1\le i\le N-j-1} 2^{-j-2}\\ &= \sum_{n+d\le j\le N-2}\ \sum_{1\le i\le N-j-1} 2^{-j-2}\\ &= \sum_{n+d\le j\le N-2} (N-j-1)2^{-j-2}\\ &= (N-n-d-1)\sum_{j\ge n+d} 2^{-j-2}\\ &\quad - 2^{-n-d-1}\left(\frac{1}{2}+2\left(\frac{1}{2}\right)^2+3\left(\frac{1}{2}\right)^3+\cdots\right)\\ &= (N-n-d-1)2^{-n-d-1} - 2^{-n-d-1} + O(N2^{-N}). \qquad (4.31)\end{aligned}$$

A similar but simpler argument gives

$$\text{Linear}[A_2] = \text{Linear}[A_3] = 2^{-n-d} + O(2^{-N}). \qquad (4.32)$$

Summarizing, by (4.31)–(4.32) the linear part of (4.29) with $A_1 \cup A_2 \cup A_3$ equals

$$\begin{aligned}\text{Linear part} &= (N-n-d+2)2^{-n-d-1} + O(N2^{-N})\\ &= \left(1-\frac{n+d-2}{N}\right)2^{-d-1} + O(N2^{-N}), \qquad (4.33)\end{aligned}$$

because $N = 2^n$. We can further simplify (4.33) to the very short form

$$\text{Linear part} = 2^{-d-1} + \text{negligible}. \qquad (4.34)$$

Next consider the contribution of the "pairwise intersections" in (4.29) with $A_1 \cup A_2 \cup A_3$. We don't want the exact solution, we just want to find the analog of the very short form (4.34). If two T-closed H-blocks overlap (more than just sharing a common T), then the intersection of the corresponding events has zero probability, so this case has no contribution in (4.29). Otherwise we have

$$\Pr[E(i_1,j_1)\cap E(i_2,j_2)] = 2^{-j_1-2-j_2-2} \text{ or } 2^{-j_1-2-j_2-1}$$

depending on whether the two T-closed H-blocks are disjoint or share a common T. It is clear that the main contribution comes from the disjoint case, the rest are negligible. Formally,

$$\begin{aligned}\text{Pairwise intersections} &= \binom{N}{2}\left(\sum_{j_1\geq n+d} 2^{-j_1-2}\right)\left(\sum_{j_2\geq n+d} 2^{-j_2-2}\right)\\ &\quad+\text{negligible}\\ &= \binom{N}{2}\left(2^{-n-d-1}\right)^2 \ +\text{negligible}\\ &= \frac{1}{2}\left(2^{-d-1}\right)^2+\text{negligible}. \qquad (4.35)\end{aligned}$$

Similarly, the contribution of the triple intersections in (4.29) with $A_1\cup A_2\cup A_3$ equals

$$\begin{aligned}\text{Triple intersections} &= \binom{N}{3}\left(\sum_{j_1\geq n+d} 2^{-j_1-2}\right)\left(\sum_{j_2\geq n+d} 2^{-j_2-2}\right)\left(\sum_{j_3\geq n+d} 2^{-j_3-2}\right)\\ &\quad+\text{negligible}\\ &= \binom{N}{3}\left(2^{-n-d-1}\right)^3+\text{negligible}\\ &= \frac{1}{3!}\left(2^{-d-1}\right)^3+\text{negligible}, \qquad (4.36)\end{aligned}$$

and so on.

Summarizing, by (4.29), (4.34)–(4.36),

$$\begin{aligned}\Pr[L=L(N)\geq \log_2 N+d] &= \lambda-\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}\mp\cdots \ + \ \text{negligible}\\ &= 1-e^{-\lambda} \ + \ \text{negligible}, \qquad (4.37)\end{aligned}$$

where $\lambda=2^{-d-1}$. Here the error term indicated by the vague term "negligible" tends to zero as d remains fixed and $N\to\infty$; I challenge the reader to double-check this fact. Thus (4.37) proves (4.25), completing the proof of Lemma 4.1. □

The goal of Lemma 4.1 was to justify the main term $\log_2 N$. But what we are really interested in is the behavior of the long runs in an *infinite* sequence of independent trials (Heads-and-Tails). Let $L_\infty(N)$ denote the length of the longest run of Heads among the first N trials (coin-tossings), $N=1,2,3,\ldots$. Lemma 4.1 describes the typical behavior for a fixed N: the longest run is $\log_2 N + O(1)$ with probability close to one. We are going to prove that, with probability one, there are infinitely many values of N such that the surplus $L_\infty(N) - \log_2 N$ tends to infinity with the rate of the iterated logarithm.

More precisely, we prove

Lemma 4.2. *With probability one, there are infinitely many values of N such that*

$$L_\infty(N) > \log_2 N + \log_2 \log N. \tag{4.38}$$

On the other hand, with probability one, for any $\varepsilon > 0$ we have the upper bound

$$L_\infty(N) < \log_2 N + (1+\varepsilon) \log_2 \log N \tag{4.39}$$

for all sufficiently large N.

Remarks. The intuitive reason behind (4.38)–(4.39) is divergence-convergence:

$$\sum_{N\ge 2} 2^{-\log_2 N - \log_2 \log N} = \sum_{N \ge 2} \frac{1}{N \log N} = \infty, \tag{4.40}$$

but

$$\sum_{N\ge 2} 2^{-\log_2 N - (1+\varepsilon)\log_2 \log N} = \sum_{N \ge 2} \frac{1}{N (\log N)^{1+\varepsilon}} < \infty. \tag{4.41}$$

Proof. The **proof of Lemma 4.2** is similar to that of Lemma 4.1; the minor difference is that the Inclusion-Exclusion formula is replaced by Chebyshev's well-known inequality. Again the non-overlapping property of the T-closed H-blocks plays a crucial technical role to simplify the calculations in the proof (see Cases 1-2-3 after (4.46)).

Let F_n denote the event that the $(n-k)$th outcome is Tails, the rth outcome is Heads with $n-k+1 \le r \le n$, and the $(n+1)$st outcome is Tails again, where $k = k(n) = \log_2 n + \log_2 \log n$ (take the lower integral part, and assume that $n > 10$). Event F_n means a particular T-closed H-block $TH\cdots HT$; the probability is 2^{-k-2}. Then by (4.40)

$$\sum_{n>10} \Pr[F_n] = \infty. \tag{4.42}$$

Let χ_n denote the characteristic function of event F_n: $\chi_n = 1$ or 0 depending on whether F_n holds or fails. The sum

$$X_M = \sum_{10<n\leq M} \chi_n \tag{4.43}$$

counts the number of times inequality (4.38) holds for some $10 < n \leq M$.

The expectation of X_M is

$$\begin{aligned} \mathbf{E}X_M &= \sum_{10<n\leq M} \Pr[F_n] = \sum_{10<n\leq M} 2^{-\log_2 n - \log_2 \log n - 2} \\ &= \sum_{10<n\leq M} \frac{1}{4n\log n} = \frac{1}{4}\log\log M \ + \ O(1). \end{aligned} \tag{4.44}$$

We want to show that the random variable X_M is typically "close" to its expected value (4.44). The standard way to do this is to apply the Chebyshev inequality. We need to compute the variance:

$$\begin{aligned} Var(X_M) = \mathbf{E}\left(\sum_{10<n\leq M} (\chi_n - E_n)\right)^2 &= \sum_{10<n\leq M} \mathbf{E}(\chi_n - E_n)^2 \\ &+ 2 \sum_{10<n_1<n_2\leq M} \mathbf{E}(\chi_{n_1} - E_{n_1})(\chi_{n_2} - E_{n_2}), \end{aligned} \tag{4.45}$$

where

$$E_n = \mathbf{E}\chi_n = \Pr[F_n]$$

is the expectation of χ_n. Clearly $(n_1 \neq n_2)$

$$\mathbf{E}(\chi_{n_1} - E_{n_1})(\chi_{n_2} - E_{n_2}) = \mathbf{E}(\chi_{n_1}\chi_{n_2} - E_{n_1}E_{n_2}), \tag{4.46}$$

so we have to study $\mathbf{E}\chi_{n_1}\chi_{n_2}$. There are three cases.

Case 1: If the corresponding T-closed H-blocks are disjoint, then $\mathbf{E}\chi_{n_1}\chi_{n_2} = E_{n_1}E_{n_2}$, which has zero contribution in (4.46).

Case 2: If the corresponding T-closed H-blocks are "touching", i.e., they share a common T, then $\mathbf{E}\chi_{n_1}\chi_{n_2} = 2E_{n_1}E_{n_2}$.

Case 3: If the corresponding T-closed H-blocks are overlapping (more than just touching), then $\Pr[F_{n_1}F_{n_2}] = 0$ (i.e., this case is impossible), which implies $\mathbf{E}(\chi_{n_1}\chi_{n_2}) = 0$. Therefore this case has negative(!) contribution in (4.46).

Let's return to (4.45). The contribution of the diagonal part

$$\sum_{10<n\le M} \mathbf{E}(\chi_n - E_n)^2$$

in (4.45) is less than $\mathbf{E}X_M$, since

$$\mathbf{E}(\chi_n - E_n)^2 = \Pr[F_n](1 - \Pr[F_n]) < \Pr[F_n].$$

Also, the contribution of Case 2 in the off-diagonal part of (4.45) is less than

$$2 \sum_{10<n\le M} \Pr\left[F_n\right] = 2\mathbf{E}X_M.$$

Thus we have the upper bound

$$(4.47) \qquad Var(X_M) = \mathbf{E}\left(\sum_{10<n\le M} (\chi_n - E_n)\right)^2 < 3\mathbf{E}X_M.$$

We apply Chebyshev's inequality

$$(4.48) \qquad \Pr\left[|X - \mathbf{E}X| \ge \lambda\right] \le \frac{Var(X)}{\lambda^2},$$

which holds for any random variable X (with finite variance) and any positive real λ: let

$$X = X_M \quad \text{and} \quad \lambda = \frac{1}{2}\mathbf{E}X_M.$$

Then by (4.47)

$$(4.49) \qquad \Pr\left[X_M \ge \frac{1}{2}\mathbf{E}X_M\right] \ge 1 - \frac{12}{\mathbf{E}X_M}.$$

In view of (4.44) we have $\mathbf{E}X_M \to \infty$ as $M \to \infty$. Combining this with (4.49) we conclude that, with probability one, inequality (4.38) has infinitely many solutions. This proves the first part of Lemma 4.2.

The second part is almost trivial. Indeed, let G_n denote the event that the $(n-\ell)$th outcome is Tails, the rth outcome is Heads with $n - \ell + 1 \le r \le n$, and the $(n+1)$st outcome is Tails again, where $\ell = \ell(n) = \log_2 n + (1+\varepsilon)\log_2 \log n$ with some fixed $\varepsilon > 0$ (again we take the lower integral part, and assume that $n > 10$). This means a particular T-closed H-block $TH \cdots HT$ that has probability $2^{-\ell-2}$. Then by (4.41)

$$\sum_{n>10} \Pr[G_n] < \infty. \tag{4.50}$$

Consider the event

$$H = \bigcap_{n>10} \bigcup_{m \geq n} G_m; \tag{4.51}$$

by (4.50) the probability of H is zero. This proves the second part of Lemma 4.2. □

A detour: the two Borel–Cantelli Lemmas The last argument is often called the "easy part of the Borel–Cantelli Lemma". The "harder part of the Borel–Cantelli Lemma" is some kind of a converse: it states that, if G_n is an infinite sequence of independent events with

$$\sum_{n} \Pr[G_n] = \infty, \tag{4.52}$$

then the probability of event H (see (4.51)) is one. Notice that with $G_n = F_n$ we cannot apply this criterion, since the events F_n are not independent if the corresponding T-closed H-blocks are touching or overlapping. This is why we couldn't apply the "harder part of the Borel–Cantelli Lemma", and had to turn to the Chebyshev inequality instead. It is important to notice that the correct proof with the Chebyshev inequality gives exactly the same divergence condition as the incorrect argument applying the "harder part of the Borel–Cantelli Lemma".

Repeating the proof of Lemma 4.2 one can easily prove the following more general convergence-divergence type result.

Lemma 4.3. *If $\varphi(N)$ is any increasing function of N for which*

$$\sum_{N} \frac{d}{\varphi(N)} = \infty,$$

then with probability one, the longest run of Heads up to N satisfies the lower bound

$$L_\infty(N) > \log_2 \varphi(N) \tag{4.53}$$

for infinitely many values of N.

On the other hand, if

$$\sum_N \frac{d}{\varphi(N)} < \infty,$$

then with probability one, the longest run of Heads up to N satisfies the upper bound

$$L_\infty(N) < \log_2 \varphi(N)$$

for all sufficiently large values of N.

Notice that in the special cases

$$\varphi(N) = N \log N \quad \text{and} \quad \varphi(N) = N(\log N)^{1+\varepsilon}$$

we get back Lemma 4.2.

Proof. Now we are ready to prove Theorem 2 (i.e., the Area Principle for slope $\sqrt{2}$). Let's return to (4.15) and (4.21)–(4.23). In view of (4.23) we have to replace the fair coin with an asymmetric discrete probability distribution: B_1, B_2, B_3, ... are independent and identically distributed random variables having values 0, 1 and "20" with the distribution

(4.54)
$$\Pr[B_i = 0] = \Pr[B_i = 1] = \sqrt{2} - 1 \text{ and } \Pr[B_i = 20] = (\sqrt{2} - 1)^2 = 3 - 2\sqrt{2},$$

and we are interested in the long runs of 0s. Let $L^*_\infty(N)$ denote the length of the longest run of 0s among B_1, B_2, ..., B_N; of course, $L^*_\infty(N)$ is a random variable. Lemma 4.3 is about the longest run of Heads in tossing a fair coin repeatedly, and for a fair coin $\Pr[Heads] = 1/2$. Since in our case

$$\Pr[B_i = 0] = \sqrt{2} - 1 = \frac{1}{1 + \sqrt{2}},$$

it is perfectly reasonable to expect the following analog of (4.53): with probability one

$$L^*_\infty(N) > \log_b \varphi(N) \tag{4.55}$$

for infinitely many values of N, where $\log_b$ denotes the base $b = 1 + \sqrt{2}$ logarithm and $\varphi(N)$ is any positive increasing function of N for which

$$\sum_N \frac{d}{\varphi(N)} = \infty. \tag{4.56}$$

The proof of (4.55)–(4.56) is the same as that of Lemma 4.3; I leave it to the reader.

If condition (4.56) applies, then by (4.55) there are infinitely many zero-blocks

$$B_k \neq 0, \qquad B_{k+1} = B_{k+2} = \cdots = B_{k+\ell} = 0 \tag{4.57}$$

where $\ell = \ell(k)$ satisfies the equation $\ell = \log_b \varphi(k+\ell)$; the base of the logarithm is $b = 1 + \sqrt{2}$. Because of independence, we have

$$\Pr[B_{k+1} = B_{k+2} = \cdots = B_{k+\ell} = 0] = b^{-\ell} = \frac{1}{\varphi(k+\ell)}. \tag{4.58}$$

There is a technical nuisance due to the slight difference between the B-indexing and the b-indexing in (4.21): this is the effect of the pairs "20". More precisely, if $B_i = 0$ then $B_i = b_j$ where $j = i + i_2$ and i_2 denotes the number of pairs $B_k = 20$ with $k < i$. By the strong law of large numbers, with probability one, $i_2/i \to 3 - 2\sqrt{2}$ as $i \to \infty$ (see (4.54)). It follows that $0 = B_i = b_j = b_{j(i)}$ implies

$$j = j(i) = (1 + (3 - 2\sqrt{2}) + o(1))i \text{ as } i \to \infty. \tag{4.59}$$

Thus we can rewrite (4.57):

$$b_{j+1} = b_{j+2} = \cdots = b_{j+\ell} = 0 \qquad \text{where } j \leq (1 + (3 - 2\sqrt{2}) + o(1))k. \tag{4.60}$$

Then by (4.9)–(4.11) there is an integer n in $q_j \leq n < q_{j+1}$ such that

$$\|n\sqrt{2} - \beta\| \leq \frac{1}{n\varphi(k+l)} \leq \frac{1}{n\varphi(\log_c n)}; \tag{4.61}$$

here $c > 1$, the base of the logarithm, is some appropriate constant.

Note that, by using the substitution $y = \log x$, we have the equidivergence property:

$$\begin{aligned} &\sum_{n=2}^{\infty} \frac{1}{n\varphi(\log n)} = \infty \Longleftrightarrow \int_2^{\infty} \frac{dx}{x\varphi(\log x)} = \infty \\ &\Longleftrightarrow \int_1^{\infty} \frac{dy}{\varphi(y)} = \infty \Longleftrightarrow \sum_{n=1}^{\infty} \frac{1}{\varphi(n)} = \infty; \end{aligned} \tag{4.62}$$

and here the base of the logarithm is irrelevant. In view of this, the technical nuisance due to the discrepancy between the B-indexing and b-indexing (see (4.59)) is also irrelevant. Therefore, (4.61)–(4.62) complete the proof of Theorem 2. □

5. The Area Principle in general: Proof of Theorem 3

We heavily use the theory of continued fractions. (Of course, this is not very surprising, since the complete solution of the homogeneous inequality (2.18), or (1.15), was determined by Euler and Lagrange exactly by using the very same tool: the theory of continued fractions.) Also, at the end of the proof, we will apply the Chebyshev inequality.

I begin with the so-called *Ostrowski representation* of integers with respect to any fixed irrational $0 < \alpha < 1$, given by the continued fraction

$$\alpha = \frac{d}{a_1 + \frac{1}{a_2 + \cdots}} = [a_1, a_2, a_3, \dots],$$

$[a_1, a_2, \dots, a_{k-1}] = p_k/q_k$ with $q_1 = 1$, $q_2 = a_1$, $q_n = a_{n-1}q_{n-1} + q_{n-2}$ for all $n \geq 3$. Since $q_n = a_{n-1}q_{n-1} + q_{n-2}$, every positive integer n can be written in the form

$$n = \sum_{i=1}^{k} d_i q_i, \ d_i \quad \text{are integers} \tag{5.1}$$

where $0 \leq d_i \leq a_i$ (see [Os22]).

An analog of the Ostrowski representation of integers can be developed for the representation of the real number β. Write

$$\theta_n = q_n\alpha - p_n, \qquad \text{then } \theta_n = a_{n-1}\theta_{n-1} + \theta_{n-2}. \tag{5.2}$$

Note that

$$\theta_n = (-1)^{n-1}|\theta_n|, \qquad \text{and} \qquad |\theta_{n-2}| = a_{n-1}|\theta_{n-1}| + |\theta_n|. \tag{5.3}$$

In the theorem we can assume without loss of generality that $0 < \alpha < 1$; so $\theta_1 = \alpha > 0$ and $\theta_2 = a_1\alpha - 1 < 0$.

Now every real number β in the interval $-\alpha \leq \beta < 1-\alpha$ of length one (any interval of length one is fine, since the theorem is about modulo one) can be written in the form

$$\beta = \sum_{i=1}^{\infty} b_i\theta_i, \ b_i \quad \text{are integers,} \tag{5.4}$$

where $0 \leq b_1 \leq a_1 - 1$ and $0 \leq b_i \leq a_i$ for $i \geq 2$. We can make representation (5.4) unique by enforcing the Extra Rule

$$b_i = a_i \quad \text{implies} \quad b_{i-1} = 0 \quad \text{for all } i \geq 2, \tag{5.5}$$

and we also require that

$$b_{2i+1} \neq a_{2i+1} \quad \text{for infinitely many } i. \tag{5.6}$$

Note that the minimum value of representation (5.4)–(5.6) is attained at

$$\begin{aligned} &a_2\theta_2 + a_4\theta_4 + a_6\theta_6 + \cdots \\ &\quad = (-\theta_1 + \theta_3) + (-\theta_3 + \theta_5) + (-\theta_5 + \theta_7) + \cdots = -\theta_1 = -\alpha, \end{aligned} \tag{5.7}$$

and similarly, the maximum value of representation (5.4)–(5.6) is attained at

$$\begin{aligned} &(a_1 - 1)\theta_1 + a_3\theta_3 + a_5\theta_5 + \cdots \\ &\quad = (a_1 - 1)\theta_1 + (-\theta_2 + \theta_4) + (-\theta_4 + \theta_6) + \cdots \\ &\quad = (a_1 - 1)\theta_1 - \theta_2 = (a_1 - 1)\alpha - (1 - a_1\alpha) = (1 - \alpha), \end{aligned} \tag{5.8}$$

but because of (5.6), equality in (5.8) cannot occur. This explains the interval $-\alpha \leq \beta < 1 - \alpha$.

Inserted Remark

Note that representation (5.4)–(5.6) was independently introduced by Cassels [Ca54], Descombes [De56] and V.T. Sós [S58], and it was constantly used by V.T. Sós in her research of studying the irregularities of the irrational rotation (see e.g. [S74] and [S83]).

By (5.1) and (5.4) (we use $\equiv$ to indicate equality modulo one)

$$\begin{aligned} n\alpha-\beta &= \sum_{i=1}^{k} d_i q_i\alpha - \sum_{i=1}^{\infty} b_i\theta_i \\ &\equiv \sum_{i=1}^{k} d_i(q_i\alpha-p_i) - \sum_{i=1}^{\infty} b_i(q_i\alpha-p_i) \\ &\equiv \sum_{i=1}^{k}(d_i-b_i)\theta_i - \sum_{j>k} b_j\theta_j \pmod 1. \end{aligned} \tag{5.9}$$

The term $\|n\alpha-\beta\|$ is particularly small if

$$d_i = b_i \quad \text{for } 1\le i\le k \tag{5.10}$$

and also

$$0 = b_{k+1} = b_{k+2} = \cdots = b_{k+\ell}, \tag{5.11}$$

meaning a relatively long zero-block of ℓ consecutive coefficients b_j—the same idea as in Section 4. By (5.9)–(5.11)

$$\|n\alpha-\beta\| \le \left|\sum_{j>k+\ell}^{\infty} b_j\theta_j\right|; \tag{5.12}$$

the larger ℓ, the better inequality (5.12).

First we need the technical

Lemma 5.1. *If* $b_m\neq 0$ *then*

$$\left|\sum_{j=m}^{\infty} b_j\theta_j\right| \le b_m|\theta_m| + |\theta_{m+1}|. \tag{5.13}$$

Proof. We have

$$\begin{aligned} &(-1)^{m-1}\left(\sum_{j=m}^{\infty} b_j\theta_j\right) \\ &\quad = b_m|\theta_m| - b_{m+1}|\theta_{m+1}| + b_{m+2}|\theta_{m+2}| - b_{m+3}|\theta_{m+3}| \pm \cdots \\ &\quad \ge b_m|\theta_m| - b_{m+1}|\theta_{m+1}| - b_{m+3}|\theta_{m+3}| - b_{m+5}|\theta_{m+5}| - \cdots \end{aligned} \tag{5.14}$$

Since $b_m \neq 0$ we have $b_{m+1} \leq a_{m+1} - 1$, and using the recurrence formula (5.3): $|\theta_{n-2}| = a_{n-1}|\theta_{n-1}| + |\theta_n|$ repeatedly, we obtain

$$\begin{aligned} b_m|\theta_m| - b_{m+1}|\theta_{m+1}| &\geq |\theta_{m+1}| + |\theta_{m+2}|, \\ |\theta_{m+2}| - b_{m+3}|\theta_{m+3}| &\geq |\theta_{m+4}|, \\ |\theta_{m+4}| - b_{m+5}|\theta_{m+5}| &\geq |\theta_{m+6}|, \end{aligned}$$

and so on. Applying these inequalities in (5.14), we have

$$(-1)^{m-1}\left(\sum_{j=m}^{\infty} b_j\theta_j\right) \geq (b_m - 1)|\theta_m| + |\theta_{m+1}|. \tag{5.15}$$

On the other hand, by a telescoping sum argument:

$$\begin{aligned} (-1)^{m-1}\left(\sum_{j=m}^{\infty} b_j\theta_j\right) &\leq b_m|\theta_m| + b_{m+2}|\theta_{m+2}| + b_{m+4}|\theta_{m+4}| + \cdots \\ &\leq b_m|\theta_m| + (|\theta_{m+1}| - |\theta_{m+3}|) \\ &\quad + (|\theta_{m+3}| - |\theta_{m+5}|) + (|\theta_{m+5}| - |\theta_{m+7}|) + \cdots \\ &= b_m|\theta_m| + |\theta_{m+1}|. \end{aligned} \tag{5.16}$$

(5.15)–(5.16) prove Lemma 5.1. □

I recall the following well-known fact from the theory of continued fraction:

$$\left|\alpha - \frac{p_m}{q_m}\right| < \frac{1}{q_m q_{m+1}} \Longleftrightarrow |\theta_m| = |q_m\alpha - p_m| < \frac{d}{q_{m+1}}. \tag{5.17}$$

By Lemma 5.1 and (5.17) we have the following upper bound in (5.12):

$$\|n\alpha - \beta\| < \frac{1 + b_{k+\ell+1}}{q_{k+\ell+2}}, \tag{5.18}$$

assuming $b_{k+\ell+1} \neq 0$ and (5.11) holds. Condition (5.11) defines an integer n such that

$$b_k q_k \leq n = \sum_{i=1}^{k} b_i q_i \leq (b_k + 2)q_k. \tag{5.19}$$

Now assume that the Area Principle fails for the homogeneous inequality (2.18); then by (5.17)

$$\psi(q_m) < \frac{1}{q_{m+1}} \quad \text{for all } m \geq m_0. \tag{5.20}$$

Let $B_1 \in \{1, \ldots, a_k\}$ be fixed with $k \geq m_0$, and, motivated by (5.19), we find a $j = j(B_1)$ such that

$$q_j < \frac{d}{\psi((B_1+2)q_k)} < q_{j+1}. \tag{5.21}$$

By (5.20)

$$\frac{d}{\psi((B_1+2)q_k)} > \frac{d}{\psi(q_k)} > q_{k+1},$$

implying $j = j(B_1) \geq k+1$. We choose a $B_2 \in \{1, \ldots, a_j\}$ such that

$$\frac{1/10}{\psi((B_1+2)q_k)} \leq \frac{q_{j+1}}{B_2} \leq \frac{d}{\psi((B_1+2)q_k)}. \tag{5.22}$$

Since $j = j(B_1) \geq k+1$, with some appropriate integer $\ell \geq 0$ we can write $j = k+1+\ell$, and define the set $S(b_k = B_1, b_{k+\ell+1} = B_2)$ as the following subset of $[-\alpha, 1-\alpha)$ (see expansion (5.4)):

$$\begin{aligned} S(b_k = B_1, b_{k+\ell+1} = B_2) = \{\beta \in [-\alpha, 1-\alpha) :\ & b_k = B_1, \\ & 0 = b_{k+1} = \cdots = b_{k+\ell},\ b_{k+\ell+1} = B_2\}. \end{aligned} \tag{5.23}$$

If $\beta \in S(b_k = B_1, b_{k+\ell+1} = B_2)$ (see (5.21)–(5.23)) then by (5.18), (5.19), (5.22) the inhomogeneous inequality

$$\|n\alpha - \beta\| = O(\psi(n)), \text{ where the implicit constant is absolute,} \tag{5.24}$$

has an integral solution n with

$$B_1 q_k \leq n \leq (B_1+2)q_k. \tag{5.25}$$

Next we compute the Lebesgue measure Meas(S) of the sets $S = S(b_k = B_1, b_{k+\ell+1} = B_2)$ defined by (5.21)–(5.23).

Lemma 5.2. *With any $B_2 \in \{1, \ldots, a_{k+\ell+1}\}$ we have*

$$\operatorname{Meas}\left(S(b_k = B_1, b_{k+\ell+1} = B_2)\right) = \begin{cases} q_k|\theta_{k+\ell+1}|, & \textit{if } B_1 \neq a_k; \\ q_{k-1}|\theta_{k+\ell+1}|, & \textit{if } B_1 = a_k. \end{cases}$$

Proof. Let $\sharp(b_1,\dots,b_{k-1})$ denote the number of permissible sequences $(b_1, \dots, b_{k-1})$ satisfying (5.4)–(5.6). Clearly $\sharp(b_1)=a_1=q_2$, $\sharp(b_1,b_2)=a_1a_2+1=q_3$, and $\sharp(b_1,\dots,b_{k-1})$ satisfies the same recurrence as q_i: $q_i=a_{i-1}q_{i-1}+q_{i-2}$, and so we have

$$\sharp(b_1,\dots,b_{k-1})=\begin{cases} q_k, & \text{if } b_k=B_1\neq a_k;\\ q_{k-1}, & \text{if } b_k=B_1=a_k.\end{cases} \tag{5.26}$$

Next we study the tail series

$$\sum_{i=k+\ell+2}^{\infty} b_i\theta_i=\tau. \tag{5.27}$$

Since $b_{k+\ell+1}=B_2\neq 0$, we have $0\le b_{k+\ell+2}\le a_{k+\ell+2}-1$. Repeating the argument (5.7)–(5.8) we have

$$(-1)^{k+\ell}\tau\le (a_{k+\ell+2}-1)|\theta_{k+\ell+2}|+|\theta_{k+\ell+3}|, \tag{5.28}$$

and also

$$(-1)^{k+\ell}\tau\ge -|\theta_{k+\ell+2}| \tag{5.29}$$

(note that (5.29) is an analog of (5.7) and (5.28) is an analog of (5.8)). It follows that the tail series (5.27) covers an interval of length

$$a_{k+\ell+2}|\theta_{k+\ell+2}|+|\theta_{k+\ell+3}|=|\theta_{k+\ell+1}|. \tag{5.30}$$

Equations (5.26) and (5.30) prove Lemma 5.2. □

Next we estimate the total sum of the measures:

$$\begin{aligned}&\sum_{\substack{(5.21)-(5.23):\\ k\ge m_0}} \text{Meas}\,(S(b_k=B_1,b_{k+\ell+1}=B_2))\\ &\ge \text{const}\sum_{k\ge m_0}\sum_{B_1=1}^{a_k}\sum_{B_2}\frac{q_k \text{ or } q_{k-1}}{q_{k+\ell+2}}\\ &\ge \text{const}\sum_{k\ge m_0}\sum_{B_1=1}^{a_k}\frac{q_k \text{ or } q_{k-1}}{q_{k+\ell+2}}\psi((B_1+2)q_k)q_{k+\ell+2}\end{aligned}$$

$$
\begin{aligned}
&= \text{const} \sum_{k \geq m_0} \left(\sum_{B_1=1}^{a_k-1} q_k \psi((B_1+2)q_k) + q_{k-1}\psi((B_1+2)q_k) \right) \\
&\geq \text{const} \sum_{n \geq q_{m_0}} \psi(n) = \infty, \qquad (5.31)
\end{aligned}
$$

where we used Lemma 5.2, (5.22), m_0 is defined by (5.20), and as usual, *const* stands for a positive absolute constant factor.

In view of (5.24)–(5.25) it suffices to show that almost all $\beta \in [-\alpha, 1-\alpha)$ are contained by infinitely many sets $S(b_k = B_1, b_{k+\ell+1} = B_2)$ defined by (5.21)–(5.23). Equation (5.31) was the first step in this direction. But we also need information about the Lebesgue measure of the pairwise intersections

$$
S(b_{k_1} = B_1, b_{k_1+\ell_1+1} = B_2) \cap S(b_{k_2} = B_3, b_{k_2+\ell_2+1} = B_4). \qquad (5.32)
$$

We can assume $k_1 < k_2$; then intersection (5.32) is the empty set, unless $k_1+\ell_1+1 < k_2$, or possibly $k_1+\ell_1+1 = k_2$, $B_2 = B_3$. Let $d = k_2-k_1-\ell_1-1$ denote the "distance"; we prove that (5.32) is exponentially close to the product rule in terms of the distance d. This means "exponentially weak dependence", a phenomenon well-known among the experts of continued fraction. For example, this fact has been constantly used by V.T. Sós in her research concerning the "strong irregularities" of the irrational rotation; see [S83]. The following useful counting lemma is taken from Sós's paper.

Lemma 5.3. *For every $r \leq t$, let $A_{r,t}(B)$ denote the number of sequences $(b_r, b_{r+1}, \ldots, b_t)$ such that*

$$
b_r = B \in \{1, \ldots, a_r\}, \quad 0 \leq b_i \leq a_i
$$

and $b_i = a_i$ implies $b_{i-1} = 0$ for every i in $r < i \leq t$. Then

$$
A_{r,t}(B) = q_{t+1}|\theta_r| + (-1)^{t-r} q_r |\theta_{t+1}|. \qquad (5.33)
$$

Proof. By definition $A_{r,r}(B) = 1$. We double-check (5.33) in the special case $t = r$ by computing the right-hand side of (5.33):

$$
\begin{aligned}
&q_{r+1}|q_r\alpha - p_r| + q_r|q_{r+1}\alpha - p_{r+1}| \\
&\quad = q_{r+1}(-1)^r(q_r\alpha - p_r) + q_r(-1)^{r+1}(q_{r+1}\alpha - p_{r+1}) \\
&\quad = (-1)^r(p_{r+1}q_r - q_{r+1}p_r) = 1,
\end{aligned}
$$

proving (5.33) in the simplest case $t = r$.

We also have $A_{r,r+1}(B) = a_{r+1}$, and

$$\begin{aligned} & q_{r+2}(-1)^r(q_r\alpha - p_r) + (-1)^{(r+1)-r} q_r(-1)^{r+2}(q_{r+2}\alpha - p_{r+2}) \\ & \quad = (-1)^r(p_{r+2}q_r - q_{r+2}p_r) \\ & \quad = (-1)^r((a_{r+1}p_{r+1} + p_r)q_r - (a_{r+1}q_{r+1} + q_r)p_r) \\ & \quad = (-1)^r a_{r+1}(p_{r+1}q_r - q_{r+1}p_r) = a_{r+1}, \end{aligned}$$

proving (5.33) for $t = r + 1$.

Since $b_i = a_i$ implies $b_{i-1} = 0$, we have the recurrence relation

$$A_{r,t}(B) = a_t A_{r,t-1}(B) + A_{r,t-2}(B) \quad \text{for all } t > r + 1. \tag{5.34}$$

Now we are ready to prove (5.33) by induction on $(t - r)$. We have

$$A_{r,t-j}(B) = q_{t-j+1}|\theta_r| + (-1)^{t-j-r} q_r|\theta_{t-j+1}|$$

for both $j = 1, 2$, and returning to (5.34), we conclude

$$\begin{aligned} A_{r,t}(B) &= a_t(q_t|\theta_r| + (-1)^{t-1-r} q_r|\theta_t|) + q_{t-1}|\theta_r| + (-1)^{t-2-r} q_r|\theta_{t-1}| \\ &= |\theta_r|(a_t q_t + q_{t-1}) + (-1)^{t-r} q_r(-a_t|\theta_t| + |\theta_{t-1}|) \\ &= q_{t+1}|\theta_r| + (-1)^{t-r} q_r|\theta_{t+1}|, \end{aligned}$$

proving (5.33), and this completes the proof of Lemma 5.3. □

Now it is easy to compute the measure of the intersection (5.32). First assume that the distance $d = k_2 - k_1 - \ell_1 - 1$ is ≥ 1. We know from the proof of Lemma 5.2 that the number of permissible sequences $(b_1, b_2, \ldots, b_{k_1-1})$ satisfying (5.4)–(5.6) is q_{k_1} if $b_{k_1} = B_1 \neq a_{k_1}$ and q_{k_1-1} if $b_{k_1} = B_1 = a_{k_1}$. By Lemma 5.3 the number of permissible sequences

$$(b_{k_1+\ell_1+1} = B_2 \neq 0, b_{k_1+\ell_1+2}, \ldots, b_{k_2-1})$$

of length d is

$$q_{k_2}|\theta_{k_1+\ell_1+1}| + (-1)^{d+1} q_{k_1+\ell_1+1}|\theta_{k_2}| \quad \text{if } b_{k_2} = B_3 \neq a_{k_2}$$

and

$$q_{k_2-1}|\theta_{k_1+\ell_1+1}| + (-1)^{d} q_{k_1+\ell_1+1}|\theta_{k_2-1}| \quad \text{if } b_{k_2} = B_3 = a_{k_2}.$$

Finally, note that, just like in Lemma 5.2, the tail series

$$\sum_{i=k_2+\ell_2+2}^{\infty} b_i\theta_i$$

completely fills out an interval of length $|\theta_{k_2+\ell_2+1}|$.

Write

$$X = S(b_{k_1} = B_1, b_{k_1+\ell_1+1} = B_2) \tag{5.35a}$$

and

$$Y = S(b_{k_2} = B_3, b_{k_2+\ell_2+1} = B_4). \tag{5.35b}$$

Lemma 5.4. *We have*

$$\frac{|\operatorname{Meas}(X \cap Y) - \operatorname{Meas}(X)\operatorname{Meas}(Y)|}{\operatorname{Meas}(X)\operatorname{Meas}(Y)} \le 2^{2-d},$$

where $d = k_2 - (k_1 + \ell_1 + 1) \ge 1$ is the "distance".

Proof. We distinguish four cases. We begin with

Case 1: Assume that $d = k_2 - k_1 - \ell_1 - 1$ is ≥ 1, $B_1 \ne a_{k_1}$, $B_3 \ne a_{k_2}$

Then we have

$$\operatorname{Meas}(X \cap Y) = q_{k_1}\left(q_{k_2}|\theta_{k_1+\ell_1+1}| + (-1)^{d+1} q_{k_1+\ell_1+1}|\theta_{k_2}|\right)|\theta_{k_2+\ell_2+1}|.$$

On the other hand, by Lemma 5.2,

$$\operatorname{Meas}(X) = q_{k_1}|\theta_{k_1+\ell_1+1}| \quad \text{and} \quad \operatorname{Meas}(Y) = q_{k_2}|\theta_{k_2+\ell_2+1}|.$$

It follows that

$$\frac{|\operatorname{Meas}(X \cap Y) - \operatorname{Meas}(X)\operatorname{Meas}(Y)|}{\operatorname{Meas}(X)\operatorname{Meas}(Y)} = \frac{q_{k_1+\ell_1+1}|\theta_{k_2}|}{q_{k_2}|\theta_{k_1+\ell_1+1}|}. \tag{5.36}$$

We need the almost trivial inequality

$$\frac{q_{i+d}}{q_i} \ge 2^{\lfloor d/2 \rfloor}, \tag{5.37a}$$

which follows from the successive application of the recurrence

$$q_i = a_{i-1}q_{i-1} + q_{i-2} \ge q_{i-1} + q_{i-2} \ge 2q_{i-2};$$

and we also need the following analog of (5.37a):

$$\frac{|\theta_i|}{|\theta_{i+d}|} \ge 2^{\lfloor d/2 \rfloor}. \tag{5.37b}$$

By (5.36)–(5.37), we have

$$\frac{|\operatorname{Meas}(X \cap Y) - \operatorname{Meas}(X)\operatorname{Meas}(Y)|}{\operatorname{Meas}(X)\operatorname{Meas}(Y)} \le 2^{1-d}, \tag{5.38}$$

where $d = k_2 - (k_1 + \ell_1 + 1) \ge 1$ is the "distance".

Inequality (5.38) justifies the term *exponentially weak dependence*, which is the reason behind the Area Principle (a "zero-one law").

Case 2: Assume that $d = k_2 - (k_1 + \ell_1 + 1) \ge 1$, $B_1 = a_{k_1}$, $B_3 = a_{k_2}$

Then (see (5.35))

$$\operatorname{Meas}(X \cap Y) = q_{k_1-1}\left(q_{k_2-1}|\theta_{k_1+\ell_1+1}| + (-1)^d q_{k_1+\ell_1+1}|\theta_{k_2-1}|\right)|\theta_{k_2+\ell_2+1}|,$$

and by Lemma 5.2,

$$\operatorname{Meas}(X) = q_{k_1-1}|\theta_{k_1+\ell_1+1}| \quad \text{and} \quad \operatorname{Meas}(Y) = q_{k_2-1}|\theta_{k_2+\ell_2+1}|.$$

Combining these facts with (5.37), we obtain

$$\frac{|\operatorname{Meas}(X \cap Y) - \operatorname{Meas}(X)\operatorname{Meas}(Y)|}{\operatorname{Meas}(X)\operatorname{Meas}(Y)} = \frac{q_{k_1+\ell_1+1}|\theta_{k_2-1}|}{q_{k_2-1}|\theta_{k_1+\ell_1+1}|} \le 2^{2-d}, \tag{5.39}$$

which is basically the same as (5.38) (we lost an irrelevant factor of 2).

It is easy to check that (5.39) remains true for the remaing two cases with $d \ge 1$: Case 3: $B_1 \ne a_{k_1}$, $B_3 = a_{k_2}$, and Case 4: $B_1 = a_{k_1}$, $B_3 \ne a_{k_2}$. In all four cases we have *exponentially weak dependence.* This completes the proof of Lemma 5.4. □

Now we are ready to complete the proof of Theorem 3: we simply use the exponentially weak dependence in a Chebyshev's inequality as follows. (The most difficult part is to find a good notation.) Let χ_{k,ℓ,B_1,B_2} denote the characteristic function of the set $S(b_k = B_1, b_{k+\ell+1} = B_2)$ defined by (5.21)–(5.23):

$$\chi_{k,\ell,B_1,B_2}(\beta) = \begin{cases} 1, & \text{if } \beta \in S(b_k = B_1, b_{k+\ell+1} = B_2); \\ 0, & \text{if } \beta \notin S(b_k = B_1, b_{k+\ell+1} = B_2). \end{cases}$$

We have a probabilistic viewpoint: the interval $-\alpha \le \beta < 1 - \alpha$ of length one is considered the whole probability space, and the usual "length" (one-dimensional Lebesgue measure), denoted by Meas(...), is the probability. So the expectation

$$\mathbf{E}\chi_{k,\ell,B_1,B_2} = \text{Meas}\left(S(b_k = B_1, b_{k+\ell+1} = B_2)\right),$$

and the sum (see (5.20))

$$\sum_{m_0 \le k \le M-2} \chi_{k,\ell,B_1,B_2}(\beta) \tag{5.40}$$

counts the number of integral solutions of the diophantine inequality

$$\|n\alpha - \beta\| = O(\psi(n)) \tag{5.41}$$

(the implicit constant in (5.41) is absolute) in the range $1 \le n \le q_M$, since by (5.25)

$$B_1 q_k \le n \le (B_1 + 2)q_k \le q_M.$$

Here M is a parameter; we choose $M \to \infty$ at the end of the proof.

To apply Chebyshev's inequality, we need to compute the variance

$$\begin{aligned}\mathbf{E}\Big(\sum_{m_0 \le k \le M-2} (\chi_{k,\ell,B_1,B_2} - \mathbf{E}\chi_{k,\ell,B_1,B_2})\Big)^2 &= \sum_{m_0 \le k \le M-2} (\chi_{k,\ell,B_1,B_2} - \mathbf{E}\chi_{k,\ell,B_1,B_2})^2 \\ &+ 2 \sum_{\substack{m_0 \le k_1 < k_2 \le M-2: \\ (k_1,\ell_1,B_1,B_2) \ne (k_2,\ell_2,B_3,B_4)}} \mathbf{E}(\chi_{k_1,\ell_1,B_1,B_2} - \mathbf{E}_1)(\chi_{k_2,\ell_2,B_3,B_4} - \mathbf{E}_2),\end{aligned} \tag{5.42}$$

where, for notational convenience, we use the brief notation

$$\mathbf{E}_1 = \mathbf{E}\chi_{k_1,\ell_1,B_1,B_2} \quad \text{and} \quad \mathbf{E}_2 = \mathbf{E}\chi_{k_2,\ell_2,B_3,B_4}.$$

Write

$$A_1 = S(b_{k_1} = B_1, b_{k_1+\ell_1+1} = B_2) \quad \text{and} \quad A_2 = S(b_{k_2} = B_3, b_{k_2+\ell_2+1} = B_4). \tag{5.43}$$

Note that with $k_1 \le k_2$ we have

$$\mathbf{E}\chi_{A_1}\chi_{A_2} = \text{Meas}(A_1 \cap A_2) = \begin{cases} 0, & \text{if } k_2 \le k_1 + \ell_1; \\ \ne 0, & \text{if } k_2 > k_1 + \ell_1 + 1 \\ & \quad \text{or } k_2 = k_1 + \ell_1 + 1, B_2 = B_3. \end{cases}$$

By (5.38)–(5.39)

$$|\mathbf{E}\chi_{A_1}\chi_{A_2} - \Pr(A_1)\Pr(A_2)| \le 2^{2-d}\Pr(A_1)\Pr(A_2), \tag{5.44}$$

where $d = k_2 - (k_1 + \ell_1 + 1) \ge 1$.

Using these facts in (5.42), we have

$$\text{Variance in (5.42)} \le \sum_{m_0 \le k_1 \le M-2} \Pr(A_1) + \sum\nolimits_1 + \sum\nolimits_2, \tag{5.45}$$

where

$$\sum\nolimits_1 = \sum_{A_1:\ m_0 \le k_1 \le M-2} \ \sum_{\substack{A_2:\ k_1+\ell_1+1=k_2 \le M-2 \\ B_2=B_3}} \Pr(A_1 \cap A_2) \tag{5.46}$$

and (5.44)

$$\sum\nolimits_2 = \sum_{A_1:\ m_0 \le k_1 \le M-2} \Pr(A_1) \left(\sum_{d \ge 1} \ \sum_{A_2:\ k_1+\ell_1+1=k_2 \le M-2} \Pr(A_2) \cdot 2^{2-d} \right). \tag{5.47}$$

Since the sets A_2 with fixed k_2 are pairwise disjoint, we have

$$\sum\nolimits_1 \le \sum_{m_0 \le k_1 \le M-2} \Pr(A_1), \tag{5.48}$$

and similarly,

$$\sum\nolimits_2 \le \sum_{m_0 \le k_1 \le M-2} \Pr(A_1) \left(\sum_{d \ge 1} 2^{2-d} \right) = 4 \sum_{m_0 \le k_1 \le M-2} \Pr(A_1). \tag{5.49}$$

Combining (5.45)–(5.49) we obtain

$$\text{Variance in (5.42)} \le 6 \sum_{m_0 \le k_1 \le M-2} \Pr(A_1). \tag{5.50}$$

By Chebyshev's inequality and (5.50), for any λ

$$\Pr\left[\sum_{m_0 \le k_1 \le M-2} \chi_{A_1} \ge \sum_{m_0 \le k_1 \le M-2} \Pr(A_1) - \lambda \right]$$

$$\geq 1-\lambda^{-2}\left(6\sum_{m_0\leq k_1\leq M-2}\Pr(A_1)\right). \tag{5.51}$$

Write

$$T=T(M)=\sum_{m_0\leq k_1\leq M-2}\Pr(A_1),$$

then by (5.31) and (5.43),

$$T=T(M)\to\infty \quad \text{as } M\to\infty. \tag{5.52}$$

We choose

$$\lambda=\lambda(M)=\frac{1}{2}T(M),$$

then by (5.51),

$$\Pr\left[\sum_{m_0\leq k_1\leq M-2}\chi_{A_1}\geq\frac{1}{2}T(M)\right]\geq 1-\frac{24}{T(M)}. \tag{5.53}$$

Taking $M\to\infty$, by (5.52)–(5.53) we obtain

$$\sum_{k\geq m_0}\chi_{k,\ell,B_1,B_2}(\beta)=\sum_{k\geq m_0}\chi_{A_1}=\infty$$

for *almost all* $\beta\in[-\alpha,1-\alpha)$, and by (5.40)–(5.41) this gives infinitely many integral solutions of the diophantine inequality

$$\|n\alpha-\beta\|=O(\psi(n)). \tag{5.54}$$

Since the implicit constant in (5.54) is absolute, the proof of Theorem 3 is complete. □

References

[Be010] Beck, J.: *Randomness of the Square-Root of Two*, manuscript of a book, 500 pages.

[Be98a] Beck, J.: Diophantine approximation and quadratic fields, *Number Theory*, Eds.: Győry/Pethő/Sós, Walter de Gruyter GmbH, Berlin – New York 1998, pp. 55–93. MR1628833

[Be98b] Beck, J.: From probabilistic diophantine approximation to quadratic fields, *Random and Quasi-Random Point Sets*, Lecture Notes in Statistics 138, Springer-Verlag New York 1998, pp. 1–49. MR1662839

[Ca51] Cassels, J.W.: On the law of the iterated logarithm, Proc. Cambridge 47 (1951), pp. 51–64. MR0040614

[Ca54] Cassels, J.W.: Über lim $x|\theta x + \alpha - y|$, Math. Ann. 127 (1954), pp. 288–304. MR0060546

[De56] Descombes, I.R.: Sur la répartition des sommets d'une ligne polygonale réguliere nonfermée, Ann. Sci. de l'École Normale Sup. 75 (1956) 284–355.

[Er42] Erdős, P.: On the law of the iterated logarithm, Annals of Math. (2) 43 (1942), pp. 419–435. MR0006630

[Fe43] Feller, W.: The general form of the so-called law of the iterated logarithm, Trans. Amer. Math. Soc. 54 (1943), pp. 373–402. MR0009263

[Kh24] Khintchine, A.: Über einen Satz der Wahrscheinlichkeitsrechnung, Fundamenta Math. 6 (1924), pp. 9–20.

[La66] Lang, S.: *Introduction to Diophantine Approximations*, Addison-Wesley 1966. MR0209227

[Os22] Ostrowski, A.: Bemerkungen zur Theorie der Diophantischen Approximationen. I. Abh. Hamburg Sem. 1 (1922), 77–99.

[S58] Sós, V.T.: On the theory of diophantine approximation II., Acta Math. Hungar., 9 (1–2) (1958), pp. 229–241. MR0095164

[S74] Sós, V.T.: On the discrepancy of the sequence $\{n\alpha\}$, Coll. Math. Soc. János Bolyai 13 (1974), pp. 359–367. MR0460265

[S83] Sós, V.T.: On strong irregularities of the distribution of $\{n\alpha\}$ sequences, Studies in Pure Math. (1983), pp. 685–700. MR0820262

József Beck
Mathematics Department, Busch Campus, Hill Center
Rutgers University, New Brunswick, NJ 08903
USA
E-mail address: jbeck@math.rutgers.edu

Received September 28, 2009