2016/17 MMTCEBES

Temas de Matemática
temas

Salvatore Cosentino
Departamento de Matemática e Aplicações - Universidade do Minho
Campus de Gualtar - 4710 Braga - PORTUGAL
gab B.4023, tel 253 604086
e-mail scosentino@math.uminho.pt
url http://w3.math.uminho.pt/~scosentino

March 3, 2017

Contents

1 More and less elementary topics in number theory

The basic reference is the classic book by Hardy & Wright [HW59]. Topics on Diophantine approximation may be found in Cassel's tract [Cas57] and Khinchin's monograph [Kh35]. A modern reference for the metric theory of Diophantine approximation is Kleinbock's notes [Kl10], and a modern reference for ergodic theory in the theory of numbers is [EW10].

1.1 Irrationals

Irrationals roots. We owe to the Greeks the "discovery" that some numbers, which solve natural geometric or arithmetic problems, are not fractions. They are *irrationals*.

The archetypical observation is the famous

Theorem 1.1 (Pythagoras)**.** *$\sqrt{2}$ is irrational.*

The proof traditionally ascribed to Pythagoras uses divisibility by 2, hence the dichotomy between even and odd numbers. More general arguments, using Gauss' *fundamental theorem of arithmetic* (or Euclid's lemma) can be used to prove that any root

$$\sqrt[n]{N}$$

is irrational, provided not trivially rational, i.e. as long as N is not the n-th power of an integer. The n-th root of N is a root of the monic polynomial $x^n - N$. Still more general is

Theorem 1.2 (Gauss, 1801)**.** *The roots of a monic polynomial*

$$x^n + a_1 x^{n-1} + \cdots + a_n$$

with integer coefficients $a_1, \ldots, a_n$ are either integer or irrational.

Other famous irrationals. Analysis (and geometry) provides further examples of irrational numbers, but the proofs are not always so easy. For example, is quite clear that numbers like

$$\log_2 3$$

are irrationals, but it took much effort to prove

Theorem 1.3 (Euler, 1737)**.** *e is irrational.*

Theorem 1.4 (Lambert, 1761)**.** *π is irrational.*

The original proofs, by Euler and Lambert, used continued fractions, and it would be interesting to look at them. Simpler analytic proofs were then discovered. A simple proof of the irrationality of e is due to Fourier (1815), and uses the power series defining the exponential. The proof quoted in [HW59] of the irrationality of π was discovered by Niven [Ni47]. Hystorical comments may be found in chapter 4 of [HW59].

Much more difficult is the following

Theorem 1.5 (Gelfond, 1929)**.** *e^π is irrational.*

Open problems. Here one could list some numbers that everybody think are transcendental, but such that nobody has been able to prove to be irrational.

1.2 Decimal representation

Decimal representation. When children we learn to represent numbers as decimals, like

$$3.14159265358979323846264338327950288419716939937510\ldots$$

Of course, there is nothing special with the number 10, it is but the number of fingers in our hands. Any other integer $d \geq 2$ would work. Representing a non-negative (for simplicity) real number $x \in \mathbb{R}_+$ in base 10 means writing x as the sum of a convergent series

$$\begin{aligned} x &= \text{“}X_m \ldots X_1 X_0 . x_1 x_2 x_3 \ldots\text{”} \\ &:= X_m \cdot 10^m + \cdots + X_1 \cdot 10 + X_0 + \frac{x_1}{10} + \frac{x_2}{10^2} + \frac{x_3}{10^3} + \ldots \\ &= \sum_{n=0}^{m} X_n \cdot 10^n + \sum_{n=1}^{\infty} x_n \cdot 10^{-n} \end{aligned}$$

where X_n and x_n are letters of the “alphabet” $\mathcal{A} := \{0, 1, 2, \ldots, 9\}$ and $m \geq 0$. Some representations terminate, i.e. have $x_n = 0$ as long as n is larger than some N, and some others are recurring, i.e. of the form

$$A.a\overline{b} := A.abbb\ldots$$

where A, a and b are given finite words in the alphabet $\mathcal{A}$ (and of course a terminating decimal is a recurring one with recurring word $\overline{0}$, indeed, the decimal representation of a reduced rational p/q terminates iff the denominator is of the form $q = 2^\alpha 5^\beta$ for some non-negative integers α and β, and this of course is an accident due to the special chosen scale 10).

The representation is unique, if we do not admit recurrent 9's, i.e. if we substitute $\ldots x_{k-1}\overline{9}$ with $\ldots (x_{k-1}+1)\overline{0}$ (where we assume $x_{k-1} \neq 9$).

The finite sum

$$[x] := X_m \ldots X_1 X_0 = \sum_{n=0}^{m} X_n \cdot 10^n \in \mathbb{Z}$$

is the *integral part* of x, the largest of those integers n such that $n \leq x$. The possibly infinite sum

$$\{x\} := 0.x_1 x_2 x_3 \cdots = \sum_{n=1}^{\infty} x_n \cdot 10^{-n} \in [0, 1)$$

is the *fractional part* of x, the difference $\{x\} = x - [x]$. Consequently, $[x] + \{x\} = x$.

Division algorithm. The iterative scheme to obtain the decimal representation of a rational number is the “division algorithm” . Consider a positive rational $x = p/q$, with $p, q \in \mathbb{N}$ and $(p, q) = 1$, with decimal representation

$$\frac{p}{q} = X_m \ldots X_1 X_0 . x_1 x_2 x_3 \ldots .$$

The integer $X := X_m \ldots X_1 X_0$ is “the number of times q is contained in p”, i.e. the unique integer such that

$$p = X \cdot q + r_0$$

for some rest r_0 which is an integer $0 \leq r_0 < q$. Hence, $p/q = X + r_0/q$ and $0 \leq r_0/q < 1$. Thus, the point r_0/q lies between $0.x_1$ and $0.x_1 + 0.1$. Multiplying by 10 and then by q this means that

$$x_1 \cdot q \leq 10 \cdot r_0 < x_1 \cdot q + q$$

or, equivalently, that x_1 is the unique integer between 0 and 9 such that

$$10 \cdot r_0 = x_1 \cdot q + r_1$$

where, again, the rest r_1 is a non-negative integer $0 \leq r_1 < q$. And so on. Hence, the digits of the decimal expansion of p/q are iteratively determined by

$$10 \cdot r_{n-1} = x_n \cdot q + r_n \qquad \text{where} \qquad 0 \leq r_n < q$$

Since the possibilities for the rests are finite, they necessarily recurr. On the other side, a simple computation using the sum of a geometric series shows that a recurring decimal is a (series converging to a) rational number.

Theorem 1.6. *The rational numbers are precisely those real numbers whose representation in base* 10 *(or any other base $d \geq 2$) is (eventually) repeating/recurring.*

Irrationals defined by decimals. Consequently, numbers represented by non-recurring decimals are irrational. This is the case, for example, of

$$0.101001000100001\cdots = \frac{1}{10} + \frac{1}{10^3} + \frac{1}{10^6} + \frac{1}{10^{10}} + \frac{1}{10^{15}} + \dots$$

Other nice example may be found in [HW59], section 9.4. Unfortunatly, it is not obvious to compute the sum of such a series!

The weight of the rationals. Actually, almost all numbers are irrational, in a precise probabilistic sense, since rationals are countable.

Consider the unit interval $I = [0, 1]$, and cut out all its rational points. What is left is a set, $I \backslash \mathbb{Q}$, whose lenght is still equal to the lenght of the unit interval. Indeed, the rationals are countable, for example those inside I may be ordered according to

$$0 \quad 1 \quad 1/2 \quad 1/3 \quad 2/3 \quad 1/4 \quad 3/4 \quad 1/5 \quad 2/5 \quad 3/5 \quad 4/5 \quad \dots$$

say $I \cap \mathbb{Q} = \{r_1, r_2, r_3, \dots\}$. Given any $\varepsilon > 0$, we may cut out a whole interval $J_n = (r_n - \ell_n/2, r_n + \ell_n/2)$ of finite lenght $\ell_n = \varepsilon/2^n$ around each r_n. The measure of what is left of the unit interval is

$$\text{lenght}\left(I \setminus (\cup_n J_n)\right) \geq 1 - \sum_n \varepsilon/2^n = 1 - \varepsilon\,.$$

Since ε may be arbitrarily small (while positive), the rationals inside the unit interval have neighborhoods of arbitrarily small lenght! Mathematicians say that

Theorem 1.7. *Rationals form a set of Lebesgue measure zero inside the real line.*

1.3 Decimals with missing digits

La biblioteca de Babel. According to Jorge Luis Borges (*La biblioteca de babel* (in *El jardím de senderos que se bifurcan*), 1941) there exists a huge library with exagonal rooms containing all possible books written combining 25 characters (letters and punctuation). The library must therefore contain, together with lot of nonsense, all possible meaningful books, and variations or translations of them. People spend their time searching between the books of the library, looking for prediction of the future, biographies of people, and all classes of useful informations.

> "... Una secta blasfema sugirió que cesaran las buscas y que todos los hombres barajaran letras y símbolos, hasta construir, mediante un improbable don del azar, esos libros canónicos. Las autoridades se vieron obligadas a promulgar órdenes severas. La secta desapareció, pero en mi niñez he visto hombres viejos que largamente se ocultaban en las letrinas, con unos discos de metal en un cubilete prohibido, y débilmente remedaban el divino desorden.... "

Expanding endomorphisms of the circle and Bernoulli shift. Here we restrict to the unit interval $[0,1)$, identified with the circle $\mathbb{T} := \mathbb{R}/\mathbb{Z}$, and consider the transformation $f : \mathbb{T} \to \mathbb{T}$ given by multipication by 10, namely $f(x+\mathbb{Z}) := 10 \cdot x + \mathbb{Z}$. If we write (in decimal notation) a generic point of the circle as $x = 0.x_1x_2x_3\cdots \in [0,1) \approx \mathbb{T}$, we see that

$$f(0.x_1x_2x_3\dots) = 0.x_2x_3x_4\dots$$

The Lebesgue measure ℓ on the circle (the σ-additive extension of the lenght of segments) is clearly invariant under f, in the sense that $\ell(f^{-1}(A)) = \ell(A)$ for all measurable $A \subset \mathbb{T}$.

More abstractly, we may consider the space $\Sigma^+ := \mathcal{A}^{\mathbb{N}}$ of infinite words $x_1x_2x_2x_3\dots$ in the alphabet $\mathcal{A} = \{0,1,2,\dots,9\}$, and the so called *Bernoulli shift* $\sigma : \Sigma^+ \to \Sigma^+$, which consists in forgetting the first letter:

$$\sigma(x_1x_2x_2x_3\dots) := x_2x_2x_3x_4\dots .$$

Σ^+ is the space representing the sucessive results of throwing a "dice" with 10 faces. The decimal representation is a surjective map $F : \Sigma^+ \to \mathbb{T}$ given by $x_1x_2x_3\dots \mapsto 0.x_1x_2x_3$, and the pull-back of Lebesgue measure coincides with the Bernoulli measure μ saying that the repeted experiment are independent and the appearence of each face of the dice is equiprobable. Multiplication by 10 corresponds to the shift, in the precise mathematical sense that

$$f \circ F = F \circ \sigma$$

(i.e. f is a "factor" of F). Since the set where F fails to be injective has zero measure, the ergodic properties of f coincide with those of the shift σ. Thus, with respect to the Lebesgue probability measure, the appearence of the digits in the decimal representation of a random number between 0 and 1 is the same thing that the successive throwing of a honest dice with 10 faces.

Decimals with missing letters. Borges was inspired by the following observation, popularized by Émile Borel as the "paradoxe des singes savants" [Bo13]:

> "Concevons qu'on ait dressé un million de singes à frapper au hasard sur les touches d'une machine à écrire et que [...] ces singes dactylographes travaillent avec ardeur dix heures par jour avec un million de machines à écrire de types variés. [...] Au bout d'un an, [leurs] volumes se trouveraient renfermer la copie exacte des livres de toute nature et de toutes langues conservs dans les plus riches bibliothèques du monde."

The proportion/probability of those numbers in $[0,1)$ which omit one given letter, say 7, in the first place of their decimal representation is $9/10$. Those which omit 7 in the first two places have probability $81/100$ ... Those which omit 7 in the first N places have probability $(9/10)^N$. Since these probabilities go to zero as $N \to \infty$, there follows that numbers with missing 7 form a set of zero probability within the unit interval. But the same reasoning can be done for numbers which omit any given finite word $b = b_1b_2\dots b_n$ in the alphabet $\mathcal{A}$. In other words, as strange as it may seems to not mathematically-educated minds, it happens that

Theorem 1.8. *Almost all decimals contain all finite words in any number of digits.*

The apparent paradox is "solved" once we try to compute the amount of time (i.e. digits) needed to really see a given "meaningful" sentence within a random infnite book in our alphabet. For example, the letters in our western alphabets are more or less 25, counting punctuation and blank space. *L'infinito* by Giacomo Leopardi,

> "Sempre caro mi fu quest'ermo colle,
> e questa siepe, che da tanta parte
> dell'ultimo orizzonte il guardo esclude.
> Ma sedendo e mirando, interminati
> spazi di là da quella, e sovrumani
> silenzi, e profondissima quïete
> io nel pensier mi fingo, ove per poco
> il cor non si spaura. E come il vento
> odo stormir tra queste piante, io quello

infinito silenzio a questa voce
vo comparando: e mi sovvien l'eterno,
e le morte stagioni, e la presente
e viva, e il suon di lei. Così tra questa
immensità s'annega il pensier mio:
e il naufragar m'è dolce in questo mare."

is made of more than 550 characters. The probability to produce it typing randomly on a computer a text of this lenght is of the order of 10^{-770}. Equivalently, if you produce random texts of lenght 550 characters, you must wait around

$$10^{770}$$

trials to see Leopardi's poem for the first time. You may want to take a look at physically meaningful numbers (the total number of fundamental particles in the universe is of the order of 10^{80}, the age of the universe is 4.32×10^{17} seconds, the Planck time, the shortest conceivable interval of time, is 5.4×10^{-44} seconds) to make sense (or not!) of the above huge number. Indeed, if every baryon in the universe were a monkey with a typewriter, typing one characterer each Planck time, ...

Normal numbers Much more is true. Lebesgue measure ℓ is ergodic w.r.t. multiplication by 10 in the unit circle (or, equivalently, the Bernoulli measure μ is ergodic w.r.t. the shift σ). Physically, this means that time averages of reasonable observables converge (almost everywhere) to their respective mean values, as time goes to infinite.

For $a = 0, 1, 2, ..., 9$, let φ_a be the characteristic function of the interval $[a/10, (a+1)/10[$, i.e. the observable which is equal to $\varphi_a(x) = 1$ if $x_1 = a$ and $\varphi_a(x) = 0$ otherwise. The time mean of φ_a is

$$\frac{1}{N}\sum_{n=0}^{N-1} \varphi_a\left(f^n(0.x_1x_2x_3\dots)\right) = \frac{1}{N}\cdot \text{card}\left\{1 \le n \le N \text{ s.t. } x_n = a\right\}$$

that is the number of a's within the first N digits of the decimal expansion of x. The limit as $N \to \infty$, if it exists, is the "asymptotic frequency" of a's contained in the expansion of x. Ergodicity of ℓ implies that there exists a set $A_a \subset [0,1[$ of Lebesgue measure one where the limit $\overline{\varphi_a}(x)$ exists and is equal to $\int \varphi_a d\ell = 1/10$. Since the intersection $A_0 \cap A_1 \cap ... \cap A_9$ has still probability one, the result is that Lebesgue almost any number $x \in [0,1[$ contains in its decimal expansion any of the letters $0, 1, 2, ..., 9$ with asymptotic frequency $1/10$.

Actually, one could repeat the same argument considering any finite word $b = b_1b_2...b_n$ in the alphabeth $\mathcal{A}$, and show that there is a set $A_b \subset [0,1[$ of probability one such that the base 10 expansion of any $x \in A_b$ contains the word b with asymptotic frequency 10^{-n}. A real number x whose base 10 expansion contains any finite word with the right asymptotic frequency is called 10-*normal* (meaning "normal in base 10"). Since finite words in the alphabeth $\mathcal{A}$ are countable, and a countable union of zero measure sets still has zero measure, we just showed that *Lebesgue almost any real number x is* 10-*normal.* Indeed, Émile Borel [Bo09] showed that

Theorem 1.9 (Borel, 1909)**.** *Lebesgue almost any real number is normal in every base $m \ge 2$.*

It is not so easy to give examples of normal numbers, actually of series whose sum is a normal number. Much more difficult (impossible?) seems to show that a "given" number, such as

$$\pi\,, \quad \sqrt{2}\,, \quad e\,, \quad \dots$$

is normal. Here we quote Mark Kac ([Ka59] pag. 18):

> "As is often the case, it is much easier to prove that an overhelming majority of objects possess a certain property that to *exhibit* even one such object. The present case is no exception. It is quite difficult to exhibit a 'normal' number! The simplest example

is the number (written in decimal notation) $x = 0.1234567891011\ldots$ where after the decimal point we write the positive integers in succession. The proof that this number is normal is by no means trivial."

1.4 Rational approximations of irrationals

Engineers' problem. It is plain that we (or machines) can only operate with rationals, and this explains why ancients developed a variety of clever methods to find good (i.e. useful for the engineers!) rational approximations of important numbers. For example, Babylonians and Greeks knew and used the approximations $\pi \simeq 25/8$ and $\pi \simeq 377/120$, respectively. More interesting is the fact that Heron of Alexandria (but probably the Babylonians) used iteration of the recursive equation $x_{n+1} = (x_n + 2/x_n)/2$ to "compute" $\sqrt{2}$ (or any other square root). The method produces super-exponentially convergent rational approximations.

Rational approximations of irrationals. In modern language, we say that rationals $\mathbb{Q}$ are dense in the real line $\mathbb{R}$ (which is, indeed, the completion of rationals w.r.t. the Euclidean metric, so that real numbers are equivalence classes of Cauchy sequences of rationals). Hence, for any $x \in \mathbb{R}$ and any "precision" $\varepsilon > 0$ we may find "rational approximations" of x, i.e. reduced fractions $p/q \in \mathbb{Q}$ solving the inequality

$$|x - p/q| < \varepsilon\,.$$

Natural questions, when x is irrational, are: how large must be the denominator q given ε ? How do the denominators grow when we take smaller and smaller ε ?

For example, we say that a number x is *approximable to order* ω if there exists a constant $\lambda = \lambda(x)$ such that the inequality

$$\left| x - \frac{p}{q} \right| < \frac{\lambda}{q^\omega}\,.$$

occurs infinitely often, i.e. admits infinite rational solutions p/q. Or, more generally, we may bound the l.h.s. above by $\phi(q)/q$ for some non-increasing function $\phi : \mathbb{N} \to \mathbb{R}_+$. We may then ask which numbers are approximable to a given order, or at least measure their relative size, i.e. their probability. Indeed, the problem is clearly invariant under integer translations, and therefore we may restrict to numbers x in the quotient unit circle $\mathbb{T} := \mathbb{R}/\mathbb{Z} \approx [0, 1)$.

Why do we care. Mathematicians are naturally fascinated by the above questions, and then by the problems that the search for their solution suggest (some of them you will see below). On the other hand, it must be said that such questions also play an important role in other areas of pure and/or applied mathematics, alias physics. A practical example referred by Khinchin ([Kh35], page 28) is Huygens' problem to construct a model of the solar system using toothed wheels. Modern more abstract examples are the problems of "small denominators", in a variety of context of pure and applied mathematics, from the classical linearization problem of Poincaré and Siegel (and then Brjuno, and then Yoccoz ...) to *KAM theory* (for Kolmogorov, Arnold and Moser) [Mar00]. Here, you may want to give some naïve examples ...

Diophantine approximation. In these contexts, the relevant notion is the converse of approximability: we say that a number x satisfies a *Diophantine condition* of order τ if there exists a constant $c = c(x)$ such that

$$\left| x - \frac{p}{q} \right| > \frac{c}{q^\tau}\,.$$

for all reduced fractons p/q.

The relation with "Diophantine equations", which originates the name *Diphantine approximation*, is explained, for example, at the end of chapter 17 of [IR90]: if the roots of a polynomial $a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$ with integer coefficients, irreducible over the rational and degree $n \geq 3$, satisfy a Diophantine condition with exponent/order $n - \varepsilon > 2$, then the Diophantine equation

$$a_n x^n + a_{n-1} x^{n-1} y + \cdots + a_0 y^n = m$$

with integer $m > 0$ has at most a finite number of (integer) solutions (x, y).

Dirichlet's box principle. The first non-trivial observation is due to Dirichlet, and it is the original application of his *box (pigeone hole) principle*:

> "if there are $n+1$ objects in n boxes, there must be at least one box which contains two (or more) of the objects".

For any real number x and any positive integer Q, there exists a reduced fraction p/q with $1 \le q < Q$ such that $|x - p/q| \le 1/qQ$. In particular,

Theorem 1.10 (Dirichlet, 1842)**.** *For any irrational number x there exist infinitely many reduced fractions p/q such that*

$$\left| x - \frac{p}{q} \right| < \frac{1}{q^2} \,.$$

Thus, all irrationals are approximable to order 2, at least. The same box principle can be used to understand simultaneous rational approximations.

Theorem 1.11. *If at least one of the numbers $x_1, x_2, \dots, x_n$ is irrational, then the inequalities*

$$\left| x_k - \frac{p_k}{q} \right| < \frac{1}{q^{1+1/n}} \qquad k = 1, 2, \dots, n \,,$$

have an infinity of solutions.

1.5 Geometry of numbers

Minkowski's convex body theorem. Also important is Minkowski's proof of Dirichlet theorem 1.10, which follows from his *convex body theorem* [Mi10, Mi27]. The most popular proof is (I think) the one presented in appendix B of [Cas57] or in chapter III of [Cas59], which is based on the following theorem [Bl14].

Theorem 1.12 (Blichfeldt, 1914)**.** *A (measurable) region $A \subset \mathbb{R}^n$ with volume $\mathrm{Vol}(A) > 1$ contains two (distinct) points such that their difference has integer coordinates.*

The proof is simple but not trivial, being based on existence and σ-additivity of Lebesgue measure, and goes as follows: if we fold all of A onto the unit hyper-cube $\square := [0,1)^n \approx \mathbb{R}^n/\mathbb{Z}^n$, at least two of the images $A_{\mathbf{n}} := \mathbf{n} + A \cap (\square - \mathbf{n})$, for $\mathbf{n} \in \mathbb{Z}^n$, must overlap for otherwise the volume of A would be $\le \mathrm{Vol}(\square) = 1$.

A subset $C \subset \mathbb{R}^n$ is convex if, whenever it contains two points $\mathbf{x}$ and $\mathbf{y}$, it contain the entire segment $\overline{\mathbf{xy}} := \{t\mathbf{x} + (1-t)\mathbf{y} \text{ with } 0 \le t \le 1\}$ between them. A subset $S \subset \mathbb{R}^n$ is *centrally symmetric* (i.e. symmetric about the origin) if, whenever it contains a point $\mathbf{x}$ it also contain its symmetric $-\mathbf{x}$. A region which is both convex and centrally symmetric is called *convex body*. In particular, together with any two points $\mathbf{x}$ and $\mathbf{y}$, a convex body contains the whole parallelepiped with vertices $\pm\mathbf{x}$ and $\pm\mathbf{y}$. Minkowski's observation that a convex body is Lebesgue measurable, and an obvious application of Blichfeldt theorem 1.12 gives the famous *Minkowski's convex body theorem*

Theorem 1.13 (Minkowski, 1896)**.** *Let $K \subset \mathbb{R}^n$ be a convex body with volume $\mathrm{Vol}(K) > 2^n$ (possibly ∞). Then K contains a point of the integer lattice $\mathbb{Z}^n \subset \mathbb{R}^n$ other than zero.*

If the convex body K is also compact, then the weaker inequality $\mathrm{Vol}(K) \ge 2^n$ is sufficient.

There exist many other proofs of Minkowski convex body theorem 1.13, three or four of them are in chapter III of [HW59]. Generalizations to lattices $\Lambda \subset \mathbb{R}^n$ other than $\mathbb{Z}^n$ are straightforward.

Minkowski's linear form theorem. Given n linear forms in $\mathbb{R}^n$, say $\mathbf{x} \mapsto \boldsymbol{\xi}_k \cdot \mathbf{x} = \xi_{k1}x_1 + \xi_{k2}x_2 + \cdots + \xi_{kn}x_n$ with $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_n \in (\mathbb{R}^n)^* \approx \mathbb{R}^n$, and n positive numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$, we may consider the problem of solving simultaneously the set of inequalites

$$|\boldsymbol{\xi}_1 \cdot \mathbf{x}| < \lambda_1 \qquad |\boldsymbol{\xi}_2 \cdot \mathbf{x}| < \lambda_2 \qquad \ldots \qquad |\boldsymbol{\xi}_n \cdot \mathbf{x}| < \lambda_n$$

for non-trivial integer vectors $\mathbf{x} \in \mathbb{Z}^n$. If $\Xi = (\xi_{ij})$ denotes the $n \times n$ matrix whose lines are the vectors $\boldsymbol{\xi}_k$'s, then the region $K \subset \mathbb{R}^n$ defined by the above inequalities (for all $\mathbf{x} \in \mathbb{R}^n$, not only integers!) is a convex body and has volume $2^n(\lambda_1\lambda_2 \ldots \lambda_n)/|\det \Xi|$. Thus, if $(\lambda_1\lambda_2 \ldots \lambda_n) > |\det \Xi|$ then there is an integer point other than zero solving the inequalities. If some of the $<$ is substituted by $\leq$ in the definition of the convex body K, and if $>$ is substituted by $\geq$ in the volume estimate, we still get existence of an integer non-trivial solution from a continuity argument.

For example, Dirichlet's theorem 1.10 corresponds to find non-trivial integer solutions $(p, q) \in \mathbb{Z}^2$ of the inequalities

$$|xq - p| \leq 1/Q \qquad \text{and} \qquad |q| < Q\,,$$

and the corresponding planar convex body has surface equal to 4.

A proof by Siegel and Mordell uses the Poisson summation formula from Fourier analysis [Si45].

Geometry of numbers. Minkowski 's convex body theorem is the foundational result of a big area called "geometry of numbers" [Mi10, Cas59]. Actually, this could be a theme on its own, following chapter XXIV of [HW59].

A nice application of the convex body theorem is a proof of *Pick's theorem* [Pi99, St99] by Murty and Thain [MT07] (but there exist many proofs of this beautiful and elementary theorem). A *lattice polygon* is a planar polygon with vertices belonging to the lattice $\mathbb{Z}^2$, i.e. with integer coordinates.

Theorem 1.14 (Pick, 1899)**.** *Let P be a convex lattice polygon with b lattice points on its boundary and i lattice points in its interior. Then the area of P is*

$$\text{Area}(P) = i + b/2 - 1\,.$$

The convex body theorem is used to prove than every "elementary triangle" T, i.e. lattice triangle with no integer points apart from its vertices, has $\text{Area}(T) = 1/2$ (the inequality $\text{Area}(T) \geq 1/2$ being trivial). Picks theorem follows decomposing a generic lattice polygon into elementary triangles and using additivity of Pick's formula.

1.6 Transcendental numbers

Algebraic numbers. *Algebraic numbers* are the roots of polynomials

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$$

with integer coefficients $a_0, \ldots, a_n \in \mathbb{Z}$. The *degree* of the algebraic number x is the minimal degree of a polynomial $p \in \mathbb{Z}[x]$ such that $p(x) = 0$. So, algebraic numbers of degree 1 are the rationals, algebraic numbers of degree 2 are the quadratic irrationals (as for example $\sqrt{2}$), and so on.

Transcendental numbers. A simple argument, using the mean value theorem, gives the

Theorem 1.15 (Liouville, 1842)**.** *A real algebraic number of degree $d \geq 2$ is not approximable to any order $\geq d$, i.e satisfies*

$$\left| x - \frac{p}{q} \right| > \frac{c}{q^d}$$

for all fractions p/q and some constant $c = c(x)$.

Liouville theorem is not optimal, but has been important because for the first time it allowed the construction of numbers which are not algebraic, called *transcendental*: it is sufficient to produce numbers which are approximable to any order. Liouville's example is

$$0.11000100000000000000000010\cdots = \frac{1}{10^{1!}} + \frac{1}{10^{2!}} + \frac{1}{10^{3!}} + \frac{1}{10^{4!}} + \dots$$

We owe the optimal result concerning algebraic numbers to the efforts of Thue, Siegel, Dyson, Gelfond, Schneider, and finally Roth [Rot55]. This is really first class mathematics (but elementary: it does not use any algebraic number theory nor deep analysis!), indeed worth a Field Medal.

Theorem 1.16 (Roth, 1955)**.** *For any irrational real algebraic number x and any $\varepsilon > 0$ there exists a constant $c = c(x, \varepsilon)$ such that*

$$\left| x - \frac{p}{q} \right| > \frac{c}{q^{2+\varepsilon}} .$$

for all fractions p/q.

Thus, algebraic numbers are not approximable to any order greater that 2 (nevertheless, observe that for quadratic irrationals Liouville estimate is better!).

Other famous transcendental numbers. We know since Cantor that algebraic numbers are countable, because rational polynomials are. This means tha most real or complex numbers are not algebraic. Much more difficult is to prove that a given number, e.g. some famous constant like e or π, is transcendental, but this is another story. Here one could include a sketch of some proofs of the following

Theorem 1.17 (Hermite, 1873)**.** *e is transcendental.*

Theorem 1.18 (Lindemann, 1882)**.** *π is transcendental.*

(the last one being generalized as the *Lindemann-Weierstrass theorem*).

1.7 Continued fractions

Continued fractions. Continued fractions constitute the fundamental tool to investigate rational approximations to real number, because they provide base-free (hence intrinsic) rational approximations, and moreover because they provide the best rational approximations, in a certain precise sense [Cas57, HW59].

Any real number $x \in \mathbb{R}$ may be uniquely (not true for rationals, but the ambiguity is small!) represented as a *continued fraction*

$$x \sim [a_0; a_1, a_2, a_3, \dots] := a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ddots}}}}$$

with $a_0 \in \mathbb{Z}$ and "partial quotients" $a_n \in \mathbb{N}$ if $n \geq 1$. This means that x is equal to the limit of the *convergents*, the finite continued fractions (hence rationals)

$$p_n/q_n = [a_0; a_1, a_2, \dots, a_n] := a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \cfrac{1}{a_n}}}}$$

as $n \to \infty$.

Euclid algorithm and Gauss map. The continued fraction converging to the given number x is given essentially by Euclid algorithm to find the m.c.d. of two integers. One starts with $a_0 = \lfloor x \rfloor$ (here the "floor" function $\lfloor x \rfloor$ returns the integer n such that $n \le x < n+1$), and write $x = a_0 + x_0$ for some $0 \le x_0 < 1$. Then define the *Gauss map* $G : (0,1] \to [0,1]$ as

$$G(x) := 1/x - \lfloor 1/x \rfloor \,,$$

(thus, $G(x)$ is the fractional part of the inverse of x) and inductively define the partial quotients and the "rests" as

$$a_{n+1} = \lfloor 1/x_n \rfloor \qquad x_{n+1} = G(x_n) \,,$$

provided all the $x_k \neq 0$. Thus,

$$\begin{aligned} x &= a_0 + x_0 \\ &= a_0 + \cfrac{1}{a_1 + x_1} \\ &= a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + x_2}} \\ &\ddots \\ &= a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \cfrac{1}{a_n + x_n}}}} \end{aligned}$$

If some $x_n = 0$ the iteration stops and x is equal to a finite continued fraction. In particular, finite continued fractions correspond to rationals (and are unique if we demand the the last non-zero partial quotient be $a_n > 1$).

Best rational approximations to reals. The convergents $p_n/q_n = [a_0; a_1, a_2, \dots, a_n]$, with $(p_n, q_n) = 1$, are obtained from the coefficients a_k's by the recursions

$$\begin{aligned} p_n &= a_n p_{n-1} + p_{n-2} \\ q_n &= a_n q_{n-1} + q_{n-2} \end{aligned} \tag{1.1}$$

given the initial conditions $p_0 = a_0$, $q_0 = 1$, and $p_{-1} = 1$, $q_{-1} = 0$ (or $p_{-2} = 0$ and $q_{-2} = 1$) . There easily follows that even and odd convergents form an increasing and decreasing sequence, respectively, with common limit $x = \lim_{n\to\infty} p_n/q_n$. We may estimate the rate of convergence (if x is irrational) as

$$\frac{1}{q_n(q_{n+1} + q_n)} < \left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_{n+1} q_n} \tag{1.2}$$

Since the denominators of the convergents grow at least as $q_n \ge 2^{n-1}$, the convergence is at least exponential. In particular, since the q_n's grow, from the second inequality we recover Dirichlet theorem.

Of fundamental importance is the fact that convergents provide the best approximations in the sense that

$$|q_n x - p_n| \le |qx - p|$$

for all reduced fractions p/q with $q \le q_n$. Thus, if we define $\|y\| := \inf_{n\in\mathbb{Z}} |y - n|$ as the distance between a real number y and the integer lattice $\mathbb{Z} \subset \mathbb{R}$, then the denominators of the convergents minimize $\|qx\|$ over all $q \le q_n$.

Moreover, using again the recursion (1.1) in (1.2) we get

$$\frac{1}{(a_{n+1} + 2)\, q_n^2} < \left| x - \frac{p_n}{q_n} \right| < \frac{1}{a_{n+1}\, q_n^2} \tag{1.3}$$

A first consequence is that there exist numbers which are approximable to any given degree $\phi(q)/q^2$, with $\phi : \mathbb{N} \to \mathbb{R}_+$, i.e. satisfying $|x - p/q| < \phi(q)/q^2$ infinitely often, for we may recursively choose the partial quotients in such a way that $a_{n+1} > 1/\phi(q_n)$. A second consequence is that the worst numbers, from the point of view of rational approximations, are those with bounded partial quotients, for if $a_n < M$ then the convergents satisfy $|x - p_n/q_n| > c/q_n^2$ for some constant $c = c(M)$ (and the worst is, of course, the number with smallest possible partial quotients, which is $[1; 1, 1, 1, \dots]$, as we already know!). These numbers are called *badly approximable*.

1.8 Modular group and Farey series

Modular group and equivalence. The *modular group* is the group $\mathrm{PSL}_2(\mathbb{Z}) = \mathrm{SL}_2(\mathbb{R})/\pm I$ of two-by-two integer matrices $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with determinant 1 modulo the equivalence relation $A \sim -A$. Rows and columns of such matrices are made of relatively prime integers, a fact that makes us suspect their relevance in number theory! The modular group acts on the Poincaré upper half-plane $\mathbb{H} := \{z \in \mathbb{C} \text{ s.t. } \Im(z) > 0\}$ as fractional linear transformations

$$x \mapsto \frac{az + b}{cz + d}$$

This actions extends by continuity to a bijection of the "ideal boundary" $\partial\mathbb{H} \approx \mathbb{R} \cup \{\infty\}$, provided we set $(a\infty + b)/(c\infty + d) = a/c$. The two-fold cover $\mathrm{SL}_2(\mathbb{Z})$ is generated by the translation $z \mapsto z + 1$, defined by the matrix $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, and the inversion $z \mapsto -1/z$, defined by the matrix $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, satisfying the relations $S^2 = 1$ and $(ST)^3 = 1$ (and one can actually show that it is isomorphic, as a group, to the free product $\mathbb{Z}_2 * \mathbb{Z}_3$).

It happens that the continued fractions of two "equivalent" numbers x and y in the ideal boundary $\partial\mathbb{H} \approx \mathbb{R} \cup \{\infty\}$, i.e. two numbers in the same $\mathrm{PSL}_2(\mathbb{Z})$ orbit, have eventually equal continued fractions representations. This means that $x \sim [a_0; a_1, a_2, \dots]$ and $y \sim [b_0; b_1, b_2, \dots]$ are related by $y = (ax + b)/(cx + d)$ with integer coefficients a, b, c, d such that $ad - bc = \pm 1$ iff there exist sufficiently large "times" n and m such that $a_{n+k} = b_{m+k}$ for all $k \geq 1$.

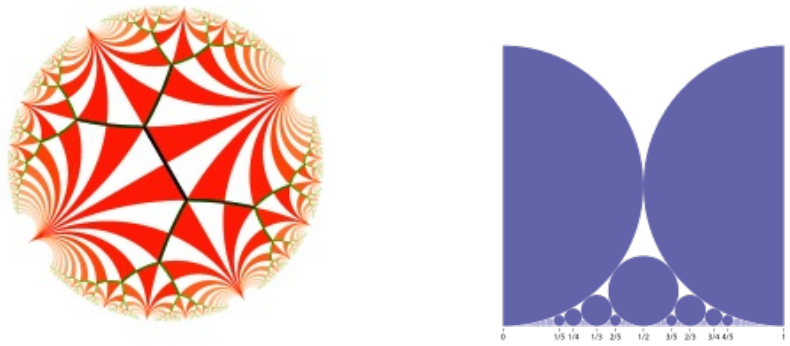
Patterns in continued fractions. In particular, since $\mathrm{PSL}_2(\mathbb{Z})$ form a group, if the continued fractions of x is eventually periodic then $x = (ax + b)/(cx + d)$ with integers a, b, c, d, and therefore x is a root of a quadratic polynomial with integer coefficients. The converse is also true, and is due to Lagrange [La70]:

Theorem 1.19 (Lagrange, 1770)**.** *Eventually periodic continued fractions correspond to quadratic irrationals.*

For example, $(\sqrt{5} + 1)/2 \sim [1; 1, 1, 1, 1, \dots]$.

Not too much is known on other possible regularities of continued fractions (but late Arnold had some conjectures, illustrated in a seminar I've seen on youtube).

Modular orbifold and Ford circles. The Poincaré upper half-plane $\mathbb{H}$ is a model of the hyperbolic plane, equipped with the Riemannian metric $dzd\bar{z}/\Im(z)^2$. Another model is the *unit disk* $\mathbb{D} := \{z \in \mathbb{C} \text{ s.t. } |z| < 1\}$, equipped with the metric $dzd\bar{z}/(1 - |z|^2)^2$. The action of fractional linear transformations $\mathrm{PSL}_2(\mathbb{R})$ on $\mathbb{H}$ is by hyperbolic isometries, and the quotient $M = \mathbb{H}/\mathrm{PSL}_2(\mathbb{Z})$ is called *modular orbifold* (it is the "moduli space" of complex tori, any complex torus being conformally equivalent to $\mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z})$ for some $\tau \in M$). The $\mathrm{PSL}_2(\mathbb{Z})$ orbit of the point $\infty \in \partial\mathbb{H}$ is the set of rationals, and the orbit of the "horoball" $H = \{z \in \mathbb{H} \text{ s.t. } \Im(z) > 1\}$ centered at ∞ is the set of *Ford disks*, which touch the real line at reduced rationals p/q and have diameter $1/q^2$. Ford disks form a packing by pairwise tangent disks (similar to a packing considered by Apollonius of Perga).



Tessellation of the modular group and Ford circles
(from http://en.wikipedia.org/wiki/Ford_circle)

A vertical line over an irrational $x \in \mathbb{R} \subset \partial\mathbb{H}$ (a hyperbolic geodesic coming from ∞ if parametrized as $t \mapsto x + ie^{-t}$) enters in infinitely many such disks. Therefore, we get one more proof of Dirichlet theorem: for any irrational x there exist infinitely many reduced fractions p/q such that

$$\left| x - \frac{p}{q} \right| < \frac{1}{2q^2} \, .$$

We now reduce the diameters of the Ford disks to $1/q^{2+\varepsilon}$ with $\varepsilon > 0$. They do not touch anymore, but leave a lot of room. Indeed, we may bound the sum of the shadows of those disk with $0 \leq p/q \leq 1$ as smaller than (forgetting the coprime conditions)

$$\sum_{q=1}^{\infty} \sum_{p=0}^{q} \frac{1}{q^{2+\varepsilon}} \leq \sum_{q=1}^{\infty} \frac{1}{q^{1+\varepsilon}} < \infty$$

There follows from the easy half of the Borel-Cantelli lemma that for almost all numbers x there exist only finitely many reduced fractions p/q such that

$$\left| x - \frac{p}{q} \right| < \frac{1}{2q^{2+\varepsilon}} \, ,$$

Therefore,

Theorem 1.20 (folklore)**.** *For all $\tau > 2$, the set of those numbers satisfying a Diophantine condition of order τ has full measure.*

This is a sort of complement to Roth theorem (which deals with the zero measure, hence more elusive, set of algebraic numbers). In particular, as Cassels wrote ([Cas57], page 119), Roth's "criterion of transcendence 'almost never' applies"!

For more about the relation between continued fractions and the hyperbolic geometry of the modular orbifold you may start with [Se85] and look for papers by Curtis McMullen.

Farey series. Given a maximal denominator $Q \in \mathbb{N}$, the set of those reduced fractions p/q with $1 \leq q \leq Q$ and $0 \leq p \leq q$ (hence in the interval $[0, 1]$), arranged in increasing size, form the *Farey sequence* of order Q, say $\mathcal{F}_Q$ (this is not a sequence, but a finite ordered set, and Farey was not a mathematician, but a geologist!). Alternatively, consider the points (q, p) of the lattice $\mathbb{Z}^2$ which are "visible" from the origin and belong to the triangle $0 < x \leq Q$ and $0 < y \leq x$. The slopes p/q of the segments joining them to the origin, arranged in increasing size, form the Farey sequence of order Q (see chapter III of [HW59], and also [Ar78], pages 110-112, or the original [Kl09], for Klein's beautiful geometric interpretation). For example,

$$\mathcal{F}_5 = \left\{ \frac{0}{1}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{1}{1} \right\}$$

Two Ford circles are tangent iff they touch the real line at points p/q and p'/q' such that $pq' - p'q = \pm 1$, hence at neighbors elements of a Farey sequence (indeed, together with the origin, the points

(q,p) and (q',p') are vertices of an elementary triangle, whose area must be $1/2$!). Indeed, neighborhood relations of elements in each Farey sequence reflect arithmetic properties according to the following elementary fact, discovered by Farey and proved by Cauchy, 1816. If the reduced fractions $p/q < p'/q'$ are neighbors of a Farey sequence, then

$$p'q - pq' = 1\,.$$

Equivalently, if the reduced fractions $p/q < p'/q' < p''/q''$ are neighbors in a Farey sequence, then

$$p'/q' = \frac{p+p''}{q+q''}\,.$$

This fact may be used, as explained in chapter XI of [HW59] (a proof using continued fractions is in [Cas57]), to prove the following

Theorem 1.21 (Hurwitz, 1891)**.** *For all irrational numbers x, there exist an infinity of rationals p/q such that*

$$\left| x - \frac{p}{q} \right| < \frac{1}{\sqrt{5}\,q^2}\,.$$

The constant $1/\sqrt{5}$ is now optimal, if we want to include all irrationals, as those numbers equivalent to the simplest continued fraction, the "ratio" $(1+\sqrt{5})/2 \sim [1;1,1,1,\dots]$.

1.9 Uniform distribution

Minimal rotations. Consider the unit circle $\mathbb{R}/\mathbb{Z} \approx [0,1)$, its points identified with classes $x+\mathbb{Z}$ or with fractional parts $\{x\} := x - \lfloor x \rfloor \in [0,1)$, for $x \in \mathbb{R}$. Any number, or "angle", $\theta \in \mathbb{R}$ defines a "rotation" of the unit circle according to

$$R_\theta(x+\mathbb{Z}) := x + \theta + \mathbb{Z}\,,$$

and the (forward) orbit of R_θ is the collection of points $R_\theta^q(x+\mathbb{Z}) = x + q\theta + \mathbb{Z}$ for $q \in \mathbb{N}$. It is clear that all orbits of a rational rotation are periodic. Dirichlet theorem says that the forward orbit of 0 of an irrational rotation, the set of points θ, 2θ, 3θ, …, $q\theta$, …, passes infinitely often $1/q$-near 0 (i.e. at a definite speed). Indeed, more is true, if we forget the estimate of the error: all orbits of an irrational rotation of the circle are dense, or, in the modern language of dynamical systems,

Theorem 1.22 (Kronecker, 1884)**.** *An irrational rotation of the circle is minimal.*

This is but a particular case of a general result by Kronecker [Kr84]. We say that the frequencies/numbers $\omega_1, \omega_2, \dots, \omega_k$ are *linearly independent over the rationals* (a physicist would say that "the vector $\omega = (\omega_1, \omega_2, \dots, \omega_k)$ is *not resonant*") if the only rational solution of the equation

$$n_1\omega_1 + n_2\omega_2 + \cdots + n_k\omega_k = 0$$

is the trivial solution $n_1 = n_2 = \cdots = n_k = 0$. An important example: the logarithms $\omega_k = \log p_k$ of different primes p_k are linearly independent, as follows from the uniqueness of prime decomposition.

Theorem 1.23 (Kronecker, 1884)**.** *If θ_1, θ_2, …, θ_n, 1, are linearly independent over the rationals, then the set of points*

$$\mathbb{Z}\,(\theta_1, \theta_2, \dots, \theta_n) + \mathbb{Z}^n$$

is dense in the torus $\mathbb{R}^n/\mathbb{Z}^n \approx [0,1)^n$.

This means that for any $x = (x_1, x_2, \dots, x_n) \in [0,1)^n \approx \mathbb{R}^n$ and any precision $\varepsilon > 0$ we can find integers $p_1, p_2, \dots, p_n$ and n such that $|n\theta_k - p_k - x_k| < \varepsilon$ for all $k = 1, 2, \dots, n$. Chapter XXIII of [HW59] contains some different proofs, my favorite being Bohr's analytic proof.

In general, the subgroup of $\mathbb{T}^n/\mathbb{Z}^n$ generated by $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ will be either finite or its closure will be a torus $\approx \mathbb{R}^k/\mathbb{Z}^k$ of dimension $k \le n$.

Billiards. Kronecker theorem is actually a statement about orbits in a square billiard: orbits are either closed, if they have rational slope, or dense. This is the beginning of another long story, starting with [KS13], and particularly active these days ...

Uniform distribution and exponential sums. Much more is true. We state for simplicity just the one-dimensional case. We say that a sequence $(x_n)_{n\in\mathbb{N}}$ is uniformly distributed in the circle $\mathbb{R}/\mathbb{Z} \approx [0,1)$ if for any interval $I \subset [0,1)$, the cardinality $C_I(n)$ of those points $x_1, x_2, \ldots, x_n$ belonging to I is

$$C_I(n) = n \cdot |I| + o(n)\,,$$

i.e., if the ratio $C_I(n)/n$ converges to the length $|I|$ of the interval. Equivalently (well, this is a theorem by Hermann Weyl!), if for any continuous observable $\varphi : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$, the Birkhoff averages converge to the mean value, i.e.

$$\frac{1}{N+1}\sum_{n=0}^{N} \varphi(x+n\theta) \to \int_0^1 \varphi(t)\,dt\,.$$

Above, we identify a function on the circle $\mathbb{R}/\mathbb{Z}$ with a $\mathbb{Z}$-periodic function on its universal covering $\mathbb{R}$. It is interesting to observe that continuity of the observable f may be substituted by Riemann integrability, but not by Lebesgue integrability (since a Lebesgue integrable function may be zero along an orbit, which is countable, without having zero mean). This last statement, when the convergence is uniform in x, is called *unique ergodicity* (w.r.t. the Lebesgue measure, see Oxtoby theorem) in the slang of dynamical systems.

Theorem 1.24 (Weyl, 1916)**.** *If θ is irrational, then the points $\{n\theta\}$ are uniformly distributed in $[0,1) \approx \mathbb{R}/\mathbb{Z}$.*

Weyl observed that, according to Weierstrass theorem, trigonometric polynomials are dense in the space of continuous functions on the circle (equipped with the sup norm), and therefore it is sufficient to check uniform distribution on the "characters" $e_\xi(x) := e^{2\pi i\xi x}$, with $\xi \in \mathbb{Z}$ (see [SS03]). Indeed, more important than the actual result above is *Weyl's criterion* relating uniform distribution to a bound on the corresponding exponential sums [We16].

Theorem 1.25 (Weyl, 1916)**.** *A sequence (x_n) of points $x_n \in [0,1) \approx \mathbb{R}/\mathbb{Z}$ is uniformly distributed (w.r.t. Lebesgue measure) iff*

$$\frac{1}{N}\sum_{n=1}^{N} e_\xi(x_n) \to 0$$

as $N \to \infty$ for all non-zero "frequencies" $\xi \in \mathbb{Z}\backslash\{0\}$.

The proof of Weyl theorem 1.24 then goes as follows : for integer $\xi \neq 0$ and irrational θ,

$$\left|\frac{1}{N+1}\sum_{n=0}^{N} e_\xi(x+n\theta)\right| = \left|\frac{1}{N+1}\sum_{n=0}^{N} e^{2\pi i\xi\theta m}\right| \leq \frac{2}{N+1}\cdot\frac{1}{|1-e^{2\pi i\xi\theta}|} \to 0$$

uniformly in x (while the time averages of e_0 are constant and equal to 1).

Today, this is just a particular case of a general statement about translations in compact abelian topological groups. Weyl himself and then Furstenberg extended this result proving uniform distribution for values of polynomials with at least one irrational coefficient, ... but this is another long story (see [EW10] and some of the n books by Terence Tao).

Today, many other example os equidistributed sequences are known. Not so difficult is, for example, to show that the sequence $\theta\sqrt{n}$ is equidistributed, provided θ is irrational. More difficult is the following

Theorem 1.26 (Vinogradov, 1935)**.** *If θ is irrational, then the points $\{p_n\theta\}$ are uniformly distributed in $[0,1) \approx \mathbb{R}/\mathbb{Z}$, where p_n denotes the n-th prime.*

Speed of equidistribution. The error may be bounded provided some "Diophantine condition" on the rotation angle θ, depending on the regularity of the test function. This is done using some Fourier analysis (actually, Sobolev spaces). Indeed, if ψ is a solution of the cohomological equation

$$\psi(x+\theta)-\psi(x)=\varphi(x)\,,$$

then the Birkhoff sums of the observable φ are telescopic, and equal to

$$\frac{1}{N}\sum_{n=0}^{N-1}\varphi(x+k\theta)=\frac{\psi(x+N\theta)-\psi(x)}{N}$$

The cohomological equation may be solved in Fourier series provided some Diophantine condition on θ, as in Arnold's version of KAM theorem for circle diffeomorphisms.

1.10 Zero-one laws

Metric theory. We may forget about the nature of the numbers that we want to approximate, but try to measure the relative sizes of those sets which are or are not approximable with the desired speed, and possibly look for results valid for almost all numbers. This is called "metric" (although a better term would be "probabilistic") Diophantine approximation. The story starts with Borel, Bernstein, and Khinchin at the beginning of the XX century.

For example, we know since Cantor that the set of algebraic numbers is countable, and therefore has zero measure. More interesting is Borel's theorem saying that the set of "bounded type" numbers (i.e. those numbers x such that their continued fraction has bounded partial quotients, say $a_n \leq M$ for some $M=M(x)$) has zero measure too [Bo09, Be12]. Thus, typical numbers have unbounded partial quotients. More precisely, we have the following dichotomy.

Theorem 1.27 (Borel-Bernstein). *Let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence of positive integers.*

- *If $\sum \frac{1}{\alpha_n}<\infty$ then for almost all reals $x\sim[a_0;a_1,a_2,\dots]$ the inequalities $a_n\geq\alpha_n$ holds for only finitely many n's*
- *If $\sum \frac{1}{\alpha_n}=\infty$ then for almost all reals $x\sim[a_0;a_1,a_2,\dots]$ the inequalities $a_n\geq\alpha_n$ holds infinitely often.*

Another relatively simple result is Khinchin's uniform estimate $c\leq\sqrt[n]{q_n}\leq C$, or, equivalently,

$$e^{an}\leq q_n\leq e^{An}$$

for the denominators of the convergents of Lebesgue almost all x (the left inequality being trivial and valid for all x, since $q_n\geq 2^{n-1}$) [Kh35].

Khinchin's zero-one law. The real big achievement is Khinchin's famous dichotomy [Kh24, Kh35]

Theorem 1.28 (Khinchin, 1924). *Let $\varphi:\mathbb{N}\to\mathbb{R}_+$ be non-increasing. Then for almost all real numbers x, the inequality*

$$\left|x-\frac{p}{q}\right|<\frac{\varphi(q)}{|q|}$$

- *holds only finitely often if $\sum_{n=1}^{\infty}\varphi(n)<\infty$,*
- *and holds infinitely often if $\sum_{n=1}^{\infty}\varphi(n)=\infty$.*

Thus, for example, for almost all x we may find an infinite number of reduced fractions p/q such that

$$\left|x-\frac{p}{q}\right|<\frac{1}{q^2\,\log q}$$

but only a finite number such that

$$\left|x-\frac{p}{q}\right|<\frac{1}{q^2\,(\log q)^{1+\varepsilon}}$$

for $\varepsilon>0$.

Borel-Cantelli and quasi-independence. Khinchin theorem is a typical 0-1 law of probability theory, and although the original proof uses continued fractions, we easily recognize a variation of the (now) classical *Borel-Cantelli lemma* [Bo09, Ca17]. Indeed, the theorem estimates the measure (i.e. probability) of the set $W(\varphi)$ of "φ-approximable" numbers, the set of those $x \in \mathbb{R}/\mathbb{Z} \approx [0,1)$ for which the inequality

$$\|qx\| < \varphi(q)$$

has infinite natural solutions q. Define $B_{p,q}(\varphi) := (p/q - \varphi(q)/q, p/q + \varphi(q)/q)$, the "balls" of radius $\phi(q)/q$ centered at p/q, and $A_q(\varphi) := \bigcup_{0\le p\le q} B_{p,q}(\varphi)$. Then

$$W(\varphi) = \limsup_{q\to\infty} A_q(\varphi) := \bigcap_{n=1}^{\infty} \bigcup_{q=n}^{\infty} A_q(\varphi)\,.$$

Therefore, the first statement of Khinchin theorem follows from the first half of the

Theorem 1.29 (Borel-Cantelli, 1909-17)**.** *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of events of the probability space $(\Omega, \mathcal{B}, \mathbb{P})$.*

- *If $\sum_n \mathbb{P}(A_n) < \infty$ then $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$.*
- *If $\sum_n \mathbb{P}(A_n) = \infty$ and if the A_n's form an independent family, then $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$.*

On the other hand, the family $A_q(\phi)$ is not independent. To establish the second half of Khinchin theorem more is needed. Some finer properties of continued fractions, as in Khinchin's original proof, or an extension of the Borel-Cantelli lemma to "quasi-independent" families , satisfying

$$\sum_{n,m=1}^{N} \mathbb{P}(A_n \cap A_m) \le \left(\sum_{n=1}^{N} \mathbb{P}(A_n)\right)^2 + C \cdot \left(\sum_{n=1}^{N} \mathbb{P}(A_n)\right)$$

A different proof, using more arithmetic, is explained in [Ha98, Kl10].

In the 80's Dennis Sullivan extended Khinchin ideas to the non-arithmetic situation of geodesic excursion in non compact hyperbolic surfaces [Su82]. The quasi-independence is then a consequence of mixing of the geodesic flow. More geometric proofs use a "shrinking targets" argument, and apply to a variety of flows on homogeneous spaces. For recent and beautiful mathematics, you may look for the papers by Margulis, Kleinbock, Shah, Stratman, Velani, Lindenstrauss ...

A nice popular texts on probabilistic independence in number theory is [Ka59]. There is also an interesting paper by Dodson, but I don't remember which one!

1.11 Gauss map and ergodicity

The Gauss map. Numbers in the unit interval $(0,1]$ are uniquely represented, i.e. "coded", by continued fractions. If we disregard rationals, which form a set of zero Lebesgue measure, we are left with infinite continued fractions $[0; a_1, a_2, a_3, \dots]$, i.e. one-sided infinite sequences $(a_1, a_2, a_3, \dots) \in \mathbb{N}^{\mathbb{N}}$. Recall that the *Gauss map* $G : (0,1] \to [0,1]$ is defined as

$$G(x) := 1/x - \lfloor 1/x \rfloor \qquad \text{if } x \neq 0$$

(but we may also define $G(0) = 0$). Observe that for any rational $r \in \mathbb{Q}$ there exists a time n such that $G^n(r) = 0$. The infinite sequence of the continued fraction expansion of $x \sim [0; a_1, a_2, a_3, \dots] \in [0,1]\backslash\mathbb{Q}$ is a coding of the orbit of x. Indeed,

$$G([0; a_1, a_2, a_3, \dots]) = [0; a_2, a_3, a_4, \dots]$$

This means that $a_n = \lfloor 1/G^{n-1}(x) \rfloor$, or, equivalently, $a_n = k$ if $G^n(x) \in \left[\frac{1}{k+1}, \frac{1}{k}\right)$. In the language of dynamical systems, the Gauss map (restricted to the full measure set of irrationals) is conjugated to the one-sided shift $\sigma : \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$, the conjugation being the continued fraction representation $x \sim [0; a_1, a_2, a_3, \dots]$. In particular, the equivalence relation coming from the action of $\mathrm{PSL}_2(\mathbb{Z})$ corresponds to "being in the same great orbit" of the Gauss map.

Ergodicity and distribution of digits. It is essentially due to Gauss himself the crucial observation that the absolutely continuous measure μ with density

$$d\mu(x) = \frac{1}{\log 2}\,\frac{1}{1+x}\,dx$$

is an invariant probability measure for G, meaning that $\mu(G^{-1}(B)) = \mu(B)$ for all Borel subsets $B \in (0,1]$. Indeed, more is true [Kn26]:

Theorem 1.30 (Knopp, 1926)**.** *The Gauss measure μ is ergodic for the Gauss map.*

Modern proofs are in [Ma87, EW10]. Ergodic means that invariant subsets have measure 0 or 1, or, equivalently, invariant observable $f : (0,1] \to \mathbb{C}$ are constant μ-a.e. (an observable f is invariant if $f \circ G = f$, i.e. if it is constant along orbits). There follows from the *Birkhoff-Khinchin ergodic theorem* (see, for example, [Wa82, Ma87, KH95]) that time-averages of integrable (i.e. in $L^1(\mu)$) observables f converge μ-a.e. and are equal to the space averages, i.e.

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(G^n(x)) = \int_0^1 f(x)\,d\mu(x) \qquad \mu \text{ - a.e.}$$

In particular, we may compute the distribution of digits in the continued fraction representation of a number in the unit interval, simply taking for f the indicator function of some digit $d \in \mathbb{N}$ in $\mathbb{N}^{\mathbb{N}} \approx (0,1]$. The result is the *Gauss-Kuzmin distribution* (conjectured by Gauss and proved by Kuzmin [Ku28], see also [Ar78]):

Theorem 1.31 (Gauss-Kuzmin, 1928)**.** *For almost every real number x, the asymptotic frequency of the digit d in the continued fraction representation $x \sim [a_0; a_1, a_2, a_3, \dots]$ is*

$$p_d = \frac{1}{\log 2}\log\left(1 + \frac{1}{d(d+2)}\right)$$

The ergodicity of the Gauss map w.r.t. the Gauss measure imply many other "surprising" results, for clever choices of the observable f. For example, if we choose $f(x) = \log(a_1)$ then the Birkhoff averages are the geometric means of the first n partial quotients. There follows that for almost all numbers $x \sim [a_0; a_1, a_2, a_3, \dots]$ the limit $\lim_{n\to\infty} \sqrt[n]{a_1 a_2 a_3 \dots a_n}$ exists and is a constant, equal to

$$\prod_{n=1}^{\infty}\left(1 + \frac{1}{n(n+2)}\right)^{\log_2 n} \simeq 2.6854\dots,$$

a number now called *Khinchin constant* [Kh35]. A similar result is: the n-th root of the denominators q_n of the convergents of almost all numbers converge to

$$\lim_{n\to\infty}\sqrt[n]{q_n} = e^{\pi^2/(12\log 2)} \simeq 3.2758\dots,$$

a number called *Khinchin-Lévy constant* [Kh24, Le29]. On the other hand, the arithmetic mean of the partial quotients is unbounded for almost all numbers.

References

[Ar78] V.I. Arnold, *Metodi geometrici della teoria delle equazioni differenziali ordinarie*, Editori Riuniti - MIR, Roma 1978.

[Be12] F. Bernstein, Über eine Anwendung der Mengenlehre auf ein aus der Theorie der säkularen Störungen herrührendes Problem, *Math. Ann.* **71** (1912), 417-439.

[Bi65] P. Billingsley, *Ergodic Theory and Information*, Wiley, 1965.

[Bl14] H.F. Blichfeldt, A New Principle in the Geometry of Numbers, with Some Applications, *Trans. Amer. Math. Soc.* **15** (1914), 227-235.

[Bo09] É. Borel, Les probabilités dénombrables et leurs applications arithmetiques, *Rend. Circ. Mat. Palermo* **27** (1909), 247-271.

[Bo13] É. Borel, Mécanique Statistique et Irréversibilité, *J. Phys. 5e série* **3** (1913), 189-196.

[Ca17] F.P. Cantelli, Sulla probabilità come limite della frequenza, *Atti Accad. Naz. Lincei* **26** (1917), 39-45.

[Cas57] J.W.S. Cassels, *An introduction to diophantine approximation*, Cambridge University Press, 1957.

[Cas59] J.W.S. Cassels, *An introduction to the geometry of numbers*, Springer, 1959.

[Do93] M. Dodson, Geometric and probabilistic ideas in the metric theory of Diophantine approximations (Russian), *Uspekhi Mat. Nauk* **48** no. 5 (1993), 77-106 [translation in *Russian Math. Surveys* **48** no. 5 (1993), 73-102].

[EW10] M. Einsiedler and T. Ward, *Ergodic Theory with a view towards Number Theory*, GTM **259**, Springer, 2010.

[HW59] G.H. Hardy and E.M. Wright, *An Introduction to the Theory of Numbers*, fourth edition, Oxford University Press, 1959.

[Ha98] G. Harman, *Metric number theory*, Oxford University Press, 1998.

[Hu91] Hurwitz, Ueber die angenäherte Darstellung der Irrationalzahlen durch rationale Brüche, *Math. Ann.* **39** (1891), 279-284.

[IR90] K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory*, Springer, 1990.

[Ka59] M. Kac, *Statistical Independence in Probability, Analysis and Number Theory*, Carus Mathematical Monographs, Mathematical Association of America, 1959.

[Kh24] A.Y. Khinchin, Einige Sätze über Kettenbrüche, mit Anwendungen auf die Theorie der Diophantischen Approximationen, *Math. Ann.* **92** (1924), 115-125.

[Kh35] A.Y. Khinchin, *Continued Fractions*, 1935 [translation by University of Chicago Press, 1954].

[KH95] A. Katok and B. Hasselblat, *Introduction to the modern theory of dynamical systems*, Encyclopedia of mathematics and its applications, Cambridge University Press 1995.

[Kl09] F. Klein, *Matemática elementar de um ponto de vista superior*, Sociedade Portuguesa de Matemática, 2009. [*Elementarmathematik von höheren Standpunkte aus*, Teubner, 1908]

[Kl10] D. Kleinbock, *Metric Diophantine approximation and dynamical systems*, Brandeis University, 2010.

[Kn26] K. Knopp, Mengentheoretische Behandlung einiger Probleme der diophantischen Approximationen und der transfiniten Wahrscheinlichkeiten, *Math. Ann.* **95** (1926), 409-426.

[Ko36] J.F. Koksma, *Diophantische Approximationen,* Ergebnisse der Mathematik **4**, Springer, 1936.

[Kr84] Kronecker, Die Periodensysteme von Funktionen Reeller Variablen, *Berliner Sitzungsberichte* (1884), 1071-1080.

[KS13] D. König, A. Szücs, Mouvement dun point abandonné à l'intérieur dun cube, *Palermo Rend.* **36** (1913), 79-90.

[Ku28] R.O. Kuzmin, Ob odnoi zadache Gaussa, *Doklady akad. nauk,* ser. A (1928), 375-380.

[La70] J.L. Lagrange, Additions au mémoire sur la résolution des équations numériques, *Mém. Acad. Royale soc. et belles-lettres Berlin* **24** (1770).

[Le29] P. Lévy, Sur les lois de probabilité dont dependent les quotients complets et incomplets d'une fraction continue, *Bull. Soc. Math.* **57** (1929), 178-194.

[Ma87] R. Mañé, *Ergodic Theory and Differentiable Dynamics*, Springer-Verlag, 1987.

[Mar00] S. Marmi, *An Introduction To Small Divisors*, 2000.

[Mi10] H. Minkowski, *Geometrie der Zahlen*, Leipzig, Teubner, 1910.

[Mi27] H. Minkowski, *Diophantische Approximationen*, Leipzig, Teubner, 1927.

[MT07] M.R. Murty and N. Thain, Pick's theorem via Minkowski's theorem, *Am. Mathematical Monthly* **114** (2007), 732-736.

[Ni47] I. Niven, A simple proof that π is irrational, *Bull. Amer. Math. Soc.* **53** (1947), 509.

[Po84] H. Poincaré, Sur une généralisation des fractions continues, *C.R.A.S. Paris* Sér. **A 99** (1884), 1014-1016 (Oeuvres V, pp. 185-187).

[Pi99] G.A. Pick, Geometrisches zur Zahlentheorie, *Sitzungber. Lotos (Prague)* **19** (1899), 311-319.

[Ro94] H.E. Rose, *A course in Number Theory*, Oxford University Press, 1994.

[Rot55] K.F. Roth, Rational approximations to algebraic numbers, *Mathematika* **2** (1955), 1-20.

[Sc80] W. Schmidt, *Diophantine approximation*, LNM vol. **785**, Springer-Verlag, 1980.

[Se85] C. Series, The modular surface and continued fractions, *J. London Math. Soc. (2)* **31** (1985), 69-80.

[Si45] C.L. Siegel, A mean value theorem in geometry of numbers, *Ann. Math.* **46** (1945), 340-347.

[Sp79] V. Sprindzuk, *Metric theory of Diophantine approximations*, John Wiley & Sons, 1979.

[SS03] E.M. Stein and R. Shakarchi, *Fourier Analysis: An Introduction*, Princeton University Press, 2003.

[St99] H. Steinhaus, *Mathematical Snapshots*, Dover, 1999.

[Su82] D. Sullivan, Disjoint spheres, approximation by imaginary quadratic numbers, and the logarithm law for geodesics, *Acta Mathematica* **149** (1982), 215-237.

[Wa82] P. Walters, *An Introduction to ergodic theory*, Graduate Texts in Math. **79**, Springer-Verlag 1982

[We16] H. Weyl, Über die Gleichverteilung von Zahlen mod. Eins, *Math. Ann.* **77** (1916), 313-352.