Metric Diophantine approximation and dynamical systems

Dmitry Kleinbock

Brandeis University, Waltham MA 02454-9110
E-mail address: kleinboc@brandeis.edu

This is a rough draft for the Spring 2007 course taught at Brandeis. Comments/corrections/suggestions welcome.

Last modified: March 29, 2007.

Here is a short **course description**:

The **main theme** is studying the way the set of rational numbers $\mathbb{Q}$ sits inside real numbers $\mathbb{R}$. This seemingly simple set-up happens to lead to quite intricate problems. The word 'metric' comes from 'measure' and implies that the emphasis will be put on properties of 'almost all' numbers; more precisely, for a certain approximation property of real numbers we will be interested in the magnitude (for example, in terms of Lebesgue measure or Hausdorff dimension) of the set of numbers having this property. The situation becomes even more interesting when one similarly considers $\mathbb{Q}^n$ sitting in $\mathbb{R}^n$; this has been the area of several exciting recent developments, which we may talk about in the second part of the course.

The **introductory part** will feature: rate of approximation of real numbers by rationals; theorems of Kronecker, Dirichlet, Liouville, Borel-Cantelli, Khintchine; connections with dynamical systems: circle rotations, hyperbolic flow in the space of lattices, geodesic flow on the modular survace, Gauss map (continued fractions).

Other topics could be (choice based upon preferences of the audience):

- Hausdorff measures and dimension, fractal sets and measures they support
- W. Schmidt's (α, β)-games, winning sets, badly approximable numbers
- ubiquitious systems, Jarnik-Besicovitch Theorem
- inhomogeneous approximation, shrinking target properties for circle rotations
- multidimensional theory, Diophantine properties of measures on $\mathbb{R}^n$, a connection with flows on $\mathrm{SL}_{n+1}(\mathbb{R})/\mathrm{SL}_{n+1}(\mathbb{Z})$

References:

- W. Schmidt, *Diophantine Approximation*, Springer, 1980.
- J.W.S. Cassels, *An introduction to Diophantine approximation*, Cambridge Univ. Press, 1957.
- G. Harman, *Metric number theory*, Oxford Univ. Press, 1998.
- E. Burger, *Exploring the number jungle: a journey into Diophantine analysis*, AMS, 2000.

1. Basic facts and definitions

1.1. ψ-approximable numbers. Our starting point is the elementary fact that $\mathbb{Q}$, the set of rational numbers, is dense in $\mathbb{R}$, the reals. In other words, every real number can be approximated by rationals, that is, for any $\alpha \in \mathbb{R}$ and any $\varepsilon > 0$ there exists $p/q \in \mathbb{Q}$ with

$$|\alpha - p/q| < \varepsilon\,. \tag{1.1}$$

The central question of this course is going to be the following: how well can various real numbers be approximated by rational numbers? Namely, how small can ε in (1.1) be chosen for varying $p/q \in \mathbb{Q}$? A natural approach is to compare the accuracy of the approximation of α by p/q to the 'complexity' of the rational number p/q, which can be measured by the size of its denominator q.

Here is a model statement along these lines.

PROPOSITION 1.1. *For any $\alpha \in \mathbb{R}$ and any $c > 0$, there exist infinitely many $(p, q) \in \mathbb{Z}^2$ such that*

$$|\alpha - p/q| < c/|q|\,, \quad \textit{i.e.}\ |q\alpha - p| < c\,. \tag{1.2}$$

The proof is elementary and is left to the reader (see Exercise 1.1; of course the proposition follows from Theorem 1.5, but the reader is invited to come up with a simpler proof). In the process of doing this exercise, or even while staring at the statement of the proposition, the following **remarks** may come to mind:

1.1.1. It makes sense to kind of multiply both sides of (1.1) by q, since in the right hand side of (1.2) one would still be able to get very small numbers. In other words, approximation of α by p/q translates into approximating integers by integer multiples of α.

1.1.2. If α is irrational, (p, q) can be chosen to be relatively prime, i.e. one gets infinitely many different rational numbers p/q satisfying (1.2). However if $\alpha \in \mathbb{Q}$ the latter is no longer true for small enough c (see Exercise 1.2). Thus it seems to be more convenient to talk about pairs (p, q) rather than $p/q \in \mathbb{Q}$, thus avoiding a necessity to consider the two cases separately.

1.1.3. If $c < 1/2$, for any q there is at most one p for which (1.2) holds. And also without loss of generality one can restrict oneself to $q > 0$.

With this in mind, let us introduce the following central

DEFINITION 1.2. Let ψ be a function $\mathbb{N} \to \mathbb{R}_+$ and let $\alpha \in \mathbb{R}$. Say that α is *ψ-approximable* (notation: $\alpha \in \mathcal{W}(\psi)$) if there exist infinitely many $q \in \mathbb{N}$ such that

$$|q\alpha - p| < \psi(q) \tag{1.3}$$

for some $p \in \mathbb{Z}$.

Note that according to this definition, rational numbers are ψ-approximable for any positive function ψ. On the other hand, if $\alpha \notin \mathbb{Q}$ and ψ is non-increasing (which in most cases will be our standing assumption), it is easy to see that one can equivalently demand the existence of infinitely many rational numbers p/q, or relatively prime pairs (p, q), satisfying (1.3). Indeed, let us define

$$\mathcal{W}'(\psi) \stackrel{\text{def}}{=} \left\{ \alpha \in \mathbb{R} \;\middle|\; \begin{array}{c} \text{(1.3) holds for infinitely many} \\ (p, q) \in \mathbb{Z}^2 \text{ with } \gcd(p, q) = 1 \end{array} \right\},$$

and state

Lemma 1.3. *For a non-increasing function ψ, one has*

$$\mathcal{W}(\psi) \smallsetminus \mathbb{Q} = \mathcal{W}'(\psi) \smallsetminus \mathbb{Q}.$$

Proof. Clearly $\mathcal{W}'(\psi) \subset \mathcal{W}(\psi)$. For the converse, take $\alpha \notin \mathbb{Q}$. If (1.3) holds for infinitely many (p, q), the ratio p/q must take infinitely many values, since (1.3) can be rewritten as $|\alpha - p/q| < \psi(q)/q$ with the right hand side tending to 0 as $q \to \infty$. Then one can write $p = p'n$, $q = q'n$, where $\gcd(p', q') = 1$, and conclude that

$$|q'\alpha - p'| = \tfrac{1}{n}|q\alpha - p| < \tfrac{1}{n}\psi(q) \le \psi(q) \le \psi(q').$$

□

See Exercise 1.2 for more on approximation properties of rational numbers.

Some more **remarks**:

1.1.4. Clearly the set of ψ-approximable numbers depends only on 'tail properties' of ψ. That is, $\mathcal{W}(\psi_1) = \mathcal{W}(\psi_2)$ if $\psi_1(x) = \psi_2(x)$ for large enough x. Hence in Definition 1.2 one can assume that $\psi(x)$ is only defined for large enough x. Also (and more generally), one clearly has $\mathcal{W}(\psi_1) \subset \mathcal{W}(\psi_2)$ if $\psi_1(x) \le \psi_2(x)$ for large enough x.

1.1.5. Without loss of generality one can only consider $\alpha \in [0, 1]$ or even $\alpha \in [0, 1/2]$, see Exercise 1.3.

1.1.6. Sometimes it will be convenient to let the domain of ψ be positive real numbers, not just integers. Clearly there is no loss of generality here, as any (non-increasing) function $\mathbb{N} \to \mathbb{R}_+$ can be extended to a (non-increasing) continuous function $\mathbb{R}_+ \to \mathbb{R}_+$.

1.1.7. Proposition 1.1 says that $\mathcal{W}(c) = \mathbb{R}$ for any $c > 0$, or, in the terminology of Exercise 1.4, $\check{\mathcal{W}}(1) = \mathbb{R}$. This is just a tip of the iceberg: our plan is to make ψ decay faster and see what happens to $\mathcal{W}(\psi)$. In fact, because of this proposition it is natural to assume, as we always will, that $\psi(x) \to 0$ as $x \to \infty$.

1.1.8. It will sometimes be convenient to use notation

$$\langle x \rangle \stackrel{\text{def}}{=} \operatorname{dist}(x, \mathbb{Z}).$$

Other sources use $\|x\|$ but I'd like to reserve the latter notation for norms in vector spaces. With this notation, assuming $\psi(x) < 1/2$ for large x, it is clear that α is ψ-approximable if and only if $\langle \alpha q\rangle < \psi(q)$ for infinitely many q.

1.2. Dynamical and geometric interpretations. Before going any further, let us discuss some restatements of the above definition.

1.2.1. *Circle rotations.* Given $\alpha \in \mathbb{R}$, let us denote by R_α the map of the unit circle S^1 to itself given by $R_\alpha(x) = x \mod 1$ (here we identify S^1 with $\mathbb{R}/\mathbb{Z}$). Then it is straightforward to see that α is ψ-approximable iff one has

$$\operatorname{dist}\big(R_\alpha^n(0), 0\big) < \psi(n) \text{ for infinitely many } n \in \mathbb{N}\,, \tag{1.4}$$

where by 'dist' we mean the metric on S^1 induced from $\mathbb{R}$. Note also that (1.4) is equivalent to

$$\text{for any } x \in S^1, \quad \operatorname{dist}\big(R_\alpha^n(x), x\big) < \psi(n) \text{ for infinitely many } n \in \mathbb{N}\,,$$

since circle rotations commute with each other.

Observe that Proposition 1.1 in this language says that R_α-orbit of any point comes back arbitrarily closely. This is trivial if α is rational, in which case every point is *periodic*, i.e. comes back exactly to the same place after a number of iterations. Bit if $\alpha \notin \mathbb{Q}$, orbit points return arbitrarily closely to the original position but never hit it. It is not hard to deduce from there (see Exercise 1.7) that every R_α-orbit in this case is *dense* in S^1, a statement known as Kronecker's Theorem.

1.2.2. *Lattice points near straight lines.* Given $\alpha \in \mathbb{R}$, let us denote by L_α the line in the xy plane given by $x = \alpha y$. Then it is straightforward to see that α is ψ-approximable iff there are infinitely many integer points $(p, q) \in \mathbb{Z}^2$ such that the distance, in the horizontal (x-axis) direction, between (p, q) and L_α is less than $\psi(|q|)$. In other words, there are infinitely many lattice points in a 'shrinking neighborhood' of L_α given by $\alpha y - \psi(|y|) < x < \alpha y + \psi(|y|)$.

1.2.3. *Lattice points in thin rectangles.* For the next interpretation we will need the following elementary

LEMMA 1.4. *Let ψ be non-increasing with $\psi(x) \to 0$ as $x \to \infty$, and let $\alpha \in \mathbb{R}$. Then α is ψ-approximable iff there exists an unbounded set of $Q > 0$ such that the system*

$$\begin{cases} |q\alpha - p| < \psi(Q) \\ \qquad |q| \le Q \end{cases} \tag{1.5}$$

has a nonzero integer solution.

PROOF. If (1.3) holds, one can take $Q = q$. Conversely, there is nothing to prove when α is rational, so one can assume that $\alpha \notin \mathbb{Q}$, in which case from the fact that $\psi(Q)$ tends to 0 as $Q \to \infty$ it follows that the set of solutions q of (1.5) is unbounded. It remains to observe that, since ψ is non-increasing, any solution of (1.5) also satisfies (1.3). $\square$

Here is a geometric way to express the conclusion of the lemma: let Λ_α be the lattice in $\mathbb{R}^2$ given by

$$\Lambda_\alpha \stackrel{\text{def}}{=} \left\{ \left(\begin{array}{c} q\alpha - p \\ q \end{array} \right) \Big| p, q \in \mathbb{Z} \right\} = \left\{ \left(\begin{array}{c} q\alpha + p \\ q \end{array} \right) \Big| p, q \in \mathbb{Z} \right\}, \tag{1.6}$$

in other words, $\Lambda_\alpha = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \mathbb{Z}^2$. Then α is ψ-approximable iff there exists an unbounded set of $Q > 0$ such that the intersection of Λ_α with the rectangle $\{|x| < 1/Q, \ |y| \le Q\}$ is nontrivial.

Perhaps it is worthwhile to make some general comments here, as this is the first occasion that we are introduced to lattices in $\mathbb{R}^2$, which will play a rather important role in this course. A *lattice* in $\mathbb{R}^2$ is a subgroup generated by two linearly independent vectors, or, equivalently, the image of $\mathbb{Z}^2$ under an invertible linear transformation of $\mathbb{R}^2$ (indeed, if Λ is generated by $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$, then $\Lambda = A\mathbb{Z}^2$, where A is the matrix with columns $\mathbf{u}, \mathbf{v}$; and in fact interchanging $\mathbf{u}, \mathbf{v}$ if necessary one can assume that $\det(A) > 0$). The quotient space $\mathbb{R}^2/\Lambda$ can be identified with the parallelogram spanned by $\mathbf{u}$ and $\mathbf{v}$, or by any other pair of generators, with opposite sides glued to each other.

The standard Lebesgue measure on $\mathbb{R}^2$, which we will denote by 'area', will naturally induce the area measure on $\mathbb{R}^2/\Lambda$. The total area of $\mathbb{R}^2/\Lambda$ is called the *discriminant*, or sometimes the *covolume*, of Λ and is denoted by $d(\Lambda)$. Clearly it is equal to $|\det(A)|$ where $\Lambda = A\mathbb{Z}^2$. A lattice Λ with $d(\Lambda) = 1$, that is, $\Lambda = A\mathbb{Z}^2$ with $A \in \mathrm{SL}_2(\mathbb{R})$, is called *unimodular*.

1.3. Dirichlet's Theorem. The next statement makes a serious improvement of Proposition 1.1.

THEOREM 1.5. *For any $\alpha \in \mathbb{R}$ and any $Q > 0$, there exist $p \in \mathbb{Z}$ and $q \in \mathbb{N}$ such that*

$$\begin{cases} |q\alpha - p| < 1/Q \\ \qquad q \le Q \end{cases}$$

We are going to consider two different proofs of this theorem, one due to Dirichlet (1842) and another one to Minkowski (1896). Both are conceptually important, especially since both allow higher-dimensional generalizations. However let us first state a straightforward corollary. It will be convenient to fix the following notation: for $c \ge 0$ and $v \ge 1$ set

$$\psi_{c,v}(x) \stackrel{\text{def}}{=} \frac{c}{x^v},$$

and also denote $\mathcal{W}(\psi_{c,v})$ by $\mathcal{W}_{c,v}$.

COROLLARY 1.6. *$\mathcal{W}_{1,1} = \mathbb{R}$; that is, for any $\alpha \in \mathbb{R}$ there exist infinitely many $q \in \mathbb{N}$ such that*

$$|\alpha - p/q| < 1/q^2 \tag{1.7}$$

for some $p \in \mathbb{Z}$, or, equivalently, $\langle \alpha q \rangle < 1/q$.

PROOF. Immediate from Lemma 1.4. □

The family of functions $\psi_{c,v}$ with $v \geq 1$ and $c > 0$ will play an important role in what follows. Namely, we are going to find out how the sets $\mathcal{W}_{c,v}$ depend on c and v, and the values $c = v = 1$ will form our reference point in the parameter plane.

Now let us get back to proving Theorem 1.5.

DIRICHLET'S PROOF OF THEOREM 1.5. Let $n = [Q]$. (Here and hereafter $[x]$ and $\{x\}$ stand for the integer and fractional parts of x.) If $\alpha = p/q$ with $1 \leq q \leq n$, there is nothing to prove. Otherwise, the $n+2$ points $0, \{x\}, \dots, \{nx\}, 1$ are pairwise distinct and divide the interval $[0,1]$ into $n+1$ subintervals. By the pigeon-hole principle (which in fact was introduced by Dirichlet on this occasion) at least one of these intervals has length not greater than $\frac{1}{n+1}$. Thus there exist integers p_1, p_2, q_1, q_2 with $0 \leq q_1 < q_2 \leq n$ and

$$|(q_1\alpha - p_1) - (q_2\alpha - p_2)| \leq \frac{1}{n+1} < 1/Q\,.$$

The proof is finished by setting $p = p_1 - p_2$ and $q = q_1 - q_2$, and by noticing that q satisfies $1 \leq q \leq n \leq Q$. □

For the other proof, we need the following very important lemma, a special case of what is sometimes called Minkowski's Convex Body Theorem.

LEMMA 1.7. *Let Λ be a lattice in $\mathbb{R}^2$, and let D be a convex centrally symmetric subset of $\mathbb{R}^2$ with $\operatorname{area}(D) > 4d(\Lambda)$. Then $D \cap \Lambda \neq \{0\}$.*

Note that 4 in the Lemma can not be replaced by a smaller number, as shown by choosing $\Lambda = \mathbb{Z}^2$ and $D = [-1+\varepsilon, 1-\varepsilon]^2$.

PROOF. Since D is convex and centrally symmetric, it coincides with the set

$$\tfrac{1}{2}D - \tfrac{1}{2}D = \{\mathbf{x} - \mathbf{y} \mid \mathbf{x}, \mathbf{y} \in D\}$$

(see Exercise 1.9). Suppose that $D \cap \Lambda = \{0\}$. Then it is not possible to find $\mathbf{x}, \mathbf{y} \in \frac{1}{2}D$ with $\mathbf{x} - \mathbf{y} \in \Lambda \smallsetminus \{0\}$. This amounts to saying that the natural projection π from $\mathbb{R}^2$ onto $\mathbb{R}^2/\Lambda$ is injective when restricted to $\frac{1}{2}D$. Hence $\frac{1}{4}\operatorname{area}(D) = \operatorname{area}(\frac{1}{2}D) \leq \operatorname{area}\big(\pi(\frac{1}{2}D)\big) \leq \operatorname{area}(\mathbb{R}^2/\Lambda) = d(\Lambda)$, a contradiction. □

MINKOWSKI'S PROOF OF THEOREM 1.5. Take Λ as in (1.6) and D given by

$$D = D_\varepsilon = \big\{(x,y) \;\big|\; |x| < 1/Q,\ |y| < Q + \varepsilon\big\}$$

for an arbitrary $\varepsilon > 0$. Since area$(D) = \frac{2}{Q} \cdot 2(Q+\varepsilon) > 4 = 4d(\Lambda)$, in view of the above lemma for any $\varepsilon > 0$ there exists a nonzero vector in $\Lambda \cap D_\varepsilon$. Letting $\varepsilon \to 0$ and using the discreteness of Λ, one gets $(x, y) \in \Lambda$ with $|x| < 1/Q$ and $|y| \leq Q$. □

1.4. Exercises.

Exercise 1.1. Give a proof of Proposition 1.1 which is more elementary than that of Theorem 1.5.

Exercise 1.2. Prove that the following are equivalent for $\alpha \in \mathbb{R}$:

(i) $\alpha \notin \mathbb{Q}$;
(ii) for any $c > 0$ there exist infinitely many rational numbers p/q such that (1.2) holds;
(iii) there exist infinitely many rational numbers p/q such that (1.7) holds.

Conclude that $\mathcal{W}'(\psi) = \mathcal{W}(\psi) \smallsetminus \mathbb{Q}$ whenever ψ is a non-increasing positive function with $\psi(x) \to 0$ as $x \to \infty$.

Exercise 1.3. Prove that the sets $\mathcal{W}(\psi)$ are invariant under translation and multiplication by integers, that is, $\alpha \in \mathcal{W}(\psi) \Rightarrow k\alpha + l \in \mathcal{W}(\psi)$ for any $k, l \in \mathbb{Z}$.

Exercise 1.4. Define

$$\hat{\mathcal{W}}(\psi) \stackrel{\text{def}}{=} \cup_{c>0} \mathcal{W}(c\psi) \quad \text{and} \quad \check{\mathcal{W}}(\psi) \stackrel{\text{def}}{=} \cap_{c>0} \mathcal{W}(c\psi)\,,$$

that is, $\alpha \in \hat{\mathcal{W}}(\psi)$ (resp., $\in \check{\mathcal{W}}(\psi)$) if for some (resp., for any) $c > 0$ there exist infinitely many $q \in \mathbb{N}$ with $\langle \alpha q \rangle < c\psi(q)$. (We will also use notation $\hat{\mathcal{W}}_{c,v} \stackrel{\text{def}}{=} \hat{\mathcal{W}}(\psi_{c,v})$ and $\check{\mathcal{W}}_{c,v} \stackrel{\text{def}}{=} \check{\mathcal{W}}(\psi_{c,v})$.)

Prove that the sets $\hat{\mathcal{W}}(\psi)$ and $\check{\mathcal{W}}(\psi)$ are invariant under translation and multiplication by rational numbers as well as under taking inverses, that is, under the transformations $\alpha \mapsto \frac{k\alpha+l}{m\alpha+n}$, where $k, l, m, n \in \mathbb{Z}$.

Exercise 1.5. Prove that for any positive non-increasing function ψ there exists another positive non-increasing function ψ' such that $\hat{\mathcal{W}}(\psi')$ is contained in $\check{\mathcal{W}}(\psi)$.

Exercise 1.6.

(a) Prove that for any positive function ψ the set $\mathcal{W}(\psi) \smallsetminus \mathbb{Q}$ is non-empty;
(b) use (a) together with Exercise 1.4 and Exercise 1.5 to conclude that $\mathcal{W}(\psi) \smallsetminus \mathbb{Q}$ is dense for any positive ψ;
(c) on the other hand, prove that $\mathbb{Q} = \bigcap_\psi \mathcal{W}(\psi)$, where the intersection is taken over all positive non-increasing functions ψ.

Exercise 1.7. Use Proposition 1.1 to show that every orbit of $R_\alpha : S^1 \to S^1$ for $\alpha \notin \mathbb{Q}$ is dense in S^1.

EXERCISE 1.8. Prove that the following are equivalent for a subgroup Λ of $\mathbb{R}^2$:

(i) Λ is a lattice in $\mathbb{R}^2$;
(ii) Λ is discrete and isomorphic to $\mathbb{Z}^2$;
(iii) Λ is discrete and has a subgroup isomorphic to $\mathbb{Z}^2$;
(iv) Λ is discrete and $\mathbb{R}^2/\Lambda$ is compact;
(v) Λ is discrete and $\mathbb{R}^2/\Lambda$ has finite area.

EXERCISE 1.9. Show that if $D \subset \mathbb{R}^2$ is convex and centrally symmetric, then $D = \frac{1}{2}D - \frac{1}{2}D$. Is the converse true? If yes, prove it, if no, give a counterexample and find an additional assumption on D which makes the statement true.

EXERCISE 1.10. Show that in Lemma 1.7 the assumption 'area$(D) > 4d(\Lambda)$' can be replaced by 'area$(D) \geq 4d(\Lambda)$ and D is compact'.

EXERCISE 1.11. Show that if in Lemma 1.7, as well as in the preceding exercise, one replaces '$4d(\Lambda)$' by '$4kd(\Lambda)$', $k \in \mathbb{N}$, it would follow that $D \cap \Lambda$ contains at least k pairs of vectors $\pm\mathbf{v}$ which are distinct from each other and from 0.

1.5. A big picture question: inhomogeneous approximation. What happens if instead of (1.3) we fix $\beta \in \mathbb{R}$ and consider solving

$$|q\alpha - \beta - p| < \psi(q) \tag{1.8}$$

for some $p \in \mathbb{Z}$? That is, instead of a linear form $q \mapsto \alpha q$ study an *affine form* $q \mapsto \alpha q - \beta$, or, instead of the imbedding $\mathbb{Q} \to \mathbb{R}$, study the way the set

$$\left\{ \tfrac{p+\beta}{q} \mid p, q \in \mathbb{Z} \right\} \tag{1.9}$$

sits inside $\mathbb{R}$?

This seems to be some kind of a translation of the previous problem but in fact it is not! For example, in §1.2.1 we have observed that the homogeneous problem corresponds to orbits of a rotation coming back to the starting point, while the new set-up describes approximating a point in S^1 by a trajectory of another point. Also, it is clear now that the case $\alpha \in \mathbb{Q}$ should be considered really separately, since in this case (1.8) can have solutions only for very special βs.

So let me ask a vaguely defined question: what stays the same and what needs to be changed in the exposition of the previous section if we allow a constant term in (1.8)? How to state (and even better, to prove) inhomogeneous analogues to some of the results above? such as Theorem 1.5 and Corollary 1.6? does either of the two proofs of Theorem 1.5 survive? Even though the question is not new, please think about it and if something interesting comes up out of it, write it down!

2. Farey series: from 1 to $1/\sqrt{5}$

Our goal in this section is to replace 1 in (1.7) by a smaller number and still get the same conclusion. It turns out that there is some room for improvement of both approaches of the preceding section.

2.1. Farey series. We start with

DEFINITION 2.1. Let $N \in \mathbb{N}$. The *Farey series* $\mathcal{F}_N$ of order N is the sequence of rational numbers in their lowest terms between 0 and 1 and with denominators $\leq N$, written in ascending order.

For example,

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

2.1.1. One can interprete elements of $\mathcal{F}_N$ geometrically by considering all the points inside the triangle $\{(x, y) \mid 0 \leq x \leq y \leq N\}$ with integer coordinates; clearly $\mathcal{F}_N$ is simple the increasing sequence of slopes of all the lines connecting those points with the origin. See Exercise 2.3 for another geometric interpretation.

We will show in this section how elements of $\mathcal{F}_N$ can be used to produce 'optimal' rational approximations to irational numbers. Here is a crucial property of Farey series that we are going to use:

PROPOSITION 2.2. *If* $\frac{m}{n} < \frac{p}{q}$ *are two successive elements of* $\mathcal{F}_N$ *for some* N*, then* $np - mq = 1$*.*

This proposition has an interesting implication:

COROLLARY 2.3. *If* $\frac{k}{l} < \frac{m}{n} < \frac{p}{q}$ *are three successive elements of* $\mathcal{F}_N$ *for some* N*, then* $\frac{m}{n} = \frac{k+p}{l+q}$*.*

PROOF. By the proposition, $np - mq = 1$ and $lm - kn = 1$, hence $m(l + q) - n(k + p) = 0$. □

See **http://www-gap.dcs.st-and.ac.uk/~history/Mathematicians/Farey.html** for the history of the latter corollary, as well as for J. Farey's somewhat controversial contribution to the subject.

Perhaps the simplest proof of Proposition 2.2 is geometric by nature. We will need a lemma which gives a criterion for two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{Z}^2$ to generate $\mathbb{Z}^2$ as a $\mathbb{Z}$-module.

LEMMA 2.4. *For* $\mathbf{u}, \mathbf{v} \in \mathbb{Z}^2$*, the following are equivalent:*

(i) $\mathbb{Z}\mathbf{u} \oplus \mathbb{Z}\mathbf{v} = \mathbb{Z}^2$*;*

(ii) *if* A *is the matrix with columns* $\mathbf{u}, \mathbf{v}$*, then* $|\det(A)| = 1$*;*

(iii) $\mathbf{u}$ *and* $\mathbf{v}$ *are not proportional, and the closed triangle with vertices* $0, \mathbf{u}, \mathbf{v}$ *contains no integer points but its vertices;*

(iv) $\mathbf{u}$ *and* $\mathbf{v}$ *are not proportional, and the closed parallelogram with vertices* $0, \mathbf{u}, \mathbf{v}, \mathbf{u} + \mathbf{v}$ *contains no integer points but its vertices.*

PROOF. From (i) it follows that the standard base vectors $\mathbf{i}, \mathbf{j}$ are integer linear combinations of $\mathbf{u}, \mathbf{v}$, in other words, A^{-1} has integer coefficients, which clearly implies (ii). Assuming (ii), one sees that A provides a one-to-one self-map of $\mathbb{Z}^2$ sending $\mathbf{i}, \mathbf{j}$ to $\mathbf{u}, \mathbf{v}$ respectively; therefore (iii) follows from the fact that the closed triangle with vertices $0, \mathbf{i}, \mathbf{j}$ contains no integer points but its vertices. The implication (iii)⇒(iv) is straightforward since $\mathbb{Z}^2$ is invariant under reflection around the line connecting $\mathbf{u}$ and $\mathbf{v}$.

It remains to assume (iv) and write any $\mathbf{w} \in \mathbb{Z}^2$ as a linear combination of $\mathbf{u}$ and $\mathbf{v}$: $\mathbf{w} = \lambda\mathbf{u} + \mu\mathbf{v}$. Then $\mathbf{w} = \mathbf{w}' + \mathbf{w}''$, where

$$\mathbf{w}' = [\lambda]\mathbf{u} + [\mu]\mathbf{v} \quad \text{and} \quad \mathbf{w}'' = \{\lambda\}\mathbf{u} + \{\mu\}\mathbf{v}\,.$$

Both $\mathbf{w}$ and $\mathbf{w}'$ have integer coordinates, hence so does $\mathbf{w}''$. Also, $\mathbf{w}''$ belongs to the parallelogram spanned by $\mathbf{u}$ and $\mathbf{v}$ but can not be equal to $\mathbf{u}$, $\mathbf{v}$ or $\mathbf{u} + \mathbf{v}$. Thus $\mathbf{w}'' = 0$, therefore $\lambda, \mu \in \mathbb{Z}$ and (i) follows. □

PROOF OF PROPOSITION 2.2. Put $\mathbf{u} = (m, n)$ and $\mathbf{v} = (p, q)$. Clearly $\mathbf{u} \neq \mathbf{v}$, and since both m, n and p, q are coprime, it follows that $\mathbf{u}$ and $\mathbf{v}$ are not proportional. It is also easy to see that the closed triangle with vertices $0, \mathbf{u}, \mathbf{v}$ can not contain any integer points except for its vertices (just think in terms of Remark 2.1.1), and an application of Lemma 2.4(iii)⇒(ii) finishes the proof. □

2.2. From 1 to 1/2 to $1/\sqrt{5}$. As another corollary of Proposition 2.2 we get the following result:

PROPOSITION 2.5. *Let $\frac{m}{n} < \frac{p}{q}$ be two successive elements of $\mathcal{F}_N$ for some $N > 1$, and let $\frac{m}{n} \le \alpha \le \frac{p}{q}$. Then*

$$\textit{either} \quad \left|\alpha - \frac{m}{n}\right| < \frac{1}{2n^2} \quad \textit{or} \quad \left|\alpha - \frac{p}{q}\right| < \frac{1}{2q^2}\,. \tag{2.1}$$

PROOF. Clearly $n \neq q$, therefore

$$\begin{aligned} &\frac{1}{2n^2} + \frac{1}{2q^2} > \frac{1}{nq} \\ \text{(Prop. 2.2)} \qquad &= \frac{np - mq}{nq} = \frac{p}{q} - \frac{m}{n} = \left(\frac{p}{q} - \alpha\right) + \left(\alpha - \frac{m}{n}\right)\,, \end{aligned}$$

which immediately implies (2.1). □

COROLLARY 2.6. *$\mathcal{W}_{1/2,1} = \mathbb{R}$; that is, for any $\alpha \in \mathbb{R}$ there exist infinitely many $q \in \mathbb{N}$ with $\langle \alpha q \rangle < 1/2q$.*

The proof is straightforward and is left to the reader, see Exercise 2.8.

However, the constant 1/2 is still not sharp. In order to produce a sharp result, we need to consider triples of successive elements of Farey series.

PROPOSITION 2.7. *Let $\frac{k}{l} < \frac{p}{q}$ be two successive elements of $\mathcal{F}_N$ for some $N > 1$, and put $m = k + p, n = l + q$. (Note that $\frac{m}{n}$ does not belong to $\mathcal{F}_N$;*

however see Exercise 2.5.) Then for every $\alpha \in [\frac{k}{l}, \frac{p}{q}]$, at least one of the following three inequalities holds:

$$\left|\alpha - \frac{k}{l}\right| < \frac{1}{\sqrt{5}l^2} \quad or \quad \left|\alpha - \frac{m}{n}\right| < \frac{1}{\sqrt{5}n^2} \quad or \quad \left|\alpha - \frac{p}{q}\right| < \frac{1}{\sqrt{5}q^2}\,. \tag{2.2}$$

Proof. Without loss of generality we may assume that $\alpha > \frac{m}{n}$ (otherwise replacing α by $1-\alpha$, $\frac{k}{l}$ by $1-\frac{p}{q}$ etc.) If none of the inequalities (2.2) hold, then

$$\alpha - \frac{k}{l} \geq \frac{1}{\sqrt{5}l^2}\,, \quad \alpha - \frac{m}{n} \geq \frac{1}{\sqrt{5}n^2}\,, \quad \frac{p}{q} - \alpha \geq \frac{1}{\sqrt{5}q^2}\,.$$

Adding the first and the third inequalities and using Proposition 2.2, we get

$$\frac{p}{q} - \frac{k}{l} = \frac{l}{q} \geq \frac{1}{\sqrt{5}}\left(\frac{1}{l^2} + \frac{1}{q^2}\right) \quad \Rightarrow \quad \sqrt{5}lq \geq l^2 + q^2\,; \tag{2.3}$$

and adding the second and the third inequalities yields

$$\frac{p}{q} - \frac{m}{n} = \frac{n}{q} \geq \frac{1}{\sqrt{5}}\left(\frac{1}{n^2} + \frac{1}{q^2}\right) \quad \Rightarrow \quad \sqrt{5}nq \geq n^2 + q^2\,. \tag{2.4}$$

Now let us add (2.3) and (2.4), and then express the result in terms of l and q only:

$$\begin{aligned} 0 &\geq l^2 + n^2 + 2q^2 - \sqrt{5}(l+n)q = l^2 + (l+q)^2 + 2q^2 - \sqrt{5}(2l+q)q \\ &= 2l^2 - 2(\sqrt{5}-1)lq + (3-\sqrt{5})q^2 = 2\big(l - \tfrac{\sqrt{5}-1}{2}q\big)^2\,. \end{aligned}$$

But this is impossible, since $\frac{\sqrt{5}-1}{2}$ is irrational. □

Corollary 2.8. *$\mathcal{W}_{1/\sqrt{5},1} = \mathbb{R}$; that is, for any $\alpha \in \mathbb{R}$ there exist infinitely many $q \in \mathbb{N}$ with $\langle \alpha q \rangle < 1/\sqrt{5}q$.*

The proof is also omitted, see Exercise 2.8.

In the next section we will exhibit real numbers α which are not in $\mathcal{W}_{c,1}$ for any $c < 1/\sqrt{5}$, thus showing that the estimate in Corollary 2.8 can not be improved further.

2.3. Exercises.

Exercise 2.1. Compute the number of elements of $\mathcal{F}_N$.

Exercise 2.2. Prove that the above number is asymptotic to $3N^2/\pi^2$ as $N \to \infty$. (Hint: look at Lemma 4.9 and Exercise 4.7.)

Exercise 2.3. For a rational number $\frac{p}{q}$, define the *Ford circle* $\mathcal{C}(\frac{p}{q})$ of $\frac{p}{q}$ to be the circle in the plane having center $(\frac{p}{q}, \frac{1}{2q^2})$ and radius $\frac{1}{2q^2}$.

(a) Prove that the intersection of the interiors of any two distinct Ford circles is empty.
(b) Prove that two Ford circles $\mathcal{C}(\frac{m}{n})$ and $\mathcal{C}(\frac{p}{q})$ are tangent if and only if $\frac{m}{n}, \frac{p}{q}$ are two adjacent elements of $\mathcal{F}_N$ for some N.

(c) Suppose that the Ford circles $\mathcal{C}(\frac{m}{n})$ and $\mathcal{C}(\frac{p}{q})$ are tangent to each other, and let $\mathcal{C}$ be the circle tangent to $\mathcal{C}(\frac{m}{n})$, $\mathcal{C}(\frac{p}{q})$ and the x-axis. Then $\mathcal{C} = \mathcal{C}(\frac{m+p}{n+q})$.

EXERCISE 2.4. Prove the converse to Proposition 2.2: If $0 \le \frac{m}{n}, \frac{p}{q} \le 1$ are such that $np - mq = 1$, then they are two successive elements of $\mathcal{F}_N$ for some N.

EXERCISE 2.5. Prove the following converse to Corollary 2.3: if $0 \le \frac{k}{l} < \frac{p}{q} \le 1$ are two successive elements of $\mathcal{F}_N$ for some N, then $\frac{k}{l}, \frac{k+p}{l+q}, \frac{p}{q}$ are three successive elements of $\mathcal{F}_M$ for some $M > N$.

EXERCISE 2.6. State and prove a version of Lemma 2.4 with $\mathbb{Z}^2$ replaced by an arbitrary lattice Λ in $\mathbb{R}^2$.

EXERCISE 2.7. Prove the following generalization of Lemma 2.4(iii)⇒(ii), known as Pick's Theorem: if P is a closed polygon whose vertices have integer coordinates, then

$$\operatorname{area}(P) = \#(P \cap \mathbb{Z}^2) - \tfrac{1}{2}\#(\partial P \cap \mathbb{Z}^2) - 1\,.$$

EXERCISE 2.8. Deduce Corollary 2.6 from Proposition 2.5 and Corollary 2.8 from Proposition 2.7.

2.4. Inhomogeneous approximation: further questions. Suppose one again wants to fix $\beta \in \mathbb{R}$ and approximate α by 'pseudo-rational' numbers of the form (1.9). Is the approach of this section at all helpful? is there anything one can prove using β-Farey series? Maybe not for all, but for some β with certain approximation properties?

3. Lower estimates

The goal of this section is to discuss real numbers α for which one can prove some 'negative' results, namely that they cannot be approximated by rational numbers with a given precision.

3.1. Quadratic irrationals. Those since a long time ago are known to be quite 'far from rational numbers'. Here is a quantitative way to express this:

LEMMA 3.1. *Suppose $\alpha \in \mathbb{R}$ is a root of a nonzero irreducible quadratic integer polynomial P with discriminant D. Then for any $c < \sqrt{D}$ the inequality*

$$|\alpha - p/q| < c/q^2 \tag{3.1}$$

(equivalently, $\langle \alpha q \rangle < c/q$) has only finitely many solutions.

PROOF. Write $P(x) = a(x-\alpha)(x-\beta)$ and note that $D = a^2(\alpha-\beta)^2$. Take p/q satisfying (3.1); certaingly $P(p/q) \neq 0$, therefore $|P(p/q)| \geq 1/q^2$. Thus one has

$$\begin{aligned} \frac{1}{q^2} &\leq \left|P\big(\frac{p}{q}\big)\right| = \left|\frac{p}{q} - \alpha\right| \cdot \left|a\big(\frac{p}{q} - \beta\big)\right| \\ &\underset{(1.7)}{<} \frac{c}{q^2}\left|a\big(\alpha - \beta + \frac{p}{q} - \alpha\big)\right| < \frac{c\sqrt{D}}{q^2} + \frac{|a|c^2}{q^4}\,, \end{aligned}$$

which clearly is impossible if $c < \sqrt{D}$ and q is large enough. □

It is not hard to see that irreducible quadratic integer polynomials cannot have discriminant smaller than 5. Howewer 5 can be attained, as is seen by taking $\alpha = \frac{\sqrt{5}-1}{2}$, the golden ratio. Thus we have proved that the complement of $\mathcal{W}_{c,1}$ is non-empty for any $c < \sqrt{5}$.

3.2. Markov constants and badly approximable numbers. More generally, let us define the *Markov constant* $\mu(\alpha)$ of $\alpha \in \mathbb{R}$ by

$$\mu(\alpha) \stackrel{\text{def}}{=} \inf\{c \mid \alpha \in \mathcal{W}_{c,1}\} = \liminf_{q\to\infty} q\langle \alpha q \rangle\,. \tag{3.2}$$

One says that α is *badly approximable* if $\mu(\alpha) > 0$, and *well approximable* otherwise. (In the notation of Exercise 1.4, the set of well approximable numbers coincides with $\check{\mathcal{W}}_{1,1}$; see also Exercise 3.4 for equivalent definitions.) With this terminology, let us summarize what we have proved so far: $\mu(\alpha) \geq 1/\sqrt{5}$ for all α; there exist at least countably many α with $\mu(\alpha) = 1/\sqrt{5}$; quadratic irrational numbers are badly approximable.

The set $\{\mu(\alpha) \mid \alpha \in \mathbb{R}\}$ of all possible values of Markov constants is called the *Markov spectrum*. Further results in this direction depend heavily on more sophisticated tools, such as contunied fractions. Here is a short list of facts:

- the set of badly approximable numbers has Lebesgue measure zero but Hausdorff dimension one (in particular, there are uncountably many of them) (this we shall prove shortly);
- the ten largest values in the Markov spectrum are

$$\begin{aligned} 1/\sqrt{5} &= 0.4472135\dots \\ 1/\sqrt{8} &= 0.3535533\dots \\ 5/\sqrt{221} &= 0.3363363\dots \\ 13/\sqrt{1517} &= 0.3337725\dots \\ 29/\sqrt{7565} &= 0.3334214\dots \\ 17/\sqrt{2600} &= 0.3333974\dots \\ 89/\sqrt{71285} &= 0.3333426\dots \\ 169/\sqrt{257045} &= 0.3333359\dots \\ 97/\sqrt{84680} &= 0.3333353\dots \\ 233/\sqrt{488597} &= 0.3333346\dots \end{aligned}$$

- as suggested by the above data, $1/3$ is the smallest accumulation points of values of Markov constants (in particular, there are only countably many values $> 1/3$), and all the values above $1/3$ are quadratic irrationals;
- below $1/3$ there are uncountably many values with various gaps, and the leftmost gap ends at *Freiman's number*

$$\theta = \frac{153640040533216 - 19623586058\sqrt{462}}{693746111282512} = 0.220856369\dots$$

  The existence of such a gap was proved by M. Hall in 1947, and thus the interval $[0, \theta]$, consisting entirely of the points in the Markov spectrum, is therefore called *Hall's ray*. The exact value of θ was computed by Russian mathematician G. Freiman in 1975.

3.3. Higher degree algebraic numbers. The main idea of the proof of Lemma 3.1 is is perfectly applicable to algebraic numbers of degree higher than 2. The next theorem is due to Liouville (1844), who used it to construct the first explicit examples of transcendental numbers.

THEOREM 3.2. *Suppose $\alpha \in \mathbb{R}$ is a root of a nonzero irreducible integer polynomial P of degree $n \geq 2$. Then for any $c < 1/|P'(\alpha)|$ the inequality*

$$|\alpha - p/q| < c/q^n \tag{3.3}$$

(equivalently, $\langle \alpha q \rangle < c/q^{n-1}$) has only finitely many solutions.

PROOF. Take any p/q satisfying (3.3); similarly to the proof of Lemma 3.1, one sees that $|P(p/q)| \geq 1/q^n$. Then, using the Mean Value Theorem, one can write

$$\frac{1}{q^n} \le \left|P\left(\frac{p}{q}\right)\right| = \left|P\left(\frac{p}{q}\right) - P(\alpha)\right| = \left|\alpha - \frac{p}{q}\right| \cdot |P'(t)| \underset{(3.3)}{<} \frac{c}{q^n}|P'(t)|\,,$$

where t is some number between α and p/q. Taking q large enough to ensure that $|P'(t)| \le 1/c$ produces a contradiction. $\square$

Thus we have proved that for any α and n as in Theorem 3.2 there exists $c > 0$ such that $\alpha \notin \mathcal{W}_{c,n-1}$.

3.4. Diophantine exponents and Liouville numbers. It will be convenient to define the *Diophantine exponent* $\omega(\alpha)$ of a real number α as follows:

$$\omega(\alpha) \stackrel{\text{def}}{=} \sup\{v \mid \alpha \in \mathcal{W}_{1,v}\} = \limsup_{q\to\infty} \frac{-\log\langle\alpha q\rangle}{\log q}\,. \tag{3.4}$$

See Exercise 3.7 for equivalent definitions. We know that $\omega(\alpha) \ge 1$ for any α (Corollary 1.6), and that $\omega(\alpha) \le n-1$ if α is algebraic of degree n (Theorem 3.2). Numbers α for which $\omega(\alpha) = \infty$ are called *Liouville* (see Exercise 3.8 for an equivalent definition). The conclusion is that Liouville irrational numbers are transcendental. Here is a concrete example:

COROLLARY 3.3. *The number*

$$\alpha = \sum_{k=1}^{\infty} \frac{1}{10^{k!}} \tag{3.5}$$

is transcendental.

PROOF. Since the decimal expansion of α is not eventually periodic, α is irrational. Take any integer $n \ge 2$ and take p/q to be equal to the nth partial sum of the series (3.5). Then $q = 1/10^{(n-1)!}$, and one has

$$\left|\alpha - \frac{p}{q}\right| = \sum_{k=n}^{\infty} \frac{1}{10^{k!}} \le \frac{2}{10^{n!}} = \frac{2}{q^n}\,.$$

Therefore α is Liouville, and hence, by Theorem 3.2, transcendental. $\square$

One immediately sees that $\omega(\alpha) = 1$ if α is badly approximable, but converse is not necessarily true, and in fact is very far from being true. One says that α is *very well approximable* if $\omega(\alpha) > 1$ (see Exercise 3.10 for an equivalent definition). We will prove later that the set of very well approximable numbers has measure zero (compare this with the full measure of well approximable numbers). One can also show that for any $v > 1$ the set $\{\alpha \mid \omega(\alpha) = v\}$ is non-empty. In fact, it is uncountable and has Hausdorff dimension $\frac{2}{v+1}$ (this was proved independently by Jarnik and Besicovitch in the 1930s).

It is an amazing (and very difficult!) theorem of Roth that Liouville's estimate '$\omega(\alpha) \le n-1$ if α is algebraic of degree n' can be drastically improved: in fact one has $\omega(\alpha) = 1$ for any algebraic α, that is, algebraic numbers happen to belong to the (full measure) set of not very well approximable numbers. See Schmidt's book for the history and proof of Roth's theorem.

3.5. Exercises.

EXERCISE 3.1. Let α be a root of an irreducible quadratic integer polynomial with discriminant bigger than 5. Show that $\mu(\alpha) \geq 1/2\sqrt{2}$.

EXERCISE 3.2. Describe explicitly all quadratic irrational α with $\mu(\alpha) = 1/\sqrt{5}$ (those in fact happen to be all $\alpha \in \mathbb{R}$ with $\mu(\alpha) = 1/\sqrt{5}$).

EXERCISE 3.3. Prove the equality in (3.2).

EXERCISE 3.4. Prove that α is well approximable $\iff$ $\forall\, c > 0$ there exists $q \in \mathbb{N}$ with $\langle \alpha q \rangle < c/q \iff \inf_q q\langle \alpha q \rangle = 0$.

EXERCISE 3.5. Compare the conclusion of Lemma 3.1 with that of the '$n = 2$' case of Theorem 3.2. Which one gives a stronger result?

EXERCISE 3.6. Prove Theorem 3.2, perhaps with a different bound for c, by modifying the proof of Lemma 3.1. Which of the proofs gives a better bound?

EXERCISE 3.7. Prove the equality in (3.4). Also show that

$$\begin{aligned}\omega(\alpha) &= \sup\{v \mid \alpha \in \hat{\mathcal{W}}_{1,v}\} = \sup\{v \mid \alpha \in \check{\mathcal{W}}_{1,v}\} \\ &= \inf\{v \mid \alpha \notin \mathcal{W}_{1,v}\} = \inf\{v \mid \alpha \notin \hat{\mathcal{W}}_{1,v}\} = \inf\{v \mid \alpha \notin \check{\mathcal{W}}_{1,v}\}\,.\end{aligned}$$

EXERCISE 3.8. Prove that α is Liouville if and only if for every $v > 1$ there exists $q \in \mathbb{N}$ with $\langle \alpha q \rangle < 1/q^v$.

EXERCISE 3.9. Modify the proof of Corollary 3.3 to conclude that the set of Liouville numbers is uncountable. Or, more generally, show that $\mathcal{W}(\psi)$ is uncountable for any positive ψ, and explain how it implies the previous assertion.

EXERCISE 3.10. Prove that α is very well approximable if and only if $\inf_q q^v\langle \alpha q \rangle = 0$ for some $v > 1$.

3.6. Inhomogeneous approximation: continuing the theme. It is known that there exists an inhomogeneous analogue of the notion of badly approximable numbers, that is, satisfying $\inf_q q\langle q\alpha + \beta \rangle > 0$. Those α can be proved to have measure zero and full Hausdorff dimension for any fixed β (although the proofs are more difficult than the homogeneous one, and the full Hausdorff dimension result does not even seem to be in the literature). However I do not know of any explicit construction of such a pair for rationally independent α and β, or of any connection between algebraicity and inhomogeneous approximation...

4. Khintchine's Theorem

Our goal in this and several subsequent sections is to study the sets of ψ-approximable numbers from the point of view of measure theory. Namely, we denote by λ the standard Lebesgue measure on $\mathbb{R}$, and attempt to compute $\lambda\big(\mathcal{W}(\psi)\big)$ for any given ψ. In fact, we will prove that only two answers are possible: either $\mathcal{W}(\psi)$ is *null* (that is, has zero Lebesgue measure) or *conull* (that is, the measure of its complement is zero). Moreover, under an additional assumption of the monotonicity of ψ we will establish an explicit criterion, due to Khintchine, for either of the alternatives to hold.

4.1. $\mathcal{W}(\psi)$ as a lim-sup set. First let us look at the way the sets $\mathcal{W}(\psi)$ are defined, and discuss simiilarly defined sets in a more general (set-theoretic and measure-theoretic) context. Suppose $\{A_n \mid n \in \mathbb{N}\}$ is a sequence of subsets of some set X. One usually defines the *lim-sup set* of $\{A_n\}$ to be

$$\limsup_n A_n \stackrel{\text{def}}{=} \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} A_n\,. \tag{4.1}$$

Equivalently, $\limsup_n A_n$ is the set of points $x \in X$ such that $x \in A_n$ for infinitely many n. (See Exercise 4.1 for a justification of the terminology.)

It is clear that one has write $\mathcal{W}(\psi)$ as a lim-sup set: if for $q \in \mathbb{N}$ and $\varepsilon > 0$ one defines

$$A(q,\varepsilon) \stackrel{\text{def}}{=} \{\alpha \in \mathbb{R} \mid \langle q\alpha\rangle < \varepsilon\}\,,$$

then one has

$$\mathcal{W}(\psi) = \limsup_q A\big(q, \psi(q)\big)\,. \tag{4.2}$$

Note that the set $A(q,\varepsilon)$ has a very transparent structure: it is simply the union of intervals

$$\big\{\alpha \in \mathbb{R} \bigm| |q\alpha - p| < \varepsilon\big\} = \left(\frac{p}{q} - \frac{\varepsilon}{q}, \frac{p}{q} + \frac{\varepsilon}{q}\right) \tag{4.3}$$

over all $p \in \mathbb{Z}$.

Similarly one can write

$$\mathcal{W}'(\psi) = \limsup_q A'\big(q, \psi(q)\big)\,, \tag{4.4}$$

where

$$\begin{aligned} A'(q,\varepsilon) &\stackrel{\text{def}}{=} \big\{\alpha \in \mathbb{R} \bigm| |q\alpha - p| < \varepsilon \text{ for some } p \in \mathbb{Z} \text{ with } \gcd(p,q) = 1\big\} \\ &= \bigcup_{\gcd(p,q)=1} \left(\frac{p}{q} - \frac{\varepsilon}{q}, \frac{p}{q} + \frac{\varepsilon}{q}\right)\,. \end{aligned} \tag{4.5}$$

As shown by Lemma 1.3, this set coincides with $\mathcal{W}(\psi)$ modulo the countable set of rationals if ψ is assumed to be non-increasing.

4.2. The Borel-Cantelli lemma and its applications. Now let us suppose that the set X carries a measure μ. More precisely, we will be talking about a *measure space* $(X, \mathcal{B}, \mu)$, where X is a set, $\mathcal{B}$ a σ-algebra of subsets of X, and μ a positive measure (that is, a countably additive function) defined on $\mathcal{B}$. The standard reference as far as measure theory is concerned will be [Rudin, Real and Complex Analysis]. Very often we will be making some simplifying assumptions, such as assuming μ to be *σ-finite* (this means that X admits a decomposition as a countable union of sets of finite measure) or *finite* ($\mu(X) < \infty$) or a *probability measure* ($\mu(X) = 1$). Naturally, in most of the applications considered below $X = \mathbb{R}$, $\mathcal{B} = \{\text{Borel subsets of } \mathbb{R}\}$, and $\mu = \lambda$. Furthermore, we have already seen (Remark 1.1.5) that all the problems concerning ψ-approximable numbers α can be reduced to $\alpha \in [0,1]$. Thus it will be natural to take $X = [0,1]$ or $S^1 = \mathbb{R}/\mathbb{Z}$, in which case λ happens to be a probability measure.

However the next statement is completely general and needs no additional assumptions.

LEMMA 4.1. *Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $\{A_n\} \subset \mathcal{B}$ be such that*

$$\sum_n \mu(A_n) < \infty \,. \tag{4.6}$$

Then $\mu(\limsup_n A_n) = 0$.

This is the easier half of the classical Borel-Cantelli Lemma (actually due to Cantelli).

PROOF. Simply write

$$\mu\big(\limsup_n A_n\big) = \mu\Big(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} A_n\Big) \le \mu\Big(\bigcup_{n=N}^{\infty} A_n\Big) \le \sum_{n=N}^{\infty} \mu(A_n)\,,$$

and notice that the right hand side tends to 0 as $N \to \infty$ in view of (4.6). □

Let us now see how the above lemma can produce a condition ensuring that $\mathcal{W}(\psi)$ is a null set. Clearly it suffices to establish that $\mathcal{W}(\psi) \cap [0,1]$ is null. Note that $A(q,\varepsilon) \cap [0,1]$ is the union of intervals (4.3) for $p = 1, \dots, q-1$ and two 'half-intervals' $\big(0, \frac{\varepsilon}{q}\big)$ and $\big(1 - \frac{\varepsilon}{q}, 1\big)$. Therefore

$$\lambda\big(A(q,\varepsilon) \cap [0,1]\big) \le 2\frac{\varepsilon}{q} + (q-1)\frac{2\varepsilon}{q} = 2\varepsilon\,.$$

Using (4.2) and applying Lemma 4.1 with $X = [0,1]$ and $\mu = \lambda$, one obtains

COROLLARY 4.2. *Let $\psi : \mathbb{N} \to \mathbb{R}_+$ be such that the series*

$$\sum_q \psi(q) \tag{4.7}$$

converges. Then $\lambda\big(\mathcal{W}(\psi)\big) = 0$.

The above statement is usually referred to as the convergence case of Khintchine's Theorem.

COROLLARY 4.3. *$\lambda(\mathcal{W}_{c,v}) = 0$ for any $c > 0$ and $v > 1$. Consequently, $\omega(\alpha) = 1$ for λ-a.e. α, that is, almost all real numbers are not very well approximable.*

PROOF. For the first assertion, simply notice that the function $\psi(x) = \frac{c}{x^v}$ satisfies (4.7) whenever $v > 1$. The second one follows since the set of very well approximable numbers can be written as a union of countably many null sets $\mathcal{W}_{1,v}$ where $v = 1 + 1/n$, $n \in \mathbb{N}$. □

Looking at our proof of Corollary 4.2, one might wonder if perhaps a stronger result can be proved if for every p/q one takes just one interval containing it, that is, restricts attention to relatively prime (p, q). Indeed, we can apply the same argument to the set (4.4), noticing that $A'(q, \varepsilon) \cap [0, 1]$ is the union of intervals (4.3) over all positive $p < q$ with $\gcd(p, q) = 1$. Therefore

$$\lambda\big(A'(q, \varepsilon) \cap [0, 1]\big) \le \varphi(q)\frac{2\varepsilon}{q}\,,$$

where φ is the *Euler function*, namely $\varphi(1) = 1$ and

$$\varphi(q) \stackrel{\text{def}}{=} \#\{1 \le p < q \mid \gcd(p, q) = 1\}$$

for $q > 1$. We arrive at the following conclusion:

COROLLARY 4.4. *Let $\psi : \mathbb{N} \to \mathbb{R}_+$ be such that the series*

$$\sum_q \frac{\varphi(q)}{q}\psi(q) \tag{4.8}$$

converges. Then $\lambda\big(\mathcal{W}'(\psi)\big) = 0$. In particular, (4.8) *and the monotonicity of ψ imply that $\lambda\big(\mathcal{W}(\psi)\big) = 0$.*

Note that the convergence of (4.8) is apriori weaker than that of (4.7), since $\varphi(q)$ is always less than q. So a natural question arises as to whether Corollary 4.4 provides more information than Corollary 4.2 in the case of a non-increasing ψ. The answer, however, is negative: for monotonic functions ψ the series (4.7) and (4.8) always converge or diverge simultaneously! (In fact, one can express this equivalence in a quantitative way, which we will do in the next subsection.)

Even more important is the fact that the convergence of these series is necessary for the conclusion of Corollaries 4.2 and 4.4, again assuming the monotonicity of ψ. Namely, the following (the divergence case of Khintchine's Theorem) holds:

THEOREM 4.5. *Let $\psi : \mathbb{N} \to \mathbb{R}_+$ be a non-increasing function such that the series* (4.7) *diverges. Then $\lambda\big(\mathcal{W}(\psi)^c\big) = 0$.*

We will complete the proof of this theorem in §6. Note that the monotonicity assumption is important: there are examples of functions ψ for which the series (4.7) converges but still $\lambda\big(\mathcal{W}(\psi)^c\big) = 0$. These examples were found in the 1940s by Duffin and Schaeffer, who proposed a (still open)

CONJECTURE 4.6. *Let $\psi : \mathbb{N} \to \mathbb{R}_+$ be an arbitrary function such that the series* (4.8) *diverges. Then* $\lambda\big(\mathcal{W}'(\psi)^c\big) = 0$.

Note also that the combination of Corollary 4.2 and Theorem 4.5 shows that the sets $\mathcal{W}(\psi)$ with non-increasing ψ are always either null or conull. This dichotomy, which in fact is true without assuming monotonicity, will be one of the steps in our proof of Theorem 4.5.

4.3. Comparing (4.7) **and** (4.8)**.** Now let us return to the discussion of the two convergence conditions. Our goal here is to prove the following

PROPOSITION 4.7. *There exists $c > 0$ such that for any positive non-increasing function ψ one has*

$$\forall\, N \in \mathbb{N}\,, \quad \sum_{q=1}^{N} \frac{\varphi(q)}{q}\psi(q) > c\sum_{q=1}^{N}\psi(q)\,. \tag{4.9}$$

In other words, assuming the monotonicity of ψ, not only the convergence/divergence of (4.7) and (4.8) are equivalent, but, furthermore, the ratio of the partial sums of the corresponding series is bounded between two uniform constants. It follows that all the work done in the proof of Corollary 4.4 (throwing away intervals whose centers had already appeared) is not needed: no extra information was gained compared to Corollary 4.2. Note that the monotonicity of ψ can be replaced by a weaker assumption that the function $x \mapsto \frac{\psi(x)}{x^v}$ is non-increasing for some $v \geq 0$, see Exercise 4.9.

For the proof, we need to look more closely at the function φ. If $p_1, \dots, p_k$ are all the prime factors of q, using the inclusion-exclusion principle one can write (Exercise 4.3) that

$$\varphi(q) = q\left(1 - \sum_{1\le i\le k}\frac{1}{p_i} + \sum_{1\le i<j\le k}\frac{1}{p_ip_j} - \cdots + (-1)^k\frac{1}{p_1\cdots p_k}\right)\,. \tag{4.10}$$

Another way of saying this is the following: introduce the *Möbius function* $\mu(\cdot)$ by

$$\mu(q) = \begin{cases} 1 & \text{if } q = 1\,;\\ 0 & \text{if } q \text{ has a squared factor;}\\ (-1)^k & \text{if } q = p_1\cdots p_k \text{ and all the primes } p_1,\dots,p_k \text{ are different.}\end{cases}$$

Then (4.10) can be written as

$$\varphi(q) = q\sum_{n\,|\,q}\frac{\mu(n)}{n}\,. \tag{4.11}$$

A nice fact about the Möbius function is that its generating function is easily expressed via the Riemann ζ-function $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$:

LEMMA 4.8. *For any $s > 1$ one has*

$$\sum_{n=1}^{\infty} \frac{\mu(n)}{n^s} = \frac{1}{\zeta(s)}\,. \tag{4.12}$$

PROOF. Using the fact that

$$\zeta(s) = \prod_{p \text{ prime}} \frac{1}{1-p^{-s}} \tag{4.13}$$

(Euler's product formula, Exercise 4.6), one can write

$$\begin{aligned}\frac{1}{\zeta(s)} &= \prod_{p \text{ prime}} (1-p^{-s}) = \sum_{k=1}^{\infty} \sum_{p_1<\cdots<p_k \text{ prime}} \mu(p_1\cdots p_k)(p_1\cdots p_k)^{-s} \\ &= \sum_{k=1}^{\infty} \sum_{p_1<\cdots<p_k \text{ prime},\, i_1,\ldots,i_k\in\mathbb{N}} \frac{\mu(p_1^{i_1}\cdots p_k^{i_k})}{(p_1^{i_1}\cdots p_k^{i_k})^{-s}} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}\,.\end{aligned}$$

□

Now we have gathered enough information to treat the case $\psi \equiv 1$ of Proposition 4.7. Namely, we have

LEMMA 4.9. $\frac{1}{N}\sum_{q=1}^{N} \frac{\varphi(q)}{q} \to \frac{6}{\pi^2}$ *as $N \to \infty$.*

Thus, roughly speaking, "on average" $\varphi(q)$ can be bounded from below by a constant times q.

PROOF. Using (4.11), write

$$\begin{aligned}\frac{1}{N}\sum_{q=1}^{N} \frac{\varphi(q)}{q} &= \frac{1}{N}\sum_{q=1}^{N}\sum_{n\,|\,q} \frac{\mu(n)}{n} = \frac{1}{N}\sum_{n=1}^{N} \frac{\mu(n)}{n}[N/n] \\ &= \sum_{n=1}^{N} \frac{\mu(n)}{n^2} + O(1/N) \to \sum_{n=1}^{\infty} \frac{\mu(n)}{n^2} \underset{(4.12)}{=} \frac{1}{\zeta(2)} = \frac{6}{\pi^2}\,.\end{aligned}$$

□

Finally, we proceed with the

PROOF OF PROPOSITION 4.7. Let us denote $\sum_{q=1}^{N} \frac{\varphi(q)}{q}$ by σ_N, with $\sigma_0 = 0$. Note that it follows from Lemma 4.9 that there exists a constant $c > 0$ such that

$$\sigma_N \geq cN \quad \text{for all } N \in \mathbb{N}\,. \tag{4.14}$$

Now rewrite the left hand side of (4.9) using summation by parts as follows:

$$\begin{aligned}\sum_{q=1}^{N}\frac{\varphi(q)}{q}\psi(q) &= \sum_{q=1}^{N}\psi(q)(\sigma_q-\sigma_{q-1}) = \sum_{q=1}^{N}\psi(q)\sigma_q - \sum_{q=1}^{N-1}\psi(q+1)\sigma_q \\ &= \psi(N)\sigma_N + \sum_{q=1}^{N-1}\sigma_q\big(\psi(q)-\psi(q+1)\big) \\ &\underset{\text{(4.14) and the monotonicity of }\psi}{\geq} c\left(N\psi(N)+\sum_{q=1}^{N-1}q\big(\psi(q)-\psi(q+1)\big)\right) = c\sum_{q=1}^{N}\psi(q)\,.\end{aligned}$$

□

4.4. Proving Khintchine's Theorem. Our proof of Theorem 4.5 will be based on the following two results, proved in the subsequent sections:

THEOREM 4.10 (Zero-one law). *For any function $\psi:\mathbb{N}\to\mathbb{R}_+$, either $\lambda\big(\mathcal{W}(\psi)\big)=0$ or $\lambda\big(\mathcal{W}(\psi)^c\big)=0$.*

THEOREM 4.11 (Duffin-Schaeffer Theorem). *Let $\psi:\mathbb{N}\to(0,1/2)$[1] be a function such that the series* (4.8) *diverges, and suppose that* (4.9) *is satisfied. Then $\lambda\big(\mathcal{W}'(\psi)\cap[0,1]\big)\geq c^2/2$.*

It is easy to see that Theorem 4.5 follows from the two theorems above: if ψ is non-increasing, it satisfies (4.9) for some $c>0$ due to Proposition 4.7, and also without loss of generality one can assume that $\psi(q)<1/2$ for all q. Therefore $\lambda\big(\mathcal{W}(\psi)\big)\geq\lambda\big(\mathcal{W}(\psi)\cap[0,1]\big)\geq\lambda\big(\mathcal{W}'(\psi)\cap[0,1]\big)\geq c^2/2>0$ by Theorem 4.11, which implies that $\lambda\big(\mathcal{W}(\psi)^c\big)=0$ in view of Theorem 4.10.

4.5. Exercises.

EXERCISE 4.1. Explain the analogy between (4.1) and the definition of $\limsup_n a_n$ where a_n is a sequence of real numbers.

EXERCISE 4.2. Invent a definition for $\liminf_n A_n$ and explain its meaning. Then find general sufficient conditions for $\liminf_n A_n$ to be either null or conull.

EXERCISE 4.3. Prove (4.10).

EXERCISE 4.4. Derive from (4.10) that

$$\varphi(p_1^{i_1}\cdots p_k^{i_k}) = (p_1^{i_1}-p_1^{i_1-1})\cdots(p_k^{i_k}-p_k^{i_k-1})\,.$$

EXERCISE 4.5. Prove the following *inversion formulas* for Euler and Möbius functions:

(a) $\sum_{n\,|\,q}\varphi(n)=q$ for every $q>1$;

[1]In fact the theorem is valid for an arbitrary function $\psi:\mathbb{N}\to\mathbb{R}_+$, but the proof is more complicated and is omitted since it is not needed for the proof of Theorem 4.5

(b) $\sum_{n\,|\,q} \mu(n) = \begin{cases} 1 \text{ if } q = 1 \\ 0 \text{ if } q > 1\,. \end{cases}$

Exercise 4.6. Prove (4.13).

Exercise 4.7. Modify the proof of Lemma 4.9 to find

(a) $\gamma_0 \stackrel{\text{def}}{=} \inf\{\gamma \in \mathbb{R} \mid \sum_{q=1}^{\infty} q^{\gamma}\varphi(q) = \infty\}$ (note that $\gamma_0 \le -1$ due to Lemma 4.9);

(b) the asymptotics of $\sum_{q=1}^{N} q^{\gamma}\varphi(q)$ as $N \to \infty$ for any $\gamma \ge \gamma_0$ (note that Exercise 2.2 is a special case).

Exercise 4.8. The "summation by parts" trick in the proof of Proposition 4.7 is a special case of the following lemma due to Abel: consider two sequences $\{a_n\}$ and $\{b_n\}$, $n \in \mathbb{N}$, and let σ_N be the N-th partial sum of $\{a_n\}$; then for any $0 < M \le N$ one has

$$\sum_{n=M}^{N} a_n b_n = (\sigma_N b_N - \sigma_{M-1} b_M) + \sum_{n=M}^{N-1} \sigma_n (b_n - b_{n+1})\,.$$

Prove it.

Exercise 4.9. Show that the conclusion of Proposition 4.7, and hence of Theorem 4.5, holds if the monotonicity of ψ is replaced by a weaker assumption that the function $x \mapsto \frac{\psi(x)}{x^v}$ is non-increasing for some $v \ge 0$. (Hint: rewrite the left hand side of (4.9) as $\sum_{q=1}^{N} \left(q^{v-1}\varphi(q)\right)\frac{\psi(q)}{q^v}$, and apply Exercises 4.7 and 4.8.)

5. Zero-one laws in dynamics and number theory

In this section we describe situations when certain subsets of a measure space are forced to be either null or conull. An example is given by Theorem 4.10, which will be proved in this section. Such situations also often occur in the context of a measure-preserving dynamical system, and it is instructive to study both settings in comparison.

5.1. Ergodicity. By a *measure preserving system* we will mean a quadruple $(X, \mathcal{B}, \mu, T)$, where $(X, \mathcal{B}, \mu)$ is a measure space, and $T : X \to X$ is a *measurable* (that is, $T^{-1}(\mathcal{B}) \subset \mathcal{B}$) and *$\mu$-preserving* ($\mu\big(T^{-1}(A)\big) = \mu(A)$ for all $A \in \mathcal{B}$) self-map of X. Note that T does not have to be one-to-one.

In all the general definitions and statements below we do not require μ to be finite, although this will be the case in all the applications and exercises. Also, we are going to avoid mentioning $\mathcal{B}$, hoping that it will be always clear from the context, so that whenever a subset of X is mentioned it is always assumed to lie in the σ-algebra of measurable sets.

One example of a measure preserving system is given in §1.2.1, where we had $T = R_\alpha$ and $X = S^1$. Obviously Lebesgue measure is preserved by R_α for any rotation angle α. For another example, take $m \in \mathbb{N}$ and consider $T_m : S^1 \to S^1$ given by

$$T_m(x) = m\,x \mod 1\,. \tag{5.1}$$

This map is not one-to-one, yet one can see that the preimage of any interval consists of m intervals, each m times shorter than the original one, hence Lebesgue measure is preserved here as well.

A useful way to look at a measure preserving system (X, μ, T) is to consider the induced action on the space $\mathcal{F}(X)$ of all measurable functions $X \to \mathbb{R}$, or on its subspaces such as $L^p(X, \mu)$, $1 \le p \le \infty$. (Here and hereafter, unless it leads to a confusion, we will identify functions with their equivalence classes modulo null sets.) Namely, we define $T_* : \mathcal{F}(X) \to \mathcal{F}(X)$ by $T_* f \stackrel{\text{def}}{=} f \circ T$. See Exercise 5.2 for a characterization of T being μ-preserving in terms of T_*.

In the study of measure preserving systems it is natural to look for the simplest, or "irreducible" ones, i.e. those which do not contain nontrivial subsystems. These are called ergodic. More precisely, $Y \subset X$ is said to be *T-invariant* if $T^{-1}(Y) = Y$. If Y is such, one can consider the restriction of T onto Y and treat it as a subsystem of (X, μ, T). The system (X, μ, T) is said to be *ergodic* if any T-invariant subset of X is either null or conull; in other words, if all its subsystems are trivial.

Equivalently, one can define ergodicity in terms of T_*, see Exercise 5.3, and in terms of the collection of all T-invariant (that is, preserved by T) measures on X, see Exercise 5.4.

5.2. Examples of ergodic systems. Here we will consider the two examples mentioned in the preceding subsection.

Proposition 5.1. *(S^1, λ, R_α) is ergodic iff $\alpha \notin \mathbb{Q}$.*

Proof. Suppose that $\alpha \notin \mathbb{Q}$. We have already shown (see Exercise 1.7) that every R_α-orbit is dense. It is easy to deduce from there that

($*$) for any nonempty interval $I \subset S^1$ there exist $n_1 < \cdots < n_k$ such that all the translates $R_\alpha^{n_i}(I)$ are disjoint and $\lambda\big(\cup_i R_\alpha^{n_i}(I)\big) > 1 - 2\lambda(I)$.

Now take $A \subset S^1$ with $\lambda(A) > 0$. Then

($**$) for any positive ε one can find an interval I with $0 < \lambda(I) < \varepsilon$ and $\lambda(A \cap I) \geq (1-\varepsilon)\lambda(I)$.

We leave the proof of ($*$) and ($**$) to the reader (Exercise 5.5). If A is R_α-invariant, then for any $n \in \mathbb{N}$ and any interval I one has $R_\alpha^n(A \cap I) = A \cap R_\alpha^n(I)$. Thus it follows from ($*$) and ($**$) that

$$\lambda(A) \geq \lambda\big(A \cap \cup_i R_\alpha^{n_i}(I)\big) \geq (1-\varepsilon)\lambda\big(\cup_i R_\alpha^{n_i}(I)\big) \geq (1-\varepsilon)(1-2\varepsilon)\,.$$

Since ε was arbitrary, we conclude that $\lambda(A) = 1$. The non-ergodicity of R_α for $\alpha \in \mathbb{Q}$ is straightforward and is left to the reader (Exercise 5.6). □

It is interesting to compare the above result, i.e. the "measure-theoretic irreducibility" of R_α, with the "topological irreducibility" or *minimality* (density of every orbit). Surprisingly, there exist examples of λ-preserving maps of S^1 which are minimal but not ergodic. On the other hand, it is not true either that ergodicity implies minimality, as can be inferred from the next proposition. See however Exercise 5.7.

Proposition 5.2. *(S^1, λ, T_m) is ergodic for any integer $m > 1$.*

Proof. For simplicity we will present the proof for the case $m = 2$, the general case is similar. Suppose that A is T_2-invariant. Then the intersection of A with $[0, \frac{1}{2})$ translated to the right by $\frac{1}{2}$ must coincide with $A \cap [\frac{1}{2}, 1)$, since both intersections are in one-to-one correspondence with $T(A) = A$. It follows that the two intersections have the same Lebesgue measure. A generalization of this argument shows that for any dyadic interval I one has

$$\lambda(A \cap I) = \lambda(A)\lambda(I)\,. \tag{5.2}$$

In other words, the function $1_A - \lambda(A)$ is orthogonal (in L^2) to 1_I for any dyadic interval I. Since linear combinations of the latter are dense in $L^2(S^1, \lambda)$, it follows that $1_A(x) - \lambda(A) = 0$ for almost every x, hence $\lambda(A)$ is either 0 or 1. □

See Exercise 5.8 for an alternative proof of both propositions. Note that even though almost all orbits of the map T_m are dense (Exercise 5.7), there exist many points with exceptional (nondense) orbits. See e.g. Exercise 5.9.

5.3. Lebesgue's Density Theorem. A nice property of the Lebesgue measure λ on $\mathbb{R}$ (or more generally on $\mathbb{R}^n$) is a possibility to extend the Fundamental Theorem of Calculus to arbitrary integrable functions. Namely, the following theorem holds:

THEOREM 5.3. *If $f \in L^1(\mathbb{R}, \lambda)$ and*

$$F(x) = \int_{(-\infty,x)} f \, d\lambda \,,$$

then F is differentiable at λ-almost every point, and furthermore $F'(x) = f(x)$ λ-a.e.

See [Rudin, Chapter 7] for the proof.

Given a subset A of $\mathbb{R}$ and a point $x \in \mathbb{R}$, let us define the *metric density* of A at x to be equal to

$$d_x(A) \stackrel{\text{def}}{=} \lim_{\varepsilon \to 0} \frac{\lambda\big(A \cap (x-\varepsilon, x+\varepsilon)\big)}{2\varepsilon} \,,$$

provided the limit exists. Applying the above theorem to the characteristic function of A, one obtains the following corollary, known as Lebesgue's Density Theorem:

COROLLARY 5.4. *The metric density of any subset A of $\mathbb{R}$ is equal to 1 at λ-almost every point of A.*

See Exercise 5.11 for examples of exceptional points, and Exercise 5.12 for an application.

5.4. Back to limsup sets. Note that for applications of Lebesgue's Density Theorem it often suffices to find a single point x in a set A of positive Lebesgue measure with $d_x(A) = 1$. Here is an example, a lemma due to Gallagher (and implicitly to Cassels).

LEMMA 5.5. *Let $\{A_n\}$ be a sequence of intervals such that*

(i) $\lambda(A_n) \to 0$ *as* $n \to \infty$,

and let a sequence $\{B_n\}$ of measurable sets and $c > 0$ be such that for all $n \in \mathbb{N}$

(ii) $B_n \subset A_n$ *and* $\lambda(B_n) \geq c\lambda(A_n)$.

Then $\lambda(\limsup_n A_n) = \lambda(\limsup_n B_n)$.

PROOF. Clearly we can discard intervals of zero length and alter the outcome only by a set of measure zero; we may thus suppose that $\lambda(A_n)$ is positive for each n. Denote $\limsup_n A_n$ by A, and write

$$A \smallsetminus \limsup_n B_n = A \smallsetminus \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} B_n = \bigcup_{N=1}^{\infty} C_N \,,$$

where we put

$$C_N \stackrel{\text{def}}{=} A \smallsetminus \bigcup_{n=N}^{\infty} B_n \,.$$

It suffices to show that $\lambda(C_N) = 0$ for each N.

Suppose, on the contrary, that $\lambda(C_N) > 0$ for some N, and take $x \in C_N$ with $d_x(C_N) = 1$. Note that $x \in A$, therefore $x \in A_n$ for infinitely many n. It follows from (i) that there exist arbitrary large values of n for which

$$\lambda(A_n \cap C_N) \geq (1 - c/2)\lambda(A_n)\,.$$

But note that C_N is disjoint from B_n when $n \geq N$, and thus, in view of (ii), $\lambda(A_n \cap C_N)$ must be not bigger than $(1 - c)\lambda(A_n)$, a contradiction. □

Corollary 5.6. *For any positive function $\psi : \mathbb{N} \to \mathbb{R}$ and any $c > 0$, one has*

$$\lambda\big(\mathcal{W}(c\psi)\big) = \lambda\big(\mathcal{W}(\psi)\big)\,, \tag{5.3}$$

and hence also $\lambda\big(\mathcal{W}(\psi)\big) = \lambda\big(\hat{\mathcal{W}}(\psi)\big) = \lambda\big(\check{\mathcal{W}}(\psi)\big)$ (see Exercise 1.4 for notation).

Proof. Clearly it suffices to consider $c < 1$, otherwise replacing $c\psi$ by ψ and c by $1/c$, and restrict attention to $[0, 1]$. Also, if $\psi(q)/q$ does not tend to zero as $q \to \infty$, one can take a sequence $q_n \to \infty$ with $c\psi(q_n) > 1$, and conclude that both (1.3) and $\langle q\alpha\rangle < c\psi(q)$ have infinitely many solutions for every α, i.e. both sets in (5.3) have empty complements.

Now suppose that $\psi(q)/q \to 0$ as $q \to \infty$. Writing $n = n(q, p)$ for the lexicographical ordering of pairs $\{(q, p) \mid q \in \mathbb{N},\, 0 \leq p \leq q\}$ and using (4.2), one can write $\mathcal{W}(\psi) \cap [0, 1] = \limsup_n A_n$ and $\mathcal{W}(c\psi) \cap [0, 1] = \limsup_n B_n$, where $A_{n(q,p)} = \left(\frac{p}{q} - \frac{\psi(q)}{q}, \frac{p}{q} + \frac{\psi(q)}{q}\right)$ and $B_{n(q,p)} = \left(\frac{p}{q} - \frac{c\psi(q)}{q}, \frac{p}{q} + \frac{c\psi(q)}{q}\right)$. All the conditions of Lemma 5.5 are satisfied, and (5.3) follows. The second part of the lemma is immediate from the countable additivity of λ. □

As a corollary from the above, we can already derive the following special case of Theorem 4.5:

Corollary 5.7. *For every $c > 0$, the set $\mathcal{W}_{c,1}$ has full Lebesgue measure; that is, λ-almost every $\alpha \in \mathbb{R}$ is well approximable.*

Proof. According to Corollary 5.6, $\lambda(\mathcal{W}_{1,1}) = \lambda(\mathcal{W}_{c,1}) = \lambda(\check{\mathcal{W}}_{1,1})$. But $\mathcal{W}_{1,1} = \mathbb{R}$ (Corollary 1.6), while $\check{\mathcal{W}}_{1,1}$ is exactly the set of well approximable numbers. □

Another application of Corollary 5.6 is the

Proof of Theorem 4.10. It is easy to see (Exercise 1.4) that the set $\hat{\mathcal{W}}(\psi) \cap [0, 1]$ is invariant under multiplication by 2 modulo 1. Therefore, by Proposition 5.2, it is either null or conull, and Corollary 5.6 establishes the same for $\mathcal{W}(\psi) \cap [0, 1]$. □

5.5. Birkhoff's Ergodic Theorem. The next theorem is one of the most fundamental results in the theory of measure-preserving transformations. We will quote it without proof, and will use it later to extract some number-theoretic information from dynamics.

THEOREM 5.8. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system, with $\mu(X) < \infty$, and let $f \in L^1(X, \mathcal{B}, \mu)$. Then the limit function*

$$\bar{f}(x) \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f\big(T^k(x)\big) \tag{5.4}$$

is defined for μ-a.e. $x \in X$, belongs to $L^1(X, \mathcal{B}, \mu)$, is T-invariant, and satisfies $\int_X \bar{f}\, d\mu = \int_X f\, d\mu$.

COROLLARY 5.9. *The following are equivalent for a measure preserving transformation T on a probability space $(X, \mathcal{B}, \mu)$:*

(i) *T is ergodic;*
(ii) *for any $f \in L^1(X, \mathcal{B}, \mu)$ one has $\bar{f} = \int_X f\, d\mu$.*

PROOF. The implication (i)⇒(ii) is straightforward from Theorem 5.8 and Exercise 5.4. The converse is left as Exercise 5.13. ☐

5.6. Exercises. Unless otherwise noted, all measure spaces below are assumed to be finite.

EXERCISE 5.1. Prove that a measurable self-map T of a measure space (X, μ) is μ-preserving iff $\mu\big(T^{-1}(A)\big) \le \mu(A)$ for all $A \subset X$.

EXERCISE 5.2. Prove that T is μ-preserving iff T_* preserves the L^p-norm for some (⇔ for all) $1 \le p < \infty$; in particular, iff $T_* : L^2 \to L^2$ is unitary. What goes wrong if $p = \infty$?

EXERCISE 5.3. Say that $Y \subset X$ is *essentially T-invariant* if $\mu\big(Y \triangle T^{-1}(Y)\big) = 0$. Prove that the following are equivalent:

(i) (X, μ, T) is ergodic;
(ii) any essentially T-invariant subset of X is either null or conull;
(iii) any T_*-invariant $f \in \mathcal{F}(X)$, or, equivalently, $\in L^p(X, \mu)$ for some (⇔ for all) $1 \le p \le \infty$, is constant (that is, 1 is a simple eigenvalue of T_*);
(iv) any eigenvalue of $T_* : \mathcal{F}(X) \to \mathcal{F}(X)$ is simple.

EXERCISE 5.4. Prove that the following are equivalent:

(i) (X, μ, T) is ergodic;
(ii) any T-invariant measure on X which is absolutely continuous with respect to μ is proportional to μ;
(iii) if $\mu = c_1\mu_1 + c_2\mu_2$ where μ_1, μ_2 are T-invariant measures on X and $c_1, c_2 \in \mathbb{R}$, then μ_1 and μ_2 are proportional.

In particular, it follows that ergodic probability measures are precisely the extremal points of the set of all invariant probability measures.

EXERCISE 5.5. Justify assertions ($*$) and ($**$) made in the proof of Proposition 5.1 (for ($**$), do not use Lebesgue's Density Theorem).

EXERCISE 5.6. Prove the "only if" part of Proposition 5.1.

EXERCISE 5.7. Let X be a second countable topological space, $T : X \to X$ a continuous map, and μ a measure on X which is positive on open sets. Suppose that (X, μ, T) is ergodic. Prove that the orbit $\{T^n(x) \mid n \in \mathbb{N}\}$ of μ-a.e. $x \in X$ is dense in X. Do not use Corollary 5.9.

EXERCISE 5.8. Prove Propositions (a) 5.1 and (b) 5.2 by writing an invariant $f \in L^2(S^1, \lambda)$ as a Fourier series and using the uniqueness of Fourier decomposition.

EXERCISE 5.9. Let m be an integer greater than 1. Prove that $x \in S^1$ is rational iff the T_m-orbit of x is *eventually periodic* (that is,

$$T_m^{n+k}(x) = T_m^n(x) \tag{5.5}$$

for some $k \in \mathbb{N}$ and all large enough n). Identify $x \in S^1$ whose T_m-orbits are: *periodic* ((5.5) holds for some k and all $n \geq 0$), *eventually stabilizing* ((5.5) holds for $k = 1$ and all large enough n). Exhibit $x \in S^1$ whose orbit is neither eventually periodic nor dense (note that the set of such x has Hausdorff dimension 1), and describe the closure of $\{T_m^n(x) \mid n \in \mathbb{N}\}$.

EXERCISE 5.10. Derive Corollary 5.4 from Theorem 5.3.

EXERCISE 5.11. Given two numbers $0 \leq a \leq b \leq 1$, construct a subset A of $\mathbb{R}$ and a point x of A such that

$$\liminf_{\varepsilon \to 0} \frac{\lambda\big(A \cap (x - \varepsilon, x + \varepsilon)\big)}{2\varepsilon} = a \quad \text{and} \quad \limsup_{\varepsilon \to 0} \frac{\lambda\big(A \cap (x - \varepsilon, x + \varepsilon)\big)}{2\varepsilon} = b\,.$$

(This includes the case when $d_x(A)$ exists and takes a given value $a \in (0, 1)$.)

EXERCISE 5.12. Suppose that $A \subset \mathbb{R}$ has the following property: for some $c > 0$ and any interval $I \subset \mathbb{R}$ one has $\lambda(A \cap I) \geq c\lambda(I)$. Prove that λ-almost all $x \in \mathbb{R}$ belong to A.

EXERCISE 5.13. Prove the implication (ii)$\Leftarrow$(i) in Corollary 5.9.

EXERCISE 5.14. Let Let $(X, \mathcal{B}, \mu, T)$ be as in Corollary 5.9, and let f be a nonnegative measurable function on X with $\int_X f \, d\mu = \infty$. Show that $\bar{f}(x)$, defined as in (5.4), is infinite for μ-a.e. $x \in X$. [Hint: approximate f by bounded functions and use Monotone Convergence Theorem.]

6. Quasi-independent Borel-Cantelli

6.1. Independence. In order to complete the proof of Theorem 4.5, we need to examine situations when the assumption

$$\sum_n \mu(A_n) = \infty \tag{6.1}$$

on a sequence $\{A_n\}$ of subsets of a measure space (X, μ) forces the set (4.1) to have positive measure. It is clear that without any additional conditions the converse to Lemma 4.1 is false: the choice of

$$A_n = (0, \tfrac{1}{n}) \subset [0, 1] \tag{6.2}$$

makes the series $\sum_n \lambda(A_n)$ divergent but $\limsup_n A_n = \varnothing$.

In this section for convenience all the measure spaces will have total measure 1.

Just for the record, let us start with the original Borel-Cantelli Lemma, which assumes an exceptionally strong condition of independence of all the sets $\{A_n\}$. Namely, let us say that the collection $\{A_n\}$ is *independent* if for every choice of $n_1 < \cdots < n_k$ one has

$$\mu\Big(\bigcap_i A_{n_i}\Big) = \prod_i \mu(A_{n_i})\,. \tag{6.3}$$

The following was proved by Borel in the end of the 19th century:

LEMMA 6.1. *Let $\{A_n\}$ be an independent collection of subsets of a probability space (X, μ) such that* (6.1) *holds. Then* $\limsup_n A_n$ *is conull.*

PROOF. It follows from (6.1) that $\sum_{n=N}^{\infty} \mu(A_n) = \infty$ for any $N \in \mathbb{N}$. The latter implies that

$$\prod_{n=N}^{\infty} e^{-\mu(A_n)} = 0 \underset{\text{(cf. Exercise 6.1)}}{\Rightarrow} \prod_{n=N}^{\infty} \big(1 - \mu(A_n)\big) = \prod_{n=N}^{\infty} \mu(A_n^c) = 0\,.$$

Because of the independence assumption (missing details are provided in Exercise 6.2), we conclude that the set $\bigcap_{n=N}^{\infty} A_n^c$ has measure zero for every N, and hence

$$0 = \mu\Big(\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty} A_n^c\Big) = \mu\Big(\big(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty} A_n\big)^c\Big) = \mu\Big(\big(\limsup_n A_n\big)^c\Big)\,.$$

□

This lemma is hard to apply to metric number theory because the independence assumption is impossible to verify for sets arising in Diophantine approximation. However it plays a major role in probability theory and has been an important step in proving the so-called "laws of large numbers".

We are going to weaken the independence assumption in two steps, proving two lemmas similar to Lemma 6.1.

6.2. Pairwise independence. It became clear in 1920s or 1930s [need to search for exact references] that the independence assumption in Lemma 6.1 can be significantly relaxed, and at the same time the conlcusion can be made much stronger. Here is an example of such a statement.

LEMMA 6.2. *Let* $\{A_n\}$ *be a collection of subsets of a probability space* (X, μ) *such that* (6.1) *holds, and such that*

$$\mu(A_m \cap A_n) = \mu(A_m)\mu(A_n) \quad \forall\, m \neq n\,. \tag{6.4}$$

Then $\limsup_n A_n$ *is conull, and, furthermore,*

$$\lim_{N\to\infty} \frac{\#\{1 \le n \le N \mid x \in A_n\}}{\sum_{n=1}^{N} \mu(A_n)} = 1 \quad \textit{for } \mu\textit{-a.e. } x \in X\,. \tag{6.5}$$

We remark that (6.4) is precisely the *pairwise independence* of the sets $\{A_n\}$, a condition much weaker than independence (Exercise 6.3). Since

$$\mu(A_m \cap A_n) - \mu(A_m)\mu(A_n) = \int_X \big(1_{A_m} - \mu(A_m)\big)\big(1_{A_n} - \mu(A_n)\big)\, d\mu\,,$$

(6.4) is equivalent to saying that the characteristic functions 1_{A_n} are *uncorrelated*, that is, their projections $1_{A_n} - \mu(A_n)$ onto the space

$$L_0^2(X,\mu) \stackrel{\text{def}}{=} \Big\{f \in L^2(X,\mu) \;\Big|\; \int_X f\, d\mu = 0\Big\}$$

are pairwise orthogonal.

It will be convenient to introduce the following notation: for $1 \le M \le N \le \infty$ let

$$S_{M,N} \stackrel{\text{def}}{=} \sum_{n=M}^{N} 1_{A_n}\,, \quad E_{M,N} \stackrel{\text{def}}{=} \sum_{n=M}^{N} \mu(A_n) = \int_X S_{M,N}\, d\mu\,.$$

We will write S_N and E_N for $S_{1,N}$ and $E_{1,N}$ respectively. Then (6.1) translates into $E_\infty = \infty$, and (6.5) into $\lim_{N\to\infty} \frac{S_N(x)}{E_N} = 1$ μ-almost everywhere, which is a quantitative strengthening of the conclusion of Lemma 6.1, i.e. of the statement that the set

$$\limsup_n A_n = \{x \in X \mid S_\infty(x) = \infty\}$$

has full measure.

PROOF OF LEMMA 6.2. Let us denote $\frac{S_N}{E_N} - 1$ by Ψ_N. Clearly $\Psi_N \in L_0^2(X,\mu)$. Let us compute its L^2-norm (denoted here by $\|\cdot\|$): $\|\Psi_N\| = \frac{1}{E_N}\|S_N - E_N\|$, and

$$\begin{aligned} \|S_N - E_N\|^2 &= \Big\|\sum_{n=1}^{N}\big(1_{A_n} - \mu(A_n)\big)\Big\|^2 \underset{(6.4)}{=} \sum_{n=1}^{N}\big\|1_{A_n} - \mu(A_n)\big\|^2 \\ &= \sum_{n=1}^{N}\big(\mu(A_n) - \mu(A_n)^2\big) \le E_N\,. \end{aligned} \tag{6.6}$$

It follows that $\Psi_N \to 0$ in $L^2(X,\mu)$. Hence (Exercise 6.4) there exists a subsequence $N_k \to \infty$ such that $\Psi_{N_k} \to 0$ μ-a.e. This already shows that for μ-a.e. $x \in X$ one has $\limsup_{N\to\infty} \frac{S_N(x)}{E_N} \geq 1$ (hence $\limsup_n A_n$ is conull, the result which we are after) and $\liminf_{N\to\infty} \frac{S_N(x)}{E_N} \leq 1$. The proof of the stronger version is left to the reader (Exercise 6.4). □

6.3. Quasi-independence. A quick analysis of the above proof shows that it has a lot of room for improvement. For convenience, let us write

$$V_{M,N} \stackrel{\text{def}}{=} \sum_{m,n=M}^{N} \mu(A_m \cap A_n) = \int_X S_{M,N}^2 \, d\mu \quad \text{and} \quad V_N = V_{1,N} \,.$$

As can be shown by a calculation similar to (6.6), it follows from (6.4) that

$$V_N \leq E_N^2 + E_N \,.$$

On the other hand, the conclusion of Lemma 6.2 still holds if (6.4) is replaced by saying that for some $C < \infty$ and $\gamma < 2$ one has

$$V_N \leq E_N^2 + C E_N^{\gamma} \quad \forall\, N \in \mathbb{N} \,. \tag{6.7}$$

This is also left to the reader (Exercise 6.5). The latter condition, unlike (6.4), can in fact be verified in many situations arising in metric number theory. Furthermore, (6.5) can be strengthened into an estimate for the error term.

However for our purposes it will suffice to have a weaker conclusion of S_N being infinite on a set of positive measure, and we proceed to state another lemma (dating back to Paley and Zygmund, 1932) which uses a condition weaker than (6.7).

Lemma 6.3. *Let $\{A_n\}$ be a collection of subsets of a probability space (X,μ) such that* (6.1) *holds, and such that for some C one has*

$$\sum_{m,n=1}^{N} \mu(A_m \cap A_n) \leq C\Big(\sum_{n=1}^{N} \mu(A_n)\Big)^2 \quad \textit{for infinitely many } N \in \mathbb{N} \,, \tag{6.8}$$

or, equivalently,

$$\liminf_{N\to\infty} \frac{E_N^2}{V_N} \geq 1/C \,. \tag{6.9}$$

Then $\mu(\limsup_n A_n) \geq 1/C$.

Observe that (6.8) with "infinitely many" replaced by "all" coincides with (6.7) when $\gamma = 2$; however in the course of doing Exercise 6.5 it should become clear that the assumption $\gamma < 2$ is crucial for the quantitative result. In fact, a choice $A_n \equiv A$ shows that the conclusion of the above lemma cannot be improved (Exercise 6.6).

PROOF OF LEMMA 6.3. It will be convenient to use the notation:

$$A_{M,N} \stackrel{\text{def}}{=} \bigcup_{n=M}^{N} A_n\,.$$

Then one can write

$$\mu(\limsup_n A_n) = \mu\Big(\bigcap_{M=1}^{\infty}\bigcup_{n=M}^{\infty} A_n\Big) = \lim_{M\to\infty}\mu\Big(\bigcup_{n=M}^{\infty} A_n\Big) = \lim_{M\to\infty}\lim_{N\to\infty}\mu(A_{M,N})\,. \tag{6.10}$$

This shows that it might be a good idea to estimate $\mu(A_{M,N})$ from below. For that, note that $S_{M,N}$ vanishes outside $A_{M,N}$, and write

$$\begin{aligned} E_{M,N}^2 &= \Big(\int_X S_{M,N}\,d\mu\Big)^2 = \Big(\int_X 1_{A_{M,N}} S_{M,N}\,d\mu\Big)^2 \\ &\underset{\text{(Cauchy-Schwarz)}}{\le} \int_X (1_{A_{M,N}})^2\,d\mu \int_X S_{M,N}^2\,d\mu = \mu(A_{M,N})V_{M,N}\,. \end{aligned}$$

Therefore

$$\mu(A_{M,N}) \ge \frac{E_{M,N}^2}{V_{M,N}}\,. \tag{6.11}$$

Thus, in view of (6.10), the desired conclusion would follow from

$$\limsup_{M\to\infty,\,N\ge M}\frac{E_{M,N}^2}{V_{M,N}} \ge 1/C\,. \tag{6.12}$$

To pass from (6.9) to (6.12), write

$$\frac{E_{M,N}^2}{V_{M,N}} \ge \frac{(E_N - E_{M-1})^2}{V_N} \ge \frac{E_N^2}{V_N} - \frac{2E_N E_{M-1}}{V_N}\,, \tag{6.13}$$

and note that $V_N \ge E_N^2$ for all N (this can be seen e.g. by looking at (6.11)). In view of this and (6.1), given any $\varepsilon > 0$ and $M \in \mathbb{N}$ one can choose $N \in \mathbb{N}$ such that the right hand side of (6.13) is not less than $\frac{E_N^2}{V_N} - \varepsilon$. This establishes (6.12) and finishes the proof. □

6.4. Back to Duffin-Schaeffer Theorem. We are now ready to start the

PROOF OF THEOREM 4.11. Take an arbitrary function $\psi : \mathbb{N} \to \mathbb{R}_+$, and write $A_q \stackrel{\text{def}}{=} A'\big(q,\psi(q)\big)$, so that, according to (4.4), $\mathcal{W}'(\psi)\cap[0,1] = \limsup_q A_q$. Our goal is to use Lemma 6.3, and clearly we would like to relate both sides of (4.9) to those of (6.8).

Recall that it was assumed that $\psi(q) < 1/2$ for all q. This forces the intervals of which A_q is the union, see (4.5), to be disjoint. Hence

$$\sum_{q=1}^{N}\lambda(A_q) = 2\sum_{q=1}^{N}\frac{\varphi(q)}{q}\psi(q)\,. \tag{6.14}$$

In order to estimate $\sum_{q,s=1}^{N} \lambda(A_q \cap A_s)$, consider a situation when a subinterval of A_q has a nonempty intersection with a subinterval of A_s for $q, s \in \mathbb{N}$. That is, when

$$\left(\frac{p}{q} - \frac{\psi(q)}{q}, \frac{p}{q} + \frac{\psi(q)}{q}\right) \bigcap \left(\frac{r}{s} - \frac{\psi(s)}{s}, \frac{r}{s} + \frac{\psi(s)}{s}\right) \neq \varnothing\,.$$

for some

$$0 < p < q \quad \text{and} \quad 0 < r < s \tag{6.15}$$

with $\gcd(p,q) = \gcd(r,s) = 1$. The latter condition ensures that the centers p/q and r/s of the intervals are distinct, hence we have

$$0 < \left|\frac{p}{q} - \frac{r}{s}\right| < \frac{\psi(q)}{q} + \frac{\psi(s)}{s}\,,$$

or, equivalently,

$$0 < |ps - rq| < s\psi(q) + q\psi(s)\,.$$

We are led to the problem of bounding the number of solutions (p, s) of

$$0 < |ps - rq| < A \tag{6.16}$$

for fixed $q, s \in \mathbb{N}$ and $A > 0$. Here is a simple lemma which will suffice for our purposes:

Lemma 6.4. *Given $q, s \in \mathbb{N}$ and $A > 0$, the number of solutions (p, r) of (6.16) and (6.15) is not greater than $2A$.*

Somewhat surprisingly, the answer turns out to be independent of q and s. We postpone the discussion of this phenomenon until Exercise 6.9, and proceed to the

Proof. Put $n = \gcd(q, s)$, $q' = q/n$ and $s' = s/n$, and rewrite (6.16) as

$$0 < \left|ps' - rq'\right| < A/n\,.$$

Hence for some a with $|a| < A/n$ one has $ps' = rq' + a$, that is,

$$ps' = rq' + a \mod q'\,.$$

Note that for any given a there is at most one solution of the above congruence in p modulo q', hence at most $q/q' = n$ solutions p with $0 < p < q$. Clearly r is determined uniquely by p and a, and there are no more than $2A/n$ possible values of a, which finishes the proof. □

We conclude that the number of pairs of overlapping intervals is at most

$$4\big(s\psi(q) + q\psi(s)\big) \le 4 \max\big(s\psi(q), q\psi(s)\big)\,,$$

and the length of the intersectino of the two intervals is no more than that of the shortest interval, that is, $\min\left(\frac{2\psi(q)}{q}, \frac{2\psi(s)}{s}\right)$. Therefore

$$\lambda(A_q \cap A_s) \le 8 \max\big(s\psi(q), q\psi(s)\big) \min\left(\frac{\psi(q)}{q}, \frac{\psi(s)}{s}\right) = 8\psi(q)\psi(s)\,,$$

and one can write

$$\sum_{q,s=1}^{N} \lambda(A_q \cap A_s) \le 8 \sum_{q,s=1}^{N} \psi(q)\psi(s) = 8\Big(\sum_{q=1}^{N} \psi(q)\Big)^2 \underset{(4.9)}{\le} 8\Big(\frac{1}{c}\sum_{q=1}^{N} \frac{\varphi(q)}{q}\psi(q)\Big)^2$$

$$\underset{(6.14)}{\le} 8\Big(\frac{1}{2c}\sum_{q=1}^{N} \lambda(A_q)\Big)^2 = \frac{2}{c^2}\Big(\frac{1}{2c}\sum_{q=1}^{N} \lambda(A_q)\Big)^2,$$

and an application of Lemma 6.3 finishes the proof. □

6.5. Exercises.

Exercise 6.1. It was used in the proof of Lemma 6.1 that $e^{-x} \ge 1 - x$ for all $0 \le x \le 1$, and therefore, given a sequence $\{x_n\} \subset [0,1]$, $\sum_n x_n = \infty$ implies $\prod_n (1 - x_n) = 0$. Prove that these two conditions are in fact equivalent.

Exercise 6.2. Prove that the independence of a collection $\{A_n\}$ implies that

(a) (6.3) is satisfied for every infinite sequence of indices $n_1 < \cdots < n_k < \dots$;
(b) the collection $\{A_n^c\}$ is independent.

Exercise 6.3. Find three subsets A_1, A_2, A_3 of $[0,1]$ of positive Lebesgue measure which are pairwise independent (with respect to λ) but have empty triple intersection (hence not independent).

Exercise 6.4. On the proof of Lemma 6.2:

(a) Prove Chebyshev's Inequality: for any $f \in L^2(X,\mu)$ and $\varepsilon > 0$ one has
$$\mu\big(\{x \in X \mid |f(x) > \varepsilon\}\big) \le \|f\|^2/\varepsilon^2 .$$
(b) Let $\{f_n\} \subset L^2(X,\mu)$ be such that $\sum_{n=1}^{\infty} \|f_n\| < \infty$. Use (a) and Lemma 4.1 to show that $f_n \to 0$ μ-a.e.
(c) Show that in the proof of Lemma 6.2 one can choose $\{N_k\}$ as follows:
$$N_k \overset{\text{def}}{=} \inf\{N \mid E_N \ge k^2\} .$$
(d) Finish the proof of Lemma 6.2 by observing that for $N_k \le N < N_{k+1}$ one has
$$\frac{S_N}{E_N} \le \frac{S_{N_{k+1}}}{E_{N_k}} = \frac{S_{N_{k+1}}}{E_{N_{k+1}}}\frac{E_{N_{k+1}}}{E_{N_k}} ,$$
together with a similar estimate from below.

Exercise 6.5. Modify the proof of Lemma 6.2 to show that (6.1) and (6.7) imply (6.5).

Exercise 6.6. Show that Lemma 6.3 is sharp by considering the case $A_n \equiv A$.

EXERCISE 6.7. Compare the two sides of (6.7) for $\{A_n\}$ as in (6.2)

EXERCISE 6.8. Identify the spots in the proof of Lemmas 6.1, 6.2 and 6.3 where the assumption $\mu(X) = 1$ was used. Show that the results are still true if it is only assumed that $\mu(X) < \infty$. What about $\mu(X) = \infty$?

EXERCISE 6.9. Describe a geometric interpretation of the statement of Lemma 6.4, and try to understand from the geometric point of view whether or not the result is to be expected.

6.6. Yet again about inhomogeneous approximation. What happens if one goes through the logic of the last three sections with (1.8) instead of (1.3), for a fixed β or almost every β? Will some of the argument survive? We will certainly return to this topic later, but it might be a good idea to think about it meanwhile...

7. Continued fractions: algebra and geometry

Our next task is to review a classical constructive approach to approximation of real numbers by rationals. Namely, it is important not only to be able to prove that certain good approximations of $\alpha \in \mathbb{R}$ exist or do not exist, but also to exhibit an algorithm of producing best possible approximations. There are many sources where the theory of continued fractions is presented from various points of view, and the current exposition does not claim any originality or superiority. My goal here is to *visualize* the main ideas related to continued fraction techniques, and use this visualization while discussing applications to Diophantine approximation.

7.1. Notation and preliminaries. Let $a_0, a_1, \dots$ be real numbers with $a_n > 0$ for $n > 0$. A *finite continued fraction* is an expression of the form

$$[a_0; a_1, \dots, a_n] \stackrel{\text{def}}{=} a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\dots + \frac{1}{a_n}}}}\,.$$

More generally, by a *continued fraction* we will mean an expression of the form

$$[a_0; a_1, \dots] \stackrel{\text{def}}{=} a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\dots}}} = \lim_{n\to\infty} [a_0; a_1, \dots, a_n]\,, \tag{7.1}$$

provided that the limit exists.

The aim of this section is to show how a real number α can be expressed in the form (7.1), where a_0 is an integer and a_n is a positive integer for any $n \geq 1$. We first deal with the case $\alpha \in \mathbb{Q}$, then describe an algorithm which associates to any irrational α an infinite sequence of integers a_n as above, and then show that the sequence of rational numbers $[a_0; a_1, \dots, a_n]$ converges to α.

Lemma 7.1. *Any rational number r has exactly two different continued fraction expansions*

$$r = [a_0; a_1, \dots, a_n] \tag{7.2}$$

with $a_0 \in \mathbb{Z}$ and $a_n \in \mathbb{N}$ for $n \geq 1$. These are $[r]$ and $[r-1; 1]$ if $r \in \mathbb{Z}$, and, otherwise, one of them has the form (7.2) *with $a_n \geq 2$, and the other one is $r = [a_0; a_1, \dots, a_n - 1, 1]$.*

Proof. Write $r = p/q$ with p, q relatively prime and $q > 0$. We argue by induction on q. If $q = 1$, clearly $r = p = [p]$ is the only expansion (7.2) with $n = 0$. If (7.2) holds with $n \geq 1$, we can write $r = a_0 + 1/[a_1; a_2, \dots, a_n]$, and since $a_1 \geq 1$ and $r \in \mathbb{Z}$, conclude that $n = 1$ and $a_1 = 1$, thus $r = a_0 + 1 = [r-1; 1]$.

Now assume that $q \geq 2$ and that the lemma holds for any r with denominator between 1 and $q - 1$. Note that a_0 must be the integer part of r. Performing the Euclidean division of p by q, one writes $p = a_0 q + s$ where $s \in \mathbb{Z}$ and $1 \leq s < q$. Thus $p/q = a_0 + s/q = a_0 + 1/(q/s) = [a_0; q/s]$, and by the induction assumption q/s has exactly two expansions, which give rise to similar expansions of r. □

7.2. The algorithm. Now let us describe an algorithm which will allow us to assocuate to any $\alpha \notin \mathbb{Q}$ an infinite sequence of integers a_n. Denoting $\alpha_0 = \alpha$, for any nonnegative integer n we define inductively the integer a_n and the real number $\alpha_{n+1} > 1$ by

$$a_n = \lfloor \alpha_n \rfloor, \quad \alpha_{n+1} = 1/\{\alpha_n\} = \frac{1}{\alpha_n - a_n}. \tag{7.3}$$

Observe that $a_n \geq 1$ for $n \geq 1$. The algorithm does not stop since α is assumed to be irrational. If α were rational, the same algorithm would terminate and associate to α a finite sequence of integers (see Exercise 7.1).

Here is an equivalent way to describe this algorithm: first let $a_0 = \lfloor \alpha \rfloor$ and $\beta_0 = \{\alpha\}$, and then for any $n \in \mathbb{N}$ inductively define

$$a_n = \lfloor 1/\beta_n \rfloor, \quad \beta_{n+1} = \{1/\beta_n\} = 1/\beta_n - a_n = T(\beta_n), \tag{7.4}$$

where T is the map from $[0, 1]$ to itself given by

$$T(x) \stackrel{\text{def}}{=} \{1/x\}, \tag{7.5}$$

called the *Gauss map*, to be discussed in more detail later. Here one has $0 < \beta_n < 1$ for any n. It easily follows from (7.3) and (7.4) that one has $\alpha_n = a_n + 1/\alpha_{n+1}$ and $\beta_n = 1/(a_{n+1} + \beta_{n+1})$ for any n, therefore by induction

$$\alpha = [a_0; a_1, \dots, a_n, \alpha_{n+1}] = [a_0; a_1, \dots, a_n + \beta_n] \quad \text{and} \quad \beta_n = 1/\alpha_{n+1}, \tag{7.6}$$

see Exercise 7.2.

Definition 7.2. Let $\alpha \in \mathbb{R}$ be irrational (resp. rational). The numbers a_n, where $n = 0, 1, \dots$ (resp. $n = 0, 1, \dots, N$ for some N) associated to α by the algorithm defined above are called the *partial quotients* of α. For any n (resp. $n \leq N$) the rational number

$$\frac{p_n}{q_n} \stackrel{\text{def}}{=} [a_0; a_1, \dots, a_n],$$

written in lowest terms, is called the *n-th convergent* of α.

Sometimes we will write $a_n = a_n(\alpha)$, $p_n = p_n(\alpha)$, $q_n = q_n(\alpha)$, as well as $\alpha_n = \alpha_n(\alpha)$, $\beta_n = \beta_n(\alpha)$, to highlight the dependence on α. We will also denote by $\mathbf{v}_n$ the vector $\begin{pmatrix} p_n \\ q_n \end{pmatrix}$.

In what follows, unless stated otherwise, we assume that $\alpha \in \mathbb{R}$ is irrational and let $a_n, p_n, q_n, \alpha_n, \beta_n, \mathbf{v}_n$ be as in Definition 7.2. However the statements below remain true for $\alpha \in \mathbb{Q}$ provided that a_n, p_n, q_n are well defined.

The next lemma provides a very efficient way both to compute the convergents and to visualize the computation. The construction is inductive and therefore it will be convenient to describe the data used in the basis of the induction in more detail. Note that by definition we have

$$p_0 = a_0 \text{ and } q_0 = 1, \text{ that is, } \mathbf{v}_0 = \begin{pmatrix} a_0 \\ 1 \end{pmatrix}. \tag{7.7}$$

We venture two more steps back, defining

$$\begin{gathered} p_{-2} = 0,\ q_{-2} = 1,\ p_{-1} = 1,\ q_{-1} = 0, \\ \text{so that } \mathbf{v}_{-1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ \mathbf{v}_{-2} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{gathered} \tag{7.8}$$

Recall also that in §**??** we denoted by L_α the line in the xy plane given by $x = \alpha y$.

Lemma 7.3. *For any nonnegative integer n:*

(a) *with one exception, a_n is equal to the maximal number of times one can add $\mathbf{v}_{n-1}$ to $\mathbf{v}_{n-2}$ so that the resulting vector and $\mathbf{v}_{n-2}$ are in the same connected component of $\mathbb{R}^2 \smallsetminus L_\alpha$;*

(b) *$\mathbf{v}_n = \mathbf{v}_{n-2} + a_n \mathbf{v}_{n-1}$, the resulting vector one encounters in part* (a)*.*

The exception in part (a) *occurs when $\alpha < 0$ and $n = 0$; then $-a_n$ is equal to the minimal number of times one needs to subtract $\mathbf{v}_{n-1}$ to $\mathbf{v}_{n-2}$ so that the resulting vector and $\mathbf{v}_{n-2}$ are in different connected component of $\mathbb{R}^2 \smallsetminus L_\alpha$.*

Proof. We assume $\alpha > 0$ and leave the exceptional case to the reader, see Exercise 7.4(a). If $n = 0$ both claims are obvious from (7.7), (7.8) and the fact that $a_0 = [\alpha]$. For brevity, instead of performing the induction step, let us derive the claims for $n = 1$, and direct the reader to Exercise 7.4(b) to transform the argument into a rigorous inductive proof.

The idea is to change variables using the linear transformation A_0 of $\mathbb{R}^2$ which sends the ordered basis $(\mathbf{v}_{-2}, \mathbf{v}_{-1})$ to $(\mathbf{v}_{-1}, \mathbf{v}_{-0})$. Since $\mathbf{v}_0$ is equal to $\mathbf{v}_{-2} + a_0 \mathbf{v}_{-1}$, one has $A_0 = \begin{pmatrix} 0 & 1 \\ 1 & a_0 \end{pmatrix}$ and $A_0^{-1} = \begin{pmatrix} -a_0 & 1 \\ 1 & 0 \end{pmatrix}$. The latter sends the line $L_\alpha = \mathbb{R} \begin{pmatrix} 1 \\ \alpha \end{pmatrix}$ to

$$\mathbb{R} \cdot A_0^{-1} \begin{pmatrix} 1 \\ \alpha \end{pmatrix} = \mathbb{R} \begin{pmatrix} \alpha - a_0 \\ 1 \end{pmatrix} = \mathbb{R} \begin{pmatrix} \{\alpha\} \\ 1 \end{pmatrix} = \mathbb{R} \begin{pmatrix} 1 \\ \alpha_1 \end{pmatrix}.$$

Since a_1 is to α_1 as a_0 is to $\alpha = \alpha_0$, the claims for $n = 1$ are derived from the case $n = 0$ by applying A_0 to all the data. □

Some **remarks**:

7.2.1. The formula in (b) is usually written in coordinates, that is

$$\begin{cases} p_n = p_{n-2} + a_n p_{n-1} \\ q_n = q_{n-2} + a_n q_{n-1} \end{cases}. \tag{7.9}$$

7.2.2. If, generalizing the notation introduced in the proof, we let $A_n \stackrel{\text{def}}{=} \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix}$, it can be easily shown by induction (Exercise 7.5) that

$$\begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} = \begin{pmatrix} \mathbf{v}_{n-1} & \mathbf{v}_n \end{pmatrix} = A_n \cdots A_1 A_0 \,. \tag{7.10}$$

Note that $\det(A_n) = -1$ for all n, that is, A_n is area-preserving and orientation-reversing.

7.2.3. I wish I knew how to draw a picture illustrating the proof, maybe I'll learn some day.

7.3. One-to-one correspondence. The next lemma lists several properties of convergents, which are easily derived from either the outcome of Lemma 7.3 or the inductive procedure of its proof.

COROLLARY 7.4. *For any nonnegative integer n:*

(a) $\begin{vmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{vmatrix} = (-1)^n$;

(b) $\begin{vmatrix} p_{n-2} & p_n \\ q_{n-2} & q_n \end{vmatrix} = (-1)^{n-1} a_n$;

(c) $\alpha - \dfrac{p_n}{q_n} > 0$ *iff n is even;*

(d) $\dfrac{p_n}{q_n} - \dfrac{p_{n-2}}{q_{n-2}} > 0$ *iff n is even.*

PROOF. Parts (a) and (b) are clear when $n = 0$, and the general case follows as in the proof of Lemma 7.3. Part (a) is also straightforward from (7.10). Part (c) is immediate from Lemma 7.3(a), since p_n/q_n is the slope of $\mathbf{v}_n$ and α is the slope of L_α. To deduce (d), note that the difference $p_n/q_n - p_{n-2}/q_{n-2}$ and the expression in (b) have opposite signs. □

Several other properties are listed in Exercise 7.6.

Now we are ready to establish that any irrational number can be uniquely written as an infinite continued fraction. Here is the main theorem of the section:

THEOREM 7.5. (a) *For any $\alpha \notin \mathbb{Q}$, $\alpha = \lim_{n\to\infty} \dfrac{p_n(\alpha)}{q_n(\alpha)} = [a_0; a_1, \dots]$.*

(b) *If $b_0, b_1, \dots$ is a sequence of integers with $b_n \geq 1$ for $n \geq 1$ such that the limit $[b_0; b_1, \dots]$ exists and is equal to α, then $b_n = a_n(\alpha)$ for each n.*

(c) *If $a_0, a_1, \dots$ is a sequence of integers with $a_n \geq 1$ for $n \geq 1$, then the limit $[a_0; a_1, \dots]$ exists and is equal to the irrational number whose partial quotients are precisely a_n.*

Proof. It follows from Corollary 7.4(d) that

$$\text{the sequence } \tfrac{p_{2n}}{q_{2n}} \text{ (resp. } \tfrac{p_{2n+1}}{q_{2n+1}}) \text{ is strictly increasing (resp. decreasing),} \tag{7.11}$$

from Corollary 7.4(c) that α belongs to the interval $\left(\frac{p_{2n}}{q_{2n}}, \frac{p_{2n+1}}{q_{2n+1}}\right)$ for any n, and from Corollary 7.4(a) that

$$\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = \frac{(-1)^{n-1}}{q_{n-1}q_n} \to 0 \text{ as } n \to \infty\,. \tag{7.12}$$

This proves (a). For (b), observe (Exercise 7.7) that

$$[b_0; b_1, \dots] \text{ cannot equal } [a_0; a_1, \dots] \text{ as soon as } \exists\, n \text{ with } a_n \neq b_n\,. \tag{7.13}$$

Finally, to prove (c) let us choose an arbitrary $N \in \mathbb{N}$ and let p_n/q_n stand for $[a_0; a_1, \dots, a_n]$, $n \leq N$. Applying Lemma 7.3 to the rational number $[a_0; a_1, \dots, a_N]$, one sees that all the properties listed in Corollary 7.4 hold in this context as well. In particular, as in the proof of (a) we can establish (7.11) and (7.12). Therefore the two sequences in (7.11) must converge to the same limit, namely to the irrational number $[a_0; a_1, \dots]$ whose partial quotients are precisely $a_0, a_1, \dots$ in view of part (b). □

To derive additional properties of continued fraction coefficients, let us write down an expression for an irrational number α in terms of α_{n+1}:

Proposition 7.6. *Let $\alpha \notin \mathbb{Q}$ and a nonnegative integer n be given. Then for any $x > 0$, one has*

$$[a_0; a_1, \dots, a_n, x] = \frac{p_{n-1} + x p_n}{q_{n-1} + x q_n}\,; \tag{7.14}$$

in particular,

$$\alpha = \frac{p_{n-1} + \alpha_{n+1} p_n}{q_{n-1} + \alpha_{n+1} q_n}\,. \tag{7.15}$$

Proof. (7.14) can be thought of as a continuous version of Lemma 7.3: arguing inductively as in the proof of that lemma, one can show that the slope of the vector $\mathbf{v}_{n-1} + x\mathbf{v}_n$ is equal to the left hand side of (7.14). The 'in particular' part is immediate from (7.6). □

It will be convenient to have a notation for the ratio of two successive denominators of the convergents of α:

$$\varphi_n \stackrel{\text{def}}{=} q_n/q_{n+1}\,. \tag{7.16}$$

It is easy to deduce from (7.9) that the sequence φ_n can be recursively defined by

$$\varphi_0 = 0,\ \varphi_n = \frac{1}{a_n + \varphi_{n-1}}\,,$$

which is equivalent to saying that

$$\varphi_n = [0; a_{n+1}, a_n, \dots, a_1]\,, \tag{7.17}$$

see Exercise 7.10.

COROLLARY 7.7. *Let $\alpha \notin \mathbb{Q}$ and a nonnegative integer n be given. Then*

(a) $q_n(q_n\alpha - p_n) = \dfrac{(-1)^n}{\alpha_{n+1} + \varphi_{n-1}} = \dfrac{(-1)^n}{[a_{n+1}; a_{n+2}, \dots] + [0; a_n, a_{n-1}, \dots, a_1]}$;

(b) $q_{n+1}(q_n\alpha - p_n) = \dfrac{(-1)^n}{1 + \beta_{n+1}\varphi_n} = \dfrac{(-1)^n}{1 + [0; a_{n+2}, \dots][0; a_{n+1}, a_n, \dots, a_1]}$.

PROOF. Using (7.15), we write

$$q_n\alpha - p_n = q_n \frac{p_{n-1} + \alpha_{n+1}p_n}{q_{n-1} + \alpha_{n+1}q_n} - p_n = \frac{p_{n-1}q_n - p_n q_{n-1}}{q_{n-1} + \alpha_{n+1}q_n},$$

thus the first equality in (a) follows from Corollary 7.4(a) and (7.16), and the second one from (7.17). To demonstrate part (b), observe that

$$q_{n-1} + \alpha_{n+1}q_n = q_{n-1} + (a_{n+1} + \beta_{n+1})q_n \underset{(7.9)}{=} q_{n+1} + \beta_{n+1}q_n,$$

and use (7.16) and (7.17) again. □

Finally let us point out that it can be easily derived from the recursive relations (7.9) that both p_n and q_n grow at least exponentially in n for any irrational α. More precisely, the following is true:

PROPOSITION 7.8. *For any $\alpha \notin \mathbb{Q}$ and $n \in \mathbb{N}$ one has $p_n \geq 2^{\frac{n}{2}-1}$ and $q_n \geq 2^{\frac{n-1}{2}}$.*

PROOF. Induction on n, see Exercise 7.8. □

See Theorem 10.4(d) for the rate of growth of p_n and q_n for α generic with respect to Lebesgue measure.

7.4. Exercises.

EXERCISE 7.1. Prove that the algorithm (7.3) terminates in finite number of steps if α is a rational number. Which of the expansions of Lemma 7.1 is produced?

EXERCISE 7.2. Prove (7.6).

EXERCISE 7.3. Show that p_n and q_n are always coprime.

EXERCISE 7.4. Finish the proof of Lemma 7.3 by:

(a) treating the case $\alpha < 0$;

(b) rigorously performing the induction step.

EXERCISE 7.5. Prove (7.10).

EXERCISE 7.6. Prove that for any nonnegative integer n:

(a) $\dfrac{p_n}{q_n} = a_0 + \displaystyle\sum_{i=1}^{n} \frac{(-1)^{n-1}}{q_{n-1}q_n}$;

(b) $\dfrac{p_n}{q_n} - \dfrac{p_{n-2}}{q_{n-2}} = (-1)^n \dfrac{q_n - q_{n-2}}{q_{n-2}q_{n-1}q_n}$.

EXERCISE 7.7. Prove (7.13).

EXERCISE 7.8. Prove Proposition 7.8 by induction on n, and then improve it by finding a sharp exponential lower estimate which works for all irrational α.

EXERCISE 7.9. Show that the map from ($\mathbb{Z} \times \mathbb{N}^{\mathbb{N}}$, product topology) to ($\mathbb{R} \smallsetminus \mathbb{Q}$, induced topology) that sends $(a_0, a_1, a_2, \dots)$ to $[a_0; a_1, a_2, \dots]$ is a homeomorphism.

EXERCISE 7.10. Prove (7.17).

EXERCISE 7.11. Prove that $\alpha = \frac{p_n + \beta_n p_{n-1}}{q_n + \beta_n q_{n-1}}$ and $q_n \alpha - p_n = \frac{(-1)^n \beta_n}{q_n + \beta_n q_{n-1}}$ for any $n \geq 0$.

EXERCISE 7.12. Instead of studying partial quotients a_n obtained as in Definition 7.2 starting from a real number α, one can adopt a purely formal approach and, given an arbitrary sequence $a_0, a_1, \dots$ of real numbers with $a_n > 0$ for $n > 0$, define p_n and q_n inductively as in (7.9). Show that the identities (a) and (b) of Corollary 7.4 will still hold.

EXERCISE 7.13. [$*$] In the generality of the previous exercise, prove that the limit in (7.1) exists if and only if the series $\sum_{n=1}^{\infty} a_n$ diverges.

8. Continued fractions and approximation

8.1. Approximating α by its convergents. The material worked out in the previous section is enough to derive plenty of wonderful applications to Diophantine approximation. First observe that Proposition 7.6(b) immediately implies that $|q_n\alpha - p_n|$ is less than $1/q_n$ for any n; in other words, any convergent p/q of α satisfies (1.7). It also follows that $|q_n\alpha - p_n| < 1/2$, and therefore

$$|q_n\alpha - p_n| = \langle q_n\alpha \rangle ,$$

for any $n \geq 1$.

The next step is to point out a similarity between Corollary 7.4(a) and the characteristic property of successive terms in the Farey series (Proposition 2.2). Namely, one can use Exercise 2.4 to conclude that for any $n \geq 1$, the fractions p_{n-1}/q_{n-1} and p_{n+1}/q_{n+1} are neighbors in $\mathcal{F}_N$ for some N. Therefore, by Proposition 2.2,

$$\begin{gathered}\text{at least one of any two successive convergents of } \alpha \\ \text{satisfies} \quad |\alpha - p/q| < 1/2q^2 .\end{gathered} \tag{8.1}$$

Further, one can follow the lines of the proof of Proposition 2.7 to prove its continued fraction analogue, namely that

$$\begin{gathered}\text{at least one of any three successive convergents of } \alpha \\ \text{satisfies} \quad |\alpha - p/q| < 1/\sqrt{5}q^2\end{gathered} \tag{8.2}$$

(Exercise 8.1).

In order to obtain further information, let us use (7.9) to write down a recursive formula for the differences $|q_n\alpha - p_n|$:

$$\begin{aligned} q_{n+1}\alpha - p_{n+1} &= (q_{n-1} + a_{n+1}q_n)\alpha - (p_{n-1} + a_{n+1}p_n) \\ &= (q_{n-1}\alpha - p_{n-1}) + a_{n+1}(q_n\alpha - p_n) , \end{aligned}$$

hence

$$q_{n-1}\alpha - p_{n-1} = (q_{n+1}\alpha - p_{n+1}) - a_{n+1}(q_n\alpha - p_n) .$$

Note that due to Corollary 7.4(c) the quantities $q_{n+1}\alpha - p_{n+1}$ and $q_n\alpha - p_n$ have opposite signs; thus taking absolute values yileds

$$|q_{n-1}\alpha - p_{n-1}| = |q_{n+1}\alpha - p_{n+1}| + a_{n+1}|q_n\alpha - p_n| . \tag{8.3}$$

This is all one needs to prove that every convergent approximates α better than the previous one:

LEMMA 8.1. *For any $\alpha \notin \mathbb{Q}$ and any nonnegative integer n, one has $a_{n+1} = \left\lfloor \frac{|q_{n-1}\alpha - p_{n-1}|}{|q_n\alpha - p_n|} \right\rfloor$, in particular,*

$$|q_n\alpha - p_n| < |q_{n-1}\alpha - p_{n-1}| . \tag{8.4}$$

PROOF. Simply divide both sides of (8.3) by $|q_n\alpha - p_n|$. □

Another useful thing to do is to turn the exact expression for the difference between $q_n\alpha$ and p_n given by Corollary 7.7 into a two-sided estimate.

LEMMA 8.2. *For any $\alpha \notin \mathbb{Q}$ and any nonnegative integer n, one has*

$$\frac{1}{q_n + q_{n+1}} < |q_n\alpha - p_n| < \frac{1}{q_{n+1}}\,. \tag{8.5}$$

PROOF. Rewrite Corollary 7.7(b) as $q_n\alpha - p_n = \frac{(-1)^n}{q_{n+1}+\beta_{n+1}q_n}$, and use the fact that $0 < \beta_{n+1} < 1$. □

8.2. Convergents as best approximatons. With a little more work, it is possible to strengthen (8.4) to the statement that every convergent approximates α better than any other fraction with a strictly smaller denominator. Let us say that a rational number p/q is a *best approximation* to α if

$$|q\alpha - p| < |q'\alpha - p'| \quad \text{for any } p' \text{ and } q' \text{ with } 0 < q' < q. \tag{8.6}$$

Note that (8.6) implies that

$$|\alpha - p/q| < |\alpha - p'/q'| \quad \text{for any } p' \text{ and } q' \text{ with } 0 < q' < q, \tag{8.7}$$

but the converse is not true, see Exercise 8.2. It is also clear that (8.6), and even (8.7), force p and q to be relatively prime.

The following can be proved:

THEOREM 8.3. *For any $\alpha \in \mathbb{R}$, a fraction p/q is a convergent of α if and only if it is a best approximation to α.*

PROOF. Let us first show that any best approximation is a convergent. Take p and q, with $q > 1$, satisfying (8.6). The process of finding n for which $p/q = p_n/q_n$ will consist of three steps.

First note that p/q must be not less than $a_0 = p_0/q_0$: indeed, since $\alpha \geq a_0$, otherwise we would have $|\alpha - a_0| < |\alpha - p/q| < |q\alpha - p|$, contradicting (8.6).

Likewise, using the fact that $\alpha \leq p_1/q_1$, we can prove that p/q must be not greater than p_1/q_1: for otherwise

$$|\alpha - \tfrac{p}{q}| > |\tfrac{p_1}{q_1} - \tfrac{p}{q}| \geq \tfrac{1}{qq_1} \quad \Rightarrow \quad |q\alpha - p| > \tfrac{1}{q_1} = \tfrac{1}{a_1} \geq |\alpha - a_0|\,.$$

Now we can be sure that p/q is between p_{n-1}/q_{n-1} and p_{n+1}/q_{n+1} for some n. Since p_{n+1}/q_{n+1} is between p_{n-1}/q_{n-1} and p_n/q_n, we can write

$$\tfrac{1}{qq_{n-1}} \leq |\tfrac{p}{q} - \tfrac{p_{n-1}}{q_{n-1}}| \leq |\tfrac{p_n}{q_n} - \tfrac{p_{n-1}}{q_{n-1}}| = \tfrac{1}{qq_{n-1}} \quad \Rightarrow \quad q \geq q_n\,.$$

On the other hand, we know that α is between p_{n+1}/q_{n+1} and p_n/q_n, therefore

$$|\alpha - \tfrac{p}{q}| \geq |\tfrac{p_{n+1}}{q_{n+1}} - \tfrac{p}{q}| \geq \tfrac{1}{qq_{n+1}} \quad \Rightarrow \quad |q\alpha - p| \geq \tfrac{1}{q_{n+1}} \underset{(8.5)}{>} |q_n\alpha - p_n|\,,$$

and (8.6) can be used to conclude that $q = q_n$.

The converse statement can be proved by induction on n. Since $q_0 = 1$, the definition of best approximation is obviously satisfied by p_0/q_0. Assume now that it is known that p_n/q_n is a best approximation for α, and let q be the smallest integer greater than q_n such that $|q\alpha - p| < |q_n\alpha - p_n|$ for some p. By the induction assumption, p/q must also be a best approximation, and hence, in view of the first part of the theorem, a convergent. Then from the choice of q it follows that $q = q_{n+1}$ and $p = p_{n+1}$. □

See Exercise 8.3 for a similar result with (8.6) replaced by (8.7).

As a corollary, we obtain a partial converse to (8.1):

LEMMA 8.4. *If p/q is a reduced fraction with $q > 0$ such that*

$$|\alpha - p/q| < 1/2q^2\,, \tag{8.8}$$

then it is a convergent of α.

PROOF. By the above theorem, it suffices to prove that p/q is a best approximation to α. Take any $\frac{p'}{q'}$ with $0 < q' < q$; it must be different from $\frac{p}{q}$ since the latter is assumed to be reduced, and one can write

$$\tfrac{1}{qq'} \le \left|\tfrac{p}{q} - \tfrac{p'}{q'}\right| \le \left|\alpha - \tfrac{p}{q}\right| + \left|\alpha - \tfrac{p'}{q'}\right| \underset{(8.8)}{<} \tfrac{1}{2q^2} + \tfrac{|q'\alpha - p'|}{q'}\,,$$

hence

$$|q'\alpha - p'| > \tfrac{1}{q} - \tfrac{q'}{2q^2} > \tfrac{1}{2q} \underset{(8.8)}{>} |q\alpha - p|\,,$$

finishing the proof. □

See Exercise 8.4 for a similar result with (8.8) replaced by (1.7).

8.3. Convergents and ψ-approximability. Let us now use Lemma 8.2 to derive an 'almost criterion' for the ψ-approximability of α:

THEOREM 8.5. *Let $\psi : \mathbb{N} \to \mathbb{R}_+$ be such that $\psi(x) < 1/2x$ for large x, and let $\alpha \notin \mathbb{Q}$ be given. Then:*

(a) $\alpha \in \mathcal{W}(\psi) \Rightarrow q_{n+1} > \frac{1}{\psi(q_n)} - q_n$ *for infinitely many n*
$\Rightarrow a_{n+1} > \frac{1}{q_n\psi(q_n)} - 2$ *for infinitely many n.*

(b) $\alpha \notin \mathcal{W}(\psi) \Rightarrow q_{n+1} \le \frac{1}{\psi(q_n)}$ *for large enough n*
$\Rightarrow a_{n+1} \le \frac{1}{q_n\psi(q_n)}$ *for large enough n.*

PROOF. If there are infinitely many solutions to (1.3), it follows from Lemma 8.4 and the assumption on ψ that one has $|q_n\alpha - p_n| < \psi(q_n)$ for infinitely many n. Using the lower estimate of (8.5), one deduces that $\frac{1}{q_n+q_{n+1}}$ is less than $\psi(q_n)$, from which the first and, with the help of (7.9), the second conclusion of (a) follows. Part (b) is proved along the same lines, using the upper bound in (8.5). □

Even though formally the above result does not give a precise description of the ψ-approximability of α in terms of continued fractions, one can still use it efficiently, since the right hand sides of the inequalities in (a) and (b) are not very different from each other. Here

Corollary 8.6. *$\alpha \notin \mathbb{Q}$ is badly approximable iff its partial quotients are uniformly bounded.*

Proof. By Theorem 8.5, $\alpha \notin \mathcal{W}_{c,1}$ implies that $a_n \leq 1/c$ for large enough n, and $\alpha \notin \mathcal{W}_{c,1}$ implies that $a_n > 1/c - 2$ for infinitely many n. □

In particular, this proves the existence of continuum many badly approximable numbers, a result mentioned in §3.2. Similarly one can describe very well approximable numbers in terms of their continued fraction expansions:

Corollary 8.7. *The Diophantine exponent of $\alpha \notin \mathbb{Q}$ is equal to*

$$\omega(\alpha) = \limsup_{n\to\infty} \frac{\log q_{n+1}}{\log q_n} = \limsup_{n\to\infty} \frac{\log a_{n+1}}{\log q_n} + 1\,. \tag{8.9}$$

The proof is straightforward and is left to the reader, see Exercise 8.7. Here is another application:

Corollary 8.8. *Let $\psi : \mathbb{N} \to \mathbb{R}_+$ be such that $\lim_{x\to\infty} x\psi(x) = 0$. Then for any $0 < c < 1$ there exist continuum many real numbers α which belong to $\mathcal{W}(\psi)$ but not $\mathcal{W}(c\psi)$.*

The proof is also left to the reader (Exercise 8.8). Note that it follows that for any $v > 1$ the set $\{\alpha \mid \omega(\alpha) = v\}$ is uncountable, as was mentioned in §3.4. See also Exercise 8.9 for another way the Diophantine exponent of α affects the growth of the denominators of its convergents.

8.4. Exercises.

Exercise 8.1. Show that Proposition 2.5 remains true if one replaces the assumption $m = k + p, n = l + q$ by $m = k + ap, n = l + aq$ for some $a \geq 1$, and use it to derive (8.2).

Exercise 8.2. Find an example of p/q and α which satisfy (8.7) but not (8.6).

Exercise 8.3. [∗] Prove that any p/q satisfying (8.7) must be either a convergent or of the form $\frac{p_{n-1}+ap_n}{q_{n-1}+aq_n}$ for some n and $0 < a < a_{n+1}$ (such fractions are called *intermediate convergents* of α). Also find a counterexample to the converse statement, that is, an intermediate convergent p/q of α which fails to satisfy (8.7).

Exercise 8.4. [∗] Prove that for any p/q satisfying (1.7) there exists n such that

$$\frac{p}{q} \in \left\{ \frac{p_n}{q_n}, \frac{p_n + p_{n+1}}{q_n + q_{n+1}}, \frac{p_{n+2} - p_{n+1}}{q_{n+2} - q_{n+1}} \right\}.$$

Exercise 8.5. Show that Lemma 8.2 can be used to derive (8.4), as well as [∗] Lemma 8.4.

Exercise 8.6. Let p/q be a convergent of α and let $0 < q' < q$. Show that $\langle q'\alpha\rangle > 1/2q$.

EXERCISE 8.7. Prove (8.9).

EXERCISE 8.8. Prove Corollary 8.8.

EXERCISE 8.9. Prove that for any $\alpha \notin \mathbb{Q}$ one has

$$\omega(\alpha) \geq \limsup_{n\to\infty} \frac{\log\log q_n}{n}$$

(in particular, if the above limsup is infinite, α has to be Liouville, and hence transcendental).

9. More about badly approximable numbers

The description of badly approximable numbers as those with uniformly bounded partial quotients turns out to be very useful in many respects. One application is in Exercise 9.1, and another one is coming up shortly.

9.1. Quadratic irrationals. Quadratic irrational numbers, as we know from Liouville's Theorem, are examples of badly approximable numbers, hence have uniformly bounded partial quotients. However a much stronger property of continued fractions of quadratic irrationals had been known long before Liouville. Namely, the following result is due to Lagrange:

THEOREM 9.1. $\alpha \in \mathbb{R}$ *is quadratic irrational if and only if the sequence of its partial quotients is eventually periodic, that is,* $a_{n+k} = a_n$ *for some* n *and all* k.

PROOF. The 'if' part is easy: indeed, the eventual periodicity of $\{a_n\}$ is equivalent to saying that $\alpha_n = \alpha_m$ for some $m \neq n$; hence, using (7.15) one can write

$$\alpha = \frac{p_{n-2} + \alpha_n p_{n-1}}{q_{n-2} + \alpha_n q_{n-1}} = \frac{p_{m-2} + \alpha_n p_{m-1}}{q_{m-2} + \alpha_n q_{m-1}}, \tag{9.1}$$

which boils down to a quadratic equation for α_n with integer coefficients.

Conversely, suppose that α satisfies a quadratic equation

$$P(\alpha) = a\alpha^2 + b\alpha + c = 0 \tag{9.2}$$

with integer coefficients. Plugging the expression for α in terms of α_n, i.e. the first equality in (9.1), into the equation, one can see that α_n will also satisfy a quadratic equation with integer coefficients, namely

$$A_n\alpha_n^2 + B_n\alpha_n + C_n = 0\,. \tag{9.3}$$

An elementary computation shows that for all n one has

$$A_n = ap_{n-1}^2 + bp_{n-1}q_{n-1} + cq_{n-1}^2 = q_{n-1}^2 P(p_{n-1}/q_{n-1}) = C_{n+1}\,.$$

Since

$$\begin{aligned} q_{n-1}^2|P(p_{n-1}/q_{n-1})| &= q_{n-1}^2|P(p_{n-1}/q_{n-1}) - P(\alpha)| \\ &\le Cq_{n-1}^2|\alpha - p_{n-1}/q_{n-1}| \underset{\text{Lemma 8.4}}{<} C/2\,, \end{aligned}$$

where C depends only on α, one concludes that the absolute values of A_n and C_n are uniformly bounded from above. Further, it is not hard to see (Exercise 9.2) that the discriminants of equations (9.2) and (9.3) coincide. Therefore $|B_n|$ are also uniformly bounded from above. This implies that all those coefficients can take only finitely many values, thus $\alpha_n = \alpha_m$ for some $m \neq n$. □

It will be convenient to introduce an equivalence relation based on the tail of the sequence of partial quotients: say that two real numbers α, β are *equivalent* ($\alpha \sim \beta$) if there exist $m, n \in \mathbb{N}$ such that $a_{n+k}(\alpha) = a_{m+k}(\beta)$ for all k, or equivalently, $\alpha_n(\alpha) = \alpha_m(\beta)$. Thus the sequence of partial quotients of a real number is eventually periodic iff it is equivalent to a number α with strictly periodic partial quotients (that is, $\alpha = \alpha_n$ for some n). See Exercise 9.3 for another (very important) description of this equivalence relation, Exercise 9.4 for more on strict vs. eventual periodicity, and Exercise 9.5, Exercise 9.6 for more on quadratic irrationals.

9.2. Markov spectrum. Our final task in this section would be to express the fact that a uniform bound on a_n makes α to be badly approximable in a more quantitative way. Namely, it follows from Lemma 8.4 that the Markov constant $\mu(\alpha)$ of $\alpha \notin \mathbb{Q}$ can be written as

$$\mu(\alpha) = \liminf_{q\to\infty} q\langle \alpha q\rangle = \liminf_{n\to\infty} q_n|q_n\alpha - p_n|\,,$$

which, in view of Proposition 7.6(b), amounts to

$$\mu(\alpha) = \liminf_{n\to\infty} \frac{1}{[a_{n+1}; a_{n+2}, \dots] + [0; a_n, a_{n-1}, \dots, a_1]}\,. \tag{9.4}$$

The latter expression can be very useful in describing real numbers α with large Markov constants. For example, one can use it to prove that

$$\alpha \sim \beta \quad\Rightarrow\quad \mu(\alpha) = \mu(\beta) \tag{9.5}$$

(Exercise 9.7). We are going to investigate the two top values of the Markov spectrum, leaving some more to the reader. Let us introduce the following notation: for an ordered k-tuple of subsets $B_1, \dots, B_k$ of $\mathbb{N}$ let us denote by $\mathcal{I}(B_1, \dots, B_k)$ the set of α such that one has $a_{n+1} \in B_1, \dots, a_{n+k} \in B_k$) for infinitely many n.

LEMMA 9.2. (a) $\alpha \in \mathcal{I}(\{3,\dots\}) \Rightarrow \mu(\alpha) < 1/3$;
(b) $\alpha \in \mathcal{I}(\{2\}) \Rightarrow \mu(\alpha) < 3/8$;
(c) $\alpha \in \mathcal{I}(\{1\},\{2\}) \Rightarrow \mu(\alpha) < 6/17$.

PROOF. Part (a) is immediate since the denominator

$$a_{n+1} + \frac{1}{a_{n+2} + \frac{1}{\cdots}} + \frac{1}{a_n + \frac{1}{\cdots + \frac{1}{a_1}}} \tag{9.6}$$

in the right hand side of (9.4) is at least 3 whenever $a_{n+1} \geq 3$. For (b), we can assume that $\alpha \in \mathcal{I}(\{2\}) \smallsetminus \mathcal{I}(\{3,\dots\})$, and thus have infinitely many n for which $a_{n+1} = 2$ and $\max(a_n, a_{n+2}) \leq 2$. This forces (9.6) to be at least $2 + 1/3 + 1/3 = 8/3$. Likewise, the assumption $\alpha \in \mathcal{I}(\{1\},\{2\})$ yields infinitely many n for which $a_{n+1} = 2$, $a_n = 1$ and $a_{n+2} \leq 2$, forcing (9.6) to be at least $2 + 1/2 + 1/3 = 17/6$. □

COROLLARY 9.3. (a) $\mu(\alpha) = 1/\sqrt{5} \iff \alpha \sim \frac{\sqrt{5}+1}{2} = [1; 1, \dots]$;
(b) $1/\sqrt{5} > \mu(\alpha) \geq 1/\sqrt{8} \iff \mu(\alpha) = 1/\sqrt{8} \iff \alpha \sim \sqrt{2} + 1 = [2; 2, \dots]$.

Proof. Since $1/\sqrt{5}$ is bigger than $3/8$, it follows from parts (a) and (b) of the above lemma that $\mu(\alpha) = 1/\sqrt{5}$ can hold only if α is not in $\mathcal{I}(\{2, \dots\})$, that is, $a_n = 1$ for large enough n. Likewise, $1/\sqrt{5} > \mu(\alpha) \geq 1/\sqrt{8} > 6/17$ implies that $a_n = 2$ for large enough n. □

Pushing the above technique a bit further, it is not hard to keep on computing the top of the Markov spectrum. The next one after $1/\sqrt{8}$ is suggested as an exercise, as well as obtaining a lower bound for $\mu(\alpha)$ over $\alpha \notin \mathcal{I}(\{3, \dots\})$.

9.3. Exercises.

Exercise 9.1. [$*$] Prove that α is badly approximable iff there exists $C > 0$ such that given a sufficiently large Q one can find a relatively prime solution p, q of the inequality (1.7) with $Q < q < CQ$.

Exercise 9.2. [$*$] Fill in the gaps in the proof of Theorem 9.1.

Exercise 9.3. [$*$] Prove that $\alpha \sim \beta$ if and only if $\beta = \frac{a\alpha+b}{c\alpha+d}$ for some $a, b, c, d \in \mathbb{Z}$ with $ad - bc = 1$.

Exercise 9.4. Describe explicitly those quadratic irrationals whose sequence of partial quotients is (a) strictly periodic with period ≤ 2; (b) eventually periodic with period 1.

Exercise 9.5. Compute the partial quotients of $\sqrt{N}$ for small values of N.

Exercise 9.6. [$*$] It follows from Exercise 9.3 that the period of the sequence of partial quotients of a root of a quadratic equation depends solely on the discriminant of the equation. How? Can you bound one in terms of the other?

Exercise 9.7. Prove (9.5).

Exercise 9.8. Justify the third entry in the table of Markov constants in §3.2, and find the tail of the continued fraction expansion of all the numbers α with this value of $\mu(\alpha)$.

Exercise 9.9. Show that the converse to Lemma 9.2(a) is not true, and find the infimum of $\mu(\alpha)$ over $\alpha \notin \mathcal{I}(\{3, \dots\})$ (that is, over all those α for which $a_n \in \{1, 2\}$ for all n). Moreover, show that the infimum is in fact the minimum, and describe the tail of the continued fraction expansion of the numbers where this munimum is attained.

10. Continued fractions and the Gauss map

10.1. Coding of trajectories. We now return to the approach to continued fractions given by (7.4), where the Gauss map T is defined by (7.5). For convenience let us assume throughout this section that $0 \le \alpha \le 1$. It follows from (7.4) that the sequence of partial quotients of an irrational α can be easily read off the sequence of iterates of T applied to α. Indeed, one has $\beta_n = T^n(\alpha)$ for all $n \ge 0$, and hence $a_n = \lfloor 1/T^n(\alpha) \rfloor$ for all $n \in \mathbb{N}$.

More precisely, note that $T(0)$ is (apriori) not defined, and it will be no harm to put $T(0) = 0$. This way we have

$$T^{-1}(0) = \{0\} \cup \{1/k \mid k \in \mathbb{N}\}\,.$$

Furthermore, each of the subintervals $\Delta_k \stackrel{\text{def}}{=} (\frac{1}{k+1}, \frac{1}{k}]$, that is, semi-open intervals bounded by points from $T^{-1}(0)$, is bijectively mapped onto $(0,1]$. It is easy to see (Exercise 10.1) that

$$\alpha \in \mathbb{Q} \Leftrightarrow T^n(\alpha) = 1/k \text{ for some } n, k \Leftrightarrow T^n(\alpha) = 0 \text{ for some } n\,. \tag{10.1}$$

Furthermore, $a_1(\alpha) = \lfloor 1/\alpha \rfloor = k$ if and only if $\alpha \in \Delta_k$, and hence

$$a_n = k \iff T^n(\alpha) \in \Delta_k \iff a_1\big(T^n(\alpha)\big)\,. \tag{10.2}$$

That is, one can generate the sequence of partial quotients of α by reading labels of intervals Δ_k where α is sent by the iterates of the Gauss map.

The map T also can help one to study the sets of numbers whose continued fraction expansion begins with a given finite sequence. More precisely, generalizing the definition of intervals Δ_k, for a fixed n-tuple of natural numbers $\mathbf{a} = (a_1, \dots, a_n)$ define

$$\Delta_{\mathbf{a}} \stackrel{\text{def}}{=} \{\alpha \in [0,1] \mid a_i(\alpha) = a_i,\ i = 1, \dots, n\}\,. \tag{10.3}$$

Lemma 10.1. *$\Delta_{\mathbf{a}}$ is precisely the semi-open interval bounded by $\frac{p_n}{q_n} = [0; a_1, \dots, a_n]$ and $\frac{p_{n-1}+p_n}{q_{n-1}+q_n} = [0; a_1, \dots, a_n, 1]$, which is a bijective image of the unit interval $[0,1]$ under the map*

$$t \mapsto \Psi_{\mathbf{a}}(t) \stackrel{\text{def}}{=} [0; a_1, \dots, a_n + t] = \frac{p_n + tp_{n-1}}{q_n + tq_{n-1}}\,. \tag{10.4}$$

Proof. The first statement follows from the fact that admissible values of $\alpha_{n+1} = \alpha_{n+1}(\alpha)$ for all α run exactly through the half-line $(1, \infty)$ (one-to-one correspondence of Theorem 7.5), and the rational boundary points are obtained by putting $x = 1$ or ∞ in (7.14). It remains to notice that (10.4) can be obtained from (7.14) by means of the substitution $x = 1/t$. □

In particular, the length of $\Delta_{\mathbf{a}}$ is equal to

$$\left| \frac{p_n}{q_n} - \frac{p_{n-1} + p_n}{q_{n-1} + q_n} \right| = \frac{1}{q_n(q_{n-1} + q_n)}\,. \tag{10.5}$$

Note also that whenever $\alpha \in \Delta_{\mathbf{a}}$, one has

$$\alpha = [0; a_1, \dots, a_n + \beta_n] = \Psi_{\mathbf{a}}\big(T^n(\alpha)\big)\,; \tag{10.6}$$

in other words, the inverse of $\Psi_{\mathbf{a}}$ restricted to $\Delta_{\mathbf{a}}$ coincides with T^n.

One can easily notice some similarities between the Gauss map and the map T_m given by (5.1). Both are expanding and piecewise continuous. Also, the complement to $T_m^{-1}(0)$ consists of m components, namely the intervals $(\frac{k}{m}, \frac{k+1}{m})$, and the labels one can generate by applying the iterates of T_m to α are precisely the m-ary digits of α. In other words, the (one-to-one except for countably many points) correspondence between α and its m-ary expansion transforms T_m into the left shift on the space $\{0, \dots, m-1\}^{\mathbb{N}}$ of infinite words in the alphabet $\{0, \dots, m-1\}$, and the (also one-to-one except for countably many points) correspondence between α and its continued fraction expansion transforms T into the left shift on the space $\mathbb{N}^{\mathbb{N}}$ of infinite words in the alphabet $\{1, 2, \dots\}$.

It is not surprising though that the analysis of the Gauss map is somewhat more complicated than that of the map T_m. Indeed, T_m is piecewise linear and finite-to-one, while T is nonlinear and infinite-to one. Yet they have a lot of important features in common.

10.2. Ergodicity. A natural invariant probability measure for T was discovered by Gauss. Namely, the following is true:

Lemma 10.2. *Let μ be a measure absolutely continuous with respect to λ, with density $\frac{1}{\log 2(1+x)}$; that is, for any* $0 \le a < b \le 1$

$$\mu\big((a,b)\big) = \frac{1}{\log 2}\int_a^b \frac{dx}{1+x} = \frac{1}{\log 2}\log\frac{1+b}{1+a}\,. \tag{10.7}$$

Then $([0,1], \mu, T)$ *is a measure preserving system.*

Proof. Note that the T-preimage of any interval $A = (a,b)$ consists of countably many intervals; more precisely, of the intervals $(\frac{1}{n+b}, \frac{1}{n+a})$, $n \in \mathbb{N}$. Therefore

$$\begin{aligned}\mu\big(T^{-1}(A)\big) &= \mu\left(\bigcup_{n=1}^{\infty}\Big(\frac{1}{n+b}, \frac{1}{n+a}\Big)\right) = \frac{1}{\log 2}\sum_{n=1}^{\infty}\log\frac{1+\frac{1}{n+a}}{1+\frac{1}{n+b}}\\ &= \frac{1}{\log 2}\sum_{n=1}^{\infty}\left(\log\frac{1+n+a}{n+a} - \log\frac{1+n+b}{n+b}\right),\end{aligned}$$

and the latter sum telescopes to the right hand side of (10.7). □

Similarly to Proposition 5.2, we have

Proposition 10.3. $([0,1], \mu, T)$ *is ergodic.*

Proof. We basically follow the strategy of the proof of Proposition 5.2. There we considered dyadic intervals, that is, the sets of numbers α whose binary expansion starts with a given finite string. Here we will work with intervals $\Delta_{\mathbf{a}}$ instead. Also, since μ is absolutely continuous with respect to λ with density bounded between two positive constants, it suffices to show that any T-invariant $A \subset [0,1]$ is either null or conull with respect to λ.

For $0 \le x < y \le 1$ and any $\mathbf{a} = (a_1, \dots, a_n)$, one can write, in view of (10.6):

$$\lambda\left(T^{-n}\big([x,y)\big) \cap \Delta_{\mathbf{a}}\right) = |\Psi_{\mathbf{a}}(y) - \Psi_{\mathbf{a}}(x)| = \left|\frac{p_n + yp_{n-1}}{q_n + yq_{n-1}} - \frac{p_n + xp_{n-1}}{q_n + xq_{n-1}}\right|$$
$$= \frac{y-x}{(q_n + xq_{n-1})(q_n + yq_{n-1})} \underset{(10.5)}{=} (y-x)\lambda(\Delta_{\mathbf{a}})\frac{q_n(q_{n-1}+q_n)}{(q_n + xq_{n-1})(q_n + yq_{n-1})}\,.$$

The last factor in the right hand side is between 1/2 and 2, hence

$$\frac{1}{2}\lambda(\Delta_{\mathbf{a}})\lambda\big([x,y)\big) \le \lambda\left(T^{-n}\big([x,y)\big) \cap \Delta_{\mathbf{a}}\right) \le 2\lambda(\Delta_{\mathbf{a}})\lambda\big([x,y)\big)\,.$$

But intervals $[x, y)$ generate the σ-algebra, therefore

$$\frac{1}{2}\lambda(\Delta_{\mathbf{a}})\lambda(A) \le \lambda\left(T^{-n}(A) \cap \Delta_{\mathbf{a}}\right) \le 2\lambda(\Delta_{\mathbf{a}})\lambda(A)\big)$$

for any measurable subset A of $[0, 1]$. If, in addition, A is T-invariant, then

$$\frac{1}{2}\lambda(\Delta_{\mathbf{a}})\lambda(A) \le \lambda\left(A \cap \Delta_{\mathbf{a}}\right) \le 2\lambda(\Delta_{\mathbf{a}})\lambda(A)\big)\,.$$

Observe that this is basically an analogue of equality (5.2), obtained in the course of the proof of ergodicity of T_m. Again, all intervals $\Delta_{\mathbf{a}}$ generate the σ-algebra, thus for any two measurable A, B one has

$$\frac{1}{2}\lambda(B)\lambda(A) \le \lambda\left(A \cap B\right) \le 2\lambda(B)\lambda(A)\big)\,.$$

By choosing $B = [0, 1] \smallsetminus A$ we obtain that $\lambda(A)\lambda([0, 1] \smallsetminus A) = 0$ □

10.3. Consequences of ergodicity. The fact that T is ergodic, together with Birkhoff's Ergodic Theorem, more precisely in the form of Corollary 5.9, can be exploited to establish various properties of the continued fraction expansion of almost every α.

We list them in the following

THEOREM 10.4. *For almost every (with respect to λ or μ) $\alpha = [0; a_1, a_2, \dots]$ the following holds:*

(a) *Each $k \in \mathbb{N}$ appears in the sequence $a_1, a_2, \dots$ with asymptotic frequency equal to $\mu(\Delta_k) = \frac{1}{\log 2}\log\frac{(k+1)^2}{k(k+2)}$;*
(b) $\lim_{n\to\infty} \frac{a_1+\cdots+a_n}{n} = \infty$*;*
(c) $\lim_{n\to\infty}(a_1 \cdots a_n)^{1/n} = \prod_{k=1}^{\infty}\left(\frac{(k+1)^2}{k(k+2)}\right)^{\log k/\log 2}$*;*
(d) $\lim_{n\to\infty} \frac{\log q_n}{n} = \frac{\pi^2}{12\log 2} = -\lim_{n\to\infty}\frac{|q_n\alpha - p_n|}{n}$.

Some comments are in order. Part (b) significantly strengthens Corollary 4.2: not only almost every α has unbounded partial quotients, but even their averages go to infinity. On the other hand, the geometric mean has a limit, approximately equal to 2.685 and known as Khintchine's constant. The limit in part (d) provides the growth rate of p_n and q_n for generic α mentioned after Proposition 7.8. Its exponential, that is, the value of $\lim_{n\to\infty} q_n^{1/n}$ for a.e. α, is usually referred to as the Khintchine-Levy constant.

Proof. For part (a), let f be the characteristic function of Δ_k Then, by (10.2), the asymptotic frequency of k is equal to

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f\big(T^k(\alpha)\big) = \bar{f}(\alpha) \underset{\text{Corollary 4.2}}{=} \int_0^1 f\,d\mu = \mu(\Delta_k)\,.$$

Part (b) follows in a similar way from Exercise 5.14. Here we take $f(x) = \lfloor 1/x \rfloor = a_1(x)$. Clearly $\int_0^1 f\,d\mu = \infty$, and it is immediate from (10.2) that the left hand side in (b) is equal to $\bar{f}(\alpha)$, which is infinite for a.e. α.

Part (c) can be attacked by considering the logarithm of the geometric mean of the coefficients a_k. Then the relevant function to consider is $f(x) = \log a_1(x)$ which is easily seen to be in L^1. Thus the logarithm of the right hand side in (c) is equal to $\int_0^1 f\,d\mu = \sum_{k=1}^{\infty} \mu(\Delta_k)\log k$.

The last part of the theorem is trickier. The reason is that the left hand side does not quite reduce to the Birkhoff-type sum, but to something which can be approximated by the latter. First observe that for any irrational α and any n

$$\begin{aligned}\frac{p_n(\alpha)}{q_n(\alpha)} &= \frac{1}{a_1 + [0; a_2, \ldots, a_n]} = \frac{1}{a_1 + \frac{p_{n-1}(T(\alpha))}{q_{n-1}(T(\alpha))}} \\ &= \frac{q_{n-1}\big(T(\alpha)\big)}{p_{n-1}\big(T(\alpha)\big) + a_1 q_{n-1}\big(T(\alpha)\big)}\,.\end{aligned} \tag{10.8}$$

But since the leftmost and the rightmost fractions above are both irreducible, it follows that their respective numerators and denominators are equal; in particular, $p_n(\alpha) = q_{n-1}\big(T(\alpha)\big)$ (Exercise 10.3). Hence $p_n = q_{n-1}\circ T$ as functions (for almost every value of the variable). Since $p_1 \equiv 1$, it follows that

$$\frac{1}{q_n} = \frac{p_n}{q_n}\cdot\frac{p_{n-1}\circ T}{q_{n-1}\circ T}\cdots\frac{p_1\circ T^{n-1}}{q_1\circ T^{n-1}}\,.$$

The above formula allows one to express $\log q_n$ as a sum of certain terms related to the iterates of T. Namely, one has

$$\begin{aligned}-\frac{1}{n}\log q_n(\alpha) &= \frac{1}{n}\sum_{k=0}^{n-1}\log\frac{p_{n-k}\big(T^k(\alpha)\big)}{q_{n-k}\big(T^k(\alpha)\big)} \\ &= \frac{1}{n}\sum_{k=0}^{n-1}\log T^k(\alpha) - \frac{1}{n}\sum_{k=0}^{n-1}\left(\log T^k(\alpha) - \log\frac{p_{n-k}\big(T^k(\alpha)\big)}{q_{n-k}\big(T^k(\alpha)\big)}\right)\,.\end{aligned} \tag{10.9}$$

The logic behind the last step is that the first term in the right hand side above, in view of the ergodic theorem, converges to

$$\int_0^1 \log x \, d\mu(x) = \int_0^1 \log x \, d\mu(x) = \frac{1}{\log 2} \int_0^1 \frac{\log x}{1+x} \, dx \,,$$

which can be shown to be equal to $-\pi^2/12 \log 2$ (Exercise 10.4). On the other hand, the second summand is an average of differences between logarithms of numbers quite close to each other, namely, $T^k(\alpha)$ and its $(n-k)$-th convergent. So there is hope that it can be shown to tend to zero. And indeed this is the case; details are left to the reader, see Exercise 10.5. It remains to observe that the second equality in (d) immediately follows from Lemma 8.2. □

10.4. Exercises.

Exercise 10.1. Prove (10.1).

Exercise 10.2. Describe, both explicitly and in terms of their continued fraction expansions, all the fixed points of T.

Exercise 10.3. Explain why it follows from the equality in (10.8) that for any n the respective numerators and denominators of the fractions on both sides are the same.

Exercise 10.4. This is an exercise involving nothing but elementary calculus - show that $\int_0^1 \frac{\log x}{1+x} \, dx = -\pi^2/12$.

Exercise 10.5. Show that the sum appearing in the second summand in the right hand side of (10.9) is uniformly bounded, thus completing the proof of Theorem 10.4.