

## Chapter 6

# Arithmetic Properties of Apollonian Gaskets

Here we study some arithmetic questions arising when we consider the curvatures of disks which constitute an Apollonian gasket.

### 6.1 The Structure of $\overline{\mathbb{Q}}$

Here we want to investigate the set  $P^1(\mathbb{Q}) = \overline{\mathbb{Q}}$  of rational numbers including the infinite point  $\infty$ . It can be called a *rational circle*.

First, think about how to parameterize  $\overline{\mathbb{Q}}$ . Every number  $r \in \overline{\mathbb{Q}}$  can be written in the form  $\frac{p}{q}$ , where  $p, q \in \mathbb{Z}$ . But the map  $\alpha : \mathbb{Z} \times \mathbb{Z} \rightarrow \overline{\mathbb{Q}}, \alpha(p, q) = \frac{p}{q}$ , is surjective but by no means injective.

We can impose the condition  $\gcd(p, q) = 1$ , that is, that  $p$  and  $q$  be relatively prime, or in other words, that the fraction  $\frac{p}{q}$  be in lowest terms. Note, however, that the set  $X$  of relatively prime pairs  $(p, q)$  is itself a rather complicated object. The map  $\alpha$ , restricted to  $X$ , will be “two-to-one”:  $\alpha^{-1}(r) = \pm(p, q)$ . And there is no natural way to choose exactly one representative from every pair  $\{(p, q), (-p, -q)\}$ . However, for all  $r = \frac{p}{q} \in \mathbb{Q}$ , we can assume  $q > 0$ . But for  $q = 0$ , there is no preference between  $p = \pm 1$ .

*Remark 6.1.* For the analytically minded reader, we can say that the construction here is similar to the Riemann surface of the function  $f(w) = \sqrt{w}$ . The map  $z \mapsto w = z^2$  has two preimages for each  $w \in \mathbb{C}^\times$ , but this double-valued function does not admit an analytic (or even continuous) single-valued branch.



*Remark 6.2.* A remarkable way to label all positive rational numbers was discovered recently by Neil Calkin and Herbert Wilf (“Recounting the Rationals,” *American Mathematical Monthly* 107 (2000), pp. 360–363). Let  $\mathbf{b}(n)$  be the number of partitions of an integer  $n \geq 0$  into powers of 2, no power of 2 being used more

than twice. Then the ratio  $r_n = \frac{b(n)}{b(n+1)}$  takes every positive rational value exactly once! The initial piece of this numeration is given in the following table:

$n$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$b(n)$	1	1	2	1	3	2	3	1	4	3	5	2	5	3	4	1	5	4
$r_n$	1	$\frac{1}{2}$	2	$\frac{1}{3}$	$\frac{3}{2}$	$\frac{2}{3}$	3	$\frac{1}{4}$	$\frac{4}{3}$	$\frac{3}{5}$	$\frac{5}{2}$	$\frac{2}{5}$	$\frac{5}{3}$	$\frac{3}{4}$	4	$\frac{1}{5}$	$\frac{5}{4}$	$\frac{4}{7}$

$n$	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	
$b(n)$	7	3	8	5	7	2	7	5	8	3	7	4	5	1	6	5	4
$r_n$	$\frac{7}{3}$	$\frac{3}{8}$	$\frac{8}{5}$	$\frac{5}{7}$	$\frac{7}{2}$	$\frac{2}{7}$	$\frac{7}{5}$	$\frac{5}{8}$	$\frac{8}{3}$	$\frac{3}{7}$	$\frac{7}{4}$	$\frac{4}{5}$	5	$\frac{1}{6}$	$\frac{6}{5}$	$\frac{5}{9}$	

It is of interest to compare this numeration with the one given by Farey series (see below).



Our next step in the study of  $\overline{\mathbb{Q}}$  is the introduction of a natural distance between points. In the following, we tacitly assume that all rational numbers are written in lowest terms.

Let us call two numbers  $r_i = \frac{p_i}{q_i}$ ,  $i = 1, 2$ , from  $\overline{\mathbb{Q}}$  **friendly** if the following equivalent conditions are satisfied:

$$(a) |p_1q_2 - p_2q_1| = 1, \quad (b) |r_1 - r_2| = \frac{1}{|q_1q_2|}. \tag{6.1.1}$$

It is worth mentioning that the friendship relation **is not** an equivalence relation:<sup>1</sup> every integer  $k$  is friendly to  $\infty$ , but only neighboring integers are friendly to each other.

Note that the group  $\text{PGL}(2, \mathbb{Z})$  acts on  $\overline{\mathbb{Q}}$  by fractional linear transformations and that this action preserves the friendship relation. We can often use this fact in our study.

**Lemma 6.1.** *The group  $\text{PSL}(2, \mathbb{Z})$  acts simply transitively on the set of all ordered pairs of friendly numbers from  $\overline{\mathbb{Q}}$ . The group  $\text{PGL}(2, \mathbb{Z})$  acts transitively but with a nontrivial stabilizer isomorphic to  $\mathbb{Z}_2$ .*

*Proof.* Let  $r_i = \frac{p_i}{q_i}$ ,  $i = 1, 2$ , be a pair of friendly numbers. Assume for definiteness that  $p_1q_2 - p_2q_1 = 1$ . We have to show that there is a unique element  $\gamma$  of  $\text{PSL}(2, \mathbb{Z})$

---

<sup>1</sup>Just as in the real life.

that sends the standard friendly pair  $(\infty, 0)$  to the given pair  $(r_1, r_2)$ . Let  $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$  be a representative of  $\gamma$  in  $\text{SL}(2, \mathbb{Z})$ . Then we have  $\gamma(0) = \frac{b}{d}$ ,  $\gamma(\infty) = \frac{a}{c}$ .

The conditions  $\gamma(\infty) = r_1$ ,  $\gamma(0) = r_2$  imply  $(a, c) = k_1 \cdot (p_1, q_1)$ ,  $(b, d) = k_2 \cdot (p_2, q_2)$ . Therefore,  $1 = \det g = ad - bc = k_1 k_2 \cdot (p_1 q_2 - p_2 q_1)^{-1} = k_1 k_2$  and  $k_1 = k_2 = \pm 1$ . Hence  $g = \pm \begin{pmatrix} p_1 & p_2 \\ q_1 & q_2 \end{pmatrix}$  is determined up to sign and defines the unique element of  $\text{PSL}(2, \mathbb{Z})$ .

The stabilizer of the pair  $(0, \infty)$  in  $\text{PGL}(2, \mathbb{Z})$  consists of classes of matrices  $\begin{pmatrix} 1 & 0 \\ 0 & \pm 1 \end{pmatrix}$ . □

**Exercise 6.1.** Describe all numbers that are (a) friendly to 0; (b) friendly to  $\infty$ ; (c) friendly to 1.

We define a **distance** in the set  $\overline{\mathbb{Q}}$  in the following way. Given two numbers  $r'$  and  $r''$ , denote by  $d(r', r'')$  the minimal  $n \in \mathbb{Z}_+$  for which there exists a chain  $r' = r_0, r_1, \dots, r_{n-1}, r_n = r''$  such that for all  $k$ , the number  $r_k$  is friendly to  $r_{k \pm 1}$  for  $1 \leq k \leq n - 1$ .

**Exercise 6.2.** (a) Show that  $(\overline{\mathbb{Q}}, d)$  is a discrete metric space on which the group  $\text{PGL}(2, \mathbb{Z})$  acts by isometries.  
 (b) Find the stabilizer of the point  $\infty$ .

**Answer.** (b) The group  $\text{Aff}(1, \mathbb{Z})$  of transformations  $r \mapsto ar + b$ ,  $a = \pm 1, b \in \mathbb{Z}$ .

**Exercise 6.3.** Compute the distances (a)  $d(\infty, n)$ ; (b)  $d(0, n)$ ; (c)  $d(0, \frac{5}{8})$ .

**Answer.** (a) 1; (b) 0 for  $n = 0$ , 1 for  $n = \pm 1$ , 2 for  $|n| > 1$ ; (c) 4.

**Exercise 6.4.** (a) Show that for every  $r', r'' \in \overline{\mathbb{Q}}$ , the distance  $d(r', r'')$  is finite.  
 (b) Is the metric space  $\overline{\mathbb{Q}}$  bounded?

**Answer.** (a) See Theorem 6.1 below; (c) No.

Rather interesting and nontrivial problems arise when we consider the geometry of balls and spheres in  $\overline{\mathbb{Q}}$ . As usual, we define a **ball** with center  $a$  and radius  $r$  as the set  $B_r(a) = \{b \in \overline{\mathbb{Q}} \mid d(a, b) \leq r\}$ . Analogously, a **sphere** is the set  $S_r(a) = \{b \in \overline{\mathbb{Q}} \mid d(a, b) = r\}$ .

**Theorem 6.1.** *The ball  $B_n(\infty)$  consists of all rational numbers that can be written as a continued fraction of length  $n$ , i.e., as*

$$r = k_1 + \frac{1}{k_2 + \frac{1}{k_3 + \frac{1}{\ddots k_{n-1} + \frac{1}{k_n}}}} \tag{6.1.2}$$

where  $k_i$  are arbitrary integers (positive or negative).

*proof.* First of all, let us show that for every  $r$  of the form (6.1.2), the distance  $d(\infty, r)$  does not exceed  $n$ . We shall do this by induction on  $n$ .

For  $n = 1$ , the result follows from Exercise 6.3. Assume that the theorem is true for all continued fractions of length  $\leq n - 1$  and consider a continued fraction of length  $n$  given by Eq. (6.1.2). Denote by  $r'$  the number  $\frac{1}{r-k_1}$ . It is clear that  $r'$  is represented by a continued fraction of length  $n - 1$ , whence  $d(\infty, r') \leq n - 1$ . Now, from the invariance of the distance with respect to shifts  $r \mapsto r + k$ ,  $k \in \mathbb{Z}$ , and with respect to the inversion  $r \mapsto r^{-1}$ , we have

$$d(\infty, r) = d(\infty, r - k_1) = d(0, r') \leq d(0, \infty) + d(\infty, r') \leq 1 + (n - 1) = n.$$

The first sign  $\leq$  is just the triangle inequality, and the second follows from Exercise 6.3(a) and from the induction hypothesis.  $\square$

The structure of spheres is a more delicate question. The “complexity” of a sphere grows with its radius.

For instance, consider  $S_1(\infty) = \mathbb{Z}$ . It is a homogeneous space with respect to the group  $\text{Aff}(1, \mathbb{Z})$ , which plays the role of the “group of rotations” around the infinite point; see Exercise 6.3(a).

The sphere  $S_2(\infty)$  consists of points  $k_1 + \frac{1}{k_2}$ , where  $k_1, k_2 \in \mathbb{Z}$  and  $k_2 \neq 0, \pm 1$ . Under the action of  $\text{Aff}(1, \mathbb{Z})$ , it splits into infinitely many orbits  $\Omega_m$ , enumerated by the number  $m = |k_2| \geq 2$ . The stabilizer of the point  $k + \frac{1}{m} \in \Omega_m$  is trivial for  $m > 2$  and contains one nonunit element  $r \mapsto 2k + 1 - r$  for  $m = 2$ .

**Problem 6.1.** Describe the orbits of  $\text{Aff}(1, \mathbb{Z})$  on the sphere  $S_k(\infty)$  for  $k > 2$ .

## 6.2 Rational Parameterization of Circles

It is well known that a circle as a real algebraic manifold is rationally equivalent to a real projective line. This means that one can establish a bijection between a circle and a projective line using rational functions with rational coefficients.

For example, the circle  $x^2 + y^2 = 1$  can be identified with a projective line with parameter  $t$  as follows:

$$x = \frac{t^2 - 1}{t^2 + 1}, \quad y = \frac{2t}{t^2 + 1}; \quad t = \frac{y}{1 - x} = \frac{1 + x}{y}. \quad (6.2.1)$$

In particular, as  $t$  runs through all rational numbers (including  $\infty$ ), the corresponding points  $(x, y)$  run through all rational points<sup>2</sup> of the circle.

From this fact one can derive the well-known description of primitive integral solutions to the equation  $x^2 + y^2 = z^2$ . Namely, in every primitive solution, exactly

---

<sup>2</sup>That is, points with rational coordinates.

one of the numbers  $x$ ,  $y$  is even. Assume that it is  $y$ ; then there are relatively prime numbers  $a$ ,  $b$  such that

$$x = a^2 - b^2, \quad y = 2ab, \quad \pm z = a^2 + b^2. \quad (6.2.2)$$

Analogously, the projectivization of the future light cone in  $\mathbb{R}^{1,3}$  is nothing but the two-dimensional sphere, which is rationally equivalent to a completed two-dimensional plane. Therefore, all future light vectors  $(t, x, y, z)$  with integral nonnegative coefficients can be written, up to positive proportionality, in the form

$$t = k^2 + l^2 + m^2, \quad x = 2km, \quad y = 2lm, \quad z = |k^2 + l^2 - m^2|. \quad (6.2.3)$$

I do not know whether any integral solution can be written exactly in the form (6.2.3) for some integers  $k, l, m$  with  $\gcd(k, l, m) = 1$ .<sup>3</sup>

Next, we take into account that on the real projective line  $\overline{\mathbb{R}}$ , there is a natural orientation. For our goals, it is convenient to define the orientation as a cyclic order for every three distinct points  $x_1, x_2, x_3 \in \overline{\mathbb{R}}$ . Geometrically, this order means that going from  $x_1$  to  $x_3$  in the positive direction, we pass  $x_2$  on our way. We shall also use the expression “ $x_2$  is between  $x_1$  and  $x_3$ .” Note that in this case,  $x_2$  is *not* between  $x_3$  and  $x_1$ .

**Exercise 6.5.** (a) Show that in the case that all  $x_1, x_2, x_3$  are finite (i.e.,  $\neq \infty$ ), the statement “ $x_2$  is between  $x_1$  and  $x_3$ ” is equivalent to the inequality

$$(x_1 - x_2)(x_2 - x_3)(x_3 - x_1) > 0.$$

(b) Which of the following are true?

- (i) 1 is between 0 and  $\infty$ ;
- (ii)  $\infty$  is between 0 and 1;
- (iii)  $-1$  is between 0 and  $\infty$ .

We now introduce a new operation<sup>4</sup> of “inserting” on  $\overline{\mathbb{R}}$ . It associates to an ordered pair of rational numbers  $(r_1, r_2)$  a third number, denoted by  $r_1 \downarrow r_2$ , such that

$$r_1 \downarrow r_2 := \frac{p_1 + p_2}{q_1 + q_2}, \quad \text{if } r_1 = \frac{p_1}{q_1}, r_2 = \frac{p_2}{q_2}, \quad (6.2.4)$$

where the signs of  $p_i$  and  $q_i$  are chosen such that  $r_1 \downarrow r_2$  is between  $r_1$  and  $r_2$ .

**Exercise 6.6.** Compute the following expressions:

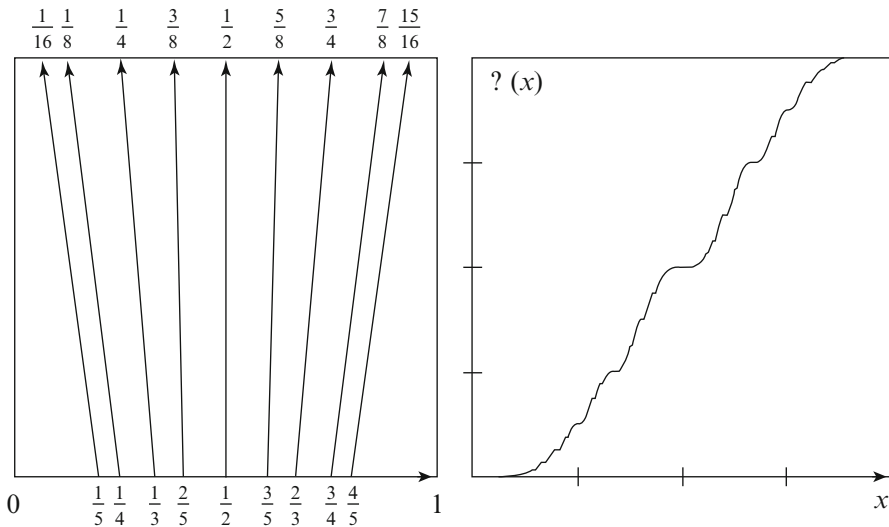
- (a)  $0 \downarrow \infty$ ; (b)  $\infty \downarrow 0$ ; (c)  $\infty \downarrow -2$ ; (d)  $1 \downarrow 2$ ; (e)  $2 \downarrow 1$ ; (f)  $\frac{1}{2} \downarrow -\frac{1}{3}$ .

**Answer.** (a) 1; (b)  $-1$ ; (c)  $-3$ ; (d)  $\frac{3}{2}$ ; (e)  $\infty$ ; (f)  $-2$ .

<sup>3</sup>As one of the reviewers pointed out, the quadruple  $(7, 2, 3, 6)$  is a counterexample, since 7 is not a sum of three squares.

<sup>4</sup>I learned from R. Borcherds that this operation is known to mathematicians in England as “English major addition.” It is also the subject of one of the standard jokes quoted in Gelfand’s seminar.





**Fig. 6.1** Graph of the function  $?$

**Exercise 6.9.** Show that  $f_k^{(n)} = f_{2k}^{(n+1)}$ , so that  $f_k^{(n)}$  actually depends only on the dyadic number  $r = \frac{k}{2^n}$ . Sometimes, we shall write  $f_r$  instead of  $f_k^{(n)}$ .

To simplify the exposition, let us consider the part of  $F^{(n)}$  between 0 and 1, i.e., members  $f_r$  with  $r$  between 0 and 1.

Note that if we change the procedure and insert between any two numbers  $a, b$  not  $a \downarrow b$ , but the arithmetic mean value  $\frac{a+b}{2}$ , we obtain at the  $n$ th step, the arithmetic progression with  $2^n + 1$  terms starting with 0 and ending by 1. The  $k$ th member of this progression is  $a_k^{(n)} = \frac{k}{2^n}$ , or in the same notation as above,  $a_r = r$  (Fig. 6.1).

Now we are prepared to define a remarkable function first introduced by Hermann Minkowski. He called it the “question mark function” and denoted it by  $?(x)$ ; see **Info E** in Part I.

**Theorem (Minkowski’s theorem).** *There exists a unique continuous and strictly increasing function  $?: [0, 1] \rightarrow [0, 1]$  such that*

$$?(a \downarrow b) = \frac{?(a) + ?(b)}{2} \quad \text{for all friendly rational numbers } a, b \in [0, 1]. \tag{6.2.5}$$

*Sketch of proof.* The formula (6.2.5) and induction over  $n$  imply that if the desired function exists, it must have the property  $?(f_k^{(n)}) = a_k^{(n)}$ . It follows that  $?(f_r) = r$  for all  $r \in \mathbb{Z}[\frac{1}{2}] \cap [0, 1]$ .

On the other hand, we can define  $?$  on  $\mathbb{Z}[\frac{1}{2}]$  by the formula  $?(f_r) = r$ . Since both sets  $\{f_k^{(n)}\}$  and  $\{a_k^{(n)}\}$  are dense in  $[0, 1]$ , the function can be extended uniquely as a monotone function from  $[0, 1]$  to  $[0, 1]$ . For example, we can put

$$?(x) = \lim_{n \rightarrow \infty} ?(x_n), \tag{6.2.6}$$

where  $\{x_n\}$  is a monotone sequence of rational numbers converging to  $x$ . □

The inverse function  $p$  to the question mark function solves the problem of computing  $f_k^{(n)}$  posed above, since for every dyadic  $r \in [0, 1]$ , we have  $f_r = p(r)$ .

On the set  $\mathbb{Z}[\frac{1}{2}] \cap [0, 1]$  of binary fractions, the function  $p(x)$  can be computed step by step using the property

$$p\left(\frac{2k+1}{2^{n+1}}\right) = p\left(\frac{k}{2^n}\right) \downarrow p\left(\frac{k+1}{2^n}\right), \tag{6.2.7}$$

which follows immediately from Eq. (6.2.5), and repeating the construction of the modified Farey series.

**Theorem 6.2.** *The function  $p := ?^{-1}$  has the following properties:*

1. (a)  $p(1-x) = 1-p(x)$ ; (b)  $p(\frac{x}{2}) = \frac{p(x)}{1+p(x)}$ ; (c)  $p(\frac{1+x}{2}) = \frac{1}{2-p(x)}$ .
2.  $(p)'(\frac{k}{2^n}) = \infty$  for every  $n$  and  $0 \leq k \leq 2^n$ .
3. For every rational nondyadic number  $r \in [0, 1]$ , the value  $p(r)$  is a quadratic irrationality, i.e., has a form  $r_1 + \sqrt{r_2}$  for some rational  $r_1, r_2$ .
4. We have the following remarkable formula:

$$p\left(\underbrace{0.0\dots 00}_{k_1} \underbrace{11\dots 11}_{l_1} \dots \underbrace{00\dots 00}_{k_n} \underbrace{11\dots 11}_{l_n} \dots\right) = \frac{1}{k_1 + \frac{1}{l_1 + \frac{1}{\ddots k_n + \frac{1}{l_n + \frac{1}{\ddots}}}}}} \tag{6.2.8}$$

where on the left-hand side, the binary system is used, while on the right-hand side, we use a continued fraction. The formula (6.2.8) works also for finite binary fractions.<sup>5</sup>

*Sketch of proof.* The relations 1(a)–(c) can be derived from the following useful fact.

---

<sup>5</sup>Guess about the form of the right-hand side of the formula in this case.



**Lemma 6.2.** *Let  $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{GL}(2, \mathbb{Z})$ . Then the transformation of  $\overline{\mathbb{Q}}$  given by*

$$r \mapsto g \cdot r := \frac{ar + b}{cr + d}$$

*commutes with the insertion operation  $\downarrow$ , i.e.,*

$$(g \cdot r_1) \downarrow (g \cdot r_2) = g \cdot (r_1 \downarrow r_2). \tag{6.2.9}$$

We leave the proof of this claim to the reader and make only two useful remarks, each of which could serve as the basis for a proof.

1. The transformations in question send friendly pairs to friendly pairs.
2. The group  $\text{GL}(2, \mathbb{Z})$  is generated by two elements:

$$g_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad g_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Now we prove relation 1(a). Consider the following diagram:

$$\begin{array}{ccc} [0, 1] & \xrightarrow{x \mapsto 1-x} & [0, 1] \\ p \downarrow & & \downarrow p \\ [0, 1] & \xrightarrow{x \mapsto 1-x} & [0, 1] \end{array} \tag{6.2.10}$$

Relation 1(a) claims that it is commutative. To check this, choose a point  $x \in [0, 1]$  that is a dyadic fraction  $r = \frac{k}{2^n} = a_r$ . Then the vertical arrow sends this number to  $p(a_r) = f_r$ , and the horizontal arrow sends  $f_r$  to  $f_{1-r}$  (check this, consulting the table above).

On the other hand, the horizontal arrow sends  $r$  to  $1 - r = a_{1-r}$ , and then the vertical arrow sends  $a_{1-r}$  to  $f_{1-r}$ . Thus for every number of the form  $\frac{k}{2^n}$ , relation 1(a) holds. By continuity, it holds everywhere.

Consider relation 1(b). It is equivalent to the commutativity of the diagram

$$\begin{array}{ccc} [0, 1] & \xrightarrow{x \mapsto x/2} & [0, \frac{1}{2}] \\ p \downarrow & & \downarrow p \\ [0, 1] & \xrightarrow{x \mapsto \frac{x}{1+x}} & [0, \frac{1}{2}] \end{array} \tag{6.2.11}$$

Here again we start with an element  $r = a_r \in [0, 1]$ . The horizontal arrow sends it to  $a_{r/2}$ , and then the vertical arrow transforms it to  $f_{r/2}$ .

On the other hand, the vertical arrow sends  $a_r$  to  $f_r$ , and we have to show that the horizontal arrow transforms it to  $f_{r/2}$ . That is, we want to check the equality

$\frac{f_r}{1+f_r} = f_{r/2}$ . For this, we observe that the transformation  $x \mapsto \frac{x}{1+x}$  maps the segment  $[0, 1]$  to the segment  $[0, \frac{1}{2}]$ . Since it belongs to  $\text{PGL}(2, \mathbb{Z})$ , it transforms the part of Farey series between  $f_0$  and  $f_1$  to the pair between  $f_0$  and  $f_{1/2}$ . Then by induction on  $n$ , we check that it sends  $f_{\frac{2^k}{2^n}}$  to  $f_{\frac{k}{2^n}}$ .

The relation 1(c) can be proved in the same way using the diagram

$$\begin{array}{ccc}
 [0, 1] & \xrightarrow{x \mapsto \frac{1+x}{2}} & [\frac{1}{2}, 1] \\
 p \downarrow & & \downarrow p \\
 [0, 1] & \xrightarrow{x \mapsto \frac{1}{2-x}} & [\frac{1}{2}, 1]
 \end{array} \tag{6.2.12}$$

The point is that affine transformations respect half-sums, while the transformations from  $\text{PGL}(2, \mathbb{Z})$  respect the insertion operation.

I recommend that the reader formulate and prove some other properties of  $?$  and  $p$  using other diagrams.

It is also useful to extend the definition of  $?$  and  $p$  to the whole set  $\overline{\mathbb{R}}$  by the formulas

$$p\left(\frac{1}{x}\right) = \frac{1}{p(x)}; \quad p(-x) = -p(x). \tag{6.2.13}$$

We shall verify property 2 only at the point  $x = 0$ . The general case  $x = \frac{k}{2^n}$  can be done similarly, or it can be reduced to the case  $x = 0$  by 1(a)–1(c).

We have  $p(0) = 0$ ,  $p(\frac{1}{2^n}) = \frac{1}{n+1}$ . So if  $\frac{1}{2^n} \leq \Delta x \leq \frac{1}{2^{n-1}}$ , we have  $\frac{1}{n+1} \leq \Delta p \leq \frac{1}{n}$ .

Therefore,  $\frac{2^{n-1}}{n+1} \leq \frac{\Delta p}{\Delta x} \leq \frac{2^n}{n}$  for  $\frac{1}{2^n} \leq \Delta x \leq \frac{1}{2^{n-1}}$  and  $p'(0) = +\infty$ .

Statement 3 follows from the formula (6.2.8). As for this formula, it can be proved for finite fractions by induction using the Farey series. Note that in the last section of Part I, we used Eq. (6.2.8) as a definition of the question mark function.

□

*Remark 6.3.* Let us interpret the function  $p := ?^{-1}$  as a distribution function for a probability measure  $\mu$  on  $[0, 1]$ : the measure of an interval  $[a, b]$  is equal to  $p(b) - p(a)$ . This measure is a weak limit<sup>6</sup> of the sequence of discrete measures  $\mu_n$ ,  $n \geq 1$ , concentrated on the subset  $F^{(n)}$ , so that the point  $f_k^{(n)}$  has the mass  $\frac{1}{2^n}$  for  $1 \leq k \leq 2^n$ .

It is clear that the support of  $\mu$  is the whole segment  $[0, 1]$  (i.e., the measure of every interval  $(a, b) \subset [0, 1]$  is positive). While for an ordinary Farey series, the measure defined in a similar way is uniform, in our case it is far from it. The detailed study of this measure is a very promising subject (see, e.g., [de Rha59]).

♡

---

<sup>6</sup>We say that a measure  $\mu$  on  $[0, 1]$  is a weak limit of the sequence of measures  $\mu_n$  if for every continuous function  $f$  on  $[0, 1]$ , we have  $\lim_{n \rightarrow \infty} \int_0^1 f(x) d\mu_n(x) = \int_0^1 f(x) d\mu(x)$ .

**Exercise 6.10.** Find the values of  $?(x)$  and  $?'(x)$  at the point  $x = \frac{1}{3}$ .

*Hint.* Using the relation  $\frac{1}{2} - \frac{1}{4} + \frac{1}{8} - \frac{1}{16} + \frac{1}{32} - \frac{1}{64} + \dots = \frac{1}{3}$ , show that

$$? \left( \frac{1}{3} - \frac{1}{3 \cdot 4^n} \right) = \frac{\Phi_{2n-1}}{\Phi_{2n+1}}, \quad ? \left( \frac{1}{3} + \frac{2}{3 \cdot 4^n} \right) = \frac{\Phi_{2n}}{\Phi_{2n+2}},$$

where  $\Phi_n$  is the  $n$ th Fibonacci number, given by the formula

$$\Phi_n = \frac{\phi^n - (-\phi)^{-n}}{\phi + \phi^{-1}},$$

where  $\phi = \frac{\sqrt{5}+1}{2} \approx 1.618\dots$  is the golden ratio.<sup>7</sup>

**Answer.**  $?\left(\frac{1}{3}\right) = \frac{3-\sqrt{5}}{2}$ ;  $?'\left(\frac{1}{3}\right) = 0$ .

**Problem 6.2.** Is it true that  $?'(x) = 0$  for all rational numbers except  $a_k^{(n)}$ ?

We can sum up the content of this section as follows: there is a monotone parameterization of all rational numbers in  $[0, 1]$  by the simpler set of all binary fractions in the same interval.

If we remove the restriction  $r \in [0, 1]$ , we get a parameterization of  $\overline{\mathbb{Q}}$  by  $\overline{\mathbb{Z}[\frac{1}{2}]}$  that preserves the cyclic order on the circle introduced above.

*Remark 6.4.* There is an interesting geometric interpretation of the Farey series and of the Minkowski question mark function. It was discovered by George de Rham [de Rha59].

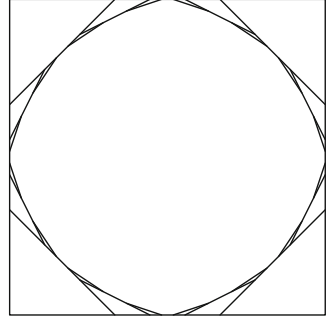
Consider the square  $[-1, 1] \times [-1, 1] \subset \mathbb{R}^2$ . Let us divide every side into three equal parts and join neighboring division points. We get an octagon with equal angles but unequal sides. Repeat this procedure: divide every side of the octagon into three equal parts and join the neighboring division points. The result will be a convex polygon with 16 sides that is contained in the octagon. Proceeding in this way, we get a nested series of convex polygons  $\Pi_n, n \geq 1$ , with  $2^{n+1}$  sides. The intersection of all these polygons is a convex domain  $D$  bounded by a  $C^2$ -smooth curve  $C$  (see Fig. 6.2). Note the following facts:

- (a) The center of each side of every  $\Pi_n$  belongs to  $C$ . Let us enumerate those that belong to the upper half of  $C$  by the numbers  $r_k = \frac{k}{2^n}, -2^n \leq k \leq 2^n$ .
- (b) Let the upper half of  $C$  be given by the equation  $y = f(x), |x| \leq 1$ . Let  $x_k$  be the  $x$ -coordinate of  $r_k$ . Then  $f'(x_k) = f'_{r_k}$ , the member of the  $n$ th Farey series.




---

<sup>7</sup>See **Info G**.

**Fig. 6.2** The de Rham curve

### 6.3 Nice Parameterizations of Disks Tangent to a Given Disk

Let  $A$  be an Apollonian gasket. Choose a disk  $D \in A$  corresponding to a Hermitian matrix  $M$  and consider those disks in  $A$  that are tangent to  $D$ .

The tangent points form a countable subset  $T \subset \partial D$ . We shall show later that one can parameterize points of  $T$  by rational numbers (including  $\infty$ ) in such a way that the natural cyclic order on  $T$ , as a part of  $\partial D$ , corresponds to the cyclic order on  $\overline{\mathbb{Q}}$ , as a part of  $\overline{\mathbb{R}}$ .

Let  $D_r$  be the disk tangent to  $D$  at the point  $t_r \in T$  and let  $M_r$  be the corresponding Hermitian matrix.

We say that a parameterization  $r \rightarrow t_r$  of  $T$  by  $\overline{\mathbb{Q}}$  is *nice* if it has the following properties:

1. If  $r = \frac{p}{q}$  in lowest terms, then

$$M_r = Ap^2 + 2Bpq + Cq^2 - M, \quad \text{where } A, B, C \text{ are fixed Hermitian matrices.}$$

2. The disk  $D_r$  is tangent to  $D_{r'}$  iff  $r = \frac{p}{q}$  and  $r' = \frac{p'}{q'}$  are friendly, i.e., iff  $|pq' - p'q| = 1$ .

Of course, conditions 1 and 2 are very strong and contain all the information about tangent disks. Therefore, the next result is rather important.

**Theorem 6.3.** *Nice parameterizations exist and have the following additional property: Let  $v_0, v_1, v_2, v_3$  be vectors in  $\mathbb{R}^{1,3}$  corresponding to matrices  $A + C, B, A - C, M$ . Then the Gram matrix of their scalar products has the form*

$$G = \|(v_i, v_j)\| = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \quad (6.3.1)$$

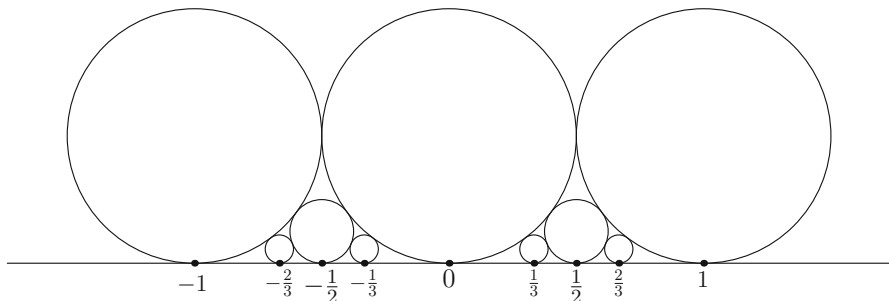


Fig. 6.3 Nice parameterization of a line in the band gasket

**First main example: band gasket.** Let  $D = \{w \in \mathbb{C} \mid \text{Im } w \leq 0\}$ ,  $D_\infty = \{w \in \mathbb{C} \mid \text{Im } w \geq 1\}$ . Let  $D_0, D_1$  be the disks of unit diameter, tangent to  $D$  at points  $0, 1$  and to  $D_\infty$  at points  $i, i + 1$  (Fig. 6.3).

Then  $\partial D = \overline{\mathbb{R}}$ ,  $T = \overline{\mathbb{Q}}$ . The tautological parameterization of  $T$  is nice, with

$$M = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad M_{\frac{p}{q}} = \begin{pmatrix} 2p^2 & -2pq - i \\ -2pq + i & 2q^2 \end{pmatrix}, \quad D_{\frac{p}{q}} : \left| w - \frac{2pq + i}{2q^2} \right| \leq \frac{1}{2q^2}.$$

**Second main example: rectangular gasket.** Let  $D = \{w \in \mathbb{C} \mid |w| \geq 1\}$  be the complement to the open unit disk, and let  $D_0$  be given by the condition  $|w - \frac{1}{2}| \leq \frac{1}{2}$ ,  $D_\infty$  by the condition  $|w + \frac{1}{2}| \leq \frac{1}{2}$ , and  $D_1$  by the condition  $|w - \frac{2i}{3}| \leq \frac{1}{3}$ .

Here  $\partial D$  is the unit circle, and a nice parameterization is  $t_r = \frac{p+iq}{p-iq}$ , so that

$$M = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad M_r = \begin{pmatrix} p^2 + q^2 - 1 & -(p + iq)^2 \\ -(p - iq)^2 & p^2 + q^2 + 1 \end{pmatrix},$$

$$D_{\frac{p}{q}} : \left| w - \frac{(p + iq)^2}{p^2 + q^2 + 1} \right| \leq \frac{1}{p^2 + q^2 + 1}.$$

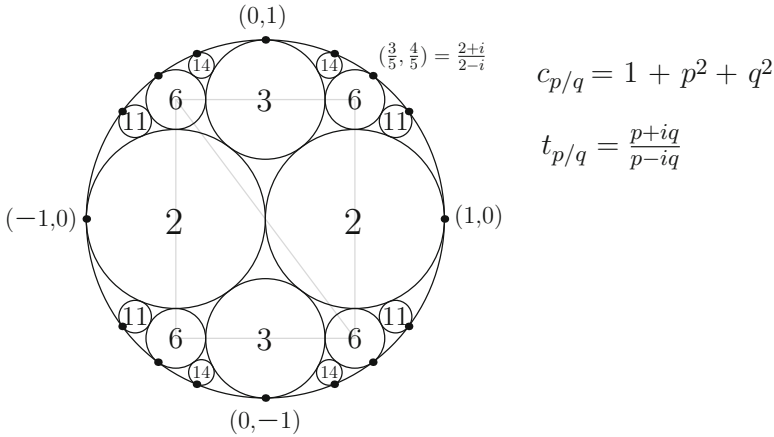
*Proof of Theorem 6.3.* Let  $D_0, D_1, D_\infty$  be any three disks from  $A$  that are tangent to  $D$  and to each other. We associate the labels  $0, 1$  and  $\infty$  to the corresponding tangent points in  $\partial D$  (Fig. 6.4).

Then, assuming that the theorem is true and the parameterization is nice, we can compute  $A, B, C$  from the equations

$$M_\infty = A - M, \quad M_0 = C - M, \quad M_1 = A + 2B + C - M.$$

We get

$$A = M + M_\infty, \quad C = M + M_0, \quad B = \frac{1}{2}(M_1 - M - M_0 - M_\infty).$$



**Fig. 6.4** Nice parameterization of the outer circle in the rectangular gasket

Then, using the property of matrices  $M_0, M_1, M_\infty,$  and  $M$ , we can check relation (6.3.1). From there, statement 2 of the theorem follows easily if we define  $M_r$  using statement 1.  $\square$

Practically, a nice parameterization can be defined step by step. Assume that the disks  $D_{r_1}$  and  $D_{r_2}$  corresponding to friendly rational numbers  $r_1$  and  $r_2$  are already defined and are tangent to  $D$  and to each other. Then we associate to  $r = r_1 \downarrow r_2$ , the disk tangent to  $D_{r_1}, D_{r_2},$  and  $D$ .

In fact, there are two such disks and two possible values of  $r = r_1 \downarrow r_2$ ; the right choice is uniquely determined by the cyclic order.

**Corollary.** *The boundary curvature of the disk tangent to  $D$  at the point  $r = \frac{p}{q}$  (in lowest terms) is given by a quadratic polynomial in  $p, q$ :*

$$c(p, q) = (c_\infty + c) \cdot p^2 + (c_1 - c_0 - c_\infty - c) \cdot pq + (c_0 + c) \cdot q^2 - c, \quad (6.3.2)$$

where  $c_i$  is the boundary curvature of the disk  $D_i$ .

*In particular, if four mutually tangent disks in an Apollonian gasket  $A$  have integral boundary curvatures, then all disks from  $A$  have this property.*

**Exercise 6.11.** For the triangular Apollonian gasket, find the curvatures of the disks tangent to the outer disk.

**Answer.**  $c(p, q) = \frac{2(p^2 - pq + q^2)}{\sqrt{3}} + 1.$

**Exercise 6.12.** Describe the canonical parameterization for the outer circle of the triangular gasket.

*Hint.* Label by 0, 1,  $\infty$  the tangent points corresponding to the three maximal inner disks.

## 6.4 Integral Apollonian Gaskets

### 6.4.1 Basic Quadruples

There are many models of Apollonian gasket for which the curvatures of all circles are integers. We call them *integral gaskets*. For each such gasket, we can choose the quadruple of disks such that corresponding boundary curvatures form an integral quadruple  $(c_1 \geq c_2 \geq c_3 \geq c_4)$  with minimal  $c_1$ . Call it a *basic quadruple*.

**Lemma 6.3.** *For a basic quadruple, we have*

$$c_4 \leq 0, \quad |c_4| < c_3 < \left(1 + \frac{2}{\sqrt{3}}\right) |c_4| \approx 2.1547 \dots \cdot |c_4|.$$

*Proof.* Let  $D_i, 1 \leq i \leq 4$ , be a quadruple of mutually tangent disks with curvatures  $c_i, 1 \leq i \leq 4$ . The first inequality has already been proved (see Remark 4.2).

Consider now Descartes’s equation (4.1.3) as a quadratic equation in  $c_1$  with given  $c_2, c_3, c_4$ . Then we get

$$c_1 = c_2 + c_3 + c_4 \pm 2\sqrt{c_2c_3 + c_3c_4 + c_4c_2}. \tag{6.4.1}$$

Since the initial quadruple is basic, we have to choose the minus sign in Eq. (6.4.1) (otherwise, we could replace  $c_1$  by a smaller quantity).

The inequality  $c_1 \geq c_2$  together with Eq. (6.4.1) gives  $c_3 + c_4 \geq 2\sqrt{c_2c_3 + c_3c_4 + c_4c_2}$ , or  $(c_3 - c_4)^2 \geq 4c_2(c_3 + c_4) \geq (c_3 + c_4)^2$ . This can be true only when  $c_4 \leq 0$ .

Finally, for nonpositive  $c_4$ , we have  $(c_3 - c_4)^2 \geq 4c_2(c_3 + c_4) \geq 4c_3(c_3 + c_4)$ , or  $3c_3^2 + 6c_3c_4 + c_4^2 \leq 4c_4^2$ . This gives  $\sqrt{3}(c_3 + c_4) \leq -2c_4$ , hence  $c_3 \leq \frac{2+\sqrt{3}}{\sqrt{3}}|c_4|$ .  $\square$

Here is a list of basic quadruples of small sizes generating nonisomorphic gaskets in order of increasing  $|c_4|$ :

- $c_4 = 0$  (1, 1, 0, 0);
- $c_4 = -1$  (3, 2, 2, -1);
- $c_4 = -2$  (7, 6, 3, -2);
- $c_4 = -3$  (13, 12, 4, -3), (8, 8, 5, -3);
- $c_4 = -4$  (21, 20, 5, -4), (9, 9, 8, -4);
- $c_4 = -5$  (31, 30, 6, -5), (18, 18, 7, -5);
- $c_4 = -6$  (43, 42, 7, -6), (15, 14, 11, -6), (19, 15, 10, -6);
- $c_4 = -7$  (57, 56, 8, -7), (20, 17, 12, -7), (32, 32, 9, -7);
- $c_4 = -8$  (73, 72, 9, -8), (24, 21, 13, -8), (25, 25, 12, -8);
- $c_4 = -9$  (91, 90, 10, -9), (50, 50, 11, -9), (22, 19, 18, -9), (27, 26, 14, -9);
- $c_4 = -10$  (111, 110, 11, -10), (39, 35, 14, -10), (27, 23, 18, -10);

$$c_4 = -11$$

(133, 132, 12, -11), (72, 72, 13, -11), (37, 36, 16, -11), (28, 24, 21, -11).

Three general formulas:

$$c_4 = -km \quad (k^2 + km + m^2, k(k + m), m(k + m), -km)$$

$$c_4 = 1 - 2k \quad (2k^2, 2k^2, 2k + 1, 1 - 2k)$$

$$c_4 = -4k \quad ((2k + 1)^2, (2k + 1)^2, 4(k + 1), -4k)$$

The reader can find many other interesting facts about integral gaskets in [G03]. However, a description of all basic quadruples is still unknown.

## Info I. The Möbius Inversion Formula

In number-theoretic computations, the *Möbius inversion formula* is frequently used. We explain here how it works.

Suppose we have a partially ordered set  $X$  with the property that for every element  $x \in X$ , there are only finitely many elements that are less than  $x$ . Let now  $f$  be any real- or complex-valued function on  $X$ . Define a new function  $F$  by the formula

$$F(x) = \sum_{y \leq x} f(y). \quad (\text{I.1})$$

**Proposition I.1.** *There exists a unique function  $\tilde{\mu}$  on  $X \times X$  with the following properties:*

1.  $\tilde{\mu}(x, y) = 0$  unless  $y \leq x$ .
2.  $\tilde{\mu}(x, x) = 1$ .
3. If the functions  $f$  and  $F$  are related by Eq. (I.1), then

$$f(x) = \sum_{y \leq x} \tilde{\mu}(x, y)F(y). \quad (\text{I.2})$$

In many applications, the set  $X$  is a semigroup of nonnegative elements in some partially ordered abelian group  $G$ , and the order relation is translation-invariant:  $y < x$  is equivalent to  $a + y < a + x$  for every  $a \in G$ . In this case,  $\mu$  is also translation-invariant,  $\tilde{\mu}(a + x, a + y) = \tilde{\mu}(x, y)$ , and hence it can be written in the form  $\mu(x - y)$ , where  $\mu$  is a function on  $G$  that is zero outside  $X$ . The inversion formula takes the form

$$f(x) = \sum_{y \leq x} \mu(x - y)F(y) \quad (\text{Möbius inversion formula}). \quad (\text{I.3})$$

We leave the proofs for the interested reader and consider only some examples that we need in our book.



*Example 1.* Let  $G = \mathbb{Z}$  with the standard order. Then the formula (I.1) takes the form  $F(n) = \sum_{m \leq n} f(m)$ , and the inversion formula is  $f(n) = F(n) - F(n - 1)$ . We see that in this case, Proposition I.1 is true and the function  $\mu$  is given by

$$\mu(n) = \begin{cases} 1 & \text{if } n = 0, \\ -1 & \text{if } n = 1, \\ 0 & \text{otherwise.} \end{cases}$$

*Example 2.*  $G = G_1 \times G_2$ , and the order on  $G$  is the product of orders on  $G_1$  and on  $G_2$ , i.e.,

$$(g_1, g_2) > (0, 0) \iff g_1 > 0 \text{ and } g_2 > 0.$$

Here the  $\mu$ -function for  $G$  is simply the product of the  $\mu$ -functions for  $G_1$  and  $G_2$ .

Note that if  $G_1$  and  $G_2$  are ordered groups, the  $G = G_1 \times G_2$  is only partially ordered.

*Example 3.*  $G = \mathbb{Q}^\times$  is the multiplicative group of nonzero rational numbers. The partial order is defined as follows:  $r_1 \leq r_2$  if the number  $\frac{r_2}{r_1}$  is an integer. So in this case,  $X = \mathbb{Z}_+$  with the order relation  $m < n$  if  $m \mid n$  ( $m$  is a divisor of  $n$ ).

It is easy to see that this partially ordered group is the direct sum of a countable number of copies of  $\mathbb{Z}$  with the usual order. Indeed, every element of  $G$  can be uniquely written in the form  $r = \prod_{k \geq 1} p_k^{n_k}$ , where  $p_k$  is the  $k$ th prime number,  $n_k \in \mathbb{Z}$ , and only finitely many of  $n_k$  are nonzero. The number  $r$  is an integer iff all  $n_k$  are nonnegative.

Therefore, the function  $\mu$  is the product of infinitely many functions from Example 1. The exact definition is as follows.

**Definition I.1.**

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1, \\ (-1)^k & \text{if } n \text{ is a product of } k \text{ distinct primes,} \\ 0 & \text{otherwise.} \end{cases}$$

Equation (I.3) in this case is the classical Möbius inversion formula

$$f(n) = \sum_{d \mid n} \mu(d) F\left(\frac{n}{d}\right). \tag{I.4}$$

As an application, we derive here the formula for the Euler  $\varphi$ -function.

Let us classify the numbers  $k \leq n$  according to the value of  $d = \gcd(k, n)$ . It is clear that  $\gcd\left(\frac{k}{d}, \frac{n}{d}\right) = 1$ . It follows that the number of those  $k$  for which  $\gcd(k, n) = d$  is equal to  $\varphi\left(\frac{n}{d}\right)$ . We have obtained the identity

$$n = \sum_{d \mid n} \varphi\left(\frac{n}{d}\right).$$

Applying the Möbius inversion formula, we get

$$\varphi(n) = \sum_{d|n} \mu(d) \cdot \frac{n}{d}, \quad \text{or} \quad \frac{\varphi(n)}{n} = \sum_{d|n} \frac{\mu(d)}{d}. \quad (1.5)$$

◇

## 6.4.2 Some Computations

A well-known unsolved problem is to compute the Hausdorff dimension of the Apollonian gasket and the Hausdorff measure of its different modifications (e.g., spherical or triangular gaskets). We know the answer to the first question to a high degree of accuracy: in [MC03], it is shown that the Hausdorff dimension of the Apollonian gasket is  $d = 1.308535???$ ... However, we have no idea about the nature of this number. For example, is it irrational? Can it be expressed in terms of some logarithms as for the Cantor set or Sierpiński gasket? Has it any interesting arithmetic properties?

Another interesting problem is to compute the total area of the disks in some Apollonian gasket that are tangent to a given disk  $D$ , e.g., to the outer disk in the rectangular or triangular gasket.

We start, however, with a slightly easier problem. Consider the first main example of the band gasket above. We want to compute the total area of the disks in the band gasket that are tangent to the real axis at the rational points of the segment  $[0, 1]$ . A more natural question, one with a simpler answer, is to compute the area of the part of the unit square with vertices  $0, 1, 1 + i, i$  covered by the disks tangent to the lower side of the square.

We know that the diameter of the disk with tangent point  $\frac{m}{n} \in [0, 1]$  is  $\frac{1}{n^2}$ . Hence its area is  $\frac{\pi}{4n^4}$ . There are  $\varphi(n)$  disks of this size. So for the area in question, we have the expression

$$A = \frac{\pi}{4} \cdot \sum_{n \geq 1} \frac{\varphi(n)}{n^4}. \quad (6.4.2)$$

This number can be expressed through the values of the Riemann  $\zeta$ -function at the points 3 and 4.

Let us use the formula for  $\varphi(n)$  obtained in Info I. The formula (6.4.2) takes the form

$$A = \frac{\pi}{4} \cdot \sum_{n \geq 1} \sum_{d|n} \frac{\mu(d)}{dn^3}.$$

We denote  $\frac{n}{d}$  by  $m$  and sum over  $d$  and  $m$ . We get

$$A = \frac{\pi}{4} \cdot \sum_{d \geq 1} \sum_{m \geq 1} \frac{\mu(d)}{m^3 d^4} = \frac{\pi}{4} \cdot \sum_{m \geq 1} \frac{1}{m^3} \cdot \sum_{d \geq 1} \frac{\mu(d)}{d^4}.$$

The sum  $\sum_{m \geq 1} \frac{1}{m^3}$  is, by definition, the value  $\zeta(3)$ . On the other hand, the sum  $\sum_{d \geq 1} \frac{\mu(d)}{d^4}$  can be written as

$$\sum_{k \geq 0} (-1)^k \cdot \sum_{1 \leq i_1 < i_2 < \dots < i_k} (p_{i_1} p_{i_2} \dots p_{i_k})^{-4} = \prod_{i \geq 1} \left(1 - \frac{1}{p_i^4}\right) = \frac{1}{\sum_{n \geq 1} \frac{1}{n^4}} = \frac{1}{\zeta(4)}.$$

Finally, we get

$$A = \frac{\pi}{4} \cdot \frac{\zeta(3)}{\zeta(4)} = \frac{45\zeta(3)}{2\pi^3} \approx 0.872284.$$

The total area of the disks tangent to the outer disk of the rectangular gasket is equal to

$$\frac{\pi}{2} \cdot \sum_{\gcd(p,q)=1} \frac{1}{(p^2 + q^2 + 1)^2}.$$

It can be expressed in terms of the  $\zeta$ -function related to the Gaussian field  $\mathbb{Q}(i)$ .

**Exercise 6.13.** Let  $\Sigma_m$  denote the sum  $\sum_{\mathbb{Z}^2 \setminus \{(0,0)\}} \frac{1}{(k^2 + l^2)^m}$ . Show that

$$\sum_{\gcd(p,q)=1} \frac{1}{(p^2 + q^2)^m} = \frac{\Sigma_m}{\zeta(2m)} \tag{6.4.3}$$

and

$$\sum_{\gcd(p,q)=1} \frac{1}{(p^2 + q^2 + 1)^2} = \sum_{m=1}^{\infty} (-1)^{m-1} \frac{m \cdot \Sigma_{m+1}}{\zeta(2m+2)}. \tag{6.4.4}$$