# Math 215 Complex Analysis

Lenya Ryzhik copy pasting from others

November 25, 2013

## 1 The Holomorphic Functions

We begin with the description of complex numbers and their basic algebraic properties. We will assume that the reader had some previous encounters with the complex numbers and will be fairly brief, with the emphasis on some specifics that we will need later.

### 1.1 The Complex Plane

#### 1.1.1 The complex numbers

We consider the set $\mathbb{C}$ of pairs of real numbers $(x, y)$, or equivalently of points on the plane $\mathbb{R}^2$. Two vectors $z_1 = (x_1, x_2)$ and $z_2 = (x_2, y_2)$ are *equal* if and only if $x_1 = x_2$ and $y_1 = y_2$. Two vectors $z = (x, y)$ and $\bar{z} = (x, -y)$ that are symmetric to each other with respect to the $x$-axis are said to be *complex conjugate* to each other. We identify the vector $(x, 0)$ with a real number $x$. We denote by $\mathbb{R}$ the set of all real numbers (the $x$-axis).

We introduce now the operations of addition and multiplication on $\mathbb{C}$ that turn it into a field. The sum of two complex numbers and multiplication by a real number $\lambda \in \mathbb{R}$ are defined in the same way as in $\mathbb{R}^2$:

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2), \quad \lambda(x, y) = (\lambda x, \lambda y).$$

Then we may write each complex number $z = (x, y)$ as

$$z = x \cdot \mathbf{1} + y \cdot i = x + iy, \tag{1.1}$$

where we denoted the two unit vectors in the directions of the $x$ and $y$-axes by $\mathbf{1} = (1, 0)$ and $i = (0, 1)$.

You have previously encountered two ways of defining a product of two vectors: the inner product $(z_1 \cdot z_2) = x_1x_2 + y_1y_2$ and the skew product $[z_1, z_2] = x_1y_2 - x_2y_1$. However, none of them turn $\mathbb{C}$ into a field, and, actually $\mathbb{C}$ is not even closed under these operations: both the inner product and the skew product of two vectors is a number, not a vector. This leads us to introduce yet another product on $\mathbb{C}$. Namely, we postulate that $i \cdot i = i^2 = -1$ and define $z_1 z_2$ as a vector obtained by multiplication of $x_1 + iy_1$ and $x_2 + iy_2$ using the usual rules of algebra with the additional convention that $i^2 = -1$. That is, we define

$$z_1 z_2 = x_1 x_2 - y_1 y_2 + i(x_1 y_2 + x_2 y_1). \tag{1.2}$$

More formally we may write

$$(x_1, y_1)(x_2, y_2) = (x_1 x_2 - y_1 y_2, x_1 y_2 + x_2 y_1)$$

but we will not use this somewhat cumbersome notation.

**Exercise 1.1** *Check that the product (1.2) turns $\mathbb{C}$ into a field, that is, the distributive, commutative and associative laws hold, and for any $z \neq 0$ there exists a number $z^{-1} \in \mathbb{C}$ so that $zz^{-1} = 1$. Hint: $z^{-1} = \dfrac{x}{x^2 + y^2} - \dfrac{iy}{x^2 + y^2}$.*

**Exercise 1.2** *Show that the following operations do not turn $\mathbb{C}$ into a field: (a) $z_1 z_2 = x_1 x_2 + iy_1 y_2$, and (b) $z_1 z_2 = x_1 x_2 + y_1 y_2 + i(x_1 y_2 + x_2 y_1)$.*

The product (1.2) turns $\mathbb{C}$ into a field (see Exercise 1.1) that is called the *field of complex numbers* and its elements, vectors of the form $z = x + iy$ are called *complex numbers.* The real numbers $x$ and $y$ are traditionally called the real and imaginary parts of $z$ and are denoted by

$$x = \mathrm{Re} z, \quad y = \mathrm{Im} z. \tag{1.3}$$

A number $z = (0, y)$ that has the real part equal to zero, is called *purely imaginary.*

The Cartesian way (1.1) of representing a complex number is convenient for performing the operations of addition and subtraction, but one may see from (1.2) that multiplication and division in the Cartesian form are quite tedious. These operations, as well as raising a complex number to a power are much more convenient in the *polar representation* of a complex number:

$$z = r(\cos\phi + i \sin\phi), \tag{1.4}$$

that is obtained from (1.1) passing to the polar coordinates for $(x, y)$. The polar coordinates of a complex number $z$ are the polar radius $r = \sqrt{x^2 + y^2}$ and the polar angle $\phi$, the angle between the vector $z$ and the positive direction of the $x$-axis. They are called the *modulus* and *argument* of $z$ are denoted by

$$r = |z|, \quad \phi = \mathrm{Arg} z. \tag{1.5}$$

The modulus is determined uniquely while the argument is determined up to addition of a multiple of $2\pi$. We will use a shorthand notation

$$\cos\phi + i \sin\phi = e^{i\phi}. \tag{1.6}$$

Note that we have not yet defined the operation of raising a number to a complex power, so the right side of (5.1) should be understood at the moment just as a shorthand for the left side. We will define this operation later and will show that (5.1) indeed holds. With this convention the polar form (1.4) takes a short form

$$z = re^{i\phi}. \tag{1.7}$$

Using the basic trigonometric identities we observe that

$$\begin{aligned} r_1 e^{i\phi_1} r_2 e^{i\phi_2} &= r_1(\cos\phi_1 + i\sin\phi_1) r_2(\cos\phi_2 + i\sin\phi_2) \\ &= r_1 r_2(\cos\phi_1\cos\phi_2 - \sin\phi_1\sin\phi_2 + i(\cos\phi_1\sin\phi_2 + \sin\phi_1\cos\phi_2)) \\ &= r_1 r_2(\cos(\phi_1+\phi_2) + i\sin(\phi_1+\phi_2)) = r_1 r_2 e^{i(\phi_1+\phi_2)}. \end{aligned} \tag{1.8}$$

This explains why notation (5.1) is quite natural. Relation (5.4) says that the modulus of the product is the product of the moduli, while the argument of the product is the sum of the arguments.

Sometimes it is convenient to consider a *compactification* of the set $\mathbb{C}$ of complex numbers. This is done by adding an ideal element that is called the point at infinity $z = \infty$. However, algebraic operations are not defined for $z = \infty$. We will call the compactified complex plane, that is, the plane $\mathbb{C}$ together with the point at infinity, the closed complex plane, denoted by $\overline{\mathbb{C}}$. Sometimes we will call $\mathbb{C}$ the open complex plane in order to stress the difference between $\mathbb{C}$ and $\overline{\mathbb{C}}$.

One can make the compactification more visual if we represent the complex numbers as points not on the plane but on a two-dimensional sphere as follows. Let $\xi$, $\eta$ and $\zeta$ be the Cartesian coordinates in the three-dimensional space so that the $\xi$ and $\eta$-axes coincide with the $x$ and $y$-axes on the complex plane. Consider the unit sphere

$$S: \quad \xi^2 + \eta^2 + \zeta^2 = 1 \tag{1.9}$$

in this space. Then for each point $z = (x, y) \in \mathbb{C}$ we may find a corresponding point $Z = (\xi, \eta, \zeta)$ on the sphere that is the intersection of $S$ and the segment that connects the "North pole" $N = (0, 0, 1)$ and the point $z = (x, y, 0)$ on the complex plane.

The mapping $z \to Z$ is called *the stereographic projection*. The segment $Nz$ may be parameterized as $\xi = tx$, $\eta = ty$, $\zeta = 1 - t$, $t \in [0, 1]$. Then the intersection point $Z = (t_0 x, t_0 y, 1 - t_0)$ with $t_0$ being the solution of

$$t_0^2 x^2 + t_0^2 y^2 + (1 - t_0)^2 = 1$$

so that $(1 + |z|^2)t_0 = 2$. Therefore the point $Z$ has the coordinates

$$\xi = \frac{2x}{1+|z|^2}, \quad \eta = \frac{2y}{1+|z|^2}, \quad \zeta = \frac{|z|^2 - 1}{1+|z|^2}. \tag{1.10}$$

The last equation above implies that $\dfrac{2}{1+|z|^2} = 1 - \zeta$. We find from the first two equations the explicit formulae for the inverse map $Z \to z$:

$$x = \frac{\xi}{1-\zeta}, \quad y = \frac{\eta}{1-\zeta}. \tag{1.11}$$

Expressions (5.9) and(5.10) show that the stereographic projection is a one-to-one map from $\mathbb{C}$ to $S\backslash N$ (clearly $N$ does not correspond to any point $z$). We postulate that $N$ corresponds to the point at infinity $z=\infty$. This makes the stereographic projection be a one-to-one map from $\bar{\mathbb{C}}$ to $S$. We will usually identify $\bar{\mathbb{C}}$ and the sphere $S$. The latter is called *the sphere of complex numbers* or *the Riemann sphere*. The open plane $\mathbb{C}$ may be identified with $S\backslash N$, the sphere with the North pole deleted.

**Exercise 1.3** *Let $t$ and $u$ be the longitude and the latitude of a point $Z$. Show that the corresponding point $z=se^{it}$, where $s=\tan(\pi/4+u/2)$.*

We may introduce two metrics (distances) on $\mathbb{C}$ according to the two geometric descriptions presented above. The first is the usual Euclidean metric with the distance between the points $z_1=x_1+iy_1$ and $z_2=x_2+iy_2$ in $\mathbb{C}$ given by

$$|z_2-z_1|=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}. \tag{1.12}$$

The second is *the spherical metric* with the distance between $z_1$ and $z_2$ defined as the Euclidean distance in the three-dimensional space between the corresponding points $Z_1$ and $Z_2$ on the sphere. A straightforward calculation shows that

$$\rho(z_1,z_2)=\frac{2|z_2-z_1|}{\sqrt{1+|z_1|^2}\sqrt{1+|z_2|^2}}. \tag{1.13}$$

This formula may be extended to $\overline{\mathbb{C}}$ by setting

$$\rho(z,\infty)=\frac{2}{\sqrt{1+|z|^2}}. \tag{1.14}$$

Note that (1.14) may be obtained from (1.13) if we let $z_1=z$, divide the numerator and denominator by $|z_2|$ and let $|z_2|\to+\infty$.

**Exercise 1.4** *Use the formula (5.9) for the stereographic projection to verify (1.13).*

Clearly we have $\rho(z_1,z_2)\leq 2$ for all $z_1,z_2\in\overline{\mathbb{C}}$. It is straightforward to verify that both of the metrics introduced above turn $\mathbb{C}$ into a metric space, that is, all the usual axioms of a metric space are satisfied. In particular, the triangle inequality for the Euclidean metric (1.12) is equivalent to the usual triangle inequality for two-dimensional plane: $|z_1+z_2|\leq|z_1|+|z_2|$.

**Exercise 1.5** *Verify the triangle inequality for the metric $\rho(z_1,z_2)$ on $\overline{\mathbb{C}}$ defined by (1.13) and (1.14)*

We note that the Euclidean and spherical metrics are equivalent on bounded sets $M\subset\mathbb{C}$ that lie inside a fixed disk $\{|z|\leq R\}$, $R<\infty$. Indeed, if $M\subset\{|z|\leq R\}$ then (1.13) implies that for all $z_1,z_2\in M$ we have

$$\frac{2}{1+R^2}|z_2-z_1|\leq\rho(z_1,z_2)\leq 2|z_2-z_1| \tag{1.15}$$

(this will be elaborated in the next section). Because of that the spherical metric is usually used only for unbounded sets. Typically, we will use the Euclidean metric for $\mathbb{C}$ and the spherical metric for $\overline{\mathbb{C}}$.

Now is the time for a little history. We find the first mention of the complex numbers as square rots of negative numbers in the book "Ars Magna" by Girolamo Cardano published in 1545. He thought that such numbers could be introduced in mathematics but opined that this would be useless: "Dismissing mental tortures, and multiplying $5+\sqrt{-15}$ by $5-\sqrt{-15}$, we obtain $25-(-15)$. Therefore the product is 40. .... and thus far does arithmetical subtlety go, of which this, the extreme, is, as I have said, so subtle that it is useless." The baselessness of his verdict was realized fairly soon: Raphael Bombelli published his "Algebra" in 1572 where he introduced the algebraic operations over the complex numbers and explained how they may be used for solving the cubic equations. One may find in Bombelli's book the relation $(2+\sqrt{-121})^{1/3}+(2-\sqrt{-121})^{1/3}=4$. Still, the complex numbers remained somewhat of a mystery for a long time. Leibnitz considered them to be "a beautiful and majestic refuge of the human spirit", but he also thought that it was impossible to factor $x^4+1$ into a product of two quadratic polynomials (though this is done in an elementary way with the help of complex numbers).

The active use of complex numbers in mathematics began with the works of Leonard Euler. He has also discovered the relation $e^{i\phi}=\cos\phi+i\sin\phi$. The geometric interpretation of complex numbers as planar vectors appeared first in the work of the Danish geographical surveyor Caspar Wessel in 1799 and later in the work of Jean Robert Argand in 1806. These papers were not widely known - even Cauchy who has obtained numerous fundamental results in complex analysis considered early in his career the complex numbers simply as symbols that were convenient for calculations, and equality of two complex numbers as a shorthand notation for equality of two real-valued variables.

The first systematic description of complex numbers, operations over them, and their geometric interpretation was given by Carl Friedreich Gauss in 1831 in his memoir "Theoria residuorum biquadraticorum". He has also introduced the name "complex numbers".

## 1.2 The topology of the complex plane

We have introduced distances on $\mathbb{C}$ and $\overline{\mathbb{C}}$ that turned them into metric spaces. We will now introduce the two topologies that correspond to these metrics.

Let $\varepsilon>0$ then an $\varepsilon$-neighborhood $U(z_0,\varepsilon)$ of $z_0\in\mathbb{C}$ in the Euclidean metric is the disk of radius $\varepsilon$ centered at $z_0$, that is, the set of points $z\in\mathbb{C}$ that satisfy the inequality

$$|z-z_0|<\varepsilon. \tag{1.16}$$

An $\varepsilon$-neighborhood of a point $z_0\in\overline{\mathbb{C}}$ is the set of all points $z\in\overline{\mathbb{C}}$ such that

$$\rho(z,z_0)<\varepsilon. \tag{1.17}$$

Expression (1.14) shows that the inequality $\rho(z,\infty)<\varepsilon$ is equivalent to $|z|>\sqrt{\dfrac{4}{\varepsilon^2}-1}$. Therefore an $\varepsilon$-neighborhood of the point at infinity is the outside of a disk centered at the origin complemented by $z=\infty$.

We say that a set $\Omega$ in $\mathbb{C}$ (or $\overline{\mathbb{C}}$) is *open* if for any point $z_0 \in \Omega$ there exists a neighborhood of $z_0$ that is contained in $\Omega$. It is straightforward to verify that this notion of an open set turns $\mathbb{C}$ and $\overline{\mathbb{C}}$ into *topological spaces*, that is, the usual axioms of a topological space are satisfied.

Sometimes it will be convenient to make use of the so called *punctured neighborhoods*, that is, the sets of the points $z \in \mathbb{C}$ (or $z \in \overline{\mathbb{C}}$) that satisfy

$$0 < |z - z_0| < \varepsilon, \quad 0 < \rho(z, z_0) < \varepsilon. \tag{1.18}$$

## 1.3 Paths and curves

**Definition 1.6** *A path $\gamma$ is a continuous map of an interval $[\alpha, \beta]$ of the real axis into the complex plane $\mathbb{C}$ (or $\overline{\mathbb{C}}$). In other words, a path is a complex valued function $z = \gamma(t)$ of a real argument $t$, that is continuous at every point $t_0 \in [\alpha, \beta]$ in the following sense: for any $\varepsilon > 0$ there exists $\delta > 0$ so that $|\gamma(t) - \gamma(t_0)| < \varepsilon$ (or $\rho(\gamma(t), \gamma(t_0)) < \varepsilon$ if $\gamma(t_0) = \infty$) provided that $|t - t_0| < \delta$. The points $a = \gamma(\alpha)$ and $b = \gamma(\beta)$ are called the endpoints of the path $\gamma$. The path is closed if $\gamma(\alpha) = \gamma(\beta)$. We say that a path $\gamma$ lies in a set $M$ if $\gamma(t) \in M$ for all $t \in [\alpha, \beta]$.*

Sometimes it is convenient to distinguish between a path and a curve. In order to introduce the latter we say that two paths

$$\gamma_1 : [\alpha_1, \beta_1] \to \overline{\mathbb{C}} \text{ and } \gamma_2 : [\alpha_2, \beta_2] \to \overline{\mathbb{C}}$$

are *equivalent* ($\gamma_1 \sim \gamma_2$) if there exists an increasing continuous function

$$\tau : [\alpha_1, \beta_1] \to [\alpha_2, \beta_2] \tag{1.19}$$

such that $\tau(\alpha_1) = \alpha_2$, $\tau(\beta_1) = \beta_2$ and so that $\gamma_1(t) = \gamma_2(\tau(t))$ for all $t \in [\alpha_1, \beta_1]$.

**Exercise 1.7** *Verify that relation $\sim$ is reflexive: $\gamma \sim \gamma$, symmetric: if $\gamma_1 \sim \gamma_2$, then $\gamma_2 \sim \gamma_1$ and transitive: if $\gamma_1 \sim \gamma_2$ and $\gamma_2 \sim \gamma_3$ then $\gamma_1 \sim \gamma_3$.*

**Example 1.8** Let us consider the paths $\gamma_1(t) = t$, $t \in [0, 1]$; $\gamma_2(t) = \sin t$, $t \in [0, \pi/2]$; $\gamma_3(t) = \cos t$, $t \in [0, \pi/2]$ and $\gamma_4(t) = \sin t$, $t \in [0, \pi]$. The set of values of $\gamma_j(t)$ is always the same: the interval $[0, 1]$. However, we only have $\gamma_1 \sim \gamma_2$. These two paths trace $[0, 1]$ from left to right once. The paths $\gamma_3$ and $\gamma_4$ are neither equivalent to these two, nor to each other: the interval $[0, 1]$ is traced in a different way by those paths: $\gamma_3$ traces it from right to left, while $\gamma_4$ traces $[0, 1]$ twice.

**Exercise 1.9** Which of the following paths: a) $e^{2\pi it}$, $t \in [0, 1]$; b) $e^{4\pi it}$, $t \in [0, 1]$; c) $e^{-2\pi it}$, $t \in [0, 1]$; d) $e^{4\pi i \sin t}$, $t \in [0, \pi/6]$ are equivalent to each other?

**Definition 1.10** *A curve is an equivalence class of paths. Sometimes, when this will cause no confusion, we will use the word 'curve' to describe a set $\gamma \in \overline{\mathbb{C}}$ that may be represented as an image of an interval $[\alpha, \beta]$ under a continuous map $z = \gamma(t)$.*

Below we will introduce some restrictions on the curves and paths that we will consider. We say that $\gamma : [\alpha, \beta] \to \overline{\mathbb{C}}$ is a *Jordan* path if the map $\gamma$ is continuous and *one-to-one*. The definition of a closed Jordan path is left to the reader as an exercise.

A path $\gamma : [\alpha, \beta] \to \mathbb{C}$ ($\gamma(t) = x(t) + iy(t)$) is *continuously differentiable* if derivative $\gamma'(t) := x'(t) + iy'(t)$ exists for all $t \in [\alpha, \beta]$. A continuously differentiable path is said to be *smooth* if $\gamma'(t) \neq 0$ for all $t \in [\alpha, \beta]$. This condition is introduced in order to avoid singularities. A path is called *piecewise smooth* if $\gamma(t)$ is continuous on $[\alpha, \beta]$, and $[\alpha, \beta]$ may be divided into a finite number of closed sub-intervals so that the restriction of $\gamma(t)$ on each of them is a smooth path.

We will also use the standard notation to describe smoothness of functions and paths: the class of continuous functions is denoted $C$, or $C^0$, the class of continuously differentiable functions is denoted $C^1$, etc. A function that has $n$ continuous derivatives is said to be a $C^n$-function.

**Example 1.11** The paths $\gamma_1$, $\gamma_2$ and $\gamma_3$ of the previous example are Jordan, while $\gamma_4$ is not Jordan. The circle $z = e^{it}$, $t \in [0, 2\pi]$ is a closed smooth Jordan path; the four-petal rose $z = e^{it} \cos 2t$, $t \in [0, 2\pi]$ is a smooth non-Jordan path; the semi-cubic parabola $z = t^2(t + i)$, $t \in [-1, 1]$ is a Jordan continuously differentiable piecewise smooth path. The path $z = t\left(1 + i \sin\left(\frac{1}{t}\right)\right)$, $t \in [-1/\pi, 1/\pi]$ is a Jordan non-piecewise smooth path.

One may introduce similar notions for curves. A *Jordan curve* is a class of paths that are equivalent to some Jordan path (observe that since the change of variables (1.19) is one-to-one, all paths equivalent to a Jordan path are also Jordan).

The definition of a smooth curve is slightly more delicate: this notion has to be invariant with respect to a replacement of a path that represents a given curve by an equivalent one. However, a continuous monotone change of variables (1.19) may map a smooth path onto a non-smooth one unless we impose some additional conditions on the functions $\tau$ allowed in (1.19).

More precisely, a smooth curve is a class of paths that may be obtained out of a smooth path by all possible re-parameterizations (1.19) with $\tau(s)$ being a continuously differentiable function with a positive derivative. One may define a piecewise smooth curve in a similar fashion: the change of variables has to be continuous everywhere, and in addition have a continuous positive derivative except possibly at a finite set of points.

Sometimes we will use a more geometric interpretation of a curve, and say that a Jordan, or smooth, or piecewise smooth curve is a set of points $\gamma \subset \mathbb{C}$ that may be represented as the image of an interval $[\alpha, \beta]$ under a map $z = \gamma(t)$ that defines a Jordan, smooth or piecewise smooth path.

## 1.4 Functions of a complex variable

### 1.4.1 Differentiability

The notion of differentiability is intricately connected to linear approximations so we start with the discussion of linear functions of complex variables.

**Definition 1.12** *A function $f : \mathbb{C} \to \mathbb{C}$ is $\mathbb{C}$-linear, or $\mathbb{R}$-linear, respectively, if*
*(a) $l(z_1 + z_2) = l(z_1) + l(z_2)$ for all $z_1, z_2 \in \mathbb{C}$,*
*(b) $l(\lambda z) = \lambda l(z)$ for all $\lambda \in \mathbb{C}$, or, respectively, $\lambda \in \mathbb{R}$.*

Thus $\mathbb{R}$-linear functions are linear over the field of real numbers while $\mathbb{C}$-linear are linear over the field of complex numbers. The latter form a subset of the former.

Let us find the general form of an $\mathbb{R}$-linear function. We let $z = x + iy$, and use properties (a) and (b) to write $l(z) = xl(1) + yl(i)$. Let us denote $\alpha = l(1)$ and $\beta = l(i)$, and replace $x = (z + \bar{z})/2$ and $y = (z - \bar{z})/(2i)$. We obtain the following theorem.

**Theorem 1.13** *Any $\mathbb{R}$-linear function has the form*

$$l(z) = az + b\bar{z}, \tag{1.20}$$

*where $a = (\alpha - i\beta)/2$ and $b = (\alpha + i\beta)/2$ are complex valued constants.*

Similarly writing $z = 1 \cdot z$ we obtain

**Theorem 1.14** *Any $\mathbb{C}$-linear function has the form*

$$l(z) = az, \tag{1.21}$$

*where $a = l(1)$ is a complex valued constant.*

**Theorem 1.15** *An $\mathbb{R}$-linear function is $\mathbb{C}$-linear if and only if*

$$l(iz) = il(z). \tag{1.22}$$

**Proof.** The necessity of (1.22) follows immediately from the definition of a $\mathbb{C}$-linear function. Theorem 1.13 implies that $l(z) = az + b\bar{z}$, so $l(iz) = i(az - b\bar{z})$. Therefore, $l(iz) = il(z)$ if and only if

$$iaz - ib\bar{z} = iaz + ib\bar{z}.$$

Therefore if $l(iz) = il(z)$ for all $z \in \mathbb{C}$ then $b = 0$ and hence $l$ is $\mathbb{C}$-linear.

We set $a = a_1 + ia_2$, $b = b_1 + ib_2$, and also $z = x + iy$, $w = u + iv$. We may represent an $\mathbb{R}$-linear function $w = az + b\bar{z}$ as two real equations

$$u = (a_1 + b_1)x - (a_2 - b_2)y, \quad v = (a_2 + b_2)x + (a_1 - b_1)y.$$

Therefore geometrically an $\mathbb{R}$-linear function is an affine transform of a plane $\mathbf{y} = A\mathbf{x}$ with the matrix

$$A = \begin{pmatrix} a_1 + b_1 & -(a_2 - b_2) \\ a_2 + b_2 & a_1 - b_1 \end{pmatrix}. \tag{1.23}$$

Its Jacobian is

$$J = a_1^2 - b_1^2 + a_2^2 - b_2^2 = |a|^2 - |b|^2. \tag{1.24}$$

This transformation is non-singular when $|a| \neq |b|$. It transforms lines into lines, parallel lines into parallel lines and squares into parallelograms. It preserves the orientation when $|a| > |b|$ and changes it if $|a| < |b|$.

However, a $\mathbb{C}$-linear transformation $w = az$ may not change orientation since its jacobian $J = |a|^2 \geq 0$. They are not singular unless $a = 0$. Letting $a = |a|e^{i\alpha}$ and recalling the geometric interpretation of multiplication of complex numbers we find that a non-degenerate $\mathbb{C}$-linear transformation

$$w = |a|e^{i\alpha}z \tag{1.25}$$

is the composition of dilation by $|a|$ and rotation by the angle $\alpha$. Such transformations preserve angles and map squares onto squares.

We note that preservation of angles characterizes $\mathbb{C}$-linear transformations. Moreover, the following theorem holds.

**Theorem 1.16** *If an $\mathbb{R}$-linear transformation $w = az + b\bar{z}$ preserves orientation and angles between three non-parallel vectors $e^{i\alpha_1}$, $e^{i\alpha_2}$, $e^{i\alpha_3}$, $\alpha_j \in \mathbb{R}$, $j = 1, 2, 3$, then $w$ is $\mathbb{C}$-linear.*

**Proof.** Let us assume that $w(e^{i\alpha_1}) = \rho e^{i\beta_1}$ and define $w'(z) = e^{-i\beta_1}w(ze^{i\alpha_1})$. Then $w'(z) = a'z + b'\bar{z}$ with

$$a' = ae^{i(\alpha_1-\beta_1)}a, \quad b' = be^{-i(\alpha_1+\beta_1)},$$

and, moreover $w'(1) = e^{-i\beta_1}\rho e^{i\beta_1} = \rho > 0$. Therefore we have $a' + b' > 0$. Furthermore, $w'$ preserves the orientation and angles between vectors $v_1 = 1$, $v_2 = e^{i(\alpha_2-\alpha_1)}$ and $v_3 = e^{i(\alpha_3-\alpha_1)}$. Since both $v_1$ and its image lie on the positive semi-axis and the angles between $v_1$ and $v_2$ and their images are the same, we have $w'(v_2) = h_2v_2$ with $h_2 > 0$. This means that

$$a'e^{i\beta_2} + b'e^{-i\beta_2} = h_2e^{i\beta_2}, \quad \beta_2 = \alpha_2 - \alpha_1,$$

and similarly

$$a'e^{i\beta_3} + b'e^{-i\beta_3} = h_3e^{i\beta_3}, \quad \beta_3 = \alpha_3 - \alpha_1,$$

with $h_3 > 0$. Hence we have

$$a' + b' > 0, \quad a' + b'e^{-2i\beta_2} > 0, \quad a' + b'e^{-2i\beta_3} > 0.$$

This means that unless $b' = 0$ there exist three different vectors that connect the vector $a'$ to the real axis, all having the same length $|b'|$. This is impossible, and hence $b' = 0$ and $w$ is $\mathbb{C}$-linear.

**Exercise 1.17** (a) Give an example of an $\mathbb{R}$-linear transformation that is not $\mathbb{C}$-linear but preserves angles between two vectors.
(b) Show that if an $\mathbb{R}$-linear transformation preserves orientation and maps some square onto a square it is $\mathbb{C}$-linear.

Now we may turn to the notion of differentiability of complex functions. Intuitively, a function is differentiable if it is well approximated by linear functions. Two different definitions of linear functions that we have introduced lead to different notions of differentiability.

**Definition 1.18** *Let $z \in \mathbb{C}$ and let $U$ be a neighborhood of $z$. A function $f : U \to \mathbb{C}$ is $\mathbb{R}$-differentiable (respectively, $\mathbb{C}$-differentiable) at the point $z$ if we have for sufficiently small $|\Delta z|$:*

$$\Delta f = f(z + \Delta z) - f(z) = l(\Delta z) + o(\Delta z), \tag{1.26}$$

*where $l(\Delta z)$ (with $z$ fixed) is an $\mathbb{R}$-linear (respectively, $\mathbb{C}$-linear) function of $\Delta z$, and $o(\Delta z)$ satisfies $o(\Delta z)/\Delta z \to 0$ as $\Delta z \to 0$. The function $l$ is called the differential of $f$ at $z$ and is denoted $df$.*

The increment of an $\mathbb{R}$-differentiable function has, therefore, the form

$$\Delta f = a\Delta z + b\overline{\Delta z} + o(\Delta z). \tag{1.27}$$

Taking the increment $\Delta z = \Delta x$ along the $x$-axis, so that $\overline{\Delta z} = \Delta x$ and passing to the limit $\Delta x \to 0$ we obtain

$$\lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \frac{\partial f}{\partial x} = a + b.$$

Similarly, taking $\Delta z = i\Delta y$ (the increment is long the $y$-axis) so that $\overline{\Delta z} = -i\Delta y$ we obtain

$$\lim_{\Delta y \to 0} \frac{\Delta f}{i\Delta y} = \frac{1}{i}\frac{\partial f}{\partial y} = a - b.$$

The two relations above imply that

$$a = \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right), \quad b = \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right).$$

These coefficients are denoted as

$$\frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right), \quad \frac{\partial f}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right) \tag{1.28}$$

and are sometimes called the formal derivatives of $f$ at the point $z$. They were first introduced by Riemann in 1851.

**Exercise 1.19** Show that (a) $\dfrac{\partial z}{\partial \bar{z}} = 0$, $\dfrac{\partial \bar{z}}{\partial \bar{z}} = 1$; (b) $\dfrac{\partial}{\partial \bar{z}}(f+g) = \dfrac{\partial f}{\partial \bar{z}} + \dfrac{\partial g}{\partial \bar{z}}$, $\dfrac{\partial}{\partial \bar{z}}(fg) = \dfrac{\partial f}{\partial \bar{z}}g + f\dfrac{\partial g}{\partial \bar{z}}$.

Using the obvious relations $dz = \Delta z$, $d\bar{z} = \Delta\bar{z}$ we arrive at the formula for the differential of $\mathbb{R}$-differentiable functions

$$df = \frac{\partial f}{\partial z}dz + \frac{\partial f}{\partial \bar{z}}d\bar{z}. \tag{1.29}$$

Therefore, all functions $f = u+iv$ such that $u$ and $v$ have usual differentials as functions of two real variables $x$ and $y$ turn out to be $\mathbb{R}$-differentiable. This notion does not bring any essential new ideas to analysis. The complex analysis really starts with the notion of $\mathbb{C}$-differentiability.

The increment of a $\mathbb{C}$-differentiable function has the form

$$\Delta f = a\Delta z + o(\Delta z) \tag{1.30}$$

and its differential is a $\mathbb{C}$-linear function of $\Delta z$ (with $z$ fixed). Expression (1.29) shows that $\mathbb{C}$-differentiable functions are distinguished from $\mathbb{R}$-differentiable ones by an additional condition

$$\frac{\partial f}{\partial \bar{z}} = 0. \tag{1.31}$$

If $f = u + iv$ then (1.28) shows that

$$\frac{\partial f}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) + \frac{i}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)$$

so that the complex equation (1.31) may be written as a pair of real equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \tag{1.32}$$

The notion of complex differentiability is clearly very restrictive: while it is fairly difficult to construct an example of a continuous but nowhere real differentiable function, most trivial functions turn out to be non-differentiable in the complex sense. For example, the function $f(z) = x + 2iy$ is nowhere $\mathbb{C}$-differentiable: $\dfrac{\partial u}{\partial x} = 1$, $\dfrac{\partial v}{\partial y} = 2$ and conditions (1.32) fail everywhere.

**Exercise 1.20** 1. Show that $\mathbb{C}$-differentiable functions of the form $u(x) + iv(y)$ are necessarily $\mathbb{C}$-linear.
2. Let $f = u+iv$ be $\mathbb{C}$-differentiable in the whole plane $\mathbb{C}$ and $u = v^2$ everywhere. Show that $f = \text{const}$.

Let us consider the notion of a derivative starting with that of the directional derivative. We fix a point $z \in \mathbb{C}$, its neighborhood $U$ and a function $f : U \to \mathbb{C}$. Setting $\Delta z = |\Delta z|e^{i\theta}$ we obtain from (1.27) and (1.29):

$$\Delta f = \frac{\partial f}{\partial z}|\Delta z|e^{i\theta} + \frac{\partial f}{\partial \bar{z}}|\Delta z|e^{-i\theta} + o(\Delta z).$$

We divide both sides by $\Delta z$, pass to the limit $|\Delta z| \to 0$ with $\theta$ fixed and obtain the derivative of $f$ at the point $z$ in direction $\theta$:

$$\frac{\partial f}{\partial z_\theta} = \lim_{|\Delta z|\to 0, \arg z = \theta} \frac{\Delta f}{\Delta z} = \frac{\partial f}{\partial z} + \frac{\partial f}{\partial \bar{z}} e^{-2i\theta}. \tag{1.33}$$

This expression shows that when $z$ is fixed and $\theta$ changes between 0 and $2\pi$ the point $\dfrac{\partial f}{\partial z_\theta}$ traverses twice a circle centered at $\dfrac{\partial f}{\partial z}$ with the radius $\left|\dfrac{\partial f}{\partial \bar z}\right|$.

Hence if $\dfrac{\partial f}{\partial \bar z} \neq 0$ then the directional derivative depends on direction $\theta$, and only if $\dfrac{\partial f}{\partial \bar z} = 0$, that is, if $f$ is $\mathbb{C}$-differentiable, all directional derivatives at $z$ are the same.

Clearly, the derivative of $f$ at $z$ exists if and only if the latter condition holds. It is defined by

$$f'(z) = \lim_{\Delta z \to 0} \frac{\Delta f}{\Delta z}. \tag{1.34}$$

The limit is understood in the topology of $\mathbb{C}$. It is also clear that if $f'(z)$ exists then it is equal to $\dfrac{\partial f}{\partial z}$. This proposition is so important despite its simplicity that we formulate it as a separate theorem.

**Theorem 1.21** *Complex differentiability of $f$ at $z$ is equivalent to the existence of the derivative $f'(z)$ at $z$.*

**Proof.** If $f$ is $\mathbb{C}$-differentiable at $z$ then (1.30) with $a = \dfrac{\partial f}{\partial z}$ implies that

$$\Delta f = \frac{\partial f}{\partial z}\Delta z + o(\Delta z).$$

Then, since $\displaystyle\lim_{\Delta z \to 0} \frac{o(\Delta z)}{\Delta z} = 0$, we obtain that the limit $f'(z) = \displaystyle\lim_{\Delta z \to 0} \frac{\Delta f}{\Delta z}$ exists and is equal to $\dfrac{\partial f}{\partial z}$. Conversely, if $f'(z)$ exists then by the definition of the limit we have

$$\frac{\Delta f}{\Delta z} = f'(z) + \alpha(\Delta z),$$

where $\alpha(\Delta z) \to 0$ as $\Delta z \to 0$. Therefore the increment $\Delta f = f'(z)\Delta z + \alpha(\Delta z)\Delta z$ may be split into two parts so that the first is linear in $\Delta z$ and the second is $o(\Delta z)$, which is equivalent to $\mathbb{C}$-differentiability of $f$ at $z$.$\Box$

The definition of the derivative of a function of a complex variable is exactly the same as in the real analysis, and all the arithmetic rules of dealing with derivatives translate into the complex realm without any changes. Thus the elementary theorems regarding derivatives of a sum, product, ratio, composition and inverse function apply verbatim in the complex case. We skip their formulation and proofs.

We should have convinced ourselves that the notion of $\mathbb{C}$-differentiability is very natural. However, as we will see later, $\mathbb{C}$-differentiability at one point is not sufficient to build an interesting theory. Therefore we will require $\mathbb{C}$-differentiability not at one point but in a whole neighborhood.

**Definition 1.22** *A function $f$ is holomorphic (or analytic) at a point $z \in \mathbb{C}$ if it is $\mathbb{C}$-differentiable in a neighborhood of $z$.*

**Example 1.23** The function $f(z) = |z|^2 = z\bar{z}$ is clearly $\mathbb{R}$-differentiable everywhere in $\mathbb{C}$. However, $\dfrac{\partial f}{\partial \bar{z}} = 0$ only at $z = 0$, so $f$ is only $\mathbb{C}$-differentiable at $z = 0$ but is not holomorphic at this point.

The set of functions holomorphic at a point $z$ is denoted by $\mathcal{O}_z$. Sums and products of functions in $\mathcal{O}_z$ also belong to $\mathcal{O}_z$, so this set is a ring. We note that the ratio $f/g$ of two functions in $\mathcal{O}_z$ might not belong to $\mathcal{O}_z$ if $g(z) = 0$.

Functions that are $\mathbb{C}$-differentiable at all points of an open set $D \subset \mathbb{C}$ are clearly also holomorphic at all points $z \in D$. We say that such functions are holomorphic in $D$ and denote their collection by $\mathcal{O}(D)$. The set $\mathcal{O}(D)$ is also a ring. In general a function is holomorphic on a set $M \subset \mathbb{C}$ if it may extended to a function that is holomorphic on an open set $D$ that contains $M$.

Finally we say that $f$ is holomorphic at infinity if the function $g(z) = f(1/z)$ is holomorphic at $z = 0$. This definition allows to consider functions holomorphic in $\overline{\mathbb{C}}$. However, the notion of derivative at $z = \infty$ is not defined.

## 1.5 Geometric and Hydrodynamic Interpretations

The differentials of an $\mathbb{R}$-differentiable and, respectively, a $\mathbb{C}$-differentiable function at a point $z$ have form

$$df = \frac{\partial f}{\partial z}dz + \frac{\partial f}{\partial \bar{z}}d\bar{z}, \quad df = f'(z)dz. \tag{1.35}$$

The Jacobians of such maps are given by (see (1.24))

$$J_f(z) = \left|\frac{\partial f}{\partial z}\right|^2 - \left|\frac{\partial f}{\partial \bar{z}}\right|^2, \quad J_f(z) = |f'(z)|^2. \tag{1.36}$$

Let us assume that $f$ is $\mathbb{R}$-differentiable at $z$ and $z$ is not a critical point of $f$, that is, $J_f(z) \neq 0$. The implicit function theorem implies that locally $f$ is a homeomorphism, that is, there exists a neighborhood $U$ of $z$ so that $f$ maps $U$ continuously and one-to-one onto a neighborhood of $f(z)$. Expressions (1.36) show that in general $J_f$ may have an arbitrary sign if $f$ is just $\mathbb{R}$-differentiable. However, the critical points of a $\mathbb{C}$-differentiable map coincide with the points where derivative vanishes, while such maps preserve orientation at non-critical points: $J_f(z) = |f'(z)'|^2 > 0$.

Furthermore, an $\mathbb{R}$-differentiable map is said to be conformal at $z \in \mathbb{C}$ if its differential $df$ at $z$ is a non-degenerate transformation that is a composition of dilation and rotation. Since the latter property characterizes $\mathbb{C}$-linear maps we obtain the following geometric interpretation of $\mathbb{C}$-differentiability:

*Complex differentiability of $f$ at a point $z$ together with the condition $f'(z) \neq 0$ is equivalent to $f$ being a conformal map at $z$.*

A map $f : D \to \mathbb{C}$ conformal at every point $z \in D$ is said to be conformal in $D$. It is realized by a holomorphic function in $z$ with no critical points ($f'(z) \neq 0$ in $D$). Its differential at every point of the domain is a composition of a dilation and a rotation, in particular it conserves angles. Such mappings were first considered by Euler in 1777 in relation to his participation in the project of producing geographic maps of Russia. The name "conformal mapping" was introduced by F. Schubert in 1789.

So far we have studied differentials of maps. Let us look now at how the properties of the map itself depend on it being conformal. Assume that $f$ is conformal in a neighborhood $U$ of a point $z$ and that $f'$ is continuous in $U$[1]. Consider a smooth path $\gamma : I = [0, 1] \to U$ that starts at $z$, that is, $\gamma'(t) \neq 0$ for all $t \in I$ and $\gamma(0) = z$. Its image $\gamma_* = f \circ \gamma$ is also a smooth path since

$$\gamma_*'(t) = f'[\gamma(t)]\gamma'(t), \quad t \in I, \tag{1.37}$$

and $f'$ is continuous and different from zero everywhere in $U$ by assumption.

Geometrically $\gamma'(t) = \dot{x}(t) + i\dot{y}(t)$ is the vector tangent to $\gamma$ at the point $\gamma(t)$, and $|\gamma'(t)|dt = \sqrt{\dot{x}^2 + \dot{y}^2}dt = ds$ is the differential of the arc length of $\gamma$ at the same point. Similarly, $|\gamma_*(t)|dt = ds_*$ is the differential of the arc length of $\gamma_*$ at the point $\gamma_*(t)$. We conclude from (1.37) at $t = 0$ that

$$|f'(z)| = \frac{|\gamma_*'(0|}{|\gamma'(0)|} = \frac{ds_*}{ds}. \tag{1.38}$$

Thus the modulus of $f'(z)$ is equal to the dilation coefficient at $z$ under the mapping $f$.

The left side does not depend on the curve $\gamma$ as long as $\gamma(0) = z$. Therefore under our assumptions all arcs are dilated by the same factor. Therefore a conformal map $f$ has a circle property: it maps small circles centered at $z$ into curves that differ from circles centered at $f(z)$ only by terms of the higher order.

Going back to (1.37) we see that

$$\arg f'(z) = \arg \gamma_*'(0) - \arg \gamma'(0), \tag{1.39}$$

so that $\arg f'(z)$ is the rotation angle of the tangent lines at $z$ under $f$.

The left side also does not depend on the choice of $\gamma$ as long as $\gamma(0) = z$, so that all such arcs are rotated by the same angle. Thus a conformal map $f$ preserves angles: the angle between any two curves at $z$ is equal to the angle between their images at $f(z)$.

If $f$ is holomorphic at $z$ but $z$ is a critical point then the circle property holds in a degenerate form: the dilation coefficient of all curves at $z$ is equal to 0. Angle preservation does not hold at all, for instance under the mapping $z \to z^2$ the angle between the lines $\arg z = \alpha_1$ and $\arg z = \alpha_2$ doubles! Moreover, smoothness of curves may be violated at a critical point. For instance a smooth curve $\gamma(t) = t + it^2$, $t \in [-1, 1]$ is mapped under the same map $z \to z^2$ into the curve $\gamma_*(t) = t^2(1 - t^2) + 2it^3$ with a cusp at $\gamma_*(0) = 0$.

**Exercise 1.24** Let $u(x, y)$ and $v(x, y)$ be real valued $\mathbb{R}$-differentiable functions and let $\nabla u = \dfrac{\partial u}{\partial x} + i\dfrac{\partial u}{\partial y}$, $\nabla v = \dfrac{\partial v}{\partial x} + i\dfrac{\partial v}{\partial y}$. Find the geometric meaning of the conditions $(\nabla u, \nabla v) = 0$ and $|\nabla u| = |\nabla v|$, and their relation to the $\mathbb{C}$-differentiability of $f = u + iv$ and the conformity of $f$.

[1]We will later see that existence of $f'$ implies its continuity and, moreover, existence of derivatives of all orders.

Let us now find the hydrodynamic meaning of complex differentiability and derivative. We consider a steady two-dimensional flow. That means that the flow vector field $v = (v_1, v_2)$ does not depend on time. The flow is described by

$$v = v_1(x, y) + iv_2(x, y). \tag{1.40}$$

Let us assume that in a neighborhood $U$ of the point $z$ the functions $v_1$ and $v_2$ have continuous partial derivatives. We will also assume that the flow $v$ is irrotational in $U$, that is,

$$\mathrm{curl} v = \frac{\partial v_2}{\partial x} - \frac{\partial v_1}{\partial y} = 0 \tag{1.41}$$

and incompressible:

$$\mathrm{div} v = \frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y} = 0 \tag{1.42}$$

at all $z \in U$.

Condition (1.41) implies the existence of a potential function $\phi$ such that $v = \nabla\phi$, that is,

$$v_1 = \frac{\partial \phi}{\partial x}, \quad v_2 = \frac{\partial \phi}{\partial y}. \tag{1.43}$$

The incompressibility condition (1.42) implies that there exists a stream function $\psi$ so that

$$v_2 = -\frac{\partial \psi}{\partial x}, \quad v_1 = \frac{\partial \psi}{\partial y}. \tag{1.44}$$

We have $d\psi = -v_2 dx + v_1 dy = 0$ along the level set of $\psi$ and thus $\frac{dy}{dx} = \frac{v_2}{v_1}$. This shows that the level set is an integral curve of $v$.

Consider now a complex function

$$f = \phi + i\psi, \tag{1.45}$$

that is called the complex potential of $v$. Relations (1.43) and (1.44) imply that $\phi$ and $\psi$ satisfy

$$\frac{\partial \phi}{\partial x} = \frac{\partial \psi}{\partial y}, \quad \frac{\partial \phi}{\partial y} = -\frac{\partial \psi}{\partial x}. \tag{1.46}$$

The above conditions coincide with (1.32) and show that the complex potential $f$ is holomorphic at $z \in U$.

Conversely let the function $f = \phi + i\psi$ be holomorphic in a neighborhood $U$ of a point $z$, and let the functions $\phi$ and $\psi$ be twice continuously differentiable. Define the vector field $v = \nabla\phi = \frac{\partial \phi}{\partial x} + i\frac{\partial \phi}{\partial y}$. It is irrotational in $U$ since $\mathrm{curl} v = \frac{\partial^2 \phi}{\partial x \partial y} - \frac{\partial^2 \phi}{\partial y \partial x} = 0$. It is also incompressible since $\mathrm{div} v = \frac{\partial^2 \phi}{\partial^2 x} + \frac{\partial^2 \phi}{\partial^2 y} = \frac{\partial^2 \phi}{\partial x \partial y} - \frac{\partial^2 \phi}{\partial y \partial x} = 0$. The complex potential of the vector field $v$ is clearly the function $f$.

*Therefore the function $f$ is holomorphic if and only if it is the complex potential of a steady fluid flow that is both irrotational and incompressible.*

It is easy to establish the hydrodynamic meaning of the derivative:

$$f' = \frac{\partial\phi}{\partial x} + i\frac{\partial\psi}{\partial x} = v_1 - iv_2, \tag{1.47}$$

so that the derivative of the complex potential is the vector that is the complex conjugate of the flow vector. The critical points of $f$ are the points where the flow vanishes.

**Example 1.25** Let us find the complex potential of an infinitely deep flow over a flat bottom with a line obstacle of height $h$ perpendicular to the bottom. This is a flow in the upper half-plane that goes around an interval of length $h$ that we may consider lying on the imaginary axis.

The boundary of the domain consists, therefore, of the real axis and the interval $[0, ih]$ on the imaginary axis. The boundary must be a stream line of the flow. We set it to be the level set $\psi = 0$ and will assume that $\psi > 0$ everywhere in $D$. In order to find the complex potential $f$ it suffices to find a conformal mapping of $D$ onto the upper half-plane $\psi > 0$. One function that provides such a mapping may be obtained as follows. The mapping $z_1 = z^2$ maps $D$ onto the plane without the half-line $\mathrm{Re}z_1 \geq -h^2$, $\mathrm{Im}z_1 = 0$. The map $z_2 = z_1 + h^2$ maps this half-line onto the positive semi-axis $\mathrm{Re}z_2 \geq 0$, $\mathrm{Im}z_2 = 0$. Now the mapping $w_2 = \sqrt{z_2} = \sqrt{|z_2|}e^{i(\arg z_2)/2}$ with $0 < \arg z_2 < 2\pi$ maps the complex plane without the positive semi-axis onto the upper half-plane. It remains to write explicitly the resulting map

$$w = \sqrt{z_2} = \sqrt{z_1 + h^2} = \sqrt{z^2 + h^2} \tag{1.48}$$

that provides the desired mapping of $D$ onto the upper half-plane. We may obtain the equation for the stream-lines of the flow by writing $(\phi + i\psi)^2 = (x + iy)^2 + h^2$. The streamline $\psi = \psi_0$ is obtained by solving

$$\phi^2 - \psi_0^2 = h^2 + x^2 - y^2, \quad 2\phi\psi_0 = 2xy.$$

This leads to $\phi = xy/\psi_0$ and

$$y = \psi_0\sqrt{1 + \frac{h^2}{x^2 + \psi_0^2}}. \tag{1.49}$$

The magnitude of the flow is $|v| = \left|\dfrac{dw}{dz}\right| = \dfrac{|z|}{\sqrt{|z|^2 + h^2}}$ and is equal to one at infinity. The point $z = 0$ is the critical point of the flow. One may show that the general form of the solution is

$$f(z) = v_\infty\sqrt{z^2 + h^2}, \tag{1.50}$$

where $v_\infty > 0$ is the flow speed at infinity.

## 1.6 Möbius transforms

We will later prove that the only conformal maps $\bar{\mathbb{C}} \to \bar{\mathbb{C}}$ are the ones given by rational functions. It is clear how to identify the conformal automorphisms amongst these maps, at least on the non-rigorous level. Indeed, the fact that there is only one (and simple since the map is one-to-one) solution to

$$\frac{P(z)}{Q(z)} = 0 \tag{1.51}$$

means (via the fundamental theorem of algebra that we will prove soon) that $P(z)$ is linear. Furthermore, if $P(z) \neq \text{const}$ then (1.51) has a solution different from $z = \infty$. Therefore, we can not have $P(\infty)/Q(\infty) = 0$ in that case, which means that $Q(z)$ also has to be linear. Finally, when $P(z) = P_0 = \text{const}$, one sees that $Q(z)$ is linear since the equation $P_0/Q(z) = w$ has exactly one solution for each $w \in \bar{\mathbb{C}}$. Based on this argument (which the reader for now can ignore if desired), the next lemma identifies all automorphisms of $\bar{\mathbb{C}}$.

**Lemma 1.26** *Every matrix $A \in GL(2, \mathbb{C})$ defines a transformation*

$$T_A(z) := \frac{az+b}{cz+d}, \qquad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

*which is holomorphic as a map from $\mathbb{C} \to \mathbb{C}$. It is called a fractional linear or Möbius transformation. The map $A \mapsto T_A$ only depends on the equivalence class of $A$ under the relation $A \sim B$ iff $A = \lambda B$, $\lambda \in \mathbb{C}^*$. In other words, the family of all Möbius transformations is the same as*

$$PSL(2, \mathbb{C}) := SL(2, \mathbb{C})/\{\pm \mathrm{Id}\} \tag{1.52}$$

*We have $T_A \circ T_B = T_{A \circ B}$ and $T_A^{-1} = T_{A^{-1}}$. In particular, every Möbius transformation is an automorphism of $\bar{\mathbb{C}}$.*

*Proof.* It is clear that each $T_A$ is a holomorphic map $\bar{\mathbb{C}} \to \bar{\mathbb{C}}$. The composition law $T_A \circ T_B = T_{A \circ B}$ and $T_A^{-1} = T_{A^{-1}}$ are simple computations that we leave to the reader. In particular, $T_A$ has a conformal inverse and is thus an automorphism of $\bar{\mathbb{C}}$. To prove the last claim, note that if $T_A = T_{\widetilde{A}}$ where $A, \widetilde{A} \in SL(2, \mathbb{C})$, then the derivatives also coincide:

$$T_A'(z) = \frac{ad - bc}{(cz+d)^2} = T_{\widetilde{A}}'(z) = \frac{\widetilde{a}\widetilde{d} - \widetilde{b}\widetilde{c}}{(\widetilde{c}z + \widetilde{d})^2}$$

and thus $cz + d = \pm(\widetilde{c}z + \widetilde{d})$, as $A$ and $\widetilde{A}$ obey the normalization

$$ad - bc = \widetilde{a}\widetilde{d} - \widetilde{b}\widetilde{c} = 1$$

Hence, $A$ and $\widetilde{A}$ are the same matrices in $SL(2, \mathbb{C})$ possibly up to a choice of sign, which establishes (1.52). □

Fractional linear transformations enjoy many important properties which can be checked separately for each of the following four elementary transformations.

**Lemma 1.27** *Every Möbius transformation is the composition of four elementary maps:*

- *translations* $z \mapsto z + z_0$
- *dilations* $z \mapsto \lambda z$, $\lambda > 0$
- *rotations* $z \mapsto e^{i\theta} z$, $\theta \in \mathbb{R}$
- *inversion* $z \mapsto \frac{1}{z}$

*Proof.* If $c = 0$, then $T_A(z) = \frac{a}{d} z + \frac{b}{d}$. If $c \neq 0$, then

$$T_A(z) = \frac{bc - ad}{c^2} \frac{1}{z + \frac{d}{c}} + \frac{a}{c}$$

and we are done. □

The reader will have no difficulty verifying that the transformation $z \mapsto \frac{z-1}{z+1}$ maps the right half-plane $\{\mathrm{Re} z > 0\}$ onto the unit disk $\mathbb{D} := \{|z| < 1\}$. In particular, the imaginary axis $i\mathbb{R}$ is mapped onto the unit circle and $z = 1$ is mapped to zero. Similarly, the transformation $z \mapsto \frac{2z-1}{2-z}$ maps $\mathbb{D}$ onto itself with the boundary going onto the boundary, since

$$\left| \frac{2e^{i\theta} - 1}{2 - e^{i\theta}} \right| = \left| \frac{2 - e^{-i\theta}}{2 - e^{i\theta}} \right| = 1, \quad \text{for any } \theta \in \mathbb{R}.$$

If we include all lines into the family of circles (they may thought of as circles passing through $\infty$, and their images on the unit sphere under the stereographic projection are true circles on the sphere) then these examples motivate the following lemma.

**Lemma 1.28** *Fractional linear transformations map circles onto circles.*

*Proof.* In view of Lemma 1.27, the only case requiring an argument is the inversion. Thus, let $|z - z_0| = r$ be a circle and set $w = \frac{1}{z}$. Then

$$0 = |z|^2 - 2\mathrm{Re}\,(\bar{z} z_0) + |z_0|^2 - r^2 = \frac{1}{|w|^2} - 2\frac{\mathrm{Re}\,(w z_0)}{|w|^2} + |z_0|^2 - r^2$$

If $|z_0| = r$, then one obtains the equation of a line in $w$. Note that this is precisely the case when the circle passes through the origin. Otherwise, we obtain the equation

$$0 = \left| w - \frac{\bar{z}_0}{|z_0|^2 - r^2} \right|^2 - \frac{r^2}{(|z_0|^2 - r^2)^2}$$

which is a circle. Finally, a line is given by an equation

$$2\mathrm{Re}\,(z \bar{z}_0) = a$$

which transforms into $2\mathrm{Re}\,(z_0 w) = a|w|^2$. If $a = 0$, then we simply obtain another line through the origin. Otherwise, we obtain the equation $|w - z_0/a|^2 = |z_0/a|^2$ which is a circle. □

Since
$$Tz = \frac{az+b}{cz+d} = z$$
is a quadratic equation[2] for any Möbius transform $T$, we see that $T$ can have at most two fixed points unless it is the identity.

It is also clear that every Möbius transform has at least one fixed point. The map $Tz = z+1$ has exactly one fixed point, namely $z = \infty$, whereas $Tz = \frac{1}{z}$ has two, $z = \pm 1$.

**Lemma 1.29** *A fractional linear transformation is determined completely by its action on three distinct points. Moreover, given $z_1, z_2, z_3 \in \bar{\mathbb{C}}$ distinct, there exists a unique fractional linear transformation $T$ with $Tz_1 = 0$, $Tz_2 = 1$, $Tz_3 = \infty$.*

*Proof.* For the first statement, suppose that $S, T$ are Möbius transformations that agree at three distinct points. Then $S^{-1} \circ T$ has three fixed points and is thus the identity. For the second statement, let
$$Tz := \frac{z - z_1}{z - z_3} \frac{z_2 - z_3}{z_2 - z_1}$$
in case $z_1, z_2, z_3 \in \mathbb{C}$. If any one of these points is $\infty$, then we obtain the correct formula by passing to the limit here. $\square$

**Definition 1.30** *The cross ratio of four points $z_0, z_1, z_2, z_3 \in \bar{\mathbb{C}}$ is defined as*
$$[z_0 : z_1 : z_2 : z_3] := \frac{z_0 - z_1}{z_0 - z_3} \frac{z_2 - z_3}{z_2 - z_1}$$
This concept is most relevant for its relation to Möbius transformations.

**Lemma 1.31** *The cross ratio of any four distinct points is preserved under Möbius transformations. Moreover, four distinct points lie on a circle iff their cross ratio is real.*

*Proof.* Let $z_1, z_2, z_3$ be distinct, $T$ be a Möbius transformation, and let $Tz_j = w_j$, $j = 1, 2, 3$. Then for all $z \in \mathbb{C}$, we have
$$[w : w_1 : w_2 : w_3] = [z : z_1 : z_2 : z_3] \text{ provided } w = Tz$$
The reason is that the cross ratio on the left side defines a Möbius transformation $S_1 w$ with the property that $S_1 w_1 = 0, S_1 w_2 = 1, S_1 w_3 = \infty$, whereas the right side defines a transformation $S_0$ with $S_0 z_1 = 0, S_0 z_2 = 1, S_0 z_3 = \infty$. Hence $S_1^{-1} \circ S_0 z_j = w_j$, for $j = 1, 2, 3$, which implies that $S_1^{-1} \circ S_0 = T$ as claimed, by virtue of Lemma 1.29. The second statement is an immediate consequence of the first and the fact that for any three distinct points $z_1, z_2, z_3 \in \mathbb{R}$, a fourth point $z_0$ has a real-valued cross ratio with these three iff $z_0 \in \mathbb{R}$. $\square$

We can now define what it means for two points to be symmetric relative to a circle (or line — recall that we consider lines to be circles passing through $z = \infty$).

[2] Strictly speaking, this is a quadratic equation provided $c \neq 0$; if $c = 0$ one obtains a linear equation with a fixed point in $\mathbb{C}$ and another one at $z = \infty$.

**Definition 1.32** *Let $z_1, z_2, z_3 \in \Gamma$ where $\Gamma \subset \mathbb{C}_\infty$ is a circle. We say that $z$ and $z^*$ are symmetric relative to $\Gamma$ if*

$$\overline{[z : z_1 : z_2 : z_3]} = [z^* : z_1 : z_2 : z_3].$$

Obviously, if $\Gamma = \mathbb{R}$, then $z^* = \bar{z}$. In other words, if $\Gamma$ is a line, then $z^*$ is the reflection of $z$ across that line. If $\Gamma$ is a circle of a finite radius, then the symmetric point is given by what is known in the elementary geometry as an inversion.

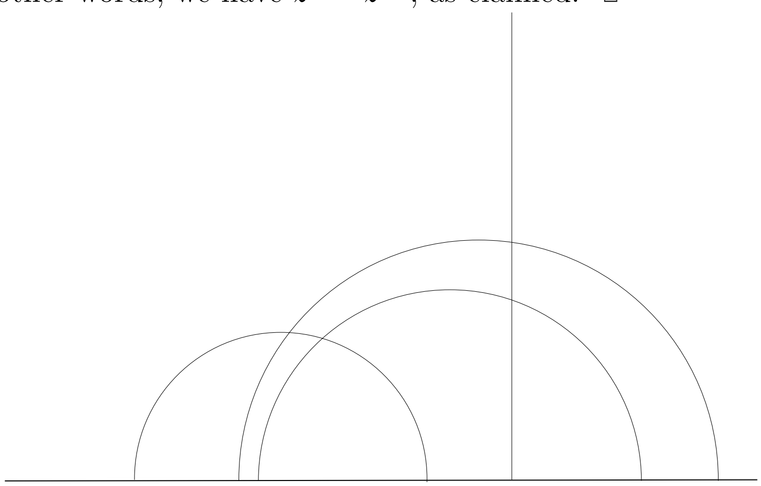
**Lemma 1.33** *Let $\Gamma = \{|z - z_0| = r\}$. Then for any $z \in \mathbb{C}_\infty$,*

$$z^* = \frac{r^2}{\bar{z} - \bar{z}_0}$$

*Proof.* It is sufficient to consider the unit circle – the general case follows by translation and dilation. If $z_{1,2,3}$ lie on the unit circle, then $\bar{z}_j = z_j^{-1}$, hence

$$\overline{[z : z_1 : z_2 : z_3]} = [\bar{z} : z_1^{-1} : z_2^{-1} : z_3^{-1}] = \frac{\bar{z} - z_1^{-1}}{\bar{z} - z_3^{-1}} \frac{z_2^{-1} - z_3^{-1}}{z_2^{-1} - z_1^{-1}} = \frac{z_1\bar{z} - 1}{\bar{z} - z_3^{-1}} \frac{z_2^{-1} - z_3^{-1}}{z_1 z_2^{-1} - 1}$$
$$= \frac{z_1\bar{z} - 1}{\bar{z} - z_3^{-1}} \frac{1 - z_2 z_3^{-1}}{z_1 - z_2} = \frac{z_1\bar{z} - 1}{\bar{z}z_3 - 1} \frac{z_3 - z_2}{z_1 - z_2} = [1/\bar{z} : z_1 : z_2 : z_3]$$

In other words, we have $z^* = \bar{z}^{-1}$, as claimed. □



Möbius transformations are important for several reasons. We already observed that they are precisely the automorphisms of the Riemann sphere (though to see that every automorphism is a Möbius transformation requires additional material). In the 19th century there was much excitement surrounding non-Euclidean geometry and there is an important connection between Möbius transformations and hyperbolic geometry: the isometries of the hyperbolic plane $\mathbb{H}$ are precisely those Möbius transformations which preserve it. Let us be more precise. Consider the upper half-plane model of the hyperbolic plane given by

$$\mathbb{H} = \{z \in \mathbb{C} \,:\, \operatorname{Im} z > 0\}, \quad ds^2 = \frac{dx^2 + dy^2}{y^2} = \frac{dz d\bar{z}}{(\operatorname{Im} z)^2}.$$

It is not hard to see that Möbius transformations that preserve the upper half-plane are given by

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

with $a, b, c, d \in \mathbb{R}$ with $ad - bc = 1$ (up to multiplication of $a, b, c, d$ by a complex number $\lambda \in \mathbb{C}^*$). Indeed, a Möbius transformation preserves the real line if and only if $a, b, c, d \in \lambda\mathbb{R}$ for some $\lambda \in \mathbb{C}^*$. Without loss of generality we may assume that $ad - bc = \pm 1$. If the determinant equals $+1$ (so that the corresponding matrix is in $PSL(2, \mathbb{R})$), then the upper half-plane is preserved, while those with a negative determinant interchange the upper and the lower half-planes. It is easy to check that $PSL(2, \mathbb{R})$ operates transitively on $\mathbb{H}$ and preserves the metric: for the latter, one simply computes that if

$$w = \frac{az+b}{cz+d}, \quad a, b, c, d \in \mathbb{R}, \quad ad - bc = 1,$$

then

$$dw = \left[\frac{a(cz+d) - (az+b)c}{(cz+d)^2}\right] dz = \frac{1}{(cz+d)^2} dz,$$

and

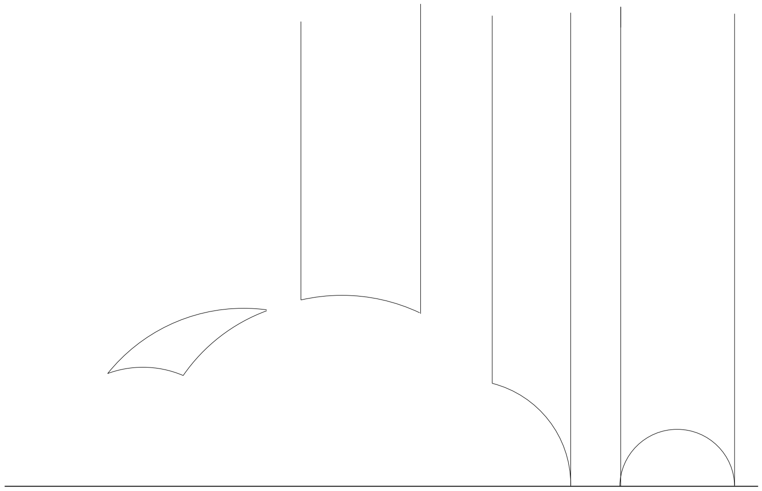
$$\begin{aligned} 2i\mathrm{Im}\, w &= \frac{az+b}{cz+d} - \frac{a\bar{z}+b}{c\bar{z}+d} = \frac{acz\bar{z} + bc\bar{z} + azd + bd - acz\bar{z} - ad\bar{z} - bcz - bd}{(cz+d)(c\bar{z}+d)} \\ &= \frac{(ad-bc)(z-\bar{z})}{(cz+d)(c\bar{z}+d)} = \frac{2i\mathrm{Im}\, z}{(cz+d)(c\bar{z}+d)}, \end{aligned}$$

hence

$$\frac{dw d\bar{w}}{(\mathrm{Im}\, w)^2} = \frac{dz d\bar{z}}{(cz+d)^2(c\bar{z}+d)^2} \frac{(cz+d)^2(c\bar{z}+d)^2}{(\mathrm{Im}\, z)^2} = \frac{d\bar{z}\, dz}{(\mathrm{Im}\, z)^2}.$$

In particular, the geodesics are preserved under the Möbius transformations from $PSL(2, \mathbb{R})$. Since the metric does not depend on $x$ it follows that all vertical lines are geodesics. In order to see what the general geodesics are, note that any two points $z_{1,2}$ in $\mathbb{H}$ lie on a unique circle $S_{12}$ that is perpendicular to the real axis. It is easy to see that there exists a Möbius transformation $T_{12}$ from $PSL(2, \mathbb{R})$ that maps one of the intersection points of $S_{12}$ and $\mathbb{R}$ to infinity and, in addition, preserves $\mathbb{H}$. As $T_{12}$ preserves angles, it maps $S_{12}$ to a line perpendicular to $\mathbb{R}$, that is, to a geodesic in $\mathbb{H}$. As Möbius transformations from $PSL(2, \mathbb{R})$ map geodesics to geodesics, it follows that $S_{12}$ itself is a geodesic. Hence, we have shown that the geodesics of $\mathbb{H}$ are precisely all circles which intersect the real line at a right angle (with the vertical lines being counted as circles of infinite radius).

It is clear from the above that the hyperbolic plane satisfies all axioms of Euclidean geometry except for the axiom of parallel lines: there are many "lines" (i.e., geodesics) passing through a point which is not on a fixed geodesic that do not intersect that geodesic. Let us now prove the famous Gauss-Bonnet theorem which describes the hyperbolic area of a triangle whose three sides are geodesics (those are called geodesic triangles).



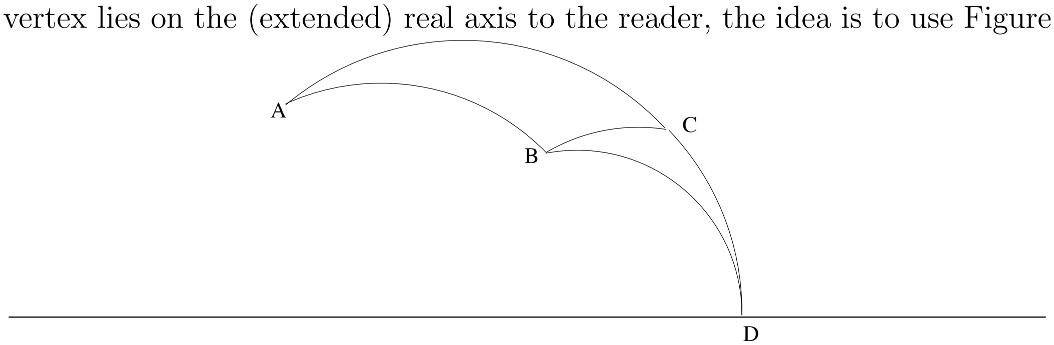
**Theorem 1.34** *Let $T$ be a geodesic triangle with angles $\alpha_1, \alpha_2, \alpha_3$, then*

$$\text{Area}(T) = \pi - (\alpha_1 + \alpha_2 + \alpha_3).$$

*Proof.* There are four essentially distinct types of geodesic triangles, depending on how many of its vertices lie on the real line. Up to equivalences via transformations in $PSL(2, \mathbb{R})$ (which are isometries and therefore also preserve the area) we see that it suffices to consider precisely those cases described in Figure 1.3. Let us start with the case in which exactly two vertices belong to $\mathbb{R}$ as shown in that figure (the second triangle from the right). Without loss of generality one vertex coincides with 1, the other with $\infty$, and the circular arc lies on the unit circle with the projection of the second finite vertex onto the real axis being $x_0$. Then

$$\text{Area}(T) = \int_{x_0}^{1} \int_{\sqrt{1-x^2}}^{\infty} \frac{dxdy}{y^2} = \int_{x_0}^{1} \frac{dx}{\sqrt{1-x^2}} = \int_{\alpha_0}^{0} \frac{d\cos\phi}{\sqrt{1-\cos^2(\phi)}} = \alpha_0 = \pi - \alpha_1$$

as desired since the other two angles are zero. By additivity of the area we can deal with the other two cases in which at least one vertex is real. We leave the case where no vertex lies on the (extended) real axis to the reader, the idea is to use Figure 1.4. $\square$



# 2 Properties of Holomorphic Functions

## 2.1 The Integral

**Definition 2.1** *Let $\gamma : I \to \mathbb{C}$ be a piecewise smooth path, where $I = [\alpha, \beta]$ is an interval on the real axis. Let a complex-valued function $f$ be defined on $\gamma(I)$ so that the function $f \circ \gamma$ is a continuous function on $I$. The integral of $f$ along the path $\gamma$ is*

$$\int_\gamma f dz = \int_\alpha^\beta f(\gamma(t))\gamma'(t)dt. \tag{2.1}$$

*The integral in the right side of (2.1) is understood to be $\int_\alpha^\beta g_1(t)dt + i\int_\alpha^\beta g_2(t)dt$, where $g_1$ and $g_2$ are the real and imaginary parts of the function $f(\gamma(t))\gamma'(t) = g_1(t) + ig_2(t)$.*

Note that the functions $g_1$ and $g_2$ may have only finitely many discontinuities on $I$ so that the integral (2.1) exists in the usual Riemann integral sense.

**Example 2.2** Let $\gamma$ be a circle $\gamma(t) = a + re^{it}$, $t \in [0, 2\pi]$, and $f(z) = (z-a)^n$, where $n = 0, \pm 1, \ldots$ is an integer. Then we have $\gamma'(t) = re^{it}$, $f(\gamma(t)) = r^n e^{int}$ so that

$$\int_\gamma (z-a)^n dz = r^{n+1} i \int_0^{2\pi} e^{i(n+1)t} dt.$$

We have to consider two cases: when $n \neq -1$ we have

$$\int_\gamma (z-a)^n dz = r^{n+1} \frac{e^{2\pi i(n+1)} - 1}{n+1} = 0,$$

because of the periodicity of the exponential function, while when $n = -1$

$$\int_\gamma \frac{dz}{z-a} = i \int_0^{2\pi} dt = 2\pi i.$$

Therefore the integer powers of $z - a$ have the "orthogonality" property

$$\int_\gamma (z-a)^n = \begin{cases} 0, & \text{if } n \neq -1 \\ 2\pi i, & \text{if } n = -1 \end{cases} \tag{2.2}$$

that we will use frequently.

**Example 2.3** Let $\gamma : I \to \mathbb{C}$ be an arbitrary piecewise smooth path and $n \neq 1$. We also assume that the path $\gamma(t)$ does not pass through the point $z = 0$ in the case $n < 0$. The chain rule implies that $\frac{d}{dt}\gamma^{n+1}(t) = (n+1)\gamma^n(t)\gamma'(t)$ so that

$$\int_\gamma z^n dz = \int_\alpha^\beta \gamma^n(t)\gamma'(t)dt = \frac{1}{n+1}[\gamma^{n+1}(\beta) - \gamma^{n+1}(\alpha)]. \tag{2.3}$$

We observe that the integrals of $z^n$, $n \neq -1$ depend not on the path but only on its endpoints. Their integrals over a closed path vanish.

Integral is invariant under a re-parameterization of the path.

**Theorem 2.4** *Let a path $\gamma_1 : [\alpha_1, \beta_1] \to \mathbb{C}$ be obtained from a piecewise smooth path $\gamma : [\alpha, \beta] \to \mathbb{C}$ by a legitimate re-parameterization, that is $\gamma = \gamma_1 \circ \tau$ where $\tau$ is an increasing piecewise smooth map $\tau : [\alpha, \beta] \to [\alpha_1, \beta_1]$. Then we have for any function $f$ that is continuous on $\gamma$ (and hence on $\gamma_1$):*

$$\int_{\gamma_1} f dz = \int_\gamma f dz. \tag{2.4}$$

**Proof.** The definition of the integral implies that

$$\int_{\gamma_1} f dz = \int_{\alpha_1}^{\beta_1} f(\gamma_1(s))\gamma_1'(s) ds.$$

Introducing the new variable $t$ so that $\tau(t) = s$ and using the usual rules for the change of real variables in an integral we obtain

$$\begin{aligned}\int_{\gamma_1} f dz &= \int_{\alpha_1}^{\beta_1} f(\gamma_1(s))\gamma_1'(s) ds = \int_\alpha^\beta f(\gamma_1(\tau(t)))\gamma_1'(\tau(t))\tau'(t) dt \\ &= \int_\alpha^\beta f(\gamma(t))\gamma'(t) dt = \int_\gamma f dz. \;\square\end{aligned}$$

This theorem has an important corollary: the integral that we defined for a path makes sense also for a *curve* that is an equivalence class of paths. More precisely, the value of the integral along any path that defines a given curve is independent of the choice of path in the equivalence class of the curve.

**Theorem 2.5** *Let $f$ be a continuous function defined on a piecewise smooth path $\gamma : [\alpha, \beta] \to \mathbb{C}$. Then the following inequality holds:*

$$\left| \int_\gamma f dz \right| \leq \int_\gamma |f||d\gamma|, \tag{2.5}$$

*where $|d\gamma| = |\gamma'(t)|dt$ is the differential of the arc length of $\gamma$ and the integral on the right side is the real integral along a curve.*

**Proof.** Let us denote $J = \displaystyle\int_\gamma f dz$ and let $J = |J|e^{i\theta}$, then we have

$$|J| = \int_\gamma e^{-i\theta} f dz = \int_\alpha^\beta e^{-i\theta} f(\gamma(t))\gamma'(t) dt.$$

The integral on the right side is a real number and hence

$$|J| = \int_\alpha^\beta \mathrm{Re}\left[e^{-i\theta} f(\gamma(t))\gamma'(t)\right] dt \leq \int_\alpha^\beta |f(\gamma(t))||\gamma'(t)| dt = \int_\gamma |f||d\gamma|. \;\square$$

**Corollary 2.6** *Let assumptions of the previous theorem hold and assume that $|f(z)| \leq M$ for a constant $M$, then*

$$\left| \int_\gamma f dz \right| \leq M|\gamma|, \tag{2.6}$$

*where $|\gamma|$ is the length of the path $\gamma$.*

Inequality (2.6) is obtained from (2.5) if we estimate the integral on the right side of (2.5) and note that $\displaystyle\int_\gamma |d\gamma| = |\gamma|$.

### 2.1.1 The anti-derivative

**Definition 2.7** *An anti-derivative of a function $f$ in a domain $D$ is a holomorphic function $F$ such that at every point $z \in D$ we have*

$$F'(z) = f(z). \tag{2.7}$$

Let us now address the existence of anti-derivative. First we will look at the question of existence of a local anti-derivative that exists in a neighborhood of a point. We begin with a theorem that expresses in the simplest form the Cauchy theorem that lies at the core of the theory of integration of holomorphic functions.

**Theorem 2.8** *(Cauchy) Let $f \in \mathcal{O}(D)$, that is, $f$ is holomorphic in $D$. Then the integral of $f$ along the oriented boundary[3] of any triangle $\Delta$ that is properly contained[4] in $D$ is equal to zero:*

$$\int_{\partial\Delta} f dz = 0. \tag{2.8}$$

**Proof.** Let us assume that this is false, that is, there exists a triangle $\Delta$ properly contained in $D$ so that

$$\left| \int_{\partial\Delta} f dz \right| = M > 0. \tag{2.9}$$

Let us divide $\Delta$ into four sub-triangles by connecting the midpoints of all sides and assume that the boundaries both of $\Delta$ and these triangles are oriented counter-clockwise. Then clearly the integral of $f$ over $\partial\Delta$ is equal to the sum of the integrals over the boundaries of the small triangles since each side of a small triangle that is not part of the boundary $\partial\Delta$ belongs to two small triangles with two different orientations so that they do not contribute to the sum. Therefore there exists at least one small triangle that we denote $\Delta_1$ so that

$$\left| \int_{\partial\Delta_1} f dz \right| \geq \frac{M}{4}.$$

[3]We assume that the boundary $\partial\Delta$ (that we treat as a piecewise smooth curve) is oriented in such a way that the triangle $\Delta$ remains on one side of $\partial\Delta$ when one traces $\partial\Delta$.

[4]A set $S$ is properly contained in a domain $S'$ if $S$ is contained in a compact subset of $S'$.

We divide $\Delta_1$ into four smaller sub-triangles and using the same considerations we find one of them denoted $\Delta_2$ so that $\left|\int_{\partial\Delta_2} f dz\right| \geq \frac{M}{4^2}$.

Continuing this procedure we construct a sequence of nested triangles $\Delta_n$ so that

$$\left|\int_{\partial\Delta_n} f dz\right| \geq \frac{M}{4^n}. \tag{2.10}$$

The closed triangles $\Delta_n$ have a common point $z_0 \in \Delta \subset D$. The function $f$ is holomorphic at $z_0$ and hence for any $\varepsilon > 0$ there exists $\delta > 0$ so that we may decompose

$$f(z) - f(z_0) = f'(z_0)(z - z_0) + \alpha(z)(z - z_0) \tag{2.11}$$

with $|\alpha(z)| < \varepsilon$ for all $z \in U = \{|z - z_0| < \delta\}$.

We may find a triangle $\Delta_n$ that is contained in $U$. Then (2.11) implies that

$$\int_{\partial\Delta_n} f dz = \int_{\partial\Delta_n} f(z_0)dz + \int_{\partial\Delta_n} f'(z_0)(z - z_0)dz + \int_{\partial\Delta_n} \alpha(z)(z - z_0)dz.$$

However, the first two integrals on the right side vanish since the factors $f(z_0)$ and $f'(z_0)$ may be pulled out of the integrals and the integrals of 1 and $z - z_0$ over a closed path $\partial\Delta_n$ are equal to zero (see Example 2.3). Therefore, we have $\int_{\partial\Delta_n} f dz = \int_{\partial\Delta_n} \alpha(z)(z - z_0)dz$, where $|\alpha(z)| < \varepsilon$ for all $z \in \partial\Delta_n$. Furthermore, we have $|z - z_0| \leq |\partial\Delta_n|$ for all $z \in \partial\Delta_n$ and hence we obtain using Theorem 2.5

$$\left|\int_{\partial\Delta_n} f dz\right| = \left|\int_{\partial\Delta_n} \alpha(z)(z - z_0)dz\right| < \varepsilon|\partial\Delta_n|^2.$$

However, by construction we have $|\partial\Delta_n| = |\partial\Delta|/2^n$, where $|\partial\Delta|$ is the perimeter of $\Delta$, so that

$$\left|\int_{\partial\Delta_n} f dz\right| < \varepsilon|\partial\Delta|^2/4^n.$$

This together with (2.10) implies that $M < \varepsilon|\partial\Delta|^2$ which in turn implies $M = 0$ since $\varepsilon$ is an arbitrary positive number. This contradicts assumption (2.9) and the conclusion of Theorem 2.8 follows. $\square$

We will consider the Cauchy theorem in its full generality in the next section. At the moment we will deduce the local existence of anti-derivative from the above Theorem.

**Theorem 2.9** *Let $f \in \mathcal{O}(D)$ then it has an anti-derivative in any disk $U = \{|z - a| < r\} \subset D$:*

$$F(z) = \int\limits_{[a,z]} f(\zeta)d\zeta, \tag{2.12}$$

*where the integral is taken along the straight segment $[a, z] \subset U$.*

**Proof.** We fix an arbitrary point $z \in U$ and assume that $|\Delta z|$ is so small that the point $z + \Delta z \in U$. Then the triangle $\Delta$ with vertices $a$, $z$ and $z + \Delta z$ is properly contained in $D$ so that Theorem 2.8 implies that

$$\int_{[a,z]} f(\zeta)d\zeta + \int_{[z,z+\Delta z]} f(\zeta)d\zeta + \int_{[z+\Delta z,a]} f(\zeta)d\zeta = 0.$$

The first term above is equal to $F(z)$ and the third to $-F(z + \Delta z)$ so that

$$F(z + \Delta z) - F(z) = \int_{[z,z+\Delta z]} f(\zeta)d\zeta. \tag{2.13}$$

On the other hand we have

$$f(z) = \frac{1}{\Delta z} \int_{[z,z+\Delta z]} f(z)d\zeta$$

(we have pulled the constant factor $f(z)$ out of the integral sign above), which allows us to write

$$\frac{F(z + \Delta z) - F(z)}{\Delta z} - f(z) = \frac{1}{\Delta z} \int_{[z,z+\Delta z]} [f(\zeta) - f(z)]d\zeta. \tag{2.14}$$

We use now continuity of the function $f$: for any $\varepsilon > 0$ we may find $\delta > 0$ so that if $|\Delta z| < \delta$ then we have $|f(\zeta) - f(z)| < \varepsilon$ for all $\zeta \in [z, z + \Delta z]$. We conclude from (2.14) that

$$\left|\frac{F(z + \Delta z) - F(z)}{\Delta z} - f(z)\right| < \frac{1}{|\Delta z|}\varepsilon|\Delta z| = \varepsilon$$

provided that $|\Delta z| < \delta$. The above implies that $F'(z)$ exists and is equal to $f(z)$. □

**Remark 2.10** We have used only two properties of the function $f$ in the proof of Theorem 2.9: $f$ is continuous and its integral over any triangle $\Delta$ that is contained properly in $D$ vanishes. Therefore we may claim that the function $F$ defined by (2.12) is a local anti-derivative of any function $f$ that has these two properties.

The problem of existence of a global anti-derivative in the whole domain $D$ is somewhat more complicated. We will address it only in the next section, and now will just show how an anti-derivative that acts along a given path may be glued together out of local anti-derivatives.

**Definition 2.11** *Let a function $f$ be defined in a domain $D$ and let $\gamma : I = [\alpha, \beta] \to D$ be an arbitrary continuous path. A function $\Phi : I \to \mathbb{C}$ is* an anti-derivative of $f$ along the path $\gamma$ *if (i) $\Phi$ is continuous on $I$, and (ii) for any $t_0 \in I$ there exists a neighborhood $U \subset D$ of the point $z_0 = \gamma(t_0)$ so that $f$ has an anti-derivative $F_U$ in $U$ such that*

$$F_U(\gamma(t)) = \Phi(t) \tag{2.15}$$

*for all $t$ in a neighborhood $u_{t_0} \subset I$.*

We note that if $f$ has an anti-derivative $F$ in the whole domain $D$ then the function $F(\gamma(t))$ is an anti-derivative along the path $\gamma$. However, the above definition does not require the existence of a global anti-derivative in all of $D$ – it is sufficient for it to exist *locally*, in a neighborhood of each point $z_0 \in \gamma$. Moreover, if $\gamma(t') = \gamma(t'')$ with $t' \neq t''$ then the two anti-derivatives of $f$ that correspond to the neighborhoods $u_{t'}$ and $u_{t''}$ need not coincide: they may differ by a constant (observe that they are anti-derivatives of $f$ in a neighborhood of the same point $z'$ and hence their difference is a constant). Therefore anti-derivative along a path being a function of the parameter $t$ might not be a function of the point $z$.

**Theorem 2.12** *Let $f \in \mathcal{O}(D)$ and let $\gamma: I \to D$ be a continuous path. Then the anti-derivative of $f$ along $\gamma$ exists and is defined up to a constant.*

**Proof.** Let us divide the interval $I = [\alpha, \beta]$ into $n$ sub-intervals $I_k = [t_k, t_k']$ so that each pair of adjacent sub-intervals overlap on an interval ($t_k < t_{k-1}' < t_{k+1} < t_k'$, $t_1 = \alpha$, $t_n' = \beta$). Using uniform continuity of the function $\gamma(t)$ we may choose $I_k$ so small that the image $\gamma(I_k)$ is contained in a disk $U_k \subset D$. Theorem 2.8 implies that $f$ has an anti-derivative $F$ in each disk $U_k$. Let us choose arbitrarily an anti-derivative of $f$ in $U_1$ and denote it $F_1$. Consider an anti-derivative of $f$ defined in $U_2$. It may differ only by a constant from $F_1$ in the intersection $U_1 \cap U_2$. Therefore we may choose the anti-derivative $F_2$ of $f$ in $U_2$ that coincides with $F_1$ in $U_1 \cap U_2$.

We may continue in this fashion choosing the anti-derivative $F_k$ in each $U_k$ so that $F_k = F_{k-1}$ in the intersection $U_{k-1} \cap U_k$, $k = 1, 2, \dots, n$. The function

$$\Phi(t) = F_k \circ \gamma(t), \quad t \in I_k, \quad k = 1, 2, \dots, n,$$

is an anti-derivative of $f$ along $\gamma$. Indeed it is clearly continuous on $\gamma$ and for each $t_0 \in I$ one may find a neighborhood $u_{t_0}$ where $\Phi(t) = F_U \circ \gamma(t)$ where $F_U$ is an anti-derivative of $f$ in a neighborhood of the point $\gamma(t_0)$.

It remains to prove the second part of the theorem. Let $\Phi_1$ and $\Phi_2$ be two anti-derivatives of $f$ along $\gamma$. We have $\Phi_1 = F^{(1)} \circ \gamma(t)$, $\Phi_2 = F^{(2)} \circ \gamma(t)$ in a neighborhood $u_{t_0}$ of each point $t_0 \in I$. Here $F^{(1)}$ and $F^{(2)}$ are two anti-derivatives of $f$ defined in a neighborhood of the point $\gamma(t_0)$. They may differ only by a constant so that $\phi(t) = \Phi_1(t) - \Phi_2(t)$ is constant in a neighborhood $u_{t_0}$ of $t_0$. However, a locally constant function defined on a connected set is constant on the whole set [5]. Therefore $\Phi_1(t) - \Phi_2(t) = \text{const}$ for all $t \in I$. □

If the anti-derivative of $f$ along a path $\gamma$ is known then the integral of $f$ over $\gamma$ is computed using the usual Newton-Leibnitz formula.

**Theorem 2.13** *Let $\gamma: [\alpha, \beta] \to \mathbb{C}$ be a piecewise smooth path and let $f$ be continuous on $\gamma$ and have an anti-derivative $\Phi(t)$ along $\gamma$, then*

$$\int_\gamma f dz = \Phi(\beta) - \Phi(\alpha). \tag{2.16}$$

[5]Indeed, let $E = \{t \in I : \phi(t) = \phi(t_0)\}$. This set is not empty since it contains $t_0$. It is open since $\phi$ is locally constant so that if $t \in E$ and $\phi(t) = \phi(t_0)$ then $\phi(t') = \phi(t) = \phi(t_0)$ for all $t'$ in a neighborhood $u_t$ and thus $u_t \subset E$. However, it is also closed since $\phi$ is a continuous function (because it is locally constant) so that $\phi(t_n) = \phi(t_0)$ and $t_n \to t''$ implies $\phi(t'') = \phi(t_0)$. Therefore $E = I$.

**Proof.** Let us assume first that $\gamma$ is a smooth path and its image is contained in a domain $D$ where $f$ has an anti-derivative $F$. Then the function $F \circ \gamma$ is an anti-derivative of $f$ along $\gamma$ and hence differs from $\Phi$ only by a constant so that $\Phi(t) = F \circ \gamma(t) + C$. Since $\gamma$ is a smooth path and $F'(z) = f(z)$ the derivative $\Phi'(t) = f(\gamma(t))\gamma'(t)$ exists and is continuous at all $t \in [\alpha, \beta]$. However, using the definition of the integral we have

$$\int_\gamma f dz = \int_\alpha^\beta f(\gamma(t))\gamma'(t) dt = \int_\alpha^\beta \Phi'(t) dt = \Phi(\beta) - \Phi(\alpha)$$

and the theorem is proved in this particular case.

In the general case we may divide $\gamma$ into a finite number of paths $\gamma_\nu : [\alpha_\nu, \alpha_{n+1}] \to \mathbb{C}$ ($\alpha_0 = \alpha < \alpha_1 < \alpha_2 < \cdots < \alpha_n = \beta$) so that each of them is smooth and is contained in a domain where $f$ has an anti-derivative. As we have just shown,

$$\int_{\gamma_\nu} f dz = \Phi(\alpha_{\nu+1}) - \Phi(\alpha_\nu),$$

and summing over $\nu$ we obtain (2.16). $\square$

**Remark 2.14** We may extend our definition of the integral to continuous paths (from piecewise smooth) by defining the integral of $f$ over an arbitrary continuous path $\gamma$ as the increment of its anti-derivative along the this path over the interval $[\alpha, \beta]$ of the parameter change. Clearly the right side of (2.16) does not change under a re-parameterization of the path. Therefore one may consider integrals of holomorphic functions over arbitrary continuous curves.

**Remark 2.15** Theorem 2.13 allows us to verify that a holomorphic function might have no global anti-derivative in a domain that is not simply connected. Let $D = \{0 < |z| < 2\}$ be a punctured disk and consider the function $f(z) = 1/z$ that is holomorphic in $D$. This function may not have an anti-derivative in $D$. Indeed, were the anti-derivative $F$ of $f$ to exist in $D$, the function $F(\gamma(t))$ would be an anti-derivative along any path $\gamma$ contained in $D$. Theorem 2.13 would imply that

$$\int_\gamma f dz = F(b) - F(a),$$

where $a = \gamma(\alpha)$, $b = \gamma(\beta)$ are the end-points of $\gamma$. In particular the integral of $f$ along any closed path $\gamma$ would vanish. However, we know that the integral of $f$ over the unit circle is

$$\int_{|z|=1} f dz = 2\pi i.$$

## 2.2 The Cauchy Theorem

We will prove now the Cauchy theorem in its general form – the basic theorem of the theory of integration of holomorphic functions (we have proved it in its simplest form in the last section). This theorem claims that the integral of a function holomorphic in some domain does not change if the path of integration is changed continuously inside the domain provided that its end-points remain fixed or a closed path remains closed. We have to define first what we mean by a continuous deformation of a path. We assume for simplicity that all our paths are parameterized so that $t \in I = [0, 1]$. This assumption may be made without any loss of generality since any path may be re-parameterized in this way without changing the equivalence class of the path and hence the value of the integral.

**Definition 2.16** *Two paths $\gamma_0 : I \to D$ and $\gamma_1 : I \to D$ with common ends $\gamma_0(0) = \gamma_1(0) = a$, $\gamma_0(1) = \gamma_1(1) = b$ are* homotopic *to each other in a domain $D$ if there exists a continuous map $\gamma(s, t) : I \times I \to D$ so that*

$$\begin{aligned} \gamma(0,t) = \gamma_0(t), \quad \gamma(1,t) = \gamma_1(t), \quad t \in I \\ \gamma(s,0) = a, \qquad \gamma(s,1) = b, \qquad s \in I. \end{aligned} \tag{2.17}$$

The function $\gamma(s_0, t) : I \to D$ defines a path inside in the domain $D$ for each fixed $s_0 \in I$. These paths vary continuously as $s_0$ varies and their family "connects" the paths $\gamma_0$ and $\gamma_1$ in $D$. Therefore the homotopy of two paths in $D$ means that one path may be deformed continuously into the other inside $D$.

Similarly two closed paths $\gamma_0 : I \to D$ and $\gamma_1 : I \to D$ are homotopic in a domain $D$ if there exists a continuous map $\gamma(s, t) : I \times I \to D$ so that

$$\begin{aligned} \gamma(0,t) = \gamma_0(t), \quad \gamma(1,t) = \gamma_1(t), \quad t \in I \\ \gamma(s,0) = \gamma(s,1), \qquad s \in I. \end{aligned} \tag{2.18}$$

Homotopy is usually denoted by the symbol $\sim$, we will write $\gamma_0 \sim \gamma_1$ if $\gamma_0$ is homotopic to $\gamma_1$.

It is quite clear that homotopy defines an equivalence relation. Therefore all paths with common end-points and all closed paths may be separated into equivalence classes. Each class contains all paths that are homotopic to each other.

A special homotopy class is that of paths homotopic to zero. We say that a closed path $\gamma$ is homotopic to zero in a domain $D$ if there exists a continuous mapping $\gamma(s, t) : I \times I \to D$ that satisfies conditions (2.18) and such that $\gamma_1(t) = \text{const}$. That means that $\gamma$ may be contracted to a point by a continuous transformation.

Any closed path is homotopic to zero in a simply connected domain, and thus any two paths with common ends are homotopic to each other. Therefore the homotopy classes in a simply connected domains are trivial.

We have introduced the notion of the integral first for a path and then verified that the value of the integral is determined not by a path but by a curve, that is, by an equivalence class of paths. The general Cauchy theorem claims that integral of a holomorphic function is determined not even by a curve but by the homotopy class of the curve. In other words, the following theorem holds.

**Theorem 2.17** *(Cauchy) Let $f \in \mathcal{O}(D)$ and $\gamma_0$ and $\gamma_1$ be two paths homotopic to each other in $D$ either as paths with common ends or as closed paths, then*

$$\int_{\gamma_0} f dz = \int_{\gamma_1} f dz. \tag{2.19}$$

**Proof.** Let $\gamma : I \times I \to D$ be a function that defines the homotopy of the paths $\gamma_0$ and $\gamma_1$. We construct a system of squares $K_{mn}$, $m, n = 1, \dots, N$ that covers the square $K = I \times I$ so that each $K_{mn}$ overlaps each neighboring square. Uniform continuity of the function $\gamma$ implies that the squares $K_{mn}$ may be chosen so small that the image of each $K_{mn}$ is contained in a disk $U_{mn} \subset D$. The function $f$ has an anti-derivative $F_{mn}$ in each of those disks (we use the fact that a holomorphic function has an anti-derivative in any disk). We fix the subscript $m$ and proceed as in the proof of Theorem 2.12. We choose arbitrarily the anti-derivative $F_{m1}$ defined in $U_{m1}$ and pick the anti-derivative $F_{m2}$ defined in $U_{m2}$ so that $F_{m1} = F_{m2}$ in the intersection $U_{m1} \cap U_{m2}$. Similarly we may choose $F_{m3}, \dots, F_{mN}$ so that $F_{m,n+1} = F_{mn}$ in the intersection $U_{m,n+1} \cap U_{mn}$ and define the function

$$\Phi_m(s,t) = F_{mn} \circ \gamma(s,t) \text{ for } (s,t) \in K_{mn},\ n = 1, \dots, N. \tag{2.20}$$

The function $\Phi_{mn}$ is clearly continuous in the rectangle $K_m = \cup_{n=1}^{N} K_{mn}$ and is defined up to an arbitrary constant. We choose arbitrarily $\Phi_1$ and pick $\Phi_2$ so that $\Phi_1 = \Phi_2$ in the intersection $K_1 \cap K_2$ [6]. The functions $\Phi_3, \dots, \Phi_N$ are chosen in exactly the same fashion so that $\Phi_{m+1} = \Phi_m$ in $K_{m+1} \cap K_m$. This allows us to define the function

$$\Phi(s,t) = \Phi_m(s,t) \text{ for } (s,t) \in K_m,\ m = 1, \dots, N. \tag{2.21}$$

the function $\Phi(s,t)$ is clearly an anti-derivative along the path $\gamma_s(t) = \gamma(s,t) : I \to D$ for each fixed $s$. Therefore the Newton-Leibnitz formula implies that

$$\int_{\gamma_s} f dz = \Phi(s,1) - \Phi(s,0). \tag{2.22}$$

We consider now two cases separately.

(a) *The paths $\gamma_0$ and $\gamma_1$ have common ends.* Then according to the definition of homotopy we have $\gamma(s,0) = a$ and $\gamma(s,1) = b$ for all $s \in I$. Therefore the functions $\Phi(s,0)$ and $\Phi(s,1)$ are locally constant as functions of $s \in I$ at all $s$ and hence they are constant on $I$. Therefore $\Phi(0,0) = \Phi(1,0)$ and $\Phi(1,0) = \Phi(1,1)$ so that (2.22) implies (2.19). $\square$

(b) The paths $\gamma_0$ and $\gamma_1$ are closed. In this case we have $\gamma(s,0) = \gamma(s,1)$ so that the function $\Phi(s,0) - \Phi(s,1)$ is locally constant on $I$, and hence this function is a constant on $I$. Therefore once again (2.22) implies (2.19).

[6] This is possible since the function $\Phi_1 - \Phi_2$ is locally constant on a connected set $K_1 \cap K_2$ and is therefore constant on this set

## 2.3 Some special cases

We consider in this section some special cases of the Cauchy theorem that are especially important and deserve to be stated separately.

**Theorem 2.18** *Let $f \in \mathcal{O}(D)$ then its integral along any path that is contained in $D$ and is homotopic to zero vanishes:*

$$\int_\gamma f dz = 0 \text{ if } \gamma \sim 0. \tag{2.23}$$

**Proof.** Since $\gamma \sim 0$ this path may be continuously deformed into a point $a \in D$ and thus into a circle $\gamma_\varepsilon = \{|z-a| = \varepsilon\}$ of an arbitrarily small radius $\varepsilon > 0$. The general Cauchy theorem implies that

$$\int_\gamma f dz = \int_{\gamma_\varepsilon} f dz.$$

The integral on the right side vanishes in the limit $\varepsilon \to 0$ since the function $f$ is bounded in a neighborhood of the point $a$. However, the left side is independent of $\varepsilon$ and thus it must be equal to zero. □

Any closed path is homotopic to zero in a simply connected domain and thus the Cauchy theorem has a particularly simple form for such domains - this is its classical statement:

**Theorem 2.19** *If a function $f$ is holomorphic in a simply connected domain $D \subset \mathbb{C}$ then its integral over any closed path $\gamma : I \to D$ vanishes.*

It is easy to deduce from the Cauchy theorem the global theorem of existence of an anti-derivative in a *simply connected* domain.

**Theorem 2.20** *Any function $f$ holomorphic in a simply connected domain $D$ has an anti-derivative in this domain.*

**Proof.** We first show that the integral of $f$ along a path in $D$ is independent of the choice of the path and is completely determined by the end-points of the path. Indeed, let $\gamma_1$ and $\gamma_2$ be two paths that connect in $D$ two points $a$ and $b$. Without any loss of generality we may assume that the path $\gamma_1$ is parameterized on an interval $[\alpha, \beta_1]$ and $\gamma_2$ is parameterized on an interval $[\beta_1, \beta]$, $\alpha < \beta_1 < \beta$. Let us denote by $\gamma$ the union of the paths $\gamma_1$ and $\gamma_2^-$, this is a closed path contained in $\gamma$, and, moreover,

$$\int_\gamma f dz = \int_{\gamma_1} f dz - \int_{\gamma_2} f dz.$$

However, Theorem 2.19 integral of $f$ over any closed path vanishes and this implies our claim[7].

[7] One may also obtain this result directly from the general Cauchy theorem using the fact that any two paths with common ends are homotopic to each other in a simply connected domain.

We fix now a point $a \in D$ and let $z$ be a point in $D$. Integral of $f$ over any path $\gamma = \widetilde{az}$ that connects $a$ and $z$ depends only on $z$ and not on the choice of $\gamma$:

$$F(z) = \int\limits_{\widetilde{az}} f(\zeta) d\zeta. \tag{2.24}$$

Repeating verbatim the arguments in the proof of Theorem 2.9 we verify that $F(z)$ is holomorphic in $D$ and $F'(z) = f(z)$ for all $z \in D$ so that $F$ is an anti-derivative of $f$ in $D$. $\Box$

The example of the function $f = 1/z$ in an annulus $\{0 < |z| < 2\}$ (see Remark 2.15) shows that the assumption that $D$ is simply connected is essential: the global existence theorem of anti-derivative does not hold in general for multiply connected domains.

The same example shows that the integral of a holomorphic function over a closed path in a multiply connected domain might not vanish, so that the Cauchy theorem in its classical form (Theorem 2.19) may not be extended to non-simply connected domains. However, one may present a reformulation of this theorem that allows such a generalization.

**Definition 2.21** *Let the boundary of a compact domain $D$[8] consist of a finite number of closed curves $\gamma_\nu$, $\nu = 0, \dots, n$. We assume that the outer boundary $\gamma_0$, that is, the curve that separates $D$ from infinity, is oriented counterclockwise while the other boundary curves $\gamma_\nu$, $\nu = 1, \dots, n$ are oriented clockwise. In other words, all the boundary curves are oriented in such a way that $D$ remains on the left side as they are traced. The boundary of $D$ with this orientation is called* the oriented boundary *and denote by $\partial D$.*

We may now state the Cauchy theorem for multiply connected domains as follows.

**Theorem 2.22** *Let a compact domain $D$ be bounded by a finite number of continuous curves and let $f$ be holomorphic in its closure $\bar{D}$. Then the integral of $f$ over its oriented boundary $\partial D$ is equal to zero:*

$$\int_{\partial D} f dz = \int_{\gamma_0} f dz + \sum_{\nu=1}^{n} \int_{\gamma_\nu} f dz = 0. \tag{2.25}$$

**Proof.** Let us introduce a finite number of cuts $\lambda_\nu^\pm$ that connect the components of the boundary of this domain. It is clear that the closed curve $\Gamma$ that consists of the oriented boundary $\partial D$ and the unions $\Lambda^+ = \cup \lambda_\nu^+$ and $\Lambda^- = \cup \lambda_\nu^-$ is homotopic to zero in the domain $G$ that contains $\bar{D}$, and such that $f$ is holomorphic in $D$. Theorem 2.18 implies that the integral of $f$ along $\Gamma$ vanishes so that

$$\int_\Gamma f dz = \int_{\partial D} f dz + \int_{\Lambda^+} f dz + \int_{\Lambda^-} f dz = \int_{\partial D} f dz$$

since the integrals of $f$ along $\Lambda^+$ and $\Lambda^-$ cancel each other.$\Box$

[8]Recall that a domain $D$ is compact if its closure does not contain the point at infinity.

## 2.4 The Cauchy Integral Formula

We will obtain here a representation of functions holomorphic in a compact domain with the help of the integral over the boundary of the domain.

**Theorem 2.23** *Let the function $f$ be holomorphic in the closure of a compact domain $D$ that is bounded by a finite number of continuous curves. Then the function $f$ at any point $z \in D$ may be represented as*

$$f(z) = \frac{1}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{\zeta - z} d\zeta, \tag{2.26}$$

*where $\partial D$ is the oriented boundary of $D$.*

The right side of (2.26) is called the Cauchy integral.

**Proof.** Let $\rho > 0$ be such that the disk $U_\rho = \{z' : |z - z'| < \rho\}$ is properly contained in $D$ and let $D_\rho = \bar{D} \backslash \bar{U}_\rho$. The function $g(\zeta) = \dfrac{f(\zeta)}{\zeta - z}$ is holomorphic in $\bar{D}_\rho$ as a ratio of two holomorphic functions with the numerator different from zero. The oriented boundary of $D_\rho$ consists of the union of $\partial D$ and the circle $\partial U_\rho = \{\zeta : |\zeta - z| = \rho\}$ oriented clockwise. Therefore we have

$$\frac{1}{2\pi i} \int_{\partial D_\rho} g(\zeta) d\zeta = \frac{1}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{\zeta - z} d\zeta - \frac{1}{2\pi i} \int_{\partial U_\rho} \frac{f(\zeta)}{\zeta - z} d\zeta.$$

However, the function $g$ is holomorphic in $\bar{D}_\rho$ (its singular point $\zeta = z$ lies outside this set) and hence the Cauchy theorem for multiply connected domains may be applied. We conclude that the integral of $g$ over $\partial D_\rho$ vanishes.

Therefore,

$$\frac{1}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{\zeta - z} d\zeta = \frac{1}{2\pi i} \int_{\partial U_\rho} \frac{f(\zeta)}{\zeta - z} d\zeta, \tag{2.27}$$

where $\rho$ may be taken arbitrarily small. Since the function $f$ is continuous at the point $z$, for any $\varepsilon > 0$ we may choose $\delta > 0$ so that

$$|f(\zeta) - f(z)| < \varepsilon \text{ for all } \zeta \in \partial U_\rho$$

for all $\rho < \delta$. Therefore the difference

$$f(z) - \frac{1}{2\pi i} \int_{\partial U_\rho} \frac{f(\zeta)}{\zeta - z} d\zeta = \frac{1}{2\pi i} \int_{\partial U_\rho} \frac{f(z) - f(\zeta)}{\zeta - z} d\zeta \tag{2.28}$$

does not exceed $\dfrac{1}{2\pi} \varepsilon \cdot 2\pi = \varepsilon$ and thus goes to zero as $\rho \to 0$. However, (2.27) shows that the left side in (2.28) is independent of $\rho$ and hence is equal to zero for all sufficiently small $\rho$, so that

$$f(z) = \frac{1}{2\pi i} \int_{\partial U_\rho} \frac{f(\zeta)}{\zeta - z} d\zeta.$$

This together with (2.27) implies (2.26). □

**Remark 2.24** If the point $z$ lies outside $\bar{D}$ and conditions of Theorem 2.23 hold then

$$\frac{1}{2\pi i}\int_{\partial D}\frac{f(\zeta)}{\zeta-z}d\zeta=0. \tag{2.29}$$

This follows immediately from the Cauchy theorem since now the function $g(\zeta)=\dfrac{f(\zeta)}{\zeta-z}$ is holomorphic in $\bar{D}$.

The integral Cauchy theorem expresses a very interesting fact: the values of a function $f$ holomorphic in a domain $\bar{G}$ are completely determined by its values on the boundary $\partial G$. Indeed, if the values of $f$ on $\partial G$ are given then the right side of (2.26) is known and thus the value of $f$ at any point $z\in D$ is also determined. This property is the main difference between holomorphic functions and differentiable functions in the real analysis sense.

**Exercise 2.25** Let the function $f$ be holomorphic in the closure of a domain $D$ that contains the point at infinity and the boundary $\partial D$ is oriented so that $D$ remains on the left as the boundary is traced. Show that then

$$f(z)=\frac{1}{2\pi i}\int_{\partial D}\frac{f(\zeta)}{\zeta-z}d\zeta+f(\infty).$$

An easy corollary of Theorem 2.23 is

**Theorem 2.26** *The value of the function $f\in\mathcal{O}(D)$ at each point $z\in D$ is equal to the average of its values on any sufficiently small circle centered at $z$:*

$$f(z)=\frac{1}{2\pi}\int_0^{2\pi}f(z+\rho e^{it})dt. \tag{2.30}$$

**Proof.** Consider the disk $U_\rho=\{z':\ |z-z'|<\rho\}$ so that $U_\rho$ is properly contained in $D$. The Cauchy integral formula implies that

$$f(z)=\frac{1}{2\pi i}\int_{\partial U_\rho}\frac{f(\zeta)}{\zeta-z}d\zeta. \tag{2.31}$$

Introducing the parameterization $\zeta=z+\rho e^{it}$, $t\in[0,2\pi]$ of $U_\rho$ and replacing $d\zeta=\rho ie^{it}dt$ we obtain (2.30) from (2.31). $\square$

The mean value theorem shows that holomorphic functions are built in a very regular fashion, so to speak, and their values are intricately related to the values at other points. This explains why these functions have specific properties that the real differentiable functions lack. We will consider many other such properties later.

Before we conclude we present an integral representation of $\mathbb{R}$-differentiable functions that generalizes the Cauchy integral formula.

**Theorem 2.27** *Let $f \in C^1(\bar{D})$ be a continuously differentiable function in the real sense in the closure of a compact domain $D$ bounded by a finite number of piecewise smooth curves. Then we have*

$$f(z) = \frac{1}{2\pi i}\int_{\partial D}\frac{f(\zeta)}{\zeta - z}d\zeta - \frac{1}{\pi}\iint_D \frac{\partial f}{\partial\bar{\zeta}}\frac{d\xi d\eta}{\zeta - z} \tag{2.32}$$

*for all $z \in D$ (here $\zeta = \xi + i\eta$ inside the integral).*

**Proof.** Let us delete a small disk $\bar{U}_\rho = \{\zeta : \ |\zeta - z| \le \rho\}$ out of $D$ and apply the Green's formula to the function $g(\zeta) = \dfrac{f(\zeta)}{\zeta - z}$ that is continuously differentiable in the domain $D_\rho = D\backslash\bar{U}_\rho$

$$\int_{\partial D}\frac{f(\zeta)}{\zeta - z}d\zeta - \int_{\partial U_\rho}\frac{f(\zeta)}{\zeta - z}d\zeta = 2i\iint_{D_\rho}\frac{\partial f}{\partial\bar{\zeta}}\frac{d\xi d\eta}{\zeta - z}{}^9. \tag{2.33}$$

The function $f$ is continuous at $z$ so that $f(\zeta) = f(z) + O(\rho)$ for $\zeta \in U_\rho$, where $O(\rho) \to 0$ as $\rho \to 0$, and thus

$$\int_{\partial U_\rho}\frac{f(\zeta)}{\zeta - z}d\zeta = f(z)\int_{\partial U_\rho}\frac{1}{\zeta - z}d\zeta + \int_{\partial U_\rho}\frac{O(\rho)}{\zeta - z}d\zeta = 2\pi i f(z) + O(\rho).$$

Passing to the limit in (2.33) and using the fact that the double integrals in (2.32) and (2.33) are convergent[10] we obtain (2.32). □

## 2.5 The Taylor series

We will obtain the representation of holomorphic functions as sums of power series (the Taylor series) in this section. Let us recall the simplest results regarding series familiar from the real analysis.

One of the main theorems of the complex analysis is

**Theorem 2.28** *Let $f \in \mathcal{O}(D)$ and let $z_0 \in D$ be an arbitrary point in $D$. Then the function $f$ may be represented as a sum of a convergent power series*

$$f(z) = \sum_{n=0}^{\infty} c_n(z - z_0)^n \tag{2.34}$$

*inside any disk $U = \{|z - z_0| < R\} \subset D$.*

---

[9] We have $\dfrac{\partial g}{\partial\bar{\zeta}} = \dfrac{1}{\zeta - z}\dfrac{\partial f}{\partial\bar{\zeta}}$ since the function $1/(\zeta - z)$ is holomorphic in $\zeta$ so that its derivative with respect to $\bar{\zeta}$ vanishes.

[10] Our argument shows that the limit $\lim\limits_{\rho\to 0}\iint_{D_\rho}\dfrac{\partial f}{\partial\bar{\zeta}}\dfrac{d\xi d\eta}{\zeta - z}$ exists. Moreover, since $f \in C^1(D)$ the double integral in (2.32) exists as can be easily seen by passing to the polar coordinates and thus this limit coincides with it.

**Proof.** Let $z \in U$ be an arbitrary point. Choose $r > 0$ so that $|z - z_0| < r < R$ and denote by $\gamma_r = \{\zeta : \ |\zeta - z_0| = r\}$ The integral Cauchy formula implies that

$$f(z) = \frac{1}{2\pi i} \int_{\gamma_r} \frac{f(\zeta)}{\zeta - z} d\zeta.$$

In order to represent $f$ as a power series let us represent the kernel of this integral as the sum of a geometric series:

$$\frac{1}{\zeta - z} = \left[ (\zeta - z_0) \left( 1 - \frac{z - z_0}{\zeta - z_0} \right) \right]^{-1} = \sum_{n=0}^{\infty} \frac{(z - z_0)^n}{(\zeta - z_0)^{n+1}}. \tag{2.35}$$

We multiply both sides by $\frac{1}{2\pi i} f(\zeta)$ and integrate the series term-wise along $\gamma_r$. The series (2.35) converges uniformly on $\gamma_r$ since

$$\left| \frac{z - z_0}{\zeta - z_0} \right| = \frac{|z - z_0|}{r} = q < 1$$

for all $\zeta \in \gamma_r$. Uniform convergence is preserved under multiplication by a continuous and hence bounded function $\frac{1}{2\pi i} f(\zeta)$. Therefore our term-wise integration is legitimate and we obtain

$$f(z) = \frac{1}{2\pi i} \int_{\gamma_r} \sum_{n=0}^{\infty} \frac{f(\zeta) d\zeta}{(\zeta - z_0)^{n+1}} (z - z_0)^n = \sum_{n=0}^{\infty} c_n (z - z_0)^n$$

where[11]

$$c_n = \frac{1}{2\pi i} \int_{\gamma_r} \frac{f(\zeta) d\zeta}{(\zeta - z_0)^{n+1}}, \quad n = 0, 1, \ldots . \square \tag{2.36}$$

**Definition 2.29** *The power series (2.34) with coefficients given by (2.36) is* the Taylor series *of the function $f$ at the point $z_0$ (or centered at $z_0$).*

The Cauchy theorem 2.17 implies that the coefficients $c_n$ of the Taylor series defined by (2.36) do not depend on the radius $r$ of the circle $\gamma_r$, $0 < r < R$.

**The Cauchy inequalities.** *Let the function $f$ be holomorphic in a closed disk $\bar{U} = \{|z - z_0| \leq r\}$ and let its absolute value on the circle $\gamma_r = \partial U$ be bounded by a constant $M$. Then the coefficients of the Taylor series of $f$ at $z_0$ satisfy the inequalities*

$$|c_n| \leq M / r^n, \quad (n = 0, 1, \ldots). \tag{2.37}$$

**Proof.** We deduce from (2.36) using the fact that $|f(\zeta)| \leq M$ for all $\zeta \in \gamma_r$:

$$|c_n| \leq \frac{1}{2\pi} \frac{M}{r^{n+1}} 2\pi r = \frac{M}{r^n}. \square$$

[11]This theorem was presented by Cauchy in 1831 in Turin. Its proof was first published in Italy, and it appeared in France in 1841. However, Cauchy did not justify the term-wise integration of the series. This caused a remark by Chebyshev in his paper from 1844 that such integration is possible only in some "particular cases".

**Exercise 2.30** Let $P(z)$ be a polynomial in $z$ of degree $n$. Show that if $|P(z)| \leq M$ for $|z| = 1$ then $|P(z)| \leq M|z|^n$ for all $|z| \geq 1$.

The Cauchy inequalities imply the interesting

**Theorem 2.31** *(Liouville[12]) If the function $f$ is holomorphic in the whole complex plane and bounded then it is equal identically to a constant.*

**Proof.** According to Theorem 2.28 the function $f$ may be represented by a Taylor series

$$f(z) = \sum_{n=0}^{\infty} c_n z^n$$

in any closed disk $\bar{U} = \{|z| \leq R\}$, $R < \infty$ with the coefficients that do not depend on $R$. Since $f$ is bounded in $\mathbb{C}$, say, $|f(z)| \leq M$ then the Cauchy inequalities imply that for any $n = 0, 1, \ldots$ we have $|c_n| \leq M/R^n$. We may take $R$ to be arbitrarily large and hence the right side tends to zero as $R \to +\infty$ while the left side is independent of $R$. Therefore $c_n = 0$ for $n \geq 1$ and hence $f(z) = c_0$ for all $z \in \mathbb{C}$. $\square$

Therefore the two properties of a function – to be holomorphic and bounded are realized simultaneously only for the trivial functions that are equal identically to a constant.

**Exercise 2.32** Prove the following properties of functions $f$ holomorphic in the whole plane $\mathbb{C}$:

(1) Let $M(r) = \sup\limits_{|z|=r} |f(z)|$, then if $M(r) = Ar^N + B$ where $r$ is an arbitrary positive real number and $A$, $B$ and $N$ are constants, then $f$ is a polynomial of degree not higher than $N$.

(2) If all values of $f$ belong to the right half-plane then $f = \text{const}$.

(3) If $\lim_{z\to\infty} f(z) = \infty$ then the set $\{z \in \mathbb{C} : \; f(z) = 0\}$ is not empty.

The Liouville theorem may be reformulated:

**Theorem 2.33** *If a function $f$ is holomorphic in the closed complex plane $\overline{\mathbb{C}}$ then it is equal identically to a constant.*

**Proof.** if the function $f$ is holomorphic at infinity the limit $\lim_{z\to\infty} f(z)$ exists and is finite. Therefore $f$ is bounded in a neighborhood $U = \{|z| > R\}$ of this point. However, $f$ is also bounded in the complement $\bar{U}^c = \{|z| \leq R\}$ since it is continuous there and the set $\bar{U}^c$ is compact. Therefore $f$ is holomorphic and bounded in $\mathbb{C}$ and thus Theorem 2.31 implies that is equal to a constant. $\square$

**Exercise 2.34** Show that a function $f(z)$ that is holomorphic at $z = 0$ and satisfies $f(z) = f(2z)$, is equal identically to a constant.

[12] Actually this theorem was proved by Cauchy in 1844 while Liouville had proved only a partial result in the same year. The wrong attribution was started by a student of Liouville who had learned the theorem at one of his lectures.

Theorem 2.34 claims that any function holomorphic in a disk may be represented as a sum of a convergent power series inside this disk. We would like to show now that, conversely, the sum of a convergent power series is a holomorphic function. Let us first recall some properties of power series that are familiar from the real analysis.

**Lemma 2.35** *If the terms of a power series*

$$\sum_{n=0}^{\infty} c_n (z-a)^n \tag{2.38}$$

*are bounded at some point $z_0 \in \mathbb{C}$, that is,*

$$|c_n (z_0 - a)^n| \leq M, \quad (n = 0, 1, \dots), \tag{2.39}$$

*then the series converges in the disk $U = \{z : \ |z-a| < |z_0 - a|\}$. Moreover, it converges absolutely and uniformly on any set $K$ that is properly contained in $U$.*

**Proof.** As in real analysis.

**Theorem 2.36** *(Abel) Let the power series (2.38) converge at a point $z_0 \in \mathbb{C}$. Then this series converges in the disk $U = \{z : \ |z-a| < |z_0 - a|\}$ and, moreover, it converges uniformly and absolutely on every compact subset of $U$.*

**Proof.** Follows immediately from lemma.

**The Cauchy-Hadamard formula.** *Let the coefficients of the power series (2.38) satisfy*

$$\limsup_{n\to\infty} |c_n|^{1/n} = \frac{1}{R}, \tag{2.40}$$

*with $0 \leq R \leq \infty$ (we set $1/0 = \infty$ and $1/\infty = 0$). Then the series (2.38) converges at all $z$ such that $|z-a| < R$ and diverges at all $z$ such that $|z-a| > R$.*

**Proof.** As in real analysis.

**Definition 2.37** *The domain of convergence of a power series (2.38) is the interior of the set $E$ of the points $z \in \mathbb{C}$ where the series converges.*

**Theorem 2.38** *The domain of convergence of the power series (2.38) is the open disk $\{|z-a| < R\}$, where $R$ is the number determined by the Cauchy-Hadamard formula.*

**Proof.** The previous proposition shows that the set $E$ where the series (2.38) converges consists of the disk $U = \{|z-a| < R\}$ and possibly some other set of points on the boundary $\{|z-a| = R\}$ of $U$. Therefore the interior of $E$ is the open disk $\{|z-a| < R\}$. □

The open disk in Theorem 2.38 is called the disk of convergence of the power series (2.38), and the number $R$ is its radius of convergence. We pass now to the proof that the sum of a power series is holomorphic.

**Theorem 2.39** *The sum of a power series*

$$f(z) = \sum_{n=0}^{\infty} c_n (z-a)^n \tag{2.41}$$

*is holomorphic in its domain of convergence.*

**Proof.** We assume that the radius of convergence $R > 0$, otherwise there is nothing to prove. Let us define the formal series of derivatives

$$\sum_{n=1}^{\infty} n c_n (z-a)^{n-1} = \phi(z). \tag{2.42}$$

Its convergence is equivalent to that of the series $\sum_{n=1}^{\infty} n c_n (z-a)^n$. However, since $\limsup_{n\to\infty} |nc_n|^{1/n} = \limsup_{n\to\infty} |c_n|^{1/n}$ the radius of convergence of the series (2.42) is also equal to $R$. Therefore this series converges uniformly on compact subsets of the disk $U = \{|z-a| < R\}$ and hence the function $\phi(z)$ is continuous in this disk.

Moreover, for the same reason the series (2.42) may be integrated term-wise along the boundary of any triangle $\Delta$ that is properly contained in $U$:

$$\int_{\partial\Delta} \phi dz = \sum_{n=1}^{\infty} n c_n \int_{\partial\Delta} (z-a)^{n-1} dz = 0.$$

The integrals on the right side vanish by the Cauchy theorem. Therefore we may apply Theorem 2.9 and Remark 2.10 which imply that the function

$$\int_{[a,z]} \phi(\zeta) d\zeta = \sum_{n=1}^{\infty} n c_n \int_{[a,z]} (\zeta - a)^{n-1} d\zeta = \sum_{n=1}^{\infty} c_n (z-a)^n$$

has a derivative at all $z \in U$ that is equal to $\phi(z)$. Once again we used uniform convergence to justify the term-wise integration above. However, then the function

$$f(z) = c_0 + \int_{[a,z]} \phi(\zeta) d\zeta$$

has a derivative at all $z \in U$ that is also equal to $\phi(z)$. $\Box$

### 2.5.1 Properties of holomorphic functions

We discuss some corollaries of Theorem 2.39.

**Theorem 2.40** *Derivative of a function $f \in \mathcal{O}(D)$ is holomorphic in the domain $D$.*

**Proof.** Given a point $z_0 \in D$ we construct a disk $U = \{|z - z_0| < R\}$ that is contained in $D$. Theorem 2.28 implies that $f$ may be represented as a sum of a converging power series in this disk. Theorem 2.39 implies that its derivative $f' = \phi$ may also be represented as a sum of a power series converging in the same disk. Therefore one may apply Theorem 2.39 also to the function $\phi$ and hence $\phi$ is holomorphic in the disk $U$. $\Box$

This theorem also implies directly the necessary condition for the existence of anti-derivative that we have mentioned in Section 2.1.1:

**Corollary 2.41** *If a continuous function $f$ has an anti-derivative $F$ in a domain $D$ then $f$ is holomorphic in $D$.*

Using Theorem 2.40 once again we obtain

**Theorem 2.42** *Any function $f \in \mathcal{O}(D)$ has derivatives of all orders in $D$ that are also holomorphic in $D$.*

The next theorem establishes uniqueness of the power series representation of a function relative to a given point.

**Theorem 2.43** *Let a function $f$ have a representation*

$$f(z) = \sum_{n=0}^{\infty} c_n (z - z_0)^n \tag{2.43}$$

*in a disk $\{|z - z_0| < R\}$. Then the coefficients $c_n$ are determined uniquely as*

$$c_n = \frac{f^{(n)}(z_0)}{n!}, \quad n = 0, 1, \dots \tag{2.44}$$

**Proof.** Inserting $z = z_0$ in (2.43) we find $c_0 = f(z_0)$. Differentiating (2.43) termwise we obtain

$$f'(z) = c_1 + 2c_2(z - z_0) + 3c_3(z - z_0)^2 + \dots$$

Inserting $z = z_0$ above we obtain $c_1 = f'(z_0)$. Differentiating (2.43) $n$ times we obtain (we do not write out the formulas for $\widetilde{c}_j$ below)

$$f^{(n)}(z) = n!c_n + \widetilde{c}_1(z - z_0) + \widetilde{c}_1(z - z_0)^2 + \dots$$

and once again using $z = z_0$ we obtain $c_n = f^{(n)}(z_0)/n!$. $\Box$

Sometimes Theorem 2.43 is formulated as follows: "Every converging power series is the Taylor series for its sum."

Comparing expressions (2.44) for the coefficients $c_n$ with their values given by (2.36) we obtain the formulas for the derivatives of holomorphic functions:

$$f^{(n)}(z_0) = \frac{n!}{2\pi i} \int_{\gamma_r} \frac{f(\zeta) d\zeta}{(\zeta - z_0)^{n+1}}, \quad n = 1, 2 \dots \tag{2.45}$$

If the function $f$ is holomorphic in a domain $D$ and $G$ is a sub-domain of $D$ that is bounded by finitely many continuous curves and such that $z_0 \in G$ then we may replace the contour $\gamma_r$ in (2.45) by the oriented boundary $\partial G$, using the invariance of the integral under homotopy of paths. Then we obtain *the Cauchy integral formula for derivatives* of holomorphic functions:

$$f^{(n)}(z) = \frac{n!}{2\pi i} \int_{\partial G} \frac{f(\zeta) d\zeta}{(\zeta - z)^{n+1}}, \quad n = 1, 2 \dots \tag{2.46}$$

These formulas may be also obtained from the Cauchy integral formula

$$f(z) = \frac{1}{2\pi i}\int_{\partial G}\frac{f(\zeta)d\zeta}{(\zeta - z)},$$

by differentiating with respect to the parameter $z$ inside the integral. Our indirect argument allowed us to bypass the justification of this operation.

**Theorem 2.44** *(Morera[13]) If a function $f$ is continuous in a domain $D$ and its integral over the boundary $\partial\Delta$ of any triangle $\Delta$ vanishes then $f$ is holomorphic in $D$.*

**Proof.** Given $a \in D$ we construct a disk $U = \{|z - a| < r\} \subset D$. The function $F(z) = \int_{[a,z]} f(\zeta)d\zeta$ is holomorphic in $U$ (see remark after Theorem 2.9). Theorem 2.40 implies then that $f$ is also holomorphic in $D$. This proves that $f$ is holomorphic at all $a \in D$. $\square$

**Remark 2.45** The Morera Theorem states the converse to the Cauchy theorem as formulated in Theorem 2.8, that is, that integral of a holomorphic function over the boundary of any triangle vanishes. However, the Morera theorem also requires that $f$ is continuous in $D$. This assumption is essential: for instance, the function $f$ that is equal to zero everywhere in $\mathbb{C}$ except at $z = 0$, where it is equal to one, is not even continuous at $z = 0$ but its integral over any triangle vanishes.

However, the Morera theorem does not require any differentiability of $f$: from the modern point of view we may say that a function satisfying the assumptions of this theorem is a generalized solution of the Cauchy-Riemann equations. The theorem asserts that any generalized solution is a classical solution, that is, it has partial derivatives that satisfy the Cauchy-Riemann equations.

**Remark 2.46** We have seen that the representation as a power series in a disk $\{|z-a| < R\}$ is a necessary and sufficient condition for $f$ to be holomorphic in this disk. However, convergence of the power series on the boundary of the disk is not related to it being holomorphic at those points. This may be sen on simplest examples. Indeed, the geometric series

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n \tag{2.47}$$

converges in the open disk $\{|z| < 1\}$. The series (2.47) diverges at all points on $\{|z| = 1\}$ since its $n$-th term does not vanish in the limit $n \to \infty$. On the other hand, the series

$$f(z) = \sum_{n=0}^{\infty} \frac{z^n}{n^2} \tag{2.48}$$

converges at all points of $\{|z| = 1\}$ since it is majorized by the convergent number series $\sum_{n=1}^{\infty} \frac{1}{n^2}$. However, its sum may not be holomorphic at $z = 1$ since its derivative $f'(z) = \sum_{n=1}^{\infty} \frac{z^{n-1}}{n}$ is unbounded as $z$ tends to one along the real axis.

[13]The theorem was proved by an Italian mathematician Giacinto Morera in 1889.

## 2.6 The Uniqueness theorem

**Definition 2.47** *A zero of the function $f$ is a point $a \in \overline{\mathbb{C}}$ where $f$ vanishes, that is, solution of $f(z) = 0$.*

Zeroes of differentiable functions in the real analysis may have limit points where the function $f$ remains differentiable, for example, $f(x) = x^2 \sin(1/x)$ behaves in this manner at $x = 0$. The situation is different in the complex analysis: zeroes of a holomorphic function must be isolated, they may have limit points only on the boundary of the domain where the function is holomorphic.

**Theorem 2.48** *Let the point $a \in \mathbb{C}$ be a zero of the function $f$ that is holomorphic at this point, and $f$ is not equal identically to zero in a neighborhood of $a$. Then there exists a number $n \in \mathbb{N}$ so that*

$$f(z) = (z - a)^n \phi(z), \tag{2.49}$$

*where the function $\phi$ is holomorphic at $a$ and is different from zero in a neighborhood of $a$.*

**Proof.** Indeed, $f$ may be represented by a power series in a neighborhood of $a$: $f(z) = \sum_{n=0}^{\infty} c_n (z - a)^n$. The first coefficient $c_0 = 0$ but not all $c_n$ are zero, otherwise $f$ would vanish identically in a neighborhood of $a$. Therefore there exists the smallest $n$ so that $c_n \neq 0$ and the power series has the form

$$f(z) = c_n (z - a)^n + c_{n+1} (z - a)^{n+1} + \ldots, \quad c_n \neq 0. \tag{2.50}$$

Let us denote by

$$\phi(z) = c_n + c_{n+1}(z - a) + \ldots \tag{2.51}$$

so that $f(z) = (z - a)^n \phi(z)$. The series (2.51) converges in a neighborhood of $a$ (it has the same radius of convergence as $f$) and thus $\phi$ is holomorphic in this neighborhood. Moreover, since $\phi(a) = c_n \neq 0$ and $\phi$ is continuous at $a$, $\phi(z) \neq 0$ in a neighborhood of $a$. $\Box$

**Theorem 2.49** *(Uniqueness) Let $f_1, f_2 \in \mathcal{O}(D)$, then if $f_1 = f_2$ on a set $E$ that has a limit point in $D$ then $f_1(z) = f_2(z)$ for all $z \in D$.*

**Proof.** The function $f = f_1 - f_2$ is holomorphic in $D$. We should prove that $f \equiv 0$ in $D$, that is, that the set $F = \{z \in D : \ f(z) = 0\}$, that contains in particular the set $E$, coincides with $D$. The limit point $a$ of $E$ belongs to $E$ (and hence to $F$) since $f$ is continuous. Theorem 2.50 implies that $f \equiv 0$ in a neighborhood of $a$, otherwise it would be impossible for $a$ to e a limit point of the set of zeroes of $f$.

Therefore the interior $F^o$ of $F$ is not empty - it contains $a$. Moreover, $F^o$ is an open set as the interior of a set. However, it is also closed in the relative topology of $D$. Indeed, let $b \in D$ be a limit point of $F^o$, then the same Theorem 2.48 implies that $f \equiv 0$ in a neighborhood of $b$ so that $b \in F^o$. Finally, the set $D$ being a domain is connected, and hence $F^o = D$ by Theorem 1.29 of Chapter 1. $\Box$

This theorem shows another important difference of a holomorphic function from a real differentiable function in the sense of real analysis. Indeed, even two infinitely differentiable functions may coincide on an open set without being identically equal to each other everywhere else. However, according to the previous theorem two holomorphic functions that coincide on a set that has a limit point in the domain where they are holomorphic (for instance on a small disk, or an arc inside the domain) have to be equal identically in the whole domain.

**Exercise 2.50** Show that if $f$ is holomorphic at $z = 0$ then there exists $n \in \mathbb{N}$ so that $f(1/n) \neq (-1)^n/n^3$.

We note that one may simplify the formulation of Theorem 2.48 using the Uniqueness theorem. That is, the assumption that $f$ is not equal identically to zero in any neighborhood of the point $a$ may be replaced by the assumption that $f$ is not equal identically to zero everywhere (these two assumptions coincide by the Uniqueness theorem).

Theorem 2.48 shows that holomorphic functions vanish as an integer power of $(z-a)$.

**Definition 2.51** *The order, or multiplicity, of a zero $a \in \mathbb{C}$ of a function $f$ holomorphic at this point, is the order of the first non-zero derivative $f^{(k)}(a)$. in other words, a point $a$ is a zero of $f$ of order $n$ if*

$$f(a) = \cdots = f^{(k-1)}(a) = 0, \quad f^{(n)}(a) \neq 0, \quad n \geq 1. \tag{2.52}$$

Expressions $c_k = f^{(k)}(a)/k!$ for the coefficients of the Taylor series show that the order of zero is the index of the first non-zero Taylor coefficient of the function $f$ at the point $a$, or, alternatively, the number $n$ in Theorem 2.48. The Uniqueness theorem shows that holomorphic functions that are not equal identically to zero may not have zeroes of infinite order.

Similar to what is done for polynomials, one may define the order of zeroes using division.

**Theorem 2.52** *The order of zero $a \in \mathbb{C}$ of a holomorphic function $f$ coincides with the order of the highest degree $(z-a)^k$ that is a divisor of $f$ in the sense that the ratio $\dfrac{f(z)}{(z-a)^k}$ (extended by continuity to $z = a$) is a holomorphic function at $a$.*

**Proof.** Let us denote by $n$ the order of zero $a$ and by $N$ the highest degree of $(z-a)$ that is a divisor of $f$. Expression (2.49) shows that $f$ is divisible by any power $k \leq n$:

$$\frac{f(z)}{(z-a)^k} = (z-a)^{n-k}\phi(z),$$

and thus $N \geq n$. Let $f$ be divisible by $(z-a)^N$ so that the ratio

$$\psi(z) = \frac{f(z)}{(z-a)^N}$$

is a holomorphic function at $a$. Developing $\psi$ as a power series in $(z-a)$ we find that the Taylor expansion of $f$ at $a$ starts with a power not smaller than $N$. Therefore $n \geq N$ and since we have already shown that $n \leq N$ we conclude that $n = N$. □

**Example 2.53** The function $f(z) = \sin z - z$ has a third order zero at $z = 0$. Indeed, we have $f(0) = f'(0) = f''(0)$ but $f'''(0) \neq 0$. This may also be seen from the representation

$$f(z) = -\frac{z^3}{3!} + \frac{z^5}{5!} + \ldots$$

**Remark 2.54** Let $f$ be holomorphic at infinity and equal to zero there. It is natural to define the order of zero at this point as the order of zero the order of zero at $z = 0$ of the function $\phi(z) = f(1/z)$. The theorem we just proved remains true also for $a = \infty$ if instead of dividing by $(z-a)^k$ we consider multiplication by $z^k$. For example, the function $f(z) = \dfrac{1}{z^3} + \dfrac{1}{z^2}$ has order 3 at infinity.

## 2.7 The Weierstrass theorem

Recall that termwise differentiation of a series in real analysis requires uniform convergence of the series in a neighborhood of a point as well as uniform convergence of the series of derivatives. The situation is simplified in the complex analysis. The following theorem holds.

**Theorem 2.55** *(Weierstrass) If the series*

$$f(z) = \sum_{n=0}^{\infty} f_n(z) \tag{2.53}$$

*of functions holomorphic in a domain $D$ converges uniformly on any compact subset of this domain then*
*(i) the sum of this series is holomorphic in $D$;*
*(ii) the series may be differentiated termwise arbitrarily many times at any point in $D$.*

**Proof.** Let $a$ be arbitrary point in $D$ and consider the disk $U = \{|z-a| < r\}$ that is properly contained in $D$. The series (2.53) converges uniformly in $U$ by assumption and thus its sum is continuous in $U$. Let $\Delta \subset U$ be a triangle contained in $U$ and let $\gamma = \partial\Delta$. Since the series (2.53) converges uniformly in $U$ we may integrate it termwise along $\gamma$:

$$\int_\gamma f(z)dz = \sum_{n=0}^{\infty} \int_\gamma f_n(z)dz.$$

However, the Cauchy theorem implies that all integrals on the right side vanish since the functions $f_n$ are holomorphic. Hence the Morera theorem implies that the function $f$ is holomorphic and part (i) is proved.

In order to prove part (ii) we once again take an arbitrary point $a \in D$, consider the same disk $U$ as in the proof of part (i) and denote by $\gamma_r = \partial U = \{|z-a| = r\}$. The Cauchy formulas for derivatives imply that

$$f^{(k)}(a) = \frac{k!}{2\pi i} \int_{\gamma_r} \frac{f(\zeta)}{(\zeta - a)^{k+1}} d\zeta. \tag{2.54}$$

The series

$$\frac{f(\zeta)}{(\zeta - a)^{k+1}} = \sum_{n=0}^{\infty} \frac{f_n(\zeta)}{(\zeta - a)^{k+1}} \tag{2.55}$$

differs from (2.53) by a factor that has constant absolute value $\frac{1}{r^{k+1}}$ for all $\zeta \in \gamma_r$. Therefore it converges uniformly on $\gamma_r$ and may be integrated termwise in (2.54). Using expressions (2.54) in (2.55) we obtain

$$f^{(k)}(a) = \frac{k!}{2\pi i} \sum_{n=0}^{\infty} \int_{\gamma_r} \frac{f_n(\zeta)}{(\zeta - a)^{k+1}} d\zeta = \sum_{n=0}^{\infty} f_n^{(k)}(a),$$

and part (ii) is proved. $\Box$

**Exercise 2.56** Explain why the series $\sum_{n=1}^{\infty} \frac{\sin(n^3 z)}{n^2}$ may not be differentiated termwise.

# 3 Properties of Harmonic functions

We will now discuss some basic properties of the harmonic functions. Most of them hold not only in two but only in higher dimensions but some are specific to two dimensions. We have already seen the man value principle: if $u(z)$ is harmonic in a disk $|z - z_0| < R$, then

$$u(z_0) = \frac{1}{2\pi} \int_0^{2\pi} u(z_0 + \rho e^{i\phi}) d\phi.$$

Here is the generalization to harmonic functions in higher dimensions.

**Theorem 3.1** *Let $U \subset \mathbb{R}^n$ be an open set and let $B(x, r)$ be a ball centered at $x \in \mathbb{R}^n$ of radius $r > 0$ contained in $U$. Assume that the function $u(x)$ satisfies $\Delta u = 0$ for all $x \in U$ and that $u \in C^2(U)$. Then we have*

$$u(x) = \frac{1}{|B(x,r)|} \int_{B(x,r)} u dy = \frac{1}{|\partial B(x,r)|} \int_{\partial B(x,r)} u dS. \tag{3.1}$$

**Proof.** Let us fix the point $x \in U$ and define

$$\phi(r) = \frac{1}{|\partial B(x,r)|} \int_{\partial B(x,r)} u(z) dS(z). \tag{3.2}$$

It is easy to see that, since $u(x)$ is continuous, we have

$$\lim_{r \downarrow 0} \phi(r) = u(x). \tag{3.3}$$

Therefore, we would be done if we knew that $\phi'(r) = 0$ for all $r > 0$ (and such that the ball $B(x, r)$ is contained in $U$). To this end, using the polar coordinates $z = x + ry$, with $y \in \partial B(0, 1)$, we may rewrite (3.2) as

$$\phi(r) = \frac{1}{|\partial B(0,1)|} \int_{\partial B(0,1)} u(x + ry) dS(y).$$

Then differentiating in $r$ gives

$$\phi'(r) = \frac{1}{|\partial B(0,1)|} \int_{\partial B(0,1)} y \cdot \nabla u(x+ry) dS(y).$$

Going back to the $z$-variables gives

$$\phi'(r) = \frac{1}{|\partial B(x,r)|} \int_{\partial B(x,r)} \frac{1}{r}(z-x) \cdot \nabla u(z) dS(z) = \frac{1}{|\partial B(x,r)|} \int_{\partial B(x,r)} \frac{\partial u}{\partial \nu} dS(z).$$

Here we used the fact that the outward normal to $B(x,r)$ at a point $z \in \partial B(x,r)$ is simply $\nu = (z-x)/r$. Using the Green's formula

$$\int_U f \Delta g dy = \int_{\partial U} f \frac{\partial g}{\partial \nu} dS - \int_U \nabla f \cdot \nabla g dy,$$

with $f = 1$ and $g = u$ gives now

$$\phi'(r) = \frac{1}{|\partial B(x,r)|} \int_{B(x,r)} \Delta u(y) dy = 0.$$

It follows that $\phi(r)$ is a constant and then (3.3) implies that

$$u(x) = \frac{1}{|\partial B(x,r)|} \int_{\partial B(x,r)} u dS, \tag{3.4}$$

which is the second identity in (3.1).

In order to prove the first equality in (3.1) we use the polar coordinates once again:

$$\frac{1}{|B(x,r)|} \int_{B(x,r)} u dy = \frac{1}{|B(x,r)|} \int_0^r \left( \int_{\partial B(x,s)} u dS \right) ds = \frac{1}{|B(x,r)|} \int_0^r u(x) n\alpha(n) s^{n-1} ds$$
$$= u(x) \frac{n\alpha(n) r^n}{\alpha(n) r^n} = u(x).$$

In the second equality above we used two facts: first, the already proved identity (3.4) about averages on spherical shells, and, second, that the area of an $(n-1)$-dimensional unit sphere is $n\alpha(n)$. Now, the proof of (3.1) is complete. $\Box$

### The maximum principle

The first consequence of the mean value property is the maximum principle that says that a harmonic function attains its maximum over any domain on the boundary and not inside the domain. Once again, in one dimension this is obvious: a linear function does not have any local extremal points.

**Theorem 3.2** (**The maximum principle**) *Let $u(x)$ be a harmonic function in a connected domain $U$ and assume that $u \in C^2(U) \cap C(\bar{U})$. Then*

$$\max_{x \in U} u(x) = \max_{y \in \partial U} u(y). \tag{3.5}$$

*Moreover, if $u(x)$ achieves its maximum at a point $x_0$ in the interior of $U$ then $u(x)$ is identically equal to a constant in $U$.*

**Proof.** Suppose that $u(x)$ attains its maximum at an interior point $x_0 \in U$, and set $M = u(x_0)$. Then for any $r > 0$ sufficiently small (so that the ball $B(x_0, r)$ is contained in $U$) we have

$$M = u(x) = \frac{1}{|B(x_0, r)|} \int_{B(x_0,r)} u dy \leq M,$$

with the equality above holding only if $u(y) = M$ for all $y$ in the ball $B(x_0, r)$. Therefore, the set $S$ of points where $u(x) = M$ is open. Since $u(x)$ is continuous, this set is also closed. Since $S$ us both open and closed in $U$, and $U$ is connected, it follows that $S = U$, hence $u(x) = M$ at all points $x \in U$. □

Of course, if we replace $u$ by $(-u)$ (which is equally harmonic), we get the minimum principle for $u$.

**Corollary 3.3** (**Strict positivity**) *Assume that $U$ is a connected domain, and $u$ solves*

$$\begin{aligned} &\Delta u = 0 \quad \text{in } U \\ &u = g \quad \text{on } \partial U. \end{aligned} \tag{3.6}$$

*Assume, in addition, that $g \geq 0$, $g$ is continuous on $\partial U$, and $g(x) \not\equiv 0$. Then $u(x) > 0$ at all $x \in U$.*

**Proof.** This is an immediate consequence of the minimum principle: $\min_{x \in \bar{U}} u(x) \geq 0$, and $u$ can not attain its minimum inside $U$, thus $u(x) > 0$ for all $x \in U$. □

### Maximum principle for subharmonic functions

We say that a function $u(x)$ is subharmonic in a domain $U$ if

$$\Delta u(x) \geq 0 \text{ for all } x \in U. \tag{3.7}$$

Inspecting the proof of the mean-value property we see that if $u(x)$ is sub-harmonic in a ball $B(x_0, R)$ and we set

$$\phi(r) = \frac{1}{|\partial B(x_0, r)|} \int_{\partial B(x_0,r)} u dS,$$

then

$$\phi'(r) = \frac{1}{|B(x_0, r)|} \int_{\partial B(x_0,r)} \Delta u dS \geq 0.$$

This means that

$$u(x_0) \leq \frac{1}{|\partial B(x_0, r)|} \int_{\partial B(x_0,r)} u dS,$$

for any $0 < r < R$. It follows that sub-harmonic functions can not attain a maximum inside a domain $U$ unless they are equal to a constant – the proof is identical to that for harmonic functions.

A typical example of a sub-harmonic function comes about as follows. Assume that $u(x)$ is harmonic: $\Delta u = 0$ and $\Phi(s)$ is a convex function. Set $v(x) = \Phi(u(x))$, then:

$$\Delta v = \nabla \cdot (\Phi'(u)\nabla u) = \Phi'(u)\Delta u + \Phi''(u)|\nabla u|^2 = \Phi''(u)|\nabla u|^2 \geq 0,$$

so that the function $v(x)$ is sub-harmonic. In particular, if $u(x)$ is harmonic then $u^2(x)$ is sub-harmonic. This fact has an important implication in complex analysis.

**Theorem 3.4** *(Maximum modulus principle) Let $f(z)$ be holomorphic in the closure of a bounded domain $D$. Then $|f(z)|$ attains its maximum on the boundary $\partial D$.*

**Proof.** Let $f(z) = u + iv(z)$, then the functions $u$ and $v$ are harmonic, whence $|f(z)|^2 = u^2 + v^2$ is subharmonic. It follows that $|f(z)|$ attains its maximum on the boundary $\partial D$. □

### The three lines theorem

A key ingredient in the proof of the Riesz interpolation theorem that we will soon encounter is the following basic result that is related to the maximum modulus principle.

**Theorem 3.5** *Let $F(z)$ be a bounded analytic function in the strip*

$$S = \{z : \ 0 \leq Rez \leq 1\},$$

*such that $|F(iy)| \leq m_0$, $|F(1+iy)| \leq m_1$, with $m_0, m_1 > 0$ for all $y \in \mathbb{R}$. Then*

$$|F(x+iy)| \leq m_0^{1-x} m_1^x \text{ for all } \ 0 \leq x \leq 1, \ y \in \mathbb{R}. \tag{3.8}$$

**Proof.** It is convenient to set

$$F_1(z) = \frac{F(z)}{m_0^{1-z} m_1^z},$$

so that $|F_1(iy)| \leq 1$, $|F_1(1+iy)| \leq 1$ and $F_1$ is uniformly bounded in $S$. It suffices to show that $|F_1(x+iy)| \leq 1$ for all $(x,y) \in S$ under these assumptions – this will give immediately (3.8). If the strip $S$ were a bounded domain, this would follow immediately from the maximum modulus principle.

Assume first that $F_1(x+iy) \to 0$ as $|y| \to +\infty$, uniformly in $x \in [0,1]$. Then $|F_1(x \pm iM)| \leq 1/2$ for all $y$ with $|y| \geq M$, and $M > 0$ large enough. The maximum modulus principle implies that $|F_1(x+iy)| \leq 1$ for $|y| \leq M$, and, since, $|F_1(x+iy)| \leq 1/2$ for all $y$ with $|y| \geq M$, it follows that $|F_1(x+iy)| \leq 1$ for all $(x,y) \in S$.

In general, set

$$G_n(z) = F_1(z) e^{(z^2-1)/n},$$

then

$$|G_n(iy)| \leq |F_1(iy)| e^{(-y^2-1)/n} \leq 1,$$

and

$$|G_n(1+iy)| \leq |F_1(1+iy)| e^{-y^2} \leq 1,$$

but in addition, $G_n$ goes to zero as $|y| \to +\infty$, uniformly in $x \in [0,1]$:

$$|G_n(x+iy)| \le |F_1(z)|e^{(x^2-y^2-1)/n} \le C_0 e^{-y^2/n},$$

with a constant $C_0$ such that $|F_1(z)| \le C_0$ for all $z \in S$. It follows from the previous part of the proof that $|G_n(z)| \le 1$, hence

$$|F_1(z)| \le e^{(1+y^2)/n},$$

for all $z \in S$ and all $n \in \mathbb{N}$. Letting $n \to +\infty$ we deduce that $|F_1(z)| \le 1$ for all $z \in S$. $\square$

**Regularity of harmonic functions**

Now, we prove that if $u(x)$ is a twice continuously differentiable harmonic function then it is infinitely differentiable – in two dimensions this is simply the reflection of the fact that a harmonic function is the real part of a holomorphic function.

**Theorem 3.6** (**Regularity**) *Let $u \in C^2(U)$ be a harmonic function in a domain $U$. Then $u$ is infinitely differentiable in $U$.*

**Proof.** The proof is via a miracle: we first define a "smoothed" version of $u$, and then verify that the "smoothed" version coincides with the original, hence original is also infinitely smooth. This is as close to a free lunch as it gets.

Consider a radial non-negative function $\eta(x) \ge 0$ that depends only on $|x|$ such that (i) $\eta(x) = 0$ for $|x| \ge 1$, (ii) $\eta(x)$ is infinitely differentiable, and (iii) $\int_{\mathbb{R}^n} \eta(x)dx = 1$. Also, for each $\varepsilon \in (0,1)$ define its rescaled version

$$\eta_\varepsilon(x) = \frac{1}{\varepsilon^n}\eta\left(\frac{x}{\varepsilon}\right).$$

It is straightforward to verify that $\eta_\varepsilon$ satisfies the same properties (i)-(iii) above. Moreover, the function

$$u_\varepsilon(x) = \int_{\mathbb{R}^n} \eta_\varepsilon(x-y)u(y)dy \tag{3.9}$$

is infinitely differentiable in the slightly smaller domain $U_\varepsilon = \{x \in: \ \mathrm{dist}(x, \partial U) > \varepsilon\}$. The reason is that we can differentiate infinitely many times under the integral sign in (3.9) – this follows from the standard multivariable calculus theorem on differentiation of integrals depending on a parameter (the variable $x$ plays the role of a parameter here). Our main claim is that, because of the mean value property, we have

$$u_\varepsilon(x) = u(x) \text{ for all } x \in U_\varepsilon. \tag{3.10}$$

This will immediately imply that $u(x)$ is infinitely differentiable in the domain $U_\varepsilon$. And, as any point $x$ from $U$ lies in $U_\varepsilon$ if $\varepsilon < \mathrm{dist}(x, \partial U)$, it follows that $u(x)$ is infinitely differentiable at all points $x \in U$.

Let us now verify (3.10):

$$u_\varepsilon(x) = \int_{\mathbb{R}^n} \eta_\varepsilon(x-y)u(y)dy = \frac{1}{\varepsilon^n}\int_U \eta\left(\frac{|x-y|}{\varepsilon}\right)u(y)dy = \frac{1}{\varepsilon^n}\int_{B(x,\varepsilon)} \eta\left(\frac{|x-y|}{\varepsilon}\right)u(y)dy.$$

The last equality holds because $\eta(z) = 0$ if $|z| \geq 1$, whence $\eta_\varepsilon(z) = 0$ if $|z| \geq \varepsilon$. Changing variables $y = x + \varepsilon z$ gives

$$u_\varepsilon(x) = \int_{B(0,1)} \eta(z)\, u(x + \varepsilon z)dz.$$

Going to the polar coordinates leads to

$$u_\varepsilon(x) = \int_0^1 \eta(r)\left[\int_{\partial B(0,1)} u(x + \varepsilon r\omega)dS(\omega)\right] r^{n-1}dr. \tag{3.11}$$

The mean value property implies that

$$\int_{\partial B(0,1)} u(x + \varepsilon r\omega)dS(\omega) = u(x)|\partial B(0,1)|.$$

Using this in (3.11), we obtain

$$u_\varepsilon(x) = u(x)\int_0^1 \eta(r)\,|\partial B(0,1)|r^{n-1}dr = u(x)\int_{B(0,1)} \eta(y)dy = u(x), \tag{3.12}$$

which is (3.10). We used the fact that $\eta$ has integral equal to one in the last step. □

This regularity property is quite fundamental and appears in one way or other for the class of elliptic equations (and not just for the Laplace equation) we will discuss later. One of their main qualitative properties is that solutions are more regular than the data prescribed, and they behave much nicer than, say, solutions of wave equations and other hyperbolic problems.

Let us now give a more quantitative estimate on how large the derivatives of the harmonic functions can be (in two dimensions these follow from the Cauchy integral formula).

**Theorem 3.7** *Let $u(x)$ be a harmonic function in a domain $U$ and let $B(y_0, r)$ be a ball contained in $U$ centered at a point $y_0 \in U$. Then there exist universal constants $C_n$ and $D_n$ that depends only on the dimension $n$ so that we have*

$$|u(y_0)| \leq \frac{C_n}{r^n}\int_{B(y_0,r)} |u(y)|dy. \tag{3.13}$$

*and*

$$|\nabla u(y_0)| \leq \frac{D_n}{r^{n+1}}\int_{B(y_0,r)} |u(y)|dy. \tag{3.14}$$

The remarkable fact about the estimate (3.14) is that we are able to estimate the size of the derivatives of a harmonic function in terms of its values – this means that a harmonic function can not oscillate (oscillation means, essentially, that the function is much smaller than its derivative).

**Proof.** First, the estimate (3.13) follows immediately from the first equality in the mean value formula (3.1). In order to obtain the derivative bound (3.14) note that if $u(x)$ is harmonic then so are the partial derivatives $\partial u/\partial x_j$, whence

$$\left|\frac{\partial u(y_0)}{\partial x_j}\right| \le \frac{1}{|B(y_0,r/2)|}\left|\int_{B(y_0,r/2)}\frac{\partial u(y)}{\partial x_j}dy\right| = \frac{1}{|B(y_0,r/2)|}\left|\int_{\partial B(y_0,r/2)}u(y)n_j(y)dy\right|, \tag{3.15}$$

where $n_j(y)$ is the $j$-th component of the outward normal. Continuing this estimate we see that (we use the fact that the area of the unit sphere is $n\alpha(n)$)

$$\left|\frac{\partial u(y_0)}{\partial x_j}\right| \le \frac{2^n}{\alpha(n)r^n}\frac{n\alpha(n)r^{n-1}}{2^{n-1}}\sup_{z\in B(y_0,r/2)}|u(z)| = \frac{2n}{r}\sup_{z\in B(y_0,r/2)}|u(z)|. \tag{3.16}$$

Now, we can use the estimate (3.13) applied at any point $z\in B(y_0,r/2)$:

$$|u(z)| \le \frac{C_n}{(r/2)^n}\int_{B(z,r/2)}|u(z')|dz'. \tag{3.17}$$

However, since $|y_0-z|\le r/2$ (this is why we took a smaller ball in (3.15)!), any such ball $B(z,r/2)$ is contained inside the ball $B(y_0,r)$, thus (3.17) implies that

$$|u(z)| \le \frac{C_n}{(r/2)^n}\int_{B(y_0,r)}|u(z')|dz'.$$

Now, it follows from (3.16) that

$$\left|\frac{\partial u(y_0)}{\partial x_j}\right| \le \frac{2n}{r}\frac{C_n}{(r/2)^n}\int_{B(y_0,r)}|u(z')|dz' = \frac{D_n}{r^{n+1}}\int_{B(y_0,r)}|u(z)|dz, \tag{3.18}$$

which is (3.14). □

Theorem 3.7 is another expression of the fact that harmonic functions do not oscillate – the first estimate says that the value of the function at a point is bounded by its averages (but we have seen that already in the mean value property), while the second bound says in a quantitative way that derivative at a point can not be large without the function being large around the point. This rules out oscillatory behavior.

### The Liouville theorem

The Liouville theorem says that a function which is harmonic in all of $\mathbb{R}^n$ is either unbounded or is identically equal to a constant.

**Theorem 3.8** *Let $u(x)$ be a harmonic bounded function in $\mathbb{R}^n$. Then $u(x)$ is equal identically to a constant.*

**Proof.** Let us assume that $|u(x)| \leq M$ for all $x \in \mathbb{R}^n$. We fix $x_0 \in \mathbb{R}^n$ and use Theorem 3.7:

$$|\nabla u(x_0)| \leq \frac{C}{r^{n+1}} \int_{B(x_0,r)} |u(y)|dy \leq \frac{C\alpha(n)r^n}{r^{n+1}} M \leq \frac{C\alpha(n)M}{r}.$$

As this is true for any $r > 0$ we may let $r \to \infty$ and conclude that $\nabla u(x_0) = 0$, thus $u(x)$ is equal identically to a constant. $\square$

This theorem is, of course, a direct generalization to higher dimensions of the familiar Liouville theorem in complex analysis.

### Harnack's inequality

Here is another way to express lack of oscillations of **nonnegative** harmonic functions – their maximum cannot be much larger than their minimum. To trivialize, consider the one-dimensional situation. Let $u(x)$ be a non-negative harmonic function on the interval $(0,1)$, that is, $u(x) = ax + b$ with some constants $a, b \in \mathbb{R}$. We claim that if $u(x) \geq 0$ for all $x \in [0,1]$ then

$$\frac{1}{3} \leq \frac{u(x)}{u(y)} \leq 3, \tag{3.19}$$

for all $x, y$ in the *smaller* interval $(1/4, 3/4)$. The constants $1/3$ and $3$ in (3.19) depend on the choice of the "smaller" interval – they would change of we would replace $(1/4, 3/4)$ by another subinterval of $[0,1]$. But once we fix the subinterval, they do not depend on the choice of the harmonic function. Let us now show that (3.19) holds for all $x, y \in (1/4, 3/4)$. Without loss of generality we may assume that $x > y$. First, consider the case $a > 0$. Then, since $u(x)$ is increasing (because $a > 0$), we have

$$1 \leq \frac{u(x)}{u(y)} \leq \frac{u(3/4)}{u(1/4)} = \frac{3a+4b}{a+4b}. \tag{3.20}$$

As $u(x) > 0$ on $[0,1]$ we know that $b > 0$ (and $a > 0$ by assumption), using this in (3.20) gives, with $c = a/b$:

$$1 \leq \frac{u(x)}{u(y)} \leq \frac{3c+4}{c+4} = 3 - \frac{8}{c+4} \leq 3.$$

On the other hand, if $a < 0$ then the function $u$ is decreasing, and

$$1 \geq \frac{u(x)}{u(y)} \geq \frac{u(3/4)}{u(1/4)} = \frac{c+4}{3c+4} = \frac{1}{3} + \frac{8}{3(3c+4)}.$$

As $u(1) > 0$ we know that $a + b > 0$, and we still have $b > 0$ since $u(0) > 0$. Thus, $c > -1$, and therefore,

$$1 \geq \frac{u(x)}{u(y)} \geq \frac{1}{3} + \frac{8}{3(3c+4)} \geq \frac{1}{3}.$$

We conclude that (3.19), indeed, holds. Geometrically, (3.19) expresses a very simple fact: if $u(3/4) \gg u(1/4)$ then the slope of the straight line connecting the points

$(1/4, u(1/4))$ and $(3/4, u(3/4))$ is too large so that it would go below the $x$-axis at $x = 0$ contradicting the assumption that the linear function is positive on the interval $(0, 1)$. On the other hand, if $u(1/4) \gg u(3/4)$ then this line would go below that $x$-axis at $x = 1$. Therefore, the condition that $u(x) > 0$ on the *larger* interval $[0, 1]$ is very important here.

Now, we turn to the general case of dimension larger than one. We say that a set $V$ is strictly contained in $U$ if $V \subset U$ and there exists $\varepsilon_0 > 0$ so that for any $x \in V$ we have $\text{dist}(x, \partial U) \geq \varepsilon_0$.

**Theorem 3.9** (**Harnack's inequality**) *Let $U$ be an open set and let $V$ be a connected compact set strictly contained in $U$. Then there exists a constant $C$ that depends on $U$ and $V$ but nothing else so that for any nonnegative harmonic function $u$ in $U$ we have*

$$\sup_{x \in V} u(x) \leq C \inf_{x \in V} u(x). \tag{3.21}$$

**Proof.** Let $r = (1/4)\text{dist}(V, \partial U)$ and choose two points $x, y \in V$ such that $|x - y| \leq r$. Then $B(x, 2r) \subset U$ so $u$ is harmonic in this ball, and the mean-value principle implies that

$$u(x) = \frac{1}{|B(x, 2r)|} \int_{B(x,2r)} u(z)dz. \tag{3.22}$$

Note also that since $|x - y| \leq r$, the ball $B(y, r)$ is contained inside $B(x, 2r)$, hence (3.22) implies that

$$u(x) \geq \frac{1}{\alpha(n)2^n r^n} \int_{B(y,r)} u(z)dz. \tag{3.23}$$

It follows, on the other hand, from the mean-value principle that

$$u(y) = \frac{1}{\alpha(n)r^n} \int_{B(y,r)} u(z)dz. \tag{3.24}$$

Putting (3.23) and (3.24) together gives

$$u(x) \geq \frac{1}{2^n} u(y). \tag{3.25}$$

Reversing the argument we can similarly conclude that

$$u(y) \geq \frac{1}{2^n} u(x), \tag{3.26}$$

hence

$$\frac{1}{2^n} u(x) \leq u(y) \leq 2^n u(x), \text{ for all } x, y \in V \text{ such that } |x - y| \leq (1/4)\text{dist}(V, \partial U). \tag{3.27}$$

Now, there exists a number $N$ so that we may cover the compact set $\bar{V}$ by $N$ balls of radius $r/2$. Then given any two points $x, y \in V$ we can connect them by a piece-wise straight line curve with no more than $N$ segments, each segment at most $r$ long. It follows that for any $x, y \in V$ we have

$$\frac{1}{2^{Nn}} u(x) \leq u(y) \leq 2^{Nn} u(x), \quad \text{for all } x, y \in V. \tag{3.28}$$

This, of course, implies (3.21) with $C = 2^{nN}$. $\square$

# 4 The Hilbert transform and the Riesz-Thorin interpolation

### The Poisson kernel

Given a Schwartz class function $f(x) \in \mathcal{S}(\mathbb{R}^n)$ define a function

$$u(x,t) = \int_{\mathbb{R}^n} e^{-2\pi t|\xi|}\hat{f}(\xi)e^{2\pi i x\xi}d\xi, \quad t \geq 0, \quad x \in \mathbb{R}^n.$$

The function $u(x,t)$ is harmonic:

$$\Delta_{x,t}u = 0 \text{ in } \mathbb{R}^{n+1}_+ = \mathbb{R}^n \times (0,+\infty),$$

and satisfies the boundary condition on the hyper-plane $t = 0$:

$$u(x,0) = f(x), \quad x \in \mathbb{R}^n.$$

We can write $u(x,t)$ as a convolution

$$u(x,t) = P_t \star f = \int P_t(x-y)f(y),$$

with

$$\hat{P}_t(\xi) = e^{-2\pi t|\xi|}.$$

Note that in one dimension

$$P_t(x) = \int e^{-2\pi t|\xi|+2\pi i x\xi}d\xi = 2\text{Re}\int_0^\infty e^{-2\pi t\xi+2\pi i x\xi}d\xi = \frac{1}{2\pi}\left[\frac{1}{t+ix} + \frac{1}{t-ix}\right] = \frac{t}{\pi(t^2+x^2)}.$$

It dimension $n > 1$ we have

$$P_t(x) = \int e^{-2\pi t|\xi|+2\pi i\xi\cdot x}d\xi = C_n\frac{t}{(t^2+|x|^2)^{(n+1)/2}}.$$

Here the constant $n$ depends only on the spatial dimension. We will focus on dimension $n = 1$ and address two questions: the first is in which sense is $f(x)$ the boundary data for $u(x,t)$ as $t \to 0$. The second is as follows: given the harmonic function $u(x,t)$ defined for $t \geq 0$ we can find its harmonic conjugate $v(t,x)$ such that $u + iv$ is a holomorphic function in the upper half plane $\{t > 0\}$. Let $g(x)$ be the boundary value of $v(x,t)$ at $t = 0$. What can we say about the map $f \to g$? Both questions are surprisingly rich, and we will begin with the second question.

### The conjugate Poisson kernel

In order to construct the conjugate harmonic function, for $f \in \mathcal{S}(\mathbb{R})$ define $u(x,t) = P_t \star f$, set $z = x + it$ and write

$$u(z) = \int_{\mathbb{R}} e^{-2\pi t|\xi|}\hat{f}(\xi)e^{2\pi i x\xi}d\xi = \int_0^\infty \hat{f}(\xi)e^{2\pi i z\xi}d\xi + \int_{-\infty}^0 \hat{f}(\xi)e^{2\pi i\bar{z}\xi}d\xi.$$

Consider the function $v(z)$ given by

$$iv(z) = \int_0^\infty \hat{f}(\xi) e^{2\pi i z \xi} d\xi - \int_{-\infty}^0 \hat{f}(\xi) e^{2\pi i \bar{z} \xi} d\xi.$$

As the function

$$u(z) + iv(z) = \int_0^\infty \hat{f}(\xi) e^{2\pi i z \xi} d\xi$$

is analytic in the upper half-plane $\{\mathrm{Im} z > 0\}$(simply because we can differentiate the integral), the function $v$ is the harmonic conjugate of $u$. It can be written as

$$v(z) = \int_{\mathbb{R}} (-i\mathrm{sgn}(\xi)) e^{-2\pi t|\xi|} \hat{f}(\xi) e^{2\pi i x \xi} d\xi = Q_t \star f, \tag{4.1}$$

with

$$\hat{Q}_t(\xi) = -i\mathrm{sgn}(\xi) e^{-2\pi t|\xi|}, \tag{4.2}$$

and

$$Q_t(x) = -i \int e^{2\pi i x \xi} \mathrm{sgn}(\xi) e^{-2\pi t|\xi|} d\xi = -\frac{i}{2\pi} \left[ \frac{1}{t - ix} - \frac{1}{t + ix} \right] = \frac{1}{\pi} \frac{x}{t^2 + x^2}.$$

The Poisson kernel and its conjugate are related by

$$P_t(x) + iQ_t(x) = \frac{i}{\pi(x + it)},$$

which is analytic in $\{\mathrm{Im}\, z \geq 0\}$. The main problem with the conjugate Poisson kernel is that it does not decay fast enough at infinity to be in $L^1(\mathbb{R})$ nor is it regular at $x = 0$ as $t \to 0$. Therefore, we need to make precise the meaning of the limit $t \to 0$ in the convolution in (4.1).

**The principle value of $1/x$**

In order to consider the limit of $Q_t$ as $t \to 0$ let us define the principal value of $1/x$ which is an element of $\mathcal{S}'(\mathbb{R})$ (the Schwartz distributions) defined by

$$\mathrm{P.V.}\frac{1}{x}(\phi) = \lim_{\varepsilon \to 0} \int_{|x| > \varepsilon} \frac{\phi(x)}{x} dx, \quad \phi \in \mathcal{S}(\mathbb{R}).$$

This is well-defined because

$$\mathrm{P.V.}\frac{1}{x}(\phi) = \int_{|x|<1} \frac{\phi(x) - \phi(0)}{x} dx + \int_{|x|>1} \frac{\phi(x)}{x} dx,$$

thus

$$\left| \mathrm{P.V.}\frac{1}{x}(\phi) \right| \leq C(\|\phi'\|_{L^\infty} + \|x\phi\|_{L^\infty}),$$

and therefore $\mathrm{P.V.}(1/x)$ is, indeed, a distribution in $\mathcal{S}'(\mathbb{R})$. The conjugate Poisson kernel $Q_t$ and the principal value of $1/x$ are related as follows.

**Proposition 4.1** *Let* $Q_t = \frac{1}{\pi} \frac{x}{t^2 + x^2}$, *then for any function* $\phi \in \mathcal{S}(\mathbb{R})$

$$\frac{1}{\pi} P.V. \frac{1}{x}(\phi) = \lim_{t \to 0} \int_{\mathbb{R}} Q_t(x) \phi(x) dx.$$

**Proof.** Let

$$\psi_t(x) = \frac{1}{x} \chi_{t<|x|}(x)$$

so that

$$\text{P.V.} \frac{1}{x}(\phi) = \lim_{t \to 0} \int_{\mathbb{R}} \psi_t(x) \phi(x) dx.$$

Note, however, that

$$\begin{aligned}
&\int (\pi Q_t(x) - \psi_t(x)) \phi(x) dx = \int_{\mathbb{R}} \frac{x \phi(x)}{x^2 + t^2} dx - \int_{|x|>t} \frac{\phi(x)}{x} dx \\
&= \int_{|x|<t} \frac{x \phi(x)}{x^2 + t^2} dx + \int_{|x|>t} \left[ \frac{x}{x^2 + t^2} - \frac{1}{x} \right] \phi(x) dx \qquad (4.3) \\
&= \int_{|x|<1} \frac{x \phi(tx)}{x^2 + 1} dx - \int_{|x|>t} \frac{t^2 \phi(x)}{x(x^2 + t^2)} dx = \int_{|x|<1} \frac{x \phi(tx)}{x^2 + 1} dx - \int_{|x|>1} \frac{\phi(tx)}{x(x^2 + 1)} dx.
\end{aligned}$$

The dominated convergence theorem implies that both integrals on the utmost right side above tend to zero as $t \to 0$. $\square$

It is important to note that the computation in (4.3) worked only because the kernel $1/x$ is odd – this produces the cancellation that saves the day. This would not happen, for instance, for a kernel behaving as $1/|x|$ near $x = 0$.

### The Hilbert transform

Motivated by the previous discussion, for a function $f \in \mathcal{S}(\mathbb{R})$, we define the Hilbert transform as

$$Hf(x) = \lim_{t \to 0} Q_t \star f(x) = \frac{1}{\pi} \lim_{\varepsilon \to 0} \int_{|y|>\varepsilon} \frac{f(x-y)}{y} dy.$$

It follows from (4.2) that

$$\widehat{Hf}(\xi) = \lim_{\varepsilon \to 0} \hat{Q}_t(\xi) \hat{f}(\xi) = -i \text{sgn}(\xi) \hat{f}(\xi). \qquad (4.4)$$

Therefore, the Hilbert transform may be extended to an isometry $L^2(\mathbb{R}) \to L^2(\mathbb{R})$, with $\|Hf\|_{L^2} = \|f\|_{L^2}$, $H(Hf) = -f$ and

$$\int (Hf)(x) g(x) dx = -\int f(x) (Hg)(x) dx. \qquad (4.5)$$

The following extension of the Hilbert transform to $L^p$-spaces for $1 < p < \infty$ is due to M. Riesz.

**Theorem 4.2** *Given $1 < p < \infty$ there exists $C_p > 0$ so that*

$$\|Hf\|_{L^p} \leq C_p\|f\|_{L^p} \text{ for all } f \in L^p(\mathbb{R}^n). \tag{4.6}$$

The proof of this theorem requires some basic interpolation theory that we will now develop. Note that the result fails at both ends: both for $p = 1$ and $p = \infty$.

### Interpolation in $L^p$-spaces

A simple example of an interpolation inequality is a bound that tells us that a function $f$ which lies in two spaces $L^{p_0}(\mathbb{R}^n, d\mu)$ and $L^{p_1}(\mathbb{R}^n, d\mu)$ has to lie also in all intermediate spaces $L^p(\mathbb{R}^n, d\mu)$ with $p_0 \leq p \leq p_1$. Indeed, if $p = \alpha p_0 + (1-\alpha)p_1$, $0 < \alpha < 1$, then, by Hölder's inequality, with $q = 1/\alpha$ and $q' = 1/(1-\alpha)$, we get

$$\int |f|^{\alpha p_0 + (1-\alpha)p_1} d\mu \leq \left(\int |f|^{p_0} d\mu\right)^{\alpha} \left(\int |f|^{p_1} d\mu\right)^{1-\alpha}.$$

### The Riesz-Thorin interpolation theorem

The Riesz-Thorin interpolation theorem deals with the following question, somewhat motivated by above. Let $(M, \mu)$ and $(N, \nu)$ be two measure spaces and consider an operator $A$ which maps $L^{p_0}(M)$ to a space $L^{q_0}(N)$, and also $L^{p_1}(M)$ to a space $L^{q_1}(N)$. More precisely, there exist operators $A_0 : L^{p_0}(M) \to L^{q_0}(N)$ and $A_1 : L^{p_1}(M) \to L^{q_1}(N)$ so that $A = A_0 = A_1$ on $L^{p_0}(M) \cap L^{p_1}(N)$. The question is whether $A$ can be defined on $L^p(M)$ with $p_0 < p < p_1$, and what is its target space. Let us define $p_t \in (p_0, p_1)$ and $q_t \in (q_0, q_1)$ by

$$\frac{1}{p_t} = \frac{t}{p_1} + \frac{1-t}{p_0}, \quad \frac{1}{q_t} = \frac{t}{q_1} + \frac{1-t}{q_0}, \quad 0 \leq t \leq 1, \tag{4.7}$$

as well as

$$k_0 = \|A\|_{L^{p_0}(M) \to L^{q_0}(N)}, \quad k_1 = \|A\|_{L^{p_1}(M) \to L^{q_1}(N)}.$$

**Theorem 4.3** *(The Riesz-Thorin interpolation theorem) For any $t \in [0,1]$ there exists a bounded linear operator $A_t : L^{p_t}(M) \to L^{q_t}(N)$ that coincides with $A$ on $L^{p_0}(M) \cap L^{p_1}(M)$ and whose operator norm satisfies*

$$\|A_t\|_{L^{p_t}(M) \to L^{q_t}(N)} \leq k_0^{1-t} k_1^t. \tag{4.8}$$

Before proving the Riesz-Thorin interpolation theorem we mention some of its implications. We already know that the Fourier transform maps $L^1(\mathbb{R}^n)$ to $L^\infty(\mathbb{R}^n)$ and $L^2(\mathbb{R}^n)$ to itself. This allows us to extend the Fourier transform to all intermediate spaces $L^p(\mathbb{R}^n)$ with $1 \leq p \leq 2$.

**Corollary 4.4** *(The Hausdorff-Young inequality) Let $1 \leq p \leq 2$, then if $f \in L^p(\mathbb{R}^n)$ then its Fourier transform $\hat{f} \in L^{p'}(\mathbb{R}^n)$ with $\frac{1}{p} + \frac{1}{p'} = 1$ and $\|\hat{f}\|_{L^{p'}} \leq \|f\|_{L^p}$.*

**Proof.** We take $p_0 = 1$, $p_1 = 2$, $q_0 = \infty$, $q_1 = 2$. Then for any $t \in [0,1]$ the corresponding $p_t$ and $q_t$ are given by

$$\frac{1}{p_t} = \frac{1-t}{1} + \frac{t}{2} = 1 - \frac{t}{2}, \quad \frac{1}{q_t} = \frac{t}{2},$$

which means that $1/p_t + 1/q_t = 1$, as claimed. Furthermore, as $\|\hat{f}\|_{L^2} = \|f\|_{L^2}$ by the Parceval identity and $\|\hat{f}\|_{L^\infty} \le \|f\|_{L^1}$, it follows that $\|\hat{f}\|_{L^{p_t} \to L^{q_t}} \le 1$. □

The next corollary of the Riesz-Thorin theorem, also sometimes called the Hausdorff-Young inequality, allows to estimate convolutions.

**Corollary 4.5** *Let $f \in L^p(\mathbb{R}^n)$ and $g \in L^q(\mathbb{R}^n)$, then $f \star g \in L^r(\mathbb{R}^n)$, and*

$$\|f \star g\|_{L^r} \le \|f\|_{L^p} \|g\|_{L^q}, \tag{4.9}$$

*with $r$ determined by*

$$\frac{1}{r} + 1 = \frac{1}{p} + \frac{1}{q}. \tag{4.10}$$

**Proof.** We do this in two steps. First, fix $g \in L^1(\mathbb{R}^n)$. Obviously, we have

$$\|f \star g\|_{L^1} \le \int |f(x-y)||g(y)|dydx = \|f\|_{L^1}\|g\|_{L^1}, \tag{4.11}$$

and

$$\|f \star g\|_{L^\infty} \le \|f\|_{L^\infty}\|g\|_{L^1}. \tag{4.12}$$

The Riesz-Thorin theorem applied to the map $f \to f \star g$ implies then that

$$\|f \star g\|_{L^p} \le \|g\|_{L^1}\|f\|_{L^p}, \tag{4.13}$$

which is a special case of (4.9) with $q = 1$ and $r = p$. On the other hand, Hölder's inequality implies that

$$\|f \star g\|_{L^\infty} \le \|f\|_{L^p}\|g\|_{L^{p'}}, \quad \frac{1}{p} + \frac{1}{p'} = 1. \tag{4.14}$$

Let us take $p_0 = 1$, $q_0 = p$, $p_1 = p'$ and $q_1 = \infty$ in the Riesz-Thorin interpolation theorem applied to the mapping $g \to f \star g$, with $f$ fixed. Then (4.13) and (4.14) imply that, for all $t \in [0,1]$,

$$\|f \star g\|_{L^r} \le \|f\|_{L^p}\|g\|_{L^q},$$

with

$$\frac{1}{q} = \frac{1}{p_t} = \frac{1-t}{1} + \frac{t}{p'},$$

and

$$\frac{1}{r} = \frac{1}{q_t} = \frac{1-t}{p} + \frac{t}{\infty}.$$

It follows that $t = 1 - p/r$, thus

$$\frac{1}{q} = 1 - (1 - \frac{p}{r}) + \frac{1}{p'}(1 - \frac{p}{r}) = \frac{p}{r} + (1 - \frac{1}{p})(1 - \frac{p}{r}) = 1 - \frac{1}{p} + \frac{1}{r},$$

which is (4.10). $\Box$

The next example arises in microlocal analysis. Given a function $a(x, \xi) \in \mathcal{S}(\mathbb{R}^{2n})$ we define a semiclassical operator

$$A(x, \varepsilon D)f = \int e^{2\pi i \xi \cdot x} a(x, \varepsilon \xi) \hat{f}(\xi) d\xi.$$

Here $\varepsilon \in (0, 1)$ is the parameter that plays the role of the Planck constant in physics and is, therefore, small.

**Corollary 4.6** *The family of operators $A(x, \varepsilon D)$, $0 < \varepsilon \leq 1$, is uniformly bounded from any $L^p(\mathbb{R}^n)$, $1 \leq p \leq +\infty$, to itself.*

**Proof.** Let us write

$$A(x, \varepsilon D)f = \int e^{2\pi i \xi \cdot x} a(x, \varepsilon \xi) \hat{f}(\xi) d\xi = \int e^{2\pi i \xi \cdot x + 2\pi i \varepsilon \xi \cdot y} \widetilde{a}(x, y) \hat{f}(\xi) d\xi dy = \int \widetilde{a}(x, y) f(x + \varepsilon y) dy,$$

where $\widetilde{a}(x, y)$ is the Fourier transform of the function $a(x, \xi)$ in the variable $\xi$. It follows that

$$\|A(x, \varepsilon D)f\|_{L^\infty} \leq \|f\|_{L^\infty} \sup_{x \in \mathbb{R}^n} \int |\widetilde{a}(x, y)| dy = C_1(a) \|f\|_{L^\infty},$$

and

$$\begin{aligned} \|A(x, \varepsilon D)\|_{L^1} &\leq \int |\widetilde{a}(x, y)| |f(x + \varepsilon y)| dy dx \leq \int (\sup_{z \in \mathbb{R}^n} |\widetilde{a}(z, y)|) |f(x + \varepsilon y)| dy dx \\ &= \|f\|_{L^1} \int (\sup_{z \in \mathbb{R}^n} |\widetilde{a}(z, y)|) dy = C_2(a) \|f\|_{L^1}. \end{aligned}$$

The Riesz-Thorin interpolation theorem implies then that for any $p \in [1, +\infty]$ there exists $C_p(a)$ which does not depend on $\varepsilon \in (0, 1]$ so that $\|A(x, \varepsilon D)\|_{L^p \to L^p} \leq C_p$. $\Box$

### The proof of the Riesz-Thorin interpolation theorem

The proof is based on the three lines theorem. First, let us define how the operator $A$ acts on $L^{p_t}(M)$ with $p_t$ as in (4.7). Given $f \in L^{p_t}(M)$ we can decompose it as

$$f(x) = f_1(x) + f_0(x), \quad f_1(x) = f(x) \chi_{|f| \leq 1}(x), \quad f_0(x) = f(x) \chi_{|f| \geq 1}(x).$$

Then, as $p_t \leq p_1$:

$$\int_M |f_1|^{p_1} d\mu = \int_M |f|^{p_1} \chi_{|f| \leq 1} d\mu \leq \int_M |f|^{p_t} \chi_{|f| \leq 1} d\mu \leq \int_M |f|^{p_t} d\mu = \|f\|_{L^{p_t}}^{p_t},$$

and, as $p_0 \le p_t$:

$$\int_M |f_0|^{p_0} d\mu = \int_M |f|^{p_t} \chi_{|f|\ge 1} d\mu \le \int_M |f|^{p_t} \chi_{|f|\ge 1} d\mu \le \int_M |f|^{p_t} d\mu = \|f\|_{L^{p_t}}^{p_t},$$

so that $f_1 \in L^{p_1}(M)$ and $f_0 \in L^{p_0}(M)$. As $A$ is defined both on $L^{p_0}(M)$ and $L^{p_1}(M)$, we can set

$$Af = Af_1 + Af_0.$$

We now need to verify that $A$ maps $L^{p_t}(M)$ to $L^{q_t}(N)$ continuously, and obtain the bound on its norm as in the theorem. Recall that the space $L^p(M)$, $1 < p \le +\infty$ is the dual space of $L^{p'}(M)$ where $p$ and $p'$ are related by

$$\frac{1}{p} + \frac{1}{p'} = 1.$$

Note also that the norm of a bounded linear functional $L_f : L^{p'}(M) \to \mathbb{R}$,

$$L_f(g) = \int_M fg d\mu, \quad f \in L^p(M),$$

is $\|L_f\| = \|f\|_{L^p}$, for all $p \in [1, +\infty]$. To see that, for $f(x) = |f(x)|e^{i\alpha(x)}$ simply take $g(x) = |f(x)|^{p/p'} \exp\{-i\alpha(x)\}$ for $1 < p < +\infty$, $g(x) = \exp\{-i\alpha(x)\}$ for $p = 1$, and $g(x) = \chi_{A_\varepsilon}(x) \exp\{-i\alpha(x)\}$, where $A_\varepsilon$ is a set of a finite measure such that $|f(x)| > (1-\varepsilon)\|f|_{L^\infty}$ on $A_\varepsilon$ for $p = +\infty$. We conclude that

$$\|f\|_{L^p} = \sup_{\|g\|_{L^{p'}}=1} \int_M fg d\mu, \quad \frac{1}{p} + \frac{1}{p'} = 1.$$

For an operator mapping $L^p$ to $L^q$ we have the corresponding representation for its norm:

$$\|A\|_{L^p(M)\to L^q(N)} = \sup_{\|f\|_{L^p(M)}=1} \|Af\|_{L^q(N)} = \sup_{\substack{\|f\|_{L^p(M)}=1 \\ \|g\|_{L^{q'}(N)}=1}} \int_N (Af) g d\nu. \tag{4.15}$$

We will base our estimate of the norm of $A : L^{p_t}(M) \to L^{q_t}(N)$ on (4.15) . Moreover, as simple functions are dense in $L^{p_t}(M)$ and $L^{q_t'}(N)$, it suffices to use in (4.15) only simple functions $f$ and $g$ with $\|f\|_{L^{p_t}(M)} = \|g\|_{L^{q_t'}(N)} = 1$, of the form

$$f(x) = \sum_{j=1}^n a_j e^{i\alpha_j(x)} \chi_{A_j}(x), \quad g(y) = \sum_{j=1}^m b_j e^{i\beta_j(y)} \chi_{B_j}(y), \quad x \in M, \ y \in N, \tag{4.16}$$

with $a_j, b_j > 0$, $\mu$-measurable sets $A_j$ and $\nu$-measurable sets $B_j$. Since $0 < t < 1$, neither $p_t$ nor $q_t'$ can be equal to $+\infty$, hence $\mu(A_j), \nu(B_j) < +\infty$.

Let us now extend the definition of $p_t$ and $q_t$ to all complex numbers $\zeta$ with $0 \le \operatorname{Re} \zeta \le 1$:

$$\frac{1}{p(\zeta)} = \frac{1-\zeta}{p_0} + \frac{\zeta}{p_1}, \quad \frac{1}{q(\zeta)} = \frac{1-\zeta}{q_0} + \frac{\zeta}{q_1}, \quad \frac{1}{q'(\zeta)} = \frac{1-\zeta}{q_0'} + \frac{\zeta}{q_1'}.$$

Fix $t \in (0,1)$ and a pair of (complex-valued) functions $f \in L^{p_t}(M)$ and $g \in L^{q_t'}(M)$ of the form (4.16). Consider a family of functions

$$u(x,\zeta) = \sum_{j=1}^{n} a_j^{p_t/p(\zeta)} e^{i\alpha_j(x)} \chi_{A_j}(x), \quad v(y,\zeta) = \sum_{j=1}^{m} b_j^{q_t'/q'(\zeta)} e^{i\beta_j(y)} \chi_{B_j}(y),$$

with $x \in M$, $y \in N$ and $0 \le \text{Re } \zeta \le 1$. Note that, when $\zeta = t$,

$$u(x,t) = f(x) \text{ and } v(y,t) = g(y). \tag{4.17}$$

As both $1/p(\zeta)$ and $1/q'(\zeta)$ are linear in $\zeta$, the functions $u(x,\zeta)$ and $v(x,\zeta)$ are analytic in $\zeta$ in the strip $S = \{\zeta : \ 0 \le \text{Re } \zeta \le 1\}$. Since $u(x,\zeta)$ and $v(y,\zeta)$ are simple functions of $x$ and $y$, respectively, vanishing outside of a set of finite measure for each $\zeta \in S$ fixed, they lie in $L^{p_0}(M) \cap L^{p_1}(M)$, and $L^{q_0'}(M) \cap L^{q_1'}(M)$, respectively. Therefore, we can define

$$F(\zeta) = \int_N (Au)(y,\zeta) v(y,\zeta) d\nu = \sum_{j=1}^{n} \sum_{k=1}^{m} a_j^{p_t/p(\zeta)} b_k^{q_t'/q'(\zeta)} \int_N (A\Psi_j)(y) e^{i\beta_k(y)} \chi_{B_k}(y) d\nu,$$

with $\Psi_j(x) = e^{i\alpha_j(x)}\chi_{A_j}(x)$. According to (4.15) and (4.17) , in order to prove that

$$\|A_t\|_{L^{p_t}(M) \to L^{q_t}(N)} \le k_0^{1-t} k_1^t, \tag{4.18}$$

it suffices to show that

$$|F(t)| \le k_0^{1-t} k_1^t. \tag{4.19}$$

The function $F(\zeta)$ is analytic and bounded in the strip $S$, as, for instance, for $\zeta = \eta + i\xi$, $0 \le \eta \le 1$:

$$\left| a_j^{p_t/p(\zeta)} \right| = \left| a_j^{p_t \zeta/p_1 + p_t(1-\zeta)/p_0} \right| = \left| a_j^{p_t \eta/p_1 + p_t(1-\eta)/p_0} \right| \le C_j < +\infty.$$

On the boundary of the strip $S$ we have the following bounds: along the line $\eta = 0$, for $z = i\xi$,

$$\begin{aligned}
\|u(x,i\xi)\|_{L^{p_0}(M)} &= \left( \int_M \sum_{j=1}^{n} \left| a_j^{[p_t(i\xi)/p_1 + p_t(1-i\xi)/p_0]p_0} \right| \chi_{A_j}(x) d\mu \right)^{1/p_0} \\
&= \left( \int_M \sum_{j=1}^{n} |a_j|^{p_t} \chi_{A_j}(x) d\mu \right)^{1/p_0} = \|f\|_{L^{p_t}(M)}^{p_t/p_0} = 1,
\end{aligned}$$

and

$$\begin{aligned}
\|v(y,i\xi)\|_{L^{q_0'}(N)} &= \left( \int_N \sum_{j=1}^{m} \left| b_j^{[q_t'(i\xi)/q_1' + q_t'(1-i\xi)/q_0']q_0'} \right| \chi_{B_j}(y) d\nu \right)^{1/q_0'} \\
&= \left( \int_N \sum_{j=1}^{m} |b_j|^{q_t'} \chi_{B_j}(y) d\nu \right)^{1/p_0} = \|g\|_{L^{q_t'}(N)}^{q_t'/q_0'} = 1.
\end{aligned}$$

It follows that

$$|F(i\xi)| \leq \|(Au)(i\xi)\|_{L^{q_0}(N)}\|v(i\xi)\|_{L^{q_0'}(N)} \leq \|A\|_{L^{p_0}(M)\to L^{q_0}(N)}\|u(i\xi)\|_{L^{p_0}(N)}\|v(i\xi)\|_{L^{q_0'}(N)} \leq k_0.$$

Similarly, we can show that along the line $\zeta = 1+i\xi$ we have $\|u(x,1+i\xi)\|_{L^{p_1}(M)} \leq 1$ and $\|v(x,1+i\xi)\|_{L^{q_1'}(N)} \leq 1$, which implies that $|F(1+i\xi)| \leq k_1$. The three lines theorem implies now that $|F(\eta+i\xi)| \leq k_0^{1-\eta}k_1^{\eta}$, hence (4.19) holds. □

### The $L^p$-bounds on the Hilbert transform

Recall that we would like to prove the following theorem.

**Theorem 4.7** *Given $1 < p < \infty$ there exists $C_p > 0$ so that*

$$\|Hf\|_{L^p} \leq C_p\|f\|_{L^p} \text{ for all } f \in L^p(\mathbb{R}). \tag{4.20}$$

**Proof.** The issue here is that we would like to use the interpolation theorem for the proof but we only have one trivial bound:

$$\|Hf\|_2 = \|f\|_2,$$

and interpolation requires two bounds! Hence, we start fishing for the second bound. We first consider $p \geq 2$. It suffices to establish (4.20) for $f \in \mathcal{S}(\mathbb{R})$. Consider a smaller set

$$\mathcal{S}_0 = \{f \in \mathcal{S} : \exists \varepsilon > 0 \text{ such that } \hat{f}(\xi) = 0 \text{ for } |\xi| < \varepsilon\}.$$

Let us show that $\mathcal{S}_0$ is dense in $L^p(\mathbb{R})$. Given any $f \in \mathcal{S}$ we'll find a sequence $g_n \in \mathcal{S}_0$ such that $\|f - g_n\|_{L^p} \to 0$ as $n \to +\infty$. For $p = 2$ this is trivial: take a smooth function $\chi(\xi)$ such that $0 \leq \chi(\xi) \leq 1$, $\chi(\xi) = 0$ for $|\xi| \leq 1$, $\chi(\xi) = 1$ for $|\xi| > 2$, and set

$$g_n(x) = \int e^{2\pi i\xi x}\hat{f}(\xi)\chi(n\xi)\,d\xi,$$

so that

$$\|f - g_n\|_{L^2}^2 \leq \int_{-2/n}^{2/n} |\hat{f}(\xi)|^2 d\xi \to 0 \text{ as } n \to +\infty. \tag{4.21}$$

On the other hand, for $p = +\infty$ we have

$$\|f - g_n\|_{L^\infty} \leq \int_{-2/n}^{2/n} |\hat{f}(\xi)| d\xi \to 0 \text{ as } n \to +\infty. \tag{4.22}$$

Interpolating between $p = 2$ and $p = +\infty$ we conclude that

$$\|f - g_n\|_{L^p} \to 0 \text{ as } n \to +\infty \tag{4.23}$$

for all $p \geq 2$, hence $\mathcal{S}_0$ is dense in $L^p(\mathbb{R})$ for $2 \leq p < +\infty$.

Given $f \in \mathcal{S}_0$, $\widehat{Hf}(\xi) = -i(\text{sgn}\xi)\hat{f}(\xi)$ is a Schwartz class function (there is no discontinuity at $\xi = 0$), thus $Hf$ is also in $\mathcal{S}(\mathbb{R})$. We may then write

$$p(x) = (f + iHf)(x) = \int_{\mathbb{R}} (1 + \text{sgn}(\xi))\hat{f}(\xi)e^{2\pi i\xi x}d\xi = 2\int_0^\infty \hat{f}(\xi)e^{2\pi i\xi x}d\xi,$$

and consider its extension to the complex plane:

$$p(z) = 2\int_0^\infty \hat{f}(\xi)e^{2\pi i\xi z}d\xi.$$

The function $p(z)$ is holomorphic in the upper half-plane $\{\text{Im}z > 0\}$ and is continuous up to the boundary $y = 0$. Since $f \in \mathcal{S}_0$ there exists $\varepsilon > 0$ so that $\hat{f}(\xi) = 0$ for $|\xi| \le \varepsilon$. Thus, $p(z)$ satisfies an exponential decay bound

$$|p(z)| \le 2e^{-2\pi\varepsilon y}\|\hat{f}\|_{L^1}, \quad z = x + iy. \tag{4.24}$$

Integrating $p^4(z)$ along the contour $C_R$ which consists of the interval $[-R, R]$ along the real axis and the semicircle $\{x^2 + y^2 = R^2,\ y > 0\}$, and passing to the limit $R \to 0$ with the help of (4.24) leads to

$$\lim_{R\to+\infty} \int_{-R}^{R} (f(x) + iHf(x))^4 dx = 0.$$

As both $f$ and $Hf$ are in $\mathcal{S}_0$, the integral above converges absolutely, hence

$$\int_{\mathbb{R}} (f(x) + iHf(x))^4 dx = 0.$$

The real part above gives

$$\int_{\mathbb{R}} (Hf(x))^4 dx = \int_{\mathbb{R}} [-f^4(x) + 2f^2(x)(Hf)^2(x)]dx \le 2\int f^2(x)(Hf)^2(x)dx$$
$$\le \int (2f^4(x) + \frac{1}{2}(Hf)^4(x))dx,$$

hence

$$\int_{\mathbb{R}} (Hf(x))^4 dx \le 4\int f^4(x)dx, \tag{4.25}$$

for any function $f \in \mathcal{S}_0$. As we have shown that $\mathcal{S}_0$ is dense in any $L^p(\mathbb{R})$, $2 \le p < \infty$, we know that (4.25) holds for all $f \in L^4(\mathbb{R})$. Therefore, the Hilbert transform is a bounded operator $L^4(\mathbb{R}) \to L^4(\mathbb{R})$. As we know that it is also bounded from $L^2(\mathbb{R})$ to $L^2(\mathbb{R})$, the Riesz-Thorin interpolation theorem implies that $\|Hf\|_{L^p} \le C_p\|f\|_{L^p}$ for all $2 \le p \le 4$.

An argument identical to the above, integrating the function $p^{2k}(z)$ over the same contour, shows that $H$ is bounded from $L^{2k}(\mathbb{R})$ to $L^{2k}(\mathbb{R})$ for all integers $k \ge 1$. It follows then from the Riesz-Thorin interpolation theorem that $\|Hf\|_{L^p} \le C_p\|f\|_{L^p}$ for all $2 \le p < +\infty$.

It remains to consider $1 < p < 2$ – this is done using the duality argument. Let $q > 2$ be the dual exponent of $p$, $1/p + 1/q = 1$. As the operator $H : L^q(\mathbb{R}) \to L^q(\mathbb{R})$ is bounded – this follows from what we have already done, as $q > 2$, so is its adjoint $H^* : L^p(\mathbb{R}) \to L^p(\mathbb{R})$ defined by $\langle H^* f, g\rangle = \langle f, Hg\rangle$, with $f \in L^p(\mathbb{R})$, $g \in L^q(\mathbb{R})$. However, identity (4.5) says that $H^* = -H$, hence the boundedness of $H^*$ implies that $H : L^p(\mathbb{R}) \to L^p(\mathbb{R})$ is also bounded. $\square$

The Hilbert transform does not map $L^1(\mathbb{R}) \to L^1(\mathbb{R})$ but we have the following result due to Kolmogorov. We say that a function $g$ belongs to weak-$L^1$: $g \in L^1_w$ if

$$P(\lambda) := |\{x : \ |g(x)| > \lambda\}| \le \frac{C}{\lambda},$$

for all $\lambda > 0$. This notion is motivated by the identity

$$\int |g(x)| dx = \int_0^\infty |\{x : \ |g(x)| > \lambda\}| d\lambda. \tag{4.26}$$

Hence, for $L^1$-functions the "tail distribution function" $P(\lambda)$ is integrable while for $L^1_w$-functions it may barely miss being integrable. Note that if $g \in L^1(\mathbb{R})$ then

$$\lambda |\{x : \ |g(x)| > \lambda\}| \le \int |g(x)| dx = \|g\|_1,$$

thus all functions in $L^1$ lie in $L^1_w$. On the other hand, for instance, the function $g(x) = 1/x$ in $\mathbb{R}$ lies in $L^1_w(\mathbb{R})$ but not in $L^1(\mathbb{R})$. We have the following theorem.

**Theorem 4.8** *Let $f \in L^1(\mathbb{R})$, then there exists $C > 0$ so that for any $\lambda > 0$ the following estimate holds:*

$$m\{x : \ |Hf(x)| \ge \lambda\} \le \frac{C}{\lambda} \int_{\mathbb{R}} |f(x) dx.$$

We will not prove this theorem here. However, we note that the Marcinkiewicz interpolation theorem (a generalization of the Riesz-Thorin theorem) allows us to conclude from Theorem 4.8 and the obvious boundedness of the Hilbert trasform from $L^2$ to $L^2$ that the Hilbert transform is bounded from any $L^p(\mathbb{R})$ to $L^p(\mathbb{R})$ providing an alternative proof of Theorem 4.7.

The Marcinkiewicz theorem says the following. First, let us generalize (4.26) to $p > 1$. Generally, for an increasing differentiable function $\phi(s)$ we have the relation

$$\int \phi(|f(x)|) dx = \int_0^\infty \phi'(\lambda) P(\lambda) dx. \tag{4.27}$$

To see that, we simply write

$$\int \phi(|f(x)|) dx = \int \int_0^{|f(x)|} \phi'(\lambda) d\lambda dx = \int_0^\infty \phi'(\lambda) |\{x : \ |f(x)| > \lambda\} d\lambda = \int_0^\infty \phi'(\lambda) P(\lambda) dx.$$

As a consequence, for $p > 1$ we have

$$\|f\|_p^p = p \int_0^\infty \lambda^{p-1} P(\lambda) d\lambda. \tag{4.28}$$

We say that an operator $T$ from $L^p$ to $L^q$ (with $q < \infty$) is weak $(p, q)$ if

$$|\{y : \ |Tf(y)| > \lambda\}| \leq \left(\frac{C\|f\|_p}{\lambda}\right)^q,$$

and we say that $T$ is weak $(p, \infty)$ if $T$ is bounded from $L^p$ to $L^\infty$. Note that if $T$ is bounded from $L^p$ to $L^q$ then it is weak $(p, q)$, because if we set

$$E_\lambda = \{y : \ |Tf(y)| > \lambda\},$$

then

$$|E_\lambda| = \int_{E_\lambda} 1 \leq \int_{E_\lambda} \left|\frac{Tf(x)}{\lambda}\right|^q \leq \frac{\|Tf\|_q^q}{\lambda^q} \leq \left(\frac{C\|f\|_p}{\lambda}\right)^q.$$

**Theorem 4.9** *(Marcinkiewicz Interpolation Theorem) Let $1 \leq p_0 < p_1 \leq \infty$, and let $T$ be a sublinear operator that is weak $(p_0, p_0)$and weak $(p_1, p_1)$. Then $T$ is a bounded operator from any $L^p$ to $L^p$ with $p_0 < p < p_1$.*

The remarkable fact is that weak bounds at the end-points imply a strong bound for intermediate values of $p$.

# 5 Harmonic functions on $\mathbb{D}$

We now consider the properties of harmonic functions on the unit disk $\mathbb{D} = \{|z| < 1\}$.

## 5.1 The Poisson kernel

We have discussed above the Hilbert transform. Recall how it was defined: we take a function $f$, extend it to a harmonic function $u(x, t)$ in the upper half plane, find the conjugate to $u$ harmonic function $v(x, t)$ and consider the limit $t \to 0$. Then "$Hf(x) = v(x, 0)$, at least in the $L^p$-sense. Here, we will be concerned with the pointwise convergence of the harmonic extension $u(x, t)$ to the function $f(x)$ itself. We will also consider the problem on the unit disk rather than on the upper half plane but these questions are equivalent for these two domains.

Let us derive a candidate solution formula for the above problem: given a function $f$ on the boundary of $\mathbb{D}$ find a harmonic function $u$ on $\mathbb{D}$ which attains these boundary values. This Dirichlet problem is formulated too vaguely – much of what we will do now will be devoted to a proper interpretation of what we mean by *attaining* the boundary values and what kind of regularity we wish $u$ to satisfy on all of $\mathbb{D}$. For the moment, let us proceed heuristically. Starting with the Fourier series for the function $f$:

$$f(\theta) = \sum_{n \in \mathbb{Z}} \hat{f}(n) e^{2\pi i n \theta},$$

we observe that one harmonic extension to the interior is given by

$$u(z) = \sum_{n\in\mathbb{Z}} \hat{f}(n) z^n = \sum_{n\in\mathbb{Z}} \hat{f}(n) r^n e^{2\pi i n\theta}, \qquad z = re^{2\pi i\theta}$$

This is singular at $z = 0$, though, in case $\hat{f}(n) \neq 0$ for some $n < 0$. Since both $z^n$ and $\overline{z^n}$ are (complex) harmonic, we can avoid the singularity by defining

$$u(z) = \sum_{n=0}^{\infty} \hat{f}(n) z^n + \sum_{n=-\infty}^{-1} \hat{f}(n) \bar{z}^{|n|} \tag{5.1}$$

which at least formally is a solution of our Dirichlet problem.

Inserting $z = re^{2\pi i\theta}$ and

$$\hat{f}(n) = \int_0^1 e^{-2\pi i n\varphi} f(\varphi)\, d\varphi$$

into (5.1) yields

$$u(re^{2\pi i\theta}) = \int_0^1 \sum_{n\in\mathbb{Z}} r^{|n|} e^{2\pi i n(\theta-\varphi)} f(\varphi)\, d\varphi =: \int_0^1 P_r(\theta-\varphi)\, f(\varphi)\, d\varphi$$

where the *Poisson kernel*

$$P_r(\theta) := \sum_{n\in\mathbb{Z}} r^{|n|} e^{2\pi i n\theta} = \frac{1-r^2}{1-2r\cos(2\pi\theta)+r^2}$$

as can be verified by explicit summation. This is a formal answer, and our goal is to understand how it can be interpreted. We start with some properties of $P_r$.

**Lemma 5.1** *The function $u(z) = P_r(\theta)$, $z = re^{2\pi i\theta}$ is a positive harmonic function on $\mathbb{D}$. It satisfies*

$$\int_0^1 P_r(\theta)\, d\theta = 1,$$

*for any $0 \leq r < 1$, and for any (complex) Borel measure $\mu$ on $\mathbb{T}$,*

$$u(z) = (P_r * \mu)(\theta)$$

*defines a harmonic function on $\mathbb{D}$.*

*Proof.* The fact that integral of the Poisson kernel equals one, follows immediately from its defining series. The fact that a convolution with a measure is harmonic follows from the formula for $P_r(\theta)$. □

The Poisson kernel close to the boundary is an approximation of identity in the following sense.

**Definition 5.2** *A sequence $\Phi_n \subset L^\infty(\mathbb{T})$ is called* an approximate identity *provided*

*(A1)* $\int_0^1 \Phi_n(\theta)\, d\theta = 1$ *for all* $n$

*(A2)* $\sup_n \int_0^1 |\Phi_n(\theta)|\, d\theta < \infty$

*(A3) for all* $\delta > 0$ *one has* $\int_{|x|>\delta} |\Phi_n(\theta)|\, d\theta \to 0$ *as* $n \to \infty$.

*The same definition applies, with obvious modifications, to families of the form* $\{\Phi_t\}_{0<t<1}$ *(with* $n \to \infty$ *replaced by* $t \to 1-$*).*

A standard example is the box kernel

$$\left\{ \frac{1}{2\varepsilon} \chi_{[-\varepsilon,\varepsilon]} \right\}_{0<\varepsilon<\frac{1}{2}}$$

in the limit $\varepsilon \to 0$. The main example for us right now is, of course, the Poisson kernel $\{P_r\}_{0<r<1}$. We leave it to the reader to check that it satisfies (A1)–(A3) as $r \to 1$. The significance of approximate identities lies in the following.

**Lemma 5.3** *For any approximate identity* $\Phi_n$ *one has*

1. *If* $f \in C(\mathbb{T})$, *then* $\|\Phi_n * f - f\|_\infty \to 0$ *as* $n \to \infty$
2. *If* $f \in L^p(\mathbb{T})$ *where* $1 \le p < \infty$, *then* $\|\Phi_n * f - f\|_p \to 0$ *as* $n \to \infty$.

*These statements carry over to approximate identities* $\Phi_t$, $0 < t < 1$ *simply by replacing* $n \to \infty$ *with* $t \to 1$.

*Proof.* For the proof of the first statement, note that, since $\mathbb{T}$ is compact, $f$ is uniformly continuous. Given $\varepsilon > 0$, let $\delta > 0$ be such that

$$\sup_x \sup_{|y|<\delta} |f(x-y) - f(x)| < \varepsilon$$

Then, by (A1)–(A3), we have

$$\begin{aligned} |(\Phi_n * f)(x) - f(x)| &= \Big| \int_{\mathbb{T}} (f(x-y) - f(x)) \Phi_n(y)\, dy \Big| \\ &\le \sup_{x\in\mathbb{T}} \sup_{|y|<\delta} |f(x-y) - f(x)| \int_{\mathbb{T}} |\Phi_n(t)|\, dt + \int_{|y|\ge\delta} |\Phi_n(y)| 2\|f\|_\infty\, dy < C\varepsilon \end{aligned}$$

provided $n$ is large.

For the second part, fix $f \in L^p$. Let $g \in C(\mathbb{T})$ with $\|f - g\|_p < \varepsilon$. Then

$$\begin{aligned} &\|\Phi_n * f - f\|_p \le \|\Phi_n * (f-g)\|_p + \|f-g\|_p + \|\Phi_n * g - g\|_p \\ &\le \left( \sup_n \|\Phi_n\|_1 + 1 \right) \|f-g\|_p + \|\Phi_n * g - g\|_\infty \end{aligned}$$

where we have used Hausdorff-Young's inequality

$$\|f_1 * f_2\|_p \leq \|f_1\|_1 \|f_2\|_p$$

to obtain the first term on the right-hand side. Using (A2), the assumption on $g$, as well as the first part finishes the proof. $\square$

An immediate consequence is the following simple and fundamental result.

**Theorem 5.4** *Let $f \in C(\mathbb{T})$. The unique harmonic function $u$ on $\mathbb{D}$, with $u \in C(\overline{\mathbb{D}})$ and $u = f$ on $\mathbb{T}$ is given by $u(z) = (P_r * f)(\theta)$, $z = re(\theta)$.*

*Proof.* Uniqueness follows from the maximum principle. For the existence, we observed before that $u(z) := (P_r * f)(\theta)$ with $|z| < 1$ is harmonic on $\mathbb{D}$. By Lemma 5.3, we have

$$\sup_{\theta \in \mathbb{T}} |u(re^{2\pi i\theta}) - f(\theta)| \to 0, \text{ as } r \to 1-.$$

This implies that we can extend $u$ continuously to $\bar{\mathbb{D}}$ by setting it equal to $f$ on $\mathbb{T}$. $\square$

## 5.2 Hardy classes of harmonic functions

Next, we wish to reverse this process and understand which classes of harmonic functions on $\mathbb{D}$ assume boundary values on $\mathbb{T}$. Moreover, we need to clarify which boundary values arise here and what we mean by "assume". Particularly important classes known as the "little" Hardy spaces are as follows:

**Definition 5.5** *For any $1 \leq p \leq \infty$ define*

$$h^p(\mathbb{D}) := \Big\{ u : \mathbb{D} \to \mathbb{C} \text{ harmonic} \;\Big|\; \sup_{0<r<1} \int_0^1 |u(re^{2\pi i\theta})|^p \, d\theta < \infty \Big\}$$

*with the norm*

$$|||u|||_p := \sup_{0<r<1} \|u(re(\cdot))\|_{L^p(\mathbb{T})}$$

By the mean value property, any positive harmonic function belongs to the space $h^1(\mathbb{D})$: then the $L^1$-norm of $u$ over any circle equals to its value in the center and is thus bounded. Among those, the most important example for us is $P_r(\theta) \in h^1(\mathbb{D})$. Observe that this function has boundary values $P_r \to \delta_0$ (the Dirac mass at $\theta = 0$) as $t \to 1-$, where the convergence is in the sense of distributions. This already shows that even for a very smooth function in $h^1(\mathbb{D})$ the boundary value may not be a function but only a distribution or a measure. In what follows, $\mathcal{M}(\mathbb{T})$ denotes the complex-valued Borel measures and $\mathcal{M}^+(\mathbb{T}) \subset \mathcal{M}(\mathbb{T})$ the positive Borel measures.

**Theorem 5.6** *There is a one-to-one correspondence between $h^1(\mathbb{D})$ and $\mathcal{M}(\mathbb{T})$ given by $\mu \in \mathcal{M}(\mathbb{T}) \mapsto F_r(\theta) := (P_r * \mu)(\theta)$. Under this map, any $\mu \in \mathcal{M}^+(\mathbb{T})$ relates uniquely to a positive harmonic function. Furthermore,*

$$\|\mu\| = \sup_{0<r<1} \|F_r\|_1 = \lim_{r \to 1} \|F_r\|_1 \tag{5.2}$$

*and the following properties hold:*

1. *The measure $\mu$ is absolutely continuous with respect to Lebesgue measure ($\mu \ll d\theta$) if and only if $\{F_r\}$ has a limit in $L^1(\mathbb{T})$ as $r \to 1$. If so, then $d\mu = f\,d\theta$ where $f$ is the $L^1$-limit of $F_r$.*

2. *The following are equivalent for $1 < p \leq \infty$:*

$$\begin{aligned}
&(i)\ \ d\mu = f d\theta \text{ with } f \in L^p(\mathbb{T})\\
&(ii)\ \{F_r\}_{0<r<1} \ \text{ is } L^p\text{- bounded}\\
&(iii)\ \{F_r\} \text{ converges in } L^p \text{ if } 1 < p < \infty,\\
&\ \text{and in weak-}* \text{ sense in } L^\infty \text{ if } p = \infty, \text{ both as } \ r \to 1.
\end{aligned}$$

3. *The following are equivalent: (i) $f$ is continuous, (ii) $F_r = P_r * f$ extends to a continuous function on $\overline{\mathbb{D}}$, (iii) $F_r$ converges uniformly as $r \to 1-$.*

This theorem identifies $h^1(\mathbb{D})$ with $\mathcal{M}(\mathbb{T})$, and $h^p(\mathbb{D})$ with $L^p(\mathbb{T})$ for $1 < p \leq \infty$. Note that there is an improvement of regularity: even if the boundary valus is a measure, the restriction of the function to any circle of radius $r < 1$ lies in $L^1(\mathbb{T})$. Moreover, $h^\infty(\mathbb{D})$ contains the subclass of harmonic functions that can be extended continuously onto $\overline{\mathbb{D}}$; this subclass is the same as $C(\mathbb{T})$. Before proving the theorem we present two simple lemmas. In what follows we use the notation $F_r(\theta) := F\big(re^{2\pi i\theta}\big)$.

**Lemma 5.7**

*(i) If $F \in C(\overline{\mathbb{D}})$ and $\triangle F = 0$ in $\mathbb{D}$, then $F_r = P_r * F_1$ for any $0 \leq r < 1$.*

*(ii) As a function of $r \in (0,1)$ the norms $\|F_r\|_p$ are non-decreasing for any $1 \leq p \leq \infty$.*

*Proof.* Part (1) is a restatement of Theorem 5.4. For (2), first note that for any $0 < s \leq r < 1$ we have

$$F_s(\theta) = (P_{s/r} * F_r)(\theta),$$

by a simple rescaling. Hausdorff-Young's inequality then implies

$$\|F_s\|_p \leq \|P_{s/r}\|_1 \|F_r\|_p = \|F_r\|_p$$

as claimed. □

**Lemma 5.8** *Let $F \in h^1(\mathbb{D})$. Then there exists a unique measure $\mu \in \mathcal{M}(\mathbb{T})$ such that $F_r = P_r * \mu$.*

*Proof.* Since the unit ball of $\mathcal{M}(\mathbb{T})$ is weak-$*$ compact, there exists a subsequence $r_j \to 1$ with $F_{r_j} \to \mu$ in weak-$*$ sense to some $\mu \in \mathcal{M}(\mathbb{T})$. We firat claim that

$$F_r = P_r * \mu. \tag{5.3}$$

To see that, note that for any $0 < r < 1$, we have

$$P_r * \mu = \lim_{j\to\infty} (P_r * F_{r_j}) = \lim_{j\to\infty} F_{rr_j} = F_r$$

by Lemma 5.7. Next, in order to show that such $\mu$ is unique, consider any $f \in C(\mathbb{T})$, then, again by Lemma 5.7, and (5.3):

$$\langle F_r, f \rangle = \langle P_r * \mu, f \rangle = \langle \mu, P_r * f \rangle \to \langle \mu, f \rangle$$

as $r \to 1$. This shows that, in the weak-$*$ sense,

$$\mu = \lim_{r \to 1} F_r \tag{5.4}$$

which implies uniqueness of $\mu$. $\square$

**Proof of Theorem 5.6.** If $\mu \in \mathcal{M}(\mathbb{T})$, then $P_r * \mu \in h^1(\mathbb{D})$ by Lemma 5.1. Conversely, given $F \in h^1(\mathbb{D})$ then by Lemma 5.8 there is a unique $\mu$ so that $F_r = P_r * \mu$. This gives the one-to-one correspondence between $h^1(\mathbb{D})$ and $\mathcal{M}(\mathbb{T})$. Moreover, (5.4) and Lemma 5.7 show that

$$\|\mu\| \leq \limsup_{r \to 1} \|F_r\|_1 = \sup_{0<r<1} \|F_r\|_1 = \lim_{r \to 1} \|F_r\|_1 \, .$$

Since one also has

$$\sup_{0<r<1} \|F_r\|_1 \leq \sup_{0<r<1} \|P_r\|_1 \|\mu\| = \|\mu\| \, ,$$

equality (5.2) follows. If $f \in L^1(\mathbb{T})$ and $d\mu = f d\theta$, then Lemma 5.3 shows that $F_r \to f$ in $L^1(\mathbb{T})$. Conversely, if $F_r \to f$ in the sense of $L^1(\mathbb{T})$, then because of (5.4) we necessarily have $d\mu = f d\theta$ which proves the first part of Theorem 5.6. The other parts are proved similarly, and we omit the details – one invokes Lemma 5.3, part (2) for $1 < p < \infty$ and Lemma 5.3 part (1) if $p = \infty$. $\square$

Let us make a remark on the Hilber transform on the circle: it is given by the convolution with the kernel $Q_r(\theta)$ which is the *harmonic conjugate* of $P_r(\theta)$. It is easy to find $Q_r(\theta)$ since

$$P_r(\theta) = \operatorname{Re}\left(\frac{1+z}{1-z}\right)$$

and therefore

$$Q_r(\theta) = \operatorname{Im}\left(\frac{1+z}{1-z}\right) = \frac{2r\sin(2\pi\theta)}{1 - 2r\cos(2\pi\theta) + r^2}$$

Observe that $\{Q_r\}_{0<r<1}$ is *not* an approximate identity, since $Q_1(\theta) = \cot(\pi\theta)$ which is not the density of a measure – it behaves like $1/(\pi\theta)$ close to $\theta = 0$. The Hilbert transform on the circle is the map which is formally defined as follows:

$$f \mapsto u_f \mapsto \widetilde{u}_f \mapsto \widetilde{u}_f\big|_{\mathbb{T}}$$

where $u_f$ denotes the harmonic extension to $\mathbb{D}$ and $\widetilde{u}_f$ its harmonic conjugate. From the preceding, $Q_1$ is the kernel of the Hilbert transform. But we will not discuss the results for the Hilbert transform disk as they are similar to those in a half space.

## 5.3 Almost everywhere convergence to the boundary data

Finally, we turn to the issue of almost everywhere convergence of $P_r * f$ to $f$ as $r \to 1$. The main idea here is to mimic the proof of the Lebesgue differentiation theorem, which says that for any $f \in L^1(\mathbb{R}^n)$ we have

$$\frac{1}{|B(x,r)|} \int_{B(x,r)} f(y) dy \to f(x) \text{ a.e. in } \mathbb{R}^n. \tag{5.5}$$

In fact, the proof we describe below gives the proof of (5.5) as well. In particular, we need as a tool the Hardy-Littlewood maximal function $Mf$, which is defined (for the torus) as follows:

$$Mf(x) = \sup_{x \in I \subset \mathbb{T}} \frac{1}{|I|} \int_I |f(y)|\, dy$$

where $I \subset \mathbb{T}$ is an (open) interval and $|I|$ is the length of $I$. It is convenient to think of $Mf(x)$ via the box kernel:

$$Mf(x) = \sup_n \left[ n \int_{-1/(2n)}^{1/(2n)} |f(y)| dy \right] = \sup_n (\chi_n * |f|)(x),$$

where $\chi_n(x) = n\chi_{[-1/2n,1/2n]}(x)$ is the box kernel.

The most basic facts concerning this (sublinear) operator are contained in the following result.

**Proposition 5.9** *The Hardy-Littlewood maximal function $M$ is bounded from $L^1$ to weak $L^1$, i.e.,*

$$|\{x \in \mathbb{T} | Mf(x) > \lambda\}| \leq \frac{C}{\lambda} \|f\|_1$$

*for all $\lambda > 0$. For any $1 < p \leq \infty$, $M$ is bounded on $L^p$.*

*Proof.* Fix some $\lambda > 0$ and any compact set $K$ such that

$$K \subset \{x \mid Mf(x) > \lambda\} \tag{5.6}$$

There exists a finite cover $\{I_j\}_{j=1}^N$ of $K$ by open arcs $I_j$ such that

$$\int_{I_j} |f(y)|\, dy > \lambda |I_j| \tag{5.7}$$

for each $j$. We now pass to a more convenient sub-cover (this is known as Wiener's covering lemma and is very much like Vitali's covering lemma). Select an arc of maximal length from $\{I_j\}$; call it $J_1$. Maximality of $I_1$ implies that any $I_j$ such that $I_j \cap J_1 \neq \emptyset$ satisfies $I_j \subset 10 \cdot J_1$ where $10 \cdot J_1$ is the arc with the same center as $J_1$ and ten times the length (if $10 \cdot J_1$ has length larger than 1, then set $10 \cdot J_1 = \mathbb{T}$). Next, remove all arcs from $\{I_j\}_{j=1}^N$ that intersect $J_1$. Let $J_2$ be one of the remaining ones with maximal length. Continuing in this fashion we obtain arcs $\{J_\ell\}_{\ell=1}^L$ which are pair-wise disjoint and so that

$$\bigcup_{j=1}^N I_j \subset \bigcup_{\ell=1}^L 10 \cdot J_\ell$$

In view of (5.6) and (5.7) therefore,

$$\mathrm{meas}(K) \le \mathrm{meas}\left(\bigcup_{\ell=1}^L 10 \cdot J_\ell\right) \le 10 \sum_{\ell=1}^L \mathrm{meas}(J_\ell) \le \frac{10}{\lambda} \sum_{\ell=1}^L \int_{J_\ell} |f(y)|\, dy \le \frac{10}{\lambda} \|f\|_1$$

as claimed. To prove the $L^p$ statement, one interpolates the weak $L^1$ bound with the trivial $L^\infty$ bound

$$\|Mf\|_\infty \le \|f\|_\infty$$

by means of Marcinkiewicz's interpolation theorem. $\square$

We now introduce a class of approximate identities which can be dominated by the box kernels. The importance of this idea is that it allows us to dominate the maximal function associated with an approximate identity by the Hardy-Littlewood maximal function, see Lemma 5.11 below.

**Definition 5.10** *Let $\Phi_n$ be an approximate identity as in Definition 5.2. We say that it is* radially bounded *if there exist functions $\Psi_n$ on $\mathbb{T}$ so that the following additional property holds (it is convenient to think of the torus as the interval $[-1/2, 1/2]$ rather than $[0, 1]$ here):*

*(A4) $|\Phi_n| \le \Psi_n$, $\Psi_n$ is even and decreasing, i.e., $\Psi_n(\phi) \le \Psi_n(\theta)$ for $0 \le \phi \le \theta \le \frac{1}{2}$, for all $n \ge 1$. Finally, we require that $\sup_n \|\Psi_n\|_1 < \infty$.*

We have the following domination lemma.

**Lemma 5.11** *If $\Phi_n$ satisfies (A4), then for any $f \in L^1(\mathbb{T})$ one has*

$$\sup_n |(\Phi_n * f)(x)| \le Mf(x) \left(\sup_n \|\Psi_n\|_1\right) \tag{5.8}$$

*for all $x \in \mathbb{T}$.*

Note that the left side of (5.8) can be thought of as the maximal function associated to the approximation of identity $\Phi_n$ in the same way as the Hardy-Littlewood maximal function is associated to the box kernel.

*Proof.* It suffices to show the following statement: let a non-negative function $K(x)$ defined on the torus $\mathbb{T} = [-1/2, 1/2]$ be even and decreasing on $[0, 1/2]$. Then for any $f \in L^1(\mathbb{T})$ we have

$$|(K * f)(x)| \le \|K\|_1 Mf(x) \tag{5.9}$$

Indeed, assume that (5.9) holds. Then

$$\sup_n |(\Phi_n * f)(x)| \le \sup_n (\Psi_n * |f|)(x) \le (\sup_n \|\Psi_n\|_1) Mf(x)$$

and the lemma follows. The idea behind (5.9) is to show that a positive, even and decreasing function $K$ can be written as an average of box kernels, i.e., for some positive measure $\mu$ we have

$$K(\phi) = \int_0^{\frac{1}{2}} \chi_{[-\theta,\theta]}(\phi)\, d\mu(\theta). \tag{5.10}$$

Let us check that

$$d\mu(\theta) = -dK(\theta) + K(1/2)\, \delta_{\theta=\frac{1}{2}}$$

is a suitable choice: the right side of (5.10) defines an even function of $\phi$ and for $0 \le \phi \le 1/2$ we have

$$\int_0^{\frac{1}{2}} \chi_{[-\theta,\theta]}(\phi)\, d\mu(\theta) = \int_0^{\frac{1}{2}} \chi_{[-\theta,\theta]}(\phi)[-dK(\theta) + K(1/2)\delta(\theta - 1/2)]$$
$$= \int_\phi^{1/2} [-dK(\theta) + K(1/2)\delta(\theta - 1/2)] = K(\phi) - K(1/2) + K(1/2) = K(\phi).$$

Note that (5.10) implies that

$$\int_0^1 K(\phi) d\phi = 2 \int_0^{1/2} K(\phi) d\phi = 2 \int_0^{1/2} \left( \int_\phi^{1/2} d\mu(\theta) \right) d\phi = 2 \int_0^{\frac{1}{2}} \theta\, d\mu(\theta).$$

Moreover, by (5.10), we have

$$|(K * f)(\phi)| = \Big| \int_0^{\frac{1}{2}} \big(\frac{1}{2\theta}\chi_{[-\theta,\theta]} * f\big)(\phi)\, 2\theta\, d\mu(\theta) \Big| \le \int_0^{\frac{1}{2}} Mf(\phi) 2\theta\, d\mu(\theta) = Mf(\phi)\|K\|_1,$$

which is (5.9). □

Finally, we can properly address the question of whether $P_r * f \to f$ in the almost everywhere sense for $f \in L^1(\mathbb{T})$. The idea is as follows: the pointwise convergence is clear from Lemma 5.3 for continuous $f$. This suggests approximating $f \in L^1$ by a sequence of continuous functions $g_n$, in the $L^1$ norm. This leads to the problem of an interchange of limits, namely $r \to 1$ and $n \to \infty$. As always in such a situation, we require some form of uniform control, which is furnished by the Hardy–Littlewood maximal function.

**Theorem 5.12** *If $\Phi_n$ satisfies (A1)–(A4), then for any $f \in L^1(\mathbb{T})$ one has $\Phi_n * f \to f$ almost everywhere as $n \to \infty$.*

*Proof.* Pick $\varepsilon > 0$ and let $g \in C(\mathbb{T})$ with $\|f - g\|_1 < \varepsilon$. By Lemma 5.3, with $h = f - g$

one has ($|\cdot|$ denotes the Lebesgue measure of a set)

$$
\begin{aligned}
&\left|\{x \in \mathbb{T} \text{ s.t. } \limsup_{n\longrightarrow\infty} |\left(\Phi_n * f\right)(x) - f(x)| > \sqrt{\varepsilon}\,\}\right| \\
&= \left|\{x \in \mathbb{T} \text{ s.t. } \limsup_{n\longrightarrow\infty} |\left(\Phi_n * h\right)(x) - h(x)| > \sqrt{\varepsilon}\,\}\right| \\
&\le \left|\{x \in \mathbb{T} \text{ s.t. } \limsup_{n\to\infty} |(\Phi_n * h)(x)| > \sqrt{\varepsilon}/2\}\right| + \left|\{x \in \mathbb{T} \text{ s.t. } |h(x)| > \sqrt{\varepsilon}/2\}\right| \\
&\le \left|\{x \in \mathbb{T} \text{ s.t. } \sup_{n} |(\Phi_n * h)(x)| > \sqrt{\varepsilon}/2\}\right| + \left|\{x \in \mathbb{T} \text{ s.t. } |h(x)| > \sqrt{\varepsilon}/2\}\right| \\
&\le \left|\{x \in \mathbb{T} \mid CMh(x) > \sqrt{\varepsilon}/2\}\right| + \left|\left[x \in \mathbb{T} \mid |h(x)| > \sqrt{\varepsilon}/2\right]\right| \\
&\le C\sqrt{\varepsilon}
\end{aligned}
$$

We used Lemma 5.11 in the next to last inequality, while in final step we used Proposition 5.9 as well as the Chebyshev inequality (recall that $\|h\|_1 < \varepsilon$). □

As a first corollary, we obtain the classical Lebesgue differentiation theorem: we have

$$\frac{1}{2\varepsilon}\int_{x-\varepsilon}^{x+\varepsilon} f(y)dy \to f(x) \text{ a.e. on } \mathbb{T}$$

if $f \in L^1(\mathbb{T})$ (this theorem holds, of course, in any dimension, not just one).

We also get the almost everywhere convergence of of the Poisson integrals $P_r * f \to f$ for any $f \in L^1(\mathbb{T})$ as $r \to 1-$. In view of Theorem 5.6 we of course would like to know whether a similar statement holds for measures instead of $L^1$ functions. It turns out that $P_r * \mu \to f$ almost everywhere where $f$ is the density of the absolutely continuous component of $\mu$ in the Lebesgue decomposition. A most important example here is $P_r$ itself! Indeed, its boundary measure is $\delta_0$ and the almost everywhere limit is identically zero. Hence, in the almost everywhere limit we lose a lot of information – the singular part of the boundary measure. An amazing fact that we will prove next, known as the F. & M. Riesz theorem, states that there is no such loss in the class $h^1(\mathbb{D}) \cap \mathcal{H}(\mathbb{D})$, that is, for analytic harmonic functions in $h^1(\mathbb{D})$. Indeed, we will see that any such function is the Poisson integral of an $L^1$ function rather than a measure.

## 5.4 The F. and M. Riesz theorem

In this section we will prove the following theorem due to F. and M. Riesz.

**Theorem 5.13** *If $\mu \in \mathcal{M}(\mathbb{T})$ satisfies*

$$\hat{\mu}(n) = \int_0^1 e^{-2\pi i n\theta} d\mu(\theta),$$

*for all $n < 0$, then $\mu$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{T}$.*

An important role in the proof of this theorem will be played by subharmonic functions. Previously, we have defined sub-harmonic functions by the inequality $\Delta\phi \ge 0$. We will first need to broaden this notion, requiring $\phi$ to be only continuous.

**Definition 5.14** *Let $\Omega$ be a domain in $\mathbb{R}^2$. We say that a function $\phi$ which takes values in $\mathbb{R} \cup \{-\infty\}$ is sub-harmonic if it is continuous and for each $z \in \Omega$ there exists $r_z$ so that for all $0 < r < r_z$ we have*

$$f(z) \leq \int_0^1 f(z + re^{2\pi i\theta})d\theta. \tag{5.11}$$

We summarize the basic properties of subharmonic functions in the following proposition.

**Proposition 5.15** *(i) Let $f$ and $g$ be subharmonic in $\Omega$, then $u(z) = \max(f(z), g(z))$ is also subharmonic in $\Omega$.*
*(ii) Let $f \in C^2(\Omega)$, then $f$ is subharmonic in $\Omega$ if and only if $\Delta f \geq 0$ in $\Omega$.*
*(iii) If $f$ is subharmonic, and $\phi$ convex and increasing then $\phi(f(z))$ is subharmonic.*
*(iv) If $F$ is analytic in $\Omega$ then $\log|F|$ and $|F|^\alpha$, with $\alpha > 0$, are subharmonic.*

The important aspect of part (iv) is that the functions $\log u$ and $u^\alpha$ (with $0 < \alpha < 1$) are concave not convex but nevertheless $\log|F|$ and $|F|^\alpha$ are sub-harmonic.

*Proof.* (i) follows immediately from (5.11) for $f$ and $g$. As for (ii), we have already shown that $\Delta f \geq 0$ implies the sub-mean value property (5.11). The converse follows from our proof of the mean value property for harmonic functions, as well: if we fix $z \in \Omega$, and set

$$\phi(r) = \int_0^1 f(z + re^{2\pi i\theta})d\theta,$$

then we have shown that

$$\phi'(r) = \frac{1}{2\pi r}\int_0^r \int_0^{2\pi} \Delta u(z + \rho e^{i\phi})\rho d\rho d\phi.$$

Therefore, if $\Delta u(z) < 0$, we would have $\phi'(r) < 0$ for small $r$ contradicting the local sub-mean value property. Property (iii) can be seen as follows; if $f$ is subharmonic and $\phi(z)$ is increasing then

$$\phi(f(z)) \leq \phi\left(\int_0^1 f(z + re^{2\pi i\theta})d\theta\right).$$

Next, if, in addition, $\phi$ is convex, we have

$$\phi\left(\int_0^1 f(z + re^{2\pi i\theta})d\theta\right) \leq \int_0^1 \phi(f(z + re^{2\pi i\theta})d\theta,$$

by Jensen's inequality. Together, the last two inequalities show that $\phi(f(z))$ is subharmonic. Finally, (iv) is shown as follows: if $F(z_0) \neq 0$ then $\log F(z)$ is analytic in disk around $z_0$, hence its real part $\log|F(z)|$ is not just subharmonic but actually harmonic at $z_0$. On the other hand, if $F(z_0) = 0$ so that $|\log F(z_0)| = -\infty$ then there is nothing to prove. Finally, we can write $|F|^\alpha(z) = \phi(\log|F|)$, where $\phi(s) = \exp(\alpha s)$ is increasing and convex. □

The maximum principle holds for the "extended" notion of a subharmoninc function as well.

**Proposition 5.16** *Let $f \in C(\bar{\Omega})$ be subharmonic and $u \in C(\bar{\Omega})$ be harmonic. If $f \leq u$ on the boundary $\partial\Omega$ then $f \leq u$ in $\Omega$.*

The proof is verbatim the same as before so we leave it to the reader.

The next proposition allows us to pass from the "local" sub-mean value property to all balls contained in the region where the function $f$ is sub-harmonic.

**Proposition 5.17** *Let $f$ be subharmonic in a domain $\Omega$ and let $\overline{B(z,r)} \subset \Omega$, then*

$$f(z) \leq \int_0^1 f(z + re^{2\pi i\theta})d\theta.$$

**Proof.** Let $n \geq 1$ and set $g_n(z) = \max(f(z), -n)$ – the function $g_n(z)$ is also sub-harmonic. Next, consider the function $u_n(z)$ which is harmonic in $B(z,r)$ and coincides with $g_n(z)$ on the circle $\partial B(z,r)$. The maximum principle and the definition of $g_n(z)$ imply that

$$f(z) \leq g_n(z) \leq u_n(z) = \int_0^1 g_n(z + \rho e^{2\pi i\theta})d\theta.$$

The monotone convergence theorem implies now that

$$f(z) \leq \int_0^1 f(z + \rho e^{2\pi i\theta})d\theta,$$

and the proof is complete. □

We now define the radial maximal function that will play crucial role in the rest of this section.

**Definition 5.18** *Let $F(z)$ be a function defined on $\mathbb{D}$, then the radial maximal function $F^* : \mathbb{T} \to \mathbb{R}$ is defined as*

$$F^*(\theta) = \sup_{0<r<1} |F(re^{2\pi i\theta})|.$$

Recall that we have already shown (see Lemma 5.11) that if $u \in h^1(\mathbb{D})$ is harmonic, and $\mu$ is its boundary measure: $u(re^{2\pi i\theta}) = P_r * \mu$, then, for all $0 < r < 1$, we have

$$|u(re^{2\pi i\theta})| \leq CM\mu(\theta),$$

that is, $u^*(\theta) \leq CM\mu(\theta)$. Here $M\mu$ is the Hardy-Littlewood maximal function of the measure $\mu$. The same result holds for sub-harmonic functions.

**Proposition 5.19** *Let $g$ be sub-harmonic in $\mathbb{D}$, $g \geq 0$, and*

$$|||g|||_1 := \sup_{0<r<1} \int_0^1 g(re^{2\pi i\theta})d\theta < +\infty. \tag{5.12}$$

*Then, (i) for all $\lambda > 0$ we have*

$$|\theta \in \mathbb{T} : \ g^*(\theta) > \lambda| \leq \frac{C}{\lambda}|||g|||_1, \tag{5.13}$$

*(ii) if for some $1 < p \le \infty$ we have*

$$|||g|||_p := \left(\sup_{0<r<1} \int_0^1 g(re^{2\pi i\theta}) d\theta\right)^{1/p} < +\infty,$$

*then $\|g^*\|_{L^p(\mathbb{T})} \le C_p |||g|||_p$.*

*Proof.* (i) The uniform bound (5.12) implies that there exists a sequence $r_n \to 1$ so that $g_{r_n}$ weak-* converges to a measure $\mu \in \mathcal{M}(\mathbb{T})$, and

$$\|\mu\| \le |||g|||_1. \tag{5.14}$$

Moreover, for ever $0 < s < 1$ we have

$$g_s(\theta) = \lim_{n\to+\infty} g_{r_n s}(\theta),$$

while (as $g$ is sub-harmonic)

$$g_{r_n s}(\theta) \le P_s * g_{r_n} \to P_s * \mu.$$

We conclude that $g_s \le P_s * \mu$. Now, Lemma 5.11 shows that

$$g_s(\theta) \le CM\mu(\theta), \tag{5.15}$$

and (5.13) follows from the weak $L^1$-bound for the maximal function in Proposition 5.9, as well as (5.14).

(ii) If, in addition, $|||g|||_p < +\infty$, for some $1 < p \le +\infty$, then $\mu$ has a density $f \in L^p(\mathbb{T})$: $d\mu = f d\theta$, and $\|f\|_{L^p(\mathbb{T})} \le |||g|||_p$. Then the $L^p$-bound on the maximal function in Proposition 5.9, together with (5.15) implies that

$$\|g^*\|_{L^p(\mathbb{T})} \le C\|f\|_{L^p(\mathbb{T})} \le C|||g|||_p,$$

completing the proof. □

This gives us the following first version of F. and M. Riesz theorem (the result is not true without the analyticity assumption).

**Proposition 5.20** *Let $F \in h^1(\mathbb{D})$ be analytic, then $F^* \in L^1(\mathbb{T})$.*

*Proof.* As $F$ is analytic, it follows from part (iv) of Proposition 5.15 that $|F|^{1/2}$ is a subharmonic function. As $F \in h^1(\mathbb{D})$, we have $|||F^{1/2}|||_2 < +\infty$, from which we conclude, using Proposition 5.19 that $(|F|^{1/2})^* \in L^2(\mathbb{T})$. But we have, obviously, $(|F|^{1/2})^* = (F^*)^{1/2}$, hence $F^* \in L^1(\mathbb{T})$. □

Next, recall that any function $F \in h^1(\mathbb{D})$ has the form $F_r = P_r * \mu$, where $\mu$ is a measure on $\mathbb{T}$. It has the Lebesgue decomposition $\mu = \mu_{ac} + \mu_s$, where $\mu_{ac}$ is absolutely continuous with respect to the Lebesgue measure $d\theta$, and $\mu_s$ and $d\theta$ are mutually singular. The measure $\mu_{ac}$ has the form $f(\theta)d\theta$ with $f \in L^1(\mathbb{T})$. A useful exercise is to show that $P_r * \mu \to f$ a.e., or, equivalently, $P_r * \mu_s \to 0$ a.e. In other words, $F(re^{2\pi i\theta})$ has a pointwise limit as $r \to 1$ for a.e. $\theta \in \mathbb{T}$. The reason is that if $\mu_s$ is singular with respect to the Lebesgue measure, then for a.e. $\theta \in \mathbb{T}$ we have

$$\frac{1}{2\varepsilon}\mu_s([\theta - \varepsilon, \theta + \varepsilon]) \to 0 \text{ as } \varepsilon \to 0.$$

In particular, as we have already mentioned, the Poisson kernel itself satisfies $P_r(\theta) \to 0$ as $r \to 1$ a.e. The next result shows that "for analytic functions this can not happen".

**Proposition 5.21** *Assume $F \in h^1(\mathbb{D})$ and $F$ is analytic, and let $f(\theta) = \lim_{r\to 1} F(re^{2\pi i\theta})$, then $F_r = P_r * f$ for all $0 < r < 1$.*

*Proof.* We have $F_r \to f$ for a.e. $\theta \in \mathbb{T}$, and $|F_r| \le F^* \in L^1(\mathbb{T})$. The Lebesgue dominated convergence theorem implies that $F_r \to f$ in $L^1(\mathbb{T})$. Now, Theorem 5.6 (part (1)) implies that $F_r = P_r * f$. □

Finally, we can prove Theorem 5.13: if a measure $\mu$ on the torus satisfies $\hat{\mu}(n) = 0$ for all $n < 0$, then $\mu$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{T}$.

*Proof.* (Of Theorem 5.13). Assume that

$$\hat{\mu}(n) = 0 \text{ for all } n < 0, \tag{5.16}$$

and set

$$F(re^{2\pi i\theta}) = \sum_{n=0}^{\infty} r^n \hat{\mu}(n) e^{2\pi n i\theta} = \sum_{n\in\mathbb{Z}} r^{|n|} \hat{\mu}(n) e^{2\pi i n\theta} = P_r * \mu(\theta).$$

Note, that $|\hat{\mu}(n)| \le \|\mu\|$, so the above definition makes sense for all $r$ because of (5.16), and the function $F_r(\theta)$ is analytic. Proposition 5.21 implies that $F_r$ has an $L^1$-limit $f$ as $r \to 1$, and $d\mu = f d\theta$. □

The last theorem by Riesz brothers is as follows.

**Theorem 5.22** *Let $F$ by analytic in $\mathbb{D}$ and $L^1$-bounded, that is, $F \in h^1(\mathbb{D})$. Assume that $F \not\equiv 0$ and set $f = \lim_{r\to 1} F_r$, then $f$ can not vanish on a set of positive measure.*

*Proof.* The idea is to show that $\log|f| \in L^1(\mathbb{T})$, so that, in particular, $\log|f|$ is finite a.e. First, note that $\log_+|f| \le |f| \in L^1(\mathbb{T})$ by Proposition 5.20. Furthermore, if $F(0) \ne 0$, then, as $\log|F|$ is sub-harmonic, we

$$-\infty < -\log|F(0)| \le \int_{\mathbb{T}} \log|F_r(\theta)| d\theta,$$

for any $0 < r < 1$. However, Fatou's lemma shows that the last inequality implies

$$-\infty < -\log|F(0)| \le \int_{\mathbb{T}} \log|f(\theta)| d\theta, \tag{5.17}$$

after passing to $r \to 1$ and recalling that $F_r(\theta) \to f(\theta)$ a.e., and $\log_+|f| \in L^1(\mathbb{T})$. Finally, if $F(0) = 0$ and there exists a point $z_0 \in \mathbb{D}$ such that $F(z_0) \ne 0$, we simply consider an automorphism of the unit disk to itself that maps $z_0 \to 0$ and repeat the previous argument. □

# 6 The Fourier transform and holomorphic functions

We now turn to further connections between the analytic functions and the Fourier transform. We will be following the Stein-Shakarchi book.

## 6.1 The Fourier transform of moderately decaying functions

The Fourier transform of a function $f(x)$ defined on $\mathbb{R}$ is

$$\hat{f}(\xi) = \int f(x) e^{-2\pi i x \xi} dx, \tag{6.1}$$

defined for $\xi \in \mathbb{R}$. A convenient class of functions that relate the Fourier and complex analyses are as follows: given $a > 0$ we denote by $\mathcal{F}_a$ the class of all functions that satisfy two conditions: (i) $f$ is holomorphic in the strip

$$S_a = \{z \in \mathbb{C} : \ |Im(z)| < a\},$$

and (ii) there exists a constant $A > 0$ (that depends on $f$) so that

$$|f(x+iy)| \leq \frac{A}{1+x^2}, \text{ for all } x \in \mathbb{R} \text{ and } |y| < a.$$

For example, the Gaussian $f(z) = e^{-\pi z^2}$ belongs to $\mathcal{F}_a$ for all $a > 0$. On the other hand, the function

$$f(z) = \frac{1}{z^2 + c^2},$$

with $c > 0$ belongs to $\mathcal{F}_a$ only for $0 < a < c$. We will also denote by $\mathcal{F}$ the class of functions that belong to $\mathcal{F}_a$ for some $a > 0$.

The first result relates the exponential decay of the Fourier transform and the moderate decay of the function itself.

**Theorem 6.1** *Let $f \in \mathcal{F}_a$ for some $a > 0$, then for any $0 \leq b < a$ there exists $B$ so that $|\hat{f}(\xi)| \leq Be^{-2\pi b|\xi|}$.*

*Proof.* The case $b = 0$ is simple: if $f \in \mathcal{F}_a$ for some $a > 0$ then its restriction to the real axis satisfies $f \in L^1(\mathbb{R})$, which means that $|\hat{f}(\xi)| \leq \|f\|_1$.

When $0 < b < a$ let us first assume that $\xi > 0$ and consider the function $g(z) = f(z)e^{-2\pi i \xi z}$ (with $\xi$ fixed) and integrate it over the boundary of the rectangle that is formed by the points $(-R, 0)$, $(R, 0)$, $(R, -b)$, and $(-R, -b)$ (connected in that order). The integral over the vertical sides goes to zero as $R \to +\infty$ because of the uniform decay of $f$:

$$\left| \int_{-b}^{0} f(-R+iy) e^{-2\pi i \xi(-R+iy)} dy \right| \leq \frac{bA}{1+R^2} \to 0 \text{ as } R \to +\infty,$$

with a similar estimate over the interval $[R-ib, R]$. The Cauchy theorem implies, therefore, that the integrals over the two horizontal sides are equal in the limit $R \to +\infty$:

$$\int_{-\infty}^{\infty} f(x)e^{-2\pi i\xi x}dx = \int_{-\infty}^{\infty} f(x-ib)e^{-2\pi i\xi(x-ib)}dx,$$

or

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x-ib)e^{-2\pi i\xi(x-ib)}dx \tag{6.2}$$

This expression (interesting in itself) leads to the estimate

$$|\hat{f}(\xi)| \le \int_{-\infty}^{\infty} \frac{A}{1+x^2}e^{-2\pi\xi b}dx \le Be^{-2\pi b\xi},$$

as claimed. The proof for $\xi < 0$ is identical except the real line is shifted up by $b$. □

The above theorem relates the decay of the Fourier transform to the possibility of extending $f$ as an analytic function in a wide strip. The ultimate step in this direction will be asking for which functions we may have $\hat{f}$ have compact support (this is "ultimate" decay), and this is what we will soon study.

First, we establish the Fourier inversion formula using the complex analysis tools.

**Theorem 6.2** *If $f \in \mathcal{F}$ then the Fourier inversion formula holds:*

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi)e^{2\pi i x\xi}d\xi. \tag{6.3}$$

*Proof.* As $f \in \mathcal{F}$, the Fourier transform $\hat{f}(\xi)$ is exponentially decaying by Theorem 6.1, so the right side of (6.3) makes sense. As $f \in \mathcal{F}$, we can find $a > 0$ so that $f \in \mathcal{F}_a$ and choose $b \in (0, a)$. For $\xi > 0$ we will then use expression (6.2):

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x-ib)e^{-2\pi i\xi(x-ib)}dx, \tag{6.4}$$

so that the integral as in (6.3) but over $\xi > 0$ can be written as

$$\begin{aligned}
&\int_0^{\infty} \hat{f}(\xi)e^{2\pi i x\xi}d\xi = \int_0^{\infty} d\xi \int_{-\infty}^{\infty} ds f(s-ib)e^{-2\pi i\xi(s-ib)}e^{2\pi i x\xi} \\
&= \int_{-\infty}^{\infty} ds f(s-ib) \int_0^{\infty} d\xi e^{-2\pi i\xi(s-x-ib)} = \int_{-\infty}^{\infty} f(s-ib)\frac{1}{2\pi i(s-x-ib)}ds \\
&= \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{f(s-ib)}{s-ib-x}ds = \frac{1}{2\pi i}\int_{L_1} \frac{f(\zeta)d\zeta}{\zeta - x}.
\end{aligned}$$

Here $L_1$ is the horizontal line $\{y = -ib\}$, oriented from the left to the right. A similar computation shows that for $\xi < 0$ we have

$$\int_{-\infty}^{0} \hat{f}(\xi)e^{2\pi i x\xi}d\xi = \frac{1}{2\pi i}\int_{L} \frac{f(\zeta)d\zeta}{\zeta - x},$$

where $L_2$ is the horizontal line $y = ib$ oriented from the right to the left. Let us now consider the rectangular contour $\Gamma_R$ that connects counterclockwise the vertices $-R+ib$, $-R-ib$, $R-ib$ and $R+ib$. The Cauchy theorem implies that

$$f(x) = \frac{1}{2\pi i}\int_{\Gamma_R}\frac{f(\zeta)d\zeta}{\zeta - x}. \tag{6.5}$$

On the other hand, as in the proof of Theorem 6.1, the integral over the vertical sides vanishes as $R \to +\infty$, and the integral over the horizontal lines becomes the sum of the integrals over $L_1$ and $L_2$, which gives, passing to $R \to +\infty$ in (6.5):

$$f(x) = \frac{1}{2\pi i}\int_{L_1}\frac{f(\zeta)d\zeta}{\zeta - x} + \frac{1}{2\pi i}\int_{L_2}\frac{f(\zeta)d\zeta}{\zeta - x} = \int_0^\infty \hat f(\xi)e^{2\pi i x\xi}d\xi + \int_{-\infty}^0 \hat f(\xi)e^{2\pi i x\xi}d\xi$$
$$= \int_{-\infty}^\infty \hat f(\xi)e^{2\pi i x\xi}d\xi,$$

and the proof is complete. □

## 6.2 The Paley-Wiener theorem

We will need to use the inversion formula for the Fourier transform under slightly different conditions. We say that $f$ is of moderate decrease if $f$ and $\hat f$ satisfy

$$|f(x)| \le \frac{A}{1+|x|^2}, \quad |\hat f(\xi)| \le \frac{B}{1+|\xi|^2}. \tag{6.6}$$

**Theorem 6.3** *Assume that $f$ and $g$ are of moderate decrease and continuous, then*

$$\int_{\mathbb{R}^n} f(x)\hat g(x)dx = \int_{\mathbb{R}^n}\hat f(x)g(x)dx, \tag{6.7}$$

*and*

$$f(x) = \int \hat f(\xi)e^{2\pi i x\xi}d\xi. \tag{6.8}$$

**Proof.** We begin with a lemma that is one of the cornerstones of the probability theory.

**Lemma 6.4** *Let $f(x) = e^{-\pi|x|^2}$, then $\hat f(x) = f(x)$.*

**Proof.** The proof is a glimpse of how useful the Fourier transform is for differential equations and vice versa: the function $f(x)$ satisfies an ordinary differential equation

$$u' + 2xu = 0, \tag{6.9}$$

with the boundary condition $u(0) = 1$. However, (6.9), together with the formula for the Fourier transform of the derivative $f'$:

$$\hat{f'}(\xi) = 2\pi i\xi\hat f(\xi),$$

implies that $\hat{f}$ satisfies the same differential equation (6.9), with the same boundary condition $\hat{f}(0) = 0$. It follows that $f(x) = \hat{f}(x)$ for all $x \in \mathbb{R}$. $\Box$

We continue with the proof of Theorem 6.3. The Parceval identity can be verified directly using Fubini's theorem:

$$\int f(x)\hat{g}(x)dx = \int f(x)g(\xi)e^{-2\pi i\xi\cdot x}dxd\xi = \int \hat{f}(\xi)g(\xi)d\xi,$$

as $f$, $\hat{f}$, $g$ and $\hat{g}$ are all integrable. Finally, we prove the inversion formula using a rescaling argument. For any $\lambda > 0$ we have

$$\int_{\mathbb{R}^n} f(x)\hat{g}(\lambda x)dx = \int_{\mathbb{R}^{2n}} f(x)g(\xi)e^{-2\pi i\lambda\xi\cdot x}dx = \int \hat{f}(\lambda\xi)g(\xi)d\xi = \frac{1}{\lambda^n}\int_{\mathbb{R}^n} \hat{f}(\xi)g\left(\frac{\xi}{\lambda}\right)d\xi.$$

Multiplying by $\lambda$ and changing variables on the left side we obtain

$$\int f\left(\frac{x}{\lambda}\right)\hat{g}(x)dx = \int \hat{f}(\xi)g\left(\frac{\xi}{\lambda}\right)d\xi.$$

Letting now $\lambda \to \infty$ using the Lebesgue dominated convergence theorem gives

$$f(0)\int \hat{g}(x)dx = g(0)\int \hat{f}(\xi)d\xi, \tag{6.10}$$

for all continuous functions $f$ and $g$ of moderate decrease. Taking $g(x) = e^{-\pi|x|^2}$ in (6.10) and using Lemma 6.4 leads to

$$f(0) = \int f(\xi)d\xi. \tag{6.11}$$

The inversion formula (6.8) now follows if we apply (6.11) to a shifted function $f_y(x) = f(x+y)$, because

$$\hat{f}_y(\xi) = \int f(x+y)e^{-2\pi i\xi\cdot x}dx = e^{2\pi i\xi\cdot y}\hat{f}(\xi),$$

so that

$$f(y) = f_y(0) = \int \hat{f}_y(\xi)d\xi = \int e^{2\pi i\xi\cdot y}\hat{f}(\xi)d\xi,$$

which is (6.8). $\Box$

Theorem 6.1 has the following "partial converse".

**Theorem 6.5** *Assume that $\hat{f}(\xi)$ satisfies $|\hat{f}(\xi)| \le Ae^{-2\pi a|\xi|}$ for some $a, A > 0$. Then $f(x)$ is the restriction to the real axis of a function holomorphic in the strip $S_a = \{|Imz| < a\}$.*

*Proof.* The function $f$ can be extended to any sub-strip $S_b$, $0 < b < a$, as

$$f(z) = \int_{-\infty}^{\infty} \hat{f}(\xi)e^{2\pi i\xi z}d\xi.$$

The integral converges absolutely for any $z \in S_a$:

$$|f(z)| \leq \int_{-\infty}^{\infty} |\hat{f}(\xi)| e^{2\pi|\xi||Imz|} d\xi \leq A \int_{-\infty}^{\infty} e^{-2\pi(a-|Imz|)|\xi|} d\xi.$$

In addition, the functions

$$f_n(z) = \int_{-n}^{n} \hat{f}(\xi) e^{2\pi i\xi z} d\xi$$

are entire simply because they can be differentiated. Finally, for any $z \in S_a$ we have

$$|f(z) - f_n(z)| \leq A \int_{|\xi|>n} e^{-2\pi(a-|Imz|)|\xi|} d\xi \to 0,$$

as $n \to +\infty$, uniformly in any sub-strip $\|\,|Imz| < b\}$ for any $b \in (0, a)$. It follows that $f_n(z)$ converges uniformly to $f(z)$ in any such sub-strip, hence, as $f_n(z)$ are holomorphic,. so is $f(z)$. □

**Corollary 6.6** *If $|\hat{f}(\xi)| \leq Ae^{-2\pi a|\xi|}$ for some $a > 0$, and $f$ vanishes on a non-empty open interval then $f \equiv 0$.*

*Proof.* Any such $f$ is a restriction of a holomorphic function, whence the uniqueness theorem implies the claim. □

As a particularly important example, it follows that it is impossible that both $f$ and $\hat{f}$ are compactly supported. Hence, it is natural to ask for which functions it is possible to have compactly supported Fourier transform.

**Theorem 6.7** *Assume that $f$ is continuous and has moderate decrease. Then $\hat{f}$ is supported inside an interval $-M \leq \xi \leq M$ if and only if $f$ is a restriction to the real line of an entire function that satisfies $|f(z)| \leq Ae^{2\pi M|z|}$ for some $A > 0$.*

*Proof.* First, if $\hat{f}$ is supported in $[-M, M]$ then both $f$ and $\hat{f}$ have moderate decrease, thus

$$f(x) = \int_{-M}^{M} \hat{f}(\xi) e^{2\pi i\xi x} d\xi.$$

Therefore, $f$ can be extended to the whole complex plane by setting

$$f(z) = \int_{-M}^{M} \hat{f}(\xi) e^{2\pi i\xi z} d\xi.$$

This function is entire and

$$|f(z)| \leq \int_{-M}^{M} |\hat{f}(\xi)| e^{2\pi M|Imz|} d\xi \leq Ce^{2\pi M|z|}.$$

The other direction is much less trivial. We will assume that $f$ is an entire function, and make progressively weaker assumptions on $f$, eventually getting to $|f(z)| \leq Ae^{2\pi M|z|}$, and show that each one guarantees that $\hat{f}(\xi) = 0$ for $|\xi| \geq M$.

First, assume that

$$|f(x+iy)| \le \frac{Ae^{2\pi M|y|}}{1+x^2}. \tag{6.12}$$

Let $\xi > M$, then, using (6.4) we get, for any $y > 0$:

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i\xi x}dx = \int_{-\infty}^{\infty} f(x-iy)e^{-2\pi i\xi(x-iy)}dx.$$

It follows that

$$|\hat{f}(\xi)| \le A\int_{-\infty}^{\infty} \frac{e^{2\pi My-2\pi\xi y}}{1+x^2}dx \le Ce^{2\pi My-2\pi\xi y}.$$

As $\xi > M$, the right side above can be made arbitrarily small by taking $y$ large. We conclude that $\hat{f}(\xi) = 0$. A nearly identical argument proves this for $\xi < -M$.

Now, instead of (6.12) assume that

$$|f(x+iy)| \le Ae^{2\pi M|y|}, \tag{6.13}$$

which is still stronger than $|f(z)| \le Ae^{2\pi M|z|}$, which is what we ultimately need. However, assumption (6.13) can be reduced to (6.12) as follows. Take $\xi > M$, and set

$$f_\varepsilon(z) = \frac{f(z)}{(1+i\varepsilon z)^2},$$

with $\varepsilon > 0$ small. The function $f_\varepsilon(z)$ satisfies

$$|f_\varepsilon(x+iy)| = \frac{|f(z)|}{(1-\varepsilon y)^2+\varepsilon^2x^2} \le \frac{A_\varepsilon e^{2\pi M|z|}}{1+x^2}, \text{ for } y \le 0, \tag{6.14}$$

with $A_\varepsilon = A/\varepsilon^2$. That is, $f_\varepsilon(z)$ satisfies (6.12) in the lower half-plane. Note that

$$|\hat{f}_\varepsilon(\xi) - \hat{f}(\xi)| \le \int_{-\infty}^{\infty} |f_\varepsilon(x) - f(x)|dx \le \int_{-\infty}^{\infty} |f(x)| \left|\frac{1}{(1+i\varepsilon x)^2} - 1\right| dx \to 0,$$

as $\varepsilon \to 0$ by the Lebesgue dominated convergence theorem, as $f$ is of moderate decrease and $|1+i\varepsilon x| \ge 1$ for all $x \in \mathbb{R}$. Observe that in showing that (6.12) implies $\hat{f}(\xi) = 0$ for $\xi > M$ we only used (6.12) in the lower half plane. Therefore, as $f_\varepsilon(z)$ satisfies (6.14), we conclude that $\hat{f}_\varepsilon(\xi) = 0$ for $\xi > M$ and all $\varepsilon \in (0,1)$. It follows that $\hat{f}(\xi) = 0$ also. The argument for $\xi < 0$ is similar except the factor $1/(1+i\varepsilon z)^2$ in the definition of $f_\varepsilon$ should be replaced by $1/(1-i\varepsilon z)^2$.

The last step is to show that condition (6.12) holds under the assumptions of the theorem. More precisely, we will show that if $|f(x)| \le 1$ for all $x \in \mathbb{R}$ and $|f(z)| \le e^{2\pi M|z|}$ for all $z \in \mathbb{C}$, then

$$|f(x+iy)| \le e^{2\pi M|y|}. \tag{6.15}$$

This will follow from the following lemma (Phragmén-Lindelöf theorem), which is another version of the three lines theorem that we have seen before.

**Lemma 6.8** *Let $F$ be a function holomorphic in the sector*

$$S = \{z : \ -\pi/4 < arg\ z < \pi/4\},$$

*and continuous up to the boundary of $S$. Assume that $|F(z)| \leq 1$ on the boundary of $S$, and that $|F(z)| \leq Ae^{c|z|}$ for all $z \in S$, with some constants $A, c > 0$. Then $|F(z)| \leq 1$ for all $z \in S$.*

Note that some restriction on the growth of $F(z)$ is necessary – for example, the function $F(z) = e^{z^2}$ satisfies

$$F(x \pm ix) = e^{\pm 2i|x|^2},$$

but is unbounded on the real axis.

Let us first finish the proof of the Paley-Wiener theorem, and then return to the proof of Lemma 6.8. We need to show that the conditions $|f(x)| \leq 1$ for all $x \in \mathbb{R}$ and

$$|f(x+iy)| \leq e^{2\pi M|x+iy|}$$

imply that

$$|f(x+iy)| \leq e^{2\pi M|y|}.$$

Of course, the Phragmén-Lindelöf principle applies to any quadrant, not just the one specified in Lemma 6.8. We will apply it to the first quadrant $Q = \{x > 0,\ y > 0\}$. On its boundary we have $|f(x)| \leq 1$ and $|f(iy)| \leq e^{2\pi M|y|}$. In order to remove the growth on the $y$-axis we set

$$F(z) = f(z)e^{2\pi iMz}.$$

Then we have on the boundary of $Q$: $|F(x)| = |f(x)| \leq 1$, as well as

$$|F(iy)| = |f(iy)|e^{-2\pi My} \leq 1, \text{ for any } y > 0.$$

We also have the growth condition inside $Q$:

$$|F(z)| = |f(x+iy)|e^{-2\pi My} \leq e^{2\pi M|z|}.$$

We conclude from Lemma 6.8 that $|F(z)| \leq 1$ everywhere in $Q$, which means

$$|f(x+iy)| \leq \left|e^{-2\pi iM(x+iy)}\right| = e^{2\pi My}.$$

The argument for the other three quadrants is very similar. This finishes the proof of the Paley-Wiener theorem except for the proof of Lemma 6.8. □

*Proof.* (Of Lemma 6.8). The proof is very similar to that of the three lines theorem. Consider the function

$$F_\varepsilon(z) = F(z)e^{-\varepsilon z^{3/2}}.$$

Note that the function $z^{3/2}$ is holomorphic in $S$: it is given by (for $z = re^{i\theta}$, $-\pi < \theta < \pi$)

$$z^{3/2} = r^{3/2}e^{3i\theta/2}.$$

The exponential is bounded by

$$\left|e^{-\varepsilon z^{3/2}}\right| = e^{-\varepsilon r^{3/2}\cos(3\theta/2)}.$$

The key point is that if $-\pi/4 < \theta < \pi/4$ then

$$-\frac{\pi}{2} < -\frac{3\pi}{8} < \frac{3\theta}{2} < \frac{3\pi}{8} < \frac{\pi}{2},$$

which means that $\cos(3\theta)/2 > \alpha_0 > 0$ in this quadrant, and $F_\varepsilon(z)$ decays as $|z| \to +\infty$ at least as $e^{-\varepsilon\alpha_0|z|^{3/2}}$. As a consequence, $F_\varepsilon(z)$ is bounded in $S$, as the growth of $F(z)$ at infinity is at most $e^{c|z|}$. Moreover, as $|F_\varepsilon(z)| \le |F(z)|$, we have $|F(z)| \le 1$ on the boundary of $S$. Now, as in the proof of the three lines theorem we may conclude that $|F_\varepsilon(z)| \le 1$: take $R$ sufficiently large so that $|F_\varepsilon(z)| \le 1$ for all $z \in S$ with $|z| > R$. Then on the boundary of the region $S_R = \{z \in S : \ |z| \le R\}$ we have $|F_\varepsilon(z)| \le 1$. The maximum principle implies that $|F_\varepsilon(z)| \le 1$ in all of $S_R$. Therefore, $F(z)$ satisfies

$$|F(z)| = |F_\varepsilon(z)e^{\varepsilon z^{3/2}}| \le e^{\varepsilon|z|^{3/2}},$$

and letting $\varepsilon \to 0$ we conclude that $|F(z)| \le 1$ for all $z \in S$. □

## 6.3 The du Bois-Reymond example

We will now give an example of a continuous periodic function whose Fourier series diverges at the point $x = 0$ (of course we can move this point to an arbitrary point on the circle) (this material is taken from M. Pinsky's book). We will look for $f(x)$ in the form

$$f(x) = \sum_{k=1}^{\infty} e^{2\pi i N_k x}\frac{B_k(x)}{k^2}. \tag{6.16}$$

The coefficients $B_k(x)$ themselves will have the form

$$B_k(x) = \sum_{j=-m_k}^{m_k} a_j e^{2\pi i j x}.$$

The main point here is to choose appropriately the integers $N_k$ and $m_k$, as well as the coefficients $a_j$. To be very concrete we will take $a_k$ to be the Fourier coefficients of the function $u(x) = 1/2 - x$, $0 < x < 1$, extended periodically (which leads to a discontinuous function at $x = 0$):

$$a_k = \int_0^1 (\frac{1}{2} - x)e^{-2\pi i k x}dx = \frac{xe^{-2\pi i k x}}{2\pi i k}\Big|_{x=0}^{x=1} = \frac{1}{2\pi i k}.$$

Note that the partial sums

$$\sum_{k=-N}^{N} a_k e^{2\pi i k x}$$

are uniformly bounded in $x$ for all $N$. This is seen as follows:

$$\sum_{k=-N}^{N} a_k e^{2\pi ikx} = \int_0^1 (\frac{1}{2}-y)\frac{\sin[(2N+1)\pi(x-y)]}{\sin[\pi(x-y)]}dy$$
$$= \int_0^1 u(x-y)\frac{\sin[(2N+1)\pi y)]}{\sin(\pi y)}dy \sim \int_0^1 u(x-y)\frac{\sin[(2N+1)\pi y)]}{\pi y}dy$$
$$= \int_0^x (\frac{1}{2}-x+y)\frac{\sin[(2N+1)\pi y)]}{\pi y}dy + \int_x^1 (y-x-\frac{1}{2})\frac{\sin[(2N+1)\pi y)]}{\pi y}dy \leq M,$$

as there exists a constant $C$ such that for any $(a,b)$ we have

$$\left|\int_a^b \frac{\sin y}{y}dy\right| \leq C.$$

Therefore, $|B_k(x)| \leq M$, and the series (6.16) converges uniformly, hence the sum $f(x)$ is a continuous function. Let us choose $N_k$ and $m_k$ so that

$$N_{k+1} - m_{k+1} > N_k + m_k.$$

This implies that the frequencies coming from the term $e^{iN_kx}B_k(x)$ and those coming from $e^{iN_{k+1}x}B_{k+1}(x)$ do not overlap. This allows us to compute the Fourier coefficients of $f$:

$$f_n = \int_0^1 f(x)e^{-2\pi inx}dx = \sum_{k=1}^{\infty}\frac{1}{k^2}\int_0^1 B_k(x)e^{2\pi iN_kx}e^{-2\pi inx}dx.$$

As all $B_k(x)$ involve different frequencies, all integrals in the right side above will vanish except possibly for one $k$ (if it exists) that satisfies $|n-N_k| \leq m_k$. That is, if there exists $k$ such that $N_k - m_k < n < N_k + m_k$, then

$$f_n = \frac{a_{n-N_k}}{k^2},$$

and otherwise $f_n = 0$. Consider then the partial sum $S_{N_k}f$ of the Fourier series for $f$:

$$S_{N_k}f(0) = \sum_{n=-N_k}^{N_k} f_n = \sum_{j=1}^{k-1}\sum_{n=N_j-m_j}^{N_j+m_j}\frac{a_{n-N_j}}{j^2} + \sum_{n=N_k-m_k}^{N_k}\frac{a_{n-N_k}}{k^2}$$
$$= \sum_{j=1}^{k-1}\frac{1}{j^2}\sum_{n=-m_j}^{m_j} a_n + \frac{1}{k^2}\sum_{j=1}^{m_k} a_{-j} = \sum_{j=1}^{k-1}\frac{B_j(0)}{j^2} + \frac{1}{k^2}\sum_{j=1}^{m_k} a_{-j}.$$

The first term in the right side above converges as $k \to +\infty$, since $|B_j(0)| \leq M$. However, the second term satisfies, from our choice of $a_j$ a lower bound:

$$\frac{1}{k^2}\left|\sum_{j=1}^{m_k} a_{-j}\right| \geq \frac{C\log m_k}{k^2}.$$

Therefore, in order for the Fourier series of the function $f$ to diverge at $x=0$ we may choose $m_k$ so that $(\log m_k)/k^2 \to +\infty$, for instance, we may take $m_k = 2^{k^4}$. After choosing $m_k$ we choose $N_k$ so that $N_{k+1} - m_{k+1} > N_k + m_k$. This completes the proof of the du Bois-Raymond example.

# 7 Entire functions

This material is also taken from the Stein-Shakrachi book.

## 7.1 Counting zeros of an entire function

The fundamental theorem of algebra shows that there is a link between the growth of a polynomial and the number of zeros it can have: a polynomial of order $n$ has $n$ zeros, – in particular, it can not have more than $n$ zeros, – growth restricts the possible number of zeros. A convenient tool to count the zeros is given by Jensen's formula. Let $f$ be analytic in a disk $D_R = \{|z| \leq R\}$, and let $n(r)$ be the number of zeros $f$ has in the disk $D_r$, $0 < r < R$ (with multiplicities). Our goal is to relate $n(R)$ to the values of $f$ on the circle $\{|z| = R\}$. We assume in this section that $f(0) \neq 0$, and $f$ has no zeros on the circle $\{|z| = R\}$ . Let us denote all zeros in $D_R$ by $z_1, \dots, z_N$ and make the following observation:

$$\int_0^R n(r)\frac{dr}{r} = \sum_{k=1}^N \log\left|\frac{R}{z_k}\right|. \tag{7.1}$$

To see that (7.1) holds, note that the right side can be written as

$$\sum_{k=1}^N \log\left|\frac{R}{z_k}\right| = \sum_{k=1}^N \int_{|z_k|}^R \frac{dr}{r},$$

and the right side above equals to the left side of (7.1) – it is important that $f(0) \neq 0$, so that $n(r) = 0$ for $r$ sufficiently small.

The next step is to re-write the right side of (7.1) as (this is known as Jensen's formula)

$$\sum_{k=1}^N \log\left(\frac{|z_k|}{R}\right) = \log|f(0)| - \int_0^1 \log|f(Re^{2\pi i\theta})|d\theta. \tag{7.2}$$

This is verified as follows. First, if $f(z)$ has no zeros in $D_R$ then $\log f(z)$ is a holomorphic function, and $\log|f(z)|$ is its real part, hence harmonic. Therefore, (7.2) holds for such $f$ by the mean-value property of harmonic functions. In the general case, we may write $f(z)$ as

$$f(z) = (z - z_1)(z - z_2)\dots(z - z_N)g(z),$$

with a holomorphic function $g(z)$ that never vanishes in $D_R$ (and for which (7.2) holds by the above argument). As $\log|z_1 z_2| = \log|z_1| + \log|z_2|$, it follows that we only need to establish (7.2) for linear functions $f(z) = z - w$, with $|w| < R$. Then, (7.2) has the form

$$\log\left(\frac{|w|}{R}\right) = \log|w| - \int_0^1 \log|Re^{2\pi i\theta} - w|d\theta, \tag{7.3}$$

which is equivalent to

$$\int_0^1 \log|e^{2\pi i\theta} - a|d\theta = 0, \tag{7.4}$$

for all $a$ with $|a| < 1$. This may, in turn, be rewritten as

$$\int_0^1 \log |ae^{2\pi i\theta} - 1| d\theta = 0, \quad \text{for all } a \in \{|z| < 1\}. \tag{7.5}$$

The function $p(z) = 1 - az$ does not vanish in the unit disk, hence its logarithm is an analytic function in $\{|z| < 1\}$ hence the left side of (7.7) equals to $\log |p(0)| = 0$, whence (7.2) holds.

Together, (7.1) and (7.2) imply that if $f(0) \neq 0$, and $f$ has no zeros on the circle $\{|z| = R\}$, then

$$\int_0^R n(r)\frac{dr}{r} = \int_0^1 \log |f(Re^{2\pi i\theta})| d\theta - \log |f(0)|. \tag{7.6}$$

In order to relate the number of zeros to the growth of an entire function at infinity, it is convenient to make the following definition: we say that an entire function has an order of growth $\rho$ if for any $s > \rho$ we can find two constants $A_s$ and $B_s$ so that

$$|f(z)| \leq A_s e^{B_s|z|^s}, \text{ for all } z \in \mathbb{C}.$$

**Theorem 7.1** *Let $f$ be an entire function of an order of growth $\rho$, then for any $s > \rho$ we have $n(r) \leq C_s(1 + r)^s$, and its zeros $z_k$ (such that $z_k \neq 0$) satisfy*

$$\sum_{k=1}^{\infty} \frac{1}{|z_k|^s} < \infty, \quad \textit{for all } s > \rho. \tag{7.7}$$

*Proof.* We may assume without loss of generality that $f(0) \neq 0$ – otherwise, we simply consider $g(z) = f(z)/z^m$, where $m$ is the order of vanishing of $f$ at $z = 0$ – this does not change the order of $f$, not modify the sum in (7.7). If $f(0) \neq 0$, we have

$$\int_0^R n(r)\frac{dr}{r} = \int_0^1 \log |f(Re^{2\pi i\theta}) d\theta - \log |f(0)|. \tag{7.8}$$

It follows that

$$\int_R^{2R} n(r)\frac{dr}{r} \leq \int_0^1 \log |f(2Re^{2\pi i\theta}) d\theta - \log |f(0)|.$$

As $n(r)$ is increasing, we have

$$\int_R^{2R} n(r)\frac{dr}{r} \geq n(R)\int_R^{2R} \frac{dr}{r} = n(R)\log 2.$$

The growth condition on $f$ means that the right side in (7.8) can be bounded, for all $s > \rho$, as

$$\int_0^1 \log |f(Re^{2\pi i\theta}) d\theta \leq \int_0^1 \log |Ae^{BR^s}| d\theta \leq C(1 + R)^s.$$

For the second statement we can write, using the dyadic blocks, for any $s > \rho$:

$$\sum_{k=1}^{\infty} \frac{1}{|z_k|^s} = \sum_{j=0}^{\infty} \sum_{2^j \leq |z_k| < 2^{j+1}} \frac{1}{|z_k|^s} \leq \sum_{j=0}^{\infty} \frac{1}{2^{js}} n(2^{j+1}) \leq C \sum_{j=0}^{\infty} \frac{1}{2^{js}}(1 + 2^{j+1})^{s'} < +\infty.$$

Here we have chosen $s' \in (\rho, s)$. □

The simple example $f(z) = \sin z$ shows that the condition $s > \rho$ can not be removed in the theorem. Indeed, we have $|\sin z| \leq e^{|z|}$ so $\rho = 1$. However, the zeros are of the form $z = \pi n$, with $n \in \mathbb{Z}$, so that

$$\sum_{j=1}^{\infty} \frac{1}{|z_j|^{\rho}} = +\infty$$

in this case.

## 7.2 Entire functions with prescribed zeros

Now, we ask the following question: given a sequence $z_k \in \mathbb{C}$, can we find an entire function that has exactly these zeros? If the set $\{z_k\}$ is finite, the answer is simple:

$$f(z) = (z - z_1)(z - z_2) \dots (z - z_N).$$

It turns out that if the set $\{z_k\}$ is infinite then the answer is, basically, the same. Moreover, the obvious obstruction – if the set $\{z_k\}$ has a limit point, then $f$ can only be identically equal to zero by the uniqueness theorem, is the only obstruction. In other words, given any set $\{z_k\}$ with no limit points in $\mathbb{C}$, there exists an entire function that vanishes exactly at $z_k$.

In order to construct such entire functions, we need to use infinite products. Let us recall the following basic result on the infinite products.

**Lemma 7.2** *If* $\sum_{n=1}^{\infty} |a_n| < +\infty$*, then the product* $\prod_{n=1}^{\infty}(1 + a_n)$ *converges. Moreover, the product is non-zero unless one of* $a_n + 1 = 0$ *for some* $n$*.*

*Proof.* As the series $\sum a_n$ is absolutely convergent, we may assume without loss of generality that all $|a_n| < 1/2$. Therefore, we may choose one branch of $\log(1 + z)$ that is holomorphic in the disk $\{|z| < 1/2\}$ and write

$$\prod_{n=1}^{N}(1 + a_n) = \exp(\sum_{n=1}^{N} \log(1 + a_n)).$$

Note that $|\log(1 + z)| \leq 2|z|$ for $|z| < 1/2$. It follows that the series

$$\sum_{n=1}^{\infty} \log(1 + a_n)$$

converges to a limit $A$, proving the first assertion. Moreover, the limit of the infinite product is non-zero since it is the exponential of the limit of the series above. □
This result generalizes to the product of holomorphic functions.

**Proposition 7.3** *Let $F_n$ be a sequence of holomorphic functions in a domain $\Omega$. Assume that there exists a sequence $c_n > 0$ so that*

$$\sum_{n=1}^{\infty} c_n < +\infty,$$

*and $|F_n(z) - 1| \le c_n$ for all $n \ge 1$ and all $z \in \Omega$. Then the product $\prod_{n=1}^{\infty} F_n(z)$ converges to a holomorphic function in $\Omega$, and if $F_n(z)$ does not vanish for any $n$ then*

$$\frac{F'(z)}{F(z)} = \sum_{n=1}^{\infty} \frac{F_n'(z)}{F(z)}.$$

The proof uses Lemma 7.2 and the fact that the sum of a uniformly converging series of holomorphic functions is itself holomorphic.

We now describe Weierstrass' construction of an entire function with prescribed zeros.

**Theorem 7.4** *Given any sequence $a_n \in \mathbb{C}$ with $|a_n| \to \infty$ as $n \to \infty$ there exists an entire function $f(z)$ whose set of zeros coincides with $\{a_n\}$. Any other function of such form is given by $f(z)e^{g(z)}$ with an entire function $g(z)$.*

The second part of the theorem is easy: if $f_1$ and $f_2$ are two entire functions with the same set of zeros (with multiplicities) then $f_1/f_2$ is an entire function that vanishes nowhere, hence it must have the form $e^{g(z)}$ with an entire function $g(z)$.

In order to construct the required entire function it is tempting to set $f(z)$ to be the product

$$\prod_{n=1}^{\infty} \left(1 - \frac{z}{a_n}\right).$$

It is not clear, however, why the product would converge, hence we need to correct this expression. Imagine that we have constructed functions $E_n(z)$ such that the only zero of $E_n(z)$ is $z = 1$, and we have

$$|1 - E_n(z)| \le \frac{C}{2^{n+1}}, \text{ for } |z| < 1/2. \tag{7.9}$$

Consider then the function

$$f(z) = z^m \prod_{m=1}^{\infty} E_n\left(\frac{z}{a_n}\right),$$

where $m$ is the (prescribed) order of vanishing at $z = 0$. We claim that $f(z)$ satisfies the requirements of the theorem. First, we have $f(a_n) = 0$. Moreover, for any $R > 0$ the product

$$\prod_{|a_n| \ge 2R} E_n\left(\frac{z}{a_n}\right)$$

is then a holomorphic function in $|z| < R$, and does not vanish in that disk. Hence, $f$ is a holomorphic function in all of $\mathbb{C}$ and vanishes only at $a_n$.

Thus, we only need to exhibit $E_n(z)$ that vanish only at $z = 1$ and satisfy (7.9). They are defined as:

$$E_0(z) = 1 - z, \quad E_k(z) = (1 - z) \exp\left\{z + \frac{z^2}{2} + \cdots + \frac{z^k}{k}\right\}. \tag{7.10}$$

In order to verify that (7.9) holds, note that for $|z| < 1/2$ we have

$$\log E_k(z) = \log(1 - z) + z + \frac{z^2}{2} + \cdots + \frac{z^k}{k} = -\sum_{n=k+1}^{\infty} \frac{z^n}{n},$$

or

$$|\log E_k(z)| \le |z|^{k+1} \sum_{n=k+1}^{\infty} \frac{|z|^{n-k-1}}{n} \le |z|^{k+1} \sum_{n=0}^{\infty} |z|^n \le C|z|^{k+1} \le \frac{C}{2^{k+1}}.$$

It follows that

$$|1 - E_k(z)| = |1 - e^{\log E_k(z)}| \le C|\log E_k(z)| \le \frac{C}{2^{k+1}},$$

and the proof of the theorem is complete.

## 7.3 Hadamard's factorization theorem

Recall that a function has an order of growth $\rho$ if $\rho$ is the smallest number such that for all $s > \rho$ we have

$$|f(z)| \le A_s e^{B_s |z|^s}.$$

We have already proved that if $f$ has growth of order $\rho$ then for any $s > \rho$ we have $n(r) \le C_s(1 + r)^s$, and its zeros $a_1, \dots, a_n, \dots$ satisfy

$$\sum_{n=1} \frac{1}{|a_n|^s} < +\infty.$$

We will prove the following.

**Theorem 7.5** *Let $f$ be entire and have growth order $\rho_0$, with zeros $a_1, \dots, a_n, \dots$, and let $k$ be an integer such that $k \le \rho_0 < k + 1$, then $f(z)$ has the representation*

$$f(z) = e^{P(z)} z^m \prod_{n=1}^{\infty} E_k\left(\frac{z}{a_n}\right). \tag{7.11}$$

*Here $P(z)$ is a polynomial of degree at most $k$, and $m$ is the order of zero of $f$ at $z = 0$.*

The main improvement here compared to the Weierstrass factorization is that the factors $E_k(z/a_n)$ have a constant order $k$, and the overall exponential factor is a polynomial. The reason this is possible is as follows. Consider the function

$$E(z) = z^m \prod_{n=1}^{\infty} E_k\left(\frac{z}{a_n}\right). \tag{7.12}$$

Previously, when we considered the function

$$E_1(z) = z^m \prod_{n=1}^{\infty} E_n\left(\frac{z}{a_n}\right),$$

in order to show that the function $E_1(z)$ is entire, we used the estimate

$$|1 - E_n(z)| \leq C|z|^{n+1} \tag{7.13}$$

for $|z| < 1/2$, which is summable in $n$. Now, however, we know, in addition, that $|a_n|$ have to grow at a certain rate: the series

$$\sum_{n=1}^{\infty} \frac{1}{|a_n|^{k+1}} < +\infty$$

converges. Therefore, we use the bound (7.13) with $n = k$ only: for any fixed $z$ and sufficiently large $n$ we have

$$\left|1 - E_k\left(\frac{z}{a_n}\right)\right| \leq C\left|\frac{z}{a_n}\right|^{k+1},$$

hence for any $R > 0$, and $|z| < R$ the series

$$\sum_{|a_n| \geq 2R} \left|1 - E_k\left(\frac{z}{a_n}\right)\right|$$

is majorized by the series

$$\sum_{|a_n| \geq 2R} \left|\frac{z}{a_n}\right|^{k+1},$$

which converges. Hence, the function $E(z)$ given by (7.12) is entire and has the same zeros (with multiplicity) as $f$. Therefore, the ratio $f(z)/E(z)$ can be written as

$$\frac{f(z)}{E(z)} = e^{g(z)},$$

with a holomorphic function $g(z)$. Our task is to show that $g(z)$ has to be a polynomial of degree at most $k$. We will do this by showing that the ratio $f(z)/E(z)$ does not grow too fast. This relies on the following two lemmas.

**Lemma 7.6** *For any $s > \rho_0$ there exists a sequence $r_m \to +\infty$ so that*

$$\left|\prod_{n=1}^{\infty} E_k\left(\frac{z}{a_n}\right)\right| \geq e^{-c|z|^s}, \tag{7.14}$$

*for $z$ with $|z| = r_m$.*

**Lemma 7.7** *Assume that $g$ is an entire function and there exists a sequence $r_n \to +\infty$ so that*

$$Re\, g(z) \leq cr_n^s, \quad \text{for all } z \text{ with } |z| = r_n. \tag{7.15}$$

*Then $g$ is a polynomial of degree less or equal to $s$.*

The end of the proof of Hadamard's theorem is as follows: for any $s > \rho_0$ we have

$$|f(z)| \leq A_s e^{B|z|^s},$$

and by Lemma 7.6 there exists a sequence $r_m \to +\infty$ so that

$$|E(z)| \geq c_1 e^{-c|z|^s}, \text{ for all } z \text{ with } |z| = r_m.$$

Therefore, we have

$$e^{Reg(z)} = \left|\frac{f(z)}{E(z)}\right| \leq C_1 e^{C_2|z|^s} \text{ for all } z \text{ with } |z| = r_m,$$

or

$$Reg(z) \leq C|z|^s, \text{ for all } z \text{ with } |z| = r_m.$$

This, by Lemma 7.7, implies that $g(z)$ is a polynomial of degree less or equal to $s$. As $s$ is an arbitrary number larger than $\rho_0$ and $k \leq \rho_0 < k+1$, $g(z)$ is a polynomial of degree at most $k$.

Therefore, to finish the proof we need to prove Lemmas 7.6 and 7.7. We begin with Lemma 7.7 which is shorter. As $g(z)$ is an entire function, we can write it as

$$g(z) = \sum_{n=0} a_n z^n,$$

and our goal is to show that $a_n = 0$ for $n > s$. Observe that for $n \geq 0$ we have,

$$a_n r^n = \int_0^1 g(re^{2\pi i\theta})e^{-2\pi in\theta}d\theta, \quad n \geq 0,$$

while for $n < 0$ these integrals vanish:

$$\int_0^1 g(re^{2\pi i\theta})e^{-2\pi in\theta}d\theta = 0, \quad n < 0.$$

It follows that

$$a_n r^n = 2\int_0^1 Reg(re^{2\pi i\theta})e^{-2\pi i n\theta}d\theta, \quad n > 0,$$

while for $n = 0$ we simply have

$$2Rea_0 = \int_0^1 Reg(re^{2\pi i\theta})d\theta.$$

We conclude that for all $n > 0$, any $C > 0$ an any $r > 0$ we have

$$a_n = \frac{2}{r^n}\int_0^1 \left[Reg(re^{2\pi i\theta}) - Cr^s\right] e^{-2\pi i n\theta}d\theta, \quad n > 0.$$

Now, for $c$ and $r_m$ as in the assumption of the Lemma, we have then

$$\begin{aligned}|a_n| &\leq \frac{2}{r_m^n}\int_0^1 \left|Reg(r_m e^{2\pi i\theta}) - Cr_m^s\right| d\theta = \frac{2}{r_m^n}\int_0^1 \left(Cr_m^s - Reg(r_m e^{2\pi i\theta})\right) d\theta \\ &\leq \frac{C'}{r_m^{n-s}} - \frac{2Rea_0}{r_m^n} \to 0 \text{ as } r_m \to +\infty,\end{aligned}$$

for all $n > s$, and the proof of Lemma 7.7 is complete.

The proof of Lemma 7.6 is quite a bit longer. An obvious potential obstacle is that $E_k(a/z_n) = 0$ when $z = a_n$. Therefore, the conclusion of the lemma can not hold for all $z$ and the best we can hope for is to prove it for points separated away from $a_n$. The main step is to prove the lower bound for $E_1(z)$ outside of "forbidden disks" centered at $a_n$, and then prove that the union of forbidden disks does not cover too many circles centered at zero. More precisely, we will show that for any $s$ such that $\rho_0 < s < k+1$ we have

$$\left|\prod_{n=1}^{\infty} E_k\left(\frac{z}{a_n}\right)\right| \geq e^{-c|z|^s}, \text{ for all } z \text{ such that } |z - a_n| > \frac{1}{|a_n|^{k+1}} \text{ for all } n \geq 1. \quad (7.16)$$

This is sufficient because the series

$$\sum_{n=1}^{\infty} \frac{1}{|a_n|^{k+1}} < +\infty$$

converges. Indeed, we can choose $N$ so that

$$\sum_{n=N}^{\infty} \frac{1}{|a_n|^{k+1}} < \frac{1}{10}.$$

It follows that for all integers $L$ we can find $r \in (L, L+1)$ so that none of the points on the circle $\{|z| = r\}$ intersect any of the forbidden disks with $n \geq N$, and the claim of

Lemma 7.6 will follow. Hence, we concentrate on the proof of (7.16). For a given $z$ we will consider separately

$$I = \prod_{|a_n|>2|z|} E_k\left(\frac{z}{a_n}\right)$$

and

$$II = \prod_{|a_n|\le 2|z|} E_k\left(\frac{z}{a_n}\right). \tag{7.17}$$

In the first term we will use the fact that for $|z| < 1/2$ we can write

$$E_k(z) = \exp\left\{\log(1-z) + \sum_{n=1}^{k} \frac{z^n}{n}\right\} = \exp\left\{-\sum_{n=k+1}^{\infty} \frac{z^n}{n}\right\} \ge e^{-c|z|^{k+1}}.$$

This gives

$$|I| \ge \prod_{|a_n|>2|z|} e^{-c|z/a_n|^{k+1}} \ge \exp\left\{-c|z|^{k+1} \sum_{|a_n|>2|z|} \frac{1}{|a_n|^{k+1}}\right\}.$$

However, for $|a_n| > 2|z|$ and $s < k+1$ we have

$$\frac{1}{|a_n|^{k+1}} = \frac{1}{|a_n|^s}\frac{1}{|a_n|^{k+1-s}} \le \frac{C}{|a_n|^s}\frac{1}{|z|^{k+1-s}}.$$

As the series

$$\sum_{n=1}^{\infty} \frac{1}{|a_n|^s} < +\infty$$

is summable, we deduce that

$$|I| \ge \exp\left\{-c|z|^s\right\}.$$

Finally, in order to bound the second term given by (7.17) we note that for $|z| > 1/2$ we have

$$\begin{aligned} |E_k(z)| &= |1-z| \left|\exp(z + \frac{z^2}{2} + \cdots + \frac{z^k}{k})\right| \ge |1-z| \exp(-|z + \frac{z^2}{2} + \cdots + \frac{z^k}{k}|) \\ &\ge |1-z| \exp(-c|z|^k). \end{aligned} \tag{7.18}$$

It follows that

$$|II| \ge \prod_{|a_n|\le 2|z|} \left|1 - \frac{z}{a_n}\right| \prod_{|a_n|\le 2|z|} e^{-c|z/a_n|^k}. \tag{7.19}$$

The second product above is estimated as before, by writing

$$\sum_{|a_n|\le 2|z|} \left|\frac{z}{a_n}\right|^k = |z|^k \sum_{|a_n|\le 2|z|} |a_n|^{s-k} \frac{1}{|a_n|^s} \le C|z|^k \sum_{|a_n|\le 2|z|} |z|^{s-k} \frac{1}{|a_n|^s} \le C|z|^s,$$

whence

$$\prod_{|a_n|\le 2|z|} e^{-c|z/a_n|^k} \ge e^{-c|z|^s}.$$

Finally, we come to the first term in the right side of (7.19) that forces us to keep $z$ outside of the forbidden disks (it is clear from its form that some assumption on the distance from $z$ to $a_n$ is needed). Note that, if $|a_n - z| > 1/|a_n|^{k+1}$ (that is, $z$ is outside all of the forbidden disks), we have

$$\prod_{|a_n|\le 2|z|} \left|1-\frac{z}{a_n}\right| = \prod_{|a_n|\le 2|z|} \left|\frac{a_n - z}{a_n}\right| \ge \prod_{|a_n|\le 2|z|} \frac{1}{|a_n|^{k+2}}.$$

Observe that (recall that $n(r)$ is the number of zeros of $f$ inside the disk $\{|z|\le r\}$)

$$(k+2)\sum_{|a_n|\le 2|z|} \log|a_n| \le (k+2)n(2|z|)\log(2|z|).$$

Theorem 7.1 implies that $n(r) \le C(1+r)^{s'}$ for any $s' > \rho_0$, thus

$$(k+2)\sum_{|a_n|\le 2|z|} \log|a_n| \le C(k+2)(1+|z|)^{s'}\log(2|z|) \le C|z|^s,$$

if we take $\rho_0 < s' < s$. As a consequence, we have

$$\prod_{|a_n|\le 2|z|} \left|1-\frac{z}{a_n}\right| \ge \prod_{|a_n|\le 2|z|} \frac{1}{|a_n|^{k+2}} \ge e^{-c|z|^s},$$

and the proof of Theorem 7.5 is complete!

# 8 The Basics of the Geometric Theory

This section (taken from the Shabat book) recalls the basics of the geometric theory of functions of a complex variable.

## 8.1 The Argument Principle

Let the function $f$ be holomorphic in a punctured neighborhood $\{0 < |z-a| < r\}$ of a point $a \in \mathbb{C}$. We assume also that $f$ does not vanish in this neighborhood. The logarithmic residue of the function $f$ at the point $a$ is the residue of the logarithmic derivative

$$\frac{f'(a)}{f(a)} = \frac{d}{dz}\mathrm{Ln} z \tag{8.1}$$

of this function at the point $a$.

Apart from isolated singular points the function $f$ may have a non-zero logarithmic residue at its zeros. Let $a \in \mathbb{C}$ be a zero of order $n$ of a function $f$ holomorphic at $a$.

Then we have $f(z) = (z-a)^n \phi(z)$ in a neighborhood $U_a$ of $a$ with the function $\phi$ holomorphic and different from zero in $U_a$. Therefore we have in $U_a$

$$\frac{f'(z)}{f(z)} = \frac{n(z-a)^{n-1}\phi(z) + (z-a)^n \phi'(z)}{(z-a)^n \phi(z)} = \frac{1}{z-a} \cdot \frac{n\phi(z) + (z-a)\phi'(z)}{\phi(z)}$$

with the second factor holomorphic in $U_a$. Hence it may be expanded into the Taylor series with the zero order term equal to $n$. Therefore we have in $U_a$

$$\frac{f'(z)}{f(z)} = \frac{1}{z-a}\left\{n + c_1(z-a) + c_2(z-a)^2 + \dots\right\} = \frac{n}{z-a} + c_1 + c_2(z-a) + \dots \quad (8.2)$$

This shows that the logarithmic derivative has a pole of order one with residue equal to $n$ at the zero of order $n$ of $f$: *the logarithmic residue at a zero of a function is equal to the order of this zero.*

If $a$ is a pole of $f$ of the order $p$ then $1/f$ has a zero of order $p$ at this point. Observing that

$$\frac{f'(z)}{f(z)} = -\frac{d}{dz}\mathrm{Ln}\frac{1}{f(z)},$$

and using (8.2) we conclude that the logarithmic derivative has residue equal to $-p$ at a pole of order $p$: *the logarithmic residue at a pole is equal to the order of this pole with the minus sign.*

Those observations allow to compute the number of zeros and poles of meromorphic functions. We adopt the convention that a pole and a zero are counted as many times as their order is.

**Theorem 8.1** *Let the function $f$ be meromorphic in a domain $D \subset \mathbb{C}$ and let $G$ be a domain properly contained in $D$ with the boundary $\partial G$ that is a continuous curve. Let us assume that $\partial G$ contains neither poles nor zeros of $f$ and let $N$ and $P$ be the total number of zeros and poles of $f$ in the domain $G$, then*

$$N - P = \frac{1}{2\pi i}\int_{\partial G} \frac{f'(z)}{f(z)} dz. \quad (8.3)$$

**Proof.** The function $f$ has only finitely many poles $a_1, \dots, a_l$ and zeros $b_1, \dots, b_m$ in $G$ since $G$ is properly contained in $D$. The function $g = f'/f$ is holomorphic in a neighborhood of $\partial G$ since the boundary of $G$ does not contain poles or zeros. Applying the Cauchy theorem on residues to $g$ we find

$$\frac{1}{2\pi i}\int_{\partial G} \frac{f'}{f} dz = \sum_{\nu=1}^{l} \mathrm{res}_{a_\nu} g + \sum_{\nu=1}^{m} \mathrm{res}_{b_\nu} g. \quad (8.4)$$

However, according to our previous remark,

$$\mathrm{res}_{a_\nu} g = n_\nu, \quad \mathrm{res}_{b_\nu} g = p_\nu.$$

Here $n_\nu$ and $p_\nu$ are the order of zero $a_\nu$ and pole $b_\nu$, respectively. Using this in (8.4) and counting the multiplicities of zeros and poles we obtain (8.3) since $N = \sum n_\nu$ and $P = \sum p_\nu$. $\square$

The theorem that we have just proved has a geometric interpretation. Let us parameterize $\partial G$ as $z = z(t)$, $\alpha \le t \le \beta$ and denote by $\Phi(t)$ the anti-derivative of $\dfrac{f'}{f}$ along this path. The Newton-Leibnitz formula implies that

$$\int_{\partial G} \frac{f'(z)}{f(z)} dz = \Phi(\beta) - \Phi(\alpha). \tag{8.5}$$

However, clearly, $\Phi(t) = \ln[f(z(t))]$, where ln denotes any branch of the logarithm that varies continuously along the path $\partial G$. It suffices to choose a branch of $\arg f$ that varies continuously along $\partial G$ since $\mathrm{Ln} f = \ln|f| + i\mathrm{Arg} f$ and the function $\ln|f|$ is single-valued. The increment of $\ln|f|$ along a closed path $\partial G$ is equal to zero and thus

$$\Phi(\beta) - \Phi(\alpha) = i\{\arg f(z(\beta)) - \arg f(z(\alpha))\}.$$

We denote the increment of the argument of $f$ in the right side by $\Delta_{\partial G} f$ and re-write (8.5) as

$$\int_{\partial G} \frac{f'}{f} dz = i\Delta_{\partial G} \arg f.$$

Theorem 8.1 may now be expressed as

**Theorem 8.2** *(The argument principle) Under the assumptions of Theorem 8.1 the difference between the number of zeros $N$ and the number of poles $P$ of a function $f$ in a domain $G$ is equal to the increment of the argument of this function along the oriented boundary of $G$ divided by $2\pi$:*

$$N - P = \frac{1}{2\pi} \Delta_{\partial G} \arg f. \tag{8.6}$$

Geometrically the right side of (8.6) is the total number of turns the vector $w = f(z)$ makes around $w = 0$ as $z$ varies along $\partial G$. Let us denote by $\partial G^*$ the image of $\partial G$ under the map $f$, that is, the path $w = f(z(t))$, $\alpha \le t \le \beta$. Then this number is equal to the total number of times the vector $w$ rotates around $w = 0$ as it varies along $\partial G^*$. This number is called the winding number of $\partial G^*$ around $w = 0$, we will denote it by $\mathrm{ind}_0 \partial G^*$. The argument principle states that

$$N - P = \frac{1}{2\pi} \Delta_{\partial G} \arg f = \mathrm{ind}_0 \partial G^*. \tag{8.7}$$

**Remark 8.3** *We may consider the $a$-points of $f$, solutions of $f(z) = a$ and not only its zeros: it suffices to replace $f$ by $f(z) - a$ in our arguments. If $\partial G$ contains neither poles nor $a$-points of $f$ then*

$$N_a - P = \frac{1}{2\pi i} \int_{\partial G} \frac{f'(z)}{f(z) - a} dz = \frac{1}{2\pi} \Delta_{\partial G} \arg\{f(z) - a\}, \tag{8.8}$$

*where $N_a$ is the number of a-points of $f$ in the domain $D$. Passing to the plane $w = f(z)$ and introducing the index of the path $\partial G^*$ around the point $a$ we may re-write (8.8) as*

$$N_a - P = \frac{1}{2\pi}\Delta_{\partial G} \arg\{f(z) - a\} = ind_a \partial G^*. \tag{8.9}$$

The next theorem is an example of the application of the argument principle.

**Theorem 8.4** *(Rouche[14]) Let the functions $f$ and $g$ be holomorphic in a closed domain $\bar{G}$ with a continuous boundary $\partial G$ and let*

$$|f(z)| > |g(z)| \quad \textit{for all } z \in \partial G. \tag{8.10}$$

*Then the functions $f$ and $f + g$ have the same number of zeros in $G$.*

**Proof.** Assumption (8.10) shows that neither $f$ nor $f + g$ vanish on $\partial G$ and thus the argument principle might be applied to both of these functions. Moreover, since $f \neq 0$ on $\partial G$, we have $f + g = f\left(1 + \frac{g}{f}\right)$ and thus we have with the appropriate choice of a branch of the argument:

$$\Delta_{\partial G} \arg(f + g) = \Delta_{\partial G} \arg f + \Delta_{\partial G} \arg\left(1 + \frac{g}{f}\right). \tag{8.11}$$

However, since $\left|\frac{g}{f}\right| < 1$ on $\partial G$, the point $\omega = \frac{g}{f}$ lies in $\{|\omega| < 1\}$ for all $z \in \partial G$. Therefore the vector $w = 1 + \omega$ may not turn around zero and hence the second term in the right side of (8.11) vanishes. Therefore, $\Delta_{\partial G} \arg(f + g) = \Delta_{\partial G} \arg f$ and the argument principle implies the statement of the theorem. □

The Rouche theorem is useful in counting the zeros of holomorphic functions. In particular it implies the main theorem of algebra in a very simple way.

**Theorem 8.5** *Any polynomial $P_n$ of degree $n$ has exactly $n$ roots in $\mathbb{C}$.*

**Proof.** All zeros of $P_n$ must lie in a disk $\{|z| < R\}$ since $P_n$ has a pole at infinity. Let $P_n = f + g$ where $f = a_0 z^n$, $a_0 \neq 0$ and $g = a_1 z^{n-1} + \cdots + a_n$, then, possibly after increasing $R$, we may assume that $|f| > |g|$ on $\{|z| = R\}$ since $|f| = |a_0|R^n$ while $g$ is a polynomial of degree less than $n$. The Rouche theorem implies that $P_n$ has as many roots in $\{|z| < R\}$ as $f = a_0 z^n$, that is, exactly $n$ of them. □

## 8.2 The Open Mapping Theorem

**Theorem 8.6** [15] *If a function $f$ holomorphic in a domain $D$ is not equal identically to a constant then the image $D^* = f(D)$ is also a domain.*

[14]Eugene Rouche (1832-1910) was a French mathematician.
[15]This theorem was proved by Riemann in 1851.

**Proof.** We have to show that $D^*$ is connected and open. Let $w_1$ and $w_2$ be two arbitrary points in $D^*$ and let $z_1$ and $z_2$ be some pre-images of $w_1$ and $w_2$, respectively. Since the domain $D$ is path-wise connected there exists a path $\gamma : [\alpha, \beta] \to D$ that connects $z_1$ and $z_2$. Its image $\gamma^* = f \circ \gamma$ connects $w_1$ and $w_2$ and is a path since the function $f$ is continuous. Moreover, it is clearly contained in $D^*$ and hence the set $D^*$ is path-wise connected.

Let $w_0$ be an arbitrary point in $D^*$ and let $z_0$ be a pre-image of $w_0$. There exists a disk $\{|z - z_0| < r\}$ centered at $z_0$ that is properly contained in $D$ since $D$ is open. After decreasing $r$ we may assume that $\{|z - z_0| \le r\}$ contains no other $w_0$-points of $f$ except $z_0$: since $f \ne$ const its $w_0$ points are isolated in $D$. We denote by $\gamma = \{|z - z_0| = r\}$ the boundary of this disk and let

$$\mu = \min_{z \in \gamma} |f(z) - w_0|. \tag{8.12}$$

Clearly $\mu > 0$ since the continuous function $|f(z) - w_0|$ attains its minimum on $\gamma$, so that if $\mu = 0$ then there would exist a $w_0$-point of $f$ on $\gamma$ contrary to our construction of the disk.

Let us now show that the set $\{|w - w_0| < \mu\}$ is contained in $D^*$. Indeed, let $w_1$ be an arbitrary point in this disk, that is, $|w_1 - w_0| < \mu$. Then we have

$$f(z) - w_1 = f(z) - w_0 + (w_0 - w_1), \tag{8.13}$$

and, moreover, $|f(z) - w_0| \ge \mu$ on $\gamma$. Then, since $|w_0 - w_1| < \mu$, the Rouche theorem implies that the function $f(z) - w_1$ has as many roots inside $\gamma$ as $f(z) - w_0$. Hence it has at least one zero (the point $z_0$ may be a zero of order higher than one of $f(z) - w_0$). Thus the function $f$ takes the value $w_1$ and hence $w_1 \in D^*$. However, $w_1$ is an arbitrary function in the disk $\{|w - w_0| < \mu\}$ and hence this whole disk is contained in $D^*$ so that $D^*$ is open. □

**Exercise 8.7** Let $f$ be holomorphic in $\{\mathrm{Im} z \ge 0\}$, real on the real axis and bounded. Show that $f \equiv$ const.

A similar but more detailed analysis leads to the solution of the problem of local inversion of holomorphic functions. This problem is formulated as follows.

A holomorphic function $w = f(z)$ is defined at $z_0$, find a function $z = g(w)$ analytic at $w_0 = f(z_0)$ so that $g(w_0) = z_0$ and $f(g(w)) = w$ in a neighborhood of $w_0$.

We should distinguish two cases in the solution of this problem:

I. The point $z_0$ is not a critical point: $f'(z_0) \ne 0$. As in the proof of the open mapping theorem we choose a disk $\{|z - z_0| \le r\}$ that contains no $w_0$-points except $z_0$, and define $\mu$ according to (8.12). Let $w_1$ be an arbitrary point in the disk $\{|w - w_0| < \mu\}$. Then the same argument (using (8.13) and the Rouche theorem) shows that the function $f$ takes the value $w_1$ as many times as $w_0$. However, the value $w_0$ is taken only once and, moreover, $z_0$ is a simple zero of $f(z) - w_0$ since $f'(z_0) \ne 0$.

Therefore the function $f$ takes all values in the disk $\{|w - w_0| < \mu\}$ once in the disk $\{|z - z_0| < r\}$. In other words, the function $f$ is a local bijection at $z_0$.

Then the function $z = g(w)$ is defined in the disk $\{|w - w_0| < r\}$ so that $g(w_0) = z_0$ and $f \circ g(w) = w$. Furthermore, derivative $g'(w)$ exists at every point of the disk $\{|w - w_0| < r\}$:

$$g'(w) = \frac{1}{f'(z)} \tag{8.14}$$

and thus $g$ is holomorphic in this disk[16].

II. The point $z_0$ is a critical point: $f'(z_0) = \cdots = f^{(p-1)}(z_0) = 0$, $f^{(p)} \neq 0$, $p \geq 2$. Repeating the same argument as before choosing a disk $\{|z - z_0| < r\}$ that contains neither $w_0$-points of $f$ nor zeros of the derivative $f'$ (we use the uniqueness theorem once again). As before, we choose $\mu > 0$, take an arbitrary point $w_1$ in the disk $\{|w - w_0| < \mu\}$ and find that $f$ takes the value $w_1$ as many times as $w_0$. However, in the present case the $w_0$-point $z_0$ has multiplicity $p$: $z_0$ is a zero of order $p$ of $f(z) - w_0$. Furthermore, since $f'(z) \neq 0$ for $0 < |z - z_0| < r$ the value $w_1$ has to be taken at $p$ different points. Therefore, the function $f$ takes each value $p$ times in $\{|z - z_0| < r\}$.

The above analysis implies the following

**Theorem 8.8** *Condition $f'(z_0) \neq 0$ is necessary and sufficient for the local invertibility of a holomorphic function $f$ at the point $z_0$.*

**Remark 8.9** The general inverse function theorem of the real analysis implies that the assumption $f'(z_0) \neq 0$ is sufficient for the local invertibility since the Jacobian $J_f(z) = |f'(z)|^2$ of the map $(x, y) \to (u, v)$ is non-zero at this point. However, for an arbitrary differentiable map to be locally invertible one needs not $J_f(z) \neq 0$ to hold. This may be seen on the example of the map $f = x^3 + iy$ that has Jacobian equal to zero at $z = 0$ but that is nevertheless one-to-one.

**Remark 8.10** The local invertibility condition $f'(z) \neq 0$ for all $z \in D$ is not sufficient for the global invertibility of the function in the whole domain $D$. This may be seen on the example of $f(z) = e^z$ that is locally invertible at every point in $\mathbb{C}$ but is not one-to-one in any domain that contains two points that differ by $2k\pi i$ where $k \neq 0$ is an integer.

## 8.3 The maximum modulus principle and the Schwartz lemma

The maximum modulus principle is expressed by the following theorem.

**Theorem 8.11** *If the function $f$ is holomorphic in a domain $D$ and its modulus $|f|$ achieves its (local) maximum at a point $z_0 \in D$ then $f$ is constant.*

**Proof.** We use the open mapping theorem. If $f \neq$ const then it maps $z_0$ into a point $w_0$ of the domain $D^*$. There exists a disk $\{|w - w_0| < \mu\}$ centered at $w_0$ that is contained in $D^*$. There must be a point $w_1$ in this disk so that $|w_1| > |w_0|$. The value $w_1$ is taken

[16]Expression (8.14) shows that in order for derivative to exist we need $f' \neq 0$. Using continuity of $f'$ we may conclude that $f' \neq 0$ in the disk $\{|z - z_0| < r\}$, possibly decreasing $r$ if needed.

by the function $f$ in a neighborhood of the point $z_0$ which contradicts the fact that $|f|$ achieves its maximum at this point. $\square$

Taking into account the properties of continuous functions on a closed set the maximum modulus principle may be reformulated as

**Theorem 8.12** *If a function $f$ is holomorphic in a domain $D$ and continuous in $\bar{D}$ then $|f|$ achieves its maximum on the boundary $\partial D$.*

**Proof.** If $f = \text{const}$ in $D$ (and hence in $\bar{D}$ by continuity) the statement is trivial. Otherwise if $f \neq \text{const}$ then $|f|$ may not attain its maximum at the points of $D$. However, since this maximum is attained in $\bar{D}$ it must be achieved on $\partial D$. $\square$

**Exercise 8.13** 1. Let $P(z)$ be a polynomial of degree $n$ in $z$ and let $M(r) = \max\limits_{|z|=r} |P(z)|$. Show that $M(r)/r^n$ is a decreasing function.

2. Formulate and prove the maximum principle for the real part of a holomorphic function.

A similar statement for the minimum of modulus is false in general. This may be seen on the example of the function $f(z) = z$ in the disk $\{|z| < 1\}$ (the minimum of $|f|$ is attained at $z = 0$). However, the following theorem holds.

**Theorem 8.14** *Let a function $f$ be holomorphic in a domain $D$ and not vanish anywhere in $D$. Then $|f|$ may attain its (local) minimum in $D$ only if $f = \text{const}$.*

For the proof of this theorem it suffices to apply Theorem 8.11 to the function $g = 1/f$ that is holomorphic since $f \neq 0$.

A simple corollary of the maximum modulus principle is

**Lemma 8.15** *(The Schwartz lemma[17]) Let a function $f$ be holomorphic in the unit disk $U = \{|z| < 1\}$, satisfy $|f(z)| \leq 1$ for all $z \in U$ and $f(0) = 0$. Then we have*

$$|f(z)| \leq |z| \tag{8.15}$$

*for all $z \in U$. Moreover, if the equality in (8.15) holds for at least one $z \neq 0$ then it holds everywhere in $U$ and in this case $f(z) = e^{i\alpha} z$, where $\alpha$ is a real constant.*

**Proof.** Consider the function $\phi(z) = f(z)/z$, it is holomorphic in $U$ since $f(0) = 0$. Let $U_r = \{|z| < r\}$, $r < 1$ be an arbitrary disk centered at zero. The function $\phi(z)$ attains its maximum in $U_r$ on its boundary $\gamma_r = \{|z| = r\}$ according to Theorem 8.12. However, we have $|\phi| \leq 1/r$ on $\gamma_r$ since $|f| \leq 1$ by assumption. Therefore we have

$$|\phi(z)| \leq 1/r \tag{8.16}$$

everywhere in $U_r$. We fix $z \in U$ and observe that $z \in U_r$ for $r > |z|$. Therefore (8.16) holds for any given $z$ with all $r > |z|$. We let $r \to 1$, and passing to the limit $r \to 1$ we obtain $|\phi(z)| \leq 1$ or $|f(z)| \leq |z|$. This proves the inequality (8.15).

[17]Hermann Schwartz (1843-1921) was a German mathematician, a student of Weierstrass. This important lemma has appeared in his papers of 1869-70.

Let us assume that equality in (8.15) holds for some $z \in U$, then $|\phi|$ attains its maximum equal to 1 at this point. Then $\phi$ is equal to a constant so that $\phi(z) = e^{i\alpha}$ and $f(z) = e^{i\alpha}z$. □

The Schwartz lemma implies that a holomorphic map $f$ that maps the disk $\{|z| < 1\}$ into the disk $\{|w| < 1\}$ and that takes the center to the center, maps any circle $\{|z| = r\}$ inside the disk $\{|w| < r\}$. The image of $\{|z| = r\}$ may intersect $\{|w| = r\}$ if and only $f$ is a rotation around $z = 0$.

**Exercise 8.16** *1. Show that under the assumptions of the Schwartz lemma we have $|f'(0)| \le 1$ and equality is attained if and only if $f(z) = e^{i\alpha}z$.*
*2. Let $f \in \mathcal{O}(D)$, $f : U \to U$ and $f(0) = \cdots = f^{(k-1)}(0) = 0$. Show that then $|f(z)| \le |z|^k$ for all $z \in U$.*

## 8.4 The Riemann Theorem

Any holomorphic one-to-one function defined in a domain $D$ defines a conformal map of this domain since the above assumptions imply that $f$ has no critical points in $D$. We have encountered such maps many times before. Here we consider a more difficult and important for practical purposes problem:

*Given two domains $D_1$ and $D_2$ find a one-to-one conformal map $f : D_1 \to D_2$ of one of these domains onto the other.*

**Definition 8.17** *A conformal one-to-one map of a domain $D_1$ onto $D_2$ is said to be a (conformal) isomorphism, while the domains $D_1$ and $D_2$ that admit such a map are isomorphic (or conformally equivalent). Isomorphism of a domain onto itself is called a (conformal) automorphism.*

It is easy to see that the set of all automorphisms $\phi : D \to D$ of a domain $D$ forms a group that is denoted $\mathrm{Aut}D$. The group operation is the composition $\phi_1 \circ \phi_2$, the unity is the identity map and the inverse is the inverse map $z = \phi^{-1}(w)$.

The richness of the group of automorphisms of a domain allows to understand the richness of the family of the conformal maps onto it of a different domain, as may be seen from the next

**Theorem 8.18** *Let $f_0 : D_1 \to D_2$ be a fixed isomorphism. Then any other isomorphism of $D_1$ onto $D_2$ has the form*

$$f = \phi \circ f_0 \tag{8.17}$$

*where $\phi$ is an automorphism of $D_2$.*

**Proof.** First, it is clear that all maps of the form of the right side of (8.17) are isomorphisms from $D_1$ onto $D_2$. Furthermore, if $f : D_1 \to D_2$ is an arbitrary isomorphism then $\phi = f \circ f_0^{-1}$ is a conformal map of $D_2$ onto itself, that is, an automorphism of $D_2$. Then (8.17) follows. □

In the sequel we will only consider simply connected domains $D$. We will distinguish three special domains that we will call canonical: the closed plane $\overline{\mathbb{C}}$, the open plane $\mathbb{C}$

and the unit disk $\{|z| < 1\}$. We have previously found the group of all fractional-linear automorphisms of those domains. However, the following theorem holds.

**Theorem 8.19** *Any conformal automorphism of a canonical domain is a fractional-linear transformation.*

**Proof.** Let $\phi$ be automorphism of $\overline{\mathbb{C}}$. There exists a unique point $z_0$ that is mapped to infinity. Therefore $\phi$ is holomorphic everywhere in $\mathbb{C}$ except at $z_0$ where it has a pole. This pole has multiplicity one since in a neighborhood of a pole of higher order the function $\phi$ could not be one-to-one. Therefore since the only singularities of $\phi$ are poles $\phi$ is a rational function. Since it has only one simple pole, $\phi$ should be of the form $\phi(z) = \dfrac{A}{z - z_0} + B$ if $z_0 \neq \infty$ and $\phi(z) = Az + B$ if $z_0 = \infty$. The case of the open complex plane $\mathbb{C}$ is similar.

Let $\phi$ be an arbitrary automorphism of the unit disk $U$. Let us denote $w_0 = \phi(0)$ and consider a fractional linear transformation

$$\lambda : \; w \to \frac{w - w_0}{1 - \bar{w}_0 w}$$

of the disk $U$ that maps $w_0$ into 0. The composition $f = \lambda \circ \phi$ is also an automorphism of $U$ so that $f(0) = 0$. Moreover, $|f(z)| < 1$ for all $z \in U$. Therefore the Schwartz lemma implies that $|f(z)| \leq |z|$ for all $z \in U$. However, the inverse map $z = f^{-1}(w)$ also satisfies the assumptions of the Schwartz lemma and hence $|f^{-1}(w)| \leq |w|$ for all $w \in U$ that in turn implies that $|z| \leq |f(z)|$ for all $z \in U$. Thus $|f(z)| = |z|$ for all $z \in U$ so that the Schwartz lemma implies that $f(z) = e^{i\alpha} z$. Then $\phi = \lambda^{-1} \circ f = \lambda^{-1}(e^{i\alpha} z)$ is also a fractional-linear transformation. $\square$

Taking into account our results on the Möbius transformations we obtain the complete description of all conformal automorphisms of the canonical domains.

(I) The closed complex plane:

$$\mathrm{Aut}\overline{\mathbb{C}} = \left\{ z \to \frac{az + b}{cz + d}, \; ad - bc \neq 0 \right\}. \tag{8.18}$$

(II) The open plane:

$$\mathrm{Aut}\mathbb{C} = \{ z \to az + b, \; a \neq 0 \}. \tag{8.19}$$

(III) The unit disk:

$$\mathrm{Aut}U = \left\{ z \to e^{i\alpha} \frac{z - a}{1 - \bar{a}z}, \; |a| < 1, \alpha \in \mathbb{R} \right\}. \tag{8.20}$$

It is easy to see that different canonical domains are not isomorphic to each other. Indeed, the closed complex plane $\overline{\mathbb{C}}$ is not even homeomorphic to $\mathbb{C}$ and $U$ and hence it may not be mapped conformally onto these domains. The domains $\mathbb{C}$ and $U$ are homeomorphic but there is no conformal map of $\mathbb{C}$ onto $U$ since such a map would have to be realized by an entire function such that $|f(z)| < 1$ which has then to be equal to a constant by the Liouville theorem.

A domain that has no boundary (boundary is an empty set) coincides with $\overline{\mathbb{C}}$. Domains with boundary that consists of one point are the plane $\overline{\mathbb{C}}$ without a point which are clearly conformally equivalent to $\mathbb{C}$ (even by a fractional linear transformation). The main result of this section is the Riemann theorem that asserts that any simply connected domain $D$ with a boundary that contains more than one point (and hence infinitely many points since boundary of a simply connected domain is connected) is conformally equivalent to the unit disk $U$.

This theorem will be presented later while at the moment we prove the uniqueness theorem for conformal maps.

**Theorem 8.20** *If a domain $D$ is conformally equivalent to the unit disk $U$ then the set of all conformal maps of $D$ onto $U$ depends on three real parameters. In particular there exists a unique conformal map $f$ of $D$ onto $U$ normalized by*

$$f(z_0) = 0, \ \ \arg f'(z_0) = \theta, \tag{8.21}$$

*where $z_0$ is an arbitrary point of $D$ and $\theta$ is an arbitrary real number.*

**Proof.** The first statement follows from Theorem 8.18 since the group Aut$U$ depends on three real parameters: two coordinates of the point $a$ and the number $\alpha$ in (8.20).

In order to prove the second statement let us assume that there exist two maps $f_1$ and $f_2$ of the domain $D$ onto $U$ normalized as in (8.21). Then $\phi = f_1 \circ f_2^{-1}$ is an automorphism of $U$ such that $\phi(0) = 0$ and $\arg f'(0) = 0$. Expression (8.20) implies that then $a = 0$ and $\alpha = 0$, that is $\phi(z) = z$ and $f_1 = f_2$. In order to prove the Riemann theorem we need to develop some methods that are useful in other areas of the complex analysis.

### The compactness principle

**Definition 8.21** *A family $\{f\}$ of functions defined in a domain $D$ is locally uniformly bounded if for any domain $K$ properly contained in $D$ there exists a constant $M = M(K)$ such that*

$$|f(z)| \leq M \text{ for all } z \in K \text{ and all } f \in \{f\}. \tag{8.22}$$

*A family $\{f\}$ is locally equicontinuous if for any $\varepsilon > 0$ and any domain $K$ properly contained in $D$ there exists $\delta = \delta(\varepsilon, K)$ so that*

$$|f(z') - f(z'')| < \varepsilon \tag{8.23}$$

*for all $z', z'' \in K$ so that $|z' - z''| < \delta$ and all $f \in \{f\}$.*

**Theorem 8.22** *If a family $\{f\}$ of holomorphic functions in a domain $D$ is locally uniformly bounded then it is locally equicontinuous.*

**Proof.** Let $K$ be a domain properly contained in $D$. Let us denote by $2\rho$ the distance between the closed sets $\bar{K}$ and $\partial D$, and let

$$K^{(\rho)} = \bigcup_{z_0 \in K} \{z : \ |z - z_0| < \rho\}$$

be a $\rho$-enlargement of $K$. The set $K^{(\rho)}$ is properly contained in $D$ and thus there exists a constant $M$ so that $|f(z)| \leq M$ for all $z \in K^{(\rho)}$ and $f \in \{f\}$. Let $z'$ and $z''$ be arbitrary points in $K$ so that $|z' - z''| < \rho$. The disk $U_\rho = \{z : \ |z - z'| < \rho\}$ is contained in $K^{(\rho)}$ and hence $|f(z) - f(z')| < 2M$ for all $z \in U_\rho$. The mapping $\zeta = \dfrac{1}{\rho}(z - z')$ maps $U_\rho$ onto the disk $|\zeta| < 1$ and the function

$$g(\zeta) = \frac{1}{2M}\left\{f(z' + \zeta\rho) - f(z')\right\}$$

satisfies the assumptions of the Schwartz lemma.

This lemma implies that $|g(\zeta)| \leq |\zeta|$ for all $\zeta$, $|\zeta| < 1$, which means

$$|f(z) - f(z')| \leq \frac{2M}{\rho}|z - z'| \text{ for all } z \in U_\rho. \tag{8.24}$$

Given $\varepsilon > 0$ we choose $\delta = \min\left(\rho, \dfrac{\varepsilon\rho}{2M}\right)$ and obtain from (8.24) that $|f(z') - f(z'')| < \varepsilon$ for all $f \in \{f\}$ provided that $|z' - z''| < \delta$. $\square$

**Definition 8.23** *A family of functions $\{f\}$ defined in a domain $D$ is compact in $D$ if any sequence $f_n$ of functions of this family has a subsequence $f_{n_k}$ that converges uniformly on any domain $K$ properly contained in $D$.*

**Theorem 8.24** *(Montel[18]) If a family of functions $\{f\}$ holomorphic in a domain $D$ is locally uniformly bounded then it is compact in $D$.*

**Proof.** (a) We first show that if a sequence $f_n \subset \{f\}$ converges at every point of an everywhere dense set $E \subset D$ then it converges uniformly on every compact subset $K$ of $D$. We fix $\varepsilon > 0$ and the set $K$. Using equicontinuity of the family $\{f\}$ we may choose a partition of $D$ into squares with sides parallel to the coordinate axes and so small that that for any two points $z', z'' \in K$ that belong to the same square and any $f \in \{f\}$ we have

$$|f(z') - f(z'')| < \frac{\varepsilon}{3}. \tag{8.25}$$

The set $K$ is covered by a finite number of such squares $q_p$, $p = 1, \ldots, P$. Each $q_p$ contains a point $z_p \in E$ since the set $E$ is dense in $D$. Moreover, since the sequence$\{f_n\}$ converges on $E$ there exists $N$ so that

$$|f_m(z_p) - f_n(z_p)| < \frac{\varepsilon}{3} \tag{8.26}$$

for all $m, n > N$ and all $z_p$, $p = 1, \ldots, P$.

Let now $z$ be an arbitrary point in $K$. Then there exists a point $z_p$ that belongs to the same square as $z$. We have for all $m, n > N$:

$$|f_m(z) - f_n(z)| \leq |f_m(z) - f_m(z_p)| + |f_m(z_p) - f_n(z_p)| + |f_n(z_p) - f_n(z)| < \varepsilon$$

[18]Paul Montel (1876-1937) was a French mathematician.

due to (8.25) and (8.26). The Cauchy criterion implies that the sequence $\{z_n\}$ converges for all $z \in K$ and convergence is uniform on $K$.

(b) Let us show now that any sequence $\{f_n\}$ has a subsequence that converges at every point of a dense subset $E$ of $D$. We choose $E$ as the set $z = x + iy \in D$ with both coordinates $x$ and $y$ rational numbers. This set is clearly countable and dense in $D$, let $E = \{z_\nu\}_{\nu=1}^{\infty}$.

The sequence $f_n(z_1)$ is bounded and hence it has a converging subsequence $f_{k1} = f_{n_k}(z_1)$, $k = 1, 2, \dots$. The sequence $f_{n1}(z_2)$ is also bounded so we may extract its subsequence $f_{k2} = f_{n_k 1}$, $k = 1, 2, \dots$. The sequence $f_{n2}$ converges at least at the points $z_1$ and $z_2$. Then we extract a subsequence $f_{k3} = f_{n_k 2}$ of the sequence $f_{n2}(z_3)$ so that $f_{n3}$ converges at least at $z_1$, $z_2$ and $z_3$. We may continue this procedure indefinitely. It remains to choose the diagonal sequence

$$f_{11}, f_{22}, \dots, f_{nn}, \dots$$

This sequence converges at any point $z_p \in E$ since by construction all its entries after index $p$ belong to the subsequence $f_{np}$ that converges at $z_p$.

Parts (a) and (b) together imply the statement of the theorem. □

The Montel theorem is often called the compactness principle.

**Exercise 8.25** Show that any sequence $\{f_n\}$ of functions holomorphic in a domain $D$ with $\mathrm{Re} f_n \geq 0$ everywhere in $D$ has a subsequence that converges locally uniformly either to a holomorphic function or to infinity.

**Definition 8.26** *A functional $J$ of a family $\{f\}$ of functions defined in a domain $D$ is a mapping $J : \{f\} \to \mathbb{C}$, that is, $J(f)$ is a complex number. A functional $J$ is continuous if given any sequence of functions $f_n \in \{f\}$ that converges uniformly to a function $f_0 \in \{f\}$ on any compact set $K \subset D$ we have*

$$\lim_{n \to \infty} J(f_n) = J(f_0).$$

**Example 8.27** Let $\mathcal{O}(D)$ be the family of all functions $f$ holomorphic in $D$ and let $a$ be an arbitrary point in $D$. Consider the $p$-th coefficient of the Taylor series in $a$:

$$c_p(f) = \frac{f^{(p)}(a)}{p!}.$$

This is a functional on the family $\mathcal{O}(D)$. Let us show that it is continuous. If $f_n \to f_0$ uniformly on every compact set $K \subset D$, we may take $K$ to be the circle $\gamma = \{|z - a| = r\} \subset D$. Then, given any $\varepsilon > 0$ we may find $N$ so that $|f_n(z) - f_0(z)| < \varepsilon$ for all $n > N$ and all $z \in \gamma$. The Cauchy formula for $c_p$

$$c_p = \frac{1}{2\pi i} \int_\gamma \frac{f(z)}{(z-a)^{n+1}} dz$$

implies that

$$|c_p(f_n) - c_p(f_0)| \leq \frac{\varepsilon}{r^n}$$

for all $n > N$ which in turn implies the continuity of the functional $c_p(f)$.

**Definition 8.28** *A compact family of functions $\{f\}$ is sequentially compact if the limit of any sequence $f_n$ that converges uniformly on every compact subset $K \subset D$ belongs to the family $\{f\}$.*

**Theorem 8.29** *Any functional $J$ that is continuous on a sequentially compact family $\{f\}$ is bounded and attains its least upper bound. That is, there exists a function $f_0 \in \{f\}$ so that we have*

$$|J(f_0)| \geq |J(f)|$$

*for all $f \in \{f\}$.*

**Proof.** We let $A = \sup\limits_{f \in \{f\}} |J(f)|$ – this number may be equal to infinity. By definition of the supremum, there exists a sequence $f_n \in \{f\}$ so that $|J(f_n)| \to A$. Since $\{f\}$ is a sequentially compact family there exists a subsequence $f_{n_k}$ that converges to a function $f_0 \in \{f\}$. Continuity of the functional $J$ implies that

$$|J(f_0)| = \lim_{k \to \infty} |J(f_{n_k})| = A.$$

This means that first $A < \infty$ and second, $|J(f_0)| \geq |J(f)|$ for all $f \in \{f\}$. $\square$

We will consider below families of univalent functions in a domain $D$. The following theorem is useful to establish sequential compactness of such families.

**Theorem 8.30** *(Hurwitz[19]) Let a sequence of functions $f_n$ holomorphic in a domain $D$ converge uniformly on any compact subset $K$ of $D$ to a function $f \neq const$. If $f(z_0) = 0$, then given any disk $U_r = \{|z - z_0| < r\}$ there exists $N$ so that all functions $f_n$ vanish at some point in $U_r$ when $n > N$.*

**Proof.** The Weierstrass theorem implies that $f$ is holomorphic in $D$. The uniqueness theorem implies that there exists a punctured disk $\{0 < |z - z_0| \leq \rho\} \subset D$ where $f \neq 0$ (we may assume that $\rho < r$). We denote $\gamma = \{|z - z_0| = \rho\}$ and $\mu = \min\limits_{z \in \gamma} |f(z)|$, and observe that $\mu > 0$. However, $f_n$ converges uniformly to $f$ on $\gamma$ and hence there exists $N$ so that

$$|f_n(z) - f(z)| < \mu$$

for all $z \in \gamma$ and all $n > N$. The Rouche theorem implies that for such $n$ the function $f_n = f + (f_n - f)$ has as many zeros (with multiplicities) as $f$ inside $\gamma$, that is, $f_n$ has at least one zero inside $U_\rho$. $\square$

**Corollary 8.31** *If a sequence of holomorphic and univalent functions $f_n$ in a domain $D$ converges uniformly on every compact subset $K$ of $D$ then the limit function $f$ is either a constant or univalent.*

**Proof.** Assume that $f(z_1) = f(z_2)$ but $z_1 \neq z_2$, $z_{1,2} \in D$ and $f \not\equiv \text{const}$. Consider a sequence of functions $g_n(z) = f_n(z) - f_n(z_2)$ and a disk $\{|z - z_1| < r\}$ with $r < |z_1 - z_2|$. The limit function $g(z)$ vanishes at the point $z_1$. Hence according to the Hurwitz theorem all functions $f_n$ starting with some $N$ vanish in this disk. This, however, contradicts the assumption that $f_n(z)$ are univalent. $\square$

[19] Adolf Hurwitz (1859-1919) was a German mathematician, a student of Weierstrass.

### The Riemann theorem

**Theorem 8.32** *Any simply connected domain $D$ with a boundary that contains more than one point is conformally equivalent to the unit disk $U$.*

**Proof.** The idea of the proof is as follows. Consider the family $S$ of holomorphic and univalent functions $f$ in $D$ bounded by one in absolute value, that is, those that map $D$ into the unit disk $U$. We fix a point $a \in D$ and look for a function $f$ that maximizes the dilation coefficient $|f'(a)|$ at the point $a$. Restricting ourselves to a sequentially compact subset $S_1$ of $S$ and using continuity of the functional $J(f) = |f'(a)|$ we may find a function $f_0$ with the maximal dilation at the point $a$. Finally we check that $f_0$ maps $D$ onto $U$ and not just into $U$ as other functions in $S$.

Such a variational method when one looks for a function that realizes the extremum of a functional is often used in analysis.

(i) Let us show that there exists a holomorphic univalent function in $D$ that is bounded by one in absolute value. By assumption the boundary $\partial D$ contains at least two points $\alpha$ and $\beta$. The square root $\sqrt{\dfrac{z-\alpha}{z-\beta}}$ admits two branches $\phi_1$ and $\phi_2$ that differ by a sign. Each one of them is univalent in $D$[20] since the equality $\phi_\nu(z_1) = \phi_\nu(z_2)$ ($\nu = 1$ or 2) implies

$$\frac{z_1-\alpha}{z_1-\beta} = \frac{z_2-\alpha}{z_2-\beta} \tag{8.27}$$

which implies $z_1 = z_2$ since fractional linear transformations are univalent. The two branches $\phi_1$ and $\phi_2$ map $D$ onto domains $D_1^* = \phi_1(D)$ and $D_2^* = \phi_2(D)$ that have no overlap. Otherwise there would exist two points $z_{1,2} \in D$ so that $\phi_1(z_1) = \phi_2(z_2)$ which would in turn imply (8.27) so that $z_1 = z_2$ and then $\phi_1(z_1) = -\phi_2(z_2)$. This is a contradiction since $\phi_\nu(z) \neq 0$ in $D$.

The domain $D_2^*$ contains a disk $\{|w - w_0| < \rho\}$. Hence $\phi_1$ does not take values in this disk. Therefore the function

$$f_1(z) = \frac{\rho}{\phi_1(z) - w_0} \tag{8.28}$$

is clearly holomorphic and univalent in $D$ and takes values inside the unit disk: we have $|f_1(z)| \leq 1$ for all $z \in D$.

(ii) Let us denote by $S$ the family of functions that are holomorphic and univalent in $D$, and are bounded by one in absolute value. This family is not empty since it contains the function $f_1$. It is compact by the Montel theorem. The subset $S_1$ of the family $S$ that consists of all functions $f \in S$ such that

$$|f'(a)| \geq |f_1'(a)| > 0 \tag{8.29}$$

at some fixed point $a \in D$ is sequentially compact. Indeed Corollary 8.31 implies that the limit of any sequence of functions $f_n \in S_1$ that converges on any compact subset $K$

[20]In general we may define a univalent branch of $\sqrt{\dfrac{z-a}{z-b}}$ in a domain $D$ if neither $a$ nor $b$ are in $D$.

of $D$ may be only a univalent function (and hence belong to $S_1$) or be a constant but the latter case is ruled out by (8.29).

Consider the functional $J(f) = |f'(a)|$ defined on $S_1$. It is a continuous functional as was shown in Example 8.27. Therefore there exists a function $f_0 \in S$ that attains its maximum, that is, such that

$$|f'(a)| \leq |f_0'(a)| \tag{8.30}$$

for all $f \in S$.

(iii) The function $f_0 \in S_1$ maps $D$ conformally into the unit disk $U$. Let us show that $f_0(a) = 0$. Otherwise, the function

$$g(z) = \frac{f_0(z) - f_0(a)}{1 - \overline{f_0(a)} f_0(z)}$$

would belong to $S_1$ and have

$$|g'(a)| = \frac{1}{1 - |f_0(a)|^2} |f_0'(a)| > |f_0'(a)|,$$

contrary to the extremum property (8.30) of the function $f$.

Finally, let us show that $f_0$ maps $D$ onto $U$. Indeed, let $f_0$ omit some value $b \in U$. Then $b \neq 0$ since $f_0(a) = 0$. However, the value $b^* = 1/b$ is also not taken by $f_0$ in $D$ since $|b^*| > 1$. Therefore one may define in $D$ a single valued branch of the square root

$$\psi(z) = \sqrt{\frac{f_0(z) - b}{1 - \bar{b} f_0(z)}} \tag{8.31}$$

that also belongs to $S$: it is univalent for the same reason as in the square root in part (i), and $|\psi(z)| \leq 1$. However, then the function

$$h(z) = \frac{\psi(z) - \psi(a)}{1 - \overline{\psi(a)} \psi(z)}$$

also belongs to $S$. We have $|h'(a)| = \dfrac{1 + |b|}{2\sqrt{|b|}} |f_0'(a)|$. However, $1 + |b| > 2\sqrt{|b|}$ since $|b| < 1$ and thus $h \in S_1$ and $|h'(a)| > |f_0'(a)|$ contrary to the extremal property of $f_0$. $\square$

The Riemann theorem implies that any two simply connected domains $D_1$ and $D_2$ with boundaries that contain more than one point are conformally equivalent. Indeed, as we have shown there exist conformal isomorphisms $f_j : D_j \to U$ of these domains onto the unit disk. Then $f = f_2^{-1} \circ f_1$ is a conformal isomorphism between $D_1$ and $D_2$. Theorem 8.20 implies that an isomorphism $f : D_1 \to D_2$ is uniquely determined by a normalization

$$f(z_0) = w_0, \quad \arg f'(z_0) = \theta, \tag{8.32}$$

where $z_0 \in D_1$, $w_0 \in D_2$ and $\theta$ is a real number.

# 9 Elliptic functions

We will be following here the book by Stein and Shakarchi, Ahlfors and the Whittaker-Watson classic.

## The fundamental parallelogram

An elliptic function is a doubly-periodic meromorphic function. More precisely, there exist two complex numbers $\omega_1$ and $\omega_2$ so that

$$f(z+\omega_1) = f(z), \text{ and } f(z+\omega_2) = f(z), \tag{9.1}$$

for all $z \in \mathbb{C}$. In order for this definition to be interesting, $\omega_1$ and $\omega_2$ should be linearly independent over $\mathbb{R}$. Otherwise, as the reader can check, if the ratio $\omega_1/\omega_2$ is a (real) rational number, condition (9.1) simply means that the function $f(z)$ is periodic. On the other hand, if the ratio $\omega_1/\omega_2$ is an irrational real number, the function $f$ has to be identically equal to a constant.

The periods of a doubly-periodic function $f$ form a module: if $\omega$ is a period, then $n\omega$ is a period for any $n \in \mathbb{C}$ and if $\omega_1$ and $\omega_2$ are periods, then $\omega_1 + \omega_2$ is also a period. It is easy to see that unless $f$ is constant, the module $M$ of its periods consists of discrete points, and we have the following.

**Proposition 9.1** *A discrete module of complex numbers consists either of $\omega = 0$ alone, of the integral multiples $n\omega$, $n \in \mathbb{Z}$ of a given complex number $\omega$, or of all linear combinations $n_1\omega_1 + n_2\omega_2$, with $n_1, n_2 \in \mathbb{Z}$ of two complex numbers $\omega_1$, $\omega_2$ with a non-real ratio $\omega_1/\omega_2$.*

*Proof.* If $M$ contains a non-zero number, it has to contain a number $\omega_1$ that has the smallest absolute value of all $\omega \in M$. Assume that $M$ also contains a number $\omega$ that is not an integer multiple of $\omega_1$. Choose the smalest in absolute value number $\omega_2 \neq \omega_1$ which is in $M$ but is not a multiple of $\omega_1$. We claim that any $\omega \in M$ has the form $\omega = n_1\omega_1 + n_2\omega_2$ with $n_1, n_2 \in \mathbb{Z}$. Indeed, we know that any $\omega$ can be written as

$$\omega = \lambda_1\omega_1 + \lambda_2\omega_2,$$

with $\lambda_1, \lambda_2 \in \mathbb{R}$. Choose integers $n_1$ and $n_2$ so that $|\lambda_1 - n_1| \leq 1/2$ and $|\lambda_2 - n_2| \leq 1/2$. As $\omega$ is in $M$, so is

$$\omega' = \omega - n_1\omega_1 - n_2\omega_2.$$

It follows that

$$|\omega'| < \frac{1}{2}|\omega_1| + \frac{1}{2}|\omega_2| \leq |\omega_2|.$$

The first inequality above is strict since $\omega_1$ and $\omega_2$ are linearly independent over $\mathbb{R}$. It follows from the way $\omega_2$ was chosen that either $\omega' = 0$ (and hence $\omega$ has the required form), or $\omega$ is a real multiple of $\omega_1$. Then $\omega$ has to be an integer multiple of $\omega_1$ – this follows from how $\omega_1$ was chosen. □

Note that simply by defining $F(z) = f(\omega_1 z)$ by obtain a function $F(z)$ that satisfies $F(z+1) = F(z)$, and any other period $\omega$ of $F$ satisfies $|\omega| \geq 1$. It will be often convenient to adopt this normalization, and we will denote $\tau = \omega_2/\omega_1$. The lattice of periods in this case is given by

$$\Lambda = \{n + m\tau : \ n, m \in \mathbb{Z}\}.$$

Its fundamental parallelogram is

$$P_0 = \{z \in \mathbb{C} : \ z = a + b\tau, \quad \text{with } 0 \leq a < 1 \text{ and } 0 \leq b < 1\}.$$

The behavior of $f$ in all of $\mathbb{C}$ is completely determined by its behavior in $P_0$. Indeed, we say that $z \sim w$ ($z$ is congruent to $w$) if $z - w \in \Lambda$. We claim that for each $z \in \mathbb{C}$ there exists a unique $w_0 \in P_0$ such that $z \sim w_0$. It is constructed as in the proof of Proposition 9.1: choose $a, b \in \mathbb{R}$ so that $z = a + b\tau$, and let $n, m$ be the greatest integers that are less or equal to $a$ and $b$, respectively. Then $w = a - n + (b - m)\tau$ is congruent to $z$ and lies in $P_0$. For uniqueness, assume that $z \sim w = a + b\tau$ and $z \sim w' = a' + b'\tau$, with both $w$ and $w'$ in $P_0$. Then $w \sim w'$, so we have

$$w - w' = n + m\tau,$$

with integer $n$ and $m$, but we also have

$$w - w' = a - a' + (b - b')\tau,$$

with $0 \leq a, a' < 1$ and $0 \leq b, b' < 1$. It follows that $n = a - a' = 0$ and $m = b - b' = 0$. Therefore, the values of $f$ everywhere in $\mathbb{C}$ are uniquely determined by its values on $P_0$.

## The basic properties of the elliptic functions

A consequence of the above observation is Liouville's theorem.

**Theorem 9.2** *An entire elliptic function is a constant.*

*Proof.* If $f$ is an entire elliptic function then $f$ is bounded on $P_0$, hence it is bounded on all of $\mathbb{C}$ hence it is a constant. □

Therefore, an elliptic function must have poles in the fundamental parallelogram. As usual, we will count them with multiplicities. As elliptic functions are meromorphic (their only singularities are poles), it can have only finitely many poles inside $P_0$. They can lie on the boundary of $P_0$ but since there are only finitely many of poles and zeros in $P_0$, we may always consider a shift: $F(z) = f(z+a)$ so that the shifted function $F(z)$ has neither zeros nor poles on the boudnary of $P_0$.

**Theorem 9.3** *The sum of residues of an elliptic function is zero.*

*Proof.* We may assume without loss of generality that there are no poles on the boundary of $P_0$. Then, the sum of the residues of $f$ inside $P_0$ is

$$\sum \text{Res} f = \frac{1}{2\pi} \int_{\partial P_0} f dz,$$

with $P_0$ traced counterclockwise, and the sum taken over all poles of $f$. However, due to periodicity of $f$, the integrals over the opposite sides cancel each other, hence

$$\sum \text{Res} f = 0,$$

and we are done. □

**Corollary 9.4** *An elliptic function has at least two poles in $P_0$ (counting with multiplicity).*

Indeed, $f$ has to have at least one pole in $P_0$, and satisfy

$$\sum \text{Res} f = 0.$$

It follows that it can not have a single pole of order one.

We say that an elliptic function has order $m$ if it has $m$ poles in $P_0$ (counted with their multiplicity).

**Theorem 9.5** *An elliptic function of order $m$ has $m$ zeros (with multiplicities).*

*Proof.* Without loss of generality we may assume that $f$ has neither poles nor zeros on the boundary of $P_0$. The argument principle implies that

$$N_p - N_z = \frac{1}{2\pi} \int_{\partial P_0} \frac{f'(z)}{f(z)} dz.$$

Here $N_p$ and $N_z$ are the number of poles and zeros of $f$ inside $P_0$, respectively. Periodicity of $f$ once again implies that the integrals over the opposite sides of the fundamental parallelogram cancel each other, whence $N_p = N_z$. □

**Theorem 9.6** *The zeros $a_1, \dots, a_n$ and the poles $b_1, \dots, b_n$ of an elliptic function satisfy*

$$a_1 + \cdots + a_n = b_1 + \cdots + b_n,$$

*modulo the period lattice.*

*Proof.* Once again, without loss of generality we may assume that $f$ has neither poles nor zeros on the boundary of $P_0$. Consider the integral

$$\frac{1}{2\pi} \int_{\partial P_0} \frac{z f'(z) dz}{f(z)}.$$

It is given by (recall that the residue of the function $f'(z)/f(z)$ at a zero $a_k$ is the multiplicity of $a_k$, and its residue at a pole $b_k$ is the order of $b_k$):

$$\frac{1}{2\pi i} \int_{\partial P_0} \frac{z f'(z) dz}{f(z)} = a_1 + \cdots + a_n - b_1 - \cdots - b_n.$$

Let us now inspect the integral on each side of $P_0$, and without loss of generality we assume that $P_0$ has the form (that is, $f(z)$ has basic periods 1 and $\tau$)

$$P_0 = \{z \in \mathbb{C} : \; z = a + b\tau, \quad \text{with } 0 \le a < 1 \text{ and } 0 \le b < 1\},$$

and its four sides are $(0,1), (1, 1+\tau), (1+\tau, \tau), (\tau, 0)$. Let us look at

$$\frac{1}{2\pi i}\left(\int_0^1 - \int_\tau^{a+\tau}\right)\frac{zf'(z)dz}{f(z)} = -\frac{\tau}{2\pi i}\int_0^1 \frac{f'(z)dz}{f(z)}.$$

Except for the factor of $\tau$, the right side is winding number of the curve traced by $f(z)$ as $z$ varies from $z = 0$ to $z = 1$. As $f(0) = f(1)$, this curve is closed, hence the winding number is an integer. A similar argument applies to

$$\frac{1}{2\pi i}\left(\int_1^{1+\tau} - \int_0^{\tau}\right)\frac{zf'(z)dz}{f(z)} = -\frac{1}{2\pi i}\int_1^{1+\tau} \frac{f'(z)dz}{f(z)}.$$

It follows that

$$a_1 + \cdots + a_n - b_1 - \cdots - b_n = m_1 + m_2\tau,$$

with some integers $m_{1,2}$, and the proof is complete. □

### The Weierstrass $\wp$ function

We have shown that no elliptic function can have one simple pole, and the next simplest step would be to construct an elliptic function that has one double pole. If we were to construct a 1-periodic function, not a doubly periodic function, a natural choice would be

$$F(z) = \sum_{n=-\infty}^{\infty} \frac{1}{z+n},$$

which is 1-periodic by inspection, and has a pole at all integers. The problem is that the series defining $F(z)$ does not converge absolutely. The solution to this would be to re-write $F(z)$ as

$$F(z) = \frac{1}{z} + \sum_{n=1}^{\infty}\left(\frac{1}{z+n} + \frac{1}{z-n}\right),$$

which converges absolutely. Another trick would be to write

$$\frac{1}{z+n} + \frac{1}{z-n} = \left(\frac{1}{z+n} - \frac{1}{n}\right) + \left(\frac{1}{z-n} + \frac{1}{n}\right),$$

which would lead to

$$F(z) = \frac{1}{z} + \sum_{n=1}^{\infty}\left(\frac{1}{z+n} - \frac{1}{n}\right) + \sum_{n=1}^{\infty}\left(\frac{1}{z+(-n)} - \frac{1}{(-n)}\right) = \frac{1}{z} + \sum_{n\neq 0}\left(\frac{1}{z+n} - \frac{1}{n}\right).$$

The idea is to construct an elliptic function with a double pole at zero following the second recipe above. Let $\Lambda$ be the period lattice, if we try to define

$$\sum_{\omega\in\Lambda}\frac{1}{(z+\omega)^2},$$

we get a series that does not converge absolutely. The remedy is to consider instead

$$\wp(z)=\frac{1}{z^2}+\sum_{\omega\in\Lambda^*}\left(\frac{1}{(z+\omega)^2}-\frac{1}{\omega^2}\right). \tag{9.2}$$

Here $\Lambda^*$ is the period lattice without the point $(0,0)$. More explicitly, we may write

$$\wp(z)=\frac{1}{z^2}+\sum_{(n,m)\neq(0,0)}\left(\frac{1}{(z+n+m\tau)^2}-\frac{1}{(n+m\tau)^2}\right). \tag{9.3}$$

Our task is now to show that (1) the series defining the function $\wp(z)$ converges, (2) the function $\wp(z)$ is meromorphic and has a double pole at all $\omega\in\Lambda$, and only at those points, and (3) that $\wp(z)$ is doubly periodic with the periods $\omega_1=1$ and $\omega_2=\tau$.

In order to see that the series converges, consider a disk $\{|z|\leq R\}$ and notice that for any $z$ with $|z|<R$ we can write

$$\wp(z)=\frac{1}{z^2}+\sum_{0<|\omega|\leq 2R}\left(\frac{1}{(z+\omega)^2}-\frac{1}{\omega^2}\right)+\sum_{|\omega|>2R}\left(\frac{1}{(z+\omega)^2}-\frac{1}{\omega^2}\right). \tag{9.4}$$

The first term above has finitely many terms and is meromorphic in $\{|z|<R\}$: it has double poles at the lattice points $\omega$ with $|\omega|<R$. The second term is holomorphic in $\{|z|<R\}$: there exists $C>$ so that for any $|z|<R$, $|\omega|>2R$ we have

$$\left|\frac{1}{(z+\omega)^2}-\frac{1}{\omega^2}\right|=\left|\frac{z^2+2z\omega}{(z+\omega)^2\omega^2}\right|\leq\frac{CR}{|\omega|^3},$$

and we have (see Stein-Shakarchi for a detailed proof of convergence of this series):

$$\sum_{\omega\in\Lambda^*}\frac{1}{|\omega|^3}<+\infty.$$

Therefore, indeed, $\wp(z)$ is well-defined and has double poles exactly at the points $z\in\Lambda$.

Let us now check that $\wp(z)$ is doubly-periodic. This is done in an indirect way. The derivative $\wp'(z)$ is

$$\wp'(z)=-2\sum_{\omega\in\Lambda}\frac{1}{(z-\omega)^3}. \tag{9.5}$$

Note that this series converges absolutely and clearly defines a doubly periodic function. It follows that there exist two constants $a$ and $b$ so that

$$\wp(z+1)-\wp(z)=a,\quad \wp(z+\tau)-\wp(z)=b,\ \text{for all } z\in\mathbb{C}.$$

The definition of $\wp(z)$ implies that it is an even function. Then, taking $z = -1/2$ implies that $a = 0$ while taking $z = -\tau/2$ implies that $b = 0$.

Let us now investigate the derivative $\wp'(z)$. First, as $\wp(z)$ is even, its derivative is both odd and doubly periodic. It follows that

$$\wp'(1/2) = -\wp'(-1/2) = -\wp'(-1/2+1) = -\wp'(1/2),$$

hence $\wp'(1/2) = 0$, and similarly, we can show that $\wp'(\tau/2) = \wp'((1+\tau)/2) = 0$. As the function $\wp'(z)$ is elliptic and has a single triple pole, it follows that it has no other zeros except for $1/2$, $\tau/2$ and $(1+\tau)/2$. Let us set

$$e_1 = \wp(1/2), \quad e_2 = \wp(\tau/2), \quad e_3 = \wp\left(\frac{1+\tau}{2}\right).$$

Equation $\wp(z) = e_1$ has a double root at $z = 1/2$, and similarly the equation $\wp(z) = e_2$ has a double root at $z = \tau/2$, while $\wp(z) = e_3$ has a double root at $z = (1+\tau)/2$. The numbers $e_{1,2,3}$ have to be distinct for otherwise, say, the equation $\wp(z) = e_1$ would have four solutions (with multiplicities) which would imply that the elliptic function $\wp(z) - e_1$ would have four poles in the fundamental domain (also with multiplicity), while it only has two poles.

**Theorem 9.7** *The function $(\wp')^2$ is a cubic polynomial in $\wp$:*

$$(\wp'(z))^2 = 4(\wp(z) - e_1)(\wp(z) - e_2)(\wp(z) - e_3). \tag{9.6}$$

*Proof.* The functions

$$F(z) = (\wp(z) - e_1)(\wp(z) - e_2)(\wp(z) - e_3)$$

and $(\wp'(z))^2$ have the same roots in the fundamental domain, with the same multiplicity two. In addition, they both have poles of order six at the vertices of the lattice $\Lambda$. The function $F(z)/(\wp'(z))^2$ is, therefore, a holomorphic double-periodic function, whence a constant $B$. To find $B$, note that for $z$ close to zero we have

$$\wp(z) = \frac{1}{z^2} + \ldots, \quad \wp'(z) = -\frac{2}{z^3},$$

which means that $B = 1/4$. □

As a consequence, we have the expression for $\wp(z)$ as the inverse of an elliptic integral:

$$z - z_0 = \int_{\wp(z_0)}^{\wp(z)} \frac{dw}{\sqrt{4(w-e_1)(w-e_2)(w-e_3)}}.$$

**Proposition 9.8** *Every even elliptic function with periods $1$ and $\tau$ is a rational function of $\wp(z)$.*

*Proof.* First, if $F(z)$ has a zero or a pole on the lattice $\Lambda$, we may consider $F_1(z) = F(z)\wp^m(z)$, with an integer $m$ so that $F_1(z)$ has no pole or zero on the lattice. The function $F$ is even, hence if $F(a) = 0$ then $F(-a) = 0$. Moreover, as in the argument for $\wp'(z)$, a zero $a$ has an even order if $a$ is a half-period. The same is true for $\wp(z) - \wp(a)$ – it has an even order zero at $a$ if and only if $a$ is a half-period. Therefore, if $a_1, -a_1, a_2, -a_2, \ldots, a_n, -a_n$ are the zeros of $F$ then the product

$$(\wp(z) - \wp(a_1))(\wp(z) - \wp(-a_1)) \ldots (\wp(z) - \wp(a_n))(\wp(z) - \wp(-a_n))$$

has exactly the same zeros as $F$, with the same multiplicities. The same argument applies to the poles $b_1, -b_1, \ldots, b_n, -b_n$. As a consequence, the ratio $F/G$, with

$$G(z) = \frac{(\wp(z) - \wp(a_1))(\wp(z) - \wp(-a_1)) \ldots (\wp(z) - \wp(a_n))(\wp(z) - \wp(-a_n))}{(\wp(z) - \wp(b_1))(\wp(z) - \wp(-b_1)) \ldots (\wp(z) - \wp(b_n))(\wp(z) - \wp(-b_n))},$$

is a bounded elliptic function, hence a constant. □

**Theorem 9.9** *Every elliptic function is a rational function of $\wp$ and $\wp'$.*

*Proof.* Recall that $\wp'(z)$ is an odd function. Given an elliptic function $f(z)$ we write it as $f(z) = h(z) + g(z)$ with an even elliptic function $h$ and an odd elliptic function $g$. It follows from the previous proposition that $h$ is a rational function of $\wp$, but also that the (even) ratio $g/\wp'$ is a rational function of $\wp$. □

We now get the addition rule for the Weierstrass function. Consider the equations

$$\wp'(z) = A\wp(z) + B, \quad \wp'(y) = A\wp(y) + B,$$

which determine $A$ and $B$ as functions of $z$ and $y$ (unless $\wp(z) = \wp(y)$, that is, unless $z = \pm y (\mathrm{mod}(1, \tau))$. Next, consider the function

$$\wp'(\zeta) - A\wp(\zeta) - B$$

as a function of $\zeta$. It is an elliptic function that has a triple pole at $\zeta = 0$, hence it has exactly three zeros. As the sum of all zeros equals to the sum of all poles (modulo the period lattice), the sum of all zeros equals to zero. Two of the zeros are $z$ and $y$, hence the third zero is equal to $-z - y$ (modulo the period lattice), that is:

$$\wp'(-z - y) = A\wp(-z - y) + B.$$

It follows (using the fact that $\wp$ is even and $\wp'$ is odd) that the following determinant vanishes:

$$\begin{vmatrix} \wp(z) & \wp'(z) & 1 \\ \wp(y) & \wp'(y) & 1 \\ \wp(z + y) & -\wp'(z + y) & 1 \end{vmatrix} = 0.$$

The derivatives that appear above can be expressed as functions of $\wp(z)$, $\wp(y)$ and $\wp(z + y)$ using the differential equation for $\wp'$. This express algebraically $\wp(z + y)$ in terms of $\wp(z)$ and $\wp(y)$.

# 10 The Theta functions

## The Jacobi theta function

The Jacobi theta function is defined as

$$\Theta(z|\tau) = \sum_{n=-\infty}^{\infty} e^{\pi i n^2 \tau} e^{2\pi i n z}. \tag{10.1}$$

Here $z \in \mathbb{C}$ and $\operatorname{Im}\tau > 0$, so that the integral converges. Note that when $\tau = it$ the function $\Theta(z, it)$ becomes the heat kernel

$$\Theta(z|it) = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 t} e^{2\pi i n z},$$

providing a link between the complex analysis, number theory and PDEs.

Our first goal is to obtain a different expression for $\Theta(z|\tau)$, known as the product formula.

**Theorem 10.1** *For all $z \in \mathbb{C}$ and $\tau \in \mathbb{H} = \{\operatorname{Im}\tau > 0\}$ we have*

$$\Theta(z|\tau) = \prod_{n=1}^{\infty}(1-q^{2n})(1+q^{2n-1}e^{2\pi i z})(1+q^{2n-1}e^{-2\pi i z}), \tag{10.2}$$

*where $q = e^{\pi i \tau}$.*

This will be done in several steps. First, we establish some basic properties of $\Theta(z|\tau)$.

**Proposition 10.2** *The function $\Theta(z|\tau)$ satisfies the following properties:*
*(i) $\Theta$ is entire in $z \in \mathbb{C}$ and holomorphic in $\tau \in \mathbb{H}$.*
*(ii) $\Theta(z+1|\tau) = \Theta(z|\tau)$,*
*(iii) $\Theta(z+\tau|\tau) = \Theta(z|\tau)e^{-\pi i \tau}e^{-2\pi i z}$.*
*(iv) $\Theta(z|\tau) = 0$ whenever $z = 1/2 + \tau/2 + n + m\tau$ and $n, m \in \mathbb{Z}$.*

*Proof.* To prove the first claim, note that if $\operatorname{Im}\tau = t \geq t_0 > 0$, and $|z| \leq R$, then

$$\left| e^{2\pi i n z} e^{\pi i n^2 \tau} \right| \leq e^{2\pi |n| R - \pi n^2 t_0},$$

hence the series defining $\Theta(z|\tau)$ is absolutely convergent. Therefore, $\Theta(z|\tau)$ is holomorphic both in $z$ and $\tau$ on the sets $\{|z| \leq R,\ \operatorname{Im}\tau \geq t_0\}$, and the conclusion of part (i) follows. The property (ii) is imply the consequence of the 1-periodicity of the exponentials $e^{2\pi i n z}$, $n \in \mathbb{Z}$. Finally, to prove (iii) we need to compute:

$$\begin{aligned}\Theta(z+\tau|\tau) &= \sum_{n=-\infty}^{\infty} e^{\pi i n^2 \tau} e^{2\pi i n (z+\tau)} = \sum_{n=-\infty}^{\infty} e^{\pi i (n^2+2n)\tau} e^{2\pi i n z} \\ &= e^{-\pi i \tau} \sum_{n=-\infty}^{\infty} e^{\pi i (n+1)^2 \tau} e^{2\pi i n z} = e^{-\pi i \tau} e^{2\pi i z} \Theta(z|\tau).\end{aligned}$$

Finally, property (iv) will follow from (ii) and (iii) if we show that

$$\Theta\left(\frac{1+\tau}{2}|\tau\right) = 0. \tag{10.3}$$

This is seen as follows: we have

$$\Theta(\frac{1}{2}+\frac{\tau}{2}|\tau) = \sum_{n=-\infty}^{\infty} e^{\pi i n^2\tau} e^{2\pi i n(1/2+\tau/2)} = \sum_{n=-\infty}^{\infty} (-1)^n e^{\pi i(n^2+n)\tau}$$

$$= \sum_{n\geq 0}(-1)^n e^{\pi i(n^2+n)\tau} + \sum_{n=-\infty}^{-1} (-1)^n e^{\pi i(n^2+n)\tau}$$

$$= \sum_{n\geq 0}(-1)^n e^{\pi i(n^2+n)\tau} + \sum_{n=-\infty}^{0} (-1)^{n-1} e^{\pi i((n-1)^2+n-1)\tau}$$

$$= \sum_{n\geq 0}(-1)^n e^{\pi i(n^2+n)\tau} + \sum_{n\geq 0}^{0}(-1)^{n-1} e^{\pi i((n+1)^2-n-1)\tau} = 0,$$

and we are done. □

The next step in the proof of Theorem 10.1 is to show that the right side of (10.2) satisfies exactly the same properties as $\Theta(z|\tau)$. Let us set

$$\Pi(z|\tau) = \prod_{n=1}^{\infty}(1-q^{2n})(1+q^{2n-1}e^{2\pi iz})(1+q^{2n-1}e^{-2\pi iz}), \tag{10.4}$$

with $q = e^{\pi i\tau}$.

**Proposition 10.3** *The function $\Pi(z|\tau)$ satisfies the following properties:*
*(i) $\Pi$ is entire in $z \in \mathbb{C}$ and holomorphic in $\tau \in \mathbb{H}$.*
*(ii) $\Pi(z+1|\tau) = \Pi(z|\tau)$,*
*(iii) $\Pi(z+\tau|\tau) = \Pi(z|\tau)e^{-\pi i\tau}e^{-2\pi iz}$.*
*(iv) $\Pi(z|\tau) = 0$ whenever $z = 1/2+\tau/2+n+m\tau$ and $n, m \in \mathbb{Z}$. Moreover, these are simple zeros of $\Pi(z|\tau)$, and $\Pi(z|\tau)$ has no other zeros.*

*Proof.* First, if $t = \operatorname{Im}\tau \geq t_0 > 0$ and $|z| \leq R$, then $|q| < e^{-\pi t_0}$, and

$$|(1-q^{2n})(1+q^{2n-1}e^{2\pi iz})(1+q^{2n-1}e^{-2\pi iz}) - 1| \leq C|q|^{2n-1}e^{2\pi R},$$

hence the infinite product converges and defines a function that is entire in $z$ and holomorphic for $\tau \in \mathbb{H}$. Furthermore, 1-periodicity of $\Pi(z|\tau)$ in $z$ follows immediately from its definition. In order to prove (iii) we compute, using $q^2 = e^{2\pi i\tau}$:

$$\Pi(z+\tau|\tau) = \prod_{n=1}^{\infty}(1-q^{2n})(1+q^{2n-1}e^{2\pi i(z+\tau)})(1+q^{2n-1}e^{-2\pi i(z+\tau)})$$

$$= \prod_{n=1}^{\infty}(1-q^{2n})(1+q^{2n+1}e^{2\pi iz})(1+q^{2n-3}e^{-2\pi iz}) = \Pi(z|\tau)\frac{1+q^{-1}e^{-2\pi iz}}{1+qe^{2\pi iz}}$$

$$= \Pi(z|\tau)q^{-1}e^{-2\pi iz},$$

which proves (iii). In order to see (iv) recall that an infinite product vanishes if and only if one of the factors vanishes. As $|q| < 1$ for $\tau \in \mathbb{H}$, the only possibility for $\Pi(z|\tau)$ to vanish is that

$$1 + q^{2n-1}e^{2\pi iz} = 0,$$

or

$$1 + q^{2n-1}e^{-2\pi iz} = 0.$$

Since $q = e^{\pi i\tau}$, these can be re-written as

$$(2n-1)\tau + 2z = 1 \ (\mathrm{mod}2), \ \text{or} \ (2n-1)\tau - 2z = 1 \ (\mathrm{mod}2),$$

that is,

$$z = \frac{1}{2} + \frac{\tau}{2} - n\tau \ (\mathrm{mod}1), \ \text{or} \ z = -\frac{1}{2} - \frac{\tau}{2} + n\tau \ (\mathrm{mod}1),$$

which is the set of zeros claimed in (iv). It is easy to check that they are all simple zeros. □

Now, we can prove Theorem 10.1. Let us fix $\tau \in \mathbb{H}$. Consider the ratio $F(z) = \Theta(z|\tau)/\Pi(z|\tau)$. According to the last two propositions, this function is entire and is doubly periodic, with periods 1 and $\tau$. It follows that $F(z)$ is a constant, which we will denote by $c(\tau)$:

$$\Theta(z|\tau) = c(\tau)\Pi(z|\tau). \tag{10.5}$$

Our goal is to show that $c(\tau) \equiv 1$. Taking $z = 1/2$ in (10.5) (hence $e^{2\pi iz} = e^{-2\pi iz} = -1$) gives

$$\sum_{n=-\infty}^{\infty} e^{\pi in^2\tau}(-1)^n = c(\tau)\prod_{n=1}^{\infty}(1-q^{2n})(1-q^{2n-1})(1-q^{2n-1}),$$

or

$$\sum_{n=-\infty}^{\infty}(-1)^n q^{n^2} = c(\tau)\prod_{n=1}^{\infty}(1-q^{2n})(1-q^{2n-1})(1-q^{2n-1}) = c(\tau)\prod_{n=1}^{\infty}(1-q^n)(1-q^{2n-1}).$$

The last identity is obtained simply by noticing that $2n-1$ runs over all odd positive integers, and $2n$ over all even ones. It follows that

$$c(\tau) = \frac{\sum_{n=-\infty}^{\infty}(-1)^n q^{n^2}}{\prod_{n=1}^{\infty}(1-q^n)(1-q^{2n-1})}. \tag{10.6}$$

Next, we take $z = 1/4$ in (10.5), with $e^{2\pi iz} = i$:

$$\Theta(1/4|\tau) = \sum_{n=-\infty}^{\infty} q^{n^2} i^n = \sum_{n\ even} q^{n^2} i^n = \sum_{n=-\infty}^{\infty} q^{4n^2}(-1)^n.$$

And we also have

$$\Pi(\frac{1}{4}, \tau) = \prod_{n=1}^{\infty}(1-q^{2n})(1+iq^{2n-1})(1-iq^{2n-1}) = \prod_{n=1}^{\infty}(1-q^{2n})(1+q^{4n-2})$$

$$= \prod_{n=1}^{\infty}(1-q^{4n})(1-q^{4n-2})(1+q^{4n-2}) = \prod_{n=1}^{\infty}(1-q^{4n})(1-q^{8n-4}).$$

We conclude that

$$c(\tau) = \frac{\sum_{n=-\infty}^{\infty}(-1)^n q^{4n^2}}{\prod_{n=1}^{\infty}(1-q^{4n})(1-q^{8n-4})}. \tag{10.7}$$

Comparing (10.6) and (10.7) and recalling that $q(4\tau) = q^4(\tau)$, we conclude that

$$c(\tau) = c(4\tau), \tag{10.8}$$

and, as consequence, $c(\tau) = c(4^k\tau)$ for all $k \in \mathbb{Z}$. We also have $q(4^k\tau) \to 0$ as $k \to +\infty$, for any $\tau \in \mathbb{H}$. It follows from (10.7) that $c(\tau) = 1$, and the proof of Theorem 10.1 is complete.

Theorem 10.1 also shows a link between the $\Theta$ function and the Weierstrass $\wp$ function.

**Corollary 10.4** *For each $\tau \in \mathbb{H}$ fixed, the function*

$$(\log\Theta(z|\tau))'' = \frac{\Theta(z|\tau)\Theta''(z|\tau) - (\Theta'(z|\tau))^2}{\Theta(z|\tau)^2} \tag{10.9}$$

*is an elliptic function of order* 2 *with periods* 1 *and* $\tau$, *and with a double pole at* $z = 1/2 + \tau/2$.

*Proof.* Let

$$F(z) = (\log(\Theta(z|\tau))' = \frac{\Theta'(z|\tau)}{\Theta(z|\tau)},$$

then Proposition 10.3 implies that $F(z+1) = F(z)$, and

$$F(z+\tau) = F(z) - 2\pi i,$$

so that $F''(z)$ is doubly periodic. Since $\Theta(z|\tau)$ vanishes only at $z = 1/2 + \tau/2$ in the fundamental parallelogram, and that zero is simple, the function $F(z)$ has a single pole at this point, meaning that $F''(z)$ has a double pole there. □

As a consequence of this corollary, $F(z) = \wp(z - \tau - 1/2) + b_\tau$, with some constant $b_\tau$ that can be computed in terms of the values of the first several terms of the Taylor series for $\Theta(z|\tau)$ at $z = 1/2 + \tau/2$.

## Generating functions and partitions of integers

We will now study some applications of the theta functions to the number theory. An important tool in this study is the generating function: given a sequence $F_n$ we may often infer its properties from the behavior of the function

$$F(x) = \sum_{n=0}^{\infty} F_n x^n.$$

A simple example of this comes from the Fibonacci sequence defined by $F_0 = F_1 = 1$, and

$$F_n = F_{n-1} + F_{n-2}.$$

It is easy to check that the function

$$F(x) = \sum_{n=0}^{\infty} F_n x^n$$

satisfies

$$F(x) = x^2 F(x) + xF(x) + x,$$

so that

$$F(x) = \frac{x}{1 - x - x^2} = \frac{A}{1 - \alpha x} + \frac{B}{1 - \beta x},$$

where

$$\alpha = \frac{1+\sqrt{5}}{2}, \quad \beta = \frac{1-\sqrt{5}}{2}, \quad A = \frac{1}{\alpha - \beta}, \quad B = \frac{1}{\beta - \alpha}.$$

It follows that the general formula for the Fibonacci sequence is

$$F_n = A\alpha^n + B\beta^n.$$

We will now use this method to find the formula for the partition function defined as follows: for an integer $n$ we denote by $p(n)$ the number of ways $n$ can be written as a sum of positive integers. For example, $p(1) = 1$, $p(2) = 2$ and $p(3) = 3$ because

$$3 = 3 + 0 = 2 + 1 = 1 + 1 + 1,$$

while $p(4) = 5$, because

$$4 = 4 + 0 = 3 + 1 = 2 + 2 = 1 + 1 + 1 + 1 = 1 + 1 + 2,$$

and so on. By convention, we set $p(0) = 0$.

**Theorem 10.5** *(Euler) If $|x| < 1$ then*

$$\sum_{n=0}^{\infty} p(n) x^n = \prod_{k=1}^{\infty} \frac{1}{1 - x^k}. \tag{10.10}$$

Note that

$$\frac{1}{1 - x^k} = 1 + O(x^k),$$

hence the product in the right side of (10.10) converges. Moreover, as each term can be written as

$$\frac{1}{1 - x^k} = \sum_{m=0}^{\infty} x^{km},$$

when we multiply them out, the coefficient in front of $x^n$ is equal exactly to $p(n)$ and Euler's formula follows.

Next, let $p_o(n)$ be the number of partitions of $n$ into odd parts, and $p_u(n)$ be the number of partitions of $n$ into unequal parts. An argument similar to the proof of Theorem 10.5 shows that the generating function for $p_o(n)$ is

$$\sum_{n=0}^{\infty} p_o(n)x^n = \prod_{n=1}^{\infty} \frac{1}{1-x^{2n-1}},$$

while the generating function for $p_u(n)$ is

$$\sum_{n=0}^{\infty} p_u(n)x^n = \prod_{n=1}^{\infty}(1+x^n).$$

It is easy to see that these two generating functions are the same:

$$\prod_{n=1}^{\infty} \frac{1}{1-x^{2n-1}} = \prod_{n=1}^{\infty}(1+x^n). \tag{10.11}$$

This is because

$$\prod_{n=1}^{\infty}(1+x^n)\prod_{n=1}^{\infty}(1-x^n) = \prod_{n=1}^{\infty}(1-x^{2n}),$$

and

$$\prod_{n=1}^{\infty}(1-x^{2n})\prod_{n=1}^{\infty}(1-x^{2n-1}) = \prod_{n=1}^{\infty}(1-x^n).$$

Hence, we have proved

**Proposition 10.6** *Let $p_o(n)$ be the number of partitions of $n$ into odd parts, and $p_u(n)$ be the number of partitions of $n$ into unequal parts, then $p_o(n) = p_u(n)$.*

Next we show the following. An integer $n$ is pentagonal if it is of the form $n = k(3k-1)/2$, with $k \in \mathbb{Z}_*$, or, equivalently, numbers of the form $n = k(3k+1)/2$, $k \in \mathbb{Z}_*$.

**Theorem 10.7** *(Euler) Let $p_{e,u}(n)$ be the number of partitions of $n$ into an even number of unequal parts, and $p_o(n)$ be the number of partitions of $n$ into am odd number of unequal parts, then $p_{e,u}(n) = p_{o,u}(n)$ unless $n$ is a pentagonal number, in which case*

$$p_{e,u}(n) - p_{o,u}(n) = (-1)^k, \quad \textit{if } n = k(3k+1)/2. \tag{10.12}$$

*Proof.* The proof based on two key observations. First, we have, for $|x| < 1$:

$$\prod_{n=1}^{\infty}(1-x^n) = \sum_{n=1}^{\infty}[p_{e,u}(n) - p_{o,u}(n)]x^n, \tag{10.13}$$

and

$$\prod_{n=1}^{\infty}(1-x^n) = \sum_{k=-\infty}^{\infty}(-1)^k x^{k(3k+1)/2}. \tag{10.14}$$

Expression (10.13) follows from multiplying out the terms in the product in the left side: this gives monomials of the form

$$(-1)^r x^{n_1+\cdots+n_r},$$

where all $n_k$, $k = 1, \ldots, r$ are different. Therefore, each partition of $n$ into an even number of unequal parts contributes $+1$ to the coefficient in front of $x^n$, while each partition of $n$ into an odd number of unequal parts contributes $-1$ to this coefficient, leading to the right side of (10.13).

Next, we prove (10.14). Set $x = e^{2\pi i u}$, with $u \in \mathbb{H}$, then we can write

$$\prod_{n=1}^{\infty}(1 - x^n) = \prod_{n=1}^{\infty}(1 - e^{2\pi i n u}),$$

and further, with $q = e^{3\pi i u}$ and $z = 1/2 + u/2$ we can re-write this as

$$\begin{aligned}
\prod_{n=1}^{\infty}(1 - x^n) &= \prod_{n=1}^{\infty}(1 - e^{2\pi i n u}) = \prod_{n=1}^{\infty}(1 - e^{2\pi i 3 n u})(1 - e^{2\pi i(3n-1)u})(1 - e^{2\pi i(3n-2)u}) \\
&= \prod_{n=1}^{\infty}(1 - q^{2n})(1 + q^{2n-1}e^{2\pi i z})(1 + q^{2n-1}e^{-2\pi i z}) = \Theta(z|\tau = 3u) = \sum_{n=-\infty}^{\infty} e^{\pi i n^2 \tau} e^{2\pi i n z} \\
&= \sum_{n=-\infty}^{\infty} e^{\pi i n^2 3u} e^{2\pi i n(1/2+u/2)} = \sum_{n=-\infty}^{\infty} (-1)^n e^{\pi i(3n^2+n)u} = \sum_{n=-\infty}^{\infty} (-1)^n x^{n(3n+1)/2},
\end{aligned}$$

which is nothing but (10.14). □

## Writing integers as sums of squares

### Integers that can be written as a sum of two squares

We now address the question of which integers can be written as a sum of squares, starting with just the sum of two squares. We let $r_2(n)$ be the number of ways an integer can be written as a sum of two squares, including obvious repetitions: for example, $r_2(5) = 8$ because

$$5 = (\pm 1)^2 + (\pm 2)^2 = (\pm 2)^2 + (\pm 1)^2.$$

It is easy to see that $r_2(3) = 0$ or $r_2(7) = 0$, so not all integers can be written as a sum of two squares.

**Theorem 10.8** *An integer $n$ can be written as a sum of two squares if and only if every prime $p_j$ of the form $4k + 3$ that occurs in its prime number factorization*

$$n = p_1^{a_1} p_2^{a_2} \ldots p_r^{a_r},$$

*has an even exponent $a_r$.*

This result will follow from the following. Let $d_1(n)$ be the number of divisors of $n$ of the form $4k+1$ and $d_3(n)$ the number of its divisors of the form $4k+3$.

**Theorem 10.9** *For all $n \geq 1$ we have $r_2(n) = 4(d_1(n) - d_3(n))$.*

Theorem 10.9 implies the conclusion of Theorem 10.8 as follows. First, if $n$ is a prime number of the form $n = 4k+1$, then $d_1(n) = 2$, $d_3(n) = 0$, so $r_2(n) = 8$, meaning that $n$ has a unique decomposition into the sum of two squares up to the trivial repetitions. Next, if $n$ is a number of the form $n = q^a$, with a prime number $q = 4k+3$, then $n$ has $a+1$ divisors. Furthermore, if $a$ is even then $d_1(n) = a/2 + 1$ and $d_3(n) = a/2$, so that $r_2(n) = 1$, while if $a$ is odd, then $d_1(n) = d_3(n) = (1+a)/2$, and $r_2(n) = 0$. For a general number

$$n = p_1^{a_1} p_2^{a_2} \dots p_r^{a_r},$$

a divisor of the form

$$m = p_1^{b_1} p_2^{b_2} \dots p_r^{b_r}, \quad 0 \leq b_j \leq a_j, \tag{10.15}$$

has the form $m = 4k+1$ if and only if $p = 2$ does not appear in the prime number factorization of $m$, and the sum of the exponents $b_s$ for all primes $p_s = 4k+3$ that appear in (10.15) is even, while $m$ has the form $m = 4k+3$ if that sum is odd. It is easy to see that to have $d_1(n) > d_3(n)$ we need all $a_j$ to be even if $p_j = 4k+3$.

The proof of Theorem 10.8 is rather long. We will consider the function

$$\theta(\tau) = \Theta(0|\tau) = \sum_{n=-\infty}^{\infty} e^{\pi i n^2 \tau} = \sum_{n=-\infty}^{\infty} q^{n^2}, \quad \tau \in \mathbb{H}, \quad q = e^{\pi i \tau}.$$

It follows that

$$(\theta(\tau))^2 = \left( \sum_{n_1=-\infty}^{\infty} q^{n_1^2} \right) \left( \sum_{n_2=-\infty}^{\infty} q^{n_2^2} \right) = \sum_{(n_1,n_2)\in\mathbb{Z}^2} q^{n_1^2+n_2^2} = \sum_{n=-\infty}^{\infty} r_2(n) q^n,$$

that is, $r_2(n)$ is the generating sequence for $\theta(\tau)^2$. Next, we make the following observation.

**Lemma 10.10** *The identity $r_2(n) = 4(d_1(n) - d_3(n))$, $n \geq 1$, is equivalent to the identities*

$$\theta(\tau)^2 = 2 \sum_{n=-\infty}^{\infty} \frac{1}{q^n + q^{-n}} = 1 + 4 \sum_{n=1}^{\infty} \frac{q^n}{1+q^{2n}}, \quad q = e^{\pi i \tau}, \quad \tau \in \mathbb{H}. \tag{10.16}$$

*Proof.* Both series in (10.16) converge because $|q| < 1$ and they are equal to each other because

$$\frac{1}{q^n + q^{-n}} = \frac{|q|^{|n|}}{1+q^{2|n|}}.$$

To prove the first equality we note that the right side of (10.16) can be written as

$$1 + 4 \sum_{n=1}^{\infty} \frac{q^n}{1+q^{2n}} = 1 + 4 \sum_{n=1}^{\infty} \frac{q^n(1-q^{2n})}{1-q^{4n}} = 1 + 4 \sum_{n=1}^{\infty} \left( \frac{q^n}{1-q^{4n}} - \frac{q^{3n}}{1-q^{4n}} \right).$$

The first term above can be written as

$$\sum_{n=1}^{\infty} \frac{q^n}{1-q^{4n}} = \sum_{n=1}^{\infty}\sum_{m=0}^{\infty} q^n q^{4mn} = \sum_{k=1}^{\infty} d_1(k) q^k,$$

while the second is

$$\sum_{n=1}^{\infty} \frac{q^{3n}}{1-q^{4n}} = \sum_{n=1}^{\infty}\sum_{m=0}^{\infty} q^{3n} q^{4mn} = \sum_{k=1}^{\infty} d_3(k) q^k,$$

finishing the proof of the lemma. □

It is convenient to set

$$\mathcal{C}(\tau) = 2 \sum_{m=-\infty}^{\infty} \frac{1}{q^n + q^{-n}} = \sum_{n=-\infty}^{\infty} \frac{1}{\cos(n\pi\tau)}, \quad q = e^{\pi i \tau}, \quad \tau \in \mathbb{H}.$$

Lemma 10.10 shows that our main task in the proof of Theorem 10.8 is to prove that $\mathcal{C}(\tau) = \theta^2(\tau)$.