Wolfgang Fischer | Ingo Lieb

A Course in Complex Analysis

From Basic Results to Advanced Topics

TEXTBOOK

VIEWEG+TEUBNER

Wolfgang Fischer | Ingo Lieb

A Course in Complex Analysis

Wolfgang Fischer | Ingo Lieb

A Course in Complex Analysis

From Basic Results to Advanced Topics

TEXTBOOK

VIEWEG+
TEUBNER

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Prof. Dr. Wolfgang Fischer
University of Bremen
Faculty 3 - Mathematics
Bibliothekstraße 1
D-28359 Bremen
Germany

fischer@math.uni-bremen.de

Prof. Dr. Ingo Lieb
University of Bonn
Mathematical Institute
Endenicher Allee 60
D-53115 Bonn
Germany

ilieb@math.uni-bonn.de

Large parts of this text are translated from the book Fischer, W./Lieb, I.: Einführung in die Komplexe Analysis, Vieweg+Teubner Verlag, 2010.
Translation: Jan Cannizzo

1st Edition 2012

Editorial Office: Ulrike Schmickler-Hirzebruch | Barbara Gerlach

Vieweg+Teubner Verlag is a brand of Springer Fachmedien.
Springer Fachmedien is part of Springer Science+Business Media.
www.viewegteubner.de

Cover design: KünkelLopka Medienentwicklung, Heidelberg
Printing company: AZ Druck und Datentechnik, Berlin
Printed on acid-free paper
Printed in Germany

ISBN 978-3-8348-1576-7

Preface

Among the most important tools of mathematics are the elementary functions – rational, trigonometric, hyperbolic and exponential functions, logarithms .... Their close relation only becomes apparent when one admits complex numbers as their arguments. This leads to developing complex analysis, i.e. calculus with complex numbers: the subject of the present book. Our choice of topics and manner of presentation have been determined by the following considerations:

1. The theory should rapidly lead to a deeper understanding of the elementary functions as well as to new classes of functions (higher functions). We thus present the elementary functions in the first chapter, study their deeper properties in the third chapter, and finally use the powerful methods of complex analysis worked out in chapters II to IV to introduce several non-elementary functions: elliptic functions, the Gamma- and the Zeta-function, and the modular map λ. These chapters (V and VII) contain a proof of the prime number theorem – perhaps the most striking application of complex analysis! – as well as a description of plane cubics in terms of elliptic functions, and a proof of Picard's theorem on essential singularities.

2. In order to circumvent topological difficulties we start with a local version of Cauchy's integral theorem – see chapter II – which suffices to build up most of the theory. A global theorem is then established using winding numbers; we follow Dixon's elegant argument. The residue theorem with its important applications then follows easily.

3. Functions of several complex variables naturally belong within the conceptual framework of complex analysis – a view which we share, e.g., with authors like H. Kneser, R. Narasimhan et al. We present these functions in the various chapters at the appropriate places and study them more deeply in the sixth chapter. Basic results that are covered are the Weierstrass preparation theorem and the solution of the Cousin problems in the entire space.

4. The geometric point of view has proven especially fruitful in complex analysis. It dominates our last chapter where we prove Riemann's mapping theorem, discuss hyperbolic geometry and introduce the modular map using a very general version of Schwarz's reflection principle.

Large parts of our text are translated from the original German version in [FL] which concentrates on the most elementary and basic results of complex analysis. We have considerably extended this text for the English version by the more advanced topics mentioned above. So the book should be accessible after a one year calculus class (which is assumed to include the definition of complex numbers), and it ought to take the reader from this foundation to fairly sophisticated topics – as expressed in the title.

Complex analysis is the creation of the great mathematicians of the 19th century; quite a few of its chapters have assumed their final shape and are presented in the same form all over the literature. We have of course followed this tradition. Wherever possible, we have given historical comments to exhibit the origin of important results; but we are quite unable to write a history of the subject. Both authors have made their first acquaintance with the field in lectures by H. Grauert to whom we owe deep gratitude. As to the actual text: the suggestion of translating and extending our textbook [FL] is due to Ulrike Schmickler-Hirzebruch (Vieweg+Teubner Verlag) and Dierk Schleicher (Jacobs University Bremen). The difficult task of translation was carefully and competently executed by Jan Cannizzo (now a graduate student at Ottawa University). Mathematical advice has come from our colleagues D. Schleicher, M. Range (Albany), J. Michel (Calais) et al. Daniel Fischer compiled the final LaTeX file and, moreover, suggested many improvements. We are sincerely grateful for all the help we received.

Bonn & Bremen, August 2011 *W. Fischer, I. Lieb*

Contents

Chapter I.

Analysis in the complex plane

The fundamental concept of holomorphic function is introduced via complex differentiability in section I.1. The relation between real and complex differentiability is then discussed, leading to the characterization of holomorphic functions by the Cauchy-Riemann differential equations (I.2). Power series are important examples of holomorphic functions (I.3); we here apply real analysis to show their holomorphy, although Chapter II will open a simpler way. In particular, the real exponential and trigonometric functions can be extended via power series to holomorphic functions on the whole complex plane; we discuss these functions without recurring to the corresponding real theory (I.4). Section I.5 presents an essential tool of complex analysis, viz. integration along paths in the plane. In I.6 we carry over the basic theory to functions of several complex variables.

The interpretation of complex numbers as points of the plane was given about 1800 by Gauss, Argand, and Wessel (independently); the term "Gaussian plane" is still in use. – Since 1820, Cauchy systematically investigated functions of a complex variable; he discussed the concept of complex differentiability (1841), as did Riemann (1851); both of them obtained the Cauchy-Riemann differential equations. The connection to real differentiability of functions of 2 or $2n$ real variables could, of course, only be clarified after the latter concept had been precisely formulated; this happened surprisingly late (Stolz 1893) [Sto]. – The "quotient-free" definition we use can be found in Carathéodory [Ca]. It offers technical advantages in complex as well as in real analysis, in particular in connection with Wirtinger derivatives, introduced by Poincaré and Wirtinger (~ 1900). – Holomorphic functions of several variables were investigated in the 19^{th} century (Cauchy, Jacobi, Riemann, Weierstrass); a systematic theory was only developed in the 20^{th} century.

0. Notations and basic concepts

Complex analysis develops differential and integral calculus for functions of one or several complex variables and applies these tools to elementary and non-elementary functions. The fundamental topological and analytic concepts of real analysis, quite naturally, play again an essential role: we recall some of them here and fix the meaning of some words.

The *field* $\mathbb{C}$ *of complex numbers* yields coordinates in the plane; we therefore talk about the *complex plane*. The *absolute value* (or *modulus*) $|z|$ of a complex number z measures its euclidean distance from 0; it satisfies the usual properties of a valuation. For a complex number $z = x+iy$, with $x = \operatorname{Re} z$ and $y = \operatorname{Im} z$ real and i the *imaginary unit*, the *complex conjugate* is defined as

$$\bar{z} = x - iy;$$

W. Fischer, I. Lieb, *A Course in Complex Analysis*, DOI 10.1007/978-3-8348-8661-3_1,

the map $z \mapsto \bar{z}$ is an automorphism of $\mathbb{C}$ (reflection in the real axis), and we have

$$|z| = \sqrt{z\bar{z}}.$$

By $D_\varepsilon(z_0)$ or $U_\varepsilon(z_0)$ we denote the *open disk* of radius ε and centre z_0:

$$D_\varepsilon(z_0) = U_\varepsilon(z_0) = \{z \in \mathbb{C} : |z - z_0| < \varepsilon\};$$

$\mathbb{D} = D_1(0)$ is the *unit disk*, the *unit circle* $\{z \in \mathbb{C} : |z| = 1\}$ is occasionally denoted by $\mathbb{S}$. We write $\mathbb{H}$ for the upper half plane $\{z : \operatorname{Im} z > 0\}$ and $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$.

Let us now turn to some topological concepts. $U_\varepsilon(z_0)$ is also called the ε-*neighbourhood* of z_0. An arbitrary *neighbourhood* of z_0 is a set U which contains an ε-neighbourhood. A set is *open* if it is a neighbourhood of each of its points; complements of open sets are *closed*. The *interior* $\overset{\circ}{M}$ of a set M is the largest open set contained in M; the smallest closed set containing M is its *closure* $\overline{M}$; the (topological) *boundary* of M is $\partial M = \overline{M} \setminus \overset{\circ}{M}$. The intersection of an open set U with an arbitrary set M is called *relatively open in* M or more simply *open in* M; *relatively closed* sets are defined correspondingly.

Convergent sequences and their *limits* are introduced in the usual way via neighbourhoods; let us note the rules

$$\lim_{\nu\to\infty} (z_\nu + w_\nu) = \lim_{\nu\to\infty} z_\nu + \lim_{\nu\to\infty} w_\nu,$$
$$\lim_{\nu\to\infty} z_\nu w_\nu = \lim_{\nu\to\infty} z_\nu \lim_{\nu\to\infty} w_\nu,$$
$$\lim_{\nu\to\infty} \frac{1}{z_\nu} = \frac{1}{\lim\limits_{\nu\to\infty} z_\nu};$$

the last formula is valid if $\lim z_\nu$ is different from 0. A point $z_0 \in \mathbb{C}$ is an *accumulation point* of $M \subset \mathbb{C}$, if there is a sequence z_ν in $M \setminus \{z_0\}$ with $\lim z_\nu = z_0$.

A set K which is *bounded* – i.e. $|z| \leqslant R$ for some R and all $z \in K$ – and closed is called *compact*; in this case each sequence in K contains convergent subsequences with limit in K. The latter property is equivalent to compactness. – By

$$U \subset\subset V$$

we will denote that the closure of U is compact and contained in V; for short: U is *relatively compact* in V.

Our main subject is the study of *complex valued functions* $f\colon M \to \mathbb{C}$, where M is a (usually open) subset of $\mathbb{C}$ or – in the case of several complex variables – $\mathbb{C}^n$. A function f is *continuous* at $z_0 \in M$, if for each neighbourhood V of $w_0 = f(z_0)$ there is a neighbourhood U of z_0 with $f(U \cap M) \subset V$. This property can also be expressed as

$$f\big(\lim_{\nu\to\infty} z_\nu\big) = \lim_{\nu\to\infty} f(z_\nu)$$

for all convergent sequences $z_\nu \in M$ with limit z_0. Any complex valued function – as any complex number – can be decomposed into its *real* and *imaginary parts*:

$$f = g + ih,$$

where g and h are real valued; f is continuous if and only if both g and h are. The *composition* of continuous functions is again continuous.

Examples of continuous functions are *polynomials* in z and $\bar{z}$:

$$f(z) = \sum_{\nu,\mu=0}^{N} a_{\nu\mu} z^\nu \bar{z}^\mu;$$

the *coefficients* $a_{\nu\mu}$ are complex numbers. The case of polynomials in z alone –

$$f(z) = \sum_{\nu=0}^{N} a_\nu z^\nu$$

is of particular importance. Other essential examples are *paths* $u\colon [a,b] \to \mathbb{C}$, i.e. continuous maps from closed finite intervals into the complex plane. The image set $u([a,\, b])$ is the *trace* of the path, denoted by $\operatorname{Tr} u$, the points $u(a)$ and $u(b)$ are the *initial* and *end points*, respectively; u *connects* its initial point with its endpoint. A set M is *(pathwise) connected* if any two points of M can be connected by a path whose trace lies in M. A connected open set is called a *domain*. An open set U is a domain if and only if no decomposition of U into disjoint nonempty open subsets exists.

Images of compact or connected sets under continuous functions are again compact or connected, resp. The corresponding statement for open or closed sets is false: it is not the images but the inverse images of open resp. closed sets that are open or closed, resp.

To conclude this sketchy reminder of things known, let us point out that all the above notions make sense for subsets of $\mathbb{R}^n$ and maps between such sets. One replaces the absolute value by the *euclidean norm* or any other norm and defines accordingly. We will make use of this remark without further mention.

1. Holomorphic functions

We begin with the central notion of complex analysis.

Definition 1.1. *A complex-valued function f defined on an open set $U \subset \mathbb{C}$ is (complex) differentiable at $z_0 \in \mathbb{C}$ if there exists a function Δ on U, continuous at z_0, such that*

$$f(z) = f(z_0) + \Delta(z)(z - z_0) \tag{1}$$

holds for all $z \in U$*. If* f *is complex differentiable at all points* $z_0 \in U$*, then we say that* f *is holomorphic on* U*. We say that* f *is holomorphic at* z_0 *if there exists an open neighbourhood of* z_0 *on which* f *is holomorphic.*

We call the number $\Delta(z_0)$, which is uniquely determined by (1), the value of the derivative of f at z_0, i.e.

$$\Delta(z_0) = f'(z_0) = \frac{df}{dz}(z_0).$$

If a function f is complex differentiable on all of U, then the values of its derivative at points $z \in U$ define

$$f'(z) = \frac{df}{dz}(z)$$

as a function on all of U. Accordingly, we may define the higher derivatives

$$\begin{aligned} f''(z) &= \frac{d}{dz} f'(z) = \frac{d^2 f}{dz^2}(z), \\ &\vdots \\ f^{(n)}(z) &= \frac{d}{dz}\frac{d^{n-1} f}{dz^{n-1}}(z) = \frac{d^n f}{dz^n}(z), \end{aligned}$$

provided they exist. In the next chapter, we will show that if f is holomorphic, then all of these higher derivatives do in fact exist! – Note that we will occasionally put $f^{(0)} = f$.

Examples:

i. If $f(z) \equiv c$ (i.e. if f is constant), then $f(z) = f(z_0) + 0(z - z_0)$, so that $f'(z) \equiv 0$.
ii. For $f(z) \equiv z$, we have $f(z) = f(z_0) + 1(z - z_0)$; thus $f'(z) \equiv 1$.
iii. Unlike the first two examples, the function $f(z) = \overline{z}$ is nowhere complex differentiable: If, as per (1), it were true that

$$\overline{z} - \overline{z}_0 = \Delta(z)(z - z_0),$$

where Δ is continuous at z_0, then for all z for which $z - z_0$ is real and nonzero, we would have

$$z - z_0 = \Delta(z)(z - z_0),$$

so that $\Delta(z) \equiv 1$, whereas for all z for which $z - z_0$ is imaginary and nonzero – thus $\overline{z} - \overline{z}_0 = -(z - z_0)$ – we would have

$$-(z - z_0) = \Delta(z)(z - z_0),$$

so that $\Delta(z) \equiv -1$. It follows that Δ cannot be continuous at z_0.

Example *iii* is surprising: an everywhere continuous but nowhere complex differentiable function! And yet the definition of complex differentiability is copied from the definition of real differentiability in one variable – and finding continuous nowhere differentiable functions on $\mathbb{R}$ is rather difficult. We will clarify this situation in the following section. For the moment, however, let us make use of the formal equality between the definitions of real and complex differentiability for functions of one variable in order to set up the differentiation rules familiar from real analysis. All proofs follow their real counterparts verbatim and are therefore left to the reader.

Proposition 1.1. *Any function that is complex differentiable at a point z_0 is continuous at z_0.*

Proposition 1.2. *Let f and g be complex differentiable at the point z_0. Then the functions $f+g$, fg, and $1/f$ (provided $f(z_0) \neq 0$) are complex differentiable at z_0, and we have*

$$\begin{aligned}(f+g)'(z_0) &= f'(z_0) + g'(z_0)\\ (fg)'(z_0) &= f'(z_0)g(z_0) + f(z_0)g'(z_0)\\ \left(\frac{1}{f}\right)'(z_0) &= -\frac{f'(z_0)}{f(z_0)^2}.\end{aligned}$$

The set of holomorphic functions on an open set U thus forms a ring (or, more precisely, a $\mathbb{C}$-algebra), which we denote by $\mathcal{O}(U)$.

Proposition 1.3 (The chain rule)**.** *Let $f\colon U \to V$ and $g\colon V \to \mathbb{C}$ be mappings of open subsets of $\mathbb{C}$ that are complex differentiable at $z_0 \in U$ and $f(z_0) = w_0 \in V$, respectively. Then*

$$g \circ f\colon U \to \mathbb{C}$$

is complex differentiable at z_0, and

$$(g \circ f)'(z_0) = g'(w_0)f'(z_0).$$

We introduce the following concepts concerning inverse functions.

Definition 1.2. *A map $f\colon U \to V$ between open subsets of $\mathbb{C}$ is called biholomorphic if it is bijective and holomorphic and, moreover, its inverse f^{-1} is holomorphic.*

We now show:

Proposition 1.4. *Let $f\colon U \to \mathbb{C}$ be a holomorphic function with a nonvanishing derivative. Then*

i. For every $z_0 \in U$, there exists a neighbourhood $U(z_0)$ such that $z = z_0$ is the only solution of the equation $f(z) = f(z_0)$ in $U(z_0)$.

ii. The map f is open (meaning that the images of open sets under f are open).

Proof: Without loss of generality, let $z_0 = 0$ and $f(z_0) = 0$. Claim *i* is almost trivial: Since f is holomorphic, it can be decomposed as

$$f(z) = \Delta(z)z, \tag{2}$$

where Δ is continuous at 0 and takes the value

$$\Delta(0) = f'(0) \neq 0$$

there. Thus Δ is nonzero in a neighbourhood of 0, and $0 = \Delta(z)z$ implies $z = 0$. We postpone the proof of the second claim to the following section. □

As in single-variable real calculus, the continuity of the inverse function (see *ii* above) yields

Proposition 1.5. *A map $f\colon U \to V$ is biholomorphic if and only if it is bijective and holomorphic and its derivative is nonvanishing. If this is the case, then*

$$\left(f^{-1}\right)'(w) = \frac{1}{f'(z)}, \quad \textit{where} \quad w = f(z).$$

The previous two propositions are special cases of the more general inverse function theorem from real analysis – cf. [Ru]; their proofs, however, are especially simple. We will strengthen both theorems considerably in the next chapter, showing that *bijective holomorphic maps are biholomorphic.*

Combining the examples of this section with the above rules, we obtain:

Proposition 1.6. *Polynomials in z, i.e. functions of the form*

$$p(z) = \sum_{\nu=0}^{n} a_\nu z^\nu,$$

are holomorphic on $\mathbb{C}$; their derivatives are again polynomials, namely

$$p'(z) = \sum_{\nu=1}^{n} \nu a_\nu z^{\nu-1}.$$

Proposition 1.7. *Rational functions, i.e. quotients of polynomials*

$$f(z) = \frac{p(z)}{q(z)},$$

are holomorphic everywhere except at the zeros of their denominators, and their derivatives are again rational functions.

Polynomials and rational functions constitute the most basic classes of functions; we will investigate them in more detail in Chapter III.

Exercises

1. Show that a function f is complex differentiable at z_0 in the sense of Def. 1.1 if and only if the limit

$$\lim_{z\to z_0} \frac{f(z)-f(z_0)}{z-z_0}$$

exists.

2. Find the derivative of $f(z) = (az+b)/(cz+d)$.
3. Where are the functions $\operatorname{Re} z$, $\operatorname{Im} z$, $|z|$, and $|z|^2$ complex differentiable?
4. Prove that every real-valued holomorphic function on $\mathbb{C}$ is constant.

2. Real and complex differentiability

Given that complex-valued functions of one complex variable can be viewed as maps (of subsets) of $\mathbb{R}^2$ into $\mathbb{R}^2$, the notion of (real) differentiability can be applied to them. A suitable formulation is:

A function

$$f = g + ih\colon U \to \mathbb{C}, \tag{1}$$

where $U \subset \mathbb{C}$ is open, is real differentiable at $z_0 = x_0 + iy_0 \in U$ if there exist functions $\Delta_1, \Delta_2\colon U \to \mathbb{C}$, continuous at z_0, such that

$$f(z) = f(z_0) + (x-x_0)\Delta_1(z) + (y-y_0)\Delta_2(z) \tag{2}$$

holds for all $z = x+iy \in U$. The functions Δ_1 and Δ_2 are not unique, but their values at z_0 are:

$$\begin{aligned} \Delta_1(z_0) &= f_x(z_0) = \frac{\partial f}{\partial x}(z_0) \\ \Delta_2(z_0) &= f_y(z_0) = \frac{\partial f}{\partial y}(z_0). \end{aligned} \tag{3}$$

It is immediate that this condition is equivalent to the differentiability of the real functions g and h, with

$$f_x = g_x + ih_x, \qquad f_y = g_y + ih_y. \tag{4}$$

We want to avoid the decomposition into real and imaginary parts and free ourselves of the real coordinates x and y. To this end, we note that

$$x - x_0 = \frac{1}{2}(z - z_0 + \overline{z} - \overline{z}_0), \qquad y - y_0 = \frac{1}{2i}(z - z_0 - (\overline{z} - \overline{z}_0)), \tag{5}$$

and substitute (5) into (2). An easy computation now yields:

The function $f\colon U \to \mathbb{C}$ *is real differentiable at* $z_0 \in U$ *if there exist functions* $\Delta, \mathrm{E}\colon U \to \mathbb{C}$ *that are continuous at* z_0 *and for which*

$$f(z) = f(z_0) + (z - z_0)\Delta(z) + (\overline{z} - \overline{z}_0)\mathrm{E}(z). \tag{2'}$$

The values $\Delta(z_0)$ and $\mathrm{E}(z_0)$ are uniquely determined by f.

We indeed obtain Δ and E from (2) and (5):

$$\Delta = \frac{1}{2}(\Delta_1 - i\Delta_2), \qquad \mathrm{E} = \frac{1}{2}(\Delta_1 + i\Delta_2). \tag{6}$$

Definition 2.1. *The values* $\Delta(z_0)$ *and* $\mathrm{E}(z_0)$ *in the decomposition* (2′) *are called the Wirtinger derivatives of* f *at* z_0, *denoted*

$$\begin{aligned}\Delta(z_0) &= \frac{\partial f}{\partial z}(z_0) = f_z(z_0)\\ \mathrm{E}(z_0) &= \frac{\partial f}{\partial \overline{z}}(z_0) = f_{\overline{z}}(z_0).\end{aligned}$$

In light of (3), we obtain from (6)

$$f_z = \frac{1}{2}(f_x - i f_y) \tag{7}$$

$$f_{\overline{z}} = \frac{1}{2}(f_x + i f_y). \tag{8}$$

The connection between real and complex differentiability is now immediately apparent from (2′):

Theorem 2.1. *A function* $f\colon U \to \mathbb{C}$ *is complex differentiable at* $z_0 \in U$ *if and only if it is real differentiable at* z_0 *and*

$$\frac{\partial f}{\partial \overline{z}}(z_0) = 0.$$

If this is the case, then $f'(z_0) = \dfrac{\partial f}{\partial z}(z_0)$.

Proof: If f is complex differentiable, we have the decomposition (1) from I.1, which is precisely (2′) with $\mathrm{E}(z) \equiv 0$; in particular, $f_{\overline{z}}(z_0) = 0$. Moreover, we have

$$f'(z_0) = \Delta(z_0) = f_z(z_0).$$

Conversely, if $f_{\overline{z}}(z_0) = 0$, then for $z \neq z_0$ we may write (2′) as

$$f(z) = f(z_0) + (z - z_0)\left(\Delta(z) + \frac{\overline{z} - \overline{z}_0}{z - z_0}\mathrm{E}(z)\right).$$

Since $\mathrm{E}(z_0) = f_{\overline{z}}(z_0) = 0$ and

$$\left|\frac{\overline{z} - \overline{z}_0}{z - z_0}\right| = 1,$$

the function

$$\Delta(z) + \frac{\overline{z} - \overline{z}_0}{z - z_0}\mathrm{E}(z)$$

is continuous at z_0. It follows that f is complex differentiable at z_0. □

Definition 2.2. *The differential operator*

$$\frac{\partial}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)$$

is called the Cauchy-Riemann operator.

In requiring that a function be complex differentiable we thus simultaneously require that a certain partial differential equation be satisfied. In other words, *holomorphic functions are the differentiable solutions of the Cauchy-Riemann equation*

$$\frac{\partial f}{\partial \overline{z}}(z) \equiv 0. \tag{9}$$

It is therefore plausible that holomorphic functions possess special properties that real differentiable functions in general do not – whether in one or in two variables.

The following "real" interpretation of the Cauchy-Riemann equations is especially important:

Let f be a twice differentiable function that satisfies (9). Then

$$\frac{\partial^2 f}{\partial z \partial \overline{z}}(z) \equiv 0. \tag{10}$$

A simple calculation shows that

$$\frac{\partial^2}{\partial z \partial \overline{z}} = \frac{1}{4}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) = \frac{1}{4}\Delta. \tag{11}$$

The operator Δ is called the *Laplace operator*, and solutions of the equation $\Delta f = 0$ are called harmonic functions. Since Δ is a real differential operator – i.e. $\Delta\overline{f} = \overline{\Delta f}$ – a function is harmonic precisely when both its real and imaginary parts are harmonic.

Proposition 2.2. *Twice differentiable holomorphic functions are harmonic, as are their real and imaginary parts.*

As we will see, the above differentiability requirement is superfluous – see Thm. II.3.4. Conversely, we will show that real harmonic functions are locally the real (or imaginary) parts of holomorphic functions – see VII.5. This connection between Laplace's equation and complex analysis is decisive in dealing with many questions. It only arises, however, in the theory of one complex variable. In this text, we will use it only in VII.6.

Before moving on, we give the Cauchy-Riemann equations in real form. Taking $f = g + ih$, they are as follows:

$$g_x = h_y, \qquad g_y = -h_x.$$

The gradients and, consequently, the equipotential lines of g and h are thus perpendicular to one another at all points where the gradients do not vanish.

As a further consequence of Thm. 2.1, we note:

Proposition 2.3. *If G is a domain, $f\colon G \to \mathbb{C}$ is holomorphic, and $f' \equiv 0$, then f is constant.*

The reason is that the Cauchy-Riemann equations then yield $f_z = 0$, $f_{\overline{z}} = 0$, $f_x = 0$, and $f_y = 0$.

We are now ready to finish the proof of Prop. 1.4. Let us note that if f and g are holomorphic functions, then $f\overline{g}$ is real differentiable, and

$$\frac{\partial}{\partial z}\big(f(z)\overline{g}(z)\big) = f'(z)\overline{g}(z). \tag{12}$$

Now we consider again the situation of Prop. 1.4. By part i, there is an $r > 0$ such that $z = 0$ is the only zero of f on $\overline{D_r(0)}$; in particular, $|f(z)| > 0$ on $\partial D_r(0) = \{z\colon |z| = r\}$, and since $\partial D_r(0)$ is compact, there is an $\varepsilon > 0$ such that

$$|f(z)| \geqslant 2\varepsilon \tag{13}$$

for $|z| = r$. We will show that

$$D_\varepsilon(0) \subset f\big(D_r(0)\big), \tag{14}$$

which implies the claim. For an arbitrary point $w_1 \in D_\varepsilon(0)$, the function

$$g(z) = |f(z) - w_1|^2$$

is real differentiable and, due to (13), satisfies the following inequalities:

$$\begin{aligned} g(0) &= |w_1|^2 < \varepsilon^2 \\ g(z) &= |f(z) - w_1|^2 > \varepsilon^2, \text{ for } |z| = r. \end{aligned} \tag{15}$$

Since g is continuous, it attains its minimum value on $\overline{D_r(0)}$, and by (15), this minimum must be attained in the interior of $\overline{D_r(0)}$, at, say, $z_1 \in D_r(0)$. All partial derivatives of g must vanish at z_1, so that, in particular,

$$0 = g_z(z_1) = f'(z_1)\big(\overline{f(z_1)} - \overline{w}_1\big). \tag{16}$$

Since $f'(z_1) \neq 0$, we have $w_1 = f(z_1)$. □

We need a special case of the chain rule for Wirtinger derivatives. The general case is left to the reader as Ex. 1.

Lemma 2.4. *Let $f\colon U \to \mathbb{C}$ be a real differentiable function defined on an open set $U \subset \mathbb{C}$, and let $w\colon [a,b] \to U$ be a differentiable map (i.e. a differentiable path in U). Then for all $t \in [a,b]$,*

$$\frac{\partial}{\partial t}(f \circ w)(t) = \begin{pmatrix} f_z(w(t)) & f_{\overline{z}}(w(t)) \end{pmatrix} \begin{pmatrix} \dot{w}(t) \\ \dot{\overline{w}}(t) \end{pmatrix} = f_z(w(t))\dot{w}(t) + f_{\overline{z}}(w(t))\dot{\overline{w}}(t).$$

Here we denote $\frac{\partial w}{\partial t}$ by $\dot{w}$ and use matrix multiplication.

We will draw on this lemma in giving a further interpretation of complex differentiability. First, let us recall some facts from linear algebra:

A complex linear map $l\colon \mathbb{C} \to \mathbb{C}$ is characterized by the conditions

$$l(z+w) = l(z) + l(w) \tag{17}$$

$$l(rz) = rl(z), \tag{18}$$

where $z, w, r \in \mathbb{C}$; if (18) holds only for real r, then l is called real linear. Every complex linear map is of the form

$$w = l(z) = az$$

for some unique number $a \in \mathbb{C}$, and every real linear map is of the form

$$w = l(z) = az + b\overline{z} = \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} z \\ \overline{z} \end{pmatrix} \tag{19}$$

for some unique pair $\begin{pmatrix} a & b \end{pmatrix} \in \mathbb{C}^2$. Thus l is complex linear if and only if $b = 0$.

Definition 2.3. *Let $f\colon U \to \mathbb{C}$ be real differentiable at the point z_0. The real linear map from $\mathbb{C}$ to $\mathbb{C}$ determined by the vector $\begin{pmatrix} f_z(z_0) & f_{\overline{z}}(z_0) \end{pmatrix}$ as per (19) is called the tangent map of f at z_0.*

We now have:

Proposition 2.5. *A function f is complex differentiable at a point z_0 if and only if its tangent map at z_0 is complex linear.*

Consider a differentiable path w whose initial point is z_0, i.e. a path $w\colon [a,b] \to \mathbb{C}$ such that $w(a) = z_0$, and assume that $\dot{w}(a) \neq 0$. Then we may interpret $\dot{w}(a)$ as the direction vector of w at a. If $v\colon [a,b] \to \mathbb{C}$ is another such path with direction vector $\dot{v}(a)$, then the angle between w and v, denoted $\sphericalangle(w,v)$, is by definition the oriented angle between the vectors $\dot{w}(a)$ and $\dot{v}(a)$, i.e. $\sphericalangle(\dot{w}(a), \dot{v}(a))$. A real differentiable function f maps the paths w and v to paths $f \circ w$ and $f \circ v$, respectively, that cross each other at an angle

$$\sphericalangle(f \circ w, f \circ v) = \sphericalangle\left(\frac{\partial}{\partial t}(f \circ w)(a), \frac{\partial}{\partial t}(f \circ v)(a)\right)$$

at the point $f(z_0)$. Now if f is complex differentiable at z_0 and $f'(z_0) \neq 0$, then by Lemma 2.4,

$$\sphericalangle\left(\frac{\partial}{\partial t}(f \circ w)(a), \frac{\partial}{\partial t}(f \circ v)(a)\right) = \sphericalangle\big(f'(z_0)\dot{w}(a), f'(z_0)\dot{v}(a)\big) = \sphericalangle\big(\dot{w}(a), \dot{v}(a)\big),$$

so that the image paths $f \circ w$ and $f \circ v$ cross at the same angle. Lemma 2.4 implies the converse of this statement as well.

Definition 2.4. *A map $f\colon U \to V$ is called conformal, or angle-preserving, if it is differentiable and has a differentiable inverse and if for all $z \in U$ and any differentiable paths w and v, both with initial point $z \in U$, we have*

$$\sphericalangle(w,v) = \sphericalangle(f \circ w, f \circ v).$$

We have thus shown:

Theorem 2.6. *A map $f\colon U \to V$ is conformal if and only if it is biholomorphic.*

Exercises

1. Formulate and prove the chain rule for Wirtinger derivatives. Furthermore, show that

$$\overline{\frac{\partial f}{\partial z}} = \frac{\partial \overline{f}}{\partial \overline{z}}.$$

2. Which of the following real functions are real parts of holomorphic functions?

$$x^3 - y^3, \quad x^3y - xy^3, \quad e^x \cos x, \quad e^x \cos y.$$

3. Let the map $f = g + ih\colon U \to \mathbb{C}$ be real differentiable. The real and complex Jacobian matrices of f are

$$J_f^{\mathbb{R}} = \begin{pmatrix} g_x & g_y \\ h_x & h_y \end{pmatrix} \quad \text{and} \quad J_f^{\mathbb{C}} = \begin{pmatrix} f_z & f_{\overline{z}} \\ \overline{f}_z & \overline{f}_{\overline{z}} \end{pmatrix},$$

respectively. Show that $\det J_f^{\mathbb{R}} = \det J_f^{\mathbb{C}}$. In particular, show that if f is holomorphic, then this determinant is equal to $|f'|^2$.

3. Uniform convergence and power series

Already in real analysis, important elementary functions are generated from simpler functions – typically polynomials – via limiting processes. This technique can naturally be transferred to complex analysis, allowing us to extend elementary real analytic functions on the real line to the complex plane and leading to a deeper understanding of their properties. Let us first recall some fundamental concepts.

An infinite series

$$\sum_{\nu=0}^{\infty} a_\nu$$

of complex numbers is *convergent* if and only if the sequence

$$s_n = \sum_{\nu=0}^{n} a_\nu$$

of its partial sums is convergent; the limit s of the sequence s_n is then by definition the sum of the series:

$$s = \sum_{\nu=0}^{\infty} a_\nu .$$

The series is *absolutely convergent* if the series whose terms are the absolute values $|a_\nu|$ is convergent. In this case, the series $\sum_{\nu=0}^{\infty} a_\nu$ is in fact *unconditionally convergent*, meaning that it will remain convergent (with the same sum) upon any reordering of its terms. If $|a_\nu| \leqslant |b_\nu|$ for all but finitely many ν, then the absolute convergence of $\sum_{\nu=0}^{\infty} b_\nu$ implies the absolute convergence of $\sum_{\nu=0}^{\infty} a_\nu$ (the *comparison test*). Likewise, the *ratio* and *root tests* familiar from real analysis are at our disposal. Both are derived via a comparison with the geometric series, which we now give for complex z.

Proposition 3.1. *The geometric series converges (even absolutely) precisely when* $|z| < 1$*; for such* z*, we have*

$$\sum_{\nu=0}^{\infty} z^\nu = \frac{1}{1-z}.$$

This, together with the ratio test, allows us to see easily that the series

$$\sum_{\nu=1}^{\infty} \nu\, z^{\nu-1}, \quad \sum_{\nu=2}^{\infty} \nu(\nu-1) z^{\nu-2}, \ldots, \quad \sum_{\nu=k}^{\infty} \nu(\nu-1)\cdots(\nu-k+1) z^{\nu-k} \tag{1}$$

are absolutely convergent for $|z| < 1$; their sums will be computed later.

Let us now consider sequences and series of functions. A sequence f_ν of functions defined on $M \subset \mathbb{C}$ *converges pointwise* to the function f, written

$$\lim_{\nu\to\infty} f_\nu = f \text{ or } f_\nu \to f,$$

if for all $z \in M$, we have $f_\nu(z) \to f(z)$. Much more important than pointwise convergence, however, is *uniform convergence*:

Definition 3.1. *A sequence of functions $f_\nu \colon M \to \mathbb{C}$ converges uniformly on M to the function f if for every $\varepsilon > 0$, there is an index ν_0 such that for all $\nu \geqslant \nu_0$ and all $z \in M$,*

$$|f_\nu(z) - f(z)| < \varepsilon.$$

The sequence f_ν converges locally uniformly to f on an open set U if for every point $z_0 \in U$, there is a neighbourhood $V(z_0) \subset U$ on which f_ν converges uniformly to f.

Locally uniform convergence is equivalent to *compact convergence*, meaning uniform convergence on every compact subset $K \subset U$. From real analysis, we have:

Proposition 3.2. *The limit of a uniformly or locally uniformly convergent sequence of continuous functions is continuous.*

Proposition 3.3. *Let f_ν be a sequence of holomorphic functions that converges to f on an open set U, and assume that the derivatives f'_ν are continuous and converge locally uniformly to a function g. Then f is holomorphic, and $f' = g$.*

Proof: We have $f_{\nu,z} = f'_\nu$ and $f_{\nu,\overline{z}} = 0$, which implies that all partial derivatives with respect to x and y are continuous and converge locally uniformly. The proposition now follows from standard results of differential calculus. □

We will obtain Prop. 3.3 in a stronger form and by simpler means (namely without recourse to real analysis) in Chapter II.

As usual, definitions and results about sequences of functions can be translated into analogous statements concerning *series* of functions. For us, the most important such definition is the following:

Definition 3.2. *An infinite series $\sum_{\nu=0}^{\infty} f_\nu$ of functions defined on a set M is absolutely uniformly convergent on M if the series of moduli $\sum_{\nu=0}^{\infty}|f_\nu|$ is uniformly convergent on M.*

The following proposition gives criteria for absolute uniform convergence.

Proposition 3.4.

i. (The Cauchy convergence test) The series $\sum_{\nu=0}^{\infty} f_\nu$ *is absolutely uniformly convergent on* M *if and only if for every* $\varepsilon > 0$*, there is an index* n_0 *such that for all* $n, m \geqslant n_0$ *(with* $m \geqslant n$*) and all* $z \in M$*,*

$$\sum_{\nu=n}^{m} |f_\nu(z)| < \varepsilon.$$

ii. (The majorant test) If $\sum_{\nu=0}^{\infty} a_\nu$ *is a convergent series with positive terms and if for almost all* ν *and all* $z \in M$ *we have*

$$|f_\nu(z)| \leqslant a_\nu,$$

then $\sum_{\nu=0}^{\infty} f_\nu$ *is absolutely uniformly convergent on* M*.*

We introduce *absolute locally uniform convergence* following Def. 3.1. Prop. 3.2 and 3.3 then hold for the limits of absolutely locally uniformly convergent series.

Before moving on to important examples, we note that the above definitions and theorems remain meaningful and true when applied to mappings of subsets of $\mathbb{R}^n$ (or $\mathbb{C}^n$) into $\mathbb{R}^m$ (or $\mathbb{C}^m$). We only need to replace the absolute value with the euclidean norm (or any other norm).

An infinite series of the form

$$P(z - z_0) = \sum_{\nu=0}^{\infty} a_\nu (z - z_0)^\nu,$$

where $a_\nu \in \mathbb{C}$ and $z_0 \in \mathbb{C}$, is called a *power series with base point* z_0 *and coefficients* a_ν. As an example, consider the geometric series

$$\sum_{\nu=0}^{\infty} z^\nu = \frac{1}{1 - z} \tag{2}$$

(see Prop. 3.1); it is not difficult to see that it converges absolutely locally uniformly to the function $(1 - z)^{-1}$ in the unit disk $\mathbb{D} = \{z\colon |z| < 1\}$. Making use of the majorant test, we get the following:

Proposition 3.5. *Assume that for some point* $z_1 \neq 0$*, the terms of the power series* $\sum_{\nu=0}^{\infty} a_\nu z^\nu$ *are bounded, that is* $|a_\nu z_1^\nu| \leqslant M$ *independently of* ν*. Then the series converges absolutely locally uniformly in the disk* $D_{|z_1|}(0) = \{z\colon |z| < |z_1|\}$*.*

Proposition 3.6. *For every power series* $P(z) = \sum_{\nu=0}^{\infty} a_\nu z^\nu$*, there is a well-defined number* $0 \leqslant r \leqslant \infty$ *such that* $P(z)$ *converges absolutely locally uniformly in the disk* $D_r(0)$ *and diverges for* $|z| > r$*.*

Definition 3.3. *The number introduced in Prop. 3.6 is called the radius of convergence of the power series* $P(z)$*, and* $D_r(0)$ *is called its disk of convergence.*

Here we regard ∞ as a number. If $r = 0$, we say that $P(z)$ is *nowhere convergent*, and if $r = \infty$, so that $D_r(0) = \mathbb{C}$, we say that $P(z)$ *converges everywhere*. We can of course immediately carry over the above statements to power series $P(z - z_0)$ with an arbitrary base point: the disk of convergence will then be centred at z_0.

Proof: It suffices to prove Prop. 3.5, since Prop. 3.6 is a direct consequence of it. Thus, let

$$|a_\nu||z_1|^\nu \leqslant M$$

for all ν. We choose z_2 such that $0 < |z_2| < |z_1|$ and for $|z| \leqslant |z_2|$ obtain

$$|a_\nu||z|^\nu \leqslant |a_\nu||z_2|^\nu = |a_\nu||z_1|^\nu \left|\frac{z_2}{z_1}\right|^\nu \leqslant Mq^\nu,$$

where $q = |z_2/z_1| < 1$. By Prop. 3.4.*ii* and Prop. 3.1, this guarantees absolute uniform convergence in the disk $D_{|z_2|}(0)$. □

With a bit of extra effort, one can also prove the *Cauchy-Hadamard formula*

$$r = \frac{1}{\limsup\limits_{\nu\to\infty} \sqrt[\nu]{|a_\nu|}} \tag{3}$$

for the radius of convergence. We leave this to the reader in Ex. 2.

By Prop. 3.2, a power series converges to a continuous function – which we also denote by $P(z)$ – in its disk of convergence. But more is true:

Theorem 3.7. *The sum of a power series*

$$P(z) = \sum_{\nu=0}^{\infty} a_\nu z^\nu \tag{4}$$

is holomorphic in its disk of convergence $D_r(0)$. Its derivative is

$$P'(z) = \sum_{\nu=1}^{\infty} \nu a_\nu z^{\nu-1}, \tag{5}$$

and the radius of convergence of P' coincides with that of P.

Proof: The series (5) is obtained from (4) by termwise differentiation. To apply Prop. 3.3, let us show that (5) converges locally uniformly in the disk of convergence of (4).

As in the proof of Prop. 3.5, we choose two points $0 < |z_2| < |z_1|$, where $z_1 \in D_r(0)$. The convergence of (4) now implies that

$$|a_\nu||z_1|^{\nu-1} \leqslant M$$

independently of ν. If $|z| \leqslant |z_2|$, then

$$|\nu a_\nu z^{\nu-1}| \leqslant \nu |a_\nu||z_1|^{\nu-1}\left|\frac{z_2}{z_1}\right|^{\nu-1} \leqslant M\nu q^{\nu-1},$$

where $q = |z_2/z_1|$. Comparing this with the first of the series in (1) gives the uniform convergence of (5) in the disk $|z| \leqslant |z_2|$. Moreover it is easy to see that the radius of convergence of (5) cannot be larger than that of (4) – we leave this to the reader. □

It follows that power series are in fact infinitely often complex differentiable in their disks of convergence – all of their derivatives are again convergent power series and are therefore holomorphic. We have

$$P^{(k)}(z) = \sum_{\nu=k}^{\infty} \nu(\nu-1)\cdots(\nu-k+1)a_\nu z^{\nu-k}.$$

Corollary 3.8 (The identity theorem for power series)**.**

i. *Let* $P(z) = \sum_{\nu=0}^{\infty} a_\nu z^\nu$ *be a convergent power series. Then*

$$a_\nu = \frac{P^{(\nu)}(0)}{\nu!}.$$

ii. *If* $P(z) \equiv 0$ *in a neighbourhood of* 0*, then* $a_\nu = 0$ *for all* ν*.*

iii. *If* $P(z) = \sum_{\nu=0}^{\infty} a_\nu z^\nu$ *and* $Q(z) = \sum_{\nu=0}^{\infty} b_\nu z^\nu$ *are convergent power series such that* $P(z) = Q(z)$ *in a neighbourhood of* 0*, then* $a_\nu = b_\nu$ *for all* ν*.*

Proof: Claim *i* is a direct consequence of the remarks preceding the corollary, *ii* follows from *i*, and *iii* follows from *ii*. □

In Chapter II, we will prove a general theorem that includes the preceding propositions. For now, however, let us apply our results in order to compute the sums (1). Repeated differentiation of the geometric series (2), namely

$$\frac{d^k}{dz^k}\frac{1}{1-z} = \frac{k!}{(1-z)^{k+1}},$$

gives us the series (1). Thus,

$$\sum_{\nu=k}^{\infty} \nu(\nu-1)\cdots(\nu-k+1)z^{\nu-k} = \frac{k!}{(1-z)^{k+1}}.$$

To conclude, we ask a few questions whose answers, although they could be given now, will be easier to come by in the next chapter.

a) Let z_0 be a point that belongs to the disk of convergence of a convergent power series

$$f(z) = \sum_{\nu=0}^{\infty} a_\nu z^\nu.$$

Is there a power series expansion

$$f(z) = \sum_{\nu=0}^{\infty} b_\nu (z - z_0)^\nu$$

of f about z_0? What would be the coefficients of this new series, and what would be its radius of convergence?

b) Does a rational function necessarily admit a power series expansion about every point in its domain of definition? How would one compute the coefficients of such a series?

Exercises

1. Prove that, on open subsets of $\mathbb{R}^n$, locally uniform convergence is equivalent to compact convergence. Does this equivalence hold on arbitrary subsets of $\mathbb{R}^n$?
2. Prove the Cauchy-Hadamard formula (3).
3. Show that if the limit $\lim_{\nu\to\infty}|a_\nu/a_{\nu+1}|$ exists, then it is equal to the radius of convergence of the series $\sum_{\nu=0}^{\infty} a_\nu z^\nu$.
4. Determine the radii of convergence of the following series:
$$\sum_{\nu=0}^{\infty} \nu^k z^\nu, \quad \sum_{\nu=0}^{\infty} \frac{z^\nu}{\nu!}, \quad \sum_{\nu=0}^{\infty} \nu! z^\nu, \quad \sum_{\nu=0}^{\infty} \frac{(2\nu)!}{2^\nu \nu!} z^\nu, \quad \sum_{\nu=0}^{\infty} \frac{(2\nu)!}{(\nu!)^2} z^\nu.$$
5. Let a_ν be a decreasing sequence of real numbers that converges to 0, and suppose that the radius of convergence of the series $\sum_{\nu=0}^{\infty} a_\nu z^\nu$ is 1. Show that for every $\delta > 0$ the series converges uniformly on $\overline{\mathbb{D}} \setminus D_\delta(1)$ (so that 1 is the only point on $\partial\mathbb{D}$ at which it could diverge).
Hint: Estimate the sum $(1 - z)\sum_{\nu=m}^{n} a_\nu z^\nu$.
6. Show that the series $\sum_{\nu=0}^{\infty} a_\nu$ converges absolutely if and only if each of the four subseries consisting, respectively, of those terms a_ν that lie in the quadrant where $\operatorname{Re} z > 0$ and $\operatorname{Im} z \geqslant 0$, those that lie in the quadrant where $\operatorname{Re} z \leqslant 0$ and $\operatorname{Im} z > 0$, and so on, converges.

4. Elementary functions

In real analysis, the elementary functions (exponential functions, trigonometric functions, and hyperbolic functions) exhibit very different behaviour (periodic, aperiodic, bounded, or unbounded) and seem to have little to do with one another. Considering them as functions of a complex variable, however, makes their close relationship apparent: All of them stem from the complex exponential function.

The power series

$$\exp z = \sum_{\nu=0}^{\infty} \frac{z^\nu}{\nu!} \tag{1}$$

converges absolutely locally uniformly on all of $\mathbb{C}$ and therefore defines a holomorphic function, the *(complex) exponential function*. We have

$$\exp 0 = 1, \tag{2}$$

and we define *Euler's number* to be

$$e = \exp 1 = \sum_{\nu=0}^{\infty} \frac{1}{\nu!}. \tag{3}$$

Since the coefficients of the series (1) are real, we have

$$\exp \overline{z} = \overline{\exp z}. \tag{4}$$

In particular, the complex exponential function is real for real z and coincides with the real exponential function on $\mathbb{R}$. The termwise differentiation of (1) gives

$$\frac{d}{dz} \exp z = \exp z, \tag{5}$$

the *differential equation of the exponential function*. Everything that follows is a consequence of the preceding identities.

Let $w \in \mathbb{C}$ be fixed. Then (5) implies that

$$\frac{d}{dz} \exp(w+z) \exp(-z) = \exp(w+z) \exp(-z) - \exp(w+z) \exp(-z) = 0.$$

Hence,

$$\exp(w+z) \exp(-z) \equiv \text{const}. \tag{6}$$

Setting $z = 0$ shows that this constant equals $\exp w$; setting $w = 0$ gives, in view of (2),

$$\exp z \exp(-z) = 1, \tag{7}$$

and after multiplying (6) with $\exp z$, we get

Proposition 4.1 (The addition rule)**.**

$$\exp(z+w) = \exp z \exp w,$$
$$\exp z \neq 0.$$

For $n \in \mathbb{Z}$, the addition rule reads

$$\exp n = (\exp 1)^n = e^n.$$

We will from now on write

$$\exp z = e^z \tag{8}$$

for an arbitrary number $z \in \mathbb{C}$ as well.

The identity (4) implies that

$$|e^z| = |e^{z/2}|^2 = e^{(z+\overline{z})/2} = e^{\operatorname{Re} z}. \tag{9}$$

The exponential function thus maps vertical lines $\operatorname{Re} z = c$ to circles; in particular, it maps the line $i\mathbb{R} = \{z\colon \operatorname{Re} z = 0\}$ to $\mathbb{S} = \{w\colon |w| = 1\}$. Moreover, if $z = x$ is real, then by (9), $e^x = |e^x| > 0$.

In summary:

Lemma 4.2. *The exponential function is a group homomorphism from $\mathbb{C}$ into $\mathbb{C}^*$ that maps the subgroup $\mathbb{R}$ into $\mathbb{R}_{>0}$ and the subgroup $i\mathbb{R}$ into $\mathbb{S}$.*

Note that the group operation on $\mathbb{C}$, $\mathbb{R}$, and $i\mathbb{R}$ is addition and the group operation on $\mathbb{C}^*$, $\mathbb{R}_{>0}$, and $\mathbb{S}$ is multiplication. Furthermore:

Proposition 4.3. *The three homomorphisms $\exp\colon \mathbb{C} \to \mathbb{C}^*$, $\exp\colon \mathbb{R} \to \mathbb{R}_{>0}$, and $\exp\colon i\mathbb{R} \to \mathbb{S}$ are surjective.*

Proof: By Prop. 1.4 and (5), the exponential map is open; let U be its image in $\mathbb{C}^*$. If $a \in \mathbb{C}^* \setminus U$ (which is closed), then the coset

$$aU = \{aw\colon w \in U\}$$

is also open, and

$$U \cap aU = \varnothing.$$

Indeed, if there were a number $b \in U \cap aU$, then for some numbers $z, w \in \mathbb{C}$, we would have

$$a\,e^z = b = e^w,$$

so that $a = e^{w-z} \in U$. It follows that $\mathbb{C}^* \setminus U$ is open and therefore empty, since $\mathbb{C}^*$ is connected. This shows that the first of the above homomorphisms is surjective. Now let $u \in \mathbb{R}_{>0}$. There exists a number $z = x + iy \in \mathbb{C}$ such that $u = e^z$, so that, by (9),

$$u = |u| = e^{\operatorname{Re} z} = e^x,$$

which shows that the second of the above homomorphisms is surjective. Finally, if $|w| = 1$ and $w = e^z$, then $e^{\operatorname{Re} z} = e^x = 1$, and since e^x is increasing (its derivative is positive), we have $x = 0$, which completes the proof. □

Next, let us consider the kernel N of the homomorphism $\exp\colon \mathbb{C} \to \mathbb{C}^*$, namely

$$N = \{z\colon e^z = 1\}.$$

Since there exists a number $z^* \in \mathbb{C}$ such that $e^{z^*} = -1$, $2z^*$ is a nonzero element of N. We already know that $N \subset i\mathbb{R}$. By Prop. 1.4, there is a neighbourhood U of 0 such that $U \cap N = \{0\}$, and since N is closed, there is a smallest positive number p such that $ip \in N$. We define the number π to be $p/2$:

Definition 4.1. *The number π is the smallest positive number for which $e^{2\pi i} = 1$.*

It is now clear that N consists of the integer multiples of $2\pi i$: if N contained a number $q \notin \{2k\pi i\colon k \in \mathbb{Z}\}$, then by adding some integer multiple of $2\pi i$, one could produce a number $q_0 = iq'$ in N such that $0 < q' < 2\pi$. – The addition rule now tells us that N consists exactly of the periods of e^z, i.e.

$$e^{z+2k\pi i} = e^z, \tag{10}$$

where $z \in \mathbb{C}$ and $k \in \mathbb{Z}$. We thus have:

Proposition 4.4.

i. The exponential function is periodic with period $2\pi i$.

ii. If a domain contains at most one member of each congruence class modulo $2\pi i$, then the exponential function maps it biholomorphically onto its image in $\mathbb{C}^*$.

Let us summarize the key mapping properties of the exponential function.

1. *The mapping $t \mapsto e^t$ maps $\mathbb{R}$ onto $\mathbb{R}_{>0}$ bijectively.*

2. *The mapping $t \mapsto e^{it}$ maps $[0, 2\pi)$ onto $\mathbb{S}$ bijectively.*

A horizontal line $z = x + iy_0$ is thus mapped bijectively onto the open ray L_{y_0} that begins at 0 and passes through the point e^{iy_0} on the unit circle. A vertical line $z = x_0 + iy$ is mapped onto the circle centred at 0 and of radius e^{x_0}; an interval of length less than 2π on this line is mapped injectively into this circle.

Every half-open horizontal strip

$$S_{y_0} = \{z = x + iy\colon y_0 \leqslant y < y_0 + 2\pi\}$$

is mapped bijectively onto $\mathbb{C}^*$. The line $z = x + iy_0$ is mapped to the ray L_{y_0}, and the remaining open strip (i.e. S_{y_0} minus its lower boundary) is mapped biholomorphically onto the "slit" plane $\mathbb{C}^* \setminus L_{y_0}$. For $y_0 = -\pi$, one thus obtains the slit plane $\mathbb{C}^* \setminus \mathbb{R}_{<0}$ cut along the negative real axis.

We introduce the remaining elementary functions in terms of the exponential function via the following *Euler formulas*:

Definition 4.2 (Trigonometric and hyperbolic functions)**.**

i. *The cosine function:* $\cos z = \frac{1}{2}(e^{iz} + e^{-iz})$.

ii. *The sine function:* $\sin z = \frac{1}{2i}(e^{iz} - e^{-iz})$.

iii. *The hyperbolic cosine function:* $\cosh z = \frac{1}{2}(e^{z} + e^{-z})$.

iv. *The hyperbolic sine function:* $\sinh z = \frac{1}{2}(e^{z} - e^{-z})$.

Properties of these functions can immediately be derived from the corresponding properties of the exponential function. We summarize them in the following proposition.

Proposition 4.5.

i. *The functions* $\cos z$ *and* $\sin z$ *are periodic with period* 2π*;* $\cosh z$ *and* $\sinh z$ *are periodic with period* $2\pi i$*.*

ii.
$$\frac{d}{dz}\sin z = \cos z \qquad \frac{d}{dz}\cos z = -\sin z$$
$$\frac{d}{dz}\sinh z = \cosh z \qquad \frac{d}{dz}\cosh z = \sinh z$$

iii.
$$\begin{aligned}\sin(z+w) &= \sin z\cos w + \cos z\sin w\\ \cos(z+w) &= \cos z\cos w - \sin z\sin w\\ \sinh(z+w) &= \sinh z\cosh w + \cosh z\sinh w\\ \cosh(z+w) &= \cosh z\cosh w + \sinh z\sinh w\end{aligned}$$

iv. *For each of the above functions,* $f(\overline{z}) = \overline{f(z)}$*.*

v.
$$\cos z = \sum_{\nu=0}^{\infty}(-1)^{\nu}\frac{z^{2\nu}}{(2\nu)!} \qquad \sin z = \sum_{\nu=0}^{\infty}(-1)^{\nu}\frac{z^{2\nu+1}}{(2\nu+1)!}$$
$$\cosh z = \sum_{\nu=0}^{\infty}\frac{z^{2\nu}}{(2\nu)!} \qquad \sinh z = \sum_{\nu=0}^{\infty}\frac{z^{2\nu+1}}{(2\nu+1)!}$$

vi.
$$e^{iz} = \cos z + i\sin z$$

vii.
$$\begin{aligned}\sin^2 z + \cos^2 z &\equiv 1\\ \cosh^2 z - \sinh^2 z &\equiv 1.\end{aligned}$$

It follows from v that, for real z, all of the above functions coincide with the trigonometric and hyperbolic functions familiar from calculus. Formula *vi* gives the decomposition of e^z into its real and imaginary parts (it is also called Euler's formula):

$$e^{x+iy} = e^{x}(\cos y + i\sin y). \tag{11}$$

Concerning the zeros of the trigonometric functions, we have:

Proposition 4.6. *The zeros of the sine function are the real numbers* $k\pi$*, and the zeros of the cosine function are the numbers* $\frac{\pi}{2} + k\pi$*, where* $k \in \mathbb{Z}$*.*

Proof: Let us restrict our attention to the cosine function. The equation

$$\cos z = \frac{1}{2}(e^{iz} + e^{-iz}) = 0$$

implies that

$$e^{2iz} + 1 = 0. \tag{12}$$

The only solutions of (12) are the aforementioned ones. □

In real analysis, one sometimes defines π via the condition that $\pi/2$ is the smallest positive zero of the cosine function. The above theorem shows that our definition of π gives the same number as this more elementary definition. The connection between π and the circumference of a circle will be derived in the following section.

As an application of the preceding formulas and definitions, let us exhibit the *nth roots of unity*, i.e. the n solutions of the equation

$$X^n = 1$$

in $\mathbb{C}$. Clearly,

$$\zeta_n = e^{2\pi i/n} = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}$$

is such a solution; all the solutions are the powers ζ_n^h, for $0 \leqslant h \leqslant n-1$. These numbers all lie on the circle $\mathbb{S}$ and span a regular n-gon one of whose vertices is 1. They form a cyclic group of order n (under multiplication); in the exercises, the reader is asked to show that all finite subgroups (and in fact all closed proper subgroups) of $\mathbb{S}$ are groups of nth roots of unity. Consequently, every complex number $z = |z|\,e^{it}$ has nth roots, which can be written in the form

$$\sqrt[n]{|z|}\,e^{it/n}\,\zeta_n^h, \tag{13}$$

where $0 \leqslant h \leqslant n-1$. For $z \neq 0$, all of these roots are distinct.

We conclude with a list of the remaining trigonometric and hyperbolic functions:

$$\tan z = \frac{\sin z}{\cos z}, \qquad \cot z = \frac{\cos z}{\sin z}, \tag{14}$$

$$\tanh z = \frac{\sinh z}{\cosh z}, \qquad \coth z = \frac{\cosh z}{\sinh z}. \tag{15}$$

They are holomorphic everywhere except at the zeros of their denominators and have periods π or, in the hyperbolic case, πi.

Concluding remark: In this section, we extended the elementary real analytic functions e^x, $\cos x$, etc. to holomorphic functions on the complex plane. Prop. II.4.2 will show that this extension can be carried out in only one way and is therefore not arbitrary.

Exercises

1. Let $\zeta \neq 1$ be an nth root of unity. Show that

$$1 + 2\zeta + 3\zeta^2 + \ldots + n\zeta^{n-1} = \frac{n}{\zeta - 1}.$$

2. Prove that every closed proper subgroup of $\mathbb{S}$ is finite cyclic and thus consists of nth roots of unity.

3. a) Show that $|\sin z| \leqslant \sinh|z|$ and $|\cos(z)| \leqslant \cosh|z|$. Investigate the behaviour of the trigonometric and hyperbolic functions as $\operatorname{Re} z \to \infty$ and $\operatorname{Im} z \to \infty$.

   b) Where do the functions $\sin z$, $\cos z$, $\tan z$, and $\cot z$ take on real values? Where do they take on imaginary values?

4. Show that the function $\tan z$ never takes on the values $\pm i$, and that therefore $\frac{d}{dz}\tan z \neq 0$ everywhere. Show that the tangent function maps the strip $S_0 = \{z\colon -\pi/2 < \operatorname{Re} z < \pi/2\}$ biholomorphically onto $\mathbb{C} \setminus \{it\colon t \in \mathbb{R},\ |t| \geqslant 1\}$.

5. Show that the radius of convergence of the power series

$$f(z) = \sum_{\nu=0}^{\infty} z^{\nu!}$$

   is 1, and that for any fixed $\alpha \in \mathbb{Q}$, $f(r\,e^{2\pi i\,\alpha})$ is unbounded as $r \to 1$.

5. Integration

We will now introduce the most important tool in complex analysis: the integration of complex-valued functions along suitable curves ("paths of integration") in the complex plane. That is, we will define

$$\int_\gamma f(z)\,dz, \tag{1}$$

where f is a – generally continuous – function and $\gamma\colon [a,b] \to \mathbb{C}$ is a piecewise continuously differentiable path in the domain of f. To the reader who is already familiar with path integrals: (1) is the integral of the special Pfaffian form $f(z)\,dz$ along the path γ. But we now define things "ab ovo", assuming (almost) no background knowledge.

The integral of the function $f = g + ih\colon [a,b] \to \mathbb{C}$, where $[a,b]$ is a closed real interval, is the complex number

$$\int_a^b f(t)\,dt = \int_a^b g(t)\,dt + i\int_a^b h(t)\,dt. \tag{2}$$

It exists if f is integrable, e.g. if f is piecewise continuous:

Definition 5.1.

i. *The function* $f\colon [a,b] \to \mathbb{C}$ *is piecewise continuous if there exists a partition*

$$a = t_0 < \ldots < t_n = b \tag{3}$$

of the interval $[a,b]$ *into subintervals* $[t_{\nu-1}, t_\nu]$ *such that the restriction of* f *to the open intervals* $]t_{\nu-1}, t_\nu[$ *is continuous and admits a continuous extension to the endpoints* $t_{\nu-1}$ *and* t_ν.

ii. *The function* f *is piecewise continuously differentiable if the derivative* f' *exists everywhere except at the points of a partition* (3) *and can be extended to a piecewise continuous function.*

Note that the functional

$$I(f) = \int_a^b f(t)\,dt$$

is complex linear and satisfies $I(\overline{f}) = \overline{I(f)}$. We thus have:

$$I(\operatorname{Re} f) = \operatorname{Re} I(f), \qquad I(\operatorname{Im} f) = \operatorname{Im} I(f). \tag{4}$$

The following inequality is essential:

Lemma 5.1.

$$\left|\int_a^b f(t)\,dt\right| \leqslant \int_a^b |f(t)|\,dt.$$

Proof: Choose a number $\alpha \in \mathbb{R}$ such that

$$e^{i\alpha} \int_a^b f(t)\,dt \geqslant 0.$$

Then using $|e^{i\alpha}| = 1$ and (4), we have

$$\begin{aligned}\left|\int_a^b f(t)\,dt\right| &= \left|e^{i\alpha}\int_a^b f(t)\,dt\right| = \operatorname{Re}\left(e^{i\alpha}\int_a^b f(t)\,dt\right) = \operatorname{Re}\int_a^b e^{i\alpha} f(t)\,dt \\ &= \int_a^b \operatorname{Re}(e^{i\alpha} f(t))\,dt \leqslant \int_a^b \left|e^{i\alpha} f(t)\right|\,dt = \int_a^b |f(t)|\,dt.\end{aligned}$$

□

Of course, the central theorems of integral calculus hold for complex-valued functions as well. Let us record the following two results for later use.

Proposition 5.2.

i. *Let f be continuously differentiable on $[a,b]$. Then*

$$\int_a^b f'(t)\,dt = f(b) - f(a).$$

ii. *(Substitution rule) Let f be piecewise continuous on $[a,b]$, and let $h\colon [c,d] \to [a,b]$ be a continuous, nondecreasing, and piecewise continuously differentiable bijection. Then*

$$\int_a^b f(s)\,ds = \int_c^d f(h(t))h'(t)\,dt. \tag{5}$$

(The integrand on the right hand side of (5) may be undefined at finitely many points, but this does not matter!)

We are now ready to define the paths along which we will integrate:

Definition 5.2. *Let $M \subset \mathbb{C}$ be a subset of the complex plane. A path of integration in M is a continuous and piecewise continuously differentiable map $\gamma\colon [a,b] \to M$.*

The interval $[a,b]$ is called the *parameter interval*, and γ runs from its *initial point* $\gamma(a)$ to its *end point* $\gamma(b)$. The image $\gamma[a,b]$ is called the *trace* of the path γ and is denoted $\operatorname{Tr}\gamma$. We borrow the following from real analysis:

Proposition 5.3. *Paths of integration are rectifiable, and their length is given by*

$$L(\gamma) = \int_a^b |\gamma'(t)|\,dt. \tag{6}$$

We now come to the key definition:

Definition 5.3. *Let $f\colon M \to \mathbb{C}$ be continuous on $M \subset \mathbb{C}$, and let $\gamma\colon [a,b] \to M$ be a path of integration in M. The integral of f along γ is defined as*

$$\int_\gamma f(z)\,dz = \int_a^b f(\gamma(t))\gamma'(t)\,dt. \tag{7}$$

The integrand on the right hand side might again be undefined at finitely many points, but it is nevertheless piecewise continuous.

Examples:

i. Let $a, b \in \mathbb{C}$, and let $\gamma\colon [0,1] \to \mathbb{C}$ be defined as

$$\gamma(t) = a + t(b-a).$$

We will denote this path by $[a,b]$ and refer to it as the *segment joining a to b*. If a and b are real and $a < b$, then it is immediate that

$$\int\limits_{[a,b]} f(z)\,dz = \int\limits_a^b f(t)\,dt.$$

ii. For $z_0 \in \mathbb{C}$ and $r > 0$, put

$$\kappa(r; z_0)(t) = z_0 + r\,e^{it}, \tag{8}$$

where $-\pi \leqslant t \leqslant \pi$. We call this path of integration the *(positively oriented) circle* with centre z_0 and radius r; the trace of the path (8) is precisely $\partial D_r(z_0)$.

Prop. 5.3 tells us that the length of the above circle is

$$L\left(\kappa(r; z_0)\right) = \int\limits_{-\pi}^{\pi} |ir\,e^{it}|\,dt = 2\pi r.$$

We thus see that the number π that we defined in terms of the period of the exponential function in the previous section coincides with the number π familiar from geometry.

iii. An especially important example is:

$$\int\limits_{\kappa(r;z_0)} \frac{dz}{z - z_0} = \int\limits_{-\pi}^{\pi} \frac{ir\,e^{it}}{r\,e^{it}}\,dt = 2\pi i. \tag{9}$$

Definition 5.4. *A parameter transformation is a map*

$$h\colon [c,d] \to [a,b]$$

between real intervals with the following properties:

i. It is a continuous, piecewise continuously differentiable bijection.

ii. There exists a number $\delta > 0$ such that $h'(t) \geqslant \delta$ for all t for which $h'(t)$ is defined.

It follows that h is increasing and that its inverse h^{-1} is a parameter transformation as well (this time from $[a,b]$ to $[c,d]$). Moreover, the composition of two parameter transformations is again a parameter transformation. If $\gamma\colon [a,b] \to \mathbb{C}$ is a path of integration, then $\gamma \circ h \colon [c, d] \to \mathbb{C}$ is a path of integration with the same initial point, end point, and trace as γ. We say that $\gamma \circ h$ is a *reparametrization* of γ.

Suppose now that f is continuous on the trace of γ. By (5) and (7),

$$\int_\gamma f(z)\,dz = \int_a^b f\big(\gamma(t)\big)\gamma'(t)\,dt = \int_c^d f\big(\gamma(h(s))\big)(\gamma \circ h)'(s)\,ds = \int_{\gamma\circ h} f(z)\,dz. \quad (10)$$

Integrals are therefore invariant under reparametrization.

In light of this, *we will identify two paths of integration if they are reparametrizations of each other.* To be pedantic: a path of integration is an equivalence class of (parametrized) paths of integration with respect to the aforementioned equivalence relation. The maps γ and $\gamma \circ h$ are thus parametrizations of the same path of integration.

The substitution rule also controls the behaviour of path integrals under holomorphic mappings. If γ is a path of integration in the open set $U \subset \mathbb{C}$ and $h \colon U \to V$ is a holomorphic function, then $h \circ \gamma$ is a path of integration in V. It follows that if f is continuous on V, then

$$\int_{h\circ\gamma} f(w)\,dw = \int_a^b f\Big(h\big(\gamma(t)\big)\Big)h'\big(\gamma(t)\big)\gamma'(t)\,dt = \int_\gamma (f \circ h)h'(z)\,dz. \quad (11)$$

Note the extra factor h' that appears in the integrand!

Paths of integration may, in a natural way, be combined, subdivided, and reversed:

a) Let γ_1 and γ_2 be two paths of integration such that the end point of γ_1 is the initial point of γ_2. We may – reparametrizing if necessary – assume that γ_1 and γ_2 are defined on the intervals $[a, b]$ and $[b, c]$. Then by setting

$$\gamma(t) = \begin{cases} \gamma_1(t), & t \in [a, b] \\ \gamma_2(t), & t \in [b, c], \end{cases}$$

we define a new path of integration γ. If f is a continuous function, then clearly

$$\int_\gamma f(z)\,dz = \int_{\gamma_1} f(z)\,dz + \int_{\gamma_2} f(z)\,dz. \quad (12)$$

We therefore also write (in this order!)

$$\gamma = \gamma_1 + \gamma_2.$$

b) If $\gamma\colon [a,b] \to \mathbb{C}$ is a path of integration and $a = t_0 < \ldots < t_n = b$ is a partition of $[a,b]$, then the restriction $\gamma_\nu = \gamma\big|[t_{\nu-1}, t_\nu]$ is also a path of integration, and

$$\gamma = \gamma_1 + \ldots + \gamma_n.$$

The subpaths γ_ν arise upon *subdividing* γ.

c) Let $\gamma\colon [0,1] \to \mathbb{C}$ again be a path of integration. We define the path

$$-\gamma\colon [0,1] \to \mathbb{C}$$

by $-\gamma(t) = \gamma(1-t)$ and call $-\gamma$ the *reverse* of γ. The trace stays the same, whereas the initial and end points are switched. Clearly,

$$\int\limits_{-\gamma} f(z)\,dz = -\int\limits_{\gamma} f(z)\,dz. \tag{13}$$

We will write $\gamma_1 - \gamma_2$ as shorthand for $\gamma_1 + (-\gamma_2)$.

Let us continue our list of examples:

iv. A path $\gamma\colon [a,b] \to \mathbb{C}$ such that $\gamma(t) \equiv c$ is a *constant path*. Any integral along γ is zero.

v. A path is called *closed* if its initial and end points coincide. The circle $\kappa(r; z_0)$, for example, is a closed path, as is the boundary $\partial\Delta$ of the triangle Δ with vertices a, b, and c:

$$\partial\Delta = [a,b] + [b,c] + [c,a].$$

vi. Example *iii* (formula (9)) shows that the value of an integral depends on the path of integration: if it did not, the integral over every closed path of integration would be zero (i.e. equal to the integral over a constant path). Two further typical examples of integrals over the circle $\kappa(1;0)\colon t \mapsto e^{it}$, where $-\pi \leqslant t \leqslant \pi$, are:

$$\int\limits_{\kappa(1;0)} z^n\,dz = i\int\limits_{-\pi}^{\pi} e^{i(n+1)t}\,dt = 0 \text{ for } n \neq -1,$$

$$\int\limits_{\kappa(1;0)} \overline{z}\,dz = \int\limits_{\kappa(1;0)} \frac{dz}{z} = 2\pi i.$$

vii. Instead of specifying a particular path along we which we want to integrate, we will often specify its trace when it is clear how it is to be parametrized. We may thus write the integrals in *vi* as $\int_{\mathbb{S}}$ or $\int_{|z|=1}$.

Translating Lemma 5.1 into the language of path integrals, we obtain:

Proposition 5.4 (The standard estimate)**.** *Let f be continuous on the trace of the path of integration γ. Then*

$$\left|\int_\gamma f(z)\,dz\right| \leqslant \max_{z\in\operatorname{Tr}\gamma} |f(z)| \cdot L(\gamma).$$

Proof: Since $\operatorname{Tr}\gamma$ is compact, $|f|$ is bounded on it, say by M. Lemma 5.1 then implies

$$\begin{aligned}\left|\int_\gamma f(z)\,dz\right| &= \left|\int_a^b f(\gamma(t))\gamma'(t)\,dt\right| \leqslant \int_a^b |f(\gamma(t))|\,|\gamma'(t)|\,dt \\ &\leqslant M\int_a^b |\gamma'(t)|\,dt = ML(\gamma).\end{aligned}$$

□

The following proposition is an immediate consequence of this estimate:

Proposition 5.5. *Let f_ν be a sequence of continuous functions that converges uniformly to a function f on the trace $\operatorname{Tr}\gamma$ of a path of integration γ. Then f is continuous, and*

$$\int_\gamma f(z)\,dz = \lim_{\nu\to\infty}\int_\gamma f_\nu(z)\,dz.$$

(Limit and integral signs are thus interchangeable.)

Proof: We already know that f is continuous. We then have

$$\begin{aligned}\left|\int_\gamma f(z)\,dz - \int_\gamma f_\nu(z)\,dz\right| &= \left|\int_\gamma (f(z)-f_\nu(z))\,dz\right| \\ &\leqslant \max_{z\in\operatorname{Tr}\gamma}\big|f(z)-f_\nu(z)\big|\cdot L(\gamma) \to 0.\end{aligned}$$

□

Two further propositions will be important in the chapters to come. Both of them are known results from the theory of integration in $\mathbb{R}^n$. We therefore state them without proof:

Proposition 5.6. *Let γ be a path of integration in $\mathbb{C}$, let $M \subset \mathbb{R}^n$, and let $f\colon \operatorname{Tr}\gamma \times M \to \mathbb{C}$ be a continuous function. Then:*

i. The function

$$F(\mathbf{x}) = \int\limits_{\gamma} f(z, \mathbf{x})\, dz$$

is continuous on M.

ii. If M is open and the partial derivative of f with respect to x_ν is continuous on $\operatorname{Tr}\gamma \times M$, then F is continuously differentiable with respect to x_ν, and

$$\frac{\partial F}{\partial x_\nu}(\mathbf{x}) = \int\limits_{\gamma} \frac{\partial f}{\partial x_\nu}(z, \mathbf{x})\, dz.$$

iii. If $M \subset \mathbb{C}$ is open and, for every $z \in \operatorname{Tr}\gamma$, f is complex differentiable with respect to $w \in M$ and the derivative $f_w(z, w)$ is continuous on $\operatorname{Tr}\gamma \times M$, then F is holomorphic in the variable w, and

$$F'(w) = \int\limits_{\gamma} f_w(z, w)\, dz.$$

Claim *iii* follows from *ii*, and *i* follows from the standard estimate in the same way as Prop. 5.5 (and can be seen as a generalization of Prop. 5.5).

Finally, the Fubini theorem (about interchanging the order of integration):

Proposition 5.7. *Let α and β be two paths of integration, and suppose the function f is continuous on $\operatorname{Tr}\alpha \times \operatorname{Tr}\beta$. Then*

$$\int\limits_{\alpha} \Big(\int\limits_{\beta} f(z, w)\, dw \Big) dz = \int\limits_{\beta} \Big(\int\limits_{\alpha} f(z, w)\, dz \Big) dw.$$

Moreover, the theory of integration tells us that the following statement can be viewed as a special case of Prop. 5.7:

Proposition 5.8. *Let $a_{\nu\mu}$, where $(\nu, \mu) \in \mathbb{N} \times \mathbb{N}$, be complex numbers. If there exists a bound M, independent of n and m, for which*

$$\sum_{\mu=0}^{m} \sum_{\nu=0}^{n} |a_{\nu\mu}| \leqslant M,$$

the series $\sum_{\nu=0}^{\infty} \big(\sum_{\mu=0}^{\infty} a_{\nu\mu} \big)$ and $\sum_{\mu=0}^{\infty} \big(\sum_{\nu=0}^{\infty} a_{\nu\mu} \big)$ are convergent, with the same sum.

Exercises

1. Let $f\colon [a, b] \to \mathbb{C}$ be a continuous function. Prove that

$$\Big| \int\limits_a^b \operatorname{Re} f(t)\, dt \Big| \leqslant \Big| \int\limits_a^b f(t)\, dt \Big| \quad \text{and} \quad \Big| \int\limits_a^b \operatorname{Im} f(t)\, dt \Big| \leqslant \Big| \int\limits_a^b f(t)\, dt \Big|.$$

2. Determine the trace of the path $\gamma(t) = a\,e^{it} + b\,e^{-it}$, where $0 \leqslant t \leqslant 2\pi$ and $a > b > 0$, and compute the integrals
$$\int_\gamma z\,dz \quad \text{and} \quad \int_\gamma z^2\,dz.$$
3. Compute $\int_{\kappa(r;0)} \operatorname{Re} z\,dz$ and $\int_{[a,b]} \operatorname{Re} z\,dz$, where $a, b \in \mathbb{C}$.

6. Several complex variables

The foundation laid in the first sections supports more than just the theory of holomorphic functions on the plane: we can easily extend our definitions to functions of several complex variables. In doing so, the proofs and results remain largely unchanged, although the theory of several complex variables eventually leads to questions that do not arise in the theory of one complex variable. Now for the details!

Given a point in the n-dimensional complex vector space
$$\mathbb{C}^n = \{\mathbf{z} = (z_1, \ldots, z_n) \colon z_\nu \in \mathbb{C}\},$$
we will denote the real and imaginary parts of its coordinates z_ν by x_ν and y_ν, respectively. Moreover, we equip $\mathbb{C}^n$ with the "maximum norm"
$$|\mathbf{z}| = \max_{1 \leqslant \nu \leqslant n} |z_\nu|$$
instead of the euclidean norm. We will consider complex-valued functions $f \colon U \to \mathbb{C}$ defined on an open set $U \subset \mathbb{C}^n$.

Definition 6.1. *A function $f \colon U \to \mathbb{C}$ is complex differentiable at the point $\mathbf{z}_0 \in U$ if there exist n functions $\Delta_1, \ldots, \Delta_n \colon U \to \mathbb{C}$, continuous at $\mathbf{z}_0$, for which*
$$f(\mathbf{z}) - f(\mathbf{z}_0) = \sum_{\nu=1}^{n} \Delta_\nu(\mathbf{z})(z_\nu - z_\nu^0) \tag{1}$$
holds for all $\mathbf{z} \in U$. If f is complex differentiable at every point $\mathbf{z}_0 \in U$, then f is said to be holomorphic on U. We say that f is holomorphic at $\mathbf{z}_0$ if it is holomorphic on a neighbourhood of $\mathbf{z}_0$.

The numbers $\Delta_\nu(\mathbf{z}_0)$ are uniquely determined by (1); they are by definition the values of the *complex partial derivatives* of f with respect to z_ν. We write:
$$\Delta_\nu(\mathbf{z}_0) = \frac{\partial f}{\partial z_\nu}(\mathbf{z}_0) = f_{z_\nu}(\mathbf{z}_0).$$
If f is holomorphic on all of U, the maps
$$\mathbf{z} \mapsto \frac{\partial f}{\partial z_\nu}(\mathbf{z}), \quad \mathbf{z} \in U,$$

define the complex partial derivatives of f as functions on U; we denote them by

$$f_{z_\nu} = \frac{\partial f}{\partial z_\nu}.$$

If they are again holomorphic, one may take *higher partial derivatives*

$$\frac{\partial^2 f}{\partial z_\nu \partial z_\mu}, \ldots, \frac{\partial^{k_1+\ldots+k_n} f}{\partial z_1^{k_1} \cdots \partial z_n^{k_n}}.$$

If we now define real differentiability as cleverly as we did in I.2, namely in terms of the relation

$$f(\mathbf{z}) - f(\mathbf{z_0}) = \sum_{\nu=1}^{n} \Delta_\nu(\mathbf{z})(z_\nu - z_\nu^0) + \sum_{\nu=1}^{n} \mathrm{E}_\nu(\mathbf{z})(\overline{z}_\nu - \overline{z}_\nu^0), \tag{2}$$

where Δ_ν and E_ν are continuous at $\mathbf{z}_0$, then as before we conclude:

The values $\Delta_\nu(\mathbf{z}_0)$ and $\mathrm{E}_\nu(\mathbf{z}_0)$ are uniquely determined by f and $\mathbf{z}_0$. We call them the *Wirtinger derivatives* of f and denote them by

$$\frac{\partial f}{\partial z_\nu} \text{ and } \frac{\partial f}{\partial \overline{z}_\nu}. \tag{3}$$

The same computation as in I.2 shows their relationship with the partial derivatives of f with respect to the real coordinates x_ν and y_ν:

$$\begin{aligned} f_{z_\nu} &= \frac{\partial f}{\partial z_\nu} = \frac{1}{2}\left(\frac{\partial f}{\partial x_\nu} - i\frac{\partial f}{\partial y_\nu}\right) \\ f_{\overline{z}_\nu} &= \frac{\partial f}{\partial \overline{z}_\nu} = \frac{1}{2}\left(\frac{\partial f}{\partial x_\nu} + i\frac{\partial f}{\partial y_\nu}\right). \end{aligned} \tag{4}$$

Likewise, the following is a consequence of the discussion in I.2:

Theorem 6.1. *A function $f\colon U \to \mathbb{C}$ is holomorphic if and only if it is real differentiable and satisfies the differential equations*

$$\frac{\partial f}{\partial \overline{z}_\nu} = 0, \quad 1 \leqslant \nu \leqslant n. \tag{5}$$

If this is the case, the Wirtinger derivatives f_{z_ν} coincide with the complex partial derivatives of f (which is why we denote them in the same way!).

The system (5) is called the *system of Cauchy-Riemann equations* in n variables.

A holomorphic function f in n variables is clearly holomorphic in each variable; more generally, if one fixes k variables, then f is holomorphic in the remaining $n-k$ variables.

The usual rules for finding partial derivatives evidently apply to functions of several complex variables as well. By observing that constant functions and the coordinate functions z_ν – but not $\overline{z}_\nu$! – are holomorphic, one obtains the following:

Proposition 6.2. *The set of holomorphic functions on an open set U forms a $\mathbb{C}$-algebra that contains the polynomial algebra $\mathbb{C}[z_1, \dots, z_n]$.*

Furthermore, the quotient rule for derivatives implies:

Proposition 6.3. *A quotient of holomorphic functions is holomorphic everywhere except at the zeros of its denominator. In particular, a rational function (i.e. a quotient of polynomials) is holomorphic at every point in its domain of definition.*

In order to formulate the chain rule, we need the following definition.

Definition 6.2. *Let $F = (f_1, \dots, f_m)$ be a mapping of an open set $U \subset \mathbb{C}^n$ into $\mathbb{C}^m$. Then we say that F is holomorphic if all of the coordinate functions f_μ are holomorphic.*

To be more explicit: $F(z_1, \dots, z_n) = (w_1, \dots, w_m)$, where

$$w_\mu = f_\mu(z_1, \dots, z_n), \quad 1 \leqslant \mu \leqslant m.$$

Let all f_μ be holomorphic. Then if f is a function that is holomorphic in a neighbourhood of a point $\mathbf{w}_0 = F(\mathbf{z}_0)$, one obtains for the function $f \circ F$ in a neighbourhood of $\mathbf{z}_0$:

Chain rule. *The function $f \circ F$ is holomorphic at $\mathbf{z}_0$, and*

$$\frac{\partial (f \circ F)}{\partial z_\nu}(\mathbf{z}_0) = \sum_{\mu=1}^{m} \frac{\partial f}{\partial w_\mu}(F(\mathbf{z}_0)) \frac{\partial f_\mu}{\partial z_\nu}(\mathbf{z}_0). \tag{6}$$

Formula (6) is typically written more concisely as

$$\frac{\partial (f \circ F)}{\partial z_\nu} = \sum_{\mu=1}^{m} \frac{\partial f}{\partial w_\mu} \frac{\partial w_\mu}{\partial z_\nu}. \tag{6'}$$

The proof is obtained as in the case of one variable. Alternatively, one can derive this formula from the corresponding formula from real analysis.

To conclude, let us turn our attention to the integration of functions of several variables. The only case we will need in establishing basic results is the following:

Let $\gamma_\nu \colon [0,1] \to \mathbb{C}$ be paths of integration, where $1 \leqslant \nu \leqslant n$. Then

$$\Gamma(t_1, \dots, t_n) = (\gamma_1(t_1), \dots, \gamma_n(t_n)) \tag{7}$$

defines a mapping of the n-dimensional cube

$$Q^n = \{(t_1, \dots, t_n) \colon 0 \leqslant t_\nu \leqslant 1\} \subset \mathbb{R}^n$$

into $\mathbb{C}^n$. We call Γ a *parametrized surface of integration* in $\mathbb{C}^n$; its image $\Gamma(Q^n)$ is the *trace* of Γ. For a continuous function f on the trace of Γ, we define

$$\int_{\Gamma} f(z_1, \ldots, z_n)\, dz_1 \ldots dz_n = \int_{\gamma_n} \cdots \int_{\gamma_2} \left(\int_{\gamma_1} f(z_1, \ldots, z_n)\, dz_1 \right) dz_2 \ldots dz_n. \tag{8}$$

Thus one first fixes the variables $z_2, \ldots, z_n$ and integrates the function

$$z_1 \mapsto f(z_1, z_2, \ldots, z_n)$$

over γ_1, obtaining a continuous function (see I.5) in the variables $z_2, \ldots, z_n$, say $F(z_2, \ldots, z_n)$. One then integrates the function

$$z_2 \mapsto F(z_2, \ldots, z_n),$$

where $z_3, \ldots, z_n$ are fixed, over γ_2, and so on. By Prop. 5.7, the order in which one integrates does not matter.

For the reader who has studied differential forms: (8) gives the integral of the complex $(n, 0)$-form

$$f(z_1, \ldots, z_n)\, dz_1 \wedge \ldots \wedge dz_n$$

over the n-dimensional surface of integration $\Gamma \subset \mathbb{C}^n = \mathbb{R}^{2n}$, i.e.

$$\int_{\Gamma} f(\mathbf{z})\, dz_1 \wedge \ldots \wedge dz_n.$$

It is possible to integrate in far greater generality, but this is needed only in the further development of the theory of several complex variables, which is beyond the scope of this book.

Exercises

1. Explicitly derive the Cauchy-Riemann equations from the definition of complex differentiability.
2. Derive the chain rule (6).
3. A map $F\colon U \to V$, where U and V are open subsets of $\mathbb{C}^n$, is biholomorphic if it is bijective and holomorphic and its inverse F^{-1} is holomorphic as well. Show that F is biholomorphic if it is bijective and holomorphic and F^{-1} is real differentiable.

Chapter II.

The fundamental theorems of complex analysis

Holomorphic functions differ fundamentally from real differentiable functions: they are infinitely often (real and complex) differentiable (II.3, II.7), they even admit power series expansions (II.4), their entire behaviour is determined by their values on arbitrarily small open sets (II.4, II.7), and they satisfy powerful convergence theorems and estimates (II.5). All of these properties are consequences of the Cauchy integral theorem and the integral representations that arise from it (II.1–3). Meromorphic functions extend the class of holomorphic functions (II.6); their study leads to the notion of isolated singularities and to generalizations of power series obtained by allowing negative powers (Laurent series). In addition to the phenomena that occur in the theory of functions of one complex variable, a fundamentally new phenomenon enters the picture in higher dimensions: the simultaneous holomorphic continuation of all holomorphic functions from a given domain to a larger one (II.7). Here the Cauchy integral formula (in one variable!) is again the decisive tool.

The Cauchy integral theorem was already known to Gauss (1811). It was proved independently of Gauss by Cauchy (1825) and by Weierstrass (1842). Goursat's proof (1900) does not require continuity of the derivative; the elegant technique of using triangular paths of integration, which we employ in the second section, is due to Pringsheim (1901). The Cauchy integral formulas were published by Cauchy in 1831 in the case of a circular path of integration and, much like we do here, were used by him to develop the theory of complex analysis. The beginnings of the theory of isolated singularities can likewise be traced to him. The theory of normal families (Thm. 5.3) was developed by Montel (1912). The remaining material in this chapter, with the exception of II.7, is "classical" – it was no doubt known in the second half of the nineteenth century. Work in the theory of several complex variables (II.7) began with Cauchy; the results were known to complex analysts of the late nineteenth century – with one (important!) exception: the use of Cauchy's integral in holomorphically continuing functions of several complex variables is due to Hartogs (1906). It is with him that the modern theory of functions of several complex variables begins (see [FG] and [Li]).

1. Primitive functions

The fundamental theorem of calculus shows that indefinite integration can be seen as the inverse of differentiation. We now investigate this question in the context of functions of a *complex* variable.

Definition 1.1. *Let $f\colon G \to \mathbb{C}$ be a continuous function defined on a domain $G \subset \mathbb{C}$. A function $F\colon G \to \mathbb{C}$ is called a primitive of f if it is holomorphic and satisfies $F' = f$.*

W. Fischer, I. Lieb, *A Course in Complex Analysis*, DOI 10.1007/978-3-8348-8661-3_2,

Example:
On its disk of convergence, a power series

$$\sum_{n=0}^{\infty} a_n (z - z_0)^n$$

has the primitive

$$\sum_{n=0}^{\infty} \frac{a_n}{n+1} (z - z_0)^{n+1}$$

(cf. I.3); in particular, the primitives of the functions $\exp z$, $\sin z$, and $\cos z$ are the ones we already know. The function $(z - z_0)^{-n}$, where $n \in \mathbb{N}$, is holomorphic on $\mathbb{C} \setminus \{z_0\}$, and for $n \neq 1$,

$$\frac{1}{1-n} (z - z_0)^{1-n}$$

is a primitive. If $n = 1$, then it does not have a primitive defined on $\mathbb{C} \setminus \{z_0\}$, as we shall soon see.

If F is a primitive of the function $f\colon G \to \mathbb{C}$, then it is easy to evaluate line integrals of f: if $\gamma\colon [a, b] \to G$ is continuously differentiable, then, since $(f \circ \gamma)\gamma'$ is continuous, the fundamental theorem of calculus implies that

$$\int_{\gamma} f(z)\, dz = \int_a^b f(\gamma(t))\gamma'(t)\, dt = \int_a^b (F \circ \gamma)'(t)\, dt = F(\gamma(b)) - F(\gamma(a)). \tag{1}$$

If γ is only piecewise continuously differentiable, we choose a partition

$$a = t_0 < \ldots < t_n = b$$

for which every $\gamma_\nu = \gamma\big|[t_{\nu-1}, t_\nu]$ is continuously differentiable and apply (1) to the subpaths γ_ν:

$$\begin{aligned} \int_{\gamma} f(z)\, dz &= \sum_{\nu=1}^{n} \int_{\gamma_\nu} f(z)\, dz \\ &= \sum_{\nu=1}^{n} \Big(F\big(\gamma(t_\nu)\big) - F\big(\gamma(t_{\nu-1})\big) \Big) = F\big(\gamma(b)\big) - F\big(\gamma(a)\big). \end{aligned}$$

We have thus shown:

Proposition 1.1. *If F is a primitive of $f\colon G \to \mathbb{C}$, then for every path of integration γ in G from z_0 to z_1, we have*

$$\int_\gamma f(z)\,dz = F(z_1) - F(z_0).$$

Thus, in this case the value of the integral depends only on the endpoints of the path of integration. In particular, if γ is a closed path of integration in G, then

$$\int_\gamma f(z)\,dz = 0. \tag{2}$$

In the real setting, every continuous function defined on an interval has a primitive. In the complex setting, functions as simple as $z \mapsto \operatorname{Re} z$ or $z \mapsto |z|$ do not (see Ex. 2). Not even the function $z \mapsto 1/z$, which is holomorphic on $\mathbb{C} \setminus \{0\}$, has a primitive there – this follows from (2) and the fact that

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

(Concerning notation, cf. I.5, Example *vii*).

The necessary condition (2) for the existence of a primitive is also sufficient:

Proposition 1.2. *If f is continuous on a domain $G \subset \mathbb{C}$ and satisfies*

$$\int_\gamma f(z)\,dz = 0$$

for every closed path of integration γ in G, then f has a primitive on G.

Proof: Let $a \in G$ be fixed. For every $z \in G$, choose a path of integration γ_z in G from a to z and set

$$F(z) = \int_{\gamma_z} f(\zeta)\,d\zeta.$$

Let us show that F is complex differentiable at every point $z_0 \in G$ and satisfies $F'(z_0) = f(z_0)$. If z is a point that is sufficiently close to z_0, then $[z_0, z] \subset G$, and $\gamma_{z_0} + [z_0, z] - \gamma_z$ is a closed path in G. By assumption, we have

$$\int_{\gamma_{z_0}} f(\zeta)\,d\zeta + \int_{[z_0,z]} f(\zeta)\,d\zeta - \int_{\gamma_z} f(\zeta)\,d\zeta = 0,$$

so that

$$F(z) - F(z_0) = \int_{[z_0,z]} f(\zeta)\, d\zeta = \int_0^1 f\big(z_0 + t(z-z_0)\big)(z-z_0)\, dt = (z-z_0)A(z),$$

where

$$A(z) = \int_0^1 f\big(z_0 + t(z-z_0)\big)\, dt.$$

We have $A(z_0) = f(z_0)$; moreover, A is continuous at z_0, because the above integrand is a continuous function of z (in a neighbourhood U of z_0). □

Note that the primitive F depends on a but not on the choice of the paths γ_z. If $\widetilde{\gamma}_z$ is another path in G from a to z, then $\gamma_z - \widetilde{\gamma}_z$ is closed, so that

$$\int_{\gamma_z} f(\zeta)\, d\zeta - \int_{\widetilde{\gamma}_z} f(\zeta)\, d\zeta = \int_{\gamma_z - \widetilde{\gamma}_z} f(\zeta)\, d\zeta = 0.$$

If we restrict ourselves to domains of a special shape, we may weaken the conditions of Prop. 1.2. Rather than considering arbitrary closed paths, it suffices to consider triangular paths.

Definition 1.2. *A domain $G \subset \mathbb{C}$ is called star-shaped if there exists a point $a \in G$ such that, for every $z \in G$, the segment $[a,z]$ is contained in G.*

Convex domains are of course star-shaped.

Proposition 1.3. *Let $G \subset \mathbb{C}$ be a star-shaped domain (with respect to $a \in G$), and let $f\colon G \to \mathbb{C}$ be continuous. If for every closed triangle $\Delta \subset G$ one of whose vertices is a, we have*

$$\int_{\partial\Delta} f(\zeta)\, d\zeta = 0,$$

then f has a primitive.

Proof: We claim that

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

is a primitive of f. If $z_0 \in G$ and z is close enough to z_0 such that $[z_0, z] \subset G$, then the triangle Δ with vertices a, z, and z_0 lies inside G, and

$$0 = \int_{\partial\Delta} f(\zeta)\,d\zeta = \int_{[a,z]} f(\zeta)\,d\zeta - \int_{[z_0,z]} f(\zeta)\,d\zeta - \int_{[a,z_0]} f(\zeta)\,d\zeta.$$

Thus,

$$F(z) - F(z_0) = \int_{[z_0,z]} f(\zeta)\,d\zeta,$$

and the claim follows by the argument used in the proof of Prop. 1.2. $\square$

Remark: If G is an arbitrary domain and $\int_{\partial\Delta} f(\zeta)\,d\zeta = 0$ holds for all closed triangles $\Delta \subset G$, then f has *local primitives*, i.e. every point $a \in G$ has a neighbourhood U such that $f\big|U$ has a primitive. It suffices to take for U a disk that is centred at a and contained in G and then to apply Prop. 1.3.

Exercises

1. Let γ be a path of integration from $i+1$ to $2i$. Compute the integrals of the following functions over γ:

   $$\cos(1+i)z, \quad iz^2 + 1 - 2iz^{-2}, \quad (z+1)^3, \quad z\exp(iz^2).$$

2. Show that the functions $z \mapsto \operatorname{Re} z$ and $z \mapsto |z|$ do not have primitives on $\mathbb{C}$.
3. Show that if a function $f\colon G \to \mathbb{C}$ has local primitives, then for every closed triangle Δ in G, we have $\int_{\partial\Delta} f(z)\,dz = 0$.

2. The Cauchy integral theorem

We now come to the central theorem of complex analysis – virtually all of the important results of this chapter and the next make direct or indirect use of the contents of this section.

Theorem 2.1 (Goursat's lemma)**.** *Let f be holomorphic in a neighbourhood of a closed triangle Δ. Then*

$$\int_{\partial\Delta} f(z)\,dz = 0.$$

Holomorphic functions thus satisfy the conditions of Prop. 1.3.

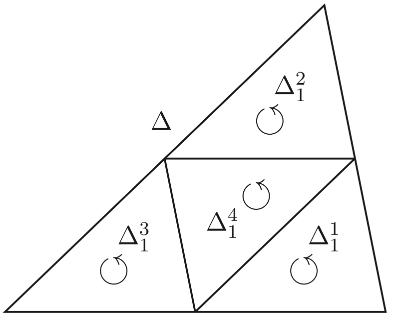


Figure 1. Subdivision of a triangle

Proof: Let us subdivide Δ into four subtriangles $\Delta_1^1, \dots, \Delta_1^4$ by connecting the midpoints of the edges of Δ (see Fig. 1). Every segment that joins two midpoints belongs to the boundary of exactly two subtriangles; thus, if we form the sum

$$\sum_{k=1}^{4} \int_{\partial \Delta_1^k} f(z)\, dz,$$

the integrals over these segments cancel, and we have

$$\int_{\partial \Delta} f(z)\, dz = \sum_{k=1}^{4} \int_{\partial \Delta_1^k} f(z)\, dz,$$

so that

$$\left| \int_{\partial \Delta} f(z)\, dz \right| \leqslant 4 \max_{1 \leqslant k \leqslant 4} \left| \int_{\partial \Delta_1^k} f(z)\, dz \right|.$$

Among the triangles Δ_1^k, choose one such that the absolute value of its boundary integral is maximal, and call it Δ_1. We then have

$$\left| \int_{\partial \Delta} f(z)\, dz \right| \leqslant 4 \left| \int_{\partial \Delta_1} f(z)\, dz \right|.$$

Now apply the same procedure to the triangle Δ_1, yielding a triangle Δ_2 such that

$$\left| \int_{\partial \Delta_1} f(z)\, dz \right| \leqslant 4 \left| \int_{\partial \Delta_2} f(z)\, dz \right|.$$

Continuing in this way, we obtain a sequence of nested triangles

$$\Delta = \Delta_0 \supset \Delta_1 \supset \dots \supset \Delta_n \supset \dots$$

for which

$$\left|\int_{\partial\Delta} f(z)\,dz\right| \leqslant 4^n \left|\int_{\partial\Delta_n} f(z)\,dz\right|. \tag{1}$$

Moreover, the perimeters $L(\partial\Delta_n)$ and diameters $\operatorname{diam}\Delta_n$ satisfy

$$L(\partial\Delta_n) = \frac{1}{2}L(\partial\Delta_{n-1}) = \ldots = \frac{1}{2^n}L(\partial\Delta) \tag{2}$$

$$\operatorname{diam}\Delta_n = \frac{1}{2^n}\operatorname{diam}\Delta. \tag{3}$$

Since all Δ_n are compact, their intersection is nonempty, and by (3), it consists of a single point z_0, i.e.

$$\bigcap_{n\geqslant 0} \Delta_n = \{z_0\}.$$

The function f is complex differentiable at z_0, so that

$$f(z) = f(z_0) + (z - z_0)f'(z_0) + (z - z_0)A(z),$$

where A is a continuous function on Δ that vanishes at z_0. This allows us to estimate the integral of f as follows: The linear function

$$f(z_0) + (z - z_0)f'(z_0)$$

has a primitive, so its integral over the closed path $\partial\Delta_n$ is zero. The standard estimate then yields

$$\begin{aligned}\left|\int_{\partial\Delta_n} f(z)\,dz\right| &= \left|\int_{\partial\Delta_n} (z - z_0)A(z)\,dz\right| \\ &\leqslant L(\partial\Delta_n)\max_{z\in\Delta_n}\bigl|(z - z_0)A(z)\bigr| \\ &\leqslant L(\partial\Delta_n)\operatorname{diam}\Delta_n \max_{z\in\Delta_n}|A(z)|.\end{aligned} \tag{4}$$

Combining (4) with (1), (2), and (3), we have

$$\begin{aligned}\left|\int_{\partial\Delta} f(z)\,dz\right| &\leqslant 4^n \cdot 2^{-n} \cdot 2^{-n} L(\partial\Delta)\operatorname{diam}\Delta \cdot \max_{z\in\Delta_n}|A(z)| \\ &= L(\partial\Delta)\operatorname{diam}\Delta \cdot \max_{z\in\Delta_n}|A(z)|.\end{aligned}$$

The continuous function $A(z)$ vanishes at z_0, hence

$$\lim_{n\to\infty}\max_{z\in\Delta_n}|A(z)| = 0,$$

from which it follows that

$$\int_{\partial\Delta} f(z)\,dz = 0,$$

as claimed. □

Prop. 1.3 and Goursat's lemma imply that, on a star-shaped domain, holomorphic functions have primitives. This is the "fundamental theorem of complex analysis":

Theorem 2.2 (The Cauchy integral theorem for star-shaped domains)**.** *Let f be a holomorphic function defined on a star-shaped domain G. Then*

$$\int_{\gamma} f(z)\,dz = 0$$

for every closed path of integration γ in G.

A holomorphic function defined on a domain G that is not star-shaped may or may not have a primitive defined on all of G, but it does have local primitives, because every point in G has a convex neighbourhood contained in G.

Exercises

1. Let f be continuous on $\overline{\mathbb{D}}$ and holomorphic on $\mathbb{D}$. Show that

$$\int_{\partial\mathbb{D}} f(\zeta)\,d\zeta = 0.$$

Hint: Consider the functions $f_r(z) = f(rz)$ as $r \uparrow 1$.

2. a) Let $a \neq 0$ be a real number. Using the formula $\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$, compute the integral

$$\int_{-\infty}^{\infty} \exp(-x^2 - 2iax)\,dx.$$

Hint: Integrate $\exp(-z^2)$ over the boundary of the rectangle with vertices $\pm R$ and $\pm R + ia$, and let $R \to \infty$.

b) Using the result from part a), compute the following integral for $\lambda \in \mathbb{R}$:

$$\int_{-\infty}^{\infty} \exp(-x^2)\cos(\lambda x)\,dx.$$

3. Compute the "Fresnel integrals"

$$\int_0^{\infty} \cos(x^2)\,dx = \sqrt{\pi/8} = \int_0^{\infty} \sin(x^2)\,dx.$$

Hint: Apply the Cauchy integral theorem to sectors with centre 0 and corners given by R and $e^{i\pi/4}R$, where $R \to \infty$.

4. Show that, for $0 \leqslant a < 1$,

$$\int_0^{\infty} e^{-(1-a^2)x^2}\cos(2ax^2)\,dx = \frac{\sqrt{\pi}}{2(1+a^2)} \quad \text{and} \quad \int_0^{\infty} e^{-(1-a^2)x^2}\sin(2ax^2)\,dx = \frac{a\sqrt{\pi}}{2(1+a^2)}.$$

3. The Cauchy integral formula

We will derive an integral representation from the Cauchy integral theorem that shows how the values of a holomorphic function in the interior of a disk are determined by its values on the boundary circle.

Theorem 3.1 (The Cauchy integral formula). *Let G be a domain, let $z_0 \in G$, and let $D = D_r(z_0) \subset\subset G$. Then for every holomorphic function f on G,*

$$f(z) = \frac{1}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{\zeta - z}\, d\zeta \quad \text{for all } z \in D. \tag{1}$$

Proof: For a sufficiently small $\varepsilon > 0$, $U = D_{r+\varepsilon}(z_0)$ is a convex (and thus star-shaped) neighbourhood of z_0 contained in G. For a fixed $z \in D$, consider the function

$$g(\zeta) = \begin{cases} \dfrac{f(\zeta) - f(z)}{\zeta - z} & \text{for } \zeta \neq z \\ f'(z) & \text{for } \zeta = z \end{cases}$$

on U. It is certainly holomorphic on $U \setminus \{z\}$, and although it may not be holomorphic at the point $\zeta = z$, it is continuous there. We will nevertheless apply the Cauchy integral theorem to g – this is justified by Cor. 3.3 below. We then have

$$0 = \int_{\partial D} g(\zeta)\, d\zeta = \int_{\partial D} \frac{f(\zeta)\, d\zeta}{\zeta - z} - \int_{\partial D} \frac{f(z)\, d\zeta}{\zeta - z}.$$

The theorem now follows provided we show that

$$\int_{\partial D} \frac{d\zeta}{\zeta - z} = 2\pi i \tag{2}$$

for $z \in D$. This will be done after Cor. 3.3.

Lemma 3.2. *Let $\Delta \subset \mathbb{C}$ be a closed triangle, and let f be holomorphic in a neighbourhood of Δ with the possible exception of a point $z_0 \in \Delta$, at which f is continuous. Then*

$$\int_{\partial \Delta} f(z)\, dz = 0.$$

Proof: a) Suppose z_0 is a vertex of Δ. The function f is continuous and hence bounded on the compact set Δ, say $|f(z)| \leqslant M$. For a given $\varepsilon > 0$, split Δ into three subtriangles, as shown in the left of Fig. 2, in such a way that $L(\partial\Delta_1) < \varepsilon$. Then by Thm. 2.1, the integrals over $\partial\Delta_2$ and $\partial\Delta_3$ vanish, and we have

$$\left| \int_{\partial \Delta} f(z)\, dz \right| = \left| \int_{\partial \Delta_1} f(z)\, dz \right| \leqslant L(\partial\Delta_1) \cdot M < \varepsilon M.$$

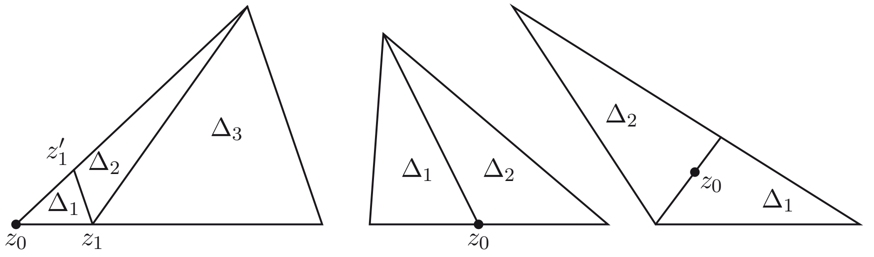


Figure 2. Splitting of triangles

b) Suppose now that z_0 lies on an edge of Δ but is not a vertex. Split Δ as shown in the middle of Fig. 2. By part a), we obtain

$$\int_{\partial\Delta} f(z)\,dz = \int_{\partial\Delta_1} f(z)\,dz + \int_{\partial\Delta_2} f(z)\,dz = 0.$$

c) If z_0 belongs to the interior of Δ, then split Δ as shown in the right of Fig. 2. By part b), we again have

$$\int_{\partial\Delta} f(z)\,dz = \int_{\partial\Delta_1} f(z)\,dz + \int_{\partial\Delta_2} f(z)\,dz = 0. \qquad \square$$

One can derive the following corollary from Lemma 3.2 in the same way that one derives the Cauchy integral theorem from Goursat's lemma.

Corollary 3.3. *Let $G \subset \mathbb{C}$ be a star-shaped domain, and let $f\colon G \to \mathbb{C}$ be a continuous function that is holomorphic everywhere except possibly at z_0. Then for every closed path of integration γ in G,*

$$\int_\gamma f(z)\,dz = 0.$$

In order to prove (2), we differentiate the integral in (2) with respect to z and, by Prop. I.5.6, obtain

$$\frac{\partial}{\partial z}\int_{\partial D} \frac{d\zeta}{\zeta - z} = \int_{\partial D} \frac{d\zeta}{(\zeta - z)^2}.$$

The latter integral is equal to zero because $(\zeta - z)^{-2}$ has the primitive $\zeta \mapsto -(\zeta - z)^{-1}$ on $\mathbb{C} \setminus \{z\}$. Therefore,

$$\int_{\partial D} \frac{d\zeta}{\zeta - z}$$

is constant on D; we have already determined its value at the centre z_0 to be $2\pi i$. This completes the proof of Thm. 3.1. □

The Cauchy integral formula has many important consequences. We derive the first ones now, keeping the notation used in Thm. 3.1.

The integrand in the formula

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

is holomorphic in $z \in D$, so that, as before, we can differentiate with respect to z under the integral sign:

$$f'(z) = \frac{1}{2\pi i} \int_{\partial D} \frac{f(\zeta)\, d\zeta}{(\zeta - z)^2} \quad \text{for } z \in D. \tag{3}$$

Here the integrand is again holomorphic in $z \in D$. As before, it follows that f' is holomorphic on D, and we have

$$f''(z) = \frac{2}{2\pi i} \int_{\partial D} \frac{f(\zeta)\, d\zeta}{(\zeta - z)^3} \quad \text{for } z \in D.$$

Continuing in this way, we see that the holomorphic function f is infinitely often complex differentiable on D and that its derivatives satisfy the formula

$$f^{(n)}(z) = \frac{n!}{2\pi i} \int_{\partial D} \frac{f(\zeta)\, d\zeta}{(\zeta - z)^{n+1}} \quad \text{for } z \in D \text{ and } n = 0, 1, 2, 3, \ldots$$

Here the disk $D \subset\subset G$ was arbitrary. Since every point in G belongs to such a disk, we conclude

Theorem 3.4. *The derivative of a holomorphic function is again holomorphic. Holomorphic functions are infinitely often complex differentiable.*

Theorem 3.5 (The Cauchy integral formula)**.** *If $f\colon G \to \mathbb{C}$ is holomorphic and the disk D is relatively compact in G, then for every $z \in D$ and $n = 0, 1, 2, \ldots$,*

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

Thm. 3.4 shows that complex differentiability is a condition of far more consequence than real differentiability: The derivative of a real differentiable function need not be continuous, let alone differentiable.

With the help of Thm. 3.4, we can derive several criteria for a function to be holomorphic. The first of these is a converse of Goursat's lemma (Thm. 2.1).

Proposition 3.6 (Morera). *Let f be continuous on a domain $G \subset \mathbb{C}$, and suppose that $\int_{\partial\Delta} f(z)\,dz = 0$ for every closed triangle $\Delta \subset G$. Then f is holomorphic on G.*

Proof: Let D be an arbitrary disk in G. Given our assumptions, Prop. 1.3 implies that f has a primitive F on D. By Thm. 3.4, $f = F'$ is holomorphic on D. □

Remark: Let $f\colon G \to \mathbb{C}$ be continuous on the domain G and holomorphic on $G\setminus\{z_0\}$ for some $z_0 \in G$. It follows from Lemma 3.2 that $\int_{\partial\Delta} f(z)\,dz = 0$ for every closed triangle Δ in G, so that f is holomorphic on all of G, by Morera's theorem. The condition "continuous, and holomorphic with the possible exception of one point" thus only seems to be more general than the condition "holomorphic". We required the former, however, in order to prove the Cauchy integral formula and Morera's theorem.

To conclude, we give an often used criterion that substantially sharpens the above remark.

Theorem 3.7 (The Riemann extension theorem). *Let $G \subset \mathbb{C}$ be a domain, let $z_0 \in G$, and assume that the function f is holomorphic on $G\setminus\{z_0\}$ and bounded near z_0. Then f can be uniquely extended to a holomorphic function on all of G.*

Here "f is bounded near z_0" means that there is a neighbourhood $U \subset G$ of z_0 and a bound M such that $|f(z)| \leqslant M$ for all $z \in U \setminus \{z_0\}$.

Proof: Since f is bounded near z_0, we have $\lim\limits_{z\to z_0}(z - z_0)f(z) = 0$. Therefore, the function

$$F(z) = \begin{cases} (z - z_0)f(z) & \text{for } z \neq z_0 \\ 0 & \text{for } z = z_0 \end{cases}$$

is holomorphic on $G \setminus \{z_0\}$ and continuous at z_0. In light of the above remark, F is holomorphic on all of G. We have

$$F'(z_0) = \lim_{z\to z_0} \frac{F(z) - F(z_0)}{z - z_0} = \lim_{z\to z_0} f(z);$$

in particular, the latter limit exists. This means that f can be uniquely extended to a continuous function $\hat{f}$ on G. Applying the above remark to $\hat{f}$ proves the claim. □

The extension of f to $\hat{f}$ means that the gap in the domain of definition of f can be removed. Thm. 3.7 is thus frequently referred to as *Riemann's removable singularity theorem.*

Example:
Let $f(z) = \frac{\sin z}{z}$ for $z \neq 0$. The power series of the sine function gives us

$$f(z) = 1 - \frac{1}{3!}z^2 + \dots,$$

so $\lim_{z \to 0} f(z) = 1$, and by setting $f(0) = 1$, f becomes a holomorphic function on all of $\mathbb{C}$.

Exercises

1. Using the Cauchy integral formulas, compute the following integrals:

   a) $\displaystyle\int_{|z+1|=1} \frac{dz}{(z+1)(z-1)^3}$, b) $\displaystyle\int_{|z-i|=3} \frac{dz}{z^2+\pi^2}$,

   c) $\displaystyle\int_{|z|=1/2} \frac{\exp(1-z)\,dz}{z^3(1-z)}$, d) $\displaystyle\int_{|z-1|=1} \left(\frac{z}{z-1}\right)^n dz \quad (n \geqslant 1)$.

2. Let f be holomorphic in a neighbourhood of the closed disk $\overline{D} \subset \mathbb{C}$. Show that

   $$z \mapsto \int_{\partial D} \frac{f(\zeta)\,d\zeta}{\zeta - z}$$

   determines a holomorphic function on $\mathbb{C} \setminus \overline{D}$. Which function is this?

3. Let f be continuous on the closed disk $\overline{D} \subset \mathbb{C}$ and holomorphic on D. Show that

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

   for $z \in D$ (cf. Ex. 1 in II.2).

4. Check whether the following functions can be holomorphically extended to $z_0 = 0$.

   $z \cot^2 z$; $z^2 \cot^2 z$; $z\,(e^z - 1)^{-1}$; $z^2 \sin(1/z)$.

4. Power series expansions of holomorphic functions

In Chapter I, we saw that a convergent power series determines a holomorphic function in its disk of convergence. We will now prove that a holomorphic function defined on a disk $D = D_r(z_0)$ has a power series expansion with base point z_0 and radius of convergence at least r. This is again a difference between real and complex analysis: the radius of convergence of the Taylor series of an infinitely often real differentiable function f can be 0, and even if it is positive, the series need not converge to f.

Theorem 4.1. *Let f be holomorphic on the domain $G \subset \mathbb{C}$, and let $z_0 \in G$. Then f can be represented by its Taylor series*

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n, \quad \text{where } a_n = \frac{1}{n!} f^{(n)}(z_0),$$

in a neighbourhood of z_0*. This series converges to* f *on the largest disk centred at* z_0 *that is still contained in* G*.*

It is possible that the Taylor series $T(z)$ of f converges in a larger disk $D(z_0)$ than is guaranteed by the above theorem. Whether it is true that $f(z) \equiv T(z)$ on $G \cap D(z_0)$ must then be checked on a case-by-case basis: in general, it is not true. Clarifying the relationship between $f(z)$ and $T(z)$ leads to the theory of Riemann surfaces (cf. [FL2, Fo2]).

Proof of Thm. 4.1**:** a) If a function f that is holomorphic on G can be represented as a power series $f(z) = \sum_{n=0}^{\infty} a_n(z-z_0)^n$ in a neighbourhood of z_0, then

$$f^{(k)}(z_0) = k!a_k$$

for $k = 0, 1, 2, \ldots$; in particular, these coefficients are unique (see Cor. I.3.8).

b) Choose a radius r such that $D = D_r(z_0) \subset\subset G$. In the Cauchy integral formula

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

expand the term $\dfrac{1}{\zeta - z}$ into a geometric series with respect to powers of $\dfrac{z - z_0}{\zeta - z_0}$:

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

For a fixed $z \in D$, this series converges uniformly on ∂D. Since f is bounded on ∂D, the series

$$\sum_{n=0}^{\infty} \frac{f(\zeta)}{(\zeta - z)^{n+1}} (z - z_0)^n$$

also converges uniformly on ∂D, and we may interchange the order of summation and integration in the following calculation:

$$\begin{aligned} f(z) &= \frac{1}{2\pi i} \int_{\partial D} \left(\sum_{n=0}^{\infty} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} (z - z_0)^n \right) d\zeta \\ &= \sum_{n=0}^{\infty} \left(\frac{1}{2\pi i} \int_{\partial D} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}}\, d\zeta \right) (z - z_0)^n, \end{aligned}$$

where $z \in D$. This is the sought-after expansion of f into a power series; by the Cauchy integral formula for derivatives (Thm. 3.5), the coefficients in parentheses are indeed the values $f^{(n)}(z_0)/n!$. Note that, together with part a), this calculation yields a new proof of Thm. 3.5.

c) Let $D_R(z_0)$ be the largest disk centred at z_0 that is contained in G (in case $G = \mathbb{C}$, note that $R = \infty$). Given a point $z \in D_R(z_0)$, we can choose a radius r such that $|z - z_0| < r < R$. By part b), the above representation of f as a power series is valid at z. □

Examples of power series expansions:

The coefficient formula $a_n = f^{(n)}(z_0)/n!$ is usually of little help, because higher derivatives often become intractable. For rational functions, Example *i* gives us a method for finding Taylor expansions. The remaining examples show how one can often attain at least the beginning of a Taylor series.

i. The function $f(z) = (z - a)^{-1}$ can be expanded into a power series with radius of convergence $|z_0 - a|$ about every point $z_0 \neq a$ (cf. the proof of Thm. 4.1):

$$\frac{1}{z-a} = \sum_{n=0}^{\infty} a_n (z - z_0)^n, \quad \text{where } a_n = -(a - z_0)^{-n-1}.$$

For $g(z) = (z - a)^{-2}$, we have $g(z) = -f'(z)$, so that

$$\frac{1}{(z-a)^2} = -\sum_{n=1}^{\infty} n a_n (z - z_0)^{n-1} = -\sum_{n=0}^{\infty} (n+1) a_{n+1} (z - z_0)^n.$$

Higher powers of $(z - a)^{-1}$ are handled similarly. With more complicated rational functions, one decomposes the function into partial fractions and combines the corresponding power series. In III.2, we will study partial fraction decompositions systematically; for now, let us give a concrete example, namely

$$h(z) = \frac{z+1}{(z-1)^2(z-2)} = \frac{3}{z-2} - \frac{3}{z-1} - \frac{2}{(z-1)^2}.$$

The power series expansion about $z_0 = 0$ is thus

$$\begin{aligned} h(z) &= -3 \sum_{n=0}^{\infty} 2^{-n-1} z^n + 3 \sum_{n=0}^{\infty} z^n - 2 \sum_{n=0}^{\infty} (n+1) z^n \\ &= \sum_{n=0}^{\infty} (3 - 3 \cdot 2^{-n-1} - 2(n+1)) z^n \quad \text{for } |z| < 1. \end{aligned}$$

ii. Let the functions f and g be holomorphic in a neighbourhood of z_0. Leibniz's rule for the derivatives of the product fg says that

$$(fg)^{(n)}(z_0)/n! = \sum_{m=0}^{n} \frac{f^{(m)}(z_0)}{m!} \cdot \frac{g^{(n-m)}(z_0)}{(n-m)!}.$$

If $f(z) = \sum_{n=0}^{\infty} a_n(z-z_0)^n$ and $g(z) = \sum_{n=0}^{\infty} b_n(z-z_0)^n$ are the Taylor series of f and g, then the Taylor series of fg is therefore

$$(fg)(z) = \sum_{n=0}^{\infty}\Big(\sum_{m=0}^{n} a_m b_{n-m}\Big)(z-z_0)^n.$$

One can also obtain this series by formal multiplication of the series for f and g.

iii. Suppose $f(z) = \sum_{n=0}^{\infty} a_n(z-z_0)^n$ in a neighbourhood of z_0, and assume that $f(z_0) = a_0 \neq 0$. Then $1/f$ is holomorphic near z_0, and the coefficients of the expansion

$$\frac{1}{f(z)} = \sum_{n=0}^{\infty} b_n(z-z_0)^n$$

can be determined from

$$1 = \frac{1}{f(z)}\cdot f(z) = \sum_{n=0}^{\infty}\Big(\sum_{m=0}^{n} b_m a_{n-m}\Big)(z-z_0)^n$$

by comparing coefficients; we obtain

$$b_0 a_0 = 1 \text{ and } \sum_{m=0}^{n} b_m a_{n-m} = 0 \text{ for } n \geqslant 1.$$

Thus,

$$b_0 = a_0^{-1}, \quad b_1 = -a_1 a_0^{-2}, \quad b_2 = (a_1^2 - a_0 a_2)a_0^{-3}, \quad \text{etc.}$$

iv. Let us determine the beginning of the power series expansion of $f(z) = \tan z$ about $z_0 = 0$. Note that f is an odd function, i.e. $f(-z) = -f(z)$; the derivatives of even order of f therefore vanish at the origin. It follows that the expansion has the form

$$\tan z = \sum_{n=0}^{\infty} a_{2n+1} z^{2n+1}.$$

The tangent function is holomorphic on $\mathbb{C}$ with the exception of the points $\frac{\pi}{2} + k\pi$, where $k \in \mathbb{Z}$, so that the radius of convergence of the above series is $\frac{\pi}{2}$ – cf. Ex. 5a. One can now obtain the first few coefficients by, for example, substituting the power series for $\sin z$ and $\cos z$ into the equation $\cos z \cdot \tan z = \sin z$ and comparing coefficients:

$$\big(1 - \tfrac{1}{2!}z^2 + \tfrac{1}{4!}z^4 - + \ldots\big)\big(a_1 z + a_3 z^3 + a_5 z^5 + \ldots\big) = z - \tfrac{1}{3!}z^3 + \tfrac{1}{5!}z^5 - + \ldots$$

gives us $a_1 = 1$, $a_3 - \frac{1}{2!}a_1 = -\frac{1}{3!}$, $a_5 - \frac{1}{2!}a_3 + \frac{1}{4!}a_1 = \frac{1}{5!}$, etc., so that $a_3 = \frac{2}{3!}, a_5 = \frac{16}{5!}$, etc.

In order to determine the derivatives $f^{(n)}(z_0)$ of a function f that is holomorphic at z_0, one only needs to know the values $f(z)$ on, say, a segment $(z_0 - \delta, z_0 + \delta)$ parallel to the real axis. If f and g are two functions that are holomorphic at z_0 and coincide on such a segment, then $f^{(n)}(z_0) = g^{(n)}(z_0)$ for all n. The functions f and g thus have the same Taylor series expansion about z_0, so that they in fact coincide on a neighbourhood of z_0! The following proposition sharpens this observation.

Proposition 4.2 (The identity theorem)**.** *Let G be a domain in $\mathbb{C}$, and let f and g be holomorphic functions on G. The following are equivalent:*

i. There exists a point $z_0 \in G$ and a sequence z_μ in $G \setminus \{z_0\}$ converging to z_0 such that $f(z_\mu) = g(z_\mu)$ for $\mu \geqslant 1$.

ii. $f \equiv g$.

iii. There exists a point $z_0 \in G$ such that $f^{(n)}(z_0) = g^{(n)}(z_0)$ for all $n \geqslant 0$.

Proof: By passing to the function $f - g$, we may assume that $g \equiv 0$.

a) Claims *i* and *iii* follow trivially from *ii*.

b) Let us assume that *i* holds and show that *iii* holds with the same z_0. Consider the Taylor series expansion of f about z_0:

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n.$$

We have $a_0 = f(z_0) = \lim_\mu f(z_\mu) = 0$. Suppose now that $a_0 = \ldots = a_{n-1} = 0$ for some $n \geqslant 1$. Then

$$f(z) = (z - z_0)^n (a_n + a_{n+1}(z - z_0) + \ldots) = (z - z_0)^n F(z),$$

where F is continuous (and in fact holomorphic) at z_0. Since $z_\mu - z_0 \neq 0$, we have $F(z_\mu) = 0$ for all μ, and hence $a_n = \lim_\mu F(z_\mu) = 0$. Thus, for all n,

$$f^{(n)}(z_0) = n!\, a_n = 0.$$

c) We now derive *ii* from *iii* By Thm. 4.1, $f \equiv 0$ in a neighbourhood of z_0. Consider the set

$$M = \{z_1 \in G \colon f \equiv 0 \text{ in a neighbourhood of } z_1\}.$$

By definition, M is open, and $z_0 \in M$. But M is also closed in G: If $z_2 \in G$ is a boundary point of M, then there exists a sequence z_μ in $M \setminus \{z_2\}$ converging to z_2. By part b), we then have $f^{(n)}(z_2) = 0$ for all $n \geqslant 0$, from which it follows that $f \equiv 0$ in a neighbourhood of z_2, i.e. $z_2 \in M$. Since G is a domain, $M = G$, so that $f \equiv 0$ on G. □

By the identity theorem, a holomorphic function on a domain G is completely determined by its values on any nondiscrete set in G. Here a set $M \subset G$ is called *nondiscrete* in G if G contains an accumulation point of M; if this is not the case, then M is called *discrete* in G. Properties that can be expressed via identities between holomorphic functions on G thus only need to be verified on a nondiscrete set in G, e.g. on a curve $C \subset G$, in order to be satisfied on all of G – they "propagate from C to G". Let us illustrate this important principle with a typical example. The function $f(z) = \cot \pi z$ is holomorphic on $\mathbb{C} \setminus \mathbb{Z}$ and has period 1 on $\mathbb{R} \setminus \mathbb{Z}$. This means that the holomorphic functions

$$z \mapsto f(z), \quad z \mapsto f(z+1)$$

coincide on the nondiscrete set $\mathbb{R} \setminus \mathbb{Z} \subset \mathbb{C} \setminus \mathbb{Z}$ and hence coincide everywhere. The periodicity of the real cotangent function thus implies the periodicity of the complex cotangent function.

In particular, the identity theorem implies that the extension of the *real* elementary functions $\sin\colon \mathbb{R} \to \mathbb{R}$, etc. to holomorphic functions can be accomplished in only one way, since $\mathbb{R}$ is nondiscrete in $\mathbb{C}$.

The notion of the order of a zero of a polynomial can be transferred to holomorphic functions as follows:

Definition 4.1. *A holomorphic function f has a zero of order (or multiplicity) k at z_0 if*

$$f(z_0) = f'(z_0) = \ldots = f^{(k-1)}(z_0) = 0, \quad f^{(k)}(z_0) \neq 0.$$

It assumes the value w at z_0 with order (or multiplicity) k if $f - w$ has a zero of order k at z_0.

Here one may allow $k = \infty$. If f takes on the value w with order ∞ at z_0, then $f \equiv w$ in a neighbourhood of z_0 – see Prop. 4.2.

The function $(z - z_0)^k$ has a zero of order k at z_0. More generally, it is easy to show the equivalence of the following claims:

i. *The function f has a zero of order k at z_0.*

ii. *The Taylor series expansion of f about z_0 is*

$$f(z) = \sum_{n=k}^{\infty} a_n (z - z_0)^n, \quad \textit{where } a_k \neq 0.$$

iii. *In a neighbourhood of z_0, one can write*

$$f(z) = (z - z_0)^k g(z),$$

where g is holomorphic and satisfies $g(z_0) \neq 0$.

The identity theorem shows that if G is a domain and $f\colon G \to \mathbb{C}$ is a nonconstant holomorphic function, then for every $w \in \mathbb{C}$, the points at which f takes on the value w are isolated, i.e. the set $f^{-1}(w) = \{z \in G\colon f(z) = w\}$ does not have an accumulation point in G (it may of course be empty!).

To conclude this section, we give several characterizations of the term "holomorphic function". We leave it to the reader to verify how they follow from the theorems we have proved thus far.

Theorem 4.3. *Let $f\colon U \to \mathbb{C}$ be a function defined on an open set $U \subset \mathbb{C}$. The following are equivalent:*

i. f is holomorphic.

ii. f is real differentiable and satisfies the Cauchy-Riemann equations.

iii. f admits a power series expansion about every point in U.

iv. f has local primitives.

v. f is continuous, and for every closed triangle $\Delta \subset U$, $\int_{\partial\Delta} f(z)\,dz = 0$.

Exercises

1. Determine the power series expansion of the following functions about $z_0 = 0$:

$$\frac{2z+1}{(z^2+1)(z+1)^2}, \qquad \sin^2 z, \qquad e^z \cos z.$$

Hint: It is sometimes helpful to express trigonometric functions in terms of the exponential function.

2. Assume that the power series $f(z) = \sum_0^\infty a_n z^n$ converges on $D = D_r(0)$. Show that:

– If f is real-valued on $\mathbb{R} \cap D$, then all a_n are real.

– If f is an even (odd) function, then $a_n = 0$ for all odd (even) n.

– If $f(iz) = f(z)$, then a_n can only be nonzero if n is divisible by 4.

In addition: Discuss the equation $f(\rho z) = \mu f(z)$, where $\rho, \mu \in \mathbb{C} \setminus \{0\}$ are given.

3. There are only even powers of z in the power series expansion of $f(z) = \frac{1}{\cos z}$ about 0. It is most often written in the form

$$\frac{1}{\cos z} = \sum_{n=0}^{\infty} (-1)^n \frac{E_{2n}}{(2n)!} z^{2n}.$$

The E_{2n} are called *Euler numbers*. Determine a recursion formula for the numbers E_{2n} and show that they are all integers. In addition: Show that $(-1)^n E_{2n}$ is always positive. Furthermore: Show that $(-1)^n E_{2n} \equiv 0 \mod 5$ for even $n > 0$ and that $(-1)^n E_{2n} \equiv 1 \mod 5$ for odd n.

4. Prove that if f and g are holomorphic on a domain G and $fg \equiv 0$, then $f \equiv 0$ or $g \equiv 0$.

5. a) Let the function f be holomorphic on $D_r(0)$, and suppose there exists a point $z_1 \in \partial D_r(0)$ such that $\lim\limits_{z \to z_1} f(z)$ does not exist. Show that the radius of convergence of the Taylor series expansion of f about 0 is exactly r.

b) Determine the radii of convergence of the Taylor series expansions about 0 of

$$\tan z, \qquad \frac{1}{\cos z}, \qquad \frac{z}{\sin z}, \qquad \frac{z}{e^z - 1}.$$

6. a) Suppose the domain G is symmetric with respect to the real axis and that f is holomorphic on G and real-valued on $G \cap \mathbb{R}$. Show that $f(\overline{z}) = \overline{f(z)}$ for all $z \in G$.

   b) Suppose $G = D_r(0)$ and f is holomorphic on G and real-valued on $G \cap \mathbb{R}$. Show that if f is even (odd), then the values of f on $G \cap i\mathbb{R}$ are real (imaginary). Prove this without using the power series expansion of f.

7. a) Suppose the domain G is symmetric with respect to the real axis and f is continuous on G and holomorphic on $G \setminus \mathbb{R}$. Show that f is holomorphic on all of G.

   Hints: Use Morera's theorem. In splitting up the triangle $\Delta \subset G$, one sees that the only problematic case is the one in which an edge of Δ lies on $\mathbb{R}$. In this case, approximate Δ by a triangle Δ_ε one of whose edges lies on $\mathbb{R} \pm i\varepsilon$, let ε tend to 0, and use the uniform continuity of f on Δ.

   b) Let G be as in part a), and let f be continuous on $G \cap \{z\colon \operatorname{Im} z \geqslant 0\}$, holomorphic on $G \cap \{z\colon \operatorname{Im} z > 0\}$, and real-valued on $G \cap \mathbb{R}$. Show that f can be extended to a function that is holomorphic on all of G (cf. Ex. 6a).

5. Convergence theorems, maximum modulus principle, and open mapping theorem

From the fundamental results of the previous sections, we now deduce further important properties of holomorphic functions. We begin with a useful estimate.

Proposition 5.1 (The Cauchy inequalities). *Let f be holomorphic in a neighbourhood of the closed disk $\overline{D_R(z_0)}$, and let $0 < r < R$. Then for all $z \in \overline{D_r(z_0)}$ and all $n \geqslant 0$,*

$$\left|f^{(n)}(z)\right| \leqslant C \max_{\zeta \in \partial D_R(z_0)} |f(\zeta)|, \tag{1}$$

where C is a constant that depends on n, r, and R and is given by

$$C = \frac{n!R}{(R-r)^{n+1}}. \tag{2}$$

The proof follows immediately from the Cauchy integral formula:

$$\begin{aligned}\left|f^{(n)}(z)\right| &= \left|\frac{n!}{2\pi i} \int\limits_{\partial D_R(z_0)} \frac{f(\zeta)}{(\zeta - z)^{n+1}}\, d\zeta\right| \\ &\leqslant \frac{n!}{2\pi} 2\pi R \frac{1}{(R-r)^{n+1}} \max_{\zeta \in \partial D_R(z_0)} |f(\zeta)|.\end{aligned}$$

If one sets $r = 0$ in (2), then, in particular,

$$\left|f^{(n)}(z_0)\right| \leqslant \frac{n!}{R^n} \max_{\zeta \in \partial D_R(z_0)} |f(\zeta)|. \tag{3}$$

One now easily obtains the following convergence theorem due to Weierstrass:

Theorem 5.2. *Suppose the sequence of holomorphic functions* (f_ν) *converges locally uniformly to the function* f *on a domain* G. *Then* f *is holomorphic, and the sequence of* n*th derivatives* $(f_\nu^{(n)})$ *converges locally uniformly to* $f^{(n)}$ *for all* n.

Proof: a) The function f is continuous. If Δ is a closed triangle in G, then

$$\int_{\partial\Delta} f(z)\,dz = \int_{\partial\Delta} \lim_{\nu\to\infty} f_\nu(z)\,dz = \lim_{\nu\to\infty} \int_{\partial\Delta} f_\nu(z)\,dz = 0$$

by Goursat's lemma. Thus, f is holomorphic by Morera's theorem.

b) Let $D_R(z_0) \subset\subset G$ and $0 < r < R$. By applying (1) to $f_\nu - f$, we have

$$\left|f_\nu^{(n)}(z) - f^{(n)}(z)\right| \leqslant C \max_{\zeta\in\partial D_R(z_0)} |f_\nu(\zeta) - f(\zeta)|$$

for $z \in \overline{D_r(z_0)}$. The uniform convergence of the sequence f_ν on the compact set $\partial D_R(z_0)$ therefore implies the uniform convergence of the $f_\nu^{(n)}$ on $\overline{D_r(z_0)}$. □

The above theorem again shows how complex analysis differs from real analysis: the limit of a uniformly convergent sequence of differentiable functions need not be differentiable. Moreover the theorem yields a new proof (which does not depend on real analysis) that the sum of a power series is holomorphic – cf. Thm. I.3.7.

Another fundamental convergence property is the following

Theorem 5.3 (Montel)**.** *Let* (f_ν) *be a locally uniformly bounded sequence of holomorphic functions in the domain* G. *Then* (f_ν) *has a locally uniformly convergent subsequence.*

The proof rests on a topological statement. We recall the pertinent definition: A sequence (f_ν) of complex valued functions defined on a subset $S \subset \mathbb{R}^n$ is called *equicontinuous* on S, if for every $\varepsilon > 0$ there is a $\delta = \delta(\varepsilon) > 0$ such that $\left|f_\nu(\mathbf{x}') - f_\nu(\mathbf{x}'')\right| < \varepsilon$ for all $\mathbf{x}'$, $\mathbf{x}'' \in S$ with $|\mathbf{x}' - \mathbf{x}''| < \delta$ and all ν. – The sequence is locally equicontinuous on an open subset $U \subset \mathbb{R}^n$, if each point of U has a neighbourhood where the sequence is equicontinuous. This is equivalent to equicontinuity on all compact subsets of U.

Proposition 5.4 (Ascoli-Arzelà)**.** *Let* (f_ν) *be a locally bounded and locally equicontinuous sequence of complex-valued functions on an open set* $U \subset \mathbb{R}^n$. *Then* (f_ν) *has a locally uniformly convergent subsequence.*

Proof: a) Let $M = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots\}$ be a denumerable subset of U. Then (f_ν) has a subsequence which converges pointwise on M. Namely, the values $f_\nu(\mathbf{x}_1)$ being bounded, there is a subsequence of (f_ν), which we denote by $(f_{\nu 1})$, such that the sequence of values $f_{\nu 1}(\mathbf{x}_1)$ converges. The values $f_{\nu 1}(\mathbf{x}_2)$ are also bounded, hence there is a subsequence $(f_{\nu 2})$ of $(f_{\nu 1})$ such that the values $f_{\nu 2}(\mathbf{x}_2)$ converge. Continuing like this, we get subsequences $(f_{\nu\varkappa})$, $\varkappa = 1, 2, 3, \ldots$, such that

$$f_{\nu\varkappa} \text{ is a subsequence of } f_{\nu,\varkappa-1} \text{ and } f_{\nu\varkappa}(\mathbf{x}_k) \text{ converges for } 1 \leqslant k \leqslant \varkappa.$$

Now consider the diagonal sequence $(f_{\nu\nu})_{\nu\geqslant 1}$. Since $(f_{\nu\nu})_{\nu\geqslant\varkappa}$ is a subsequence of $(f_{\nu\varkappa})$, the values $f_{\nu\nu}(\mathbf{x}_k)$ converge for every $\mathbf{x}_k$.

b) Now let (f_ν) be given as in the proposition and let $M = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots\}$ be a dense denumerable subset of U, e.g. $M = U \cap \mathbb{Q}^n$. By a) we may assume that (f_ν) converges pointwise on M. Let $V \subset\subset U$ be an open neighbourhood of some point $\mathbf{x}_0 \in U$, on which (f_ν) is equicontinuous. We show that (f_ν) converges uniformly on V. Fix $\varepsilon > 0$ and choose $\delta = \delta(\varepsilon)$ according to the definition of equicontinuity. Because $M \cap V$ is dense in V and, consequently, in $\overline{V}$, we can cover $\overline{V}$ by finitely many balls $D_\delta(\mathbf{x}_{k_\lambda})$, $\lambda = 1, \ldots, \ell$, with centres $\mathbf{x}_{k_\lambda}$ in V. As $f_\nu(\mathbf{x}_{k_\lambda})$ converges, there is an index ν_0 such that

$$\left|f_\nu(\mathbf{x}_{k_\lambda}) - f_\mu(\mathbf{x}_{k_\lambda})\right| < \varepsilon \text{ for } \lambda = 1, \ldots, \ell \text{ and } \mu, \nu \geqslant \nu_0.$$

Now let $\mathbf{x} \in V$ be arbitrary. Choose λ with $\mathbf{x} \in D_\delta(\mathbf{x}_{k_\lambda})$ and let $\nu, \mu \geqslant \nu_0$. Then

$$\left|f_\nu(\mathbf{x}) - f_\mu(\mathbf{x})\right| \leqslant \left|f_\nu(\mathbf{x}) - f_\nu(\mathbf{x}_{k_\lambda})\right| + \left|f_\nu(\mathbf{x}_{k_\lambda}) - f_\mu(\mathbf{x}_{k_\lambda})\right| + \left|f_\mu(\mathbf{x}_{k_\lambda}) - f_\mu(\mathbf{x})\right| < 3\varepsilon$$

by the choice of δ and ν_0. Hence the Cauchy criterion for uniform convergence on V is satisfied. □

Proof of Thm. 5.3**:** Local uniform boundedness of holomorphic functions f_ν implies local equicontinuity: If, on $D_r(z_0) \subset\subset G$, we have $|f_\nu(z)| \leqslant M_0$ for all ν, the Cauchy inequalities (Prop. 5.1) imply $|f'_\nu(z)| \leqslant M_1$ for $z \in \overline{D_{r/2}(z_0)}$ and all ν, with a suitable constant M_1. Hence, for $z', z'' \in \overline{D_{r/2}(z_0)}$ and all ν

$$\left|f_\nu(z') - f_\nu(z'')\right| = \left|\int_{[z',\, z'']} f'_\nu(\zeta)\, d\zeta\right| \leqslant M_1|z' - z''|,$$

so (f_ν) is equicontinuous on $\overline{D_{r/2}(z_0)}$. Now Prop. 5.4 yields the claim. □

Using different terminology, we can formulate this theorem as follows: Every locally bounded family of holomorphic functions is normal. We will prove a deeper theorem on normal families in the last chapter.

Let us turn our attention to the value distribution of a holomorphic function. As is well known, a real continuous function on the plane generally has local minima and maxima. However:

Theorem 5.5 (The maximum modulus principle)**.** *Let f be holomorphic on a domain G. If $|f|$ has a local maximum at $z_0 \in G$, then f is constant.*

Thus, local maxima of $|f|$ cannot occur if f is a nonconstant holomorphic function, and local minima can occur only at the zeros of f. Namely, if f has no zeros on G, then $1/f$ is also holomorphic on G, and the local minima of $|f|$ are the local maxima of $|1/f|$. We thus note the following consequence of the maximum modulus principle:

Proposition 5.6 (The minimum modulus principle)**.** *If f is holomorphic and nonzero on G and $|f|$ has a local minimum at $z_0 \in G$, then f is constant.*

Proof of Thm. 5.5**:** Suppose that

$$|f(z)| \leqslant |f(z_0)| \tag{4}$$

for all $z \in D_R(z_0) \subset\subset G$. The Cauchy integral formula gives us

$$f(z_0) = \frac{1}{2\pi i} \int\limits_{|z-z_0|=r} \frac{f(z)}{z - z_0}\, dz = \frac{1}{2\pi} \int\limits_0^{2\pi} f(z_0 + r\, e^{it})\, dt,$$

where $r \leqslant R$ is arbitrary. Due to (4), we have

$$|f(z_0)| \leqslant \frac{1}{2\pi} \int\limits_0^{2\pi} |f(z_0 + r\, e^{it})|\, dt \leqslant \frac{1}{2\pi} \int\limits_0^{2\pi} |f(z_0)|\, dt = |f(z_0)|,$$

so that

$$\int\limits_0^{2\pi} \big(|f(z_0)| - |f(z_0 + r\, e^{it})|\big)\; dt = 0.$$

Since the integrand is continuous and nonnegative, it must be identically zero; thus,

$$|f(z_0)| = |f(z_0 + r\, e^{it})|$$

for all $r \leqslant R$, i.e. $|f|$ is constant on $D_R(z_0)$.

Differentiating the equation

$$f \cdot \overline{f} = \text{const} \quad (\text{on } D_R(z_0))$$

with respect to z (Wirtinger derivative!), we obtain

$$f_z \cdot \overline{f} = 0.$$

If $\overline{f} \equiv 0$, then there is nothing to show; otherwise, $f_z \equiv 0$, and f is constant on $D_R(z_0)$. The identity theorem then implies that f is constant on G. □

Since a real continuous function takes on its maximum and minimum on a compact set, the previous propositions imply:

Corollary 5.7. *Let $G \subset\subset \mathbb{C}$ be a bounded domain, and let $f : \overline{G} \to \mathbb{C}$ be continuous on $\overline{G}$ and holomorphic on G.*

i. The modulus $|f|$ attains its maximum on the boundary ∂G; if f is nonconstant, then for all $z \in G$,

$$|f(z)| < \max_{z \in \partial G} |f(z)|.$$

ii. If f has no zeros in G, then $|f|$ attains its minimum on ∂G; if f is nonconstant, then for all $z \in G$,

$$|f(z)| > \min_{z \in \partial G} |f(z)|.$$

Claim *ii* can be used to prove the existence of zeros: if, in the above situation, there is a point $z_0 \in G$ such that

$$|f(z_0)| < \min_{z \in \partial G} |f(z)|,$$

then f must have a zero in G.

Theorem 5.8 (The open mapping theorem)**.** *Let f be a nonconstant holomorphic function on a domain $G \subset \mathbb{C}$. Then $f(G)$ is a domain.*

Proof: We only need to check that $f(G)$ is open. Let $w_0 = f(z_0)$, and choose a radius r such that $D = D_r(z_0) \subset\subset G$ and f does not assume the value w_0 on the boundary $\partial D_r(z_0)$ – this is certainly possible, since the points at which f equals w_0 are isolated. Then

$$\delta = \min_{z \in \partial D_r(z_0)} |f(z) - w_0| > 0.$$

If $|w - w_0| < \delta/2$, then for $z \in \partial D$, we have

$$|f(z) - w| \geqslant |f(z) - w_0| - |w - w_0| > \delta - \delta/2 = \delta/2,$$

but

$$|f(z_0) - w| < \delta/2.$$

By the minimum modulus principle (see the remark following Cor. 5.7), $f(z) - w$ has a zero in D, i.e. f takes on the value w in D. We thus see that $D_{\delta/2}(w_0) \subset f(G)$, and $f(G)$ is open. □

Nonconstant holomorphic functions are thus *open mappings*: they send open sets to open sets. We may thus improve the inverse function theorem (Prop. I.1.4) considerably (the following proof is taken from [RS]):

Theorem 5.9 (The inverse function theorem)**.** *Let $f\colon G \to G'$ be a bijective holomorphic function. Then if G is a domain, G' is a domain as well, and f^{-1} is holomorphic on G'.*

Proof: We already know that G' is a domain and that f is an open mapping and hence a homeomorphism from G to G'. The set of zeros of f' is a discrete subset M of G, since f' is holomorphic and $f' \not\equiv 0$. Thus $M' = f(M)$ is also discrete in G', and by Prop. I.1.5, the map $f\colon G \setminus M \to G' \setminus M'$ is biholomorphic. It follows that f^{-1} is holomorphic on $G' \setminus M'$ and continuous on G'. By the Riemann removable singularity theorem, f^{-1} is therefore holomorphic on all of G'. Since $f \circ f^{-1} = \mathrm{id}$, i.e. $f' \cdot (f^{-1})' \equiv 1$, we see that in fact $M = M' = \varnothing$. □

Let us explicitly note: the derivative of an injective holomorphic function f on a domain U has no zeros.

Exercises

Notation: If $f\colon D_R(z_0) \to \mathbb{C}$ is continuous, then for $0 < r < R$, we put

$$M_r(f) = M_r(f; z_0) = \max\{|f(z)|\colon |z - z_0| = r\}.$$

1. Suppose $f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$ converges on $D_R(z_0)$. Show that:
   a) $|a_n| \leqslant r^{-n} M_r(f)$ for $0 < r < R$.
   b) If f is nonconstant, then $M_r(f)$ is a strictly increasing function of r.
2. Verify that the series $\sum_{n=1}^{\infty} \frac{z^n}{1-z^n}$ defines a holomorphic function on $\mathbb{D}$, and find its power series expansion about 0.
3. Consider a polynomial $p(z) = z^n + a_{n-1}z^{n-1} + \ldots + a_0$.
   a) Show that if $|p(z)| \leqslant 1$ on the circle $|z| = 1$, then $p(z) = z^n$.
   b) Show that $r^{-n} M_r(p)$ is strictly decreasing as a function of r on $]0, +\infty[$ unless $p(z) = z^n$.
   Hint: Consider also $q(z) = 1 + a_{n-1}z + \ldots + a_0 z^n$.
4. Let G be a bounded domain, and let f be continuous on $\overline{G}$ and holomorphic on G. Show that if $|f|$ is constant on ∂G, then f is constant or f has a zero.
5. Let f and g be holomorphic on $\mathbb{D}$ and continuous and nonzero on $\overline{\mathbb{D}}$. Show that if $|f| = |g|$ on $\partial\mathbb{D}$, then $f = cg$ for some constant c. Does this remain true if one allows zeros in $\mathbb{D}$?
6. Let $f(z) = a_0 + a_m z^m + a_{m+1} z^{m+1} + \ldots$ be a convergent power series such that $a_m \neq 0$ (here $m \geqslant 1$). Prove that for a sufficiently small r, $M_r(f) > |a_0|$. Prove this without using the maximum modulus principle, and then derive the maximum modulus principle from the claim.
   Hint: In the case where $a_0 \neq 0$, one can assume that $a_0 = 1$. Then choose z_1 such that $a_m z_1^m > 0$.
7. a) Let $f(z) = 1 + \sum_1^{\infty} a_n z^n$ on $D_R(0)$. Show that if $0 < \rho < R$ and $r = \frac{\rho}{1+M_\rho(f)}$, then f has no zeros in $D_r(0)$.
   Hint: Use the fact that $|a_n| \leqslant \rho^{-n} M_\rho(f)$ to estimate $f(z) - 1$.
   b) For an arbitrary holomorphic function $f \not\equiv 0$ on $D_R(0)$, use part a) to give a radius r such that f has at most the zero $z_0 = 0$ in $D_r(0)$.
8. Let (f_ν) be a locally equicontinuous sequence of continuous functions $G \to \mathbb{C}$. Show that (f_ν) is equicontinuous on every compact $K \subset G$.

6. Isolated singularities and meromorphic functions

Functions like $\tan z = \frac{\sin z}{\cos z}$, $\frac{\sin z}{z}$, $\exp(1/z)$, or rational functions are holomorphic outside a discrete set, namely the set of zeros of their denominators. We now investigate this in general. In doing so, it is convenient to use the following terminology: If U is a neighbourhood of the point z_0, then we say that $U \setminus \{z_0\}$ is a *punctured neighbourhood* of z_0.

Definition 6.1. *Let the function f be holomorphic on a punctured neighbourhood of $z_0 \in \mathbb{C}$. Then z_0 is called an isolated singularity of f.*

We already saw an occurrence of this in II.3 and proved the Riemann extension theorem: If f is bounded near z_0, then $\lim_{z \to z_0} f(z)$ exists, and by defining $f(z_0) = \lim_{z \to z_0} f(z)$, we obtain a function that is holomorphic on all of U. Isolated singularities with this property are therefore called *removable singularities.*

In many cases, the values of a function grow to infinity as one approaches an isolated singularity.

Definition 6.2. *Let z_0 be an isolated singularity of f.*

i. If $\lim_{z \to z_0} |f(z)| = +\infty$, then z_0 is called a pole of f.

ii. If z_0 is neither a removable singularity nor a pole of f, then z_0 is called an essential singularity of f.

Here $\lim_{z \to z_0} |f(z)| = +\infty$ means that, for every $C > 0$, there is an $r > 0$ such that f is defined on $D_r(z_0) \setminus \{z_0\}$ and satisfies $|f(z)| \geqslant C$ there.

An essential singularity thus occurs precisely when there exist sequences z_n, z'_n in the domain of f such that $z_n \to z_0$, $z'_n \to z_0$, and $f(z_n)$ is bounded whereas $f(z'_n)$ is unbounded.

Let us consider the examples mentioned before: The isolated singularities of $\tan z$ are the points $z_k = \frac{\pi}{2} + k\pi$, where $k \in \mathbb{Z}$. Since $\cos z_k = 0$ and $\sin z_k \neq 0$, each z_k is a pole. As we have already seen, the origin is a removable singularity of the function $(\sin z)/z$. If $R(z) = p(z)/q(z)$ is a rational function, then the zeros z_k of the denominator $q(z)$ are isolated singularities of R, and z_k is a pole if $p(z_k) \neq 0$; in case $p(z_k) = 0$, z_k may be a removable singularity. The origin is an essential singularity of the function $f(z) = \exp(1/z)$: one has $\pm\frac{1}{n} \to 0$, but $f(\frac{1}{n}) \to +\infty$ and $f(-\frac{1}{n}) \to 0$ as $n \to \infty$.

Let us study the behaviour of a function near a pole more closely. Let f be holomorphic on a punctured neighbourhood $U \setminus \{z_0\}$ of z_0, and suppose that z_0 is a pole of f. We may assume that f is nonzero on $U \setminus \{z_0\}$ (if not, we can always make U smaller). Then $g(z) = 1/f(z)$ is holomorphic and nonzero on $U \setminus \{z_0\}$, and $\lim_{z \to z_0} |f(z)| = +\infty$ implies that $\lim_{z \to z_0} |g(z)| = 0$. By the Riemann removable singularity theorem, g can be extended to a holomorphic function on all of U – we will also denote this extension g. If k is the order of g at z_0, we may write

$$g(z) = (z - z_0)^k \widetilde{h}(z),$$

where $\widetilde{h}$ is holomorphic on U and $\widetilde{h}(z_0) \neq 0$. Thus $\widetilde{h}$ is nonzero on all of U, $h = 1/\widetilde{h}$ is holomorphic on U, and

$$f(z) = (z - z_0)^{-k} h(z)$$

on $U \setminus \{z_0\}$. We have thus shown:

Proposition 6.1. *Let f be holomorphic on a punctured neighbourhood $U \setminus \{z_0\}$ of z_0, and suppose that z_0 is a pole of f. Then there exists a natural number $k \geqslant 1$ and a holomorphic function h on U with $h(z_0) \neq 0$, such that*

$$f(z) = (z - z_0)^{-k} h(z) \tag{1}$$

on $U \setminus \{z_0\}$. The number k and the function h are uniquely determined by f.

The number k is called the *order* or *multiplicity* of the pole z_0. If $k = 1$, then z_0 is also called a *simple pole*.

As before, suppose z_0 is a pole of order k of $f\colon U \setminus \{z_0\} \to \mathbb{C}$. Then we can substitute the power series expansion $h(z) = \sum_0^\infty b_n (z - z_0)^n$ into (1) to express f as a series:

$$f(z) = (z - z_0)^{-k} \sum_0^\infty b_n (z - z_0)^n = \sum_{n=-k}^\infty a_n (z - z_0)^n,$$

where $a_n = b_{n+k}$ for $n = -k, -k+1, \ldots$. In particular, $a_{-k} = b_0 = h(z_0) \neq 0$.

The series $\sum_0^\infty b_n (z - z_0)^n$ converges locally uniformly on the largest disk $D_R(z_0)$ that is still contained in U, and $(z - z_0)^{-k}$ is bounded on every compact subset of $D_R(z_0) \setminus \{z_0\}$. Thus, our series converges locally uniformly on $D_R(z_0) \setminus \{z_0\}$, which proves most of the following proposition.

Proposition 6.2. *Let f be holomorphic on a punctured neighbourhood $U \setminus \{z_0\}$ of z_0 and have a pole of order k at z_0. Then there exists a series expansion*

$$f(z) = a_{-k}(z - z_0)^{-k} + \ldots + a_{-1}(z - z_0)^{-1} + \sum_{n=0}^\infty a_n (z - z_0)^n, \tag{2}$$

where $a_{-k} \neq 0$. It converges locally uniformly on the largest punctured disk centred at z_0 contained in U. The coefficients a_n are unique: we have

$$a_n = \frac{1}{2\pi i} \int\limits_{|z - z_0| = r} \frac{f(z)\, dz}{(z - z_0)^{n+1}} \tag{3}$$

for every r such that $\overline{D_r(z_0)} \subset U$.

Proof: The uniqueness of the coefficients a_n follows from the uniqueness of the coefficients b_m in the expansion of $h(z)$. Moreover

$$a_n = b_{n+k} = \frac{1}{2\pi i} \int\limits_{|z-z_0|=r} \frac{h(z)\,dz}{(z-z_0)^{n+k+1}} = \frac{1}{2\pi i} \int\limits_{|z-z_0|=r} \frac{f(z)\,dz}{(z-z_0)^{n+1}}. \qquad \square$$

The integral formula for the coefficients a_n formally coincides with the corresponding expression for the coefficients of the Taylor series of a holomorphic function. For $n = -1$ we have the important formula

$$a_{-1} = \frac{1}{2\pi i} \int\limits_{|z-z_0|=r} f(z)\,dz.$$

The series expansion (2) is called the *Laurent series expansion* of f about the pole z_0. The subseries $h_{z_0}(z) = \sum_{n=-k}^{-1} a_n(z-z_0)^n$ describes how f becomes infinite near z_0 and is called the *principal part* of (the Laurent series expansion of) f at z_0.

Without reference to (2), the principal part of f can be characterized as follows: $h_{z_0}(z)$ is the (unique) polynomial in $(z-z_0)^{-1}$ with no constant term for which $f(z)-h_{z_0}(z)$ is holomorphic at z_0.

Examples:

i. The origin is an isolated singularity of $\cot z = \cos z/\sin z$. Since $\cos 0 \neq 0$ and $\sin 0 = 0$, the origin is a pole. The multiplicative decomposition (1) is

$$\cot z = \frac{1}{z} h(z), \quad \text{where } h(z) = \cos z \cdot \frac{z}{\sin z}$$

(the last factor should be defined to take the value 1 at the origin, so that it is holomorphic there). The origin is thus a simple pole, and the principal part of $\cot z$ at 0 is $h(0)/z = 1/z$. Since $\cot(z+\pi) = \cot z$, all of the isolated singularities $z_m = m\pi$, where $m \in \mathbb{Z}$, are simple poles, and their principal parts are $1/(z-z_m)$. The identities $\tan z = -\cot(z-\frac{\pi}{2})$ or $\tan z = 1/\cot z$ now yield the singularities of the tangent function with their principal parts.

ii. The partial fraction decomposition of the rational function

$$R(z) = \frac{z^4+z^2+1}{z(z^2+1)} = z + \frac{1}{z} - \frac{1}{2}\left(\frac{1}{z-i} + \frac{1}{z+i}\right)$$

shows that R has simple poles at 0, i, and $-i$ with principal parts

$$\frac{1}{z}, \quad -\frac{1}{2}\frac{1}{z-i}, \quad -\frac{1}{2}\frac{1}{z+i},$$

respectively. The Laurent series expansion of R about 0 can be obtained by expanding

$$R(z) - \frac{1}{z} = z - \frac{z}{1+z^2}$$

as a power series about 0; we have

$$R(z) = \frac{1}{z} + \sum_{n=1}^{\infty} (-1)^{n+1} z^{2n+1}.$$

As z approaches a removable singularity or a pole of f, the values of f behave in a clear way. The opposite is true of essential singularities:

Proposition 6.3 (Casorati-Weierstrass)**.** *Suppose that z_0 is an essential singularity of a function f that is holomorphic on a punctured neighbourhood $U \setminus \{z_0\}$. Then for every $w_0 \in \mathbb{C}$, there is a sequence z_n in $U \setminus \{z_0\}$ such that $z_n \to z_0$ and $f(z_n) \to w_0$.*

Applying the open mapping theorem, we see that f maps $U \setminus \{z_0\}$ onto an open, dense subset of the complex plane – and the same is true of every punctured neighbourhood $V \setminus \{z_0\} \subset U \setminus \{z_0\}$! It follows immediately that f is not injective on $U \setminus \{z_0\}$: the image $f(U \setminus \overline{V})$, where $z_0 \in V$ and $V \subset\subset U$, necessarily intersects the open, dense set $f(V \setminus \{z_0\})$ nontrivially.

In fact more is true: "Picard's big theorem" (whose considerably more difficult proof will be given in the last chapter) states that *the image of every punctured neighbourhood $V \setminus \{z_0\}$ under f is the entire complex plane with the possible exception of one point.*

Proof of Prop. 6.3**:** Suppose that z_0 is an isolated singularity of f and the claim is false. Then there exist a point $w_0 \in \mathbb{C}$ and positive radii r and ε such that $|f(z) - w_0| \geqslant \varepsilon$ for all z satisfying $0 < |z - z_0| < r$. The function

$$g(z) = \frac{1}{f(z) - w_0}$$

is thus holomorphic for $0 < |z - z_0| < r$ and bounded by $1/\varepsilon$. By the Riemann removable singularity theorem, g can be extended to a holomorphic function at z_0. The relationship $f(z) = w_0 + 1/g(z)$ now shows that f has a removable singularity (if $g(z_0) \neq 0$) or a pole (if $g(z_0) = 0$) at z_0, but certainly not an essential singularity. □

The proof of Prop. 6.1 yields a characterization of the order of a pole in terms of the growth of f as it approaches the pole:

Corollary 6.4. *Let z_0 be an isolated singularity of f.*

i. *The function f has a pole of order less than or equal to k or a removable singularity at z_0 if and only if*

$$|f(z)| \leqslant C_1|z - z_0|^{-k}$$

on a punctured neighbourhood of z_0, where C_1 is some positive constant.

ii. *The function f has a pole of order greater than or equal to k at z_0 if and only if*

$$|f(z)| \geqslant C_2|z - z_0|^{-k}$$

on a punctured neighbourhood of z_0, where C_2 is some positive constant.

Functions whose only isolated singularities are poles are not much more complicated than holomorphic functions.

Definition 6.3. *A meromorphic function f on a domain G is a holomorphic function $f\colon G \setminus P_f \to \mathbb{C}$, where P_f is a discrete subset of G whose points are poles of f.*

Later, we will see that such a function can be extended to a continuous function $f\colon G \to \mathbb{C} \cup \{\infty\}$ by introducing a "point at infinity", denoted ∞, and setting $f(z_0) = \infty$ for all $z_0 \in P_f$.

Holomorphic functions on G are of course meromorphic (their set of poles is empty). Quotients f/g of holomorphic functions on G (where $g \not\equiv 0$) are meromorphic: the set of zeros of g is discrete in G, and its points are isolated singularities of f/g that are either removable or poles. In particular, rational functions are meromorphic on all of $\mathbb{C}$, as are functions like $\tan z$. On the other hand, $\exp(1/z)$ is not meromorphic on $\mathbb{C}$, since the origin is an essential singularity.

A natural question is: Is every meromorphic function on a domain G a quotient of two holomorphic functions that are holomorphic *on all of G*? The answer is "yes" – see, for example, [FL1]. At this stage, however, we will only record the weaker claim that every meromorphic function is *locally* the quotient of two holomorphic functions. To be more precise:

Let f be holomorphic everywhere on G with the exception of a discrete subset P_f. Then f is meromorphic if and only if there is a neighbourhood $U \subset G$ about any point $z_0 \in G$ such that, on $U \setminus P_f$, $f = h/g$ is the quotient of two holomorphic functions $g, h\colon U \to \mathbb{C}$.

Functions of this form are certainly meromorphic. Conversely, if f is meromorphic and z_0 is not a pole of f, then one can choose $h = f$ and $g \equiv 1$. If z_0 is a pole of order k, then by (1), $f(z) = h(z)/(z - z_0)^k$ near z_0, where h is holomorphic at z_0. – It follows that if $f \not\equiv 0$, then $1/f$ is also meromorphic; the zeros of f are the poles of $1/f$, and the poles of f are the zeros of $1/f$. Moreover the above description shows that the set of meromorphic functions on a domain G is a field – i.e. sums, differences, products, and quotients of meromorphic functions are again meromorphic ($f \equiv 0$ is the zero element, which is excluded as a denominator). The ring of holomorphic functions $\mathcal{O}(G)$ is a subring of this field, and the positive answer to the above question implies that it is the field of fractions of this ring. Here it is of course essential that one consider a domain, i.e. a *connected* open set (if U is not connected, then $\mathcal{O}(U)$ has zero divisors!).

To conclude, we remark that the identity theorem applies to meromorphic functions. We leave it to the reader to justify this claim.

Exercises

1. Determine the type of singularity that each of the following functions has at z_0. If the singularity is removable, calculate the limit of $f(z)$ as $z \to z_0$; if the singularity is a pole, find its order and the principal part of f at z_0.

   a) $\dfrac{1}{1-e^z}$ at $z_0 = 0$, b) $\dfrac{1}{z - \sin z}$ at $z_0 = 0$,

   c) $\dfrac{z\,e^{iz}}{(z^2+b^2)^2}$ at $z_0 = ib$ $(b > 0)$, d) $(\sin z + \cos z - 1)^{-2}$ at $z_0 = 0$.

2. Let f be holomorphic on $D_r(z_0) \setminus \{z_0\}$ with a pole at z_0. Show that there exists an $R > 0$ such that

   $$f(D_r(z_0) \setminus \{z_0\}) \supset \mathbb{C} \setminus \overline{D_R(0)}.$$

3. Suppose z_0 is an isolated singularity of f. Show that z_0 is not a pole of e^f. (Hint: Use the result of Ex. 2.)

4. Let f be holomorphic for $0 < |z| < r_0$. For $0 < r < r_0$, put $M_r(f) = \max\{|f(z)| \colon |z| = r\}$. Prove that:

   a) The origin is a removable singularity of f if and only if $M_r(f)$ stays bounded as $r \to 0$. If this is the case, then $M_r(f)$ is a strictly increasing function of r, provided f is nonconstant, and $\lim\limits_{r\to 0} M_r(f) = |f(0)|$.

   b) The origin is a pole of f if and only if $M_r(f) \to \infty$ as $r \to 0$ and there is a number ℓ such that $r^\ell M_r(f)$ stays bounded. The order of the pole is the minimal ℓ with this property.

   c) The origin is an essential singularity of f if and only if $r^\ell M_r(f) \to \infty$ for all $\ell \geqslant 0$ as $r \to 0$.

   d) In parts b) and c), there is an $r_1 \in \,]0, r_0[$ such that $r \mapsto M_r(f)$ is strictly decreasing on $]0, r_1]$.

5. Suppose that the functions $a_1, \ldots, a_n$ are holomorphic at z_0 and that f has an essential singularity at z_0. Show that $g = f^n + a_1 f^{n-1} + \ldots + a_n$ has an essential singularity at z_0. Furthermore, show that the claim remains true if the functions a_ν are only assumed to be meromorphic.

7. Holomorphic functions of several variables

A holomorphic function of n variables $z_1, \ldots, z_n$ is holomorphic in each individual variable and can therefore be expressed via the Cauchy integral formula by applying it to each z_ν. This leads to n successive integrations with respect to the variables z_ν. More explicitly:

Let $\mathbf{z}^0 = (z_1^0, \ldots, z_n^0) \in \mathbb{C}^n$, and let

$$D = D_{\mathbf{r}}(\mathbf{z}^0) = \big\{\mathbf{z} = (z_1, \ldots, z_n) \colon |z_\nu - z_\nu^0| < r_\nu,\ \nu = 1, \ldots, n\big\}$$

be a *polydisk* about $\mathbf{z}^0$ with *polyradius* $\mathbf{r} = (r_1, \ldots, r_n)$, where $r_\nu > 0$. Thus, D is the product of n disks of radius r_ν about z_ν^0. We put

$$T = \big\{\mathbf{z} \colon |z_\nu - z_\nu^0| = r_\nu,\ \nu = 1, \ldots, n\big\};$$

this is a product of n circles T_ν, i.e. an n-dimensional torus, and it is clearly an n-dimensional surface of integration (cf. I.6). The torus T is called the *distinguished boundary* of D.

Let f be holomorphic on the open set U, and suppose that $U \supset \overline{D}$. We choose a point $\mathbf{z} \in D$ and apply the Cauchy integral formula to the first variable:

$$f(\mathbf{z}) = \frac{1}{2\pi i} \int_{T_1} \frac{f(\zeta_1, z_2, \ldots, z_n)}{\zeta_1 - z_1}\, d\zeta_1.$$

A second application – this time with respect to the second variable – gives us

$$f(\zeta_1, z_2, \ldots, z_n) = \frac{1}{2\pi i} \int_{T_2} \frac{f(\zeta_1, \zeta_2, z_3, \ldots, z_n)}{\zeta_2 - z_2}\, d\zeta_2,$$

so that

$$f(\mathbf{z}) = \left(\frac{1}{2\pi i}\right)^2 \int_{T_1} \int_{T_2} \frac{f(\zeta_1, \zeta_2, z_3, \ldots, z_n)}{(\zeta_1 - z_1)(\zeta_2 - z_2)}\, d\zeta_2\, d\zeta_1.$$

Repeating this procedure n times yields the Cauchy integral formula for polydisks.

Theorem 7.1. *Let f be holomorphic in a neighbourhood of a closed polydisk $\overline{D}$ with distinguished boundary T. Then for every $\mathbf{z} \in D$,*

$$f(\mathbf{z}) = \left(\frac{1}{2\pi i}\right)^n \int_T \frac{f(\boldsymbol{\zeta})}{(\zeta_1 - z_1) \cdots (\zeta_n - z_n)}\, d\zeta_1 \ldots d\zeta_n.$$

More generally, we define the Cauchy integral for a continuous function h on T as

$$C_h(\mathbf{z}) = \left(\frac{1}{2\pi i}\right)^n \int_T \frac{h(\boldsymbol{\zeta})}{(\zeta_1 - z_1) \cdots (\zeta_n - z_n)}\, d\zeta_1 \ldots d\zeta_n,$$

where $\mathbf{z} \in D$. Differentiating under the integral shows that $C_h(\mathbf{z})$ is holomorphic on D and arbitrarily often differentiable with respect to all variables, and it yields the formula

$$\frac{\partial^{k_1 + \ldots + k_n}}{\partial z_1^{k_1} \cdots \partial z_n^{k_n}} C_h(\mathbf{z}) = \frac{k_1! \cdots k_n!}{(2\pi i)^n} \int_T \frac{h(\boldsymbol{\zeta})}{(\zeta_1 - z_1)^{k_1+1} \cdots (\zeta_n - z_n)^{k_n+1}}\, d\zeta_1 \ldots d\zeta_n.$$

By combining this information with Thm. 7.1, we obtain, as in II.3:

Theorem 7.2. *Holomorphic functions of several variables are infinitely often complex differentiable; all of their derivatives are again holomorphic.*

One must only note that every point in $\mathbb{C}^n$ has arbitrarily small polydisk neighbourhoods.

A series of further important results that hold for holomorphic functions of one variable can be carried over unchanged to the setting of several complex variables. Their proofs can either be given using the Cauchy integral formula for polydisks, as is done above, or by tracing them back to the corresponding claims for functions of one variable. We take the second approach.

Theorem 7.3 (The identity theorem)**.** *Let f be holomorphic on a domain G and identically zero on a nonempty open subset U of G. Then $f \equiv 0$ on all of G.*

First, a simple lemma that we will also use for the proofs to follow:

Lemma 7.4. *Let $\mathbf{a}, \mathbf{b} \in \mathbb{C}^n$ and $\mathbf{b} \neq 0$; moreover, let $\lambda\colon \mathbb{C} \to \mathbb{C}^n$ be the affine line $\lambda(t) = \mathbf{a} + t\mathbf{b}$, where $t \in \mathbb{C}$, and let L denote the image of λ. If f is holomorphic on $U \subset \mathbb{C}^n$, then $f \circ \lambda$ is holomorphic on $\lambda^{-1}(U)$.*

Indeed, it is not difficult to show that

$$\frac{d}{dt}(f \circ \lambda)(t) = \sum_{\nu=1}^{n} \frac{\partial f}{\partial z_\nu}(\lambda(t))b_\nu.$$

In short, if f is holomorphic on U, then the restriction of f to every affine line L is holomorphic on $L \cap U$. This lemma generalizes the self-evident fact that a holomorphic function of n variables is holomorphic in each individual variable.

Proof of Thm. 7.3**:** a) Let G be a polydisk. By the identity theorem in one variable (Prop. 4.2), $f \equiv 0$ on $L \cap G$ for every affine line L that intersects U. But G is the union of such lines, so $f \equiv 0$ on G.

b) Let G be an arbitrary domain, and let $G_0 = \{\mathbf{z} \in G\colon\ f \equiv 0 \text{ in a neighbourhood of } \mathbf{z}\}$. Then G_0 is open. On the other hand, if $\mathbf{z}^0$ is a limit point of G_0 in G, then there exists a polydisk D about $\mathbf{z}^0$ that intersects G_0 and is contained in G. By part a), $f \equiv 0$ on D. Therefore $\mathbf{z}^0 \in G_0$, i.e. G_0 is closed in G. It follows that $G_0 = G$. □

It is not very difficult to derive the further versions of the identity theorem, e.g. if f is holomorphic on G and $Df(\mathbf{z}^0) = 0$ for *all* complex derivatives D of any order, then $f \equiv 0$. But f can vanish on a nondiscrete set without vanishing identically – consider, for instance, the function z_1 in $\mathbb{C}^2$.

Using Lemma 7.4, we can prove the following theorem just as before:

Theorem 7.5 (The maximum modulus principle)**.** *Let f be holomorphic on the domain G, and suppose that $|f|$ has a local maximum at $\mathbf{z}^0$. Then $f(z) \equiv f(\mathbf{z}^0)$.*

Proof: By the maximum modulus principle in one variable (Thm. 5.5), $f(\mathbf{z}) \equiv f(\mathbf{z}^0)$ on all affine lines L that pass through $\mathbf{z}^0$ (that is to say, on the path component of $\mathbf{z}^0$ in $L \cap G$). It follows that $f \equiv f(\mathbf{z}^0)$ on an open neighbourhood of $\mathbf{z}^0$ and therefore on all of G, by the identity theorem. □

As in the case of one complex variable, the Weierstrass convergence theorem follows from the Cauchy integral formula:

Theorem 7.6. *Let f_ν be a locally uniformly convergent sequence of holomorphic functions on a domain G whose limit is f. Then f is holomorphic, and the sequence of derivatives*

$$D^{\mathbf{k}} f_\nu = \frac{\partial^{k_1+\ldots+k_n}}{\partial z_1^{k_1} \cdots \partial z_n^{k_n}} f_\nu$$

converges locally uniformly to the corresponding derivative $D^{\mathbf{k}} f$.

Power series expansions of holomorphic functions follow from Cauchy's integral formula as in the case of one variable, but we will not pursue this in detail because power series in more than one variable are technically more difficult to handle.

The theory of holomorphic functions of several variables has thus far shown itself to be a direct generalization of complex analysis in one variable. In the case of more than one complex variable, however, the Cauchy integral formula leads to a surprise:

Let f be holomorphic in a neighbourhood of the closed polydisk $\overline{D_{\mathbf{r}}(0)} \subset \mathbb{C}^n$, where $n > 1$, with the possible exception of the origin. By the Cauchy integral formula in one variable,

$$f(z_1, \ldots, z_n) = \frac{1}{2\pi i} \int\limits_{|\zeta_n|=r_n} \frac{f(z_1, \ldots, z_{n-1}, \zeta_n)}{\zeta_n - z_n} \, d\zeta_n$$

in case not all z_ν, where $1 \leqslant \nu \leqslant n-1$, are zero. But the integral on the right hand side depends holomorphically on $z_1, \ldots, z_n$ for *all* $\mathbf{z} \in D_{\mathbf{r}}(0)$ and thus also for $z_1 = \ldots = z_n = 0$. In this way, the right hand side extends f to a holomorphic function on all of D – the origin cannot be an isolated singularity. The fact that we have singled out the origin is of course irrelevant. Therefore:

Theorem 7.7. *A function of at least two variables that is holomorphic in a punctured neighbourhood of $\mathbf{z}^0$ can be holomorphically extended to $\mathbf{z}^0$.*

Holomorphic functions of more than one variable cannot have isolated singularities. Since an isolated zero of a holomorphic function f is an isolated singularity of $1/f$, it follows that:

Proposition 7.8. *Holomorphic functions of several variables do not have isolated zeros.*

We will learn more about the zeros of holomorphic functions of one or more variables in Chapter IV. Let us mention two consequences of Thm. 7.7 here:

1. There exist domains $G \subsetneqq \hat{G}$ in $\mathbb{C}^n$, where $n > 1$, such that $\mathcal{O}(G) = \mathcal{O}(\hat{G})$, i.e. every holomorphic function on G is the restriction of a holomorphic function on $\hat{G}$. The domains $D \setminus \{0\}$ and D serve as an example.

2. Meromorphic functions cannot be defined in terms of their behaviour at isolated singularities; they must be defined as local quotients of holomorphic functions – see the discussion following Def. 6.3.

The actual theory of several complex variables begins here (cf. [FG] and [Ra]); we will say more about it later.

Exercises

1. Let $0 \leqslant r_\nu < R_\nu$ (where $1 \leqslant \nu \leqslant n$), and let $D = D_{\mathbf{R}}(0) \setminus D_{\mathbf{r}}(0)$ be the difference of the polydisks with polyradii $\mathbf{R} = (R_1, \ldots, R_n)$ and $\mathbf{r} = (r_1, \ldots, r_n)$. Assume that $n > 1$. Show that every holomorphic function on D is the restriction of a holomorphic function on $D_{\mathbf{R}}(0)$
   Hint: Use the proof of Thm. 7.7.

Chapter III.

Functions on the plane and on the sphere

By adding a "point at infinity", denoted ∞, to the complex plane, we obtain the Riemann sphere (III.1); it allows an elegant description of meromorphic and, in particular, rational functions and an interpretation of Möbius transformations as automorphisms of the sphere (III.2,4). Important theorems about functions that are holomorphic on all of $\mathbb{C}$ ("entire functions") follow from the fact that the point ∞ is an isolated singularity of these functions (III.3). Polynomials and rational functions are investigated in detail in III.2; in particular, this section contains proofs of the fundamental theorem of algebra as well as historical notes. With the logarithm function and the functions that arise from it, we conclude our "elementary" study of the elementary functions; among other things, we describe the local mapping properties of holomorphic or meromorphic functions via root functions. Partial fraction decompositions (III.6) are an essential tool in the study of meromorphic functions; in addition to a general existence theorem, this section contains the decompositions of the functions $\cot \pi z$ and $1/\sin^2 \pi z$ and their consequences. The Weierstrass product formula (III.7) for entire functions substantially generalizes the factorisation of polynomials into linear factors. We shall use it in V.1 and V.4 to define non-elementary functions.

A large part of this chapter contains classical material whose origin dates to before 1800. The Riemann sphere is, naturally, the complex projective line, and the geometry of Möbius transformations is one-dimensional projective geometry (valid over an arbitrary field). Thm. 6.1 was established by Mittag-Leffler in 1877; Euler succeeded in summing the series $\sum n^{-2k}$ in 1740; Bernoulli numbers were introduced by Jakob Bernoulli around 1700 in order to compute power sums. The product decomposition of $\sin z$ goes back to Euler (1734); Weierstrass proved his general product theorem in 1876.

1. The Riemann sphere

By adding a new point ∞ to $\mathbb{C}$, we extend the complex plane to a compact space $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ such that meromorphic functions become continuous maps into $\widehat{\mathbb{C}}$ if one assigns to them the value ∞ at their poles.

In more detail: We set $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ and call ∞ the "point at infinity". We then extend the topology on $\mathbb{C}$ to $\widehat{\mathbb{C}}$ as follows:

Definition 1.1. *A subset U of $\widehat{\mathbb{C}}$ is called open if either $U \subset \mathbb{C}$ and U is open in $\mathbb{C}$ or $\infty \in U$ and $\widehat{\mathbb{C}} \setminus U$ is compact in $\mathbb{C}$.*

Examples of open neighbourhoods of ∞ include the complements $\widehat{\mathbb{C}} \setminus \overline{D}$ of closed disks $\overline{D} \subset \mathbb{C}$. A sequence (z_n) of complex numbers therefore converges to ∞ if and only if for every radius R only finitely many elements of the sequence are contained in

W. Fischer, I. Lieb, *A Course in Complex Analysis*, DOI 10.1007/978-3-8348-8661-3_3,

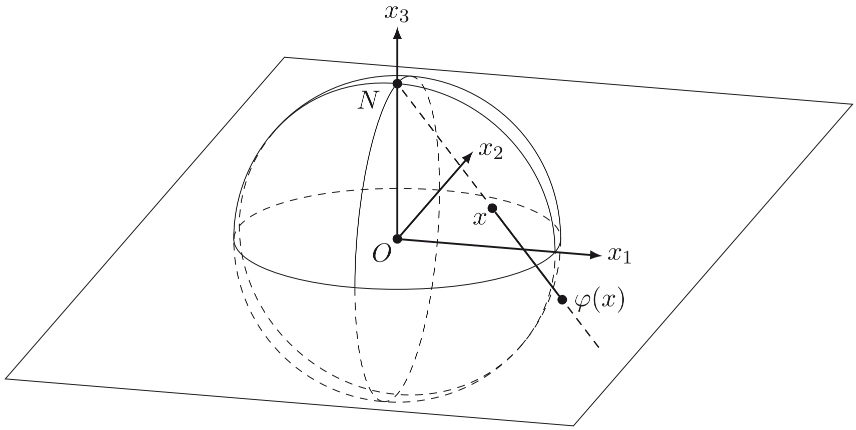


Figure 3. Stereographic projection

$D_R(0)$, i.e. if and only if $|z_n| \to \infty$ as per our former understanding. Once we have introduced open sets, the complete language of topology is at our disposal and can be applied to $\widehat{\mathbb{C}}$. Note also that we have given a precise meaning to expressions such as

$$\lim_{z\to a}(z-a)^{-1} = \infty, \quad \lim_{z\to\infty} z^2 = \infty.$$

The space $\widehat{\mathbb{C}}$ is called the *extended plane* (the one-point compactification of $\mathbb{C}$), or the *Riemann sphere.* The latter terminology is based on the following geometric model: Consider $\mathbb{R}^3$ with the coordinates x_1, x_2, and x_3, and identify $\mathbb{C}$ with the (x_1, x_2)-plane by setting $z = x_1 + ix_2$. As indicated in Fig. 3, we stereographically project the two-dimensional unit sphere

$$\mathbb{S}^2 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \colon {x_1}^2 + {x_2}^2 + {x_3}^2 = 1\}$$

from its "north pole" $N = (0, 0, 1)$ onto $\mathbb{C}$. Every point $\mathbf{x} \in \mathbb{S}^2 \setminus \{N\}$ is thus associated to the point $\varphi(\mathbf{x})$ at which the line connecting N and $\mathbf{x}$ intersects $\mathbb{C}$. This gives us a continuous bijection

$$\varphi\colon \mathbb{S}^2 \setminus N \to \mathbb{C}, \quad \varphi(x_1, x_2, x_3) = \frac{1}{1-x_3}(x_1 + ix_2),$$

and the inverse map

$$\varphi^{-1}\colon \mathbb{C} \to \mathbb{S}^2 \setminus N, \quad \varphi^{-1}(x+iy) = \frac{1}{x^2+y^2+1}(2x, 2y, x^2+y^2-1)$$

is also continuous. If we extend φ to a bijection

$$\hat\varphi\colon \mathbb{S}^2 \to \widehat{\mathbb{C}}, \quad \text{where } \hat\varphi(\mathbf{x}) = \varphi(\mathbf{x}) \quad \text{for } \mathbf{x} \neq N,\ \hat\varphi(N) = \infty,$$

then both $\hat\varphi$ and its inverse $\hat\varphi^{-1}\colon \widehat{\mathbb{C}} \to \mathbb{S}^2$ are continuous. We may thus identify $\widehat{\mathbb{C}}$ and the sphere $\mathbb{S}^2$ as topological spaces.

Proposition 1.1. *The Riemann sphere $\widehat{\mathbb{C}}$ is a compact, connected topological space.*

Proof: The space $\mathbb{S}^2 \subset \mathbb{R}^3$ has these properties. □

We may now reformulate Def. II.6.3, the definition of meromorphic functions:

A meromorphic function on a domain $G \subset \mathbb{C}$ is a continuous map $f: G \to \widehat{\mathbb{C}}$ with the following properties:

i. The set $P_f = \{z \in G\colon f(z) = \infty\}$ is discrete in G.
ii. The restriction $f: G \setminus P_f \to \mathbb{C}$ is holomorphic.

It is useful to allow ∞ to be not only a value, but an argument as well. We note that

$$\psi: \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}, \quad \psi(z) = 1/z \text{ for } z \neq 0, \infty, \quad \psi(0) = \infty, \quad \psi(\infty) = 0$$

defines a homeomorphism from $\widehat{\mathbb{C}}$ to itself; we have $\psi = \psi^{-1}$, and ψ is holomorphic on $\mathbb{C}^* = \widehat{\mathbb{C}} \setminus \{0, \infty\}$. The function ψ maps neighbourhoods of ∞ to neighbourhoods of 0 and vice versa.

Definition 1.2. *A function f defined in a neighbourhood of ∞ is called holomorphic at ∞ if the function*

$$f^* = f \circ \psi\colon \begin{cases} \zeta & \mapsto f(1/\zeta) \text{ for } \zeta \neq 0 \\ 0 & \mapsto f(\infty) \end{cases}$$

is holomorphic at the origin.

For example, the power functions $f_k(z) = z^{-k}$, where $k \in \mathbb{N}$, are holomorphic at ∞ if one sets $f_k(\infty) = 0$ – then $f_k^*(\zeta) = \zeta^k$.

Let f be a holomorphic function on $U_\varepsilon(\infty) = \{z \in \mathbb{C}\colon |z| > 1/\varepsilon\} \cup \{\infty\}$, and suppose that $f(\infty) = w_0$. Then f^* admits a power series expansion on $U_\varepsilon(0) = \psi(U_\varepsilon(\infty))$, namely

$$f^*(\zeta) = w_0 + \sum_{n=k}^{\infty} a_n \zeta^n, \quad \text{where } a_k \neq 0.$$

From this, we obtain the expansion

$$f(z) = w_0 + a_k z^{-k} + a_{k+1} z^{-k-1} + \dots,$$

which is valid on $U_\varepsilon(\infty) \setminus \{\infty\}$, and say that f takes on the value w_0 with multiplicity k at ∞.

The notion of an isolated singularity can be extended to the point ∞:

Definition 1.3. *Let f be holomorphic in a punctured neighbourhood of ∞.*

i. *The point ∞ is a removable singularity of f if f is bounded near ∞.*
ii. *The point ∞ is a pole of f if $\lim_{z\to\infty} f(z) = \infty$.*
iii. *If ∞ is neither a removable singularity nor a pole of f, then it is an essential singularity of f.*

It follows that ∞ is a singularity of type *i*, *ii*, or *iii* of f if and only if 0 is a singularity of type *i*, *ii*, or *iii*, respectively, of $f^* = f \circ \psi$. This in turn implies that the Riemann extension theorem and the Casorati-Weierstrass theorem, as well as their corollaries, apply to ∞ as an isolated singularity.

If f has a pole at ∞, then its order is naturally defined to be the order of 0 as a pole of f^*. If this order is k, we can write

$$f(z) = z^k g(z),$$

where g is a function that is holomorphic at ∞ and satisfies $g(\infty) \neq 0$. The principal part $h_\infty(z)$ of f at ∞ can then be defined as the (unique) polynomial $h_\infty(z)$ with no constant term for which $f(z) - h_\infty(z)$ is holomorphic at ∞. The degree of this polynomial coincides with the order of ∞ as a pole of f.

Examples:

i. Let $h(z) = a_{-k}(z-z_0)^{-k} + \ldots + a_{-1}(z-z_0)^{-1}$, where $a_{-k} \neq 0$. Then $\lim_{z\to\infty} h(z) = 0$, so that h is holomorphic at ∞ and vanishes there.

ii. A polynomial $p(z) = a_0 + a_1 z + \ldots + a_n z^n$, where $a_n \neq 0$, has a pole of order n at ∞. The principal part of p at ∞ is $h_\infty(z) = a_1 z + \ldots + a_n z^n$.

iii. Consider a "fractional linear" function

$$f(z) = \frac{az+b}{cz+d},$$

where $a, b, c, d \in \mathbb{C}$, $c \neq 0$, and $ad - bc \neq 0$ (otherwise f would be constant). This function is holomorphic on $\mathbb{C} \setminus \{-d/c\}$ and has a simple pole at $z_0 = -d/c$. But f is also holomorphic at ∞ if we set $f(\infty) = a/c$: The function

$$f^*(\zeta) = \frac{a/\zeta + b}{c/\zeta + d} = \frac{a + b\zeta}{c + d\zeta}$$

is holomorphic at 0, with $f^*(0) = a/c$.

iv. The function $f(z) = e^z$ has an essential singularity at ∞, because $f^*(\zeta) = e^{1/\zeta}$ has an essential singularity at 0.

Functions that are holomorphic on the entire Riemann sphere are especially simple:

Proposition 1.2. *Any function that is holomorphic on all of $\widehat{\mathbb{C}}$ is constant.*

Proof: Let $f\colon \widehat{\mathbb{C}} \to \mathbb{C}$ be holomorphic. Since $\widehat{\mathbb{C}}$ is compact, the continuous function $|f|\colon \widehat{\mathbb{C}} \to \mathbb{R}$ attains a maximum value. If this occurs at $z_0 \in \mathbb{C}$, then f is constant on $\mathbb{C}$ by the maximum modulus principle and therefore constant on $\widehat{\mathbb{C}}$ by continuity. If $|f|$ is maximal at ∞, then $|f^*|$ has a maximum at 0. It follows that f^* and therefore f are constant. □

In our definition of a meromorphic function f, we may thus allow $G \subset \widehat{\mathbb{C}}$, i.e. we may allow f to be defined at ∞. The open mapping theorem also holds in this more general situation:

Proposition 1.3. *If f is a nonconstant meromorphic function on a domain $G \subset \widehat{\mathbb{C}}$, then $f(G)$ is a domain in $\widehat{\mathbb{C}}$.*

We leave the proof to the reader as Ex. 2.

The following definition will help to simplify the terminology:

Definition 1.4. *Let G_1 and G_2 be domains in $\widehat{\mathbb{C}}$. A holomorphic mapping f of G_1 onto G_2 is a meromorphic function $f\colon G_1 \to \widehat{\mathbb{C}}$ such that $f(G_1) = G_2$.*

The fractional linear functions from Example *iii* are thus holomorphic mappings of $\widehat{\mathbb{C}}$ onto itself; we will investigate them in more detail later.

The expressions "holomorphic mapping" and "holomorphic function" are thus no longer synonymous: A holomorphic mapping $f\colon G_1 \to G_2$ is a holomorphic function if and only if $\infty \notin f(G_1)$.

Exercises

1. a) Determine the order of the zero of $\sum\limits_{n=1}^{k} b_n(z-z_0)^{-n}$ at ∞.

   b) For the following functions, determine the value $w_0 = f(\infty)$ and its multiplicity:
   $$f(z) = \frac{2z^4 - 2z^3 - z^2 - z + 1}{z^4 - z^3 - z + 1}, \quad f(z) = \frac{z^4 + iz^3 + z^2 + 1}{z^4 + iz^3 + z^2 - iz}.$$

2. Let $G \subset \widehat{\mathbb{C}}$ be a domain, and let f be a nonconstant meromorphic function on G. Show that $f(G)$ is a domain in $\widehat{\mathbb{C}}$.

2. Polynomials and rational functions

Let us first consider polynomials, i.e. functions of the form

$$p(z) = a_0 + a_1 z + \ldots + a_n z^n, \tag{1}$$

where $a_0, \ldots, a_n$ are complex constants; the degree of $p(z)$ is n, provided $a_n \neq 0$. We are interested in the existence and position of zeros of p, as well as in factorization, i.e. expressing p as a product.

We begin by studying the behaviour of polynomials for large $|z|$. Let p, as given in (1), be of degree $n \geqslant 1$. By III.1, p has a pole at ∞, i.e.

$$\lim_{z\to\infty} p(z) = \infty. \tag{2}$$

The formula

$$p(z) = a_n z^n \left(1 + \frac{a_{n-1}}{a_n} z^{-1} + \ldots + \frac{a_0}{a_n} z^{-n}\right)$$

shows that

$$\lim_{z\to\infty} \frac{p(z)}{a_n z^n} = 1, \tag{3}$$

so $a_n z^n$ is the dominant term of $p(z)$ for large $|z|$.

We will improve the banal result (3) by a more precise estimate. Set $c = \sum_{\nu=1}^{n-1} |a_\nu|$ and $q(z) = p(z) - a_n z^n = \sum_{\nu=0}^{n-1} a_\nu z^\nu$. Then for $|z| \geqslant 1$,

$$|q(z)| = \left| z^{n-1} \sum_{\nu=0}^{n-1} a_\nu z^{1-n+\nu} \right| \leqslant |z|^{n-1} \sum_{\nu=0}^{n-1} |a_\nu| |z|^{1-n+\nu} \leqslant c|z|^{n-1}$$

and

$$\begin{aligned} |a_n||z|^n \left(1 - \frac{c}{|a_n| \cdot |z|}\right) &= |a_n| \cdot |z|^n - c|z|^{n-1} \\ &\leqslant |p(z)| \leqslant |a_n||z|^n + c|z|^{n-1} \\ &= |a_n||z|^n \left(1 + \frac{c}{|a_n| \cdot |z|}\right). \end{aligned}$$

For $\varepsilon > 0$ and $R(\varepsilon) = c/\varepsilon|a_n|$, we obtain:

Proposition 2.1. *Let $p(z) = a_0 + \ldots + a_n z^n$ be a polynomial of degree $n \geqslant 1$, and let $c = \sum_{\nu=0}^{n-1} |a_\nu|$. Then for every $\varepsilon > 0$ and $|z| \geqslant \max(1, R(\varepsilon))$, we have*

$$(1-\varepsilon)|a_n||z|^n \leqslant |p(z)| \leqslant (1+\varepsilon)|a_n||z|^n.$$

If $\varepsilon < 1$, this estimate shows that p has no zeros for $|z| \geqslant \max(1, R(\varepsilon))$. Letting $\varepsilon \to 1$, we see that:

Corollary 2.2. *All zeros of p lie in the closed disk*

$$\{z \in \mathbb{C}\colon |z| \leqslant \max(1, c/|a_n|)\}.$$

The example $p(z) = z^n - 1$ shows that zeros may lie on the boundary of this disk.

We have not yet proved that an arbitrary nonconstant polynomial has zeros in $\mathbb{C}$. The "fundamental theorem of algebra" shows this to be true (see also the historical note at the end of this section):

Proposition 2.3. *Let p be a polynomial of degree $n \geqslant 1$ with complex coefficients. Then p has a zero in $\mathbb{C}$.*

First proof: By (2), we may choose an R such that $|p(z)| > |p(0)|$ for $|z| \geqslant R$; thus, $\min\limits_{|z|=R} |p(z)| > |p(0)|$. By the minimum modulus principle (Prop. II.5.6), p has a zero in $D_R(0)$. □

Second proof: Let us regard p as a meromorphic function on $\widehat{\mathbb{C}}$. By Prop. 1.3, $p(\widehat{\mathbb{C}})$ is a domain in $\widehat{\mathbb{C}}$. Since the image $p(\widehat{\mathbb{C}})$ is also closed (this follows from the compactness of $\widehat{\mathbb{C}}$), we have necessarily $p(\widehat{\mathbb{C}}) = \widehat{\mathbb{C}}$; in particular, $0 \in p(\widehat{\mathbb{C}})$. Since $p(\infty) = \infty$, we have $0 \in p(\mathbb{C})$. □

Assume that z_1 is a zero of p, and that p is of degree $n \geqslant 1$. Then

$$p(z) = (z - z_1)p_1(z),$$

where p_1 is a polynomial of degree $n - 1$. This can be seen either via polynomial division or by expanding p into powers of $z - z_1$ as

$$p(z) = p(z_1) + p'(z_1)(z - z_1) + \ldots + \frac{1}{n!}p^{(n)}(z_1)(z - z_1)^n,$$

taking into account that $p(z_1) = 0$. If $n - 1 \geqslant 1$, then p_1 has a zero z_2, so that $p_1(z) = (z - z_2)p_2(z)$ and

$$p(z) = (z - z_1)(z - z_2)p_2(z),$$

where p_2 is a polynomial of degree $n - 2$. After n steps, we have

$$p(z) = c\prod_{\nu=1}^{n}(z - z_\nu), \tag{4}$$

where the constant c is the coefficient of z^n in p. Of course, the z_ν need not be distinct. By combining repeated factors in (4) (and renumbering the zeros of p), the factorization of p can be written in the form

$$p(z) = c\prod_{\rho=1}^{r}(z - z_\rho)^{n_\rho} \tag{5}$$

with *distinct* zeros $z_1, \ldots, z_r$. The exponent n_ρ is the multiplicity (order) of the zero z_ρ; we have $\sum_{\rho=1}^{r} n_\rho = n$.

If w is an arbitrary complex number, then we can factor $p(z) - w$ as per (4):

$$p(z) - w = c\prod_{\nu=1}^{n}(z - \zeta_\nu).$$

This shows that p takes on the value w at the n (not necessarily distinct) points $\zeta_1, \ldots, \zeta_n$. The polynomial p takes on the value w at ζ with multiplicity greater than 1 if and only if $p(\zeta) = w$ and $p'(\zeta) = 0$, thus it is precisely at the at most $n-1$ zeros of p' that p assumes its value with multiplicity greater than 1.

We now turn to *rational functions* $f(z) = p(z)/q(z)$. Here p and q are polynomials, with $q \not\equiv 0$. We can (and will) assume that p and q have no common zeros. Then the zeros (of order k) of p are exactly the zeros (of order k) of f in $\mathbb{C}$; the zeros (of order k) of q give the poles (of order k) of f in $\mathbb{C}$. In order to clarify the behaviour of f at ∞, we write $p(z) = \sum_{\nu=0}^{m} a_\nu z^\nu$ and $q(z) = \sum_{\nu=0}^{n} b_\nu z^\nu$, where $a_m b_n \neq 0$, and

$$f(z) = \frac{a_m z^m + \ldots + a_0}{b_n z^n + \ldots + b_0} = z^{m-n} \frac{a_m + a_{m-1} z^{-1} + \ldots + a_0 z^{-m}}{b_n + b_{n-1} z^{-1} + \ldots + b_0 z^{-n}}.$$

Hence, if $m > n$, then f has a pole of order $m-n$ at ∞; if $m = n$, then f is holomorphic at infinity, with $f(\infty) = a_m/b_n \neq 0$; if $m < n$, then f has a zero of order $n-m$ at ∞. Moreover we see that the number of zeros of f, counting multiplicity, on the entire sphere $\widehat{\mathbb{C}}$ is $d = \max(m, n)$, and that the number of poles, again counting multiplicity, is also d.

Definition 2.1. *The degree of the rational function $f = p/q$ is*

$$d = \max(\deg p, \deg q).$$

Proposition 2.4. *A rational function $f = p/q$ of degree d takes on every value $w \in \widehat{\mathbb{C}}$ exactly d times on $\widehat{\mathbb{C}}$ (counting multiplicity).*

Proof: We may assume that $w \in \mathbb{C}$. The claim then follows from the fact that the function $f(z) - w = (p(z) - wq(z))/q(z)$ is also of degree d, and hence has d zeros. □

In integrating rational functions, partial fraction decomposition plays an important role. We prove its existence.

Proposition 2.5. *Let f be a rational function, let $z_1, \ldots, z_r$ be the distinct poles of f in $\mathbb{C}$, and let $h_1(z), \ldots, h_r(z)$ be their principal parts, respectively. If f has a pole at ∞, let $h_\infty(z)$ be its corresponding principal part; otherwise, set $h_\infty \equiv 0$. Then*

$$f(z) = \sum_{\rho=1}^{r} h_\rho(z) + c + h_\infty(z), \tag{6}$$

where c is a constant.

Proof: The function $f - \sum_{\rho=1}^{r} h_\rho - h_\infty$ is a rational function without poles in $\widehat{\mathbb{C}}$, i.e. it is holomorphic on $\widehat{\mathbb{C}}$ and therefore constant (Prop. 1.2). □

The functions $h_\rho(z)$ are of the form

$$a^{(\rho)}_{-k_\rho}(z-z_\rho)^{-k\rho}+\ldots+a^{(\rho)}_{-1}(z-z_\rho)^{-1}.$$

The individual summands are the partial fractions that give (6) its name.

Concerning h_∞ and c: the term h_∞ actually appears only when $\deg p > \deg q$. If this is the case, one can obtain $h_\infty(z)$ and c via polynomial division, which gives

$$p(z) = p^*(z)q(z) + r(z)$$

with polynomials p^* and r. Here $r \not\equiv 0$ (otherwise p and q would have common zeros or q would be constant), and the degree of r is smaller than the degree of q. Thus,

$$f(z) = \frac{r(z)}{q(z)} + p^*(z). \tag{7}$$

Since $p^*(z)$ is holomorphic on $\mathbb{C}$, f and r/q have the same principal parts h_ρ. By the preceding discussion, $r/q = \sum_{\rho=1}^{r} h_\rho$, so that $p^*(z) = c + h_\infty(z)$.

In order to determine the principal parts h_ρ, one must know the zeros z_ρ of q and their orders k_ρ. As in II.6, one can then calculate the principal part h_ρ from the beginning of the power series expansion of

$$g_\rho(z) = (z-z_\rho)^{k_\rho}\frac{r(z)}{q(z)}$$

about z_ρ. If z_ρ is a simple zero of q, then

$$h_\rho(z) = \frac{r(z_\rho)}{q'(z_\rho)}\frac{1}{(z-z_\rho)}.$$

In simple cases, ad hoc methods lead to the answer more quickly. Let us illustrate this with the function

$$f(z) = \frac{z^2-5z+6}{(z-1)^2(z+1)}.$$

We write down the partial fraction decomposition with undetermined coefficients:

$$\frac{z^2-5z+6}{(z-1)^2(z+1)} = \frac{a_2}{(z-1)^2} + \frac{a_1}{z-1} + \frac{b}{z+1},$$

then multiply both sides by the denominator $(z-1)^2(z+1)$ and obtain

$$z^2-5z+6 = a_2(z+1) + a_1(z-1)(z+1) + b(z-1)^2. \tag{8}$$

Comparing coefficients in (8) yields a system of linear equations for a_1, a_2, b. One may also substitute the zeros ± 1 of the denominator and one other value, say $z = 0$, into (8), or, alternatively, substitute the zeros ± 1 into (8) and evaluate the equation obtained by differentiating (8) at the double zero $z = 1$. One gets $a_1 = -2$, $a_2 = 1$, $b = 3$.

Rational functions are meromorphic on all of $\widehat{\mathbb{C}}$. To finish, let us prove the converse:

Proposition 2.6. *Every function that is meromorphic on all of $\widehat{\mathbb{C}}$ is rational.*

Proof: Let $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be meromorphic. Since $\widehat{\mathbb{C}}$ is compact, the discrete set of poles of f must be finite: $P_f = \{z_1, \ldots, z_r\}$ (we admit $\infty \in P_f$). Let h_ρ be the principal part of f at z_ρ, for $1 \leqslant \rho \leqslant r$. Then $f - \sum_{\rho=1}^{r} h_\rho$ is holomorphic on all of $\widehat{\mathbb{C}}$ and hence equal to a constant c, by Prop. 1.2. But then $f = \sum_{\rho=1}^{r} h_\rho + c$ is rational. $\square$

Historical note

Determining zeros of polynomials (first with "real coefficients") occupied mathematicians for centuries. The solution for quadratic polynomials is ancient and appears in the form of series of examples in Babylonian and ancient Chinese mathematics. Quadratic polynomials without real roots and cubic polynomials led to the discovery of complex numbers (Bombelli, 1560). In the sixteenth century, Italian mathematicians (Scipio del Ferro, Cardano, and Ferrari) found formulas for solving polynomial equations of degrees three and four; equations of higher degrees remained inaccessible (in the beginning of the nineteenth century, Abel and Galois proved that such equations could not in general be solved in terms involving only the usual algebraic operations and radicals). There was still the feeling, however, that the roots of such equations existed in the field of complex numbers. In 1749, Euler gave an incomplete proof of this fact; it was improved by Lagrange in 1772. Gauss was the first to give a more or less complete proof in 1799 and would go on to publish three others. These proofs used – in modern terminology – topological and complex analytical arguments. Since the development of complex analysis in the nineteenth century, many simple proofs were found. We have given two in this section, and more will follow. The word "algebra" has a different meaning today than it did in the nineteenth century, and the name "fundamental theorem of algebra" must be understood in a historical context – the theorem belongs to complex analysis.

Exercises

1. Give the partial fraction decomposition of the following functions:

   a) $\dfrac{z^5 + 2z^4 - 2z^3 - 3z^2 + z + 6}{(z^2-1)^2}$ b) $\dfrac{4z^2 + 6z + 2}{(z^2+1)(z-1)}$ c) $\dfrac{12z + 4}{(z^2+1)^2(z-1)}$.

2. a) Show that the polynomial $q(z) = z^n - \alpha_{n-1}z^{n-1} - \ldots - \alpha_0$, where $n \geqslant 1$, has exactly one positive zero provided $\alpha_\nu \geqslant 0$ (for $0 \leqslant \nu \leqslant n-1$) and at least one α_ν is strictly positive.

   b) Consider $p(z) = z^n + a_{n-1}z^{n-1} + \ldots + a_0$, where $n \geqslant 1$, $a_\nu \in \mathbb{C}$, and not all a_ν are zero. Let r be *the* positive zero of $q(z) = z^n - |a_{n-1}|z^{n-1} - \ldots - |a_0|$ – cf. part a). Show that for every zero z_0 of $p(z)$, we have $|z_0| \leqslant r$.

3. a) Let $\widetilde{q}(z) = z^n + \alpha_{n-1}z^{n-1} + \ldots + \alpha_1 z - \alpha_0$, where $\alpha_\nu \geqslant 0$ and $\alpha_0 > 0$. Show that $\widetilde{q}$ has exactly one positive zero ρ. Use 2a)!

   b) Let $p(z)$ be as in 2b), where $a_0 \neq 0$, and let ρ be the positive zero of $z^n + |a_{n-1}|z^{n-1} + \ldots + |a_1|z - |a_0|$. Show that for every zero z_0 of $p(z)$, we have $\rho \leqslant |z_0|$.

3. Entire functions

Definition 3.1. *An entire function is a function that is holomorphic on the entire complex plane. An entire function that is not a polynomial is called an entire transcendental function.*

Examples of entire transcendental functions include the exponential function, sine, and cosine, and also functions like $\exp(f(z))$, where f is any nonconstant entire function.

Considered as a function on the Riemann sphere $\widehat{\mathbb{C}}$, an entire function has an isolated singularity at ∞.

Proposition 3.1. *Let f be an entire function.*

i. If ∞ is a removable singularity of f, then f is constant.

ii. The point ∞ is a pole of order $n \geqslant 1$ of f if and only if f is a polynomial of degree n.

iii. The point ∞ is an essential singularity of f if and only if f is entire transcendental.

Proof: Let us prove *i* and *ii*. Claim *iii* then follows automatically.

For *i*: If ∞ is removable, then f can be holomorphically extended to $\widehat{\mathbb{C}}$ and is therefore constant.

For *ii*: If ∞ is a pole of order n, then the principal part h_∞ of f at ∞ is a polynomial of degree n, and $f - h_\infty$ is holomorphic on $\widehat{\mathbb{C}}$ and hence constant. The converse is clear. □

The above proposition subsumes a number of well-known theorems of complex analysis that we now state separately.

Theorem 3.2 (Liouville)**.** *Every bounded entire function is constant.*

Namely, any such function has a removable singularity at ∞.

This proposition yields a further – perhaps the best-known – proof of the fundamental theorem of algebra: If $p(z)$ is a polynomial without zeros, then $1/p(z)$ is a bounded entire function and hence constant.

Proposition 3.3. *Let f be an entire function.*

i. If for sufficiently large $|z|$, f satisfies

$$|f(z)| \leqslant C\,|z|^n,$$

where $C \geqslant 0$ and $n \in \mathbb{N}$, then f is a polynomial of degree less than or equal to n.

ii. If for sufficiently large $|z|$, f satisfies

$$|f(z)| \geqslant C\,|z|^n,$$

where $C > 0$ and $n \in \mathbb{N}$, then f is a polynomial of degree greater than or equal to n.

Proof: Applying Cor. II.6.4 to the pole 0 of $f^*(\zeta) = f(1/\zeta)$ shows that f has a pole of order less than or equal to n at ∞ in the former case, and that f has a pole of order greater than or equal to n at ∞ in the latter case. □

Proposition 3.4. *An entire function is entire transcendental if and only if for every $w \in \widehat{\mathbb{C}}$, there is a sequence $z_\nu \to \infty$ such that $f(z_\nu) \to w$.*

By the Casorati-Weierstrass theorem (Prop. II.6.3), this means precisely that f has an essential singularity at ∞.

Liouville's theorem also holds for entire functions of n variables, i.e. functions that are holomorphic on all of $\mathbb{C}^n$:

Theorem 3.2′. *Every bounded entire function on $\mathbb{C}^n$ is constant.*

As was done in II.7, the claim can be reduced to the case of one variable (Thm. 3.2).

Exercises

1. Let f and g be entire functions such that $|f| \leqslant |g|$. Show that $f = cg$ for some constant c.
2. Let f be a nonconstant entire function. Show that e^f is transcendental.
3. Given an entire function $f = \sum_{n=0}^{\infty} a_n z^n$, put $M_r = M_r(f) = \max_{|z|=r} |f(z)|$.

   a) Show that if f is transcendental, then $\lim_{r\to\infty} r^{-k} M_r = +\infty$ for all $k \in \mathbb{N}$.

   b) Investigate the limit $\lim_{r\to\infty} \frac{\log M_r}{\log r}$.
4. Consider the function $f(z) = z + e^z$. Show that for all $t \in [0, 2\pi]$, $\lim_{r\to\infty} f(r\,e^{it}) = \infty$ and that the convergence is uniform with respect to t on the sets $\{t\colon |t - \pi| \leqslant \frac{\pi}{2}\}$ and $\{t\colon |t| \leqslant \alpha\}$ for every $\alpha < \pi/2$. How does this agree with Prop. 3.4?
5. a) Let $f(z) = \exp(z^2)$ and $\alpha \in]0, \pi/4[$. Show that for $t \in [-\alpha, \alpha] \cup [\pi-\alpha, \pi+\alpha]$, $\lim_{r\to\infty} f(r\,e^{it}) = \infty$ uniformly in t, and that for $t \in \left[-\frac{\pi}{2}-\alpha, -\frac{\pi}{2}+\alpha\right] \cup \left[\frac{\pi}{2}-\alpha, \frac{\pi}{2}+\alpha\right]$, $\lim_{r\to\infty} f(r\,e^{it}) = 0$ uniformly in t.

   b) Let $p(z) = z^n + \ldots + a_0$ be a polynomial, and let $f(z) = \exp(p(z))$. Moreover let $0 < \alpha < \pi/2n$. Show that

   $$\lim_{z\to\infty} f(z) = \infty \text{ uniformly on } \{z\colon |\arg z - 2k\pi/n| \leqslant \alpha\}$$

   $$\lim_{z\to\infty} f(z) = 0 \quad \text{uniformly on } \{z\colon |\arg z - (2k+1)\pi/n| \leqslant \alpha\}$$

   for $0 \leqslant k \leqslant n-1$.

4. Möbius transformations

In this section, we will investigate rational functions of degree 1, i.e. functions of the form

$$z \mapsto \frac{az+b}{cz+d}, \tag{1}$$

where $a, b, c,$ and d are complex constants. We will always assume that $ad - bc \neq 0$ – otherwise, the function would be constant. We will most often use the letters $S, T, \ldots$ rather than $f, g, \ldots$ to denote functions of this sort and will often write Tz instead of $T(z)$.

If $c \neq 0$, then we call (1) a fractional linear transformation T; if one puts $T(-d/c) = \infty$ and $T(\infty) = a/c$, then T is a holomorphic mapping from $\widehat{\mathbb{C}}$ to itself. It is bijective, and its inverse is

$$w \mapsto \frac{dw - b}{a - cw},$$

which is again a fractional linear transformation.

If $c = 0$, we may assume that $d = 1$, so that $Tz = az + b$, with $T(\infty) = \infty$. We then call T an entire linear transformation. The transformation T is a holomorphic bijection from $\widehat{\mathbb{C}}$ to itself that maps $\mathbb{C}$ onto $\mathbb{C}$. Its inverse $w \mapsto (w - b)/a$ is also an entire linear transformation. We will call linear transformations – whether fractional or entire – *Möbius transformations.*

The composition $S \circ T = ST$ of two Möbius transformations

$$Sz = \frac{\alpha z + \beta}{\gamma z + \delta} \quad \text{and} \quad Tz = \frac{az + b}{cz + d}$$

works out to

$$(S \circ T)(z) = \frac{(\alpha a + \beta c)z + (\alpha b + \beta d)}{(\gamma a + \delta c)z + (\gamma b + \delta d)}, \tag{2}$$

which is again a Möbius transformation. If S and T are entire linear, then so is ST.

Proposition 4.1. *The set of Möbius transformations forms a group $\mathcal{M}$ of biholomorphic mappings of $\widehat{\mathbb{C}}$ onto itself. The set of entire linear transformations is the subgroup $\mathcal{M}_0 = \{T \in \mathcal{M} \colon T\infty = \infty\}$ of $\mathcal{M}$. The elements of $\mathcal{M}_0$ are biholomorphic mappings of $\mathbb{C}$ onto itself.*

The following is remarkable:

Proposition 4.2. *Möbius transformations are the only biholomorphic mappings of $\widehat{\mathbb{C}}$ onto itself, and entire linear transformations are the only biholomorphic mappings of $\mathbb{C}$ onto itself.*

Proof: a) A biholomorphic map $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a meromorphic function and hence a rational function (Prop. 2.6). Since it assumes every value exactly once, its degree is 1 and it is thus a Möbius transformation.

b) If $f \colon \mathbb{C} \to \mathbb{C}$ is biholomorphic, then ∞ cannot be an essential singularity of f; if it were, then f would not be injective – see the remark following Prop. II.6.3. It follows that f is an injective polynomial and hence of degree 1. □

Every matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}(2,\mathbb{C})$ gives rise to the Möbius transformation

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

and formula (2) shows that this assignment defines a group homomorphism from $\mathrm{GL}(2,\mathbb{C})$ onto $\mathcal{M}$. Its kernel

$$\{A \in \mathrm{GL}(2,\mathbb{C}) \colon T_A = \mathrm{id}\}$$

consists of the matrices of the form λI, where I is the identity matrix and $\lambda \in \mathbb{C}^*$. If $(\det A)^{1/2}$ denotes a square root of $\det A$, then A and $(\det A)^{-1/2}A$ yield the same Möbius transformation; the second matrix has determinant 1. This yields

Proposition 4.3. *The assignment $A \mapsto T_A$ is a surjective group homomorphism from $\mathrm{GL}(2,\mathbb{C})$ onto $\mathcal{M}$ whose kernel is $\{\lambda I \colon \lambda \in \mathbb{C}^*\}$. It also yields a group homomorphism from*

$$\mathrm{SL}(2,\mathbb{C}) = \{A \in \mathrm{GL}(2,\mathbb{C}) \colon \det A = 1\}$$

onto $\mathcal{M}$ whose kernel is $\{\pm I\}$.

We now study geometric properties of Möbius transformations. First, we note that a transformation $T \in \mathcal{M}$, $T \neq \mathrm{id}$, has at least one and at most two fixed points in $\widehat{\mathbb{C}}$, i.e. points z such that $Tz = z$:

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

leads to the equation $cz^2 + (d-a)z - b = 0$ for $z \neq \infty$, which has at most two solutions in $\mathbb{C}$. If $c = 0$, so that T is entire, then this equation has at most one solution in $\mathbb{C}$, but it is precisely in this case that ∞ is a (further) fixed point; exactly the translations $z \mapsto z + b$, where $b \neq 0$, have ∞ as their only fixed point.

Any Möbius transformation with more than two fixed points is thus equal to the identity map. It follows that a transformation $T \in \mathcal{M}$ is completely determined if one knows the images of three distinct points in $\widehat{\mathbb{C}}$ under T.

Arbitrary triples of pairwise distinct points (z_1, z_2, z_3) and (w_1, w_2, w_3) can actually be mapped to one another by a unique Möbius transformation. We may assume that $(w_1, w_2, w_3) = (0, 1, \infty)$ and immediately verify that

$$Tz = \frac{z - z_1}{z - z_3} : \frac{z_2 - z_1}{z_2 - z_3} \tag{3}$$

has the desired property. If all $z_k \neq \infty$, this is clear. If one of the z_k is ∞, one replaces the point z_k with $1/\zeta$ in (3), simplifies, and finally sets $\zeta = 0$ to again obtain a Möbius transformation. E.g. if $z_1 = \infty$, we get

$$Tz = \frac{z_2 - z_3}{z - z_3}.$$

We describe (3) by a new word:

Definition 4.1. *Let* z_1, z_2, z_3 *be three distinct points in* $\widehat{\mathbb{C}}$, *and let* $z \in \mathbb{C}$ *be arbitrary. The cross-ratio* $\mathrm{CR}(z, z_1, z_2, z_3)$ *is the value at* z *of the uniquely defined Möbius transformation that sends* (z_1, z_2, z_3) *to* $(0, 1, \infty)$. *It is given by* (3), *where* ∞ *has to be dealt with as described above.*

In summary:

Proposition 4.4. *Let* (z_1, z_2, z_3) *and* (w_1, w_2, w_3) *be two triples of distinct points in* $\widehat{\mathbb{C}}$. *Then* $z \mapsto \mathrm{CR}(z, z_1, z_2, z_3)$ *is the Möbius transformation that maps* z_1 *to* 0, z_2 *to* 1, *and* z_3 *to* ∞. *There is exactly one* $T \in \mathcal{M}$ *such that* $Tz_k = w_k$ *for* $1 \leqslant k \leqslant 3$.

The transformation T can, for example, be obtained by solving the equation

$$\mathrm{CR}(w, w_1, w_2, w_3) = \mathrm{CR}(z, z_1, z_2, z_3)$$

for w.

The cross-ratio is invariant under Möbius transformations in the following sense:

Proposition 4.5. *If* z_1, z_2, *and* z_3 *are three distinct points in* $\widehat{\mathbb{C}}$, *then for every* $T \in \mathcal{M}$, *we have*

$$\mathrm{CR}(Tz, Tz_1, Tz_2, Tz_3) = \mathrm{CR}(z, z_1, z_2, z_3).$$

Proof: The map $S(z) = \mathrm{CR}(Tz, Tz_1, Tz_2, Tz_3)$ is the composition of $z \mapsto Tz = w$ and $w \mapsto \mathrm{CR}(w, Tz_1, Tz_2, Tz_3)$ and thus a Möbius transformation. We have $Sz_1 = 0$, $Sz_2 = 1$, and $Sz_3 = \infty$, so that $Sz = \mathrm{CR}(z, z_1, z_2, z_3)$. □

In order to more conveniently formulate geometric properties of Möbius transformations, we introduce the following terminology:

Definition 4.2. *A subset* $K \subset \widehat{\mathbb{C}}$ *is called a Möbius circle if* K *is a circle in* $\mathbb{C}$ *or a union of a straight line in* $\mathbb{C}$ *with the point* ∞.

Möbius circles are precisely the images of circles on the sphere $\mathbb{S}^2 \subset R^3$ under stereographic projection $\varphi\colon \mathbb{S}^2 \to \widehat{\mathbb{C}}$, whereby circles through the north pole are mapped to straight lines (including ∞). In any case, the part of a Möbius circle that lies in $\mathbb{C}$ can be described by an equation

$$\alpha z\bar{z} + \bar{b}z + b\bar{z} + \gamma = 0, \tag{4}$$

where $\alpha, \gamma \in \mathbb{R}$, $b \in \mathbb{C}$, and $\alpha\gamma < b\bar{b}$.

Proposition 4.6. *The image of a Möbius circle under a Möbius transformation is again a Möbius circle. Given two Möbius circles* K_1 *and* K_2, *there is a* $T \in \mathcal{M}$ *such that* $T(K_1) = K_2$.

Proof: a) If T is a translation $z \mapsto z + b$ or a homothety $z \mapsto az$, the first claim is evident. Let T be the inversion $z \mapsto 1/z$. Substituting $w = 1/z$ into (4) and multiplying by $w\overline{w}$, the image of the Möbius circle described by (4) is seen to satisfy the equation

$$\alpha + bw + \overline{b}\overline{w} + \gamma w\overline{w} = 0,$$

which again describes a Möbius circle.

b) Every Möbius transformation (1) can be written as a composition of translations, dilations, rotations, and possibly an inversion. This is clear if $c = 0$. For $c \neq 0$, this follows from the formula

$$\frac{az+b}{cz+d} = \frac{bc-ad}{c^2}\left(z + \frac{d}{c}\right)^{-1} + \frac{a}{c}.$$

Part a) of the proof now implies the invariance claim.

c) A Möbius circle is uniquely determined by three distinct points that lie on it. Let z_1, z_2, and z_3 be three distinct points on K_1, let w_1, w_2, and w_3 be three distinct points on K_2, and let $T \in \mathcal{M}$ be such that $Tz_k = w_k$ for $1 \leqslant k \leqslant 3$. By what we have already proved, T maps K_1 to the Möbius circle through w_1, w_2, and w_3, which is K_2. □

Moreover, we have:

Proposition 4.7. *A point $z \in \widehat{\mathbb{C}}$ lies on the Möbius circle K determined by the points z_1, z_2, and z_3 if and only if* $\mathrm{CR}(z, z_1, z_2, z_3) \in \mathbb{R} \cup \{\infty\}$.

Proof: The transformation $T\colon z \mapsto \mathrm{CR}(z, z_1, z_2, z_3)$ maps K onto the Möbius circle that passes through 0, 1, and ∞, i.e. onto $\mathbb{R} \cup \{\infty\}$. Thus $z \in K$ if and only if $Tz \in \mathbb{R} \cup \{\infty\}$. □

A Möbius circle K divides $\widehat{\mathbb{C}} \setminus K$ into two disjoint domains G_1 and G_2: the inside and outside of K, or two open half-planes. A Möbius transformation that sends K to K' maps each of the domains determined by K biholomorphically onto one of the domains determined by K'. In particular, if we take a $T \in \mathcal{M}$ such that $T(K) = \mathbb{R} \cup \{\infty\}$, then $T(G_1) = \mathbb{H}$ and $T(G_2) = \mathbb{H}^- = \{z \in \mathbb{C}\colon \operatorname{Im} z < 0\}$, or else $T(G_2) = \mathbb{H}$ and $T(G_1) = \mathbb{H}^-$. After composing T with the map $z \mapsto -z$, if necessary, we see that:

Proposition 4.8. *Let the domain $G \subset \widehat{\mathbb{C}}$ be bounded by a Möbius circle. Then there exists a $T \in \mathcal{M}$ that maps G biholomorphically onto the upper half-plane $\mathbb{H}$.*

In order to map $\mathbb{D}$ onto $\mathbb{H}$, for example, we can use the transformation

$$S\colon z \mapsto \mathrm{CR}(z, 1, i, -1) = i\,\frac{1-z}{1+z}.$$

Here the points $1, i, -1 \in \partial\mathbb{D}$ are mapped to 0, 1, and ∞, respectively, i.e. $\partial\mathbb{D}$ is mapped onto $\mathbb{R} \cup \{\infty\}$; $S(0) = i \in \mathbb{H}$ then shows that $S(\mathbb{D}) = \mathbb{H}$.

Later, we will show that all biholomorphic mappings between disks or half-planes are Möbius transformations.

Exercises

1. a) Let $S, T \in \mathcal{M}$. Show that a point $z_1 \in \widehat{\mathbb{C}}$ is a fixed point of T if and only if Sz_1 is a fixed point of STS^{-1}.
   b) Suppose T has exactly one fixed point z_1. Show that there is an $S \in \mathcal{M}$ such that STS^{-1} is a translation. Moreover, show that for every $z \in \widehat{\mathbb{C}}$, we have $\lim T^n z = z_1$, where $T^n = T \circ \ldots \circ T$ denotes the n-fold composition of T with itself.
   c) Suppose T has exactly two fixed points z_1 and z_2. Show that there is an $S \in \mathcal{M}$ such that STS^{-1} is of the form $z \mapsto az$, where $a \in \mathbb{C}^*$, and that the pair $\{a, a^{-1}\}$ is uniquely determined by T.
   d) Show that if in part c) we have $|a| \neq 1$, then after a possible renumbering of our fixed points, we have $\lim T^n z = z_1$ for all $z \in \widehat{\mathbb{C}} \setminus \{z_2\}$. In case $|a| = 1$, show that every point in $\widehat{\mathbb{C}} \setminus \{z_1, z_2\}$ lies on a T-invariant Möbius circle.
2. a) Suppose $T \in \mathcal{M}$ has exactly one fixed point $z_1 \in \widehat{\mathbb{C}}$. Show that
   $$\{S \in \mathcal{M} \colon ST = TS\} = \{S \in \mathcal{M} \colon \ z_1 \text{ is the only fixed point of } S \text{ or } S = \mathrm{id}\}$$
   and that this is a commutative subgroup of $\mathcal{M}$.
   b) Suppose T has two distinct fixed points $z_1, z_2 \in \widehat{\mathbb{C}}$ and $T^2 \neq \mathrm{id}$. Show that
   $$\{S \in \mathcal{M} \colon ST = TS\} = \{S \in \mathcal{M} \colon Sz_1 = z_1 \text{ and } Sz_2 = z_2\}$$
   and that this is a commutative subgroup of $\mathcal{M}$.
   c) Suppose $T^2 = \mathrm{id}$ and $T \neq \mathrm{id}$. Show that T has two fixed points and $ST = TS$ if and only if S either has the same fixed points as T or permutes the fixed points of T. If the latter is the case, show that $\mathrm{CR}(z_1, w_1, z_2, w_2) = -1$, where z_1 and z_2 are the fixed points of T and w_1 and w_2 are the fixed points of S. Show that the group $\{S \in \mathcal{M} \colon ST = TS\}$ is not commutative.
3. Let f be a rational function of degree 2. Prove that there exist a transformation $T \in \mathcal{M}$ and an entire linear transformation S such that $S\,f\,T(z) = z^2$ if and only if f has a pole of order 2. Prove that there exist transformations S and T as above such that $S\,f\,T(z) = z + 1/z$ if and only if f has simple poles.

5. Logarithms, powers, and roots

Definition 5.1. *Let $z \neq 0$ be a complex number. A number $\zeta \in \mathbb{C}$ is called a logarithm of z if $e^\zeta = z$. We write $\zeta = \log z$.*

Every $z \in \mathbb{C}^*$ has infinitely many logarithms; any two logarithms of z differ by an integer multiple of $2\pi i$. We thus have

$$e^{\log z} = z$$

for *every* logarithm of z; conversely, for an *appropriate choice* of logarithm of e^z, we have

$$\log e^z = z.$$

Likewise, for an *appropriate choice* of the involved logarithms, we have the addition rule

$$\log(zw) = \log z + \log w. \tag{1}$$

If z is real and positive, then the *real* logarithm is a logarithm of z, and all other logarithms are obtained by adding multiples $2k\pi i$, where $k \in \mathbb{Z}$. Writing $z = |z|\, e^{it}$, it follows from (1) that

$$\log z = \log|z| + it, \text{ where } \log|z| \in \mathbb{R}. \tag{2}$$

The real part of a logarithm is unique; the imaginary part is unique up to integer multiples of 2π.

Definition 5.2. *For* $z \neq 0$, $\arg z = \operatorname{Im} \log z$ *is an argument of* z.

Via the logarithm, we introduce *powers with an arbitrary exponent*: For $z \neq 0$ and an arbitrary w, let

$$z^w = e^{w \log z}. \tag{3}$$

There are thus in general infinitely many values for z^w, depending on the choice of logarithm of z; they differ only by factors of the form $\exp(w \cdot 2k\pi i)$, where $k \in \mathbb{Z}$. If, in particular, $w = n \in \mathbb{Z}$, then formula (3) yields exactly one value, namely

$$z^n = e^{n \log z}; \tag{4}$$

the factors $e^{n \cdot 2\pi i k}$ are all 1. If $w = 1/n$, then (3) yields exactly n values, namely the nth roots, which can be written as

$$\sqrt[n]{z} = z^{\frac{1}{n}} = |z|^{\frac{1}{n}}\, e^{it/n}\, \zeta_j, \quad 0 \leqslant j \leqslant n-1, \tag{5}$$

where $z = |z|\, e^{it}$ and the $\zeta_j = \exp(j \cdot 2\pi i/n)$ are the nth roots of unity. The laws of exponents

$$z^{w_1+w_2} = z^{w_1} z^{w_2} \tag{6}$$

$$(z_1 z_2)^w = z_1^w z_2^w \tag{7}$$

hold in the following sense: in (6), the same choice of logarithm of z must be used in both sides, and in (7), $\log z_1 + \log z_2$ must be chosen as the value of the logarithm of the product.

With these rules, we have concluded the arithmetic of complex numbers. Here is a nice example:

$$i^i = e^{i \log i} = e^{i(\frac{\pi}{2} i + 2\pi i\, k)} = e^{-\frac{\pi}{2}}\, e^{-2\pi k}, \quad k \in \mathbb{Z}.$$

Note that all of these powers are real. Euler discovered this relation in 1746 and called it “merkwürdig” – cf. [RS].

Because of the ambiguity of the above expressions, it is not clear how one can assemble the values of, say, $\log z$ into a holomorphic function. Indeed, this is impossible on all of $\mathbb{C}^*$, but it can be done on appropriately chosen subsets of $\mathbb{C}^*$.

Definition 5.3. *A logarithm function on a domain $G \subset \mathbb{C}^*$ is a continuous function $z \mapsto \log z$ that satisfies the condition*

$$e^{\log z} = z.$$

Suppose that $\log z$ is such a function on G. By definition, it is injective. We denote its image by G'. For $z, z_0 \in G$, let

$$w = \log z, \quad w_0 = \log z_0.$$

Then $e^w = z$, $e^{w_0} = z_0$, so that

$$\frac{\log z - \log z_0}{z - z_0} = \frac{w - w_0}{e^w - e^{w_0}}.$$

As $z \to z_0$, $w \to w_0$ by the continuity of $\log z$, so that we have

$$\lim_{z \to z_0} \frac{w - w_0}{e^w - e^{w_0}} = \frac{1}{e^{w_0}} = \frac{1}{z_0}.$$

Thus, $\log z$ is holomorphic, and

$$\frac{d}{dz} \log z = \frac{1}{z}. \tag{8}$$

The image G' must be a domain on which the exponential function serves as the left inverse of the bijection $\log z$ and hence also the right inverse. We thus have:

Proposition 5.1. *If there exists a (continuous) logarithm function on a domain G, then it is in fact holomorphic, and its derivative is $1/z$. It is injective and satisfies the identities*

$$e^{\log z} \equiv z, \quad \log e^w \equiv w,$$

where $z \in G$ and $w \in G' = \log G$. The image G' is again a domain, and $\log z$ and e^w are mutually inverse biholomorphic mappings between G and G'.

Any two logarithm functions on G differ by an integer multiple of $2\pi i$. Instead of speaking of a logarithm function, we will often speak of a *branch of the logarithm*.

Proposition 5.2. *Let $G \subset \mathbb{C}^*$ be a domain. Then the following are equivalent:*

i. There exists a branch of the logarithm defined on G.

ii. There exists a $G' \subset \mathbb{C}$ that is bijectively mapped onto G under the exponential function.

iii. The function $1/z$ has a primitive on G.

iv. For every closed path of integration γ in G, we have

$$\int_\gamma \frac{dz}{z} = 0.$$

Almost all of the claims have already been proved – it only remains to show that *i* follows from *iii*. Let $l(z)$ be a primitive of $1/z$. Then

$$\frac{d}{dz}\frac{e^{l(z)}}{z} = \frac{1}{z^2}\left(\frac{z\,e^{l(z)}}{z} - e^{l(z)}\right) = 0,$$

so that $e^{l(z)} = cz$. Writing $c = e^a$, we obtain

$$e^{l(z)-a} = z,$$

i.e. $l(z) - a$ is a branch of the logarithm. □

This proposition shows that there is no branch of the logarithm defined on all of $\mathbb{C}^*$, but that there is a branch defined on every star-shaped subdomain of $\mathbb{C}^*$, e.g. on a "slit plane" $\mathbb{C} \setminus L_\varphi$, where L_φ is the ray $\{z = t\,e^{i\varphi} : 0 \leqslant t < \infty\}$.

Let us consider the different branches of $\log z$ on a subdomain $G \subset \mathbb{C}^*$. A priori, there is nothing that distinguishes one branch from the others. On the other hand, for positive real numbers, it is reasonable to choose *real* logarithms that coincide with those from elementary analysis. We thus define:

Definition 5.4. *Let $G \subset \mathbb{C}^*$ be a domain on which branches of the logarithm exist, and suppose that $G \cap \mathbb{R}_{>0}$ is connected. The principal branch of the logarithm is the branch that is real for positive real arguments.*

An especially large domain that satisfies the above condition is $\mathbb{C}^* \setminus \mathbb{R}_{<0}$, the slit plane cut along the negative real axis. Denoting the principal branch by $\operatorname{Log} z$, we have

$$\operatorname{Log} z = \log|z| + i\arg z, \quad \text{where } -\pi < \arg z < \pi. \tag{9}$$

Starting from $\operatorname{Log}'(z) = 1/z$, it is easy to find *power series expansions* of $\operatorname{Log} z$. We record only the expansion about $z_0 = 1$ in the form

$$\operatorname{Log}(1+z) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z^n \quad \text{for } |z| < 1. \tag{10}$$

The principal branch of the logarithm cannot be continuously extended to $\mathbb{C}^*$. Letting $z \in \mathbb{C}^* \setminus \mathbb{R}_{<0}$ tend to $x_0 < 0$ in the upper half-plane, we have

$$\lim \operatorname{Log}(z) = \log|x_0| + \pi i,$$

and letting z tend to x_0 in the lower half-plane, we have

$$\lim \operatorname{Log}(z) = \log|x_0| - \pi i.$$

The values of $\operatorname{Log} z$ thus jump by $2\pi i$ upon crossing the negative real axis.

A branch of the logarithm exists exactly when there exists a continuous branch of the argument function; the latter is then automatically infinitely often real differentiable. Using the functions $\log z$ and $\arg z$, we would like to interpret the integral

$$\int_\gamma \frac{dz}{z}$$

over closed paths as concretely as possible. Let $\gamma\colon [a,b] \to \mathbb{C}^*$ be a closed path.

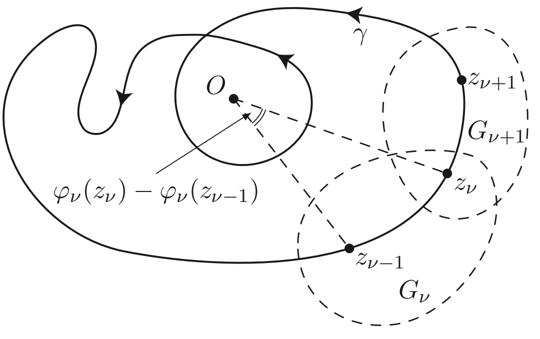


Figure 4. Interpreting the winding number

Choose a partition $a = t_0 < \ldots < t_n = b$ of the parameter interval such that every subpath $\gamma_\nu = \gamma\big|_{[t_{\nu-1},t_\nu]}$ belongs to a domain G_ν on which there exists a branch of the logarithm (the G_ν can, for instance, be chosen to be disks). We now determine branches $l_\nu(z) = \log|z| + i\varphi_\nu(z)$ on the G_ν as follows: Let l_1 be an arbitrary branch of the logarithm. For $\nu \geqslant 2$, we fix the branch l_ν via $l_\nu(\gamma(t_{\nu-1})) = l_{\nu-1}(\gamma(t_{\nu-1}))$. Then

$$\varphi\colon t \mapsto \varphi_\nu(\gamma(t)), \ \text{ where } t \in [t_{\nu-1}, t_\nu],$$

is a continuous function φ on $[a,b]$ that associates to every t an argument of $\gamma(t)$. Putting $z_\nu = \gamma(t_\nu)$, we then have

$$\int_{\gamma_\nu} \frac{dz}{z} = l_\nu(z_\nu) - l_\nu(z_{\nu-1}) = \log|z_\nu| - \log|z_{\nu-1}| + i(\varphi_\nu(z_\nu) - \varphi_\nu(z_{\nu-1})),$$

the imaginary part of which is $\varphi(t_\nu) - \varphi(t_{\nu-1})$ and therefore measures the change in the argument of $\gamma(t)$ on the subinterval $[t_{\nu-1}, t_\nu]$. Summing these integrals over all subpaths, the real parts cancel, since $z_n = z_0$, and we are left with

$$\int_\gamma \frac{dz}{z} = i(\varphi(b) - \varphi(a)).$$

Our integral thus measures the total change of the argument along the closed path γ, i.e. the number $\frac{1}{2\pi i}\int_\gamma \frac{dz}{z}$ (which is an integer!) tells us how often $\gamma(t)$ winds around the origin when t goes from a to b (here counterclockwise windings are automatically counted as positive, and clockwise windings are counted as negative).

Likewise, we of course have: If γ is a closed path of integration and $z_0 \notin \operatorname{Tr}\gamma$, then

$$\frac{1}{2\pi i}\int\limits_\gamma \frac{dz}{z-z_0}$$

is an integer that tells us how often γ winds around the point z_0. We thus define:

Definition 5.5. *Let γ be a closed path of integration in $\mathbb{C}$, and suppose $z_0 \notin \operatorname{Tr}\gamma$. Then the integer*

$$n(\gamma, z_0) = \frac{1}{2\pi i}\int\limits_\gamma \frac{dz}{z-z_0}$$

is called the winding number of γ about z_0.

Winding numbers will play a key role in the formulation the general Cauchy integral theorem in Chapter IV.

Prop. 5.2 implies:

A branch of the logarithm exists on $G \subset \mathbb{C}^*$ if and only if for every closed path of integration γ in G, the winding number $n(\gamma, 0) = 0$. This is true, for example, for all star-shaped domains in $\mathbb{C}^*$.

As with the logarithm, we will now introduce branches of root and power functions; to do so, it is easiest to exploit our considerations concerning the logarithm. In what follows, let G be a subdomain of $\mathbb{C}^*$ on which a branch of the logarithm $\log z$ exists.

Definition 5.6. *For a fixed $w \in \mathbb{C}$ and $z \in G$,*

$$z^w = e^{w \log z}$$

is a branch of the power function $z \mapsto z^w$.

Thus, $z \mapsto z^w$ is holomorphic on G, and

$$\frac{d}{dz} z^w = w z^{w-1}, \tag{11}$$

where the same branch of the logarithm is to be used on the left and right hand sides. In general, there may be infinitely many branches of z^w; if w is an integer, then there is exactly one, namely the functions z^n, $n \in \mathbb{Z}$, that we have already studied. For $n \geqslant 0$, they can be holomorphically extended to 0, and for $n < 0$, they are defined on all of $\mathbb{C}^*$ (these functions can of course be defined without recourse to logarithms). For $w = 1/n$, where $n > 1$, we obtain exactly n branches of the nth root function, which arise from one another via multiplication by an nth root of unity.

Finally, the general exponential functions

$$z \mapsto a^z, \quad a \neq 0 \text{ fixed,}$$

are defined on all of $\mathbb{C}$. They differ by factors of the form $e^{2\pi i\, kz}$.

Let us give the particularly important Taylor series expansion of the function $f(z) = (1+z)^\alpha$. This power is certainly holomorphic in z for $|z| < 1$. If one requires $f(0) = 1$, i.e. if one works with the principal branch of the logarithm, then by comparing coefficients, the relation

$$(1+z)f'(z) = \alpha f(z)$$

yields the *binomial series*

$$(1+z)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} z^n, \quad \text{for } |z| < 1, \tag{12}$$
$$\binom{\alpha}{n} = \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}, \quad \binom{\alpha}{0} = 1.$$

Next, we will investigate the question whether a nonconstant holomorphic function $f \in \mathcal{O}(G)$ has a holomorphic logarithm, i.e. whether there exists a function $g \in \mathcal{O}(G)$ such that $f = e^g$. It is clearly necessary that f have no zeros, i.e. that $f(G) \subset \mathbb{C}^*$. If a branch log of the logarithm exists on the domain $f(G)$, then one can simply set $g = \log \circ f$. Otherwise, the following is helpful:

Proposition 5.3. *Let G be a star-shaped domain in $\mathbb{C}$, and let f be holomorphic and nonzero on G. Then f has a holomorphic logarithm on G.*

Proof: The function f'/f is holomorphic on G, and since G is star-shaped, it has a primitive h on G. We have

$$(e^{-h}f)' = e^{-h}(-h'f + f') = 0,$$

so that $f = c_1 e^h = e^{h+c}$ with constants c_1 and c, and $g = h + c$ is the desired logarithm. □

The constant c in the proof is only determined up to addition of integer multiples of $2\pi i$. The notation $g = \log f$ thus refers to a unique function only upon specifying the value $\log f(z)$ at some point z.

The existence of a holomorphic logarithm of f implies the existence of *holomorphic nth roots* of f: If $f(z) = e^{g(z)}$ on G, then $h(z) = e^{g(z)/n}$ is holomorphic on G and satisfies $h^n = f$.

Remarks:

a) Each of the following conditions implies the existence of $\log f$:

i. The function f is holomorphic and nonzero on a star-shaped domain. The image of f may then be arbitrary.

ii. The image $f(G)$ is a domain on which a holomorphic logarithm of f exists; then G need not be star-shaped.

b) An nth root of f, i.e. a holomorphic function g such that $g^n = f$, can also exist if f does not have a holomorphic logarithm – see the exercises.

Using holomorphic root functions, we can now understand the local mapping properties of arbitrary holomorphic functions. For $k \in \mathbb{N}$, consider the function

$$z \mapsto z^k = w \tag{13}$$

in a disk about the origin. This function clearly maps the disk $D_r(0)$ surjectively onto the disk $D_{r^k}(0)$, whereby every point $w \in D_{r^k}(0)$ with the exception of 0 has exactly k preimages, namely the kth roots of w; the only preimage of the point 0 is 0. Moreover the map f is open and finite, i.e. the fibres $\{z\colon z^k = w\}$ are finite. In topology, maps of this kind are called (k-fold) *branched coverings.*

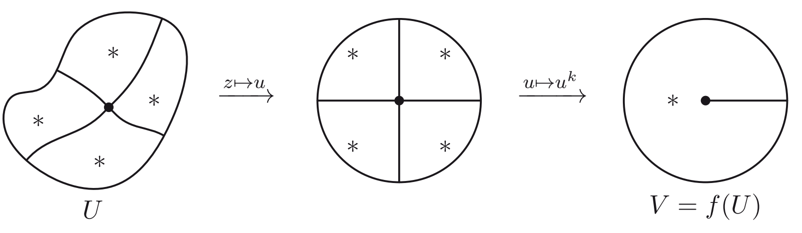


Figure 5. A branch point of order 4

Let $f\colon G \to \mathbb{C}$ be an arbitrary nonconstant holomorphic function. Upon composing with translations, we may assume that $f(0) = 0$. If k is the order of f at 0, then

$$f(z) = z^k g(z), \quad g(0) \neq 0 \tag{14}$$

for some holomorphic function g. It follows that a kth root of g, i.e. a holomorphic function h such that

$$h(z)^k \equiv g(z), \tag{15}$$

exists in a sufficiently small neighbourhood of 0. The map f may thus be described as follows:

$$z \mapsto zh(z) \stackrel{\text{def}}{=} u, \tag{16}$$

$$u \mapsto u^k. \tag{17}$$

The map (16) is biholomorphic in a neighbourhood of 0, and the map (17) is a k-fold branched covering – meaning that its fibres consist of k points with the exception of the branch point 0. We thus have:

Proposition 5.4. *Let f be a nonconstant holomorphic function, and suppose f takes on the value w_0 at z_0 with multiplicity k. Then there exist neighbourhoods U and V of z_0 and w_0, respectively, such that $f\colon U \to V$ is a k-fold branched covering with branch point z_0. In particular, $f(U) = V$, and f is open.*

We have thus found a new proof of the fact that holomorphic functions are open mappings.

Exercises

1. Let f be the branch of the logarithm on $\mathbb{C} \setminus \mathbb{R}_{\geqslant 0}$ that takes the value $-i\pi/2$ at $-i$. Determine

$$f(i), \quad f(-e), \quad f(-1-i\sqrt{3}), \quad f\big((-1-i\sqrt{3})^2\big).$$

2. a) Find a maximal domain on which holomorphic functions $\log(1-z)^2$ and $\sqrt{z+\sqrt{z}}$, respectively, can be defined.

   b) Show that a logarithm of the tangent function exists on $G = \mathbb{C} \setminus \bigcup_{k\in\mathbb{Z}} [k\pi - \frac{\pi}{2}, k\pi]$.

3. Show that there exists an entire function g such that $g(z)^2 = 1-\cos z$. (Hint: Begin with the strip $S_0 = \{z\colon -2\pi < \operatorname{Re} z < 2\pi\}$.) Does $1-\cos z$ have a holomorphic logarithm on $\mathbb{C}\setminus\{2k\pi\colon k \in \mathbb{Z}\}$?

4. The tangent function maps the strip $S_0 = \{-\pi/2 < \operatorname{Re} z < \pi/2\}$ biholomorphically onto $G_0 = \mathbb{C} \setminus \{ti\colon t \in \mathbb{R},\ |t| \geqslant 1\}$. Express the inverse function $\arctan\colon G_0 \to S_0$ using the principal branch of the logarithm and show that $\arctan w = \int_{\gamma_w} \frac{d\zeta}{1+\zeta^2}$, where γ_w is a path from 0 to w in G_0.

5. The sine function maps the strip $S_0 = \{-\pi/2 < \operatorname{Re} z < \pi/2\}$ biholomorphically onto $G_1 = \mathbb{C} \setminus \{t \in \mathbb{R}\colon |t| \geqslant 1\}$. Express the inverse function $\arcsin\colon G_1 \to S_0$ using logarithms and roots (specify the branches to be used!), and show that

$$\arcsin w = \int\limits_{\gamma_w} \frac{d\zeta}{\sqrt{1-\zeta^2}},$$

   where γ_w is a path from 0 to w in G_1 and we choose the branch of the square root function that takes the value 1 at $w = 0$.

6. Partial fraction decompositions

In this section, we want to, on the one hand, construct meromorphic functions that have poles at a given set of points. On the other hand, we will establish a series representation (partial fraction decomposition) for meromorphic functions in terms of their principal parts. To do so, we must extend the definition of compact convergence to series of meromorphic functions.

Definition 6.1. *Let $G \subset \mathbb{C}$ be a domain, and let f_i, $i \in I$, be a countable collection of meromorphic functions on G. The sum $\sum_{i \in I} f_i$ converges (absolutely) compactly on G if for every compact subset $K \subset G$, there exists a finite subset J of the index set I such that the f_i, with $i \in I \setminus J$, have no poles in K and the sum $\sum_{i \in I \setminus J} f_i$ converges (absolutely) uniformly on K.*

In this situation, the set P of poles of the functions f_i is discrete in G, and $f = \sum_{i \in I} f_i$ is a meromorphic function on G whose set of poles must be contained in P. Compact convergence is again equivalent to locally uniform convergence.

Consider a meromorphic function f on G with the discrete (and hence at most countable) set of poles $P = \{a_1, a_2, a_3, \dots\}$; assume the principal part of f at a_ν is

$$h_\nu(z) = \frac{c_1^{(\nu)}}{z - a_\nu} + \dots + \frac{c_{k(\nu)}^{(\nu)}}{(z - a_\nu)^{k(\nu)}}. \tag{1}$$

If $P = \{a_1, \dots, a_n\}$ is finite, then

$$f = g + \sum_{\nu=1}^{n} h_\nu$$

for some function $g \in \mathcal{O}(G)$, since the function $f - \sum_{\nu=1}^{n} h_\nu$ only has removable singularities. If P is infinite, then $\sum_{\nu=1}^{\infty} h_\nu$ will only converge in lucky cases; if it does, then one of course has

$$f = g + \sum_{\nu=1}^{\infty} h_\nu,$$

where $g \in \mathcal{O}(G)$. In general, one must try to find "convergence-generating summands" $p_\nu \in \mathcal{O}(G)$ such that the series

$$\sum_{\nu=1}^{\infty} (h_\nu - p_\nu)$$

converges compactly on G. It then yields a meromorphic function with the same poles and principal parts as f. We thus have

$$f = g + \sum_{\nu=1}^{\infty} (h_\nu - p_\nu), \tag{2}$$

where $g \in \mathcal{O}(G)$; we call this representation a *partial fraction decomposition* of f.

The reverse question is: Suppose we are given a discrete subset $P = \{a_1, a_2, a_3, \dots\}$ of G, as well as a principal part $h_\nu(z)$ for every $a_\nu \in P$ (i.e. $h_\nu(z)$ is a function of the form (1)). Does there exist a meromorphic function on G that has poles precisely at the points a_ν and whose principal part at each a_ν is h_ν?

If one can find $p_\nu \in \mathcal{O}(G)$ such that $\sum_{\nu=1}^{\infty}(h_\nu - p_\nu)$ converges compactly, then this series clearly solves our problem. The construction of such convergence-generating summands is possible for arbitrary domains G. We will only treat the simplest case $G = \mathbb{C}$, and prove:

Theorem 6.1 (Mittag-Leffler)**.**

i. Let $P = \{a_1, a_2, a_3, \ldots\}$ be an infinite discrete set in $\mathbb{C}$, and let $h_\nu(z)$ be a principal part at a_ν for every $a_\nu \in P$. Then there exist polynomials $p_\nu(z)$ such that the series

$$\sum_{\nu=1}^{\infty}(h_\nu - p_\nu) \tag{3}$$

converges absolutely and locally uniformly on $\mathbb{C}$. It yields a meromorphic function that has poles precisely at the points a_ν and whose principal part at each a_ν is h_ν.

ii. Let f be a meromorphic function on $\mathbb{C}$ with infinitely many poles a_ν, and let h_ν be the principal part at a_ν. Then there exists a partial fraction decomposition

$$f = g + \sum_{\nu=1}^{\infty}(h_\nu - p_\nu),$$

where g is an entire function and the p_ν are polynomials.

Proof: It suffices to find polynomials p_ν such that (3) converges. Let us assume that $0 \notin P$ and that the a_ν are ordered by increasing absolute values:

$$0 < |a_1| \leqslant |a_2| \leqslant |a_3| \leqslant \ldots .$$

We then choose positive numbers ε_ν such that $\sum_{\nu=1}^{\infty} \varepsilon_\nu < \infty$. The principal part $h_\nu(z)$ is holomorphic on $\{z\colon |z| < |a_\nu|\}$, and its Taylor series (about 0) converges compactly to h_ν. We may thus choose the polynomial $p_\nu(z)$ as a partial sum of this Taylor series such that

$$|h_\nu(z) - p_\nu(z)| \leqslant \varepsilon_\nu \text{ on } \overline{D_\nu} = \{z\colon |z| \leqslant |a_\nu|/2\}.$$

We prove convergence: Given an arbitrary radius $R > 0$, choose ν_0 such that $2R \leqslant |a_\nu|$ for $\nu \geqslant \nu_0$. Then $\overline{D_R(0)} \subset \overline{D_\nu}$ for $\nu \geqslant \nu_0$, so that for $z \in \overline{D_R(0)}$ we have

$$\sum_{\nu=\nu_0}^{\infty} |h_\nu(z) - p_\nu(z)| \leqslant \sum_{\nu=\nu_0}^{\infty} \varepsilon_\nu < \infty.$$

This shows the absolute uniform convergence of $\sum_{\nu=\nu_0}^{\infty}(h_\nu - p_\nu)$ on $\overline{D_R(0)}$. By Def. 6.1, $\sum_{\nu=1}^{\infty}(h_\nu - p_\nu)$ is then absolutely and compactly convergent on $\mathbb{C}$. In case $0 \in P$, we add h_0 to obtain

$$h_0 + \sum_{\nu=1}^{\infty}(h_\nu - p_\nu). \qquad \square$$

The p_ν are of course not unique; in dealing with a concrete case, one would choose polynomials whose degrees are as small as possible.

Let us look at two examples; afterwards, we will identify the results as well-known elementary functions.

Examples:

i. We take $P = \mathbb{Z}$; let $h_n(z) = (z-n)^{-2}$ be the given principal part at $n \in \mathbb{Z}$. For $|z| \leqslant R$ and $n \geqslant 2R$, we have $|z-n|^2 \geqslant (|n|-R)^2 \geqslant |n|^2/4$, and since $\sum_{n=1}^{\infty} n^{-2} < \infty$, the series $\sum_{|n|\geqslant 2R}(z-n)^{-2}$ is absolutely and uniformly convergent for $|z| \leqslant R$. Thus,

$$f_1(z) = \sum_{n\in\mathbb{Z}} \frac{1}{(z-n)^2} \tag{4}$$

is a meromorphic function on $\mathbb{C}$ with poles at the points $n \in \mathbb{Z}$ and principal parts $(z-n)^{-2}$.

ii. As before, take $P = \mathbb{Z}$, but suppose the principal parts at $n \in \mathbb{Z}$ are given by $h_n(z) = (z-n)^{-1}$; in particular, $h_0(z) = 1/z$. The series $\sum_{n\in\mathbb{Z}}(z-n)^{-1}$ is divergent, so we need corrective terms. The constant term $p_n(z) = -1/n$ of the Taylor expansion of $h_n(z)$ is already sufficient: For $|z| \leqslant R$ and $|n| \geqslant 2R$, we have

$$\left|\frac{1}{z-n} + \frac{1}{n}\right| = \frac{|z|}{|n|\cdot|n-z|} \leqslant \frac{2R}{|n|^2}.$$

Thus, the series

$$f_2(z) = \frac{1}{z} + \sum_{n\in\mathbb{Z}\setminus\{0\}} \left(\frac{1}{z-n} + \frac{1}{n}\right) \tag{5}$$

is absolutely and locally uniformly convergent.

We now look for a partial fraction decomposition of the function

$$f_0(z) = \frac{\pi^2}{\sin^2 \pi z}.$$

It is meromorphic on $\mathbb{C}$ with poles of order two at the points $n \in \mathbb{Z}$. The principal part of f_0 at the origin is easily computed to be $1/z^2$. The principal part at n is $(z-n)^{-2}$ due to the periodicity $f_0(z+1) = f_0(z)$, and the same holds true of the function f_1 from Example *i*. We thus have

$$\frac{\pi^2}{\sin^2 \pi z} = g(z) + \sum_{n\in\mathbb{Z}} \frac{1}{(z-n)^2}$$

for some entire function g. We claim that $g \equiv 0$:

Proposition 6.2.

$$\frac{\pi^2}{\sin^2 \pi z} = \sum_{n\in\mathbb{Z}} \frac{1}{(z-n)^2}.$$

Proof: We will employ a trick: Both f_0 and f_1 satisfy (for $2z \notin \mathbb{Z}$) the "duplication formula"

$$4h(2z) = h(z) + h\Big(z + \frac{1}{2}\Big). \tag{6}$$

Namely, we have

$$\frac{4}{\sin^2(2\pi z)} = \frac{1}{\sin^2 \pi z} + \frac{1}{\cos^2 \pi z} = \frac{1}{\sin^2 \pi z} + \frac{1}{\sin^2 \pi(z+\frac{1}{2})}$$

and

$$\begin{aligned} f_1(2z) = \sum_{n\in\mathbb{Z}} \frac{1}{(2z-n)^2} &= \sum_{m\in\mathbb{Z}} \frac{1}{(2z-2m)^2} + \sum_{m\in\mathbb{Z}} \frac{1}{(2z-2m+1)^2} \\ &= \frac{1}{4} f_1(z) + \frac{1}{4} f_1\left(z + \frac{1}{2}\right). \end{aligned}$$

The holomorphic function $g = f_0 - f_1$ thus satisfies the relation (6) on all of $\mathbb{C}$. Now let $R > 0$, $A = \{z \colon 0 \leqslant \operatorname{Re} z \leqslant 1, |\operatorname{Im} z| \leqslant R\}$, and $M = \max\{|g(z)| \colon z \in A\}$. Choose $z_0 \in A$ such that $|g(z_0)| = M$. Since $z_0/2$ and $(z_0+1)/2$ also belong to A, (6) yields

$$4M = |4g(z_0)| = |g(z_0/2) + g((z_0+1)/2)| \leqslant 2M,$$

so that $M = 0$. It follows that $g \equiv 0$ on A and hence on all of $\mathbb{C}$ by the identity theorem. □

By integrating, we obtain a partial fraction decomposition of the cotangent function from Prop. 6.2. Indeed, $f_3(z) = -\pi \cot \pi z$ is a primitive of $f_0(z) = \pi^2(\sin \pi z)^{-2}$, and, on the other hand,

$$-f_2(z) = -\frac{1}{z} - \sum_{n\in\mathbb{Z}} \left(\frac{1}{z-n} + \frac{1}{n}\right)$$

is clearly a primitive of f_1. Since $f_0 = f_1$, the functions $-f_2$ and f_3 differ only by a constant, and since both are odd functions of z, this constant is 0. We thus have

Proposition 6.3.

$$\pi \cot \pi z = \frac{1}{z} + \sum_{n\in\mathbb{Z}} \left(\frac{1}{z-n} + \frac{1}{n}\right) = \frac{1}{z} + \sum_{n=1}^{\infty} \frac{2z}{z^2 - n^2}.$$

The second sum arises from the first upon combining the terms corresponding to n and $-n$.

With the help of basic trigonometric identities like

$$\begin{aligned}
\cos \pi z &= \sin \pi\Big(z + \frac{1}{2}\Big) \\
\tan \pi z &= -\cot \pi\Big(z + \frac{1}{2}\Big) \\
\cot \pi z + \tan \pi z &= \frac{2}{\sin 2\pi z},
\end{aligned}$$

the previous propositions yield further partial fraction decompositions (the proofs are left to the reader):

Corollary 6.4. *Putting $a_n = n - 1/2$, we have*

$$\begin{aligned}
\frac{\pi^2}{\cos^2 \pi z} &= \sum_{n\in\mathbb{Z}} \frac{1}{(z-a_n)^2} \\
\pi \tan \pi z &= -\sum_{n\in\mathbb{Z}} \left(\frac{1}{z-a_n} + \frac{1}{a_n}\right) \\
\frac{\pi}{\sin \pi z} &= \frac{1}{z} + \sum_{n=1}^{\infty} (-1)^n \frac{2z}{z^2 - n^2}.
\end{aligned}$$

To conclude, we will use Prop. 6.3 to evaluate the sums

$$\zeta(2\mu) = \sum_{\nu=1}^{\infty} \frac{1}{\nu^{2\mu}}$$

for $\mu \in \mathbb{N}$. The remarkable result was found by Euler. The idea is to determine the Taylor expansion of the function $\pi z \cot \pi z$, which is holomorphic at 0, in two different ways: first by using its partial fraction decomposition, and then by expressing the cotangent function in terms of the exponential function.

By Prop. 6.3, we have

$$\pi z \cot \pi z = 1 + 2\sum_{\nu=1}^{\infty} \frac{z^2}{z^2 - \nu^2}. \tag{7}$$

We expand $z^2/(z^2 - \nu^2)$ as a geometric series

$$\frac{z^2}{z^2-\nu^2} = -\sum_{\mu=1}^{\infty} \frac{z^{2\mu}}{\nu^{2\mu}},$$

substitute this into (7), switch the order of summation, and obtain

$$\pi z \cot \pi z = 1 - 2\sum_{\mu=1}^{\infty} \left(\sum_{\nu=1}^{\infty} \frac{1}{\nu^{2\mu}}\right) z^{2\mu}. \tag{8}$$

On the other hand, we have

$$z \cot z = iz \frac{e^{2iz}+1}{e^{2iz}-1} = iz + \frac{2iz}{e^{2iz}-1} = iz + f(2iz), \tag{9}$$

where we have set

$$f(z) = \frac{z}{e^z - 1}.$$

By putting $f(0) = 1$, we make f holomorphic at 0, and we see from (9) that $f(z)+z/2$ is an even function, hence its Taylor series contains only even powers of z. We write it in the form

$$f(z) = 1 - \frac{z}{2} + \sum_{\mu=1}^{\infty} \frac{B_{2\mu}}{(2\mu)!} z^{2\mu}. \tag{10}$$

Equation (10) defines the *Bernoulli numbers* $B_2, B_4, B_6 \ldots$ One can extend this definition by putting $B_0 = 1$, $B_1 = -1/2$, and $B_{2\mu+1} = 0$ for $\mu \geqslant 1$.

The relationship $f(z)(e^z - 1) = z$ allows to derive a recursion formula for the numbers $B_{2\mu}$ (Ex. 2), which shows, in particular that the $B_{2\mu}$ are rational numbers. The first few Bernoulli numbers are:

$$B_2 = \frac{1}{6}, \quad B_4 = -\frac{1}{30}, \quad B_6 = \frac{1}{42}, \quad B_8 = -\frac{1}{30}.$$

Substituting the series (10) into (9) and replacing z with πz, one obtains

$$\pi z \cot \pi z = 1 + \sum_{\mu=1}^{\infty} (-1)^{\mu} \frac{2^{2\mu} B_{2\mu}}{(2\mu)!} \pi^{2\mu} z^{2\mu}. \tag{11}$$

Comparing coefficients in (8) and (11) then yields the following beautiful formula:

Proposition 6.5.

$$\sum_{\nu=1}^{\infty} \frac{1}{\nu^{2\mu}} = (-1)^{\mu-1} 2^{2\mu-1} \frac{B_{2\mu}}{(2\mu)!} \pi^{2\mu}.$$

For example:

$$\sum_{\nu=1}^{\infty} \frac{1}{\nu^2} = \frac{\pi^2}{6}, \quad \sum_{\nu=1}^{\infty} \frac{1}{\nu^4} = \frac{\pi^4}{90}, \quad \sum_{\nu=1}^{\infty} \frac{1}{\nu^6} = \frac{\pi^6}{945}, \quad \sum_{\nu=1}^{\infty} \frac{1}{\nu^8} = \frac{\pi^8}{9450}.$$

Let us record a few corollaries:

i. $\sum_{\nu=1}^{\infty} \frac{1}{\nu^{2\mu}}$ *is a rational multiple of* $\pi^{2\mu}$.

ii. The Bernoulli numbers $B_{2\mu}$ *have alternating signs.*

iii. $\lim\limits_{\mu\to\infty} |B_{2\mu}| = \infty$, *since* $\sum_{\nu=1}^{\infty} \frac{1}{\nu^{2\mu}} \geqslant 1$ *and* $\frac{a^n}{n!} \to 0$ *for* $a > 0$.

Exercises

1. Show that $\lim_{\mu\to\infty} |B_{2\mu}| \frac{(2\pi)^{2\mu}}{(2\mu)!} = 2$.

2. For $k \geqslant 1$, prove the recursion formulas

$$\sum_{n=0}^{k} \binom{k+1}{n} B_n = 0, \quad \sum_{\mu=0}^{k} \binom{2k+1}{2\mu} B_{2\mu} = \frac{1}{2}(2k+1).$$

3. Show that

$$\sum_{\nu=1}^{\infty} (-1)^{\nu-1} \frac{1}{\nu^{2\mu}} = (1 - 2^{1-2\mu}) \sum_{\nu=1}^{\infty} \frac{1}{\nu^{2\mu}},$$

$$\sum_{\nu=1}^{\infty} \frac{1}{(2\nu-1)^{2\mu}} = (1 - 2^{-2\mu}) \sum_{\nu=1}^{\infty} \frac{1}{\nu^{2\mu}}.$$

4. a) Derive the power series expansion of $\frac{z}{\sin z}$ about 0 using Cor. 6.4.iii.

   b) Using this result and (8), write down the power series of $\tan z$ about 0 (express the coefficients in terms of the Bernoulli numbers).

5. Use Cor. 6.4.iii to find a partial fraction decomposition of $\pi/\cos \pi z$. In analogy to the proof of Prop. 6.5, express the sums

$$\sum_{\nu=1}^{\infty} \frac{(-1)^{\nu-1}}{(2\nu-1)^{2\mu+1}}, \quad \mu = 1, 2, 3, \ldots,$$

   in terms of the Euler numbers (cf. Ex. 3 of II.4).

7. Product Expansions

If a function $f(z)$, holomorphic on a domain $G \subset \mathbb{C}$, has only finitely many zeros $a_1, \ldots, a_r$ with multiplicities n_ϱ respectively, it can be written as

$$f(z) = g(z) \prod_{1}^{r} (z - a_\varrho)^{n_\varrho},$$

$g(z)$ being holomorphic without zeros on G. We ask whether there is a similar product representation exhibiting the zeros of f in case there are infinitely many. The answer is affirmative for any G. We prove it only in the case $G = \mathbb{C}$, the proof in the general case is much more intricate.

We begin by defining infinite products of numbers and of functions, taking care of the special role of 0 with respect to multiplication.

First, let $\left(a_\nu\right)_{\nu\geqslant 1}$ be a sequence of *non-zero* complex numbers. We say that the infinite product $\prod_{1}^{\infty} a_\nu$ converges to a and write $\prod_{1}^{\infty} a_\nu = a$, if the sequence of partial products $p_n = \prod_{1}^{n} a_\nu$ converges to a and $a \neq 0$. Since $a_n = p_n/p_{n-1}$, we have the obvious necessary condition $a_n \to 1$.

Proposition 7.1. *The product* $\prod_1^\infty a_\nu$ *converges if and only if the series* $\sum_1^\infty \log a_\nu$ *converges (for a suitable choice of the logarithms).*

Note that convergence of $\sum_1^\infty \log a_\nu$ implies $a_\nu \to 1$ and that $\log a_\nu$ is the principal value for large ν.

Proof: Assume $\sum_1^\infty \log a_\nu$ converges. Then

$$p_n = \prod_1^n a_\nu = \exp\left(\sum_1^n \log a_\nu\right) \to \exp\left(\sum_1^\infty \log a_\nu\right)$$

by the continuity of the exponential function. Conversely, if $\prod_1^\infty a_\nu = a \neq 0$, we can choose a neighbourhood $U(a)$ of a on which a branch of log exists. We may assume that $\log U(a)$ is contained in a horizontal strip of width $< 2\pi$. Then, for n sufficiently large, $n \geqslant n_0$ say, $p_n \in U(a)$ and $\log p_{n+1} = \log p_n + \log a_{n+1}$ with $\log a_{n+1}$ being the principal value. Thus

$$\log p_{n+1} = \log p_{n_0} + \sum_{n_0+1}^{n+1} \log a_\nu,$$

and the convergence of $\sum_{n_0+1}^\infty \log a_\nu$ follows. □

In view of the condition $a_\nu \to 1$ we often write $a_\nu = 1 + u_\nu$. The following criterion is very useful:

Proposition 7.2. *The series* $\sum_1^\infty \log(1 + u_\nu)$ *converges absolutely if and only if* $\sum_1^\infty u_\nu$ *is absolutely convergent.*

Proof: Easy estimates of the power series expansion of $\log(1 + u)$ yield

$$\frac{2}{3}|u| \leqslant |\log(1+u)| \leqslant \frac{4}{3}|u| \quad \text{for } |u| \leqslant \frac{1}{4}.$$

The proposition follows by the majorant criterion. □

Absolute convergence of $\sum \log(1 + u_\nu)$ implies that the sum is not affected by permutation of the terms. Hence, in this case, the infinite product $\prod(1 + u_\nu)$ remains unchanged if we permute the factors. Therefore, we define $\prod(1 + u_\nu)$ to be *absolutely convergent* if $\sum u_\nu$ or, equivalently, $\sum \log(1 + u_\nu)$ converges absolutely.

We now extend the definition to arbitrary sequences a_ν of complex numbers. We call $\prod\limits_1^\infty a_\nu$ (absolutely) convergent, if there is an index ν_0 such that $a_\nu \neq 0$ for $\nu \geqslant \nu_0$, and $\prod\limits_{\nu_0}^\infty a_\nu$ converges (absolutely). We then set

$$\prod_1^\infty a_\nu = \prod_1^{\nu_0 - 1} a_\nu \cdot \lim_{n\to\infty} \prod_{\nu_0}^n a_\nu .$$

Clearly, this definition does not depend on the choice of ν_0.

Of course, we are mainly interested in infinite products of holomorphic functions. Let G be a domain in $\mathbb{C}$, and let $f_\nu = 1 + u_\nu$ be holomorphic on G. We always assume that no f_ν vanishes identically. The notion of pointwise convergence of $\prod\limits_1^\infty f_\nu(z) = \prod\limits_1^\infty \big(1 + u_\nu(z)\big)$ should be clear, but is of minor importance. The following condition implies that the product is well-behaved.

Definition 7.1. *The infinite product* $\prod(1 + u_\nu)$ *converges absolutely and locally uniformly on* G *if the series* $\sum u_\nu(z)$ *does.*

In view of Prop. 7.1 and 7.2 this implies pointwise convergence. If the condition in Def. 7.1 holds, then for any compact $K \subset G$ there is a ν_0 such that $|u_\nu(z)| < 1$, hence $f_\nu(z) \neq 0$, for all $z \in K$ and $\nu \geqslant \nu_0$. In particular, on K there are only a finite number of zeros of the factors f_ν.

Proposition 7.3. *Let* $u_\nu \in \mathcal{O}(G)$, $\nu \geqslant 1$. *If* $F(z) = \prod\limits_1^\infty \big(1 + u_\nu(z)\big)$ *converges locally uniformly and absolutely on* G*, then the sequence of the partial products* $p_n(z) = \prod\limits_1^n \big(1 + u_\nu(z)\big)$ *converges locally uniformly to* $F(z)$*, hence* $F(z)$ *is holomorphic on* G*. Moreover, permutation of the factors does not affect* $F(z)$*.*

Proof: First note

$$\big|e^z - 1\big| \leqslant 2|z| \quad \text{for } |z| \leqslant 1/2.$$

Therefore, if g_n is a sequence of continuous functions converging uniformly on a compact set K, then the g_n are uniformly bounded on K, i.e. $|g_n(z)| \leqslant M$ for all $z \in K$ and all n, and

$$\begin{aligned} \big|e^{g_n} - e^{g_m}\big| &= \big|e^{g_m}\big| \cdot \big|e^{g_n - g_m} - 1\big| \\ &\leqslant e^M \cdot 2|g_n - g_m| \end{aligned}$$

if $|g_n - g_m| \leqslant 1/2$ on K. Thus the functions e^{g_n} also converge uniformly on K.

Now, assume $\sum u_\nu$ is an absolutely and locally uniformly convergent series of holomorphic functions on G. For any compact $K \subset G$, there is a ν_1 such that $|u_\nu(z)| \leqslant 1/4$ for all $z \in K$ and $\nu \geqslant \nu_1$. By Prop. 7.2 the sequence $\sum_{\nu_1}^{n} \log\big(1+u_\nu(z)\big)$ converges uniformly on K. Therefore, the products $\prod_{\nu_1}^{n}\big(1+u_\nu(z)\big) = \exp\sum_{\nu_1}^{n}\log\big(1+u_\nu(z)\big)$ converge uniformly on K, and so do the $p_n(z)$. □

We are now ready to address the main theme of this section. Let $Z = \{a_1, a_2, a_3, \ldots\}$ be an infinite discrete subset of $\mathbb{C}$, and let n_ν, $\nu \geqslant 1$, be positive integers. We want to construct an entire function f which vanishes precisely at the a_ν and has order n_ν at a_ν.

If there is such a function f, its logarithmic derivative $h = f'/f$ will be holomorphic except for simple poles at the a_ν with residue n_ν. But a meromorphic function h with this property can be constructed by means of the Mittag-Leffler theorem, and we shall obtain the desired f by integration from h.

To this end, we assume for the moment that $0 \notin Z$ and arrange the a_ν according to increasing modulus:

$$0 < |a_1| \leqslant |a_2| \leqslant |a_3| \leqslant \ldots .$$

Thm. 6.1 then yields a series

$$\sum_{\nu=1}^{\infty} h_\nu(z) \quad \text{with} \quad h_\nu(z) = n_\nu\Big(\frac{1}{z-a_\nu} + p_\nu(z)\Big)$$

converging absolutely and locally uniformly to a meromorphic function h on $\mathbb{C}$ with the required poles and residues; the $p_\nu(z)$ are suitable polynomials.

Let $q_\nu(z)$ be the primitive of $p_\nu(z)$ with $q_\nu(0) = 0$, and define

$$f_\nu(z) = \Big[\Big(1 - \frac{z}{a_\nu}\Big)\exp q_\nu(z)\Big]^{n_\nu}.$$

This is an entire function vanishing only at a_ν (of order n_ν), satisfying $f_\nu(0) = 1$ and $f_\nu'(z)/f_\nu(z) = h_\nu(z)$.

We will show that the infinite product

$$f(z) = \prod_{1}^{\infty} f_\nu(z) = \prod_{1}^{\infty}\Big[\Big(1 - \frac{z}{a_\nu}\Big)\exp q_\nu(z)\Big]^{n_\nu}$$

converges absolutely and locally uniformly on $\mathbb{C}$, thus it solves our problem.

Fix $R > 0$ and ν_0 such that $|a_\nu| > R$ for $\nu \geqslant \nu_0$. Then $\sum_{\nu_0}^{\infty} h_\nu(z)$ is a series of holomorphic functions on $|z| \leqslant R$, converging absolutely and uniformly. The series

$$\sum_{\nu_0}^{\infty} \log f_\nu(z) = \sum_{\nu_0}^{\infty} \int_{[0,\,z]} h_\nu(\zeta)\, d\zeta$$

then converges absolutely and uniformly on $|z| \leqslant R$, too. By Prop. 7.2 and Def. 7.1 this implies absolute and uniform convergence of $\prod_{\nu_0}^{\infty} f_\nu$ on $|z| \leqslant R$.

For a complete statement of the result we use the p_ν as given in Thm. 6.1, namely as suitable partial sums of the power series expansion of $-1/(z - a_\nu)$ at $z = 0$. Furthermore, we admit $a_0 = 0$ as a possible zero of order $n_0 \geqslant 0$ by writing an additional factor z^{n_0}.

Theorem 7.4 (Weierstrass product theorem). *Let $Z = \{0, a_1, a_2, \dots\}$ be a discrete set in $\mathbb{C}$, arranged by increasing modulus. Let n_0 be a nonnegative integer, and for $\nu \geqslant 1$, let n_ν be a positive integer. Choose integers k_ν, $\nu \geqslant 1$, such that the series*

$$\sum_{1}^{\infty} n_\nu \left(\frac{1}{z - a_\nu} + \frac{1}{a_\nu} \sum_{\mu=0}^{k_\nu} \left(\frac{z}{a_\nu} \right)^{\mu} \right) \tag{1}$$

converges absolutely and locally uniformly on $\mathbb{C}$. Then the infinite product

$$f(z) = z^{n_0} \prod_{\nu=1}^{\infty} \left[\left(1 - \frac{z}{a_\nu} \right) \exp \sum_{\mu=1}^{k_\nu+1} \frac{1}{\mu} \left(\frac{z}{a_\nu} \right)^{\mu} \right]^{n_\nu} \tag{2}$$

is an entire function with zeros of order n_ν precisely at a_ν $(\nu \geqslant 0)$. □

It is reasonable to choose the k_ν as small as possible and desirable to have them all equal. E.g. if $\sum n_\nu |a_\nu|^{-1} < \infty$, $k_\nu = -1$ for all ν, i.e. $p_\nu(z) \equiv 0$ for all ν, will do.

Corollary 7.5. *Let a_ν, n_ν and f be as in the theorem. Then any entire function g with zeros of order n_ν at the a_ν (and no other zeros) can be written as $g(z) = e^{h(z)} f(z)$ with an entire function h.*

Namely, the quotient g/f is, after removing the singularities, an entire function without zeros, thus of the form $\exp h$.

Corollary 7.6. *Any meromorphic function on $\mathbb{C}$ is the quotient of two entire functions.*

Proof: Let $h \not\equiv 0$ be meromorphic on $\mathbb{C}$ with poles at a_ν of multiplicities n_ν. By Thm. 7.4 there is an entire function g with zeros of order n_ν at the a_ν (this statement is elementary if there are only finitely many a_ν). Then $f = g \cdot h$ has only removable singularities and can hence be regarded as an entire function. □

As an example, let us determine the product expansion of the entire function $\sin \pi z$. It has simple zeros at the points $\nu \in \mathbb{Z}$. As the series $\sum_{\nu \neq 0} \left(\frac{1}{z-\nu} + \frac{1}{\nu} \right)$ converges absolutely and locally uniformly on $\mathbb{C}$, Thm. 7.4 yields a product representation

$$\sin \pi z = e^{g(z)} \, z \prod_{\nu \neq 0} \left(1 - \frac{z}{\nu} \right) e^{z/\nu}. \tag{3}$$

To determine g, we form the logarithmic derivative of (3):

$$\pi \cot \pi z = g'(z) + \frac{1}{z} + {\sum}' \left(\frac{1}{z-\nu} + \frac{1}{\nu} \right).$$

By Prop. 6.3, $g' \equiv 0$, i.e. g is constant. Since $\lim_{z \to 0} \frac{\sin \pi z}{z} = \pi$, we have $e^g \equiv \pi$ and

Proposition 7.7. $\sin \pi z = \pi z \prod_{\nu \neq 0} \left(1 - \frac{z}{\nu} \right) e^{z/\nu} = \pi z \prod_{\nu=1}^{\infty} \left(1 - \frac{z^2}{\nu^2} \right).$ □

The second product results by combining the factors with index ν and $-\nu$ in the first product.

Euler found the product representation of $\sin \pi z$ in 1734 and proved it some years later. The specialization $z = 1/2$ yields a still older formula of Wallis (1655):

$$\frac{\pi}{2} = \prod_{\nu=1}^{\infty} \frac{(2\nu)^2}{(2\nu-1)(2\nu+1)} = \frac{2 \cdot 2}{1 \cdot 3} \cdot \frac{4 \cdot 4}{3 \cdot 5} \cdot \frac{6 \cdot 6}{5 \cdot 7} \cdot \dots \, .$$

A product representation of $\cos \pi z$ can be obtained on similar lines:

$$\cos \pi z = \prod_{\nu \in \mathbb{Z}} \left(1 - \frac{z}{a_\nu} \right) e^{z/a_\nu} \qquad \text{with } a_\nu = \nu - 1/2.$$

Exercises

1. Let $F(z) = \prod_1^\infty f_\nu(z)$ be a locally uniformly convergent product of holomorphic functions on G.
   a) Show that $H_n(z) = \prod_n^\infty f_\nu(z)$ converges locally uniformly to 1 on G, and hence $H_n' \to 0$ locally uniformly.
   b) $\dfrac{F'(z)}{F(z)} = \sum_1^\infty \dfrac{f_\nu'(z)}{f_\nu(z)}$, the right hand side being a locally uniformly convergent series of meromorphic functions on G.
2. Find a product representation of $e^{2\pi z} - 1$.

Chapter IV.

Integral formulas, residues, and applications

The concept of the winding number allows a general formulation of the Cauchy integral theorems (IV.1), which is indispensable for everything that follows. IV.2 presents a generalization of the Cauchy integral formula to real differentiable functions; it will play a basic role in Chapter VI. With the Laurent series expansion (IV.3) and the residue theorem (IV.4), further essential tools of complex analysis are at our disposal. They will be used to evaluate complicated integrals (IV.5) and then to study the equation $f(z) = w$, where f is a holomorphic function (IV.6). If one makes the integral formulas from sections IV.3 and IV.6 dependent on parameters, then one obtains the Weierstrass preparation theorem (IV.7), which gives fundamental information about the zeros of holomorphic functions of several variables.

The winding number was already used in complex analysis by Hadamard in 1910; it is a special case of the Kronecker index. Its systematic use for the development of complex analysis was first proposed by Artin (ca. 1944); the elegant proofs in IV.1 were developed by Dixon only in 1971 [D]. The inhomogeneous Cauchy integral formula seems to occur first in Pompeiu (1911); after 1950, it was exploited by Dolbeault and Grothendieck in the theory of several complex variables. Laurent series were introduced by Laurent in 1843 (Weierstrass already knew about them in 1841); the theory of residues (IV.4–IV.6) goes back to Cauchy. The Weierstrass preparation theorem has a complicated history; it was rediscovered and proved several times – cf. [Car]. Weierstrass published it in 1886 but knew of it since roughly 1860. The proof in IV.7, which is based on a clever application of the residue formulas from the preceding sections, is due to Stickelberger (1887) [St]; in presenting it, we follow [Fo1] and [Ra].

1. The general Cauchy integral theorem

In Chapter II, we proved the Cauchy integral theorem for star-shaped domains; it is time to get rid of this restriction. More precisely, the following question is to be answered:

For which closed paths of integration γ in an arbitrary domain G is the Cauchy integral theorem valid, i.e. for which γ do we have

$$\int_\gamma f(z)\,dz = 0$$

for every holomorphic function on G? – Equivalently: Given two paths of integration γ_1 and γ_2 from a to b in G, when does the identity

$$\int_{\gamma_1} f(z)\,dz = \int_{\gamma_2} f(z)\,dz$$

W. Fischer, I. Lieb, *A Course in Complex Analysis*, DOI 10.1007/978-3-8348-8661-3_4,

hold for all holomorphic functions on G?

It is clear that an answer will be especially useful for evaluating integrals, e.g. by changing the path of integration. Before turning to our question, let us recall Def. III.5.5.

Definition 1.1. *Let γ be a closed path of integration, and let z be a point that does not lie on γ, i.e. $z \notin \operatorname{Tr}\gamma$. The winding number of γ about z is*

$$n(\gamma, z) = \frac{1}{2\pi i} \int\limits_{\gamma} \frac{d\zeta}{\zeta - z}.$$

In the previous chapter, we saw that $n(\gamma, z)$ is always an integer. Here is a simple alternative proof:

Let $\gamma\colon [0,1] \to \mathbb{C}$ be piecewise continuously differentiable with $\gamma(0) = \gamma(1)$, and for $t \in [0,1]$, let

$$h(t) = \frac{1}{2\pi i} \int\limits_0^t \frac{\gamma'(\tau)}{\gamma(\tau) - z}\, d\tau.$$

Then $h(0) = 0$ and $h(1) = n(\gamma, z)$. Moreover,

$$h'(t) = \frac{1}{2\pi i} \frac{\gamma'(t)}{\gamma(t) - z}.$$

Thus

$$\frac{d}{dt} \frac{e^{2\pi i\, h(t)}}{\gamma(t) - z} = \frac{e^{2\pi i\, h(t)}}{\big(\gamma(t) - z\big)^2} \Big(\big(\gamma(t) - z\big) 2\pi i\, h'(t) - \gamma'(t) \Big) = 0,$$

so that

$$e^{2\pi i\, h(t)} = c(\gamma(t) - z)$$

with a nonzero constant c. Now $\gamma(0) = \gamma(1)$ implies that

$$1 = e^{2\pi i\, h(0)} = c(\gamma(0) - z) = c(\gamma(1) - z) = e^{2\pi i\, h(1)},$$

i.e.

$$e^{2\pi i\, n(\gamma, z)} = 1.$$

It follows that $n(\gamma, z)$ is an integer. □

To give a simple example, note that for $\kappa_m(t) = r\,e^{imt}$, where $0 \leqslant t \leqslant 2\pi$, the *m-times traversed circle*, we have

$$n(\kappa_m, z) = \begin{cases} m & \text{for } |z| < r \\ 0 & \text{for } |z| > r. \end{cases}$$

Here m can be negative – the circle is then traversed in the clockwise direction. Furthermore:

Lemma 1.1. *The function*

$$z \mapsto n(\gamma, z) \tag{1}$$

is locally constant and vanishes on the unbounded path component of the set $\mathbb{C} \setminus \operatorname{Tr}\gamma$.

Proof: The first claim follows from the continuity of the function (1), which can only take on integer values. Assume that $\operatorname{Tr}\gamma$ is contained in the disk $D_R(0)$. Then

$$|n(\gamma, z)| \leqslant \frac{1}{2\pi} L(\gamma) \frac{1}{|z| - R} < 1,$$

provided $|z|$ is large enough. Thus $n(\gamma, z) = 0$ for these z, and by continuity on the entire unbounded path component of $\mathbb{C} \setminus \operatorname{Tr}\gamma$. □

It will be useful to slightly generalize the definition of a closed path of integration:

Definition 1.2. *A cycle* Γ *is a formal linear combination of closed paths of integration with integer coefficients:*

$$\Gamma = n_1\gamma_1 + \ldots + n_r\gamma_r. \tag{2}$$

The integral of a continuous function defined on the trace of Γ,

$$\operatorname{Tr}\Gamma = \operatorname{Tr}\gamma_1 \cup \ldots \cup \operatorname{Tr}\gamma_r,$$

is

$$\int\limits_{\Gamma} f(z)\,dz = \sum_{\rho=1}^{r} n_\rho \int\limits_{\gamma_\rho} f(z)\,dz.$$

The winding number of Γ *about* $z \notin \operatorname{Tr}\Gamma$ *is defined to be*

$$n(\Gamma, z) = \sum_{\rho=1}^{r} n_\rho n(\gamma_\rho, z).$$

The length of Γ *is* $L(\Gamma) = \sum_{\rho=1}^{r} |n_\rho| L(\gamma_\rho)$.

A few comments:

A summand of the form $0\cdot\gamma$ can always be added or subtracted from (2) – in defining the trace of Γ, we of course count only those γ_ρ with nonzero coefficients; we always set $1\cdot\gamma=\gamma$ and $(-1)\cdot\gamma=-\gamma$. Cycles can be added:

$$\sum_{\rho=1}^{r} n_\rho\gamma_\rho + \sum_{\rho=1}^{r} m_\rho\gamma_\rho = \sum_{\rho=1}^{r}(n_\rho+m_\rho)\gamma_\rho.$$

They therefore form an abelian group. We have

$$\int_{\Gamma_1} f(z)\,dz + \int_{\Gamma_2} f(z)\,dz = \int_{\Gamma_1+\Gamma_2} f(z)\,dz, \tag{3}$$

and, in particular,

$$n(\Gamma_1+\Gamma_2,z) = n(\Gamma_1,z)+n(\Gamma_2,z) \tag{4}$$
$$n(-\Gamma,z) = -n(\Gamma,z) \tag{5}$$

(for every z for which the winding number is defined). It should be noted that the plus sign in (2) has nothing to do with the combination of paths as in I.5.

We now describe a configuration that will be important for many considerations to follow:

Let $0<r<R$ be real numbers, $z_0\in\mathbb{C}$, and

$$\Gamma=\kappa(R;z_0)-\kappa(r;z_0).$$

The *annulus*

$$K(z_0;r,R)=\{z\colon r<|z-z_0|<R\}$$

is characterized by

$$n(\Gamma,z)=1,$$

and the complement of its closure by $n(\Gamma,z)=0$. We generalize this situation as follows:

Definition 1.3. *A domain with positive boundary is a bounded domain G in $\mathbb{C}$ whose boundary ∂G is the union of the traces of $n+1$ pairwise disjoint simple closed paths of integration $\gamma_0,\dots,\gamma_n$. For the cycle*

$$\Gamma=\gamma_0-\gamma_1-\dots-\gamma_n,$$

we require

$$G=\{z\in\mathbb{C}\colon n(\Gamma,z)=1\}$$
$$\mathbb{C}\setminus\overline{G}=\{z\in\mathbb{C}\colon n(\Gamma,z)=0\}.$$

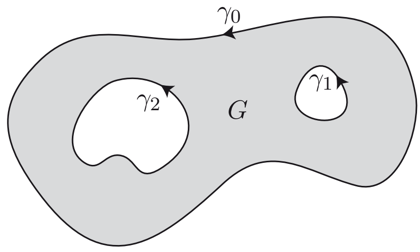


Figure 6. A domain with positive boundary

A closed path $\gamma\colon [a, b] \to \mathbb{C}$ is *simple*, if γ is injective on $[a, b[$. – The cycle Γ is also denoted by ∂G and called the boundary (or boundary cycle) of G. Intuitively, G lies "on the left hand side of its (oriented) boundary". Examples are disks, annuli, and general domains such as the one shown in Fig. 6.

Using winding numbers, we now introduce the notion of homology that will be decisive for what follows.

Definition 1.4. *A cycle Γ in G is called null-homologous in G if for every point $z \notin G$, we have*

$$n(\Gamma, z) = 0.$$

Two cycles Γ_1 and Γ_2 are called homologous in G if $\Gamma_1 - \Gamma_2$ is null-homologous. Notation: $\Gamma \sim 0$, $\Gamma_1 \sim \Gamma_2$.

The circle $\kappa(r; 0)$ is null-homologous in $\mathbb{C}$ but not in $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$.

The main result of this section is

Theorem 1.2 (The general Cauchy integral theorem and general Cauchy integral formulas)**.** *Let Γ be a null-homologous cycle in a domain G, and let f be a holomorphic function on G. Then*

i. $$\int_\Gamma f(z)\,dz = 0.$$

ii. For every $z \notin \operatorname{Tr}\Gamma$ and all $k = 0, 1, 2, \ldots$, we have

$$n(\Gamma, z) f^{(k)}(z) = \frac{k!}{2\pi i} \int_\Gamma \frac{f(\zeta)}{(\zeta - z)^{k+1}}\, d\zeta.$$

(For $z \notin G$, *ii* says that the integral on the right hand side is 0.)

Proof: We will first prove the second claim, where we may of course set $k = 0$: the general claim then follows by differentiating under the integral sign. By substituting the definition of the winding number into the formula in *ii*, it takes the form

$$\int_\Gamma \frac{f(\zeta) - f(z)}{\zeta - z}\, d\zeta = 0, \text{ for } z \in G \setminus \operatorname{Tr}\Gamma. \tag{6}$$

To prove this, we consider the integral (6) as a function of z. We will extend this function to an entire function h; the latter will then satisfy

$$\lim_{z\to\infty} h(z) = 0,$$

which, by Liouville's theorem, implies that $h(z) \equiv 0$.

Let us investigate the integrand of (6) as a function of ζ and z simultaneously, i.e.

$$g(\zeta, z) = \begin{cases} \dfrac{f(\zeta) - f(z)}{\zeta - z}, & \zeta \neq z \\ f'(z), & \zeta = z. \end{cases}$$

This function is defined on $G \times G$ and – as we will show – is continuous in both variables.

This is clear for $\zeta \neq z$. To prove continuity on the diagonal, we consider the difference

$$g(\zeta, z) - g(z_0, z_0)$$

in a neighbourhood $U_\delta(z_0) \times U_\delta(z_0)$

a) if $\zeta = z$:

$$g(z, z) - g(z_0, z_0) = f'(z) - f'(z_0),$$

b) if $\zeta \neq z$:

$$g(\zeta, z) - g(z_0, z_0) = \frac{f(\zeta) - f(z)}{\zeta - z} - f'(z_0) = \frac{1}{\zeta - z} \int_{[z,\zeta]} \big(f'(w) - f'(z_0)\big)\, dw.$$

Now, the derivative f' is continuous – see Thm. II.3.4. Therefore, for any $\varepsilon > 0$, the number δ can be chosen small enough such that

$$|f'(w) - f'(z_0)| < \varepsilon$$

for all $w \in U_\delta(z_0)$. In case a), we then have

$$|g(z, z) - g(z_0, z_0)| < \varepsilon,$$

and in case b),

$$|g(\zeta, z) - g(z_0, z_0)| \leqslant \frac{1}{|\zeta - z|}|\zeta - z|\varepsilon = \varepsilon.$$

This shows that g is continuous. We now put

$$h_0(z) = \int_\Gamma g(\zeta, z)\, d\zeta.$$

This function is continuous on G. If, moreover, γ is the boundary of a triangle that lies in G, then

$$\int_\gamma h_0(z)\, dz = \int_\gamma \int_\Gamma g(\zeta, z)\, d\zeta\, dz = \int_\Gamma \int_\gamma g(\zeta, z)\, dz\, d\zeta.$$

For ζ fixed, the function

$$z \mapsto g(\zeta, z)$$

is holomorphic for $z \neq \zeta$ and continuous at ζ and hence holomorphic everywhere. Thus

$$\int_\gamma g(\zeta, z)\, dz = 0$$

by Goursat's lemma; it follows that

$$\int_\gamma h_0(z)\, dz = 0,$$

and by Morera's theorem, h_0 is holomorphic.

We now use the fact that Γ is null-homologous. Let

$$G_0 = \{z \in \mathbb{C} \colon n(\Gamma, z) = 0\}.$$

On $G \cap G_0$, we have

$$h_0(z) = \int_\Gamma \frac{f(\zeta)}{\zeta - z}\, d\zeta \stackrel{\text{def}}{=} h_1(z),$$

but the integral allows one to define h_1 as a holomorphic function on all of G_0. Therefore, h_0 can be holomorphically extended to $G \cup G_0$ via

$$h(z) = \begin{cases} h_0(z) & \text{for } z \in G \\ h_1(z) & \text{for } z \in G_0. \end{cases}$$

But since $\Gamma \sim 0$, we have

$$G \cup G_0 = \mathbb{C},$$

i.e. h is an entire function. For large $|z|$,

$$|h(z)| = |h_1(z)| \leqslant \frac{1}{\operatorname{dist}(z, \operatorname{Tr}\Gamma)} \cdot L(\Gamma) \cdot \max_{\operatorname{Tr}\Gamma} |f|$$

holds on G_0. It follows that

$$\lim_{z \to \infty} h(z) = 0,$$

so that $h(z) \equiv 0$ by Liouville's theorem. This is what we wanted to show.

We now derive the first claim of the theorem from the second. Let a be an arbitrary point in $G \setminus \operatorname{Tr}\Gamma$. The function $F(z) = (z-a)f(z)$ is holomorphic on G and satisfies $F(a) = 0$. By part *ii* of the theorem, we then have

$$0 = n(\Gamma, a)F(a) = \frac{1}{2\pi i} \int_\Gamma \frac{F(z)}{z-a}\, dz = \frac{1}{2\pi i} \int_\Gamma f(z)\, dz,$$

so that indeed

$$\int_\Gamma f(z)\, dz = 0. \qquad \square$$

We note as an immediate consequence:

Theorem 1.3. *Let Γ_1 and Γ_2 be homologous cycles in G. Then for every holomorphic function on G,*

$$\int_{\Gamma_1} f(z)\, dz = \int_{\Gamma_2} f(z)\, dz.$$

This is because

$$0 = \int_{\Gamma_1 - \Gamma_2} f(z)\, dz = \int_{\Gamma_1} f(z)\, dz - \int_{\Gamma_2} f(z)\, dz.$$

The concept of a star-shaped domain can now be generalized:

Definition 1.5. *A domain $G \subset \mathbb{C}$ is called simply connected if every cycle in G is null-homologous.*

It is in such domains – and clearly only in such domains! – that the Cauchy integral theorem

$$\int_{\Gamma} f(z)\, dz = 0$$

holds for all holomorphic functions on G and all cycles in G. Examples include domains with positive boundary that have only one "boundary component" $\Gamma = \gamma_0$ – see Def. 1.3. For simply connected domains G – and only for such domains! – the following holds: Every holomorphic function f on G has a primitive; if f has no zeros, then f has a holomorphic logarithm and thus holomorphic powers with arbitrary exponents (cf. Prop. III.5.2).

In the case of a domain with positive boundary, Thm. 1.2 takes an especially simple and often used form:

Theorem 1.4 (Cauchy integral theorems for domains with positive boundary)**.** *Let f be holomorphic in a neighbourhood U of $\overline{G}$, where G is a domain with positive boundary. Then we have:*

i. $\displaystyle\int_{\partial G} f(z)\, dz = 0.$

ii. For $z \in G$ and all $k = 0, 1, 2, \ldots,$

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

We may assume that U is connected and note that $\partial G \sim 0$ in U. This implies *i*; the second claim follows from Thm. 1.2.*ii* and the fact that $n(\partial G, z) = 1$. □

We note without proof that one only needs to assume that f is continuous on $\overline{G}$ and holomorphic in the interior of G.

In order to check our assumptions, we have to know how to compute winding numbers. Let γ be a closed path of integration, and let $U = \mathbb{C} \setminus \operatorname{Tr}\gamma$. The set U can be decomposed into path components, among which there is exactly one unbounded component U_0. By Lemma 1.1, $n(\gamma, z)$ is constant on the path components of U and vanishes on U_0. The remaining values can be obtained using the following "right of way" rule:

Let a and b be points from different components of U, and suppose there is a path in $\mathbb{C}$ from a to b that intersects γ exactly once at a "regular" point of γ, whereby γ is crossed from right to left (with respect to the orientation of γ). Then

$$n(\gamma, b) = n(\gamma, a) + 1.$$

We refer the reader to [FL1] for a detailed explanation.

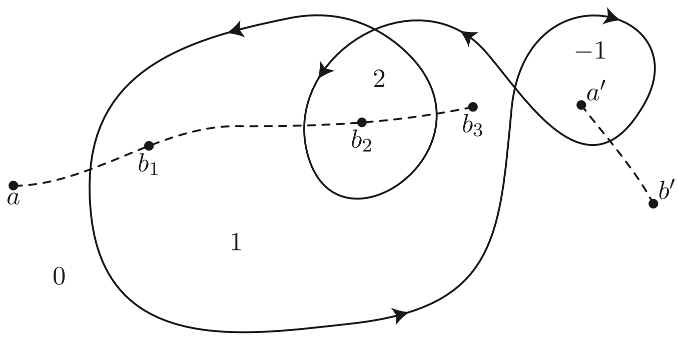


Figure 7. Determining the winding number

We also note the following intuitive fact without proof:

Every simple closed continuously differentiable path of integration γ in $\mathbb{C}$ is (if given an appropriate orientation) the boundary of a simply connected domain with positive boundary. We again refer the reader to [FL1] for a proof.

Exercises

1. Suppose the domain G is bounded by a simple closed polygonal path. Show that G is simply connected. Extend this claim to piecewise smooth boundaries.
2. Determine whether the following domains are simply connected:

$$\mathbb{C} \setminus \{0\}, \quad \mathbb{C} \setminus [-1,1], \quad \mathbb{C} \setminus \mathbb{R}_{\leqslant 0}, \quad \mathbb{C}^* \setminus \{z = e^{t(1+i)} : t \in \mathbb{R}\}.$$

3. Show that the image of a simply connected domain under a biholomorphic mapping is simply connected. Is it sufficient to assume that the mapping is locally biholomorphic?

2. The inhomogeneous Cauchy integral formula

A different method of generalising Cauchy's integral formula is based on Stokes' theorem. We consider a domain G with positive boundary ∂G as defined in the previous section, and a continuously differentiable function f defined in a neighbourhood of $\overline{G}$; f need not be holomorphic!

Let us fix a point $z \in G$ and choose a disk $D_r(z) \subset\subset G$. The complement $G \setminus \overline{D_r(z)} = G_r$ is a domain with positive boundary $\partial G - \partial D_r(z)$. Application of Stokes' theorem to the continuously differentiable 1-form

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

on G_r yields

$$\int_{\partial G_r} \omega(\zeta) = \int_{G_r} d\omega(\zeta).$$

Inserting the definitions we obtain

$$\frac{1}{2\pi i}\int_{\partial G} \frac{f(\zeta)}{\zeta - z}\, d\zeta = \frac{1}{2\pi i}\int_{\partial D_r(z)} \frac{f(\zeta)}{\zeta - z}\, d\zeta - \frac{1}{2\pi i}\int_{G_r} \frac{\partial f/\partial\bar\zeta}{\zeta - z}\, d\zeta \wedge d\bar\zeta.$$

For $r \to 0$ the surface integral tends to

$$\frac{1}{2\pi i}\int_{G} \frac{\partial f/\partial\bar\zeta}{\zeta - z}\, d\zeta \wedge d\bar\zeta$$

(since the integrand is integrable); the boundary integral over $\partial D_r(z)$ can be rewritten as

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

The first integral yields

$$f(z) \cdot \frac{1}{2\pi i}\int_{\partial D_r(z)} \frac{d\zeta}{\zeta - z} = f(z);$$

the second integral can be estimated, using $|\zeta - z| = r$, by

$$\left|\frac{1}{2\pi i}\int_{\partial D_r(z)} \frac{f(\zeta) - f(z)}{\zeta - z}\, d\zeta\right| \leqslant \max_{|\zeta - z| = r} |f(\zeta) - f(z)|,$$

which tends to zero with r. So we have established

Theorem 2.1 (Inhomogeneous Cauchy formula)**.** *Let f be a continuously differentiable function in a neighbourhood of a domain G with positive boundary ∂G. Then one has, for each point $z \in G$,*

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

Note that if f is holomorphic in G, we have again Thm. 1.4.

3. Laurent decomposition and Laurent expansion

Functions that are holomorphic on disks can be represented as power series. We now present a generalization for functions that are holomorphic on an annulus.

Let $r < R \leqslant \infty$ be real numbers, with $R > 0$. We denote by

$$K(r,R) = \{z \colon r < |z| < R\} \tag{1}$$

the annulus centred at 0 with inner radius r and outer radius R. If $r < 0$, then $K(r,R)$ is the disk $D_R(0)$, and if $r = 0$, it is the punctured disk $D_R(0) \setminus \{0\}$. If one chooses another centre, say z_0, then we write

$$K(z_0; r, R) = \{z \colon r < |z - z_0| < R\}. \tag{2}$$

The following theorems are formulated in terms of the general situation (2), but to simplify notation, we will always set $z_0 = 0$ in the proofs, i.e. we will work with the case (1).

Theorem 3.1 (Laurent decomposition)**.** *Let f be holomorphic in an annulus (2). Then there exist unique holomorphic functions f_0 on $D_R(z_0)$ and f_∞ on $\mathbb{C} \setminus \overline{D_r(z_0)}$ such that*

$$f = f_0 + f_\infty \tag{3}$$

$$\lim_{z\to\infty} f_\infty(z) = 0. \tag{4}$$

Definition 3.1. *The decomposition given by* (3) *and* (4) *is called the Laurent decomposition of f. The terms f_0 and f_∞ are called the regular part and principal part, respectively, of the decomposition.*

Laurent decompositions have already appeared in special cases in the theory of isolated singularities (II.6).

Proof of Thm. 3.1**:** a) *Uniqueness.* Let

$$0 = f_0 + f_\infty, \quad \lim_{z\to\infty} f_\infty(z) = 0 \tag{5}$$

be a Laurent decomposition of the zero function. Setting

$$\widetilde{f}(z) = \begin{cases} f_0(z) & \text{for } |z| < R \\ -f_\infty(z) & \text{for } |z| > r, \end{cases}$$

one obtains an entire function $\widetilde{f}$ that tends to 0 as $z \to \infty$. Thus $\widetilde{f}(z) \equiv 0$, and hence $f_0(z) \equiv 0 \equiv f_\infty(z)$.

b) *Existence.* Assume that $r \geqslant 0$ – the case $r < 0$ is trivial, since then $f = f_0$ and $f_\infty = 0$.

Choose r' and R' such that $r < r' < R' < R$. For z with $r' < |z| < R'$, the Cauchy integral formula gives

$$f(z) = \frac{1}{2\pi i} \int\limits_{|\zeta|=R'} \frac{f(\zeta)}{\zeta - z}\, d\zeta - \frac{1}{2\pi i} \int\limits_{|\zeta|=r'} \frac{f(\zeta)}{\zeta - z}\, d\zeta$$

$$\stackrel{\text{def}}{=} f_0(z) + f_\infty(z). \tag{6}$$

The functions f_0 and f_∞ defined by the above integrals are holomorphic for $|z| < R'$ and $|z| > r'$, respectively, and $f_\infty(z) \to 0$ as $z \to \infty$. By the Cauchy integral theorem for homologous cycles, the integrals in (6) are independent of R' and r', respectively, as long as $r < r' < |z| < R' < R$. Therefore, f_0 and f_∞ can be extended holomorphically to all of $D_R(0)$ and $\mathbb{C} \setminus \overline{D_r(0)}$, respectively, preserving the decomposition (3) and the property (4). □

Let $f = f_0 + f_\infty$ be the Laurent decomposition of a function f that is holomorphic in an annulus $K(r, R)$. The function f_0, which is holomorphic on $D_R(0)$, can be expanded into its (locally uniformly convergent) Taylor series:

$$f_0(z) = \sum_{\nu=0}^{\infty} a_\nu z^\nu, \quad |z| < R. \tag{7}$$

The principal part f_∞ is holomorphic on $\mathbb{C} \setminus \overline{D_r(0)}$ and, since $\lim_{z\to\infty} f_\infty(z) = 0$, can be extended to a holomorphic function on $\widehat{\mathbb{C}} \setminus \overline{D_r(0)}$ by setting $f_\infty(\infty) = 0$. Consequently,

$$g(z) = f_\infty\left(\frac{1}{z}\right), \quad g(0) = 0 \tag{8}$$

is holomorphic on the disk of radius $1/r$ and has the series expansion

$$g(z) = \sum_{\nu=1}^{\infty} b_\nu z^\nu \tag{9}$$

there (note that b_0 must vanish!). We rewrite (9) as

$$f_\infty(z) = g\left(\frac{1}{z}\right) = \sum_{\nu=1}^{\infty} b_\nu \frac{1}{z^\nu} = \sum_{\nu=-1}^{-\infty} a_\nu z^\nu, \tag{10}$$

where $a_\nu = b_{-\nu}$. By combining (7) and (10), we obtain:

Theorem 3.2 (Laurent expansion)**.** *Let f be holomorphic in the annulus $K(z_0; r, R)$. Then f can be expanded into an absolutely locally uniformly convergent series*

$$f(z) = \sum_{-\infty}^{\infty} a_\nu (z - z_0)^\nu. \tag{11}$$

The coefficients a_ν satisfy

$$a_\nu = \frac{1}{2\pi i} \int\limits_{|z-z_0|=\rho} f(z)(z - z_0)^{-(\nu+1)}\, dz, \tag{12}$$

where ρ can be chosen to be any number between r and R.

It remains only to show (12), but this follows immediately upon substituting (11) into the integral in (12) and exchanging the order of summation and integration. □

The sum (11) means

$$\sum_{-\infty}^{\infty} a_\nu (z - z_0)^\nu = \sum_{\nu=-1}^{-\infty} a_\nu (z - z_0)^\nu + \sum_{\nu=0}^{\infty} a_\nu (z - z_0)^\nu, \tag{13}$$

i.e. the left hand side is defined to be the right hand side of (13).

Series like

$$\sum_{-\infty}^{\infty} a_\nu (z - z_0)^\nu$$

that appear in Thm. 3.2 are called *Laurent series*; they are convergent at z if the *principal part*

$$\sum_{\nu=-1}^{-\infty} a_\nu (z - z_0)^\nu$$

and the *regular part*

$$\sum_{\nu=0}^{\infty} a_\nu (z - z_0)^\nu$$

are convergent at z. Their domain of convergence is either an annulus $K(z_0; r, R)$ or empty; in the former case, they converge compactly to a holomorphic function. Using the standard estimate, formula (12) yields the Cauchy inequalities:

Proposition 3.3. *Let f be given as*

$$f(z) = \sum_{-\infty}^{\infty} a_\nu (z - z_0)^\nu$$

in an annulus $K(z_0; r, R)$. Then for every ρ such that $r < \rho < R$, we have

$$|a_\nu| \leqslant \rho^{-\nu} \max_{|z-z_0|=\rho} |f(z)|.$$

Power series are of course special cases of Laurent series. Similar to power series expansions, Laurent expansions of rational functions can also be calculated by applying geometric series. Here is an example:

The function

$$f(z) = \frac{1}{z(z-1)^2}$$

is holomorphic on $\mathbb{C} \backslash \{0, 1\}$, and we already know its Laurent expansion in $K(0; 0, 1)$:

$$\frac{1}{z(z-1)^2} = \frac{1}{z} \frac{d}{dz} \frac{1}{1-z} = \frac{1}{z} \sum_{\nu=1}^{\infty} \nu z^{\nu-1}.$$

Accordingly, in $K(0; 1, \infty)$ one has

$$\frac{1}{z(z-1)^2} = \frac{1}{z^3} \sum_{\nu=1}^{\infty} \nu \left(\frac{1}{z}\right)^{\nu-1}.$$

To conclude, we record the identity theorem for Laurent series:

Proposition 3.4. *If the Laurent series $\sum_{-\infty}^{\infty} a_\nu z^\nu$ converges to 0 in an annulus $K(r, R)$, then all a_ν are zero.*

With these results, we return to the theory of isolated singularities. The function f has an isolated singularity at z_0 if and only if it is holomorphic in a punctured disk

$$D_R(z_0) \setminus \{z_0\} = K(z_0; 0, R).$$

It thus has the absolutely locally uniformly convergent Laurent expansion

$$f(z) = \sum_{-\infty}^{\infty} a_\nu (z - z_0)^\nu = \sum_{\nu=-1}^{-\infty} a_\nu (z - z_0)^\nu + \sum_{\nu=0}^{\infty} a_\nu (z - z_0)^\nu \tag{14}$$

in $K(z_0; 0, R)$, which we have decomposed into its principal and regular parts. The case of a finite principal part – i.e. the case when $a_\nu = 0$ for $\nu < -n$ – is the familiar case of a pole of order less than or equal to n at z_0, or a removable singularity, in case the principal part in (14) vanishes. The case of an infinite principal part – i.e. when infinitely many a_ν are nonzero for $\nu < 0$ – must therefore correspond to an essential singularity, which gives us:

Proposition 3.5. *The function f has an essential singularity at z_0 if and only if the principal part of its Laurent expansion about z_0 contains infinitely many powers of $(z - z_0)^{-1}$.*

Note that the principal part in (14) is always holomorphic on $\widehat{\mathbb{C}} \setminus \{z_0\}$.

We conclude with the following observation: Let f be holomorphic on $\mathbb{C}$ with the exception of a discrete set of singularities S. Given an arbitrary point $z_0 \in \mathbb{C}$, the Taylor series (or Laurent series) of f converges in the largest (possibly punctured) disk about z_0 that contains no points in $S \setminus \{z_0\}$. Together with the Cauchy-Hadamard formula for the radius of convergence, this statement is occasionally useful for computing limits (see Ex. 3).

Exercises

1. Determine the sets on which the following Laurent series converge:

$$\sum_{-\infty}^{\infty} 2^{-|\nu|} z^\nu, \quad \sum_{-\infty}^{\infty} \frac{(z-1)^\nu}{3^\nu + 1}, \quad \sum_{-\infty}^{\infty} \frac{z^\nu}{1+\nu^2}, \quad \sum_{-\infty}^{\infty} 2^\nu (z+2)^\nu.$$

2. Find the principal part of the Laurent expansion of

$$\frac{z-1}{\sin^2 z} \quad \text{in } 0 < |z| < \pi \qquad \text{and of} \qquad \frac{z}{(z^2+b^2)^2} \quad \text{in } 0 < |z - ib| < 2b.$$

3. a) Let $a_\nu = 3 \cdot 2^\nu + 2 \cdot 3^\nu$. Find $\lim_{\nu\to\infty} a_\nu^{1/\nu}$ by considering the series $\sum_{\nu=0}^{\infty} a_\nu z^\nu$.

   b) Compute $\limsup\limits_{\nu\to\infty} \left(\dfrac{|E_{2\nu}|}{(2\nu)!} \right)^{1/2\nu}$ (cf. II.4, Ex. 3).

   It follows that Euler numbers can become very large.

4. Let $S_\omega = \{a < \operatorname{Im}(z/\omega) < b\}$ with $\omega \in \mathbb{C}^*$, $-\infty \leqslant a < b \leqslant +\infty$, be a parallel strip. Then $w = F(z) = \exp(2\pi i z/\omega)$ maps S_ω onto an annulus $K = K(r, R)$. Let f be holomorphic on S_ω and satisfy the periodicity condition $f(z + \omega) \equiv f(z)$. Show that there is a unique holomorphic function $\tilde{f} \colon K \to \mathbb{C}$ such that $f = \tilde{f} \circ F$. Translate the Laurent series of $\tilde{f}$ into a Fourier series representation $f(z) = \sum_{-\infty}^{\infty} a_n \exp(2\pi i n z/\omega)$. Express the coefficients a_n by f.

4. Residues

Definition 4.1. *Let f be holomorphic on a domain G with the exception of a discrete set S. The residue of f at the point $z_0 \in G$ is the number*

$$\operatorname{res}_{z_0} f = \frac{1}{2\pi i} \int_{\kappa(r;z_0)} f(z)\, dz, \tag{1}$$

where r is chosen small enough that the disk $\overline{D_r(z_0)}$ lies in G and contains no points of S except possibly z_0.

Taking the above restriction into account, r can be chosen arbitrarily as long as $\overline{D_r(z_0)} \subset G$; the value of the integral (1) will not change. Similarly, the value of the integral will not change if one replaces $\kappa(r; z_0)$ with a cycle Γ such that $n(\Gamma, z_0) = 1$ and $n(\Gamma, z) = 0$ for all $z \in S \setminus \{z_0\}$.

The function f has isolated singularities at the points in S. Formula (12) from the previous section gives us:

$$\operatorname{res}_{z_0} f = a_{-1}, \tag{2}$$

where

$$f(z) = \sum_{-\infty}^{\infty} a_\nu (z - z_0)^\nu$$

is the Laurent expansion of f about z_0. In particular, the residue is 0 in the case of a removable singularity – but not only in this case! Formula (2) is sometimes also called the local residue theorem. For example, (1) and (2) immediately yield

$$\int_{\kappa(r;0)} e^{1/z}\, dz = 2\pi i,$$

a result that is even more remarkable when one considers that $\exp(1/z)$ is not an elementary integrable function. This leads to an ansatz for evaluating certain definite integrals: this method will be expanded upon in the following section. Our next goal is to "globalize" formula (2):

Theorem 4.1 (The residue theorem)**.** *Let f be holomorphic on a domain G with the possible exception of a discrete set $S \subset G$, and let Γ be a null-homologous cycle in G that does not meet any point in S. Then*

$$\frac{1}{2\pi i} \int_\Gamma f(z)\, dz = \sum_{z \in G} n(\Gamma, z) \operatorname{res}_z f. \tag{3}$$

Proof: The sum in (3) is finite, because there are only finitely many points $z \in S$ such that $n(\Gamma, z) \neq 0$. Let $z_1, \ldots, z_k$ be these points; then the cycle Γ is also null-homologous in $G \setminus S'$, where S' is the remaining set of points in S. For every $z_\varkappa$, let $h_\varkappa$ denote the principal part of the Laurent expansion of f about $z_\varkappa$:

$$h_\varkappa(z) = \sum_{\nu=-1}^{-\infty} a_\nu^\varkappa (z - z_\varkappa)^\nu. \tag{4}$$

The function $h_\varkappa$ is holomorphic on $\widehat{\mathbb{C}} \setminus \{z_\varkappa\}$, and the function

$$F(z) = f(z) - \sum_{\varkappa=1}^{k} h_\varkappa(z) \tag{5}$$

is holomorphic on $G \setminus S'$, i.e. it can be holomorphically extended to the points $z_\varkappa$. The Cauchy integral theorem now implies

$$\int_\Gamma F(z)\,dz = 0,$$

so that

$$\int_\Gamma f(z)\,dz = \sum_{\varkappa=1}^{k} \int_\Gamma h_\varkappa(z)\,dz. \tag{6}$$

We compute the integrals on the right hand side:

$$\begin{aligned}\int_\Gamma h_\varkappa(z)\,dz &= \sum_{\nu=-1}^{-\infty} a_\nu^\varkappa \int_\Gamma (z - z_\varkappa)^\nu\,dz \\ &= a_{-1}^\varkappa \cdot n(\Gamma, z_\varkappa) \cdot 2\pi i \\ &= \operatorname{res}_{z_\varkappa} f \cdot n(\Gamma, z_\varkappa) \cdot 2\pi i.\end{aligned} \tag{7}$$

Substituting (7) into (6) yields the claim. □

As an immediate corollary, we have:

Corollary 4.2. *Let G be a domain with positive boundary cycle $\partial G = \Gamma$, and let f be holomorphic in a neighbourhood of $\overline{G}$ with the exception of isolated singularities, none of which lie on the boundary. Then*

$$\frac{1}{2\pi i} \int_{\partial G} f(z)\,dz = \sum_{z \in G} \operatorname{res}_z f. \tag{8}$$

In order to apply the above formulas, one must of course be able to calculate residues. The following rules are easily proved:

$$\operatorname{res}_z(af + bg) = a \operatorname{res}_z f + b \operatorname{res}_z g, \tag{9}$$

where $a, b \in \mathbb{C}$. If f has a simple pole at z_0, then

$$\operatorname{res}_{z_0} f = \lim_{z \to z_0} f(z)(z - z_0); \tag{10}$$

if, in addition, g is holomorphic in z_0, then

$$\operatorname{res}_{z_0}(gf) = g(z_0) \operatorname{res}_{z_0} f. \tag{11}$$

For poles of order $n \geqslant 2$, the situation is slightly more complicated:

$$\operatorname{res}_{z_0} f = \frac{1}{(n-1)!} \frac{d^{n-1}}{dz^{n-1}} \big((z - z_0)^n f(z)\big)\Big|_{z=z_0}. \tag{12}$$

Formula (12) can be easily deduced from the Laurent expansion of f. From formula (10), we obtain a rule that is often useful:

Suppose f has a simple zero (so that $1/f$ has a simple pole) at z_0; then if g is holomorphic,

$$\operatorname{res}_{z_0} \frac{g(z)}{f(z)} = \frac{g(z_0)}{f'(z_0)}. \tag{13}$$

To take an easy example, let us compute the residues of the function

$$f(z) = \frac{z}{z^2+1}.$$

It has simple poles at i and $-i$. By (13), we then have

$$\operatorname{res}_i f = \frac{i}{2i} = \frac{1}{2},$$

and likewise for $-i$.

Concluding remark: Suppose $F\colon G \to G'$ is a biholomorphic mapping and f is holomorphic on G' except at certain isolated singularities. Then $f \circ F$ is holomorphic on G except at isolated singularities. If f has a singularity of a certain type at w_0 (removable, pole of order k, or essential), then $f \circ F$ has a singularity of the same type at z_0, where $F(z_0) = w_0$. But residues are not preserved by this transformation: using formula (1), it is not difficult to see that

$$\operatorname{res}_{F(z_0)} f = \operatorname{res}_{z_0}(f \circ F) \cdot F'. \tag{14}$$

This shows that the residue is actually to be associated with the differential form

$$f(w)dw. \tag{15}$$

This concept is also necessary in order to define residues at the point ∞. We refer the reader to [FL1] for a more detailed explanation.

Exercises

1. Let f be holomorphic on $\mathbb{C}$ with the exception of isolated singularities. Show that:
   i. If f is even, then $\operatorname{res}_{-z} f = -\operatorname{res}_z f$.
   ii. If f is odd, then $\operatorname{res}_{-z} f = \operatorname{res}_z f$.
   iii. If $f(z+\omega) \equiv f(z)$ for some $\omega \in \mathbb{C}$, then $\operatorname{res}_{z+\omega} f = \operatorname{res}_z f$.
   iv. If f is real on $\mathbb{R}$, then $\operatorname{res}_{\overline{z}} f = \overline{\operatorname{res}_z f}$.
2. Find the residues of the following functions at their singularities:
   $$\frac{1-\cos z}{z}, \quad \frac{z^2}{(1+z)^2}, \quad \frac{e^z}{(z-1)^3}, \quad \frac{1}{\cos z}, \quad \tan z.$$
3. Let G be a simply connected domain, and let f be holomorphic on G with the exception of a set of isolated singularities S. Show that f has a primitive (on $G \setminus S$) if and only if all residues of f vanish.

5. Residue calculus

The subject of this section is the evaluation of definite integrals over real intervals using the residue theorem (it should of course be possible to extend the integrand to a holomorphic function). A simple case occurs when one integrates a periodic function over a period-interval: by making a substitution, one can transform the integral directly into an integral over a closed path in $\mathbb{C}$. In other cases, one completes the interval of integration to a closed path in $\mathbb{C}$ by adding a "detour" γ to it. One can then apply the residue theorem, provided one has information about the integral over γ. We will exhibit a collection of methods and examples that can be adapted to situations other than those described – there exists no systematic theory.

We begin with periodic integrands. As an example, we investigate the integral

$$I = \int_0^{2\pi} \frac{dt}{a + \sin t}, \quad \text{where } a > 1.$$

Using $z = e^{it}$ and $\sin t = \frac{1}{2i}\left(z - \frac{1}{z}\right)$, we can write I as an integral over the unit circle:

$$I = \int_{|z|=1} \frac{1}{a + \frac{1}{2i}\left(z - \frac{1}{z}\right)} \frac{dz}{iz} = \int_{|z|=1} \frac{2\,dz}{z^2 + 2iaz - 1}.$$

The integrand is a rational function of z with simple poles at $z_1 = -i(a - \sqrt{a^2-1}) \in \mathbb{D}$ and $z_2 = -i(a + \sqrt{a^2-1}) \notin \overline{\mathbb{D}}$. The residue at z_1 is

$$\frac{2}{z_1 - z_2} = \frac{1}{i} \frac{1}{\sqrt{a^2-1}},$$

so that by the residue theorem

$$\int_0^{2\pi} \frac{dt}{a + \sin t} = \frac{2\pi}{\sqrt{a^2-1}}.$$

We extract the following rule:

Given a rational function $R(\cos t, \sin t)$ that stays finite for all $t \in \mathbb{R}$, the substitution $z = e^{it}$, i.e.

$$\cos t = \frac{1}{2}\left(z + \frac{1}{z}\right), \quad \sin t = \frac{1}{2i}\left(z - \frac{1}{z}\right),$$

transforms the integral $\int_0^{2\pi} R(\cos t, \sin t)\,dt$ into an integral of a rational function $\widetilde{R}(z)$ over $\partial\mathbb{D}$ that can be evaluated using the residue theorem.

Next, we consider integrals of the form $\int_{-\infty}^{\infty} f(x)\,dx$, assuming that $f(x)$ can be extended to a function $f(z)$ that is holomorphic in a neighbourhood of the closed upper half-plane $\overline{\mathbb{H}}$ with the possible exception of finitely many singularities, none of which lie on $\mathbb{R}$ – we will call such functions *admissible*.

The basic idea is this: Choose $r_1, r_2 > 0$ and paths γ in $\overline{\mathbb{H}}$ (which depend on r_1 and r_2) from r_2 to $-r_1$ such that the closed path $[-r_1, r_2] + \gamma$ encircles all singularities of f in $\mathbb{H}$ with winding number 1. The residue theorem then gives

$$\int_{-r_1}^{r_2} f(x)\,dx = 2\pi i \sum_{z\in\mathbb{H}} \operatorname{res}_z f - \int_\gamma f(z)\,dz.$$

If $f(z)$ tends to 0 sufficiently quickly as $z \to \infty$, then by appropriately choosing the paths γ, one can expect the integral on the right hand side to tend to 0 as $r_1, r_2 \to +\infty$; then

$$\int_{-\infty}^{\infty} f(x)\,dx = 2\pi i \sum_{z\in\mathbb{H}} \operatorname{res}_z f, \tag{1}$$

upon which the existence of the integral on the left hand side (as an improper Riemann integral) is proved as well.

Let us illustrate this with the example $f(z) = z^2/(1+z^4)$. Here it is clear that the integral of f over $\mathbb{R}$ exists. Choosing $r_1 = r_2 = r$ and $\gamma = \gamma_r$ as the semicircle from r to $-r$, we have

$$\left|\int_{\gamma_r} f(z)\,dz\right| \leqslant \pi r \max\bigl\{|f(z)|\colon |z| = r,\ z \in \overline{\mathbb{H}}\bigr\}.$$

Since f has a double zero at ∞, $|f(z)| \leqslant cr^{-2}$ on $\operatorname{Tr}\gamma_r$, where c is a constant; it follows that $\int_{\gamma_r} f(z)\,dz \to 0$ as $r \to \infty$. We may thus apply (1), and it remains only to compute the residues. The (simple) poles of f in $\mathbb{H}$ are $z_1 = (1+i)/\sqrt{2}$ and $z_2 = (-1+i)/\sqrt{2}$, and the denominator of f factorizes as

$$z^4 + 1 = (z - z_1)(z - z_2)(z + z_1)(z + z_2).$$

Thus,

$$\operatorname{res}_{z_1} f = \frac{z_1^2}{2z_1(z_1 - z_2)(z_1 + z_2)}, \qquad \operatorname{res}_{z_2} f = \frac{z_2^2}{2z_2(z_2 - z_1)(z_2 + z_1)}.$$

Their sum is $1/2(z_1 + z_2) = 1/(2\sqrt{2}i)$, so (1) yields

$$\int_{-\infty}^{\infty} \frac{x^2}{1+x^4}\,dx = \frac{\pi}{\sqrt{2}}.$$

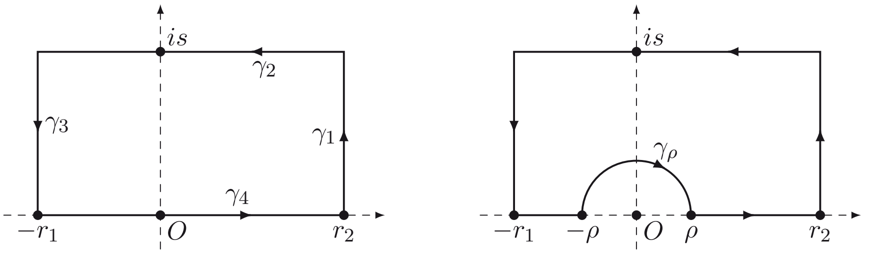


Figure 8. For the proof of (2) and (6)

Let us extract the conditions under which this "semicircle method" works:

Suppose $f(z)$ *is an admissible function,* $\int_{-\infty}^{\infty} f(x)\,dx$ *exists, and*

$$\lim_{z\to\infty} z f(z) = 0.$$

Then (1) *holds.*

For the integral

$$\int_{-\infty}^{\infty} \frac{x \sin x}{a^2 + x^2}\, dx, \quad a > 0,$$

this rule does not apply, since $|\sin it| = |\sinh t|$ grows exponentially as $t \to \infty$. Moreover the existence of the integral is not known a priori. The situation improves if we write

$$\int_{-\infty}^{\infty} \frac{x \sin x}{a^2 + x^2}\, dx = \operatorname{Im} \int_{-\infty}^{\infty} \frac{x\, e^{ix}\, dx}{a^2 + x^2},$$

because the factor e^{iz} on the right hand side becomes small for large $\operatorname{Im} z$. But the semicircle method, which is based on the standard estimate, still does not help; one needs more accurate estimates and better paths of integration.

More generally, we treat integrals of the form

$$\int_{-\infty}^{\infty} g(z)\, e^{iz}\, dz$$

and to begin with assume only that g is an admissible function. The positive numbers r_1, r_2, and s are chosen such that the rectangle with corners $-r_1$, r_2, $r_2 + is$, and $-r_1 + is$ contains all the singularities of g in $\mathbb{H}$. For our auxiliary path γ, we take the part $\gamma_1+\gamma_2+\gamma_3$ of the boundary of this rectangle (see Fig. 8) and estimate the integrals over the paths γ_j. For γ_2, the standard estimate is sufficient:

$$\left|\int_{\gamma_2} g(z)\,e^{iz}\,dz\right| \leqslant \max_{z\in\operatorname{Tr}\gamma_2} |g(z)|\cdot e^{-s}(r_1+r_2).$$

For γ_1, we have

$$\begin{aligned}\left|\int_{\gamma_1} g(z)\,e^{iz}\,dz\right| &= \left|\int_0^s g(r_2+it)\,e^{i(r_2+it)}i\,dt\right| \\ &\leqslant \max_{z\in\operatorname{Tr}\gamma_1}|g(z)|\cdot\int_0^s e^{-t}\,dt \leqslant \max_{z\in\operatorname{Tr}\gamma_1}|g(z)|,\end{aligned}$$

and the same goes for γ_3. Assuming $\lim_{z\to\infty} g(z)=0$ (this applies to the above example), then given an $\varepsilon>0$, one can choose r_1, r_2, and s large enough so that $|g(z)|<\varepsilon$ on γ, and then, possibly by further enlarging s, ensure that $e^{-s}(r_1+r_2)\leqslant 1$. Then $|\int_\gamma g(z)\,e^{iz}\,dz|<3\varepsilon$, which proves the following rule:

Suppose $g(z)$ is an admissible function and $g(z)\to 0$ as $z\to\infty$ in $\overline{\mathbb{H}}$. Then

$$\int_{-\infty}^{\infty} g(z)\,e^{iz}\,dz = 2\pi i\sum_{z\in\mathbb{H}}\operatorname{res}_z\big(g(\zeta)\,e^{i\zeta}\big). \tag{2}$$

Here the integral may be an improper integral.

These conditions are clearly satisfied if g is a rational function without poles on $\mathbb{R}$ that vanishes at ∞. Let us return to our example: The only singularity of $g(z)=z\,e^{iz}/(a^2+z^2)$ in $\mathbb{H}$ is the simple pole ia with residue $e^{-a}/2$, so by (2),

$$\int_{-\infty}^{\infty}\frac{x\sin x}{a^2+x^2}\,dx = \operatorname{Im}\int_{-\infty}^{\infty}\frac{x\,e^{ix}\,dx}{a^2+x^2} = \operatorname{Im}2\pi i\cdot e^{-a}/2 = \pi\,e^{-a}.$$

As a consequence of (2), we note that if $g(z)$ is real-valued on $\mathbb{R}$, then

$$\begin{aligned}\int_{-\infty}^{\infty} g(x)\cos x\,dx &= \operatorname{Re}\left(2\pi i\sum_{z\in\mathbb{H}}\operatorname{res}_z\big(g(\zeta)\,e^{i\zeta}\big)\right)\\ \int_{-\infty}^{\infty} g(x)\sin x\,dx &= \operatorname{Im}\left(2\pi i\sum_{z\in\mathbb{H}}\operatorname{res}_z\big(g(\zeta)\,e^{i\zeta}\big)\right).\end{aligned} \tag{3}$$

Should one need these integrals when $g(z)$ is not real-valued, one may, for example, put

$$\int_{-\infty}^{\infty} g(x)\cos x\,dx = \frac{1}{2}\int_{-\infty}^{\infty} g(z)\,e^{iz}\,dz + \frac{1}{2}\int_{-\infty}^{\infty} g(z)\,e^{-iz}\,dz.$$

Our rule can be applied to the first integral on the right hand side; in order to evaluate the second integral, one must pass to the lower half-plane (e^{-iz} becomes small as $\operatorname{Im} z \to -\infty$!). This of course requires g to be holomorphic on the entire complex plane except for finitely many singularities (none of which lie on $\mathbb{R}$) and to satisfy $\lim_{z\to\infty} g(z) = 0$. One thus obtains

$$\int_{-\infty}^{\infty} g(x)\cos x\,dx = \pi i \sum_{\operatorname{Im} z>0} \operatorname{res}_z\big(g(\zeta)\,e^{i\zeta}\big) - \pi i \sum_{\operatorname{Im} z<0} \operatorname{res}_z\big(g(\zeta)\,e^{-i\zeta}\big). \tag{4}$$

The minus sign in front of the second sum is there because the boundary of the rectangle we have to use in the lower half-plane encircles the singularities there with winding number -1. Likewise, we have

$$\int_{-\infty}^{\infty} g(x)\sin x\,dx = \pi \sum_{\operatorname{Im} z>0} \operatorname{res}_z\big(g(\zeta)\,e^{i\zeta}\big) + \pi \sum_{\operatorname{Im} z<0} \operatorname{res}_z\big(g(\zeta)\,e^{-i\zeta}\big). \tag{5}$$

The techniques developed so far do not allow us to compute the integral

$$\int_{-\infty}^{\infty} \frac{\sin x}{x}\,dx.$$

Seen as a real function, the integrand is harmless at $x = 0$, but the complex integrand e^{iz}/z has a pole of order 1 at 0. In such cases, there is a trick that will lead us to our goal – let us immediately formulate it in general terms. Let $g(z)$ be holomorphic in a neighbourhood of $\overline{\mathbb{H}}$ except for finitely many singularities in $\mathbb{H}$ and a simple pole at 0, and suppose $\lim_{z\to\infty} g(z) = 0$. We then choose $\rho > 0$ such that g has no singularities in $\overline{D_\rho(0)} \setminus \{0\}$. We proceed as in the proof of (2) but replace the subinterval $[-\rho, \rho]$ of $[-r_1, r_2]$ by the semicircle $\gamma_\rho(t) = \rho\, e^{i(\pi - t)}$, where $0 \leqslant t \leqslant \pi$ (cf. Fig. 8). As $r_1, r_2, s \to \infty$, we obtain

$$\int_{-\infty}^{-\rho} g(z)\,e^{iz}\,dz + \int_{\gamma_\rho} g(z)\,e^{iz}\,dz + \int_{\rho}^{\infty} g(z)\,e^{iz}\,dz = 2\pi i \sum_{\operatorname{Im} z>0} \operatorname{res}_z\big(g(\zeta)\,e^{i\zeta}\big).$$

But writing

$$g(z)\,e^{iz} = \frac{c}{z} + h(z),$$

in $\overline{D_\rho(0)}$, we see that

$$\int_{\gamma_\rho} g(z)\,e^{iz}\,dz = c\int_{\gamma_\rho}\frac{dz}{z} + \int_{\gamma_\rho} h(z)\,dz = c\int_0^\pi(-i)\,dt + \int_{\gamma_\rho} h(z)\,dz.$$

The first integral on the right hand side is

$$-\pi i\,\operatorname{res}_0\big(g(\zeta)\,e^{i\zeta}\big) = -\pi i\,\operatorname{res}_0 g,$$

and the last integral on the right hand side tends to zero as $\rho \to 0$. It follows that

$$\lim_{\rho\to 0}\left(\int_{-\infty}^{-\rho} g(z)\,e^{iz}\,dz + \int_\rho^\infty g(z)\,e^{iz}\,dz\right) = 2\pi i\sum_{\operatorname{Im} z>0}\operatorname{res}_z\big(g(\zeta)\,e^{i\zeta}\big) + \pi i\,\operatorname{res}_0 g. \quad (6)$$

The left hand side of this formula is also called the *Cauchy principal value* of the improper integral $\int_{-\infty}^{\infty} g(z)\,e^{iz}\,dz$. It is not difficult to extend (6) to the case when g has finitely many simple poles on $\mathbb{R}$.

In our example, we now have

$$\int_{-\infty}^{\infty}\frac{\sin x}{x}\,dx = \lim_{\rho\to 0}\operatorname{Im}\left(\int_{-\infty}^{-\rho}\frac{e^{iz}}{z}\,dz + \int_\rho^\infty\frac{e^{iz}}{z}\,dz\right) = \operatorname{Im}\left(\pi i\,\operatorname{res}_0\left(\frac{1}{\zeta}\right)\right) = \pi.$$

Note that $(\sin x)/x$ is not integrable over $\mathbb{R}$; the integral must be understood as an improper integral.

To conclude, we give a method for computing integrals of the types

$$\int_0^\infty R(x)\,dx, \qquad \int_0^\infty R(x)\log x\,dx, \qquad \int_0^\infty x^\alpha R(x)\,dx,$$

where $R(x)$ is a rational function without poles on $\mathbb{R}_{\geqslant 0}$ and with a zero of order $\geqslant 2$ at ∞ and where $0 < \alpha < 1$.

Here a new idea appears: We cut the complex plane along the *positive* real axis, choose the branch of the logarithm on $G = \mathbb{C}\setminus\mathbb{R}_{\geqslant 0}$ that satisfies $0 < \operatorname{Im}\log z < 2\pi$, and compare integrals over paths above and below the cut. Concretely, say with $\rho > 0$ small and $r > \rho$ large:

$$\int_\rho^r R(x+i\varepsilon)\log(x+i\varepsilon)\,dx \to \int_\rho^r R(x)\log x\,dx \text{ for } \varepsilon\downarrow 0$$

$$\int_\rho^r R(x-i\varepsilon)\log(x-i\varepsilon)\,dx \to \int_\rho^r R(x)(\log x + 2\pi i)\,dx \text{ for } \varepsilon\downarrow 0.$$

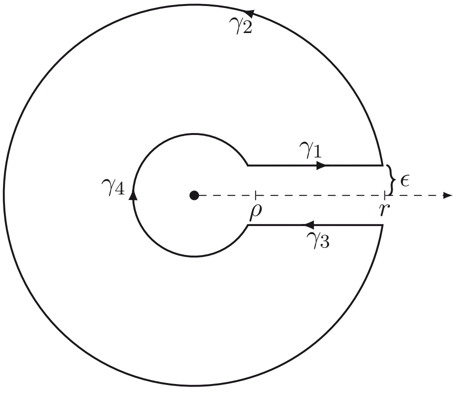


Figure 9. For the proof of (7), (8), and (11)

The difference of the integrals to the left thus tends to $-2\pi i \int_\rho^r R(x)\,dx$. In order to apply the residue theorem, we construct a closed path as shown in Fig. 9 that has $\gamma_1 = [\rho + i\varepsilon, r + i\varepsilon]$ and $\gamma_3 = [r - i\varepsilon, \rho - i\varepsilon]$ as subpaths (here γ_2 and γ_4 are circular arcs about 0, and we assume $0 < \varepsilon < \rho$). If ρ is sufficiently small and r is sufficiently large, then $\gamma = \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4$ encircles all poles of $R(z)\log z$, i.e. the poles of $R(z)$, and we have

$$\int_\gamma R(z)\log z\,dz = 2\pi i \sum_{z\in G} \operatorname{res}_z(R(\zeta)\log\zeta).$$

On the other hand, the above considerations yield

$$\lim_{\substack{\rho\to 0\\ r\to\infty}} \lim_{\varepsilon\to 0} \int_{\gamma_1+\gamma_3} R(z)\log z\,dz = -2\pi i \int_0^\infty R(z)\,dz.$$

In order to obtain a result, we must determine the limits of the integrals over γ_2 and γ_4. The function $R(z)$ is bounded near 0, and $|R(z)| \leqslant c/|z|^2$ in a neighbourhood of ∞. Since $\lim_{\rho\to 0}(\rho\log\rho) = 0$ and $\lim_{r\to\infty}(r^{-1}\log r) = 0$, the standard estimate yields

$$\lim_{\rho\to 0}\int_{\gamma_4} R(z)\log z\,dz = 0, \qquad \lim_{r\to\infty}\int_{\gamma_2} R(z)\log z\,dz = 0$$

uniformly in ε. To summarize, we have the following rule:

Let $R(z)$ be a rational function with no poles on $\mathbb{R}_{\geqslant 0}$ and with at least a double zero at ∞. Then

$$\int_0^\infty R(x)\,dx = -\sum_{z\neq 0} \operatorname{res}_z(R(\zeta)\log\zeta). \tag{7}$$

For example,

$$\int_0^\infty \frac{dx}{x^3+1} = \frac{2\pi}{3\sqrt{3}},$$

since $R(z) = (z^3+1)^{-1}$ has simple poles at -1, $z_1 = (1+i\sqrt{3})/2$, and $\overline{z}_1$ with residues $1/3$, $-(1+i\sqrt{3})/6$, and $-(1-i\sqrt{3})/6$, respectively. With $\log(-1) = \pi i$, $\log z_1 = \pi i/3$, and $\log \overline{z}_1 = 5\pi i/3$, formula (7) yields the result.

In order to find

$$\int_0^\infty R(x) \log x \, dx,$$

we integrate the function $R(z)(\log z)^2$ over our path γ, choosing the same logarithm as above. We then have

$$\lim_{\varepsilon \to 0} \int_{\gamma_1} R(z) \log^2 z \, dz = \int_\rho^r R(x) \log^2 x \, dx$$

$$\lim_{\varepsilon \to 0} \int_{\gamma_3} R(z) \log^2 z \, dz = -\int_\rho^r R(x)(\log x + 2\pi i)^2 \, dx,$$

and as before,

$$\lim_{\rho \to 0} \int_{\gamma_4} R(z) \log^2 z \, dz = 0 = \lim_{r \to \infty} \int_{\gamma_2} R(z) \log^2 z \, dz.$$

In the limit, the relation

$$\int_\gamma R(z) \log^2 z \, dz = 2\pi i \sum_{z \in G} \operatorname{res}_z(R(\zeta)(\log \zeta)^2)$$

thus yields

$$4\pi i \int_0^\infty R(x) \log x \, dx = 4\pi^2 \int_0^\infty R(x) \, dx - 2\pi i \sum_{z \neq 0} \operatorname{res}_z(R(\zeta) \log^2 \zeta). \tag{8}$$

We already have methods for computing the integral on the right hand side; it is superfluous if $R(x)$ is real-valued on $\mathbb{R}$, in which case, passing to the imaginary part, (8) implies that

$$\int_0^\infty R(x) \log x \, dx = -\frac{1}{2} \operatorname{Re} \sum_{z \neq 0} \operatorname{res}_z(R(\zeta) \log^2 \zeta). \tag{9}$$

As an example, let us show that

$$\int_0^\infty \frac{\log x\,dx}{x^2+a^2} = \frac{\pi \log a}{2a} \quad \text{for } a > 0.$$

The function $R(z) = (z^2+a^2)^{-1}$ has simple poles at ia and $-ia$ with residues $1/2ia$ and $-1/2ia$, respectively. The claim now follows from (9) by noting that $\log ia = \log a + \pi i/2$ and $\log(-ia) = \log a + 3\pi i/2$.

To conclude, we consider integrals of the form

$$\int_0^\infty x^\alpha R(x)\,dx, \text{ where } 0 < \alpha < 1.$$

Here we may allow that $R(z)$ has a simple pole at the origin without affecting the existence of the integral. We set $z^\alpha = \exp(\alpha \log z)$ on $G = \mathbb{C} \setminus \mathbb{R}_{\geqslant 0}$ with the same choice of $\log z$ as above and then integrate $z^\alpha R(z)$ over our path γ, which encircles all poles of R except possibly 0. Since

$$\lim_{\varepsilon \downarrow 0}(x+i\varepsilon)^\alpha = x^\alpha \quad \text{and} \quad \lim_{\varepsilon \downarrow 0}(x-i\varepsilon)^\alpha = e^{2\pi i\,\alpha} x^\alpha,$$

we have

$$\lim_{\varepsilon \to 0} \int_{\gamma_3} z^\alpha R(z)\,dz = -e^{2\pi i\,\alpha} \lim_{\varepsilon \to 0} \int_{\gamma_1} z^\alpha R(z)\,dz. \tag{10}$$

The integrals over γ_2 and γ_4 vanish as $r \to \infty$ and $\rho \to 0$. This is because $|R(z)| \leqslant c/|z|$ near 0 and $|R(z)| \leqslant c|z|^{-2}$ near ∞, so that the standard estimate yields

$$\left| \int_{\gamma_4} z^\alpha R(z)\,dz \right| \leqslant 2\pi\rho \cdot \rho^\alpha \cdot c\rho^{-1} \quad \text{and} \quad \left| \int_{\gamma_2} z^\alpha R(z)\,dz \right| \leqslant 2\pi r \cdot r^\alpha \cdot cr^{-2}.$$

Passing to the limit as $\varepsilon \to 0$, $\rho \to 0$, and $r \to \infty$, the residue theorem, together with (10), thus yields

$$\int_0^\infty x^\alpha R(x)\,dx = \frac{2\pi i}{1-e^{2\pi i\,\alpha}} \sum_{z \neq 0} \operatorname{res}_z(R(\zeta)\zeta^\alpha). \tag{11}$$

For example, -1 is the only pole in G of the function

$$R(z) = \frac{1}{z(z+1)},$$

its residue is -1, and $(-1)^\alpha = e^{\pi i \alpha}$. Therefore, (11) yields

$$\int_0^\infty \frac{x^{\alpha-1}}{x+1}\,dx = \frac{-2\pi i\, e^{\pi i \alpha}}{1-e^{2\pi i\,\alpha}} = \frac{\pi}{\sin \pi\alpha} \quad \text{for } 0 < \alpha < 1.$$

Exercises

1. Prove that

$$\int_{-\infty}^{\infty} \frac{dx}{\cosh x} = \pi.$$

Hint: Integrate over the boundary of the rectangle whose corners are $\pm r$ and $\pm r + i\pi$.

2. Find

$$\int_0^{\infty} \frac{x^\alpha}{1+x^n}\, dx \quad \text{for } n > \alpha + 1 > 0 \text{ and } n \geqslant 2.$$

Hint: Integrate over the boundary of the sector with corners 0, r, and $r\, e^{2\pi i/n}$.

3. Compute the following:

a) $\displaystyle\int_0^{\pi} \frac{\sin^2 x}{a+\cos x}\, dx \quad \text{for } a > 1$ b) $\displaystyle\int_0^{2\pi} \frac{dt}{1-2a\cos t + a^2}, \quad a \in \mathbb{C} \text{ and } |a| \neq 1.$

4. Compute the following:

a) $\displaystyle\int_0^{\infty} \frac{dx}{(x^2+1)(x^2+4)}$ b) $\displaystyle\int_{-\infty}^{\infty} \frac{dx}{(a^2+b^2x^2)^n} \quad (a, b > 0,\ n \geqslant 1)$

c) $\displaystyle\int_0^{\infty} \frac{\sqrt{x}\, dx}{x^2+16}$ first by reducing to (1) using the substitution $x = u^2$, then by using (11).

5. a) Compute

$$\int_{-\infty}^{\infty} \frac{x\, e^{-\pi i x/2}\, dx}{x^2 - 2x + 5}.$$

b) Compute

$$I(a) = \int_{-\infty}^{\infty} \frac{e^{iax}\, dx}{x^2+1} \quad \text{for } a \in \mathbb{R}.$$

Is $I(a)$ differentiable?

6. Show that

$$\int_{2i-\infty}^{2i+\infty} \frac{z \sin az}{z^2+1}\, dz = \pi \cosh a$$

(the integral is to be taken over the line parallel to the real axis and passing through $2i$).
Hint: It is easier to adapt the proof of (5) than to transform the above into an integral over $\mathbb{R}$.

7. Compute the Cauchy principal value of

$$\int_{-\infty}^{\infty} \frac{1-e^{itx}}{x^2}\, dx$$

for $t > 0$ and thereby

$$\int_0^{\infty} \frac{1-\cos tx}{x^2}\, dx \quad \text{and} \quad \int_0^{\infty} \frac{\sin^2 x}{x^2}\, dx.$$

8. Compute

a) $$\int_0^\infty \frac{x^\alpha\, dx}{(x+t)(x+2t)} \quad \text{for } 0<\alpha<1 \text{ and } t>0$$

b) $$\int_0^\infty \frac{x^\alpha}{1+x^{1/3}}\, dx \quad \text{for } -1<\alpha<\frac{-2}{3}.$$

9. Show that for a rational function $R(x)$ that satisfies the conditions needed for (11) and for $0<\alpha<1$,

$$\int_0^\infty R(x)x^\alpha \log x\, dx = \frac{2\pi i}{1-e^{2\pi i\,\alpha}} \sum_{z\neq 0} \operatorname{res}_z(R(\zeta)\zeta^\alpha \log\zeta) + \frac{\pi^2}{\sin^2(\pi\alpha)} \sum_{z\neq 0} \operatorname{res}_z(R(\zeta)\zeta^\alpha).$$

For z^α, choose $0<\arg z<2\pi$.

10. Using Ex. 9, compute

$$\int_0^\infty \frac{\log x}{(x^2+1)\sqrt{x}}\, dx.$$

6. Counting zeros

By the identity theorem, the fibres $f^{-1}(w)$ of a nonconstant holomorphic function are discrete subsets of the domain of f. The residue theorem provides information about the number of points in such a fibre. This is what we will carry out in this section.

Let $f\colon G\to\mathbb{C}$ be holomorphic on a domain G, and suppose f takes the value $w_0=f(z_0)$ at the point $z_0\in G$ with multiplicity $k>0$. Then

$$f(z) = w_0 + (z-z_0)^k g(z), \tag{1}$$

g is holomorphic at z, and $g(z_0)\neq 0$. Thus,

$$\frac{f'(z)}{f(z)-w_0} = \frac{k}{z-z_0} + \frac{g'(z)}{g(z)}, \tag{2}$$

and we have

$$\operatorname{res}_{z_0} \frac{f'(z)}{f(z)-w_0} = k. \tag{3}$$

For meromorphic functions that have a pole of order k at z_0, we have for arbitrary $w\in\mathbb{C}$

$$f(z)-w = \frac{g(z)}{(z-z_0)^k} \tag{4}$$

with a holomorphic function g such that $g(z_0)\neq 0$. Therefore,

$$\frac{f'(z)}{f(z)-w} = \frac{-k}{z-z_0} + \frac{g'}{g}, \tag{5}$$

i.e.

$$\operatorname{res}_{z_0} \frac{f'(z)}{f(z)-w} = -k. \tag{6}$$

The residue theorem now yields:

Proposition 6.1. *Let f be meromorphic in a domain G with poles $b_1, b_2, \ldots$ of multiplicities $k(b_\mu)$, where $\mu = 1, 2, \ldots$. Let $w \in \mathbb{C}$, and suppose f takes on the value w at the points $a_1, a_2, \ldots$ with multiplicities $k(a_\mu)$, again where $\mu = 1, 2, \ldots$. Then for every null-homologous cycle Γ in G that does not pass through any of these points,*

$$\frac{1}{2\pi i}\int_\Gamma \frac{f'(z)}{f(z)-w}\,dz = \sum_\mu n(\Gamma, a_\mu)k(a_\mu) - \sum_\mu n(\Gamma, b_\mu)k(b_\mu).$$

By (3) and (6), the right hand side is the "total residue" of $f'(z)/(f(z)-w)$ in the domain $G_0 = \{z\colon n(\Gamma, z) \neq 0\}$. The sum is finite, because G_0 is relatively compact in G.

The situation is especially clear in the case of a domain with positive boundary:

Corollary 6.2. *Let G be a domain with positive boundary, and suppose f is meromorphic in a neighbourhood of $\overline{G}$ and does not assume w and ∞ on ∂G. Then*

$$\frac{1}{2\pi i}\int_{\partial G} \frac{f'(z)}{f(z)-w}\,dz = N(w) - N(\infty),$$

where $N(w)$ and $N(\infty)$ are the number of points in G at which f takes on the value w and the number of poles of f in G (counting multiplicity), respectively.

Let us again consider a holomorphic function $f\colon G \to \mathbb{C}$ that takes on the value w_0 at z_0 with multiplicity k. Then there exists a neighbourhood $U = U_\delta(z_0)$ such that z_0 is the only point of $\overline{U_\delta(z_0)}$ where f has the value w_0 and moreover is the only point at which f' (possibly) vanishes. We thus have

$$N(w_0) = k = \frac{1}{2\pi i}\int_{\partial U_\delta(z_0)} \frac{f'(z)}{f(z)-w_0}\,dz$$

in U. For values w that are sufficiently close to w_0, say $w \in V$ for some neighbourhood of w_0, the integral

$$N(w) = \frac{1}{2\pi i}\int_{\partial U_\delta(z_0)} \frac{f'(z)}{f(z)-w}\,dz$$

exists and is a continuous function of w. By Cor. 6.2, it is integer-valued and thus constant, and therefore equal to k. This gives

Proposition 6.3. *If a holomorphic function f takes on the value w_0 at z_0 with multiplicity k, then there exist neighbourhoods U and V of z_0 and w_0, respectively, with the following property: Every value $w \in V$ is assumed by f at exactly k distinct points in U with the exception of the value w_0, which is assumed by f exactly at z_0.*

That the equation $f(z) = w$ has exactly k distinct solutions in V follows from the fact that $f'(z) \neq 0$ for $z \neq z_0$. Setting

$$U' = U \cap f^{-1}(V),$$

$f\colon U' \to V$ is surjective and a k-fold branched covering of V with exactly one branch point z_0. If $k = 1$, then f is biholomorphic. We already derived this result using a different argument in III.5.

As a further easy corollary of Prop. 6.1, we have:

Proposition 6.4 (Rouché's theorem)**.** *Let f and g be holomorphic in a neighbourhood of $\overline{G}$, where G is a domain with positive boundary, and suppose*

$$|f(z) - g(z)| < |f(z)|$$

on ∂G. Then f and g have the same number of zeros (counting multiplicity) in G.

Proof: For $0 \leqslant \lambda \leqslant 1$,

$$|f(z) + \lambda(g(z) - f(z))| > 0$$

on ∂G. By Cor. 6.2, the number N_λ of zeros of $f + \lambda(g - f)$ in G is given by

$$N_\lambda = \frac{1}{2\pi i} \int\limits_{\partial G} \frac{f'(z) + \lambda(g'(z) - f'(z))}{f(z) + \lambda(g(z) - f(z))}\, dz$$

and is thus continuous in λ and, since it is an integer, constant. Putting $\lambda = 0$ and $\lambda = 1$, one obtains the number of zeros of f and g, respectively. □

The following application of Cor. 6.2 is important in the theory of holomorphic mappings:

Proposition 6.5. *Let $f_\nu\colon G \to \mathbb{C}$ be a locally uniformly convergent sequence of injective holomorphic functions. Then the limit function f is either injective or constant.*

Proof: Let w be a value that is assumed by f at two different points z_0 and z_1. If f is nonconstant, then there exist disjoint compact disks $\overline{D_r(z_0)}$ and $\overline{D_r(z_1)}$ in G such that f takes on the value w in these disks only at the points z_0 and z_1. By Cor. 6.2, we have

$$N_0 = \frac{1}{2\pi i} \int\limits_{\partial D_r(z_0)} \frac{f'(z)}{f(z) - w}\, dz \geqslant 1$$

$$N_1 = \frac{1}{2\pi i} \int\limits_{\partial D_r(z_1)} \frac{f'(z)}{f(z) - w}\, dz \geqslant 1$$

in these disks. On the other hand,

$$N_0 = \lim_{\nu \to \infty} \frac{1}{2\pi i} \int\limits_{\partial D_r(z_0)} \frac{f_\nu'(z)}{f_\nu(z) - w}\, dz,$$

and likewise for N_1. But the integrals on the right hand side give the number of points at which f_ν takes on the value w in the disks $D_r(z_0)$ and $D_r(z_1)$ and can therefore converge to a nonzero value only in one case, i.e. *either* in the case of $\partial D_r(z_0)$ *or* in the case of $\partial D_r(z_1)$. This contradiction shows that f must be constant: $f(z) \equiv w$. □

To conclude, we use the tools developed here to supply a proposition that will be useful in the next section.

Proposition 6.6. *Suppose f is a meromorphic function on the disk $D_R(0)$ that has exactly k zeros and k poles (counting multiplicity), and suppose that all of these points lie in the disk $D_r(0)$, where $r < R$. Then there is a holomorphic logarithm of f in the annulus $K(r, R)$.*

Proof: Let γ be a closed path of integration in $K(r, R)$. Then γ is null-homologous in $D_R(0)$, and by Prop. 6.1,

$$\frac{1}{2\pi i} \int\limits_{\gamma} \frac{f'(z)}{f(z)}\, dz = \sum n(\gamma, a_\mu) k(a_\mu) - \sum n(\gamma, b_\mu) k(b_\mu), \tag{7}$$

where a_μ and b_μ are the zeros and poles of f with their corresponding multiplicities. But

$$n(\gamma, a_\mu) = n(\gamma, b_\mu) \stackrel{\text{def}}{=} n_0$$

independently of μ – the right hand side of (7) is thus equal to $n_0(k - k) = 0$. The function f'/f thus has a primitive g, and as before

$$e^g = cf, \quad c \in \mathbb{C}^*,$$

so that $g - a$, where $e^a = c$, is a holomorphic logarithm of f. □

Finally, let us record the formula

$$\log f(z) = \int_{z_0}^{z} \frac{f'(\zeta)}{f(\zeta)}\, d\zeta + \text{const}, \tag{8}$$

valid in the above situation. The integral is extended over an arbitrary path of integration in $K(r, R)$.

Exercises

1. Suppose f is holomorphic in the domain G and Γ is a null-homologous cycle in G that does not pass through any zeros of f. Let a_μ be the zeros of f with multiplicities k_μ. Show that

$$\frac{1}{2\pi i} \int_\Gamma z\, \frac{f'(z)}{f(z)}\, dz = \sum_\mu n(\Gamma, a_\mu)\, k_\mu\, a_\mu .$$

2. Replace the zeros in Ex. 1 with the points at which f takes on a value w and derive an analogous formula. If, in particular, f is injective, give an integral formula for the inverse function f^{-1}.
3. Let $\lambda > 1$. Show that the equation $e^{-z} + z = \lambda$ has exactly one solution in the half-plane $\operatorname{Re} z > 0$. Show that it is real.
4. Find the number of zeros of f in the domains G:

$$\begin{aligned} f(z) &= 2z^4 - 5z + 2, & G &= \{z\colon |z| > 1\} \\ f(z) &= z^5 + iz^3 - 4z + i, & G &= \{z\colon 1 < |z| < 2\}. \end{aligned}$$

5. Prove the fundamental theorem of algebra using Rouché's theorem.

7. The Weierstrass preparation theorem

The results and methods of Sections 3 and 6 also provide insight into the distribution of zeros of holomorphic functions of several variables. Up to now, we only know that the zeros of such functions are never isolated – see II.7.

We start by proving "parameter dependent" versions of Thm. 3.1 and Prop. 6.1 and 6.6.

Proposition 7.1 (Laurent decomposition). *Let $U \subset \mathbb{C}^n$ be an open set, $K = K(r, R) \subset \mathbb{C}$ an annulus, and $f\colon U \times K \to \mathbb{C}$ a holomorphic function. Then there exist unique holomorphic functions f_0 on $U \times D_R(0)$ and f_∞ on $U \times (\mathbb{C} \setminus \overline{D_r(0)})$ such that*

$$f = f_0 + f_\infty \quad \text{and} \quad \lim_{w \to \infty} f_\infty(\mathbf{z}, w) = 0.$$

Here we have denoted the coordinates in $U \subset \mathbb{C}^n$ by $\mathbf{z}$ and the coordinate in $K \subset \mathbb{C}$ by w. For $n = 0$, the proposition reduces to Thm. 3.1.

To prove existence, we use formula (6) from IV.3; in our case, it is

$$f(\mathbf{z}, w) = \frac{1}{2\pi i} \int\limits_{|\zeta|=R'} \frac{f(\mathbf{z}, \zeta)}{\zeta - w} \, d\zeta - \frac{1}{2\pi i} \int\limits_{|\zeta|=r'} \frac{f(\mathbf{z}, \zeta)}{\zeta - w} \, d\zeta, \tag{1}$$

where $\mathbf{z} \in U$ and $r' < |w| < R'$. The two integrals above are simultaneously holomorphic in $\mathbf{z}$ and w for $|w| < R'$ and $|w| > r'$ and yield the desired decomposition as before. The proof of uniqueness follows exactly as in the proof of Thm. 3.1. □

One also obtains the following proposition upon analysing a proof, namely the proof of Prop. 6.1:

Proposition 7.2. *Let $U \subset \mathbb{C}^n$ be a domain, let $D_R(0) \subset \mathbb{C}$ be the disk of radius R, and let $f\colon U \times D_R(0) \to \mathbb{C}$ be a holomorphic function of the $n+1$ variables $\mathbf{z} \in U$ and $w \in D_R(0)$ that is nonzero for $r \leqslant |w| < R$ and $\mathbf{z} \in U$. Then the number of zeros (counting multiplicity) of the functions $w \mapsto f(\mathbf{z}, w)$ is independent of $\mathbf{z} \in U$.*

Indeed, the number of zeros of $w \mapsto f(\mathbf{z}, w)$ is given by

$$N(\mathbf{z}) = \frac{1}{2\pi i} \int\limits_{\partial D_r} \frac{f_w(\mathbf{z}, w)}{f(\mathbf{z}, w)} \, dw, \tag{2}$$

and this function is continuous in $\mathbf{z}$ and integer-valued, hence constant. □

In the situation of the above theorem, we call the number of zeros $N(\mathbf{z})$ the *order* of the function f with respect to w.

The "parameter dependent" version of Prop. 6.6 reads as follows:

Proposition 7.3. *Let $f\colon U \times D_R(0) \to \mathbb{C}$ be a holomorphic function of order k with respect to w, and suppose f is nonzero for $r \leqslant |w| < R$. Then there exists a nonzero holomorphic function c on U and a holomorphic function h on $U \times K(r, R)$ such that*

$$w^k \cdot e^{h(\mathbf{z}, w)} = c(\mathbf{z}) f(\mathbf{z}, w).$$

Proof: For every $\mathbf{z} \in U$, the function

$$w \mapsto g(\mathbf{z}, w) = \frac{f(\mathbf{z}, w)}{w^k} \tag{3}$$

satisfies the conditions of Prop. 6.6 with respect to w. Therefore,

$$\frac{g_w(\mathbf{z}, w)}{g(\mathbf{z}, w)}$$

has a primitive with respect to w, namely

$$h(\mathbf{z}, w) = \int_{w_0}^{w} \frac{g_\zeta(\mathbf{z}, \zeta)}{g(\mathbf{z}, \zeta)}\, d\zeta, \tag{4}$$

on $U \times K(r, R)$. The integral (4) depends holomorphically on $\mathbf{z}$ and w. Differentiating with respect to w yields

$$\frac{\partial}{\partial w} \frac{e^{h(\mathbf{z},w)}}{g(\mathbf{z}, w)} \equiv 0, \tag{5}$$

so that for every $\mathbf{z} \in U$, there exists a constant $c(\mathbf{z}) \neq 0$ such that

$$e^{h(\mathbf{z},w)} = c(\mathbf{z}) g(\mathbf{z}, w). \tag{6}$$

The identity (6) implies that c depends holomorphically on $\mathbf{z}$. Substituting (3) into (6) yields the claim. □

The main result of this section is

Theorem 7.4 (The Weierstrass preparation theorem)**.** *Let f be a holomorphic function on $U \times D_R(0)$ whose order with respect to w is k, and let $U \subset \mathbb{C}^n$ be a domain. Then there exist a nonzero holomorphic function e on $U \times D_R(0)$ and a polynomial $\omega \in \mathcal{O}(U)[w]$ of degree k with leading coefficient 1 such that*

$$f(\mathbf{z}, w) = e(\mathbf{z}, w)\omega(\mathbf{z}, w). \tag{7}$$

The functions e and ω are uniquely determined by (7) and the above conditions.

Proof: Choose r such that f is nonzero for $r \leqslant |w| < R$, and apply Prop. 7.3: there exist holomorphic functions $c(\mathbf{z})$ and $h(\mathbf{z}, w)$ on U and $U \times K(r, R)$, respectively, such that

$$w^k\, e^{h(\mathbf{z},w)} = c(\mathbf{z}) f(\mathbf{z}, w) \tag{8}$$

and $c(\mathbf{z})$ is nonzero. Now apply Prop. 7.1 to obtain the Laurent decomposition of h:

$$h(\mathbf{z}, w) = h_0(\mathbf{z}, w) + h_\infty(\mathbf{z}, w), \tag{9}$$

where $h_0 \in \mathcal{O}(U \times D_R(0))$ and $h_\infty \in \mathcal{O}(U \times \mathbb{C} \setminus \overline{D_r(0)})$, and also

$$\lim_{w \to \infty} h_\infty(\mathbf{z}, w) = 0. \tag{10}$$

Substituting this into (8) yields

$$f(\mathbf{z}, w) = \frac{1}{c(\mathbf{z})}\, e^{h_0(\mathbf{z},w)}\, w^k\, e^{h_\infty(\mathbf{z},w)}. \tag{11}$$

We now set

$$e(\mathbf{z}, w) = \frac{1}{c(\mathbf{z})}\, e^{h_0(\mathbf{z},w)} \tag{12}$$

and investigate the remaining factors in (11). It follows from (10) that

$$\lim_{w\to\infty} e^{h_\infty(\mathbf{z},w)} = 1; \tag{13}$$

for fixed $\mathbf{z}$, we therefore have the Laurent expansion

$$e^{h_\infty(\mathbf{z},w)} = 1 + \sum_{\nu=1}^{\infty} a_\nu(\mathbf{z}) w^{-\nu}, \tag{14}$$

whose coefficients are given by integrals that depend holomorphically on $\mathbf{z}$ and are thus holomorphic on U. We write

$$\begin{aligned} w^k\, e^{h_\infty(\mathbf{z},w)} &= w^k \left(1 + \sum_{\nu=1}^{k} a_\nu(\mathbf{z}) w^{-\nu}\right) + w^k \sum_{\nu=k+1}^{\infty} a_\nu(\mathbf{z}) w^{-\nu} \\ &= \omega(\mathbf{z}, w) + R_\infty(\mathbf{z}, w) \end{aligned}$$

and obtain

$$f(\mathbf{z}, w) = e(\mathbf{z}, w)\, (\omega(\mathbf{z}, w) + R_\infty(\mathbf{z}, w)), \tag{15}$$

which is the desired decomposition if we can show that $R_\infty(\mathbf{z}, w) \equiv 0$. But by (15), we have

$$0 = \frac{f(\mathbf{z}, w)}{e(\mathbf{z}, w)} - \omega(\mathbf{z}, w) - R_\infty(\mathbf{z}, w) \stackrel{\text{def}}{=} R_0(\mathbf{z}, w) - R_\infty(\mathbf{z}, w) \tag{16}$$

on $U \times K(r, R)$, and this is the Laurent decomposition of the zero function! Since the Laurent decomposition is unique,

$$R_0(\mathbf{z}, w) = 0, \quad R_\infty(\mathbf{z}, w) = 0,$$

which proves the existence of the decomposition (7).

Uniqueness is almost trivial: for every $\mathbf{z}$, the functions

$$\begin{aligned} &w \mapsto f(\mathbf{z}, w) \\ &w \mapsto \omega(\mathbf{z}, w) \end{aligned}$$

have the same zeros. This determines $\omega(\mathbf{z}, w)$ and therefore also $e(\mathbf{z}, \omega)$. □

The preparation theorem contains a great deal of information about the local structure of the set of zeros of a holomorphic function of several complex variables. In the remainder of this section, we will describe such sets of zeros.

Let f be holomorphic in a connected neighbourhood W of $0 \in \mathbb{C}^{n+1}$ such that $f \neq 0$ and $f(0) = 0$. Then f cannot be identically zero on every complex line that passes through 0. Choosing a complex line L such that $f\big|L \not\equiv 0$ as the w-axis, then after a linear change of coordinates and possibly shrinking W, we may assume the following:

We have $f \in \mathcal{O}\left(U \times D_R(0)\right)$, where U is a connected neighbourhood of $0 \in \mathbb{C}^n$ and the only zero of $f(0, w)$ is 0, which is a zero of order k. Without loss of generality, let f be of order k on $U \times D_R(0)$.

The conditions of Thm. 7.4 are satisfied:

$$f(\mathbf{z}, w) = e(\mathbf{z}, w)\omega(\mathbf{z}, w),$$

where e and ω are as in the theorem. In particular, the set of zeros of f coincides with that of ω.

Here

$$\omega(\mathbf{z}, w) = w^k + a_1(\mathbf{z})w^{k-1} + \ldots + a_k(\mathbf{z}), \tag{17}$$

where $a_\varkappa$ is holomorphic on U and, by our assumption about f at 0, $a_\varkappa(0) = 0$ for $1 \leqslant \varkappa \leqslant k$.

From now on, we therefore only have to consider sets of zeros of polynomials (17) – so-called *Weierstrass polynomials* – where we may assume that they are contained in $U \times D_R(0) = U \times D$, where U is a domain in $\mathbb{C}^n$. Clearly, for every $\mathbf{z} \in U$, there are at least one and at most k points $(\mathbf{z}, w)$ such that $\omega(\mathbf{z}, w) = 0$. In other words: the holomorphic projection

$$\begin{aligned} p\colon U \times D &\to U \\ (\mathbf{z}, w) &\mapsto \mathbf{z} \end{aligned}$$

maps the set of zeros

$$M = \{(\mathbf{z}, w) \colon \omega(\mathbf{z}, w) = 0\} \tag{18}$$

onto U, and the fibres $\left(p\big|M\right)^{-1}(\mathbf{z})$ have at most k elements.

We want to study the projection $p\big|M$ more closely. Let us make the abbreviation $H = \mathcal{O}(U)$, and let K denote the field of fractions of the integral domain H. Moreover, assume that

$$\omega(\mathbf{z}, w) \in H[w] \subset K[w]$$

is an irreducible polynomial over K. We then set

$$\omega' = \omega_w(\mathbf{z}, w),$$

so ω' is a polynomial of degree $k-1$ in $H[w]$. Since ω is irreducible, ω and ω' are coprime in $K[w]$ and thus satisfy an equation

$$a\omega + b\omega' = 1,$$

for some $a, b \in K[w]$. Multiplying by a common denominator of the coefficients of a and b yields an identity

$$A\omega + B\omega' = C,$$

where $A, B \in H[w]$ and $C \in H$. To be precise,

$$A(\mathbf{z}, w)\omega(\mathbf{z}, w) + B(\mathbf{z}, w)\omega_w(\mathbf{z}, w) \equiv C(\mathbf{z}).$$

Outside of the set of zeros of C in U, i.e. outside of a nowhere dense set, $\omega(\mathbf{z}, w)$ and $\omega_w(\mathbf{z}, w)$ thus have no common zeros, which means that the polynomial $\omega(\mathbf{z}, w)$ can be decomposed into k linear factors: the fibre $\left(p|M\right)^{-1}(\mathbf{z})$ has exactly k points. Let $(\mathbf{z}_0, w_0)$ be such a point. In a neighbourhood $V(\mathbf{z}_0) \times D(w_0)$, the function ω thus has order 1 with respect to $w - w_0$: by the Weierstrass preparation theorem, we can factor ω:

$$\omega(\mathbf{z}, w) = e_1(\mathbf{z}, w)\omega_1(\mathbf{z}, w),$$

with $e_1(\mathbf{z}, w) \neq 0$ on $V(\mathbf{z}_0) \times D(w_0)$ and

$$\omega_1(\mathbf{z}, w) = w - w_0 + A_0(\mathbf{z}),$$

where $A_0(\mathbf{z})$ is holomorphic on V and satisfies $A_0(\mathbf{z}_0) = 0$. The set of zeros of ω in a neighbourhood of $(\mathbf{z}_0, w_0)$ is thus given by an equation

$$w = w_0 - A_0(\mathbf{z}).$$

Its projection onto $V(\mathbf{z}_0)$ is clearly a homeomorphism.

To summarize:

Proposition 7.5. *Let $U \subset \mathbb{C}^n$ be a domain and*

$$\omega(\mathbf{z}, w) = w^k + a_1(\mathbf{z})w^{k-1} + \ldots + a_k(\mathbf{z})$$

an irreducible polynomial over the field of fractions of $H = \mathcal{O}(U)$, with $a_\varkappa(\mathbf{z}) \in H$ ($\varkappa = 1, \ldots, k$), and let $M = \{(\mathbf{z}, w) \colon \omega(\mathbf{z}, w) = 0\}$ be its set of zeros. Then there exists a holomorphic function C on U, not identically zero, such that for all $\mathbf{z}$ with $C(\mathbf{z}) \neq 0$, the projection

$$p \colon M \to U, \quad p(\mathbf{z}, w) = \mathbf{z}$$

is a locally homeomorphic mapping. All fibres $\left(p|M\right)^{-1}(\mathbf{z})$ with $C(\mathbf{z}) \neq 0$ contain exactly k points. If $(\mathbf{z}_0, w_0)$ is such a point, then in a neighbourhood of $V(\mathbf{z}_0) \times D(w_0)$ the set M is the graph of a holomorphic function:

$$M \cap (V(\mathbf{z}_0) \times D(w_0)) = \{(\mathbf{z}, w) \colon \mathbf{z} \in V(\mathbf{z}_0),\ w = w_0 + A_0(\mathbf{z})\}.$$

The Weierstrass preparation theorem can be used to show much more, but we will content ourselves with the above application.

Exercises

1. Let U be a neighbourhood of 0 in $\mathbb{C}^n$ and D a disk about 0 in $\mathbb{C}$, and suppose $f(\mathbf{z}, w)$ is holomorphic on $U \times D$ with $f(0,0) = 0$ and $f(0,w) \not\equiv 0$. Show that there exists a neighbourhood $U_0 \subset U$ of 0, radii $0 < r < R$, and a number $k \geqslant 1$ such that f is nonzero on $U_0 \times (D_R(0) \setminus \overline{D_r(0)})$ and $w \mapsto f(\mathbf{z}, w)$ has exactly k zeros in $D_r(0)$ (counting multiplicity) for $\mathbf{z} \in U_0$. One can thus apply Thm. 7.4 to $f\big|U \times D_R(0)$.
2. Suppose the function $f(\mathbf{z}, w)$ has order k with respect to w in $U \times D_R(0)$ and $f(0,w)$ has a zero of order k at $w = 0$. Show that if $\varphi\colon U \to D_R(0)$ is a function that assigns to every $\mathbf{z}$ a zero of $w \mapsto f(\mathbf{z}, w)$, then φ is continuous at $0 \in U$.
3. Suppose f is holomorphic in a neighbourhood U of $0 \in \mathbb{C}^n$, $f(0) = 0$, and $f \not\equiv 0$. Show that there exists a linear bijection $L\colon \mathbb{C}^n \to \mathbb{C}^n$, a neighbourhood V of $0 \in \mathbb{C}^{n-1}$, and numbers $0 < r < R$ such that $g = f \circ L$ is of order $k \geqslant 1$ on $V \times D_R(0)$, i.e. for $(z_1, \ldots, z_{n-1}) \in V$, $z_n \mapsto g(z_1, \ldots, z_{n-1}, z_n)$ has exactly k zeros in $D_R(0)$ that all lie in $D_r(0)$.

Chapter V.

Non-elementary functions

The theory of the preceding chapters permits the construction and investigation of new transcendental functions. The Γ-function, interpolating the factorials, is perhaps the most important non-elementary function (V.1). Riemann's ζ-function (V.2,3) and its generalizations play an eminent role in number theory and algebraic geometry. It is the main tool in most proofs of the prime number theorem (V.2). Elliptic functions, i.e. functions with two independent periods, and their connection with plane cubic curves is a classical theme (V.4,5) with applications in many areas, e.g. mathematical physics and cryptography.

Beginning in 1729, Euler interpolated the factorials by an infinite product and established integral representations for it. Legendre, Gauss, and Weierstrass took up this theme. – Riemann introduced his ζ-function in a seminal paper in 1859 [Ri], which we follow in V.3. For the prime number theorem see the remarks at the end of V.2. – Elliptic functions appear as inverse functions of indefinite elliptic integrals (Gauss since 1796, Abel, Jacobi); they were a leading theme in 19^{th} century mathematics. We follow Hurwitz [HuC], who leaned on methods developed by Weierstrass; a great deal of the results go back to Eisenstein and Liouville (1844–47).

1. The Γ-function

We look for a meromorphic function on the complex plane, as simple as possible, interpolating the factorials, i.e. a function satisfying $f(n) = (n-1)!$ (for historical reasons, we do not demand $f(n) = n!$). In view of the recursive definition of the factorials,

$$0! = 1, \quad n \cdot (n-1)! = n!,$$

we require

$$f(1) = 1, \quad z \cdot f(z) = f(z+1). \tag{1}$$

By (1) we have $\lim\limits_{z\to 0} zf(z) = f(1) = 1$, therefore any solution of (1) must have a simple pole at $z = 0$ with residue 1. More generally, applying (1) $n+1$ times leads to

$$(z+n)f(z) = \frac{1}{z(z+1)\cdots(z+n-1)} f(z+n+1).$$

Taking the limit $z \to -n$ shows that f must have simple poles at $z = -n$ with $n = 0, 1, 2, \ldots$, the residues being $(-1)^n/n!$.

W. Fischer, I. Lieb, *A Course in Complex Analysis*, DOI 10.1007/978-3-8348-8661-3_5,

Therefore, if f satisfies (1), the function $g(z) = 1/f(z)$ will have simple zeros at $0, -1, -2, \ldots$ and satisfy the functional equation

$$g(1) = 1, \qquad zg(z+1) = g(z). \tag{2}$$

The Weierstrass product theorem suggests we try

$$g(z) = e^{h(z)} z \prod_1^\infty \Big(1 + \frac{z}{\nu}\Big) e^{-z/\nu} \tag{3}$$

with $h \in \mathcal{O}(\mathbb{C})$. This yields an entire function with the prescribed zeros, but we still have to show that h can be chosen so that g satisfies (2). We note that $g(1) = 1$ is equivalent to $\lim\limits_{z\to 0} g(z)/z = 1$, i.e. to $\exp h(0) = 1$. To cope with the second equation in (2) we consider the partial products

$$\begin{aligned} g_n(z) &= e^{h(z)} z \prod_1^n \Big(1 + \frac{z}{\nu}\Big) e^{-z/\nu} \\ &= \frac{1}{n!} \exp\Big(h(z) - z \sum_1^n \frac{1}{\nu}\Big) \prod_0^n (z + \nu). \end{aligned}$$

Then

$$\begin{aligned} \frac{zg_n(z+1)}{g_n(z)} &= (z+n+1) \exp\Big(h(z+1) - h(z) - \sum_1^n \frac{1}{\nu}\Big) \\ &= \Big(1 + \frac{1+z}{n}\Big) \exp\Big(h(z+1) - h(z) + \log n - \sum_1^n \frac{1}{\nu}\Big). \end{aligned}$$

Recall now that $\gamma := \lim\limits_{n\to\infty} \Big(\sum_1^n \frac{1}{\nu} - \log n\Big) = 0.5772\ldots$ exists (cf. Ex. 3); this is the so-called Euler's constant.

Therefore

$$1 = \frac{zg(z+1)}{g(z)} = \lim_{n\to\infty} \frac{zg_n(z+1)}{g_n(z)} = \exp\big(h(z+1) - h(z) - \gamma\big).$$

In other words: g satisfies (2) if and only if

$$\exp\big(h(z+1) - h(z) - \gamma\big) = 1 \text{ and } \exp h(0) = 1. \tag{4}$$

The simplest solution of (4) is $h(z) = \gamma z$, leading to the following definition:

Definition 1.1. *The function*

$$\Gamma(z) = e^{-\gamma z} \frac{1}{z} \prod_1^\infty \Big(1 + \frac{z}{\nu}\Big)^{-1} e^{z/\nu}$$

is called the Gamma function.

Summarizing:

Proposition 1.1. $\Gamma(z)$ *is meromorphic and without zeros throughout the complex plane. It satisfies the functional equation*

$$\Gamma(1) = 1, \qquad z\Gamma(z) = \Gamma(z+1),$$

in particular, $\Gamma(n) = (n-1)!$ *. It has simple poles precisely at* $z = -n$, $n = 0, 1, 2, \ldots$, *the residue in* $-n$ *is* $(-1)^n/n!$ *. If* $z = x$ *is real and positive,* $\Gamma(x)$ *is positive.* □

We consider once more the entire function $g(z) = 1/\Gamma(z)$. The product representation (3) (with $h(z) = \gamma z$) implies (cf. Prop. III.7.7)

$$g(z)g(-z) = -z^2 \prod_1^\infty \Bigl(1 - \frac{z^2}{\nu^2}\Bigr) = -z\frac{\sin \pi z}{\pi}.$$

Using (2) for $-z$, i.e. $g(-z) = -zg(1-z)$, we obtain

$$g(1-z)g(z) = \frac{\sin \pi z}{\pi}.$$

Proposition 1.2. *The equation* $\Gamma(1-z)\Gamma(z) = \dfrac{\pi}{\sin \pi z}$ *holds.* □

Setting $z = \frac{1}{2}$ shows

$$\Gamma\Bigl(\frac{1}{2}\Bigr) = \sqrt{\pi},$$

and, employing the functional equation for $n = 1, 2, 3, \ldots$,

$$\Gamma\Bigl(\frac{1}{2} + n\Bigr) = 2^{-n}\bigl(1 \cdot 3 \cdot \ldots \cdot (2n-1)\bigr)\sqrt{\pi},$$
$$\Gamma\Bigl(\frac{1}{2} - n\Bigr) = (-1)^n 2^n \bigl(1 \cdot 3 \cdot \ldots \cdot (2n-1)\bigr)^{-1}\sqrt{\pi}.$$

We now deduce two other useful representations of the Γ-function, originating with Euler and Gauss. We begin by rewriting

$$\begin{aligned} g_n(z) &= \frac{1}{n!}\exp\Bigl(z\Bigl(\gamma - \sum_1^n \frac{1}{\nu}\Bigr)\Bigr)\prod_0^n (z+\nu) \\ &= \frac{1}{n!\,n^z}\exp\Bigl(z\Bigl(\gamma - \sum_1^n \frac{1}{\nu} + \log n\Bigr)\Bigr)\prod_0^n (z+\nu). \end{aligned}$$

Now $\Gamma(z) = \lim_{n\to\infty} \dfrac{1}{g_n(z)}$ together with the definition of γ yields immediately

Proposition 1.3. $\Gamma(z) = \lim_{n\to\infty} \dfrac{n!\,n^z}{z(z+1)\cdot \ldots \cdot (z+n)}.$ □

The fraction in this formula can be represented as an integral. Indeed, assuming $z = x \geqslant 1$ real, integrating by parts $(n+1)$ times we arrive at

$$\int_0^n \Big(1 - \frac{t}{n}\Big)^n t^{x-1} dt = n^x \int_0^1 (1-s)^n s^{x-1} ds = \frac{n!\, n^x}{x(x+1)\cdot \ldots \cdot (x+n)}.$$

In view of taking the limit $n \to \infty$, we introduce the functions

$$\varphi_n(t) = \Big(1 - \frac{t}{n}\Big)^n \text{ for } 0 \leqslant t \leqslant n, \quad \varphi_n(t) = 0, \text{ for } t \geqslant n.$$

The φ_n converge to e^{-t} pointwise, furthermore $\varphi_n(t) \leqslant e^{-t}$ (the function $\varphi_n(t)e^t$ is nonincreasing for $t \geqslant 0$). Therefore, by Lebesgue's theorem on dominated convergence, applied to the sequence $\varphi_n(t)t^{x-1}$, we obtain

$$\int_0^\infty e^{-t} t^{x-1} dt = \lim_{n\to\infty} \int_0^\infty \varphi_n(t) t^{x-1} dt = \lim_{n\to\infty} \frac{n!\, n^x}{x(x+1)\cdot \ldots \cdot (x+n)}.$$

Proposition 1.4. *For* $\operatorname{Re} z > 0$ *we have*

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt. \tag{5}$$

Proof: We just proved the equation for z real $\geqslant 1$. If we show the integral in (5) to be a holomorphic function in $\operatorname{Re} z > 0$, the claim will follow from the identity theorem. Now

$$I_n(z) = \int_{1/n}^n e^{-t} t^{z-1} dt$$

is clearly holomorphic in $\mathbb{C}$. Using $|t^{z-1}| = t^{x-1}$ one verifies easily that the I_n converge uniformly on every vertical strip $0 < c_1 \leqslant \operatorname{Re} z \leqslant c_2 < \infty$. By Weierstrass's convergence theorem, $I(z) = \lim I_n(z)$ is holomorphic on $\operatorname{Re} z > 0$. □

Remark: Of course, Prop. 1.4 follows directly from Lebesgue's theorem on interchanging limits; we preferred to use the weaker Prop. I.5.6.

Another useful identity is Legendre's duplication formula:

Proposition 1.5. $\Gamma(2z) = \dfrac{1}{\sqrt{\pi}} 2^{2z-1} \Gamma(z)\Gamma(z + \frac{1}{2})$.

Proof: We use the Weierstrass product to compute the logarithmic derivative of Γ:

$$\frac{d}{dz}\log\Gamma(z) = -\gamma - \frac{1}{z} + \sum_1^\infty\Big(\frac{1}{\nu} - \frac{1}{z+\nu}\Big), \tag{6}$$

$$\frac{d^2}{dz^2}\log\Gamma(z) = \sum_0^\infty \frac{1}{(z+\nu)^2}. \tag{7}$$

Therefore,

$$\begin{aligned}\frac{d^2}{dz^2}\log\big(\Gamma(z)\Gamma(z+\tfrac{1}{2})\big) &= \sum_0^\infty \frac{1}{(z+\nu)^2} + \sum_0^\infty \frac{1}{\big(z+\frac{1}{2}+\nu\big)^2} \\ &= \sum_0^\infty \frac{4}{(2z+2\nu)^2} + \sum_0^\infty \frac{4}{(2z+1+2\nu)^2} \\ &= 4\sum_0^\infty \frac{1}{(2z+\nu)^2} = \frac{d^2}{dz^2}\log\Gamma(2z).\end{aligned}$$

Consequently,

$$\log\big(\Gamma(z)\Gamma(z+\tfrac{1}{2})\big) - \log\Gamma(2z)$$

is a linear function $az+b$, and

$$\frac{\Gamma(z)\Gamma(z+\frac{1}{2})}{\Gamma(2z)} = e^{az+b}.$$

To determine the coefficients a and b, we set $z=1$ and $z=1/2$, obtaining

$$\frac{1}{2}\sqrt{\pi} = \Gamma\Big(\frac{3}{2}\Big) = e^{a+b} \text{ and } \sqrt{\pi} = \Gamma\Big(\frac{1}{2}\Big) = e^{(a/2)+b},$$

and finally $e^a = 1/4,\ e^b = 2\sqrt{\pi}$. □

We conclude this section with a proof of Stirling's formula, which describes the behaviour of $\Gamma(z)$ for $\operatorname{Re} z \to +\infty$ and enables one to calculate good approximations of the values $\Gamma(z)$ for $\operatorname{Re} z$ large.

We start with formula (7) and the identity

$$\int_0^\infty te^{-(z+\nu)t}dt = \frac{1}{(z+\nu)^2} \text{ for } \operatorname{Re} z > 0,\ \nu = 0, 1, 2, \dots .$$

Summation yields

$$\frac{d^2}{dz^2}\log\Gamma(z) = \int_0^\infty te^{-zt}\sum_0^\infty e^{-\nu t}dt = \int_0^\infty e^{-(z-1)t}\frac{t}{e^t-1}dt$$

or

$$\frac{d^2}{dz^2}\log\Gamma(z+1)=\int_0^\infty e^{-zt}\frac{t}{e^t-1}dt. \tag{8}$$

The interchange of summation and integration is again justified by the dominated convergence theorem. Integrating (8) formally with respect to z leads to the integral

$$\int_0^\infty e^{-zt}\frac{dt}{e^t-1},$$

which does not exist on account of the singularity at $t=0$. To avoid this and an analogous complication with a second integration, we recall (cf. III.6 (10))

$$\frac{t}{e^t-1}=1-\frac{t}{2}+\frac{1}{12}t^2-+\cdots$$

for small t and write

$$\frac{t}{e^t-1}=1-\frac{t}{2}+t^2f(t),$$

that is

$$f(t)=\Big(\frac{1}{e^t-1}-\frac{1}{t}+\frac{1}{2}\Big)\frac{1}{t}\,. \tag{9}$$

Then we have for $\operatorname{Re} z>0$

$$\begin{aligned}\int_0^\infty e^{-zt}\frac{t}{e^t-1}dt&=\int_0^\infty e^{-zt}t^2f(t)dt+\int_0^\infty e^{-zt}\Big(1-\frac{t}{2}\Big)dt\\&=\int_0^\infty e^{-zt}t^2f(t)dt+\frac{1}{z}-\frac{1}{2z^2}\,.\end{aligned}$$

Now we may integrate twice under the integral sign with respect to z and obtain

$$\frac{d}{dz}\,\log\Gamma(z+1)=\log z+\frac{1}{2z}-\int_0^\infty e^{-zt}tf(t)dt+c_1, \tag{10}$$

$$\log\Gamma(z+1)=z(\log z-1)+\frac{1}{2}\log z+\int_0^\infty e^{-zt}f(t)dt+c_1z+c_0 \tag{11}$$

with integration constants c_0 and c_1. To fix c_0 we choose the principal branch of the logarithm.

Now we determine c_0 and c_1.
For c_1: Since $\lim_{t\to\infty} tf(t) = \frac{1}{2}$, the function $tf(t)$ is bounded on $[0, +\infty[$, $|tf(t)| \leqslant M$ say, and

$$\left|\int_0^\infty e^{-zt} tf(t)dt\right| \leqslant \frac{M}{\operatorname{Re} z} \to 0$$

for $\operatorname{Re} z \to +\infty$. Therefore by (10)

$$c_1 = \lim_{\operatorname{Re} z\to\infty}\left(\frac{d}{dz}\log\Gamma(z+1) - \log z - \frac{1}{2z}\right) = \lim_{\operatorname{Re} z\to\infty}\left(\frac{d}{dz}\log\Gamma(z+1) - \log z\right).$$

We let $z \to \infty$ by integer values n. By (6) we have

$$\frac{d\log\Gamma}{dz}(n+1) - \log n = \left(-\gamma + \sum_1^n \frac{1}{\nu}\right) - \log n.$$

By the definition of γ, the limit is 0, i.e. $c_1 = 0$.

For c_0: The integral in (11) tends to zero for $\operatorname{Re} z \to \infty$, too, as f is bounded. Therefore

$$c_0 = \lim_{\operatorname{Re} z\to\infty}\left(\log\Gamma(z+1) - (z+\tfrac{1}{2})\log z + z\right),$$

or, taking exponentials,

$$c_2 := e^{c_0} = \lim_{\operatorname{Re} z\to\infty}\left(\Gamma(z+1)z^{-z-1/2}e^z\right),$$

or, replacing z by $2z$:

$$c_2 = \lim_{\operatorname{Re} z\to\infty}\left(\Gamma(2z+1)(2z)^{-2z-1/2}e^{2z}\right).$$

Into the last formula we insert a slightly modified duplication formula (Prop. 1.5), namely $\Gamma(2z+1) = \pi^{-1/2}2^{2z}\Gamma(z+1)\Gamma(z+\frac{1}{2})$. After some elementary manipulations we obtain

$$c_2 = {c_2}^2(2\pi)^{-1/2}e^{1/2}\lim_{\operatorname{Re} z\to\infty}\left(1-\frac{1}{2z}\right)^z = {c_2}^2(2\pi)^{-1/2},$$

that is $c_0 = \log c_2 = \frac{1}{2}\log 2\pi$.

This proves the first part of

Proposition 1.6. *Define $f(t)$ by (9), then the following equation holds for $\operatorname{Re} z > 0$:*

$$\log\Gamma(z+1) = \left(z+\frac{1}{2}\right)\log z - z + \frac{1}{2}\log 2\pi + \int_0^\infty e^{-zt}f(t)dt,$$

and, uniformly in $\operatorname{Im} z$,

$$\lim_{\operatorname{Re} z\to\infty}\left(\log\Gamma(z+1) - \left(z+\frac{1}{2}\right)\log z + z\right) = \frac{1}{2}\log 2\pi.$$

The last equation is equivalent to *Stirling's formula*

$$\lim_{\operatorname{Re} z\to\infty}\left(\Gamma(z+1)z^{-z-1/2}e^{z}\right)=\sqrt{2\pi}.$$

Using $\Gamma(z+1)=z\Gamma(z)$, we can write this as

$$\Gamma(z)\sim\sqrt{2\pi}z^{z-1/2}e^{-z},$$

the sign $\sim$ ("asymptotic equality") meaning that the quotient of both sides tends to 1 as $\operatorname{Re} z\to+\infty$.

Proof of Prop. 1.6, second part**:** f is bounded on $[0,+\infty[$, $|f(t)|\leqslant M$, say. Therefore

$$\left|\int_0^\infty e^{-zt}f(t)dt\right|\leqslant M\int_0^\infty e^{-t\operatorname{Re} z}dt=\frac{M}{\operatorname{Re} z}\to 0. \qquad\square$$

More precisely, the sharp inequality $0<f(t)\leqslant\dfrac{1}{12}$ holds for $t\geqslant 0$ (cf. Ex. 4). This yields

$$\left|\log\Gamma(z+1)-\left(z+\frac{1}{2}\right)\log z+z-\frac{1}{2}\log 2\pi\right|\leqslant\frac{1}{12\operatorname{Re} z}$$

for $\operatorname{Re} z>0$, and for real $x>0$

$$\sqrt{2\pi}x^{x-1/2}e^{-x}\leqslant\Gamma(x)\leqslant\sqrt{2\pi}x^{x-1/2}e^{-x+(1/12x)}.$$

As an example we take $x=1000$ and obtain

$$5912.1281\leqslant\log(1000!)\leqslant 5912.1282,$$

that is

$$4.023\cdot 10^{2567}\leqslant 1000!\leqslant 4.024\cdot 10^{2567}.$$

Exercises

1. Express $\Gamma(\frac{1}{2}+z)\Gamma(\frac{1}{2}-z)$ and $\Gamma(z)\Gamma(-z)$ by trigonometric functions. Deduce $|\Gamma(iy)|^2=\dfrac{\pi}{y\sinh\pi y}$ ($y\neq 0$, real).
2. Show $\Gamma(z+1)\Gamma(z+\frac{1}{2})=\sqrt{\pi}4^{-z}\Gamma(2z+1)$.
3. Prove: The sequence $\displaystyle\sum_1^n\frac{1}{\nu}-\log n$ is positive and decreasing (hence its limit γ exists).
4. For the function $f(t)$ defined in (9) prove the estimate $0<f(t)\leqslant\frac{1}{12}$.
Hint: $f'(t)$ is clearly negative for $t\geqslant 4$. In a neighbourhood of 0 the terms of the power series for f have alternating signs and decreasing modulus (cf. III.6.(11)).
5. Show $\displaystyle\int_0^\infty\frac{t\,dt}{e^t-1}=\frac{\pi^2}{6}$.

2. The ζ-function and the Prime Number Theorem

Riemann's ζ-function is defined by

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} \tag{1}$$

for complex $s = \sigma + it$ (this notation of the complex variable was introduced by Riemann). For the power $n^{-s} = e^{-s \log n}$ we use the real logarithm of n. Since $\left|n^{-s}\right| = n^{-\sigma}$, the series (1) converges uniformly for $\sigma \geqslant \sigma_0 > 1$, therefore locally uniformly on the half plane $\{\sigma + it\colon \sigma > 1\}$. Accordingly, (1) defines $\zeta(s)$ as a holomorphic function on $\operatorname{Re} s > 1$. We have already calculated the special values $\zeta(2n)$, $n = 1, 2, \dots$ in Prop. III.6.5.

The following product representation, found by Euler, is fundamental for the use of $\zeta(s)$ in number theory.

Proposition 2.1.

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1} \tag{2}$$

for $\operatorname{Re} s > 1$*, where the product is extended over all prime numbers.*

Proof: The product converges locally uniformly and absolutely, since $\sum p^{-s}$ is a subseries of $\sum n^{-s}$. – We number the primes in increasing order: $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, $\dots$. Then, by expanding into geometric series:

$$\prod_{p_1,\dots,p_k} (1 - p^{-s})^{-1} = \prod_{p_1,\dots,p_k} \sum_{m=0}^{\infty} \left(p^m\right)^{-s} = \sum n^{-s},$$

the right hand sum being taken over those n whose (unique!) prime factor decomposition contains only $p_1, \dots, p_k$. Thus the first missing n is p_{k+1}. Now $k \to \infty$ proves the claim. □

From (2) we infer that there are infinitely many primes and even the divergence of $\sum_p 1/p$: If there were only finitely many, the product would have a finite limit for $s \to 1$, whereas the sum tends to infinity (cf. also Ex. 1).

Denoting by $\pi(x)$ the number of of primes $p \leqslant x$, we therefore have

$$\lim_{x \to \infty} \pi(x) = \infty.$$

The old problem (cf. the historical note at the end of this section) of the order of growth of $\pi(x)$ is answered by

Theorem 2.2 (Prime Number Theorem)**.**

$$\lim_{x\to\infty} \frac{\pi(x)\log x}{x} = 1. \tag{3}$$

Using the notion of asymptotic equality introduced after Prop. 1.6, we can write (3) as

$$\pi(x) \sim \frac{x}{\log x}.$$

The remainder of this section is devoted to the proof of Thm. 2.2.

First, we derive two properties of the ζ-function. Here and later on, we will make use of an integral transformation:

Let $f\colon [1, +\infty[\to \mathbb{R}$ be a locally integrable function, assume there is a $k \in \mathbb{R}$ with $|f(x)| \leqslant Cx^k$. Then $f(x)x^{-s-1}$ is integrable for $\operatorname{Re} s > k$ and

$$\mathcal{M}_f\colon s \mapsto s\int_1^\infty f(x)x^{-s-1}dx$$

is holomorphic in this half plane. We call $\mathcal{M}_f$ the *Mellin transform* of f.

As an example, for $f(x) = x$ one easily computes

$$\mathcal{M}_f(s) = \frac{s}{s-1} = \frac{1}{s-1} + 1.$$

Taking $g(x) = [x]$ (the *Gauss bracket* $[x]$ denotes the largest integer $\leqslant x$), we obtain for $\operatorname{Re} s > 1$:

$$\begin{aligned}\mathcal{M}_g(s) &= s\int_1^\infty [x]\, x^{-s-1}dx = \sum_{n=1}^\infty s\int_n^{n+1} nx^{-s-1}dx\\ &= \sum_{n=1}^\infty n\big(n^{-s} - (n+1)^{-s}\big).\end{aligned}$$

Assuming $\operatorname{Re} s > 2$, we can decompose the last sum:

$$\begin{aligned}\mathcal{M}_g(s) &= \sum_{n=1}^\infty n^{1-s} - \sum_{n=1}^\infty (n+1-1)(n+1)^{-s}\\ &= \sum_{n=1}^\infty n^{1-s} - \sum_{n=2}^\infty n^{1-s} + \sum_{n=2}^\infty n^{-s}\\ &= 1 + \big(\zeta(s) - 1\big) = \zeta(s).\end{aligned}$$

By the identity theorem, the equation $\mathcal{M}_g(s) = \zeta(s)$ holds for $\operatorname{Re} s > 1$, too.

Combining these examples, we get for $\operatorname{Re} s > 1$

$$\mathcal{M}_{([x]-x)}(s) = \zeta(s) - \frac{1}{s-1} - 1.$$

But $[x] - x$ is bounded, so $\mathcal{M}_{([x]-x)}$ is holomorphic for $\operatorname{Re} s > 0$. Therefore

Proposition 2.3. *The Riemann ζ-function can be extended by*

$$\zeta(s) = \frac{1}{s-1} + 1 + \mathcal{M}_{([x]-x)}(s)$$

to a function meromorphic on the half plane $\operatorname{Re} s > 0$. *The extension is holomorphic save for a simple pole at $s = 1$ with residue* 1. □

The other property of $\zeta(s)$ we need is

Proposition 2.4. $\zeta(s)$ *has no zeros on the line* $\operatorname{Re} s = 1$.

Proof: Suppose $\zeta(1 + it_0) = 0$ for some real $t_0 \neq 0$. Consider the function

$$g(s) = \zeta(s)^3 \zeta(s + it_0)^4 \zeta(s + 2it_0).$$

Since the pole at $s = 1$ of the first factor is cancelled by the zero of the second, $g(s)$ is holomorphic in a neighbourhood of $s = 1$ and $g(1) = 0$. Therefore, $\log|g(s)| \to -\infty$ if $s \to 1$. Now Euler's product (2) gives for $\operatorname{Re} s > 1$

$$\begin{aligned}
\log|\zeta(s)| &= \operatorname{Re} \sum_p \log(1 - p^{-s})^{-1} \\
&= \operatorname{Re} \sum_p \sum_{\nu=1}^{\infty} \frac{1}{\nu} (p^\nu)^{-s} \\
&= \operatorname{Re} \sum_{n=1}^{\infty} a_n n^{-s}
\end{aligned}$$

where $a_n = 0$ if n is not a power of a prime and $a_n = 1/\nu$ if $n = p^\nu$. We only use $a_n \geqslant 0$.

Then, for $\sigma > 1$

$$\begin{aligned}
\log|g(\sigma)| &= 3\log|\zeta(\sigma)| + 4\log|\zeta(\sigma + it_0)| + \log|\zeta(\sigma + 2it_0)| \\
&= \sum_{n=1}^{\infty} a_n \operatorname{Re}\bigl(3n^{-\sigma} + 4n^{-\sigma - it_0} + n^{-\sigma - 2it_0}\bigr) \\
&= \sum_{n=1}^{\infty} a_n\, n^{-\sigma} \bigl[3 + 4\cos(t_0 \log n) + \cos(2t_0 \log n)\bigr].
\end{aligned}$$

The term in brackets is nonnegative since

$$3 + 4\cos\alpha + \cos 2\alpha = 3 + 4\cos\alpha + 2\cos^2\alpha - 1 = 2(1+\cos\alpha)^2.$$

Therefore $\log|g(\sigma)| \geqslant 0$, contradicting $\lim\limits_{\sigma\downarrow 1} \log|g(\sigma)| = -\infty$. □

Now we turn to some number theoretic preliminaries. We introduce the function

$$\vartheta(x) = \sum_{p\leqslant x} \log p,$$

(summation is extended over all primes $\leqslant x$), which is more amenable to analytic methods than $\pi(x)$. The connection between $\vartheta(x)$ and $\pi(x)$ is expressed by

Proposition 2.5. *If one of the limits in the following formula exists, then so does the other, and they are equal:*

$$\lim_{x\to\infty} \frac{\pi(x)\log x}{x} = \lim_{x\to\infty} \frac{\vartheta(x)}{x}.$$

We shall eventually show that the right hand side limit exists and has the value 1. This will prove the prime number theorem.

Proof: On the one hand

$$\vartheta(x) = \sum_{p\leqslant x} \log p \leqslant \sum_{p\leqslant x} \frac{\log x}{\log p} \log p = \log x \sum_{p\leqslant x} 1 = \pi(x)\log x.$$

On the other hand we can estimate

$$\pi(x) = \pi(y) + \sum_{y<p\leqslant x} 1 \leqslant \pi(y) + \frac{1}{\log y} \sum_{y<p\leqslant x} \log p \leqslant y + \frac{1}{\log y}\vartheta(x)$$

for $1 < y < x$, hence

$$\frac{\pi(x)\log x}{x} \leqslant \frac{y\log x}{x} + \frac{\log x}{\log y} \cdot \frac{\vartheta(x)}{x}.$$

Choosing (for $x \geqslant 3$) $y = \dfrac{x}{(\log x)^2}$, we obtain

$$\frac{\pi(x)\log x}{x} \leqslant \frac{1}{\log x} + \frac{\log x}{\log x - 2\log\log x} \cdot \frac{\vartheta(x)}{x}.$$

Combining the two inequalities yields the claim because

$$\lim_{x\to\infty} \frac{1}{\log x} = 0, \quad \lim_{x\to\infty} \frac{\log x}{\log x - 2\log\log x} = 1.$$ □

Complex analysis enters the scene in form of the Mellin transform of $\vartheta(x)$. We need

Lemma 2.6. *The function* $\dfrac{\vartheta(x)}{x}$ *is bounded on* $[1, +\infty[$.

Proof: We start with the following observation: If $k < p \leqslant 2k$, then p divides the binomial coefficient

$$\binom{2k}{k} = \frac{(k+1)\cdot\ldots\cdot(2k-1)\cdot 2k}{k!}.$$

Therefore $\binom{2k}{k}$ is divisible by the product of all p with $k < p \leqslant 2k$, in particular not less than this product:

$$\prod_{k<p\leqslant 2k} p \leqslant \binom{2k}{k} \leqslant 2^{2k}; \tag{4}$$

the right hand inequality results from the binomial expansion of $(1+1)^{2k}$. By (4) we have

$$\sum_{k<p\leqslant 2k} \log p \leqslant 2k \log 2.$$

Inserting $k = 1, 2, 2^2, \ldots, 2^{\ell-1}$ and adding the inequalities, we obtain

$$\sum_{p\leqslant 2^\ell} \log p \leqslant (2+4+\ldots+2^\ell)\log 2 \leqslant 2^{\ell+1}\log 2.$$

Given $x \geqslant 1$, we choose ℓ so that $2^{\ell-1} \leqslant x < 2^\ell$, and get

$$\vartheta(x) = \sum_{p\leqslant x} \log p \leqslant \sum_{p\leqslant 2^\ell} \log p \leqslant 2^{\ell+1}\log 2 \leqslant 4x\log 2. \qquad \square$$

In view of the lemma, $\mathcal{M}_\vartheta$ is holomorphic on $\operatorname{Re} s > 1$. We have

Proposition 2.7. *For* $\operatorname{Re} s > 1$,

$$\mathcal{M}_\vartheta(s) = \sum_p \frac{\log p}{p^s} = -\frac{\zeta'(s)}{\zeta(s)} - \sum_p \frac{\log p}{p^s(p^s-1)} \tag{5}$$

holds (the sums are extended over all primes p).

Proof: First, we deduce the left equation. Noting $\vartheta(x) = \vartheta(n)$ if $n \leqslant x < n+1$, we compute

$$s\int_1^\infty \vartheta(x)x^{-s-1}dx = s\sum_{n=1}^\infty \int_n^{n+1} \vartheta(x)x^{-s-1}dx = s\sum_{n=1}^\infty \vartheta(n)\int_n^{n+1} x^{-s-1}dx$$
$$= \sum_{n=1}^\infty \vartheta(n)\big(n^{-s} - (n+1)^{-s}\big).$$

By the lemma, $\sum \vartheta(n)n^{-s}$ converges for $\operatorname{Re} s > 2$. Thus, assuming $\operatorname{Re} s > 2$, we may decompose the last sum and find, using $\vartheta(0) = 0$,

$$s\int_1^\infty \vartheta(x)x^{-s-1}dx = \sum_{n=1}^\infty \vartheta(n)n^{-s} - \sum_{n=2}^\infty \vartheta(n-1)n^{-s}$$
$$= \sum_{n=1}^\infty \big(\vartheta(n) - \vartheta(n-1)\big)n^{-s}.$$

Since $\vartheta(n) - \vartheta(n-1) = \log p$ if $n = p$ is a prime and $\vartheta(n) - \vartheta(n-1) = 0$ otherwise, this proves the left equation in (5) for $\operatorname{Re} s > 2$. Both sides of the equation are holomorphic for $\operatorname{Re} s > 1$, so the equation holds on $\operatorname{Re} s > 1$, too.

Now we prove the right equation in (5). Logarithmic derivation of the Euler product

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1}$$

leads to

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_p \frac{p^{-s}\log p}{1 - p^{-s}} = \sum_p \frac{\log p}{p^s - 1}.$$

The identity

$$\sum_p \frac{\log p}{p^s - 1} - \sum_p \frac{\log p}{p^s(p^s - 1)} = \sum_p \frac{\log p}{p^s}$$

is evident. □

Equation (5) implies

$$\mathcal{M}_{(\vartheta(x)-x)}(s) = -\frac{\zeta'(s)}{\zeta(s)} - \frac{1}{s-1} - 1 - \sum_p \frac{\log p}{p^s(p^s - 1)}. \tag{6}$$

By Prop. 2.4, $-\zeta'/\zeta$ is holomorphic in a neighbourhood of $\operatorname{Re} s \geqslant 1$, excepting the pole at $s = 1$ with residue 1. The sum in (6) converges locally uniformly for $\operatorname{Re} s > \frac{1}{2}$ and is, consequently, holomorphic there. So (6) provides a holomorphic extension of $\mathcal{M}_{(\vartheta(x)-x)}$ to a neighbourhood of $\operatorname{Re} s \geqslant 1$, and we can apply the following crucial statement:

Proposition 2.8. *Let $f\colon [1, +\infty[\to \mathbb{R}$ be locally integrable and let $f(x)/x$ be bounded. If $\mathcal{M}_f$ has a holomorphic extension to a neighbourhood of $\operatorname{Re} s \geqslant 1$, then*

$$\lim_{\mu\to\infty} \int_1^\mu \frac{f(x)}{x}\,\frac{dx}{x}$$

exists.

The existence of $\lim\limits_{\mu\to\infty}\int\limits_1^\mu \Big(\frac{\vartheta(x)}{x}-1\Big)\frac{dx}{x}$ implies $\lim\limits_{x\to\infty}\frac{\vartheta(x)}{x}=1$, i.e. the prime number theorem, by

Lemma 2.9. *Let $g\colon [1,+\infty[\ \to \mathbb{R}$ be nondecreasing, assume the existence of $\lim\limits_{\mu\to\infty}\int\limits_1^\mu \Big(\frac{g(x)}{x}-1\Big)\frac{dx}{x}$. Then $\lim\limits_{x\to\infty}\frac{g(x)}{x}=1$.*

Proof: Suppose the claim is wrong. Then there exists a number $\eta>0$ and a sequence $x_\nu\to+\infty$ such that $\Big|\frac{g(x_\nu)}{x_\nu}-1\Big|\geqslant\eta$. Assume $\frac{g(x_\nu)}{x_\nu}-1\geqslant\eta$, i.e. $g(x_\nu)\geqslant(1+\eta)x_\nu$, for infinitely many ν. After passing to a subsequence, we can assume this inequality holds for all ν. Setting $\varrho=\frac{1+\eta}{1+\frac{\eta}{2}}$, we have for $x_\nu\leqslant x\leqslant\varrho x_\nu$

$$\Big(1+\frac{\eta}{2}\Big)x\leqslant(1+\eta)x_\nu\leqslant g(x_\nu)\leqslant g(x),$$

the last inequality by the monotonicity of g. Then

$$\int\limits_{x_\nu}^{\varrho x_\nu}\Big(\frac{g(x)}{x}-1\Big)\frac{dx}{x}\geqslant\frac{\eta}{2}\int\limits_{x_\nu}^{\varrho x_\nu}\frac{dx}{x}=\frac{1}{2}\eta\log\varrho>0$$

for all ν, contradicting the existence of the improper integral.
In case that $g(x_\nu)\leqslant(1-\eta)x_\nu$ for almost all ν, we can proceed similarly. □

It remains to prove Prop. 2.8. We set $g(x)=f(x)/x$ and note that the factor s, appearing in the definition of $\mathcal{M}_f$, does not affect holomorphy on $\operatorname{Re}s\geqslant1$. Thus, Prop. 2.8 is contained in

Theorem 2.10. *Let $g\colon[1,+\infty[\to\mathbb{R}$ be a bounded locally integrable function. Set, for $\operatorname{Re}s>1$,*

$$G(s)=\int\limits_1^\infty g(x)x^{-s}\,dx.$$

If G can be holomorphically extended to a neighbourhood U of the closed half plane $\operatorname{Re}s\geqslant1$, then

$$\lim_{\mu\to\infty}\int\limits_1^\mu g(x)\,\frac{dx}{x}=G(1).$$

We have denoted the extension of G again by G; this explains the right hand side of the equation. Note that we do not claim the existence of the integral

$$\int_1^\infty g(x)\,\frac{dx}{x}.$$

Proof: The functions

$$G_\mu(s) = \int_1^\mu g(x)x^{-s}\,dx$$

are holomorphic for all $s \in \mathbb{C}$; we have to show

$$\lim_{\mu\to\infty} G_\mu(1) = G(1).$$

We represent the difference $G(1) - G_\mu(1)$ by Cauchy's integral formula:

For any $R > 0$ there is a $\delta = \delta(R) > 0$ such that the circular segment $S = \{s : |s-1| \leqslant R, \operatorname{Re} s \geqslant 1-\delta\}$ is contained in U. We subdivide the positively oriented boundary γ of S into $\gamma = \gamma_1 + \gamma_2$, γ_1 being the part of γ in $\operatorname{Re} s \geqslant 1$ and γ_2 the part in $\operatorname{Re} s \leqslant 1$.

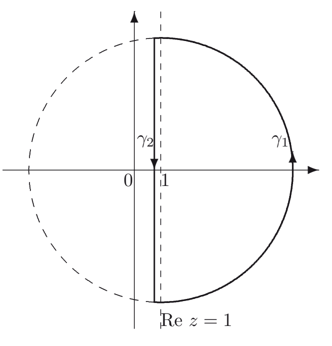


Figure 10. Path of integration

Then

$$G(1) - G_\mu(1) = \frac{1}{2\pi i}\int_\gamma \big(G(s) - G_\mu(s)\big)\frac{ds}{s-1}.$$

We hope that the integral will become small for R and μ large. Not knowing much about $G(s)$ for $\operatorname{Re} s \leqslant 1$, we decompose

$$\int_{\gamma} \big(G(s) - G_{\mu}(s)\big) \frac{ds}{s-1} = \int_{\gamma_1} \big(G(s) - G_{\mu}(s)\big) \frac{ds}{s-1} - \int_{\gamma_2} G_{\mu}(s) \frac{ds}{s-1} + \int_{\gamma_2} G(s) \frac{ds}{s-1}.$$

We may assume $|g| \leqslant 1$. Then one can estimate the integrands as follows. For $\sigma = \operatorname{Re} s > 1$,

$$\big|G(s) - G_{\mu}(s)\big| = \left|\int_{\mu}^{\infty} g(x) x^{-s} dx\right| \leqslant \int_{\mu}^{\infty} x^{-\sigma} dx = \frac{1}{\sigma - 1} \mu^{1-\sigma} = \left|\frac{2}{(s-1) + (\bar{s} - 1)} \mu^{1-s}\right|,$$

thus for $\operatorname{Re} s > 1$ and $|s-1| = R$

$$\left|\big(G(s) - G_{\mu}(s)\big) \frac{1}{s-1}\right| \leqslant \left|\frac{2}{(s-1)^2 + R^2} \mu^{1-s}\right|, \tag{7}$$

this bound is clearly bad if $s \to 1 \pm iR$.

For $\sigma = \operatorname{Re} s < 1$,

$$\big|G_{\mu}(s)\big| \leqslant \int_{1}^{\mu} x^{-\sigma} dx = \frac{1}{1-\sigma} \Big(\mu^{1-\sigma} - 1\Big) \leqslant \frac{1}{1-\sigma} \mu^{1-\sigma},$$

thus for $\operatorname{Re} s < 1$ and $|s-1| = R$

$$\left|G_{\mu}(s) \frac{1}{s-1}\right| \leqslant \left|\frac{2}{(s-1)^2 + R^2} \mu^{1-s}\right|. \tag{8}$$

This bound is again bad for $s \to 1 \pm iR$ and also for large μ.

So the straightforward estimates of the integrals do not lead to a satisfactory estimate of $G(1) - G_{\mu}(1)$. The essential trick now is to replace $G - G_{\mu}$ by $(G - G_{\mu}) \cdot h_{\mu,\,R}$, where $h_{\mu,\,R}$ is a suitable holomorphic function satisfying $h_{\mu,\,R}(1) = 1$, then

$$G(1) - G_{\mu}(1) = \frac{1}{2\pi i} \int_{\gamma} \big(G(s) - G_{\mu}(s)\big) \frac{h_{\mu,\,R}(s)}{s-1}\, ds.$$

The above estimates (7) and (8) suggest we choose

$$h_{\mu}(s) = h_{\mu,\,R}(s) = \frac{(s-1)^2 + R^2}{R^2} \mu^{s-1}.$$

Then we have

$$G(1) - G_\mu(1) = \frac{1}{2\pi i}\int_{\gamma_1}\big(G(s) - G_\mu(s)\big)\frac{h_\mu(s)}{s-1}ds$$
$$- \frac{1}{2\pi i}\int_{\gamma_2} G_\mu(s)\frac{h_\mu(s)}{s-1}ds + \frac{1}{2\pi i}\int_{\gamma_2} G(s)\frac{h_\mu(s)}{s-1}ds$$
$$=: I_1 - I_2 + I_3.$$

Applying (7) we see

$$\left|\big(G(s) - G_\mu(s)\big)\frac{h_\mu(s)}{s-1}\right| \leqslant \frac{2}{R^2}$$

on γ_1 (by continuity, this estimate also holds for $s = 1 \pm iR$). The standard estimate now yields for all μ

$$|I_1| \leqslant \frac{1}{R}.$$

To estimate I_2 we note that the integrand is holomorphic on $\mathbb{C} \setminus \{1\}$. By Cauchy's integral theorem we can replace the path γ_2 by the semicircle $|s-1| = R$, $\operatorname{Re} s \leqslant 1$ without changing the integral. On the semicircle we have by (8)

$$\left|G_\mu(s)\frac{h_\mu(s)}{s-1}\right| \leqslant \frac{2}{R^2},$$

therefore $|I_2| \leqslant \dfrac{1}{R}$ for all μ.

It remains to investigate

$$I_3 = \frac{1}{2\pi i}\int_{\gamma_2} G(s)\left(\frac{s-1}{R^2} + \frac{1}{s-1}\right)\mu^{s-1}ds.$$

G is clearly bounded on γ_2, thus

$$\left|G(s)\left(\frac{s-1}{R^2} + \frac{1}{s-1}\right)\right| \leqslant A(R,\delta) = A.$$

The standard estimate then gives

$$|I_3| \leqslant \frac{1}{2\pi}L(\gamma_2)A \cdot \sup_{\gamma_2}\mu^{\sigma-1} \leqslant \frac{RA}{2}, \tag{9}$$

which is of no use. But, if we choose $\delta_1 \in]0,\ \delta]$ and denote the part of γ_2 in $\operatorname{Re} s \leqslant 1-\delta_1$ by γ_2', we get in place of (9)

$$\left|\frac{1}{2\pi i}\int_{\gamma_2'} G(s)\left(\frac{s-1}{R^2} + \frac{1}{s-1}\right)\mu^{s-1}ds\right| \leqslant \frac{RA}{2}\cdot \sup_{\gamma_2'}\mu^{\sigma-1} = \frac{RA}{2}\mu^{-\delta_1} \tag{10}$$

and this bound is small for μ large (R, δ, δ_1 fixed). The integral over the arcs of $|s-1| = R$ above and below $[1-\delta_1, 1]$ remains to be accounted for. These arcs have length $\leqslant \pi\delta_1/2$; using $|\mu^{s-1}| \leqslant 1$ we get

$$\left| \frac{1}{2\pi i} \int_{\gamma_2 - \gamma_2'} G(s) \left(\frac{s-1}{R^2} + \frac{1}{s-1} \right) \mu^{s-1} ds \right| \leqslant \frac{1}{2\pi} \cdot \pi \delta_1 A. \tag{11}$$

Now the claim follows: Given $\varepsilon > 0$, choose $R = \dfrac{1}{\varepsilon}$ and a suitable $\delta = \delta(R)$. Then $|I_1| \leqslant \varepsilon$, $|I_2| \leqslant \varepsilon$, and we may take δ_1 so small that the right hand side of (11) becomes $\leqslant \varepsilon$. Finally, we choose μ_0 so large that the right hand side of (10) is $\leqslant \varepsilon$ for $\mu \geqslant \mu_0$. Then $|I_3| \leqslant 2\varepsilon$, so that for $\mu \geqslant \mu_0$

$$\big| G(1) - G_\mu(1) \big| \leqslant |I_1| + |I_2| + |I_3| \leqslant 4\varepsilon. \qquad \square$$

Historical Note

The classical proof that there are infinitely many primes is in Euclid (about -300). The first statement on the distribution of primes seems to be due to Euler (1737): Using the Euler product (2) he showed the divergence of $\sum 1/p$. On the basis of numerical evidence, Gauss conjectured $\pi(x) \sim x/\log x$ in 1792; in later years, he and Legendre proposed similar, but numerically better approximations (without proofs).

Riemann was the first to consider $\zeta(s)$ as a holomorphic function on $\operatorname{Re} s > 1$. He extended it to a meromorphic function on $\mathbb{C}$ (see the next section) and in 1859 found deep connections between the distribution of prime numbers and the zeros of $\zeta(s)$, of which complete proofs were only given towards the end of the 19^{th} century.

About 1850, Chebyshev proved our Prop. 2.5 and, as a first approximation to the prime number theorem, the estimate $0.92 \leqslant \vartheta(x)/x \leqslant 1.11$ for x sufficiently large. The first proofs of the theorem itself were only given in 1896 by Hadamard and de la Vallée-Poussin, independently of each other. They showed our Prop. 2.4 and then worked with very subtle estimates of $\zeta(s)$ for $\operatorname{Re} s \leqslant 1$.

The proof presented here is very much simpler than the original ones. The essential trick is done by Prop. 2.8 (or Thm. 2.10) which are due to D. J. Newman [Ne].

There are "elementary" (but by no means simple) proofs, i.e. proofs not making use of complex analysis. The first of these were published by P. Erdős and A. Selberg in 1949, again independently of each other.

Exercises

1. Show that $\sum 1/p$ diverges (sum over all prime numbers p).
2. Deduce from Prop. 2.3:

$$\lim_{s \to 1} \left(\zeta(s) - \frac{1}{s-1} \right) = \gamma,$$

   where γ is Euler's constant.

3. Let $(a_n)_{n\geqslant 1}$ be a sequence of complex numbers, satisfying an estimate $|a_n| \leqslant Cn^k$; define $f(x) = \sum\limits_{n\leqslant x} a_n$ for $x \geqslant 1$. Show

$$\mathcal{M}_f(s) = \sum_{1}^{\infty} a_n n^{-s},$$

both sides being holomorphic for $\operatorname{Re} s > k+1$.

4. Define $\Lambda\colon \mathbb{N} \to \mathbb{R}$ by $\Lambda(n) = \log p$, if n is a power of a prime p, $\Lambda(n) = 0$ otherwise. For $\psi(x) = \sum\limits_{n\leqslant x} \Lambda(n)$ show $\mathcal{M}_\psi(s) = -\zeta'(s)/\zeta(s)$.

5. Gauss proposed the approximation $\operatorname{Li}(x) = \displaystyle\int_2^x \frac{dt}{\log t}$ of $\pi(x)$. Show $\operatorname{Li}(x) \sim \dfrac{x}{\log x}$.

   Hint: an upper estimate for Li may e.g. be obtained by partial integration of $\displaystyle\int_{\sqrt{x}}^x 1\cdot\frac{1}{\log t}dt$.

3. The functional equation of the ζ-function

In the previous section we defined the Riemann ζ-function on the half plane $\{s = \sigma + it \in \mathbb{C} : \sigma > 1\}$ by

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}.$$

We proved that it can be extended meromorphically to the half plane $\sigma > 0$. Now we will show how it can be extended to a meromorphic function on the whole plane. Then we will deduce a remarkable functional equation for $\zeta(s)$. We follow the ideas of Riemann (1859).

In section 1 we saw that, for $\sigma > 0$,

$$\Gamma(s) = \int_0^\infty e^{-t} t^{s-1} dt = n^s \int_0^\infty e^{-nx} x^{s-1} dx,$$

(the last equation results by substituting $nx = t$). Thus

$$\zeta(s)\Gamma(s) = \sum_{n=1}^{\infty} \int_0^\infty e^{-nx} x^{s-1} dx = \int_0^\infty \left(\sum_{n=1}^{\infty} e^{-nx} \right) x^{s-1} dx,$$

$$\zeta(s)\Gamma(s) = \int_0^\infty \frac{x^{s-1}}{e^x - 1} dx \tag{1}$$

for $\sigma > 1$. The interchange of summation and integration is justified by the theorem on dominated convergence.

The integral in (1) does not exist for $\sigma \leqslant 1$ because of the singularity at 0.

By modification of the path of integration we shall produce an integral valid for all $s \in \mathbb{C}$. We will modify the integrand, too: On the z-plane slit along the positive real axis, i.e. on $G = \mathbb{C} \setminus \{z \in \mathbb{R} : z \geqslant 0\}$, we consider the function

$$f(z,\, s) = \frac{(-z)^{s-1}}{e^z - 1}, \tag{2}$$

with $(-z)^{s-1} = \exp[(s-1)\log(-z)]$ and $-\pi < \operatorname{Im}\log(-z) < \pi$. As a function of z, f is meromorphic on G with simple poles at $z = 2\pi i\nu$, $\nu \in \mathbb{Z} \setminus \{0\}$; as a function of s, f is holomorphic everywhere if $z \neq 2\pi i\nu$ is fixed. We integrate $f(z,\, s)$ along

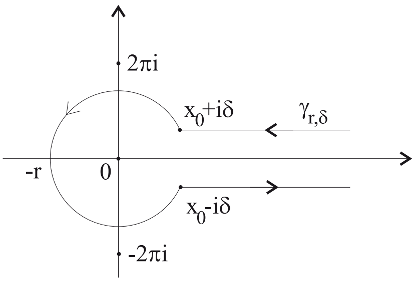


Figure 11.

the path $\gamma_{r,\delta} = \gamma_1 + \gamma_2 + \gamma_3$ shown in Fig. 11. The parameters r, δ are subject to the condition $0 < \delta \leqslant r/2 < \pi/2$; γ_1 is the ray parallel to the x-axis from infinity to $x_0 + i\delta = \sqrt{r^2 - \delta^2} + i\delta$, γ_2 is the positively oriented arc of $|z| = r$ from $x_0 + i\delta$ to $x_0 - i\delta$, and γ_3 is the ray parallel to the x-axis from $x_0 - i\delta$ to infinity. Integrals over the infinite paths γ_1 and γ_3 are defined in the obvious way as limits of integrals over finite paths.

The integral

$$I_{r,\delta}(s) = \int\limits_{\gamma_{r,\delta}} f(z,\, s)dz \tag{3}$$

exists for arbitrary $s \in \mathbb{C}$ since

$$|f(z,\, s)| \leqslant e^{|t|\pi} \frac{|z|^{\sigma-1}}{|e^z - 1|}. \tag{4}$$

The function $I_{r,\delta}(s)$ is holomorphic on $\mathbb{C}$: If we denote the part of $\gamma_{r,\delta}$ to the left of $\operatorname{Re} s = n$ by $\gamma^{(n)}$, then $I^{(n)}_{r,\delta}(s) = \int_{\gamma^{(n)}} f(z,\, s)dz$ is in $\mathcal{O}(\mathbb{C})$ by Prop. I.5.6. Using (4) one verifies easily that, for $n \to \infty$, the $I^{(n)}_{r,\delta}(s)$ converge locally uniformly – of course to $I_{r,\delta}(s)$. The Weierstrass convergence theorem yields the claim.

Moreover, $I_{r,\delta}(s)$ does not depend on the choice of parameters: If e.g. $r' < r$ and $\delta' < \delta$, there are no singularities of f between $\gamma_{r,\delta}$ and $\gamma_{r',\delta'}$. Since $f(z, s)$ tends exponentially to 0 for $\operatorname{Re} z \to +\infty$ (s fixed), the Cauchy integral theorem gives

$$I_{r,\delta}(s) = I_{r',\delta'}(s) =: I(s).$$

On the other hand, we will show

$$\begin{aligned}\lim_{r\to 0}\lim_{\delta\to 0} I_{r,\delta}(s) &= \left(e^{(s-1)\pi i} - e^{-(s-1)\pi i}\right)\int_0^\infty \frac{x^{s-1}}{e^x - 1}dx \\ &= -2i\sin(s\pi)\int_0^\infty \frac{x^{s-1}}{e^x-1}dx \qquad (5)\end{aligned}$$

if $\sigma > 1$. Inserting (1) and using $\sin(s\pi)\Gamma(s)\Gamma(1-s) = \pi$ we arrive at

Proposition 3.1. *The equation*

$$\zeta(s) = -\Gamma(1-s)\cdot\frac{1}{2\pi i}I(s) \qquad (6)$$

holds for $\operatorname{Re} s > 1$.

Proof: We have to verify (5). By our choice of the power $(-z)^{s-1}$ we have

$$\lim_{\delta\downarrow 0} f(x \pm i\delta, s) = \exp\bigl(\mp(s-1)\pi i\bigr)\frac{x^{s-1}}{e^x-1}.$$

Therefore

$$\begin{aligned}\lim_{\delta\downarrow 0}\int_{\gamma_1} f(z,\, s)dz &= -\lim_{\delta\downarrow 0}\int_{x_0}^\infty f(x+i\delta,\, s)dx \\ &= -\int_r^\infty \lim_{\delta\downarrow 0} f(x+i\delta,\, s)dx \\ &= -e^{-(s-1)\pi i}\int_r^\infty \frac{x^{s-1}}{e^x-1}dx.\end{aligned}$$

Similarly, we obtain

$$\lim_{\delta\downarrow 0}\int_{\gamma_3} f(z,\, s)dz = e^{(s-1)\pi i}\int_r^\infty \frac{x^{s-1}}{e^x-1}dx.$$

The interchange of limit and integration is legitimate, because by (4)

$$|f(x \pm i\delta,\, s)| \leqslant \text{const} \cdot \frac{x^{\sigma-1}}{e^x - 1}$$

for $x \geqslant r$ and $0 \leqslant \delta \leqslant r/2$.

Putting things together, we see

$$\lim_{\delta \downarrow 0} I_{r,\delta}(s) = 2i \sin(s-1)\pi \cdot \int_r^\infty \frac{x^{s-1}}{e^x - 1} dx + \int_{\kappa_r(0)} \frac{(-z)^{s-1}}{e^z - 1} dz. \tag{7}$$

In view of taking the limit $r \to 0$, we now assume $\operatorname{Re} s > 1$. Then

$$\int_r^\infty \frac{x^{s-1}}{e^x - 1} dx \to \int_0^\infty \frac{x^{s-1}}{e^x - 1} dx, \qquad \int_{\kappa_r(0)} \frac{(-z)^{s-1}}{e^z - 1} dz \to 0,$$

the last limit by the standard estimate. – This proves (5). □

Let us review equation (6) in Prop. 3.1. The function $\Gamma(s)$ is meromorphic on the whole plane and $I(s)$ is an entire function. So, the right hand side of (6) is a meromorphic function on $\mathbb{C}$. Therefore we may *define* $\zeta(s)$ on $\operatorname{Re} s \leqslant 1$ by this equation, thus extending the ζ-function as a meromorphic function to all of $\mathbb{C}$.

The poles of $\Gamma(1-s)$ are situated at $s = 1, 2, 3, \ldots$, since $\Gamma(s)$ has poles precisely at $s = 0, -1, -2, \ldots$. As $\zeta(s)$ is holomorphic for $\operatorname{Re} s > 1$, the poles at $s = 2, 3, \ldots$ of $\Gamma(1-s)$ must cancel against zeros of $I(s)$. On the other hand, the simple pole of $\Gamma(1-s)$ at 1 gives rise to a simple pole of $\zeta(s)$ at $s = 1$, since by Cauchy's integral theorem

$$\frac{1}{2\pi i} I(1) = \frac{1}{2\pi i} \int_{\gamma_{r,\delta}} \frac{dz}{e^z - 1} = \frac{1}{2\pi i} \int_{|z|=1} \frac{dz}{e^z - 1} = 1 \neq 0.$$

Observing $\operatorname{res}_{s=1}\ \Gamma(1-s) = -1$, we get

Proposition 3.2. *The function $\zeta(s)$, extended to $\mathbb{C}$ by* (6)*, is holomorphic everywhere save for a simple pole at $s = 1$ with residue* 1. □

The values $\zeta(-n)$, $n = 0, 1, 2, \ldots$, can be determined from the definition: We have by (7),

$$\frac{1}{2\pi i} I(-n) = \frac{1}{2\pi i} \lim_{\delta \downarrow 0} I_{r,\delta}(-n) = \frac{1}{2\pi i} \int_{|z|=r} \frac{(-z)^{-n-1}}{e^z - 1} dz,$$

and thus, by (6), inserting $\Gamma(n+1) = n!$,

$$\zeta(-n) = (-1)^n n!\Big[\frac{1}{2\pi i}\int\limits_{|z|=r}\frac{z^{-n-1}}{e^z-1}dz\Big].$$

Since $r < 2\pi$, the term in brackets is the coefficient of z^n in the Laurent expansion of $(e^z-1)^{-1}$ about zero. In III.6.10 we have written down this expansion in terms of Bernoulli numbers:

$$\frac{1}{e^z-1} = \frac{1}{z} - \frac{1}{2} + \sum_{\nu=1}^{\infty}\frac{B_{2\nu}}{(2\nu)!}z^{2\nu-1}.$$

This yields

Proposition 3.3. *We have $\zeta(0) = -1/2$, $\zeta(-n) = 0$ for $n > 0$ even, and $\zeta(-n) = -B_{n+1}/(n+1)$ for $n > 0$ odd.* □

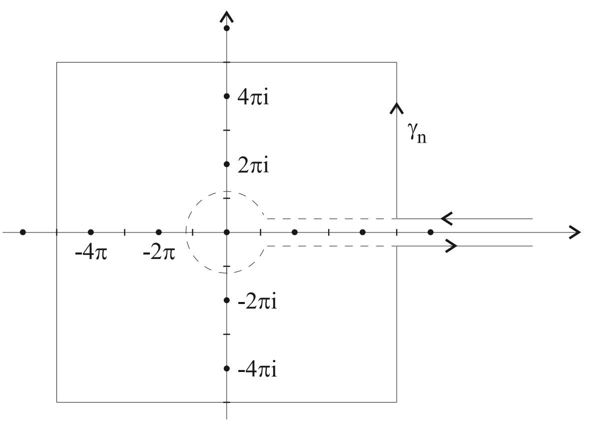


Figure 12.

Now we shall deduce the functional equation. To this end, we consider not only the path $\gamma = \gamma_{r,\delta}$, but also paths γ_n we obtain by replacing the part of γ to the left of $\operatorname{Re} z = (2n+1)\pi$ with the positively oriented boundary of the square S_n with vertices $(2n+1)\pi(\pm 1 \pm i)$, omitting, of course, the segment $[(2n+1)\pi - i\delta, (2n+1)\pi + i\delta]$ – cf. Fig. 12. The path $\gamma_n - \gamma$ encloses the poles $\pm 2\pi i\nu$, $\nu = 1, \ldots, n$, of the integrand $f(z) = (-z)^{s-1}/(e^z-1)$ (we keep s fixed).

By the residue theorem

$$\frac{1}{2\pi i}\int\limits_{\gamma_n} f(z)dz - \frac{1}{2\pi i}\int\limits_{\gamma} f(z)dz = \sum_{\substack{-n\leqslant\nu\leqslant n\\ \nu\neq 0}} \operatorname{res}_{2\pi i\nu} f. \tag{8}$$

Now the residue at $2\pi i\nu$ is $(-2\pi i\nu)^{s-1}$; since $\arg(-i) = -\pi/2$, we obtain the value $(2\pi\nu)^{s-1}\exp(-i\pi(s-1)/2)$ for $\nu>0$; similarly $(-2\pi i\nu)^{s-1}=(2\pi|\nu|)^{s-1}\exp(i\pi(s-1)/2)$ for $\nu < 0$. Therefore, we may write the sum in (8) as

$$\sum_{\nu=1}^{n}(2\pi\nu)^{s-1} \cdot 2\cos\frac{\pi(s-1)}{2} = 2^s\pi^{s-1}\sin\frac{\pi s}{2}\sum_{\nu=1}^{n}\nu^{s-1}.$$

Assuming $\operatorname{Re} s < 0$, the last sum tends to $\zeta(1-s)$ for $n \to \infty$.

On the other hand, we shall show that for $\operatorname{Re} s < 0$

$$\lim_{n\to\infty}\int_{\gamma_n} f(z)dz = 0. \tag{9}$$

Thus (8) implies

$$-\frac{1}{2\pi i}I(s) = 2^s\pi^{s-1}\zeta(1-s)\sin\frac{\pi s}{2}$$

for $\operatorname{Re} s < 0$. Combining this with (6), we obtain the claim of the following theorem in case $\operatorname{Re} s < 0$; the case $\operatorname{Re} s \geqslant 0$ then follows by the identity theorem.

Theorem 3.4. *The following functional equation holds throughout* $\mathbb{C}$*:*

$$\zeta(s) = \Gamma(1-s)\zeta(1-s)2^s\pi^{s-1}\sin\frac{\pi s}{2}. \tag{10}$$

Proof: It remains to verify (9). The denominator of $f(z) = (-z)^{s-1}/(e^z - 1)$ is bounded away from zero on the edges of the square S_n:

$$\begin{aligned}
&\left|e^z - 1\right| = \left|-e^x - 1\right| \geqslant 1 \text{ on the horizontal edges,}\\
&\left|e^z - 1\right| \geqslant e^{(2n+1)\pi} - 1 \geqslant e^{3\pi} - 1 \text{ on the right hand edge,}\\
&\left|e^z - 1\right| \geqslant 1 - e^{-(2n+1)\pi} \geqslant 1 - e^{-3\pi} \text{ on the left hand edge.}
\end{aligned}$$

Furthermore, $|z| \geqslant \text{const} \cdot n$ on the boundary of S_n, so for $\sigma = \operatorname{Re} s < 0$

$$\left|(-z)^{s-1}\right| = |z|^{\sigma-1}e^{-t\arg(-z)} \leqslant \text{const} \cdot n^{\sigma-1}.$$

Therefore, by the standard estimate

$$\left|\int_{\gamma_n'} f(z)dz\right| \leqslant \text{const} \cdot n^{\sigma} \to 0,$$

where γ_n' denotes the part of γ_n on the boundary of S_n.

The integrals of f over the infinite horizontal segments $]+\infty+i\delta, (2n+1)\pi+i\delta]$ and $[(2n+1)\pi - i\delta, +\infty - i\delta[$ tend to zero too, as $\int_\gamma f(z)dz$ exists. □

We can achieve a more symmetrical form of the functional equation if we make use of Legendre's duplication formula (Prop. 1.5) in the form

$$\Gamma(1-s) = \pi^{-1/2}2^{-s}\Gamma\Big(\frac{1-s}{2}\Big)\Gamma\Big(1-\frac{s}{2}\Big)$$

and of the equation

$$\Gamma\Big(\frac{s}{2}\Big)\Gamma\Big(1-\frac{s}{2}\Big)\sin\frac{\pi s}{2} = \pi.$$

Inserting this into (10), an easy computation yields

Corollary 3.5.

$$\pi^{-s/2}\Gamma\Big(\frac{s}{2}\Big)\zeta(s) = \pi^{-(1-s)/2}\Gamma\Big(\frac{1-s}{2}\Big)\zeta(1-s).$$

To conclude this section, let us have a look at the zeros of $\zeta(s)$. The Euler product shows $\zeta(s) \neq 0$ for $\operatorname{Re} s > 1$, by Prop. 2.4 there are no zeros on the line $\operatorname{Re} s = 1$. Now apply the functional equation (10) for $\operatorname{Re} s \leqslant 0$, i.e. $\operatorname{Re}(1-s) \geqslant 1$. The only zeros of the right hand side stem from the sine factor, so the only zeros of $\zeta(s)$ in the closed left half plane are the "trivial zeros" $s = -2, -4, -6, \ldots$ (cf. Prop. 3.3).

In other words, all "non-trivial zeros" of $\zeta(s)$ must be contained in the "critical strip"

$$\{s = \sigma + it : 0 < \sigma < 1\}.$$

The statement $\zeta(s) \neq 0$ for $\sigma = 1$ is essential in all analytical proofs of the Prime Number Theorem. More is true: Explicit knowledge of "large" zero-free regions to the left of $\sigma = 1$ leads to "good" estimates of the deviation

$$\pi(x) - \frac{x}{\log x} \qquad \text{or} \qquad \pi(x) - \operatorname{Li}(x)$$

(for details we must refer the reader to treatises on analytic number theory). Therefore the situation of the zeros of $\zeta(s)$ is of eminent number theoretical interest.

Already in his 1859 paper, Riemann stated his famous conjecture that there are infinitely many non-trivial zeros and all of them lie on the line $\sigma = \frac{1}{2}$. This conjecture – the *Riemann Hypothesis* – has been corroborated by many theoretical and numerical investigations, but up to now it remains unproven. Anyhow, Hardy showed in 1914 that an infinity of zeros lie on this line.

If the Riemann Hypothesis is correct, then

$$\left|\pi(x) - \int_2^x \frac{dt}{\log t}\right| \leqslant C x^{1/2} \log x \tag{11}$$

(cf. Ex. 2.5). Conversely, (11) implies the Riemann Hypothesis.

Exercises

1. If s is a non-real zero of the ζ-function, then so are $\bar{s}$, $1-s$, $1-\bar{s}$.
2. Let $\xi(s) = s(s-1)\pi^{-s/2}\Gamma\big(\frac{s}{2}\big)\zeta(s)$. Show that $\xi(s)$ is an entire function whose zeros are precisely the non-real zeros of $\zeta(s)$. Moreover, $\xi(1-s) = \xi(s)$, and ξ is real-valued on the line $\sigma = \frac{1}{2}$.

4. Elliptic functions

We consider functions f defined on the complex plane. A complex number ω is called a *period* of f if $f(z+\omega) \equiv f(z)$. Every function has the trivial period 0; the periods of a given function f form a subgroup G_f of the additive group $\mathbb{C}$. A constant function has $G_f = \mathbb{C}$, but the period group of a non constant *meromorphic* function f is a discrete subset of $\mathbb{C}$: If there were a convergent sequence in G_f, say $\omega_\nu \to \omega_0$, then for an arbitrary z_0, $f(z_0) = f(z_0 + \omega_\nu) = f(z_0 + \omega_0)$ would hold – the last equality by the continuity of f – and f would be constant by the identity theorem.

Now the converse question arises: Given a discrete subgroup $G \neq \{0\}$ of $\mathbb{C}$, are there non constant meromorphic functions whose period group is (or contains) G? The simplest examples of such subgroups are the infinite cyclic groups $\mathbb{Z}\omega$, $\omega \in \mathbb{C} \setminus \{0\}$, which are the period groups of $f(z) = \exp(2\pi iz/\omega)$. There is only one type of non cyclic discrete subgroup of $\mathbb{C}$, namely the lattices (cf. Ex. 1):

Definition 4.1. *A lattice Ω in $\mathbb{C}$ is an additive subgroup of $\mathbb{C}$ of the form*

$$\Omega = \{m\omega_1 + n\omega_2 : m, n \in \mathbb{Z}\},$$

where $\omega_1, \omega_2 \in \mathbb{C}$ are linearly independent over $\mathbb{R}$.

We call (ω_1, ω_2) a basis of Ω and write $\Omega = \langle\omega_1, \omega_2\rangle$. Of course, a given lattice has many bases, e.g. $\langle\omega_1 + \omega_2, \omega_2\rangle = \langle\omega_1, \omega_2\rangle$ – see Ex. 2.

If $\Omega = \langle\omega_1, \omega_2\rangle$ is a lattice, ω_1 and ω_2 do not lie on a straight line through 0, hence they span a parallelogram

$$P = P(\omega_1, \omega_2) = \{z = t_1\omega_1 + t_2\omega_2 : 0 \leqslant t_1, t_2 < 1\},$$

the (half-open) *period-parallelogram*. The translates $\omega + P = \{\omega + z : z \in P\}$, $\omega \in \Omega$, tessellate the plane: Every $z \in \mathbb{C}$ has a unique representation $z = \omega + z'$ with $\omega \in \Omega$ and $z' \in P$.

Definition 4.2. *An elliptic function with respect to the lattice Ω is a meromorphic function f on $\mathbb{C}$ with*

$$f(z+\omega) \equiv f(z) \qquad \textit{for all } \omega \in \Omega. \tag{1}$$

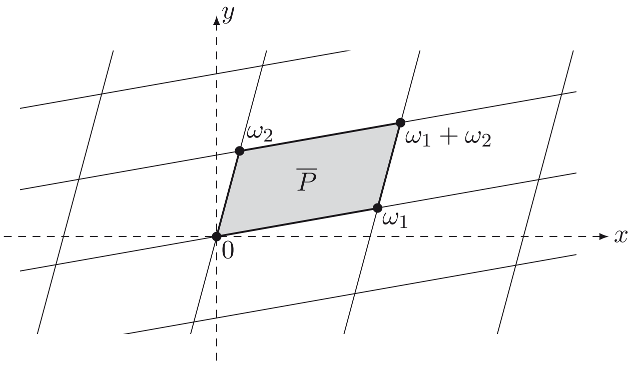


Figure 13. Lattice with period parallelogram

The period group G_f thus contains Ω, but may be larger. If $\Omega = \langle\omega_1, \omega_2\rangle$, then (1) follows from $f(z+\omega_1) \equiv f(z) \equiv f(z+\omega_2)$. Any value assumed by f is assumed in P, and also in any $a + P = \{a + z : z \in P\}$.

In the sequel, let $\Omega = \langle\omega_1, \omega_2\rangle$ denote a fixed lattice; "elliptic" will mean "elliptic with respect to Ω". We first prove some general statements on elliptic functions; the existence of non constant such functions will be established later.

Proposition 4.1. *The elliptic functions constitute a field $K(\Omega)$ containing $\mathbb{C}$ (as field of constant functions). If $f \in K(\Omega)$, then $f' \in K(\Omega)$.*

Proof: This follows immediately from (1). □

Proposition 4.2. *The only holomorphic elliptic functions are the constants.*

Proof: A holomorphic elliptic function is bounded on the compact closure of the period parallelogram, hence on $\mathbb{C}$. Liouville's theorem yields the claim. □

In other words: A non constant elliptic function must have poles. The number of poles in P is, of course, finite.

Proposition 4.3. *Let f be elliptic with poles $z_1, \ldots, z_n$ in P. Then*

$$\sum_1^n \operatorname{res}_{z_\nu} f = 0.$$

Hence a non constant $f \in K(\Omega)$ has at least two poles in P (counting multiplicity).

Proof of Prop. 4.3**:** Assume $\Omega = \langle \omega_1, \omega_2 \rangle$ with $\mathrm{Im}(\omega_2/\omega_1) > 0$. Then $\overset{\circ}{P}$ is a domain with positive boundary

$$\partial P = [0, \omega_1] + [\omega_1, \omega_1 + \omega_2] + [\omega_1 + \omega_2, \omega_2] + [\omega_2, 0].$$

We suppose no pole of f is on ∂P and apply the residue theorem:

$$\begin{aligned} 2\pi i \sum_1^n \mathrm{res}_{z_\nu} \; f &= \int_{\partial P} f(z)\,dz \\ &= \int_{[0,\,\omega_1]} f(z)\,dz + \int_{[\omega_1,\,\omega_1+\omega_2]} f(z)\,dz - \int_{[\omega_2,\,\omega_1+\omega_2]} f(z)\,dz - \int_{[0,\,\omega_2]} f(z)\,dz. \end{aligned}$$

As $f(z + \omega_1) = f(z) = f(z + \omega_2)$, the first integral cancels against the third and the second against the fourth: the sum is 0. – If some of the poles are on ∂P, we choose a near 0 such that the z_ν are in the interior of $a + P$, and integrate over $\partial(a + P)$. □

Proposition 4.4. *A non constant elliptic function f assumes every $w \in \widehat{\mathbb{C}}$ in P equally often (counting multiplicity).*

Proof: Given w, choose $a + P$ such that $f(z) \neq w, \infty$ on $\partial(a + P)$. The proof of Prop. 4.3 shows, since $f'(z)/(f(z) - w) \in K(\Omega)$,

$$\frac{1}{2\pi i} \int_{\partial(a+P)} \frac{f'(z)dz}{f(z) - w}\, dz = 0.$$

By Cor. IV.6.2, the left hand side is the difference of the number of points with $f(z) = w$ and the number of poles of f in $a + P$ (counting multiplicity). □

Definition 4.3. *The order of an elliptic function is the number of its poles in P (counting multiplicity).*

It is easily seen that this number depends only on Ω and not on the choice of the basis (ω_1, ω_2).

The device used to prove the last propositions also yields

Proposition 4.5. *Let $a_1, \ldots, a_k$ be the different zeros in P of the elliptic function f, with multiplicities $m_1, \ldots, m_k$, and let $b_1, \ldots, b_\ell$ be the different poles of f in P, with multiplicities $n_1, \ldots, n_\ell$. Then*

$$\sum_{\varkappa=1}^{k} m_\varkappa a_\varkappa - \sum_{\lambda=1}^{\ell} n_\lambda b_\lambda \in \Omega.$$

Proof: We assume that none of $a_\varkappa$, b_λ is on ∂P (otherwise we replace P by a suitable $a+P$). Then $g(z) = z\dfrac{f'(z)}{f(z)}$ has no poles on ∂P. The only singularities of g in P are simple poles, viz. at the $a_\varkappa$ with residues $m_\varkappa a_\varkappa$ and at the b_λ with residues $-n_\lambda b_\lambda$. Thus

$$\sum m_\varkappa a_\varkappa - \sum n_\lambda b_\lambda = \frac{1}{2\pi i}\int\limits_{\partial P} \frac{zf'(z)}{f(z)}\,dz.$$

Now consider the opposite segments $[0, \omega_1]$ and $[\omega_1+\omega_2, \omega_2]$ of ∂P. Since f is elliptic, one has

$$\frac{(z+\omega_2)f'(z+\omega_2)}{f(z+\omega_2)} - \frac{zf'(z)}{f(z)} = \omega_2\frac{f'(z)}{f(z)}.$$

Therefore

$$\frac{1}{2\pi i}\Bigg(\int\limits_{[0,\,\omega_1]} g(z)\,dz + \int\limits_{[\omega_1+\omega_2,\,\omega_2]} g(z)\,dz\Bigg) = -\frac{\omega_2}{2\pi i}\int\limits_{[0,\,\omega_1]}\frac{f'(z)}{f(z)}\,dz.$$

Since $f(0) = f(\omega_1)$, f maps the segment $[0, \omega_1]$ onto a closed path γ and

$$\frac{\omega_2}{2\pi i}\int\limits_{[0,\,\omega_1]}\frac{f'(z)}{f(z)}\,dz = \frac{\omega_2}{2\pi i}\int\limits_{\gamma}\frac{dw}{w} = \omega_2\cdot n(\gamma, 0) \in \mathbb{Z}\omega_2.$$

Similarly, integration over the remaining segments of ∂P leads to an integral multiple of ω_1. □

For instance, if f is of order 3 and has a triple pole at the origin, it must have three zeros $z_1, z_2, z_3 \in P$ satisfying $z_1 + z_2 + z_3 \in \Omega$ (of course, some of the zeros may coincide).

We now construct elliptic functions by means of partial fraction series. We have the obvious candidates

$$\sum_{\omega\in\Omega}\frac{1}{(z-\omega)^k} \tag{2}$$

for $k = 2, 3, \ldots$: if (2) converges to a meromorphic function f_k, we have for any $\omega_0 \in \Omega$

$$f_k(z+\omega_0) = \sum_{\omega\in\Omega}\frac{1}{\big(z-(\omega-\omega_0)\big)^k} = f_k(z),$$

as $\omega - \omega_0$ ranges over all of Ω if ω does.

Proposition 4.6. *The series* (2) *converge absolutely locally uniformly for* $k \geqslant 3$.

The proof depends on

Proposition 4.7. *The series* $\sum_{\omega\in\Omega\setminus\{0\}} \omega^{-k}$ *converges absolutely for* $k > 2$.

Summation over all lattice points $\neq 0$ occurs here and frequently in the sequel; we write for short

$$\sum{}' = \sum_{\omega\in\Omega}{}' = \sum_{\omega\in\Omega\setminus\{0\}} .$$

Proof of Prop. 4.7**:** Let $\Omega = \langle\omega_1, \omega_2\rangle$ and denote P_ℓ the parallelogram with vertices $\pm\ell\omega_1 \pm \ell\omega_2$, $\ell = 1, 2, 3, \ldots$. Then $\delta := \operatorname{dist}(\partial P_1, 0) > 0$ and $\operatorname{dist}(\partial P_\ell, 0) = \ell\delta$. The lattice points $\omega \in \partial P_\ell$ are 8ℓ in number, they satisfy $|\omega|^k \geqslant (\ell\delta)^k$. Hence, for $k > 2$ we have

$$\sum{}' |\omega|^{-k} = \sum_{\ell=1}^{\infty} \sum_{\omega\in\partial P_\ell} |\omega|^{-k} \leqslant 8\delta^{-k} \sum_{\ell=1}^{\infty} \ell^{-k+1} < +\infty. \qquad \square$$

Proof of Prop. 4.6**:** Given $R > 0$, there are only finitely many $\omega \in \Omega$ with $|\omega| < 2R$. For $|\omega| > 2R$ and $|z| \leqslant R$, the estimate

$$|z - \omega| \geqslant |\omega| - |z| \geqslant |\omega|/2$$

holds, hence, for $|z| \leqslant R$ and $k \geqslant 3$,

$$\sum_{|\omega|\geqslant 2R} |z - \omega|^{-k} \leqslant 2^k \sum_{|\omega|\geqslant 2R} |\omega|^{-k} < +\infty.$$

In view of Def. III.6.1, this yields the claim. $\square$

Prop. 4.6 exhibits elliptic functions of order k for $k = 3, 4, \ldots$. According to Prop. 4.3, a function g of order 2 may possibly exist. As (2) diverges for $k = 2$, we try to construct such a g by integration of f_3 (we could also employ the Mittag-Leffler theorem).

The function $\sum{}'(z - \omega)^{-3}$ is holomorphic on $G = (\mathbb{C} \setminus \Omega) \cup \{0\}$, all its residues are zero. Integrating it along a path in G from 0 to z yields as a primitive the absolutely and locally uniformly convergent series

$$-\frac{1}{2}\sum{}'\left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\right).$$

Hence

$$g(z) = -\frac{1}{2}\left(\frac{1}{z^2} + \sum{}'\left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\right)\right) \tag{3}$$

is a primitive of f_3 on $\mathbb{C}\setminus\Omega$ which is meromorphic on $\mathbb{C}$, having poles in the lattice points, each of order 2. It is, however, not clear whether g is periodic; the periodicity of f_3 translates to

$$g'(z+\omega)-g'(z)\equiv 0$$

for every $\omega\in\Omega$. This implies $g(z+\omega)-g(z)=C(\omega)$ with a constant $C(\omega)$. Taking a basis $(\omega_1,\,\omega_2)$ of Ω and setting $z=-\omega_j/2,\ j=1,\,2,$ we get

$$g(\omega_j/2)-g(-\omega_j/2)=C(\omega_j),\quad j=1,\,2.$$

But g is clearly an even function, hence $C(\omega_1)=C(\omega_2)=0$ and the periodicity of g follows.

We introduce the standard notation:

Definition 4.4. *The Weierstrass $\wp$-function of the lattice Ω is*

$$\wp(z)=\frac{1}{z^2}+\sum_{\omega\in\Omega}{}'\left(\frac{1}{(z-\omega)^2}-\frac{1}{\omega^2}\right). \tag{4}$$

By construction, $\wp$ is an even elliptic function of order 2 with double poles at the lattice points. Its derivative,

$$\wp'(z)=-2\sum_{\Omega}(z-\omega)^{-3}, \tag{5}$$

is an odd elliptic function of order 3.

In a way, $\wp$ is the simplest elliptic function, but together with its derivative, it generates the field $K(\Omega)$ of all elliptic functions! As any $f\in K(\Omega)$ is the sum of an even and an odd function in $K(\Omega)$:

$$f(z)=\frac{1}{2}\big(f(z)+f(-z)\big)+\frac{1}{2}\big(f(z)-f(-z)\big),$$

the claim is a consequence of the following

Proposition 4.8.

i. Every even elliptic function f is a rational function of $\wp$: $f=R(\wp)$ with $R(X)\in\mathbb{C}(X)$.

ii. Every odd elliptic function g can be written in the form $g=\wp'\cdot R(\wp)$ with $R(X)\in\mathbb{C}(X)$.

Proof: If $g\in K(\Omega)$ is odd, then $g/\wp'$ is even. Hence *ii* follows from *i*.

To prove *i* we first treat the case of an even $f\in K(\Omega)$ with poles $z_1,\,\dots,\,z_k$ in $P\setminus\{0\}$. As $\wp(z)-\wp(z_j)$ vanishes in z_j, the function $\big(\wp(z)-\wp(z_j)\big)^{m_j}f(z)$ will have a removable singularity at z_j for m_j sufficiently large. Hence we can choose exponents m_j such that

$$f_1(z) = \prod_{j=1}^{k} \big(\wp(z) - \wp(z_j)\big)^{m_j} f(z)$$

has no poles in $P \setminus \{0\}$. The proof is then completed by

Lemma 4.9. *Let f_1 be an even elliptic function with poles (if any) only in the lattice points. Then f_1 is a polynomial in $\wp$: $f_1 = a_0 + a_1\wp + \ldots + a_n\wp^n$ with $a_\nu \in \mathbb{C}$.*

Proof: This is clear if f_1 is constant. If not, consider the Laurent expansion of f_1 about 0. As f_1 is even, only even powers of z occur:

$$f_1(z) = b_{-2n}z^{-2n} + \ldots\,, \quad b_{-2n} \neq 0,$$

in particular, f_1 is of even order. Now the Laurent expansion of $\wp(z)$ about 0 begins with z^{-2} by (4). Therefore the coefficient of z^{-2n} in the Laurent expansion of

$$f_2(z) = f_1(z) - b_{-2n}\big(\wp(z)\big)^n$$

vanishes: The even elliptic function f_2 is of lower order than f_1 and has singularities at most in the lattice points. Induction on the order proves the lemma. □

In particular, we can apply the lemma to the function $(\wp')^2$, which is even and of order 6. Accordingly, there is an equation

$$(\wp')^2 = a_0 + a_1\wp + a_2\wp^2 + a_3\wp^3,$$

a differential equation for $\wp$! To explicitly determine the coefficients a_ν as in the proof of the lemma, we need the first few terms of the Laurent expansion of $\wp$ about 0. We write

$$\wp(z) = \frac{1}{z^2} + c_2z^2 + c_4z^4 + \ldots\,; \tag{6}$$

the constant term vanishes by (4). We then have

$$\begin{aligned}
\wp^3(z) &= \frac{1}{z^6} + 3c_2\frac{1}{z^2} + 3c_4 + \ldots\,, \\
\wp'(z) &= \frac{-2}{z^3} + 2c_2z + 4c_4z^3 + \ldots\,, \\
\big(\wp'(z)\big)^2 &= \frac{4}{z^6} - 8c_2\frac{1}{z^2} - 16c_4 + \ldots\,,
\end{aligned}$$

whence

$$\wp'^2(z) - 4\wp^3(z) = -20c_2 z^{-2} - 28c_4 + \dots\ ,$$
$$\wp'^2(z) - 4\wp^3(z) + 20c_2\wp(z) = -28c_4 + \dots\ .$$

The right hand side is a holomorphic elliptic function, hence is the constant $-28c_4$. Thus $\wp$ satisfies the differential equation

$$\wp'^2 = 4\wp^3 - 20c_2\wp - 28c_4.$$

It is desirable to express the coefficients c_2, c_4 or even all $c_{2\nu}$ in (6) in terms of the lattice Ω. To this end we consider

$$h(z) = \sum_1^\infty c_{2\nu} z^{2\nu} = \wp(z) - \frac{1}{z^2} = \sum{}' \left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\right).$$

We have for $\mu \geqslant 1$

$$h^{(\mu)}(z) = (-1)^\mu (\mu+1)! \sum{}' (z-\omega)^{-\mu-2}.$$

Since $(2\nu)!\, c_{2\nu} = h^{(2\nu)}(0)$, the equation

$$c_{2\nu} = (2\nu+1) \sum{}' \omega^{-2\nu-2}. \tag{7}$$

follows. – We summarize:

Theorem 4.10. *The Weierstrass $\wp$-function of the lattice Ω satisfies the differential equation*

$$\wp'^2 = 4\wp^3 - g_2\wp - g_3, \tag{8}$$

with $g_2 = g_2(\Omega) = 60 \sum{}' \omega^{-4}$ *and* $g_3 = g_3(\Omega) = 140 \sum{}' \omega^{-6}$.

The lattice Ω is determined uniquely by the numbers $g_2(\Omega)$ and $g_3(\Omega)$ (cf. Ex. 11), they are therefore called the invariants of Ω.

We note two formulas resulting from (8) by differentiation:

$$\begin{aligned} \wp'' &= 6\wp^2 - \frac{1}{2}g_2 = 6\wp^2 - 10c_2 \\ \wp''' &= 12\wp\wp'. \end{aligned} \tag{9}$$

To factorize the polynomial

$$q(X) = 4X^3 - g_2X - g_3$$

on the right hand side of (8), we have to determine the zeros of $\wp'$. To this end, we fix a basis (ω_1, ω_2) of Ω and define

$$\varrho_1 = \omega_1/2, \quad \varrho_2 = (\omega_1+\omega_2)/2, \quad \varrho_3 = \omega_2/2;$$

these are the points $z \in P$ satisfying $2z \in \Omega$, $z \notin \Omega$.

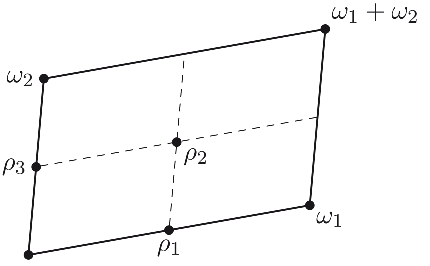


Figure 14. The points ϱ_j

Lemma 4.11. *An odd $g \in K(\Omega)$ has zeros or poles (of odd multiplicity) in $0, \varrho_1, \varrho_2, \varrho_3$. An even $f \in K(\Omega)$ assumes its values in these points with (even) multiplicity > 1.*

Proof: If $z_0 \in \{0,\, \varrho_1,\, \varrho_2,\, \varrho_3\}$ and $g \in K(\Omega)$ is odd with $g(z_0) \neq \infty$, we have, since $2z_0 \in \Omega$,

$$-g(z_0) = g(-z_0) = g(-z_0 + 2z_0) = g(z_0),$$

i.e. $g(z_0) = 0$. For even $f \in K(\Omega)$ the derivative f' is odd, hence $f(z_0) = \infty$ or $f'(z_0) = 0$. – The statements about the parity of the multiplicities are left to the reader. □

In particular, $\wp'$ has zeros in $\varrho_1,\, \varrho_2,\, \varrho_3$; as $\wp'$ is of order 3, these are all simple and are the only zeros of $\wp'$ in P. Setting

$$\wp(\varrho_i) = e_i, \quad i = 1,\, 2,\, 3,$$

the e_i are exactly the values assumed by $\wp$ with multiplicity 2. They are all distinct, since $\wp$ assumes every value only twice in P. Hence $e_1,\, e_2,\, e_3$ are the three distinct zeros of $q(X)$:

$$q(X) = 4(X - e_1)(X - e_2)(X - e_3), \tag{10}$$

and the differential equation of $\wp$ may be rewritten:

Theorem 4.12.

$$\left(\wp'\right)^2 = 4\prod_1^3 (\wp - e_j). \tag{11}$$

By elementary algebra, (10) or (11) imply the equations

$$\begin{aligned} e_1 + e_2 + e_3 &= 0 \\ -4(e_1e_2 + e_2e_3 + e_3e_1) &= g_2 \\ 4e_1e_2e_3 &= g_3. \end{aligned}$$

The discriminant

$$\Delta = 16(e_1 - e_2)^2(e_2 - e_3)^2(e_3 - e_1)^2$$

of $q(X)$ can be computed to

$$\Delta = {g_2}^3 - 27{g_3}^2.$$

As the e_j are all distinct, we have $\Delta \neq 0$.

We mentioned above that the lattice Ω is characterized by its invariants $g_2(\Omega)$, $g_3(\Omega)$. One can show that for any pair of complex numbers g_2, g_3 with ${g_2}^3 - 27{g_3}^2 \neq 0$ there exists a lattice Ω with $g_2 = g_2(\Omega)$, $g_3 = g_3(\Omega)$ – see [FL2, HuC].

As a consequence of the above we have an algebraic description of $K(\Omega)$:

Proposition 4.13. *$K(\Omega)$ is isomorphic to the quotient field of the polynomial ring $\mathbb{C}(X)[Y]$ over the rational function field $\mathbb{C}(X)$ by the ideal generated by*

$$Y^2 - (4X^3 - g_2X - g_3).$$

Namely, the homomorphism of $\mathbb{C}(X)[Y]$ into $K(\Omega)$, given by $X \mapsto \wp$, $Y \mapsto \wp'$, is surjective by Prop. 4.8; its kernel is the given ideal by Thm. 4.10.

We now ask if there are elliptic functions with prescribed zeros and poles. More precisely, let P be the (half-open) period parallelogram of $\Omega = \langle\omega_1, \omega_2\rangle$, let $a_1, \ldots, a_k$, $b_1, \ldots, b_\ell$ be different points of P and $m_1, \ldots, m_k, n_1, \ldots, n_\ell$ positive integers. Does there exist an $f \in K(\Omega)$ with zeros exactly at the $a_\varkappa$ with multiplicities $m_\varkappa$, and poles exactly at the b_λ with multiplicities n_λ? Prop. 4.4 and 4.5 give two necessary conditions:

$$\sum_1^k m_\varkappa = \sum_1^\ell n_\lambda \tag{12}$$

$$\sum_1^k m_\varkappa a_\varkappa - \sum_1^\ell n_\lambda b_\lambda \in \Omega. \tag{13}$$

We will show that these conditions are also sufficient. To this end, we shall construct an entire function $\sigma(z)$, having simple zeros at the points of Ω (and no other zeros), and satisfying the transformation rules

$$\sigma(z + \omega_j) = \exp(\eta_j z + c_j)\sigma(z), \quad j = 1, 2 \tag{14}$$

with constants η_j, c_j.

Replacing, if necessary, one of the $a_\varkappa$'s by an equivalent point mod Ω, we may assume

$$\sum m_\varkappa a_\varkappa - \sum n_\lambda b_\lambda = 0$$

in place of (13). Then the following proposition holds:

Proposition 4.14. *With the above notation,*

$$f(z) = \frac{\prod_1^k \sigma(z - a_\varkappa)^{m_\varkappa}}{\prod_1^\ell \sigma(z - b_\lambda)^{n_\lambda}}$$

is an elliptic function with the desired zeros and poles.

Proof: For the moment we assume the existence of $\sigma(z)$. Then obviously f is meromorphic on $\mathbb{C}$ with the prescribed poles and zeros. We show the periodicity of f: Using (14), we find for $j = 1, 2$

$$\begin{aligned} f(z + \omega_j) &= \frac{\prod \sigma(z + \omega_j - a_\varkappa)^{m_\varkappa}}{\prod \sigma(z + \omega_j - b_\lambda)^{n_\lambda}} \\ &= \frac{\prod \sigma(z - a_\varkappa)^{m_\varkappa}}{\prod \sigma(z - b_\lambda)^{n_\lambda}} \exp A_j = f(z) \exp A_j \end{aligned}$$

where

$$A_j = \sum_\varkappa m_\varkappa(\eta_j(z - a_\varkappa) + c_j) - \sum_\lambda n_\lambda(\eta_j(z - b_\lambda) + c_j).$$

By (12) and (13), $A_j = 0$.

It remains to construct $\sigma(z)$. Recall

$$-\wp(z) = -\frac{1}{z^2} - \sum{}' \left(\frac{1}{(z - \omega)^2} - \frac{1}{\omega^2} \right). \tag{15}$$

An obvious primitive of $-\wp$ is

$$\zeta(z) = \frac{1}{z} + \sum{}' \left(\frac{1}{z - \omega} + \frac{1}{\omega} + \frac{z}{\omega^2} \right), \tag{16}$$

the so-called Weierstrass ζ-function (it is not related to Riemann's ζ-function!). Since $\big(\zeta(z + \omega_j) - \zeta(z)\big)' = 0$, there are constants η_1, η_2 such that

$$\zeta(z + \omega_j) = \zeta(z) + \eta_j.$$

The series in (15) converges locally uniformly and absolutely, so the same is true for the series in (16), obtained by integration. The infinite product

$$\sigma(z) = z \prod{}' \left(1 - \frac{z}{\omega}\right) \exp\left(\frac{z}{\omega} + \frac{z^2}{2\omega^2}\right)$$

has the series (16) as its logarithmic derivative and therefore, by the Weierstrass product theorem, defines an entire function, called the Weierstrass σ-function (the product $\prod'$ is extended over all $\omega \in \Omega \setminus \{0\}$). It has simple zeros at the lattice points and no other zeros.

From $\sigma'/\sigma = \zeta$ we deduce

$$\frac{\sigma'(z+\omega_j)}{\sigma(z+\omega_j)} = \zeta(z+\omega_j) = \zeta(z) + \eta_j = \frac{\sigma'(z)}{\sigma(z)} + \eta_j.$$

Upon integrating, we obtain

$$\sigma(z+\omega_j) = \sigma(z)\exp(\eta_j z + c_j)$$

with constants c_j, $j = 1, 2$. This completes the proof of Prop. 4.14, since the Weierstrass σ-function has the required properties. □

Note that $\zeta(z)$ and $\sigma(z)$ are odd functions, and that by putting $z = -\omega_j/2$ in (14) we find $c_j = (\eta_j\omega_j + 2\pi i)/2$.

Finally, we study a special case in more detail and illustrate how integration of certain algebraic functions leads to elliptic functions. We consider a rectangular lattice Ω, generated by $\omega_1 \in \mathbb{R}$, $\omega_2 \in i\mathbb{R}$. Then $\omega \in \Omega$ implies $\overline{\omega} \in \Omega$ and

$$\wp(z) = \frac{1}{z^2} + \sum{}' \Big(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\Big) \tag{4}$$

satisfies $\wp(\bar{z}) = \overline{\wp(z)}$. Hence the coefficients $c_{2\nu}$ in the Laurent expansion (6) are all real, and so are the invariants g_2, g_3. Likewise, the values of $\wp$ on the real and imaginary axes are real or ∞. The same holds for the straight lines $\varrho_1 + i\mathbb{R}$ and $\varrho_3 + \mathbb{R}$, because e.g.

$$\wp(\varrho_1 + iy) = \wp(-\varrho_1 + iy) = \wp(\varrho_1 - iy) = \overline{\wp(\varrho_1 + iy)}.$$

We study the behaviour of $\wp$ on these straight lines more closely. By (4), $\wp(t) \to +\infty$ and $\wp(it) \to -\infty$, if $t \to 0$ through positive values. As $\wp' \neq 0$ in $]0,\, \varrho_1]$, $\wp(x)$ decreases strictly from $+\infty$ to e_1 if x increases from 0 to ϱ_1. Similarly, we see that $\wp(z)$ decreases strictly from e_1 to e_2 (from e_2 to e_3 resp. from e_3 to $-\infty$) if z runs from ϱ_1 to ϱ_2 (from ϱ_2 to ϱ_3 resp. from ϱ_3 to 0). Thus, the positive boundary of the open rectangle Q with vertices 0, ϱ_1, ϱ_2, ϱ_3 is mapped bijectively onto $\mathbb{R}$, traversed from $+\infty$ to $-\infty$. We conclude that $\wp$ maps Q biholomorphically onto the lower half plane.

We now focus on the real interval $]0,\, \varrho_1]$. Here $\wp$ is a strictly decreasing function mapping $]0,\, \varrho_1]$ onto $[e_1,\, +\infty[$. Let

$$E\colon [e_1,\, +\infty[\to]0,\, \varrho_1], \quad u \mapsto E(u) = \wp^{-1}(u)$$

be the inverse function. On $]e_1,\, +\infty[$, its derivative is

$$E'(u) = \frac{1}{\wp'\big(E(u)\big)} = \frac{-1}{\sqrt{4u^3 - g_2 u - g_3}}$$

in view of the differential equation (8); the minus sign is there because E decreases. In other words: $\frac{1}{\sqrt{4u^3-g_2u-g_3}}$ has the primitive $-E$ on $]e_1,+\infty[$, and for $u_0, u_1 \in]e_1,+\infty[$ we have

$$\int_{u_0}^{u_1} \frac{du}{\sqrt{4u^3-g_2u-g_3}} = E(u_0) - E(u_1) = \wp^{-1}(u_0) - \wp^{-1}(u_1). \tag{17}$$

Summarizing: Assume the polynomial $q(X) = 4X^3 - g_2X - g_3$ has three real roots e_1, e_2, e_3 with $e_1 > e_2 > e_3$. There exists a unique lattice Ω in $\mathbb{C}$, whose invariants are the coefficients g_2, g_3 of $q(X)$, moreover, Ω is a rectangular lattice. Then (17) holds with $\wp = \wp_\Omega$.

Taking the existence of Ω for granted, it is not hard to see that Ω has to be rectangular (Ex. 8 & 9).

A more profound analysis shows: For every polynomial $q(X) = 4X^3 - g_2X - g_3$, g_2, $g_3 \in \mathbb{C}$, with non-vanishing discriminant, path integrals

$$\int_\gamma \frac{du}{\sqrt{q(u)}}$$

in the complex plane can be evaluated by means of the inverse function of a suitable Weierstrass $\wp$-function.

If q is a quadratic polynomial with distinct zeros, formulas like

$$\int (1-u^2)^{-1/2}\, du = \arcsin u$$

show that integrals of type $\int q(u)^{-1/2}\, du$ can be evaluated using inverse functions of trigonometric functions. For polynomials of degree 3 or 4 with real coefficients and distinct zeros one can show: The indefinite integral $\int q(u)^{-1/2}\, du$ has inverse functions which can be extended to the complex plane and yield elliptic functions. In this way, the theory of elliptic functions was developed by Gauss, Legendre, Abel, Jacobi, Weierstrass et al. In the context of complex analysis, the restriction to real polynomials is artificial, it means to consider only rectangular resp. rhomboid lattices (cf. Ex. 8).

Exercises

1. Every discrete subgroup $G \neq \{0\}$ of the additive group $\mathbb{C}$ is one of the following:

   α) $G = \mathbb{Z}\omega$ with $\omega \neq 0$,

   β) $G = \{m\omega_1 + n\omega_2 : m, n \in \mathbb{Z}\}$ with ω_1, ω_2 linearly independent over $\mathbb{R}$.

   Hint: Choose $\omega_1 \in G \setminus \{0\}$ with minimal absolute value. Then $G \cap \mathbb{R}\omega_1 = \mathbb{Z}\omega_1$. If $G \neq \mathbb{Z}\omega_1$, choose $\omega_2 \in G \setminus \mathbb{Z}\omega_1$ with minimal absolute value and show $G = \langle \omega_1, \omega_2 \rangle$.

2. Let (ω_1, ω_2) be a basis of a lattice Ω. Prove: $\Omega = \langle \omega_1', \omega_2' \rangle$ if and only if $\omega_1' = a\omega_1 + b\omega_2$, $\omega_2' = c\omega_1 + d\omega_2$ with integers a, b, c, d satisfying $ad - bc = \pm 1$.

3. Choose a basis (ω_1, ω_2) of the lattice Ω according to the hint in Ex. 1. Suppose $\tau = \omega_2/\omega_1 \in \mathbb{H}$ (this can be obtained, if necessary, by replacing ω_2 by $-\omega_2$). Prove $|\tau| \geqslant 1$ and $|\mathrm{Re}\,\tau| \leqslant 1/2$. Prove further that, by modifying the basis, $-1/2 \leqslant \mathrm{Re}\,\tau < 1/2$ and $|\tau| > 1$ for $0 < \mathrm{Re}\,\tau < 1/2$ can be achieved.

4. Let Ω be a lattice and $2z_1 \notin \Omega$. Find all even elliptic functions of order 2 having poles only in $z_1 + \omega$, $-z_1 + \omega \quad (\omega \in \Omega)$.

5. Let an even $f \in K(\Omega)$ be given. If z_0 is a pole of f of multiplicity m, then so is $-z_0$, and in case $2z_0 \in \Omega$, m is even. Now find a minimal set $\{z_1, \ldots, z_r\} \subset P \setminus \{0\}$ and minimal exponents such that

$$\prod_1^r \big(\wp(z) - \wp(z_j)\big)^{m_j} f(z)$$

has no poles outside the lattice points.

6. Consider the Laurent series (6) of $\wp$ and the formulas (9). Derive a recursion formula for $c_{2\nu}$, $\nu \geqslant 3$. Deduce

   (∗) There are polynomials $P_\nu(X, Y)$ with positive rational coefficients, independent of the lattice, such that $c_{2\nu} = P_\nu(c_2, c_4)$.

7. Let Ω be a lattice with invariants g_2, g_3 and Weierstrass function $\wp = \wp_\Omega$. Prove that the following conditions are equivalent (use (∗) in Ex. 6):

   (*i*) $g_2, g_3 \in \mathbb{R}$,

   (*ii*) the $c_{2\nu}$ in (6) are all real,

   (*iii*) $\wp(\bar z) = \overline{\wp(z)}$,

   (*iv*) $\omega \in \Omega$ implies $\overline{\omega} \in \Omega$.

8. A lattice satisfying the conditions of Ex. 7 is called real. Rectangular lattices and rhomboid lattices, i.e. lattices with a basis $\langle \omega_1, \overline{\omega_1} \rangle$, are real. Prove the converse: A real lattice is rectangular or rhomboid.

9. Let $\Omega = \langle \omega_1, \overline{\omega_1} \rangle$ be a rhomboid lattice, $\varrho_2 = (\omega_1 + \overline{\omega_1})/2$. Show that the values of $\wp_\Omega$ are real or ∞ on the straight lines $\mathbb{R}$, $\omega_1 + \mathbb{R}$, $\omega_1 + i\mathbb{R}$. In particular, $e_2 = \wp(\varrho_2) \in \mathbb{R}$, $e_3 = \overline{e_1} = \overline{\wp(\omega_1/2)}$. Show further that $\wp(x)$ decreases from $+\infty$ to e_2 for $0 < x \leqslant \varrho_2$. Discuss $\int_{u_0}^{u_1} (4u^3 - g_2 u - g_3)^{-1/2}\, du$ for u_0, $u_1 > e_2$ $(g_j = g_j(\Omega))$.

10. Let Ω be a square lattice, i.e. with basis ω_1, $i\omega_1$, where $\omega_1 \in \mathbb{R}$. Then the coefficients $c_{4\nu}$ in (6) vanish, $\wp$ takes purely imaginary values on the diagonals of the fundamental square, and $e_3 < e_2 = 0 < e_1 = -e_3$, $g_2 > 0$, $g_3 = 0$.

11. Deduce from (∗) in Ex. 6: If Ω and Ω' are lattices with $g_2(\Omega) = g_2(\Omega')$ and $g_3(\Omega) = g_3(\Omega')$, then $\Omega = \Omega'$.

12. Show that

$$\zeta(z+u) + \zeta(z-u) - 2\zeta(z) = \frac{\wp'(z)}{\wp(z) - \wp(u)}.$$

Hint: Both sides are elliptic functions of z (or of u).

13. An elliptic function f with only simple poles, situated at $a_1, \ldots, a_k \in P$, can be written as

$$f(z) = c_0 + \sum_{\varkappa=1}^{k} c_\varkappa \zeta(z - a_\varkappa)$$

with suitable constants $c_0, \ldots, c_k$.

14. Show that

$$\wp(z) - \wp(u) = -\frac{\sigma(z-u)\sigma(z+u)}{\sigma^2(z)\sigma^2(u)}.$$

Use this formula to derive the result of Ex. 12.

15. Let f and g be meromorphic solutions of the differential equation ${w'}^2 = 4w^3 - g_2 w - g_3$, defined on all of $\mathbb{C}$. Assume that f is even and has a pole and g is not constant. Prove $g(z) = f(z - z_0)$ with a suitable $z_0 \in \mathbb{C}$. Hint: There are z_1, $z_2 \in \mathbb{C}$ such that $f(z_1) = g(z_2) \neq \infty$. What about $f'(z_1)$ and $g'(z_2)$? What about $f^{(k)}(z_1)$ and $g^{(k)}(z_2)$, $k \geqslant 2$?

5. Elliptic functions and plane cubics

Let Ω be a lattice in $\mathbb{C}$. The function $\wp(z)$ belonging to Ω is holomorphic in $\mathbb{C} \setminus \Omega$. Together with its derivative it defines a holomorphic map

$$\varphi\colon \mathbb{C} \setminus \Omega \to \mathbb{C}^2, \quad z \mapsto \big(\wp(z),\, \wp'(z)\big).$$

By the differential equation of $\wp$ (Thm. 4.10), the image of φ is contained in the cubic curve

$$E = \{(u, v) \in \mathbb{C}^2 : v^2 = 4u^3 - g_2 u - g_3\}.$$

In fact, φ maps onto E: For a given point $(u, v) \in E$ there is a point $z \in \mathbb{C} \setminus \Omega$ with $\wp(z) = u$. Then $\wp'(z) = v$ or $\wp'(z) = -v$, that is $\varphi(z) = (u,\, v)$ or $\varphi(-z) = (u,\, v)$.

Moreover $\varphi(z_1) = \varphi(z_2)$ if and only if $z_2 - z_1 \in \Omega$: Assume $\wp(z_1) = \wp(z_2)$, then $z_2 - z_1 \in \Omega$ or $z_1 + z_2 \in \Omega$, and the condition $\wp'(z_1) = \wp'(z_2)$ excludes $z_1 + z_2 \in \Omega$ unless $\wp'(z_j) = 0$, in which case both hold (recall that $\wp$ is an even function).

We want to extend the map $\varphi\colon \mathbb{C} \setminus \Omega \to E$ to all of $\mathbb{C}$. To this end we embed $\mathbb{C}^2$ into the complex projective plane $\mathbb{P}^2(\mathbb{C})$, take the closure $\overline{E}$ of E in $\mathbb{P}^2(\mathbb{C})$ and map the points of Ω to the unique point of $\overline{E} \setminus E$.

In more detail: The points $P \in \mathbb{P}^2(\mathbb{C})$ are described by triples $(w_0,\, w_1,\, w_2) \neq (0,\, 0,\, 0)$ of complex numbers, two triples describing the same point if and only if they differ by a scalar factor $\neq 0$. We write $P = [w_0 : w_1 : w_2]$ and call the w_j homogeneous coordinates of P.

Now we map the affine plane $\mathbb{C}^2$ injectively into $\mathbb{P}^2(\mathbb{C})$ by

$$(u,\, v) \mapsto [1 : u : v]$$

and identify $\mathbb{C}^2$ with its image. The complement of $\mathbb{C}^2$ in $\mathbb{P}^2(\mathbb{C})$ is the “line at infinity” $\{P \in \mathbb{P}^2(\mathbb{C}) : w_0(P) = 0\}$.

The embedding $\mathbb{C}^2 \hookrightarrow \mathbb{P}^2(\mathbb{C})$ takes E to the set of points with $w_0(P) \neq 0$ whose homogeneous coordinates satisfy

$$w_0 {w_2}^2 = 4{w_1}^3 - g_2 w_1 {w_0}^2 - g_3 {w_0}^3. \tag{1}$$

This equation makes sense for $w_0 = 0$, too; the only solution with $w_0 = 0$ is $P_0 = [0 : 0 : 1]$.

The set $\overline{E} = E \cup \{P_0\}$ of *all* solutions is called a projective cubic (in Weierstrass normal form).

Now we can extend the map $\varphi\colon \mathbb{C}\setminus\Omega \to E$ to a continuous map $\varphi\colon \mathbb{C} \to \overline{E}$ by defining $\varphi(\omega) = P_0$ for $\omega \in \Omega$. Note that, in a neighbourhood of ω, φ can be written as

$$\varphi(z) = [(z-\omega)^3 : (z-\omega)^3\wp(z) : (z-\omega)^3\wp'(z)];$$

the homogeneous coordinates of $\varphi(z)$ thus are holomorphic functions of z.

For the extended map $\varphi\colon \mathbb{C} \to \overline{E}$ it remains true that $\varphi(z_1) = \varphi(z_2)$ if and only if $z_2 - z_1 \in \Omega$. Hence φ induces a bijection of the quotient group $\mathbb{C}/\Omega$ onto $\overline{E}$.

$\mathbb{C}$ (or $\mathbb{C}/\Omega$) is an abelian group under addition. By φ, we can transfer this structure to $\overline{E}$, simply by defining

$$P + Q := \varphi\big(\varphi^{-1}(P) + \varphi^{-1}(Q)\big)$$

for P, $Q \in \overline{E}$. Thus P_0 is the neutral element of the group $\overline{E}$.

The group structure thus defined on $\overline{E}$ may seem rather artificial. But it admits of a nice geometric interpretation, which will eventually show that it does not depend on the parametrization of $\overline{E}$ by $(\wp, \wp')$, but only on the geometry of the projective cubic.

To this end, we need one more property of the $\wp$-function, namely the following addition theorem.

Proposition 5.1. *Let z_1, $z_2 \in \mathbb{C}\setminus\Omega$, $\wp(z_1) \neq \wp(z_2)$. Then*

$$\wp(z_1 + z_2) = -\wp(z_1) - \wp(z_2) + \frac{1}{4}\left(\frac{\wp'(z_1) - \wp'(z_2)}{\wp(z_1) - \wp(z_2)}\right)^2.$$

If, in this formula, we let z_2 tend to z_1, assuming $2z_1 \notin \Omega$, i.e. $\wp'(z_1) \neq 0$, we arrive at

Corollary 5.2. *For $2z \notin \Omega$, we have*

$$\wp(2z) = -2\wp(z) + \frac{1}{4}\left(\frac{\wp''(z)}{\wp'(z)}\right)^2 = -2\wp(z) + \frac{1}{4}\left(\frac{12\wp(z)^2 - g_2}{2\wp'(z)}\right)^2.$$

Proof of Prop. 5.1**:** We assume z_1, z_2 lie in a period parallelogram of Ω. Consider the function

$$f(z) = \wp'(z) - a\wp(z) - b,$$

the constants a and b being chosen such that $f(z_1) = f(z_2) = 0$, in particular

$$a = \frac{\wp'(z_1) - \wp'(z_2)}{\wp(z_1) - \wp(z_2)}.$$

Now f is an elliptic function of order three with poles only at the lattice points. Consequently, there is a third zero z_3 of f in the period parallelogram. By Prop. 4.5, $z_1 + z_2 + z_3 \in \Omega$. In other words, $f(-z_1 - z_2) = 0$, that is

$$\wp'(-z_1 - z_2) = a\wp(-z_1 - z_2) + b$$

or

$$-\wp'(z_1 + z_2) = a\wp(z_1 + z_2) + b.$$

We write $p_j = \wp(z_j)$, $p'_j = \wp'(z_j)$ for $j = 1, 2$, and $p_3 = \wp(z_1 + z_2)$, $p'_3 = \wp'(z_1 + z_2)$. Then

$$p'_j = ap_j + b \text{ for } j = 1, 2, \quad -p'_3 = ap_3 + b;$$

the points (p_1, p'_1), (p_2, p'_2), $(p_3, -p'_3)$ lie on the line

$$L = \{(u, v) \in \mathbb{C}^2 : v = au + b\}.$$

On the other hand, the differential equation for $\wp$ (Thm. 4.10) gives

$$4{p_j}^3 - g_2 p_j - g_3 = {p'_j}^2 = (ap_j + b)^2.$$

Thus the numbers p_1, p_2, p_3 are zeros of the cubic polynomial

$$g(w) = 4w^3 - g_2 w - g_3 - (aw + b)^2.$$

We now assume $p_3 \neq p_1, p_2$; given z_1, this excludes only finitely many values of z_2 in the period parallelogram. Then p_1, p_2, p_3 are the three different zeros of $g(w) = 4(w - p_1)(w - p_2)(w - p_3)$. Comparing the coefficients of w^2, we see

$$4(p_1 + p_2 + p_3) = a^2.$$

This is the formula of the proposition. By continuity, it also holds for the previously excluded values of z_2. □

We now use the information contained in the proof to geometrically describe the addition on $\overline{E}$. Let $z_1, z_2, z_3 = z_1 + z_2 \notin \Omega$ and $p_j = \wp(z_j)$, $p'_j = \wp'(z_j)$. Then the points $P = (p_1, p'_1)$, $Q = (p_2, p'_2)$ and $R' = (p_3, -p'_3)$ are the intersection points of the affine cubic E with the line $L = \{(u, v) : v = au + b\}$, where

$$a = \frac{p'_1 - p'_2}{p_1 - p_2} \text{ if } p_1 \neq p_2, \; a = \frac{12{p_1}^2 - g_2}{2p'_1} \text{ if } p_1 = p_2, \text{ and } b = p'_1 - ap_1 \tag{2}$$

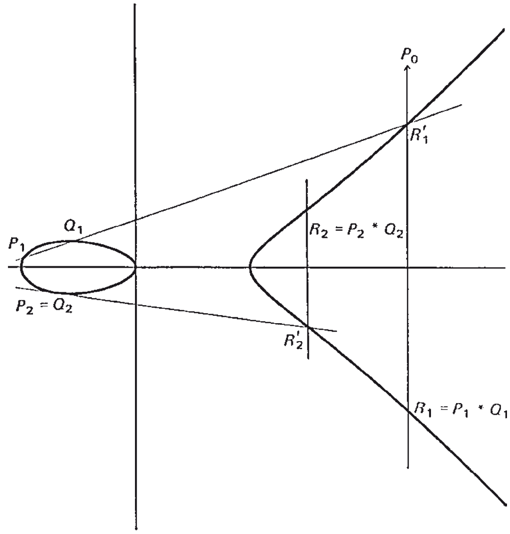


Figure 15. Addition on a cubic in Weierstrass normal form

(for the case $p_1 = p_2$ see Ex. 1). Moreover, $R = P + Q = (p_3, p_3')$ is the intersection of E with the vertical line $u = p_3$ through R'. Passing to the projective plane, this line is the (affine part of the) line $w_1 = p_3 w_0$ which intersects $\overline{E}$ in $P_0 = [0 : 0 : 1]$ and, of course, in R' and R.

Note that a complex line in $\mathbb{P}^2(\mathbb{C})$ intersects the cubic $\overline{E}$ in three points, counting multiplicity, i.e. a simple point of tangency is counted twice, an inflexion point thrice.

Thus, we have the following construction:

Given $P, Q \in \overline{E}$, let R' be the third point of intersection of the line through P and Q with $\overline{E}$. Then $R = P + Q$ is the third point of intersection of the line through P_0 and R' with $\overline{E}$. (In case $P = Q$ or $P_0 = R'$ take the tangential line to $\overline{E}$.)

We have established the construction under the assumption $z_1, z_2, z_1 + z_2 \notin \Omega$, i.e. $P, Q, R \neq P_0$. The reader may verify that it holds in the other cases, too. The parametrization $\varphi \colon \mathbb{C} \to \overline{E}$ does not enter into the geometric description of the addition on $\overline{E}$, nor into the algebraic description

$$p_3 = -p_1 - p_2 + \frac{a^2}{4}, \quad p_3' = -(ap_3 + b), \tag{3}$$

a and b being given by (2). The group laws can be verified geometrically or by computation. Therefore, any nonsingular projective cubic in Weierstrass normal form, i.e. given by (1) with ${g_2}^3 - 27{g_3}^2 \neq 0$, can be made an abelian group in this manner.

Our geometric construction of the addition on the nonsingular cubic $\overline{E}$ does not depend on the explicit equation of $\overline{E}$ either. It can be shown that, choosing a point

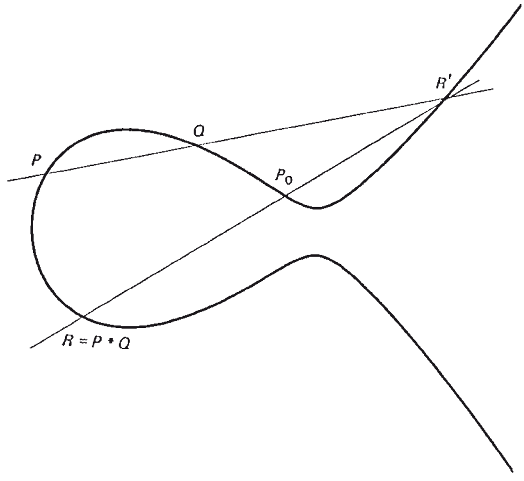


Figure 16. Addition on an arbitrary cubic

$P_0 \in \overline{E}$ arbitrarily, the construction makes $\overline{E}$ into an abelian group with P_0 as the zero element.

Exercises

1. Prove an addition theorem and a duplication formula for $\wp'$. Use the latter to verify that the tangent to E at $P = \big(\wp(z),\, \wp'(z)\big)$, $\wp'(z) \neq 0$, intersects E in $-(P+P) = \big(\wp(2z),\, -\wp'(2z)\big)$.

2. Let the curve C in $\mathbb{C}^2$ be given by the equation

$$b_2 v^2 + b_1 v + b_0 = g(u), \quad b_2 \neq 0,$$

g being a polynomial of degree 3 with 3 distinct roots. Show that by an affine linear bijection, C can be transformed into a curve in Weierstrass normal form ($v^2 = 4u^3 - g_2 u - g_3$, ${g_2}^3 - 27{g_3}^2 \neq 0$). Assuming the fact that there is a lattice Ω in $\mathbb{C}$ with $g_2(\Omega) = g_2$ and $g_3(\Omega) = g_3$, show that C can be parametrized by elliptic functions in $K(\Omega)$.

3. Let $C = \{(u,\, v) \in \mathbb{C}^2 : v^2 = f(u)\}$, f being a polynomial of degree 4 with 4 distinct roots. Show the map

$$z = (u - u_0)^{-1}, \quad w = v(u - u_0)^{-2},$$

u_0 suitably chosen, transforms C into a cubic $w^2 = g(z)$, g a cubic polynomial with distinct zeros. Then, applying Ex. 2, parametrize C by elliptic functions.

Chapter VI.

Meromorphic functions of several variables

This chapter presents the analogues of the Mittag-Leffler and Weierstrass theorems for functions of several complex variables. To this end it develops fundamental methods of multivariable complex analysis that reach far beyond the applications we are going to give here. – *Meromorphic functions* of several variables are defined as local quotients of holomorphic functions (VI.2); the definition requires some information on zero sets of holomorphic functions (VI.1). After introducing *principal parts* and *divisors* we formulate the main problems that arise: *To find a meromorphic function with i) a given principal part* (first Cousin problem) *ii) a given divisor* (second Cousin problem); *iii) to express a meromorphic function as a quotient of globally defined holomorphic functions* (Poincaré problem). These problems are solved on *polydisks* – bounded or unbounded, in particular on the whole space – in VI.6–8. The essential method is a constructive solution of the *inhomogeneous Cauchy-Riemann equations* (VI.3 and 5) based on the one-dimensional inhomogeneous Cauchy formula – see Chapter IV.2. Along the way, various extension theorems for holomorphic functions are proved (VI.1 and 4). Whereas the first Cousin problem can be completely settled by these methods, the second requires additional topological information which is discussed in VI.7, and for the Poincaré problem one needs some facts on the ring of convergent power series which we only quote in VI.8.

The main results on principal parts and divisors go back to P. Cousin 1895 and H. Poincaré 1883. It was remarked somewhat later (Gronwall 1917) that the second Cousin problem meets with a topological obstruction whose nature was finally cleared up by K. Oka in 1939 [Ok, Ra]. It was also Oka who solved these problems on the class of domains where they are most naturally posed: on *domains of holomorphy* of which polydisks are the simplest example [Ok]. The solution of the inhomogeneous Cauchy-Riemann equation in VI.5 is given by a method due to S. Bochner [Bo]; the result is usually referred to as Dolbeault's lemma, because P. Dolbeault exploited it systematically in his study of complex manifolds. The connection of compactly supported solutions and holomorphic extension theorems was discovered by L. Ehrenpreis in 1961 [Eh]; the *Kugelsatz* in VI.4 is due to F. Hartogs. Hartogs' work of 1906 ff [Ha] can be seen as the beginning of modern complex analysis in n variables. The language of *cocycles* and their solutions that we have used throughout the last sections was worked out, in this context, by H. Cartan and J. P. Serre around 1950. All the above problems can be solved on arbitrary *plane* domains [FL1].

A modern comprehensive exposition of the theory on general domains of holomorphy – even on *Stein manifolds* – can be found in [GR], [Hö], and [Ra]. Our presentation follows Hörmander and Range. For historical aspects see also [Li].

1. Zero sets of holomorphic functions

Let f be a function holomorphic on some subdomain G of $\mathbb{C}^n$; we assume $f \not\equiv 0$. The zero set

$$V(f) = \{\mathbf{z} : f(\mathbf{z}) = 0\}$$

W. Fischer, I. Lieb, *A Course in Complex Analysis*, DOI 10.1007/978-3-8348-8661-3_6,

of f is clearly a relatively closed nowhere dense subset of G. We can say more:

Proposition 1.1. *$G \setminus V(f)$ is connected.*

This will be deduced from the following fundamental result:

Theorem 1.2 (Riemann's extension theorem)**.** *If a function h is holomorphic on $G \setminus V(f)$ and locally bounded on G, then it extends to a holomorphic function $\widehat{h}$ on all of G:*

$$\widehat{h} \in \mathcal{O}(G), \quad \widehat{h}|G \setminus V(f) = h.$$

Proof: Since $V(f)$ is nowhere dense in G, the extension $\widehat{h}$, if it exists, is uniquely determined by h. So we only have to find a local holomorphic extension (near an arbitrary point $\mathbf{z}_0 \in V(f)$). We choose coordinates $\mathbf{z} = (\mathbf{z}', z_n)$, $\mathbf{z}' \in \mathbb{C}^{n-1}$, $z_n \in \mathbb{C}$, such that $\mathbf{z}_0 = 0 = (0', 0)$ and $f(0', z_n) \not\equiv 0$. Because f is continuous, there are polydisks $D' \subset \mathbb{C}^{n-1}$ and $D_n \subset \mathbb{C}$ with centres $0'$ and 0, respectively, with the following properties: $D = D' \times D_n \subset\subset G$ and $f(\mathbf{z}) \neq 0$ on $D' \times \partial D_n$. Therefore, for each $\mathbf{z}' \in D'$, the function $z_n \mapsto f(\mathbf{z}', z_n)$ is holomorphic on D_n and not identically zero; its zeros are consequently isolated in D_n. This shows that the function $h(\mathbf{z}', z_n)$ is, for fixed $\mathbf{z}' \in D'$, holomorphic in z_n, as long as $(\mathbf{z}', z_n) \notin V(f)$, and bounded on D_n; Riemann's extension theorem in one variable yields a holomorphic (in z_n) extension $\widehat{h}(\mathbf{z}', z_n)$ to all of D_n. It remains to show that $\widehat{h}$ is holomorphic on D as a function of $\mathbf{z} = (\mathbf{z}', z_n)$. But this follows from the Cauchy integral representation

$$\widehat{h}(\mathbf{z}', z_n) = \frac{1}{2\pi i} \int\limits_{\partial D_n} \frac{h(\mathbf{z}', \zeta_n)}{\zeta_n - z_n} \, d\zeta_n \ :$$

the right hand side is holomorphic in $(\mathbf{z}', z_n)$. □

The proof of Prop. 1.1 is now easy: if $G \setminus V(f)$ could be decomposed into two open non-empty sets U_0 and U_1,

$$G \setminus V(f) = U_0 \cup U_1, \quad U_0 \cap U_1 = \varnothing,$$

the function $f = 0$ on U_0 and $= 1$ on U_1 would be holomorphic on $G \setminus V(f)$ but clearly not holomorphically extendible to G.

We will consider a slightly more general situation.

Definition 1.1.

i. An analytic hypersurface S of a domain G is a non-empty subset $S \subset G$ with the following property: for each $\mathbf{z}_0 \in G$ there exists an open neighbourhood U of $\mathbf{z}_0$ and a holomorphic function f on U, nowhere $\equiv 0$, such that

$$S \cap U = V(f) = \{\mathbf{z} \in U : f(\mathbf{z}) = 0\}.$$

ii. *A subset $M \subset G$ is called thin, if it is relatively closed in G and if for each $\mathbf{z} \in M$ there is a neighbourhood U and a holomorphic function $f \not\equiv 0$ on U with $M \cap U \subset V(f)$.*

Analytic hypersurfaces are clearly closed in G, hence thin. The same proof as above carries Prop. 1.1 over to thin sets:

Proposition 1.3. *If M is a thin subset of a domain G then $G \setminus M$ is connected.* □

Exercises

1. The $2n$-dimensional (Lebesgue-)measure of a thin set is zero. Proof! Hint: Apply the Weierstrass preparation theorem.
2. The function z_2^{-1} cannot be holomorphically extended to all of $\mathbb{C}^2$. Is it locally integrable? Is it locally square integrable?

2. Meromorphic functions

Meromorphic functions of one complex variable were defined, in Chapter II, as functions which are holomorphic up to isolated singularities; the singularities were required to be poles. Since, in more than one variable, there are no isolated singular points, we use the alternative characterisation of meromorphic functions as local quotients of holomorphic functions for our definition.

Definition 2.1. *A meromorphic function on a domain $G \subset \mathbb{C}^n$ is a pair (f, M), where M is a thin set in G and f a holomorphic function on $G \setminus M$ with the following property: for each point $\mathbf{z}_0 \in G$ there is a neighbourhood U of $\mathbf{z}_0$ and there are holomorphic functions g and h on U, such that $V(h) \subset M$ and*

$$f(\mathbf{z}) = \frac{g(\mathbf{z})}{h(\mathbf{z})} \quad \text{for } \mathbf{z} \in U \setminus M. \tag{1}$$

Examples:

a) $f = g/h$, $M = V(h)$, where g and h are holomorphic on G and $h \not\equiv 0$, is a meromorphic function.

b) In particular, holomorphic functions are meromorphic.

The representation (1) is of course not unique. A closer study of the ring of convergent power series allows to define a representation (1) by a reduced fraction which is essentially unique. We do not need that here. We therefore introduce, for general (f, M), the set

$$P = \{\mathbf{z} \in G : h(\mathbf{z}) = 0 \text{ for all } (g, h) \text{ with } (1)\}$$

Then P is obviously contained in M and closed, and f can be extended to a holomorphic function $\widehat{f}$ on $G \setminus P$. We identify (f, M) with $(\widehat{f}, P)$.

Definition 2.2. *The set P as defined above is called the polar set of f.*

Since P is uniquely determined by the holomorphic function f on $G \setminus M$, we will denote meromorphic functions (f, M) simply by f, assuming, if necessary, that f is holomorphically continued to all of $G \setminus P$. It can even be shown that the polar set is a hypersurface or empty, but this requires a more detailed study of the ring of convergent power series. The polar set can alternatively be described as the set of points where f is unbounded.

We continue our examples:

c) The function $f(z_1, z_2) = z_1/z_2$ is meromorphic in $\mathbb{C}^2$, with the polar set $P = V(z_2) = \{(z_1, z_2) : z_2 = 0\}$. Note that f is unbounded at all points of P, but that $|f(z_1, z_2)| \to \infty$ only for $(z_1, z_2) \to (a, 0)$ with $a \neq 0$. This is a general phenomenon: a meromorphic function need not yield a continuous map into the Riemann sphere $\widehat{\mathbb{C}}$ – except in the case of one variable.

d) The most common way of defining a meromorphic function is the following: Let $\{U_i : i \in I\}$ be an open covering of G and $g_i, h_i \in \mathcal{O}(U_i)$ satisfy

i. h_i nowhere $\equiv 0$

ii. $g_i h_j \equiv g_j h_i$ on $U_{ij} = U_i \cap U_j$.

Then, setting

$$f = \frac{g_i}{h_i} \text{ on } U_i,$$

we obtain a well-defined meromorphic function f on G. It is holomorphic outside the thin set

$$M = \{\mathbf{z} : h_i(\mathbf{z}) = 0 \text{ for all } i \text{ with } \mathbf{z} \in U_i\}.$$

Let us now state the identity theorem for meromorphic functions:

Proposition 2.1. *If two meromorphic functions f_1 and f_2 on G coincide on a non-empty open set where both are holomorphic, they are identical on G.*

Proof: $G \setminus (P_1 \cup P_2)$, where P_j is the polar set of f_j, is connected; consequently, $f_1 \equiv f_2$ there. This says that $(f_1, P_1 \cup P_2)$ and $(f_2, P_1 \cup P_2)$ are the same meromorphic function and implies, in fact, that $P_1 = P_2$. □

Addition or multiplication of meromorphic functions at the points where they are holomorphic clearly yield meromorphic functions: the polar set of the resulting function is contained in the union of the polar sets of the summands resp. factors. Also, *if f is meromorphic and does not vanish identically, $1/f$ is again a meromorphic function.* Namely, let $f = g/h$ on an open connected set $U \subset G$; then $g \not\equiv 0$, and $1/f = h/g$

defines a meromorphic function on U. Since G can be covered by open sets as above, we are in the situation of example d) to define $1/f$.

Note that the polar set Q of $1/f$ is locally contained in the zero set $V(g)$, and that

$$Q \cap (G \setminus P) = V(f),$$

where P is the polar set of f. All this is summed up in

Proposition 2.2. *The meromorphic functions on a domain G form a field, denoted by $\mathcal{M}(G)$.* □

At this point we ask an important – and deep – question: *Is $\mathcal{M}(G)$ the quotient field of $\mathcal{O}(G)$?* The answer is positive for polydisks (including $\mathbb{C}^n$) and more general classes of domains, but not for arbitrary domains.

Exercises

1. Consider the function $f(z_1, z_2) = z_1/z_2$. Prove: the set of accumulation points of sequences $f(\mathbf{z}_j)$, $\mathbf{z}_j \to 0$, is the Riemann sphere.
2. (A more precise description of the above situation) Let $f\colon \mathbb{C}^2 \setminus \{0\} \to \widehat{\mathbb{C}}$ be given by $f(\mathbf{z}) = z_1/z_2$. Let $M \subset (\mathbb{C}^2 \setminus \{0\}) \times \widehat{\mathbb{C}}$ be its graph. Consider M as a subset of $\mathbb{C}^2 \times \widehat{\mathbb{C}}$ and show that its closure $\overline{M}$ is $M \cup (\{0\} \times \widehat{\mathbb{C}})$. Introduce homogeneous coordinates ζ_1, ζ_2 on $\widehat{\mathbb{C}}$ and describe $\overline{M}$ by a homogeneous quadratic polynomial in $z_1, z_2, \zeta_1, \zeta_2$.

3. The inhomogeneous Cauchy-Riemann equation in dimension 1

Holomorphic functions of n variables are solutions of the homogeneous Cauchy-Riemann equations

$$\frac{\partial f}{\partial \bar z_\nu} = 0, \qquad \nu = 1, \ldots, n. \tag{1}$$

In the next sections we shall construct holomorphic or meromorphic functions with prescribed additional properties by the following method: We will, in a first step, construct a smooth but not holomorphic solution with the required additional properties, say f. Then f does not satisfy (1), that is

$$\frac{\partial f}{\partial \bar z_\nu} = f_\nu \neq 0.$$

In a second step we will find a solution of the inhomogeneous Cauchy-Riemann system

$$\frac{\partial u}{\partial \bar z_\nu} = f_\nu, \tag{2}$$

such that $f - u$ still has the required properties; the function $f - u$ is then clearly holomorphic. – This method is based on a careful study of (2); note that (2) can only be solved if the right-hand side satisfies the *integrability condition*

$$\frac{\partial f_\nu}{\partial \bar{z}_\mu} = \frac{\partial f_\mu}{\partial \bar{z}_\nu}, \quad \nu, \mu = 1, \dots, n. \tag{3}$$

This condition is automatically fulfilled if the f_ν are given as above.

The main work will be done in one variable, so from now on we take $n = 1$. The integrability condition (3) is then empty.

Theorem 3.1. *Let $D' \subset\subset D$ be two disks in $\mathbb{C}$ and $G \subset \mathbb{R}^k$ a domain. There is a linear operator*

$$T\colon \mathcal{C}^\infty(D \times G) \to \mathcal{C}^\infty(D' \times G) \tag{4}$$

with

$$\frac{\partial Tf}{\partial \bar{z}} = f | D' \times G \tag{5}$$

and

$$\frac{\partial}{\partial t_j} Tf = T \frac{\partial f}{\partial t_j}, \quad j = 1, \dots, k. \tag{6}$$

(We have denoted the variable in $\mathbb{C}$ by z, the variables in $\mathbb{R}^k$ by t_j. $\mathcal{C}^\infty$ is the space of smooth - i.e. infinitely differentiable - functions. – The theorem holds with the same proof for any pair of domains $D' \subset\subset D \subset \mathbb{C}$, but we only need it in the above situation.)

Proof: 0) We choose a smooth real-valued function φ with compact support in D, $\varphi \equiv 1$ on D', and define, for $f \in \mathcal{C}^\infty(D \times G)$, $z \in D'$, $\mathbf{t} \in G$:

$$Tf(z, \mathbf{t}) = \frac{1}{2\pi i} \int_{\mathbb{C}} \frac{\varphi(\zeta) f(\zeta, \mathbf{t})}{\zeta - z} \, d\zeta \wedge d\bar{\zeta}. \tag{7}$$

This operator will be shown to have the required properties.

1) The substitution $w = \zeta - z$ leads to

$$Tf(z, \mathbf{t}) = \frac{1}{2\pi i} \int_{\mathbb{C}} \frac{\varphi(w+z) f(w+z, \mathbf{t})}{w} \, dw \wedge d\overline{w}, \tag{8}$$

which immediately shows that Tf is smooth on $\mathbb{C} \times G$; differentiation under the integral (7) yields (6).

2) The inhomogeneous Cauchy formula – see IV.2 – gives for the compactly supported function φf

$$\varphi(z)f(z, \mathbf{t}) = \frac{1}{2\pi i}\int\limits_{\mathbb{C}} \frac{\partial/\partial\bar\zeta[\varphi(\zeta)f(\zeta, \mathbf{t})]}{\zeta - z}\, d\zeta \wedge d\bar\zeta; \tag{9}$$

a boundary integral does not occur because $\varphi(\zeta)f(\zeta, \mathbf{t}) \equiv 0$ for $|\zeta|$ large enough. – On the other hand, differentiation of (8) with respect to $\bar z$ leads to

$$\begin{aligned} \frac{\partial}{\partial\bar z}Tf(z, \mathbf{t}) &= \frac{1}{2\pi i}\int\limits_{\mathbb{C}} \frac{\partial/\partial\bar z[\varphi(z+w)f(z+w, \mathbf{t})]}{w}\, dw \wedge d\overline{w} \\ &= \frac{1}{2\pi i}\int\limits_{\mathbb{C}} \frac{\partial/\partial\bar\zeta[\varphi(\zeta)f(\zeta, \mathbf{t})]}{\zeta - z}\, d\zeta \wedge d\bar\zeta. \end{aligned} \tag{10}$$

Comparison of (9) and (10) shows the claim: for $z \in D'$ one has

$$f(z, \mathbf{t}) = \varphi(z)f(z, \mathbf{t}) = \frac{\partial}{\partial\bar z}Tf(z, \mathbf{t}). \qquad \square$$

Remark: *If f is holomorphic in some of the parameters, then Tf is holomorphic in the same parameters.*

In fact, assume f holomorphic in t_0 – so now G is a subdomain of $\mathbb{C} \times \mathbb{R}^\ell$ with coordinates t_0 and $\mathbf{t}$ – then in view of (6)

$$\frac{\partial}{\partial\bar t_0}Tf(z, t_0, \mathbf{t}) = T\frac{\partial}{\partial\bar t_0}f(z, t_0, \mathbf{t}) = 0.$$

Exercises

1. Justify in detail the differentiation under the integral sign used in the proof of Thm. 3.1.

4. The Cauchy-Riemann equations with compact support

We will solve, for $n = 1, 2, \ldots,$ the Cauchy-Riemann differential equations

$$\frac{\partial u}{\partial\bar z_\nu} = f_\nu, \quad \nu = 1, \ldots, n, \tag{1}$$

where the f_ν are smooth functions in $\mathbb{C}^n$ *with compact support* satisfying the *integrability condition*

$$\frac{\partial f_\nu}{\partial\bar z_\mu} = \frac{\partial f_\mu}{\partial\bar z_\nu}, \quad \nu, \mu = 1, \ldots, n. \tag{2}$$

The main result is

Theorem 4.1. (1) *has a smooth solution u. If $n > 1$, then u can be chosen with compact support.*

More precisely: *If $f_\nu \equiv 0, \nu = 1, \ldots, n$, outside a compact set K then there is, in case $n > 1$, a solution which vanishes on the unbounded component of the complement of K.* – Note that solutions are of course not unique: we can always add a holomorphic function to obtain a new solution from a given one.

Proof: We define

$$u(z_1, \ldots, z_n) = \frac{1}{2\pi i} \int\limits_{\mathbb{C}} \frac{f_1(\zeta, z_2, \ldots, z_n)}{\zeta - z_1} d\zeta \wedge d\bar{\zeta}; \tag{3}$$

this is a smooth function on $\mathbb{C}^n$ satisfying – see the previous section –

$$\frac{\partial u}{\partial \bar{z}_1} = f_1.$$

In fact, (3) is the solution Tf_1 of section 3 constructed for a sufficiently large disk D' and $D = \mathbb{C}$. Now, for $\nu > 1$,

$$\frac{\partial u}{\partial \bar{z}_\nu} = \frac{1}{2\pi i} \int\limits_{\mathbb{C}} \frac{\partial f_1/\partial \bar{z}_\nu(\zeta, z_2, \ldots, z_n)}{\zeta - z_1} d\zeta \wedge d\bar{\zeta} = \frac{1}{2\pi i} \int\limits_{\mathbb{C}} \frac{\partial f_\nu/\partial \bar{\zeta}(\zeta, z_2, \ldots, z_n)}{\zeta - z_1} d\zeta \wedge d\bar{\zeta},$$

because of (2). Since f_ν has compact support, the last integral is, by the inhomogeneous Cauchy integral formula, $f_\nu(z_1, \ldots, z_n)$. So u solves (1). – Moreover, let the support of the functions f_ν be contained in a compact set K, and let U_0 be the unbounded component of $\mathbb{C}^n \setminus K$. If $n > 1$, there is an affine hyperplane E contained in U_0. The function u is holomorphic outside K, in particular its restriction to E is an entire function of $n-1$ variables. Now, the integration in (3) only has to extend over a bounded domain (again because the support of f_1 is compact). This shows that $u(z) \to 0$ for $|z| \to \infty$. By Liouville's theorem, $u|E \equiv 0$. As E could be chosen arbitrarily in U_0, the identity theorem gives $u \equiv 0$ on U_0. □

Remark: For $n = 1$ the above solution need not have compact support – see Ex. 1.

A striking consequence of the above theorem is

Theorem 4.2 (Hartogs' Kugelsatz)**.** *Let K be a compact subset of an open set $U \subset \mathbb{C}^n$, with $n > 1$, and suppose that $U \setminus K$ is connected. Then any function f holomorphic on $U \setminus K$ is the restriction to $U \setminus K$ of a holomorphic function $\widehat{f}$ on U.*

Proof: We choose a smooth real-valued function φ with compact support in U, $\varphi \equiv 1$ in a neighbourhood of K, and set, for $f \in \mathcal{O}(U \setminus K)$,

$$\tilde{f} = (1 - \varphi) f. \tag{4}$$

This function is smooth on all of U – where f is not defined, the right-hand side of (4) is 0. The derivatives

$$g_\nu = \frac{\partial \tilde{f}}{\partial \bar{z}_\nu} \tag{5}$$

are smooth in $\mathbb{C}^n$, with support contained in the support of φ, because outside that support $\tilde{f} = f$ is holomorphic. They obviously satisfy the integrability conditions (2) and therefore there exists a solution u of

$$\frac{\partial u}{\partial \bar{z}_\nu} = g_\nu, \tag{6}$$

which vanishes in the unbounded component of the complement of the support of φ. But the boundary of U belongs to that component. So

$$\widehat{f} = \tilde{f} - u \tag{7}$$

coincides with f on a non-empty open set in $U \setminus K$, hence on all of $U \setminus K$, and is holomorphic on U because of (5) and (6). □

The theorem applies in particular to a spherical shell: this explains the name.

Exercises

1. Let $f \in \mathcal{C}^\infty(\mathbb{C})$ be a smooth function with compact support; set $g = f_{\overline{z}}$. Compute the integral

$$\int_{\mathbb{C}} g(z)\, dx\, dy$$

with the help of Stokes' theorem. Use the information obtained that way to construct a compactly supported smooth function g in the plane that has no solution f of $f_{\overline{z}} = g$ with compact support.

5. The Cauchy-Riemann equations in a polydisk

The problem of solving the Cauchy-Riemann equations is more difficult if one drops the assumption of compact support; we study it here in polydisks. In all that follows, D, D', D_0, ... will stand for polydisks in $\mathbb{C}^n$ centred at the origin. Let the functions f_ν, for $\nu = 1, \ldots, n$, be smooth on D and satisfy the integrability conditions

$$\frac{\partial f_\nu}{\partial \bar{z}_\mu} = \frac{\partial f_\mu}{\partial \bar{z}_\nu}, \quad \nu, \mu = 1, \ldots, n. \tag{1}$$

We want to solve

$$\frac{\partial u}{\partial \bar{z}_\nu} = f_\nu, \quad \nu = 1, \ldots, n, \tag{2}$$

by a smooth function u on D.

Proposition 5.1. *Let $D_0 \subset\subset D$ be relatively compact in D. Then* (2) *can be solved by a function $u \in \mathcal{C}^\infty(D_0)$.*

Proof: We denote by $A_k(D')$ the set of n-tuples of smooth functions

$$(f) = (f_1, \ldots, f_n)$$

on D' satisfying $f_\nu \equiv 0$ for $\nu > k$ and condition (1). Here, D' stands for any polydisk contained in D. Let us prove, for arbitrary pairs of polydisks $D_0 \subset\subset D'$, the claim in case the right-hand side of (2) belongs to $A_k(D')$; the case $k = n$ is what we want.

We proceed by induction with respect to k. If $k = 0$, then $(f) = (0)$, and the function $u \equiv 0$ solves. Now consider $(f) \in A_k(D')$, and assume the claim for $k - 1 \geqslant 0$. So

$$(f) = (f_1, \ldots, f_k, 0, \ldots, 0).$$

By (1) we have for $\mu > k$:

$$\frac{\partial f_k}{\partial \bar{z}_\mu} = \frac{\partial f_\mu}{\partial \bar{z}_k} \equiv 0,$$

since $f_\mu \equiv 0$. Hence f_k is holomorphic in the variables $z_{k+1}, \ldots, z_n$. We now choose the solution v of

$$\frac{\partial v}{\partial \bar{z}_k} = f_k \tag{3}$$

constructed in section 3 on a polydisk D'' with $D_0 \subset\subset D'' \subset\subset D'$:

$$v(z_1, \ldots, z_n) = T f_k(z_1, \ldots, z_n).$$

Then v is holomorphic in $z_{k+1}, \ldots, z_n$, so

$$\frac{\partial v}{\partial \bar{z}_\mu} = 0 = f_\mu, \quad \mu > k. \tag{4}$$

The system $(g) = (g_1, \ldots, g_n)$ with

$$g_j = f_j - \frac{\partial v}{\partial \bar{z}_j}$$

belongs to $A_{k-1}(D'')$ and satisfies the integrability conditions: in fact, for $j > k$ we have $f_j = 0 = \partial v / \partial \bar{z}_j$, and for $j = k$ we have $g_k = 0$ in view of (3); finally (1) is satisfied for (g) because it is satisfied for (f) and for the derivatives of v. The induction hypothesis yields a solution $w \in \mathcal{C}^\infty(D_0)$ with

$$\frac{\partial w}{\partial \bar{z}_j} = g_j;$$

setting $u = v + w$ on D_0 we thus solve (2). – This concludes the induction and proves our claim. □

An approximation argument will finally give a solution of the Cauchy-Riemann system on the whole polydisk:

Theorem 5.2. *The system*

$$\frac{\partial u}{\partial \bar z_\nu} = f_\nu, \quad \nu = 1, \ldots, n, \tag{2}$$

with $f_\nu \in \mathcal{C}^\infty(D)$ has a solution $u \in \mathcal{C}^\infty(D)$ if and only if it satisfies the integrability conditions

$$\frac{\partial f_\nu}{\partial \bar z_\mu} = \frac{\partial f_\mu}{\partial \bar z_\nu}, \quad \nu, \mu = 1, \ldots, n. \tag{1}$$

Proof: We choose a sequence of polydisks

$$D_0 \subset\subset D_1 \subset\subset D_2 \subset\subset \ldots \subset\subset D$$

with

$$\bigcup_{\varkappa \geqslant 0} D_\varkappa = D$$

and functions $u'_\varkappa \in \mathcal{C}^\infty(D_\varkappa)$ which solve (2) on $D_\varkappa$. If the sequence $u'_\varkappa$ – which is defined on each fixed D_k for all $\varkappa \geqslant k$ – were convergent on D_k, we would simply define our solution u as the limit of that sequence. So our task is to modify the $u'_\varkappa$ in order to obtain a convergent sequence. We set $u_0 = u'_0$, $u_1 = u'_1$ and assume that we have found solutions $u_\varkappa$ of (2) on $D_\varkappa$, for $\varkappa = 1, \ldots, k$, such that

$$|u_\varkappa - u_{\varkappa-1}|_{D_{\varkappa-2}} < 2^{1-\varkappa}. \tag{5}$$

Here $|\cdot|_M$ denotes the supremum norm of a function on M. Now u'_{k+1} and u_k both solve (2) on D_k and consequently differ by a holomorphic function f_k on D_k. Power series development of f_k around 0 yields a polynomial p_k such that

$$|f_k - p_k|_{D_{k-1}} = |u'_{k+1} - u_k - p_k|_{D_{k-1}} < 2^{-k}.$$

Setting

$$u_{k+1} = u'_{k+1} - p_k,$$

we get a new solution of (2) on D_{k+1} which now satisfies (5) for $\varkappa = k+1$. Let us now define, for $z \in D$,

$$u(z) = \lim_{\varkappa \to \infty} u_\varkappa(z).$$

The limit is well-defined – see our remark above. Moreover, for $\varkappa \geqslant \lambda \geqslant k+2$ and $z \in D_k$

$$\begin{aligned} |u_\varkappa(z) - u_\lambda(z)| &\leqslant |u_\varkappa(z) - u_{\varkappa-1}(z)| + \cdots + |u_{\lambda+1}(z) - u_\lambda(z)| \\ &\leqslant 2^{-\varkappa+1} + \cdots + 2^{-\lambda} \leqslant 2 \cdot 2^{-\lambda}; \end{aligned}$$

so the limit exists uniformly on D_k. This shows that u is a continuous function on D. Now, on D_k we have

$$u - u_k = \lim_{\varkappa \geqslant k+2} (u_\varkappa - u_k),$$

and the terms on the right-hand side are holomorphic on D_k – which implies that their uniform limit is holomorphic as well. Hence, $u - u_k$ and therefore u is smooth on D_k, and u thus solves (2) because it differs from the solution u_k by a holomorphic function. □

6. Principal parts: the first Cousin problem

Let $G \subset \mathbb{C}^n$ be a domain.

Definition 6.1. *A Cousin-I-distribution on G is a system (f_i, U_i), $i \in I$, where the U_i are open subsets of G which form a covering of G and the f_i are meromorphic functions on U_i such that*

$$f_{ij} = f_j - f_i \in \mathcal{O}\,(U_{ij})\,, \tag{1}$$

i.e. f_{ij} is holomorphic on $U_{ij} = U_i \cap U_j$.

If $U_{ij} = \varnothing$, condition (1) is of course void. The *principal part* of a meromorphic function can be defined as follows: *two meromorphic functions f and g have the same principal part if their difference is holomorphic.* In particular, their polar sets coincide in that case. So in the above definition, the principal parts of f_i and f_j coincide where both functions are defined. – Instead of Cousin-I-distribution we could also speak of a distribution of principal parts or simply of a principal part on G. Note that on $\mathbb{C}$ the Mittag-Leffler data define naturally a Cousin-I-distribution, and vice versa – see exercises.

Definition 6.2. *A solution of a Cousin-I-distribution (f_i, U_i) is a meromorphic function f on G such that $f - f_i$ is holomorphic on U_i for all i.*

In other words: f should be a meromorphic function with the given principal parts. The Mittag-Leffler theorem gives a solution of any Cousin-I-distribution in the complex plane. In general, on an arbitrary domain in $\mathbb{C}^n$, with $n > 1$, not every Cousin-I-distribution is soluble; for $n = 1$, it is – see [FL1]. We will show that any Cousin-I-distribution on a polydisk is soluble. So from now on, we will choose for G a (bounded or unbounded) polydisk $D \subset \mathbb{C}^n$. The main work will be done in proving the next theorem.

Theorem 6.1. *Let $\{U_i : i \in I\}$ be an open covering of the polydisk D. Suppose that for each pair of indices $i, j \in I$ with $U_{ij} \neq \varnothing$ a holomorphic function $g_{ij} \in \mathcal{O}(U_{ij})$ is given such that the following conditions are fulfilled:*

$$g_{ij} = -g_{ji} \tag{2}$$

$$g_{jk} - g_{ik} + g_{ij} = 0 \text{ on } U_{ijk}. \tag{3}$$

Then there are holomorphic functions $g_i \in \mathcal{O}(U_i)$ with

$$g_j - g_i = g_{ij} \text{ on } U_{ij}. \tag{4}$$

We have used the notation $U_{ij} = U_i \cap U_j,\ \ U_{ijk} = U_i \cap U_j \cap U_k$. If $U_{ijk} = \varnothing$, then condition (3) is void.

Proof: Let us choose a partition of unity subordinate to the covering, i.e. functions $\varphi_i \in \mathcal{C}^\infty(D)$, $0 \leqslant \varphi_i \leqslant 1$, $\operatorname{supp}\varphi_i \subset U_i$, such that the system of supports of the φ_i is locally finite, and

$$\sum_{i \in I} \varphi_i(z) \equiv 1. \tag{5}$$

Then for each i

$$h_i = \sum_{k \in I} \varphi_k g_{ki} \tag{6}$$

is a well-defined smooth function on U_i, and we have on U_{ij}

$$\frac{\partial h_i}{\partial \bar{z}_\nu} - \frac{\partial h_j}{\partial \bar{z}_\nu} = \sum_{k \in I} \frac{\partial \varphi_k}{\partial \bar{z}_\nu}(g_{ki} - g_{kj}) = \sum_{k \in I} \frac{\partial \varphi_k}{\partial \bar{z}_\nu} g_{ji} = \left(\frac{\partial}{\partial \bar{z}_\nu} \sum_{k \in I} \varphi_k\right) g_{ji} = 0 \tag{7}$$

because of (3) and (5). Hence, for each ν, the function

$$F_\nu(z) = \frac{\partial h_i}{\partial \bar{z}_\nu}(z), \quad z \in U_i, \tag{8}$$

is well-defined on all of D, and the F_ν satisfy, in view of their definition (8) as derivatives, the integrability conditions for the Cauchy-Riemann system. The previous section yields a smooth function $u \in \mathcal{C}^\infty(D)$ with

$$\frac{\partial u}{\partial \bar{z}_\nu} = F_\nu. \tag{9}$$

Now set on U_i

$$g_i = h_i - u. \tag{10}$$

Then $g_i \in \mathcal{O}(U_i)$ and

$$g_j - g_i = h_j - h_i = \sum_{k \in I} \varphi_k (g_{kj} - g_{ki}) = \left(\sum_{k \in I} \varphi_k\right) g_{ij} = g_{ij},$$

again by (2), (3) and (5). So the g_i solve (4). □

Let us point out that we only used solubility of the Cauchy-Riemann equations in the proof. It is worthwhile to state this explicitly as

Theorem 6.2. *If G is a domain where the Cauchy-Riemann equations for smooth data are always soluble, then for all data $\left(U_i,\, g_{ij}\right)$ as in Thm. 6.1 there are holomorphic functions $g_i \in \mathcal{O}\left(U_i\right)$ with $g_{ij} = g_j - g_i$.* □

It is equally worthwhile to introduce a new word in order to express the above results, and also for later use:

Definition 6.3. *The data $\left(U_i,\, g_{ij}\right)$ with the properties (2) and (3) of Thm. 6.1 is called an $\mathcal{O}$-cocycle; the $\left(U_i,\, g_i\right)$ satisfying (4) are an $\mathcal{O}$-solution of this cocycle.*

So we can say more conveniently: *An arbitrary $\mathcal{O}$-cocycle on a polydisk has an $\mathcal{O}$-solution; the same result holds on domains where the Cauchy-Riemann system is soluble.*

From here we deduce easily

Theorem 6.3. *Any Cousin-I-distribution on a polydisk – more generally: on a domain satisfying the assumption of Thm. 6.2 – is soluble.*

Proof: Let $(f_i,\, U_i)$ be such a distribution. Then the f_{ij} given by (1) define an $\mathcal{O}$-cocycle, which therefore has an $\mathcal{O}$-solution (g_i). Let us now set

$$f = f_i - g_i \text{ on } U_i. \tag{11}$$

In view of (4) the definition is independent of the choice of i and yields a meromorphic function on D which, by (11), solves the distribution. □

Exercises

1. Let $G \subset \mathbb{C}^n$ be an arbitrary domain and $f_j, j = 1, \ldots, n$, be smooth functions on G satisfying the integrability conditions (1) from section 5. Show that every point $a \in G$ has a neighbourhood U such that there is a smooth function u on U with

$$\frac{\partial u}{\partial \overline{z_j}} = f_j | U.$$

Choose an open covering U_i with corresponding solutions u_i of the above equation. Consider the differences $u_j - u_i$ on U_{ij}. From here, state and prove a converse to Thm. 6.2.

7. Divisors: the second Cousin problem

We now carry over Weierstrass' product theorem to higher dimensions. We shall even consider a slightly more general situation. In all that follows, G will be a domain in $\mathbb{C}^n$; later on it will be taken to be a polydisk. $\mathcal{M}^*(U)$ denotes the set of meromorphic functions on the open set U which nowhere vanish identically, $\mathcal{O}^*(U)$ is the set of holomorphic functions on U without zeros. Both sets are multiplicative groups: the groups of units in the rings $\mathcal{M}(U)$ resp. $\mathcal{O}(U)$.

Definition 7.1. *A divisor on G is a system*

$$\Delta = \big(U_i,\, f_i\big)_{i\in I},\ I \text{ an index set,}$$

where the $U_i \subset G$ form an open covering of G and the f_i are elements of $\mathcal{M}^*(U_i)$, such that for all $i,\, j \in I$ with $U_{ij} = U_i \cap U_j \neq \varnothing$ one has

$$\frac{f_j}{f_i} = g_{ij} \in \mathcal{O}^*(U_{ij})\,. \tag{1}$$

Two divisors Δ and Δ' are – by definition! – equal if their union $\Delta \cup \Delta'$ is a divisor.

Condition (1) means, intuitively, that on U_{ij} the polar sets of f_i and f_j coincide, including *multiplicities*, and that also the zero sets of f_i and f_j are identical, and that the two functions vanish there with the same multiplicity. By sticking to the above definition one circumvents the need to explicitly define the notion of multiplicity. – A *positive* divisor is a divisor given by holomorphic functions f_i satisfying (1).

If $f \not\equiv 0$ is a meromorphic function on G, then $(G,\, f)$ is a divisor. We call these divisors *principal* – notation: $\operatorname{div} f$. Now suppose that $\Delta = \big(U_i,\, f_i\big)$ is an arbitrary divisor. Then clearly,

$$\Delta = \operatorname{div} f, \qquad f \in \mathcal{M}^*(G)\,,$$

if and only if *the quotients f/f_i are holomorphic without zeros on U_i, for all i.* We define:

Definition 7.2. *A solution of Δ is a meromorphic function f with $\Delta = \operatorname{div} f$.*

Such a solution – if it exists – is obviously determined up to multiplication with an element of $\mathcal{O}^*(G)$. In the case of a positive divisor it is a holomorphic function with prescribed zeros (including multiplicities): that is just what is given in the Weierstrass product theorem.

Divisors – more precisely, the $\big(U_i,\, f_i\big)$ satisfying (1) – are also called Cousin-II-distributions. The second Cousin problem can now be briefly stated:

Is each divisor on G a principal divisor?

The problem is by now completely understood for large classes of domains (domains of holomorphy); here we shall give a positive answer in the case of polydisks. Our discussion will show that certain topological conditions play a role, conditions which are fulfilled on a polydisk but not in general. The method of proof reaches beyond our immediate aim.

We start with a topological result:

Lemma 7.1. *Let $f\colon U \to \mathbb{C}^*$ be a continuous function in a simply connected domain $U \subset \mathbb{R}^k$. Then f has a continuous logarithm, i.e. there is a continuous function F on U with*

$$f = \exp F. \tag{2}$$

In fact, the universal cover of the punctured plane $\mathbb{C}^*$ is the exponential map $\exp\colon \mathbb{C} \to \mathbb{C}^*$, so f lifts to a continuous map F to $\mathbb{C}$.

Before proving the multiplicative analogue of Thm. 6.1 we introduce a convenient terminology:

Definition 7.3.

i. An $\mathcal{O}^*$-cocycle in a domain G is a data $g = \big(U_i,\, g_{ij}\big)$, where the U_i, $i \in I$, are an open covering of G and $g_{ij} \in \mathcal{O}^*(U_{ij})$ such that

$$g_{ij} = {g_{ji}}^{-1}, \tag{3}$$
$$g_{jk}{g_{ik}}^{-1}g_{ij} = 1 \text{ on } U_{ijk}, \tag{4}$$

provided $U_{ijk} \neq \varnothing$.

ii. An $\mathcal{O}^*$-solution h of g is a system $h_i \in \mathcal{O}^*(U_i)$ of holomorphic functions with

$$g_{ij} = \frac{h_j}{h_i} \text{ on } U_{ij}.$$

If we replace $\mathcal{O}^*$ by $\mathcal{C}^*$, the space of continuous functions without zeros, we obtain in the same way the notion of a $\mathcal{C}^*$-cocycle and a $\mathcal{C}^*$-solution. Since an $\mathcal{O}^*$-cocycle is also a $\mathcal{C}^*$-cocycle, we can speak of $\mathcal{C}^*$-solutions of an $\mathcal{O}^*$-cocycle.

The analogue of Thm. 6.1 and 6.2 now comes up with a surprise:

Theorem 7.2. *Let D be a polydisk – or, more generally, a domain where the Cauchy-Riemann system is always soluble. Then an $\mathcal{O}^*$-cocycle has an $\mathcal{O}^*$-solution if and only if it has a $\mathcal{C}^*$-solution.*

Proof: Let $\big(U_i,\, g_{ij}\big)$ be the given cocycle.

1) We first assume that the U_i are simply connected. By our assumption,

$$g_{ij} = c_j c_i^{-1} \tag{5}$$

with non-vanishing continuous functions c_i, c_j on U_i and U_j, respectively.
By Lemma 7.1, for each i,

$$c_i = \exp b_i,$$

b_i continuous on U_i. Now, on the intersection U_{ij}:

$$g_{ij} = \exp(b_j - b_i),$$

which means that

$$b_{ij} = b_j - b_i \tag{6}$$

is holomorphic. The b_{ij} are clearly an $\mathcal{O}$-cocycle and Thm. 6.1 provides us with a solution $a_i \in \mathcal{O}(U_i)$:

$$b_{ij} = a_j - a_i \tag{7}$$

on U_{ij}. Let us now set $h_i = \exp a_i$. Then

$$\frac{h_j}{h_i} = \exp(a_j - a_i) = \exp b_{ij} = \frac{c_j}{c_i} = g_{ij},$$

as required.

2) The general case will now be reduced to the above special case. Let the $\mathcal{O}^*$-cocycle $g = \big(U_i,\, g_{ij}\big)_{i\in I}$ be given with a $\mathcal{C}^*$-solution $c = \big(U_i,\, c_i\big)_{i\in I}$. We choose a covering V_α, $\alpha \in A$, of simply connected open sets which is *finer* than the covering by the U_i, i.e. each V_α is contained in some U_i, and there is a *refinement map* $\tau\colon A \to I$ such that always $V_\alpha \subset U_{\tau(\alpha)}$. We now set

$$g'_{\alpha\beta} = g_{\tau(\alpha)\tau(\beta)}; \quad c'_\alpha = c_{\tau(\alpha)},$$

restricted to $V_{\alpha\beta}$ and V_α resp. Then $\big(V_\alpha,\, g'_{\alpha\beta}\big)$ is again an $\mathcal{O}^*$-cocycle with a $\mathcal{C}^*$-solution $\big(V_\alpha,\, c'_\alpha\big)$, and by the first part of the proof we obtain an $\mathcal{O}^*$-solution $h'_\alpha \in \mathcal{O}^*(V_\alpha)$ with

$$g'_{\alpha\beta} = \frac{h'_\beta}{h'_\alpha}. \tag{8}$$

We now define $h_i \in \mathcal{O}^*(U_i)$ by

$$h_i = h'_\alpha g_{\tau(\alpha)i} \text{ on } U_i \cap V_\alpha. \tag{9}$$

On $U_i \cap V_{\alpha\beta}$ we have in view of (4):

$$h'_\alpha g_{\tau(\alpha)i} {h'_\beta}^{-1} g_{i\tau(\beta)} = g'_{\beta\alpha} g_{\tau(\alpha)\tau(\beta)} = g'_{\beta\alpha} g'_{\alpha\beta} = 1,$$

so (9) defines h_i on U_i uniquely. The definition yields on $U_{ij} \cap V_\alpha$

$$h_j {h_i}^{-1} = h'_\alpha g_{\tau(\alpha)j} {h'_\alpha}^{-1} g_{i\tau(\alpha)} = g_{\tau(\alpha)j} g_{i\tau(\alpha)} = g_{ij};$$

so (9) solves our problem. □

As in the previous section we point out that, among the various properties of D, we only need the solubility of the Cauchy-Riemann equations for smooth data. We state this explicitly as

Theorem 7.3. *If $G \subset \mathbb{C}^n$ is a domain where the Cauchy-Riemann system is soluble, then an $\mathcal{O}^*$-cocycle on G has an $\mathcal{O}^*$-solution if and only if it has a $\mathcal{C}^*$-solution.*

To express this more briefly we introduce

Definition 7.4. *A domain $G \subset \mathbb{C}^n$ has the Oka property if each $\mathcal{C}^*$-cocycle has a $\mathcal{C}^*$-solution.*

It now follows easily

Theorem 7.4. *Let G be a domain with the Oka property where the Cauchy-Riemann system is soluble. Then all divisors on G are principal.*

Proof: Let $\Delta = \big(U_i,\, f_i\big)$ be a divisor. By the Oka property there are continuous functions without zeros c_i on U_i such that

$$\frac{f_j}{f_i} =: g_{ij} = \frac{c_j}{c_i}.$$

The previous theorem gives us holomorphic functions $h_i \in \mathcal{O}^*(U_i)$ with

$$g_{ij} = \frac{h_j}{h_i}.$$

Then

$$f := \frac{f_i}{h_i} \text{ on } U_i$$

is a well-defined meromorphic function on G with divisor Δ. □

In order to apply this to the polydisk we need

Theorem 7.5. *Polydisks have the Oka property.*

We do not give the – purely topological – proof, but refer the reader to [Ra]. The main consequence now is

Theorem 7.6. *All divisors on a polydisk are principal.* □

8. Meromorphic functions revisited

We go back to the quotient representation of meromorphic functions. We have to use certain algebraic properties of the ring of convergent power series which can be deduced from the Weierstrass preparation theorem which has been proved in Chapter IV; the details of the deduction, however, will not be given in our book. So this is one instance where we rely on a bit more than just the previous arguments.

Let f be holomorphic in a domain G. If $\mathbf{z}_0 \in G$, then, as a consequence of Cauchy's integral formula, f can be developed into a convergent power series in $\mathbf{z} - \mathbf{z}_0$; let us call this series $f_{\mathbf{z}_0}$. We now need

Theorem 8.1. *The ring H of convergent power series of n variables is factorial.*

Moreover, again as a consequence of the Weierstrass preparation theorem, we have

Proposition 8.2. *If f and g are holomorphic and their power series $f_{\mathbf{z}_0}$ and $g_{\mathbf{z}_0}$ are coprime, then for all points $\mathbf{z}$ in a sufficiently small neighbourhood of $\mathbf{z}_0$ the series $f_{\mathbf{z}}$ and $g_{\mathbf{z}}$ are also coprime.*

For the proofs we refer to [Hö].

The upshot of the previous statements is: *A meromorphic function f can always locally be represented by quotients g/h of holomorphic functions with $g_{\mathbf{z}}$ and $h_{\mathbf{z}}$ coprime for all $\mathbf{z}$ in a sufficiently small open set.*

We apply this information to divisors. Let us first note that a divisor Δ given as

$$\Delta = \big(U_i,\, f_i\big)_{i\in I}$$

can equally well be given as

$$\Delta = \big(V_j,\, g_j\big)_{j\in J},$$

where the V_j are a refinement of the covering U_i and

$$g_j = f_{\sigma(j)}|V_j,$$

with $\sigma\colon J \to I$ a *refinement map*, i.e. $V_j \subset U_{\sigma(j)}$. Different refinement maps yield different g_j but the same divisor. So two divisors Δ and Γ can always be given by a Cousin-II-distribution defined over the same open covering: just pass to a common refinement of the original coverings! Now, if

$$\Delta = \big(U_i,\, f_i\big)_{i\in I},$$
$$\Gamma = \big(U_i,\, g_i\big)_{i\in I}$$

are given, we define their product as

$$\Delta\Gamma = \big(U_i,\, f_i g_i\big)_{i\in I}.$$

It is easy to check that this is again a divisor, and that the divisors on G form an abelian group under this multiplication; for instance

$$\Delta^{-1} = \big(U_i,\, 1/f_i\big)_{i\in I}.$$

Let now

$$\Delta = \big(U_i,\, f_i\big)_{i\in I}$$

be a divisor. We can choose the U_i so small that we have a representation

$$f_i = g_i/h_i, \quad g_i,\, h_i \in \mathcal{O}\,(U_i)$$

with $g_{i\mathbf{z}}$ and $h_{i\mathbf{z}}$ coprime for each $\mathbf{z} \in U_i$.

This implies that

$$\Delta_+ = \big(U_i,\, g_i\big)_{i\in I}$$

and

$$\Delta_- = \big(U_i,\, h_i\big)_{i\in I}$$

are again – necessarily positive – divisors. We show it for Δ_+:

On U_{ij} there are holomorphic non-vanishing functions $a_{ij} \in \mathcal{O}^*(U_{ij})$ with

$$f_j = a_{ij} f_i \text{ on } U_{ij}.$$

So

$$g_j = a_{ij}\frac{h_j}{h_i} g_i.$$

Since $g_{i\mathbf{z}}$ and $h_{i\mathbf{z}}$ are coprime, the right-hand side can only be holomorphic if h_j/h_i is holomorphic – without zeros, because we can apply the same argument to h_i/h_j. This shows

$$a_{ij}\frac{h_j}{h_i} \in \mathcal{O}^*(U_{ij})$$

and verifies our claim for Δ_+. Hence

Proposition 8.3. *Every divisor is the quotient of positive divisors.* □

Now let $f \not\equiv 0$ be a meromorphic function on a polydisk D. Its divisor decomposes

$$\operatorname{div} f = \Delta_+/\Delta_-$$

into the quotient of two positive divisors. Since these are principal – by what we know by now – we find holomorphic functions g and h on D with

$$\operatorname{div} g = \Delta_+, \quad \operatorname{div} h = \Delta_-.$$

This implies

$$\operatorname{div} f = \operatorname{div}(g/h),$$

and so f and g/h differ by a function $a \in \mathcal{O}^*(D)$.

$$f = a \cdot \frac{g}{h}.$$

Hence

Theorem 8.4. *Meromorphic functions on a polydisk are quotients of globally defined holomorphic functions; the field $\mathcal{M}(D)$ is the quotient field of the ring $\mathcal{O}(D)$.* □

Exercises

1. Use the Mittag-Leffler and Weierstrass theorems from Chapter III to explicitly solve the Cousin and Poincaré problems in the plane. (This has been mentioned in the main text without explanation of the details.)

Chapter VII.

Holomorphic maps: Geometric aspects

We study holomorphic, in particular biholomorphic, maps between domains in $\mathbb{C}$ and, in one case, in $\mathbb{C}^n$, $n > 1$. These maps are for $n = 1$ *conformal* (angle and orientation preserving); so we shall use the terms *biholomorphic* and *conformal* interchangeably in this case. For $n > 1$ we consistently use *biholomorphic*. *Automorphisms* of domains, i.e. biholomorphic self-maps, are determined for disks resp. half-planes, the entire plane, and the sphere: they form groups consisting of Möbius transformations (VII.1). The proof of this fact relies on an important growth property of bounded holomorphic functions: the Schwarz lemma 1.3. Because the automorphism group of the unit disk (or upper half plane) acts transitively, it gives rise – according to F. Klein's Erlangen programme – to a geometry, which turns out to be the *hyperbolic (non-euclidean)* geometry (VII.2 and 3). The unit disk is conformally equivalent to almost all simply connected plane domains: Riemann's mapping theorem, proved in VII.4. For $n > 1$ even the immediate generalisations of the disk – the *polydisk* and the *unit ball* – are not biholomorphically equivalent (VII.4). Riemann's mapping theorem can be generalised to the general uniformization theorem (VII.4): a special but exceedingly useful case of this is the *modular map* λ which we introduce in VII.7. Its construction uses tools that are also expedient for other purposes: *harmonic functions* (with a solution of the Dirichlet problem for disks) and Schwarz's reflection principle (VII.5 and 6). The existence of λ finally yields two important classical results: Montel's and Picard's "big theorems".

Schwarz's lemma was established by H. A. Schwarz in 1869, its invariant form by G. Pick in 1915. C. Carathéodory [Ca] recognised its importance for the theory of bounded holomorphic functions where it is basic for most estimates. It also plays an important role in higher dimensions – see [La, Si]. For notes on the history of non-euclidean geometry we refer to the corresponding VII.3. – Riemann proved his mapping theorem in 1851 by potential theoretic methods which at that time had not yet been fully justified. Our proof – like most modern presentations – uses Montel's theorem. – The non-equivalence of the ball and the polydisk was discovered by H. Poincaré in 1907; our proof follows Range [Ra] who ascribes it to R. Remmert and K. Stein [RSt]. – The theory of harmonic functions is centuries old – we do not attempt to go into its history. The *reflection principle* – crucial for many questions on conformal mappings – was worked out and abundantly used by Schwarz from 1869 on. – The modular map is a special instance of an automorphic function. These functions are intimately tied up with elliptic functions and form an important part of 19^{th} and 20^{th} century mathematics: F. Klein, H. Poincaré ... : its history requires a book of its own. For a broader exposition we refer to [FL2] and [HuC]. The uniformization theorem 4.2 was established by Poincaré and P. Koebe at the beginning of the 20^{th} century. It is most satisfactorily dealt with in the framework of Riemann surfaces – see e.g. [FL2, Fo2].

1. Holomorphic automorphisms

Biholomorphic mappings of a domain $G \subset \widehat{\mathbb{C}}$ onto itself are called (holomorphic) *automorphisms* of G. A trivial example of such an automorphism is the identity mapping $\mathrm{id}\colon G \to G$; for some domains, this is the only automorphism. If f and g

W. Fischer, I. Lieb, *A Course in Complex Analysis*, DOI 10.1007/978-3-8348-8661-3_7,

are automorphisms of G, then so are $f \circ g$ and f^{-1}: The composition of mappings yields a group structure.

Definition 1.1. *Let G be a domain in $\widehat{\mathbb{C}}$. The group of biholomorphic mappings from G onto itself is called the automorphism group of G and is denoted* $\operatorname{Aut} G$.

We already determined the automorphism groups of $\widehat{\mathbb{C}}$ and $\mathbb{C}$ in Prop. III.4.2:

$\operatorname{Aut}\widehat{\mathbb{C}}$ *is the group of all Möbius transformations, and* $\operatorname{Aut}\mathbb{C}$ *is the group of entire linear transformations.*

The following proposition is useful in determining further automorphism groups:

Proposition 1.1. *Let $F\colon G_1 \to G_2$ be a biholomorphic mapping of domains in $\widehat{\mathbb{C}}$. If $f \in \operatorname{Aut} G_1$, then $F_*(f) = FfF^{-1} \in \operatorname{Aut} G_2$, and the mapping $F_*\colon \operatorname{Aut} G_1 \to \operatorname{Aut} G_2$ is a group isomorphism.*

We leave the easy verification to the reader.

To take an example, consider the biholomorphic mapping

$$S\colon \mathbb{H} \to \mathbb{D}, \quad z \mapsto \frac{i-z}{i+z}. \tag{1}$$

It follows that $\operatorname{Aut}\mathbb{H}$ and $\operatorname{Aut}\mathbb{D}$ are isomorphic! We will determine these groups, beginning with the upper half-plane. It will be seen that they consist of Möbius transformations – cf. III.4 for their elementary properties.

The translations $z \mapsto z + b$, where b is real, and the homotheties $z \mapsto \lambda z$, where $\lambda > 0$, are Möbius transformations that belong to $\operatorname{Aut}\mathbb{H}$. Every $z \in \mathbb{H}$ can be moved to the imaginary axis by such a translation and then moved to the point i with a homothety. The group $\operatorname{Aut}\mathbb{H} \cap \mathcal{M}$ thus acts transitively on $\mathbb{H}$, i.e. for every $z_1, z_2 \in \mathbb{H}$, there is a $T \in \operatorname{Aut}\mathbb{H} \cap \mathcal{M}$ such that $Tz_1 = z_2$.

We will show that all automorphisms of $\mathbb{H}$ that fix i are Möbius transformations:

$$\mathcal{F}_i = \{f \in \operatorname{Aut}\mathbb{H}\colon f(i) = i\} \subset \mathcal{M}. \tag{$*$}$$

It then easily follows that $\operatorname{Aut}\mathbb{H} \subset \mathcal{M}$: Let $h \in \operatorname{Aut}\mathbb{H}$, with $h(i) = z_0$. Then there exists a $T \in \operatorname{Aut}\mathbb{H} \cap \mathcal{M}$ such that $Tz_0 = i$. Therefore, i is a fixed point of $T \circ h$, and by $(*)$, we have $T \circ h \in \mathcal{M}$, so that $h \in \mathcal{M}$ as well.

Let us determine the Möbius transformations T with $T(\mathbb{H}) = \mathbb{H}$. This is equivalent to $T(\mathbb{R} \cup \{\infty\}) = \mathbb{R} \cup \{\infty\}$ (in which case $T(\mathbb{H}) = \mathbb{H}$ or $T(\mathbb{H}) = -\mathbb{H}$) and $T(i) \in \mathbb{H}$. For

$$Tz = \frac{\alpha z + \beta}{\gamma z + \delta}, \tag{2}$$

$T(\mathbb{R} \cup \{\infty\}) = \mathbb{R} \cup \{\infty\}$ is the same as saying that the coefficients α, β, γ, and δ can be chosen to be real – note that they are only determined up to a common factor by T! This is because we can write T as a cross-ratio:

$$Tz = \mathrm{CR}(z, z_1, z_2, z_3) = \frac{z - z_1}{z - z_3} : \frac{z_2 - z_1}{z_2 - z_3}, \tag{3}$$

where z_1, z_2, and z_3 are the preimages of 0, 1, and ∞ under T. If $\mathbb{R} \cup \{\infty\}$ is T-invariant, then $z_1, z_2, z_3 \in \mathbb{R} \cup \{\infty\}$, and (3) yields the formula (2) with real coefficients. The converse is trivial.

If the coefficients in (2) are real, then one computes

$$\operatorname{Im} T(i) = \frac{\alpha\delta - \beta\gamma}{\gamma^2 + \delta^2},$$

so that the statement $Ti \in \mathbb{H}$ is equivalent to the determinant $\alpha\delta - \beta\gamma$ being positive. We thus have:

Proposition 1.2. *The group* $\operatorname{Aut} \mathbb{H}$ *is the group of Möbius transformations*

$$Tz = \frac{\alpha z + \beta}{\gamma z + \delta}$$

such that $\alpha, \beta, \gamma, \delta \in \mathbb{R}$ *and* $\alpha\delta - \beta\gamma > 0$.

Thus, $\operatorname{Aut} \mathbb{H}$ is isomorphic to the group $\mathrm{SL}(2, \mathbb{R})/\{\pm I\}$, cf. Prop. III.4.3.

To completely prove this proposition, it remains to establish the claim $(*)$. This is made easier by transferring the problem to $\mathbb{D}$ via the isomorphism

$$S_*\colon \operatorname{Aut} \mathbb{H} \to \operatorname{Aut} \mathbb{D},\ T \mapsto STS^{-1},$$

where S is given by (1). Since $S(i) = 0$, the stabilizer $\mathcal{F}_i$ of i in $\operatorname{Aut} \mathbb{H}$ is sent to the stabilizer $\mathcal{F}_0$ of 0 in $\operatorname{Aut} \mathbb{D}$, and since S is in $\mathcal{M}$, it suffices to show that

$$\mathcal{F}_0 = \{g \in \operatorname{Aut} \mathbb{D} \colon g(0) = 0\} \subset \mathcal{M}. \tag{$*'$}$$

The proof of $(*')$ depends on a growth property of bounded holomorphic functions that is also important in many other contexts:

Theorem 1.3 (The Schwarz lemma). *Let* $f\colon \mathbb{D} \to \mathbb{D}$ *be a holomorphic function with* $f(0) = 0$. *Then* $|f(z)| \leqslant |z|$ *for all* $z \in \mathbb{D}$ *and* $|f'(0)| \leqslant 1$. *If* $|f(z_0)| = |z_0|$ *at some point* $z_0 \in \mathbb{D} \setminus \{0\}$ *or if* $|f'(0)| = 1$, *then* f *is a rotation, i.e.* $f(z) = e^{i\lambda} z$, *where* $\lambda \in \mathbb{R}$.

Proof: The function $g\colon \mathbb{D} \to \mathbb{C}$ given by

$$g(z) = \frac{f(z)}{z} \quad \text{for } z \neq 0, \quad g(0) = f'(0)$$

is holomorphic on $\mathbb{D}$, since $f(0) = 0$. For $|z| \leqslant r < 1$, the maximum modulus principle implies that

$$|g(z)| \leqslant \max_{|\zeta|=r} |g(\zeta)| < \frac{1}{r},$$

where we have used that $|f(\zeta)| < 1$. Letting $r \to 1$, we obtain $|g(z)| \leqslant 1$ for $z \in \mathbb{D}$, which is the first claim of the proposition. It follows from $|f(z_0)| = |z_0|$ or $|f'(z_0)| = 1$ that $|g|$ attains a maximum at z_0 or 0, respectively. But then g is a constant of absolute value 1. □

Using the Schwarz lemma, we can easily determine the group $\mathcal{F}_0$. Let $g \in \operatorname{Aut}\mathbb{D}$ be such that $g(0) = 0$. Then $|g'(0)| \leqslant 1$ by Thm. 1.3. The point 0 is also a fixed point of the inverse map g^{-1}, so that $|(g^{-1})'(0)| \leqslant 1$ as well. Using $g'(0) \cdot (g^{-1})'(0) = 1$, it follows that $|g'(0)| = 1$, and by Thm. 1.3, $g(z) = e^{i\lambda} z$, where $\lambda \in \mathbb{R}$. Conversely, all of these rotations of course belong to $\mathcal{F}_0$. We have thus shown:

The group $\mathcal{F}_0$ is the group of rotations about the origin; in particular, $\mathcal{F}_0 \subset \mathcal{M}$.

Prop. 1.2 is thus completely proved. Moreover, $S \in \mathcal{M}$ implies that $\operatorname{Aut}\mathbb{D} = S_*(\operatorname{Aut}\mathbb{H})$ consists of Möbius transformations as well.

In order to explicitly determine the elements of $\operatorname{Aut}\mathbb{D}$, one can compute STS^{-1} for $T \in \operatorname{Aut}\mathbb{H}$ – we leave this to the reader. One obtains

Proposition 1.4. *The group* $\operatorname{Aut}\mathbb{D}$ *is the group of Möbius transformations that can be written in the form*

$$Tz = \frac{az + b}{\overline{b}z + \overline{a}}, \tag{4}$$

where $a, b \in \mathbb{C}$ and $a\overline{a} - b\overline{b} > 0$.

Let us interpret (4) in a geometric way. Since $a\overline{a} - b\overline{b} > 0$, we have $a \neq 0$ and $|b/a| < 1$. We write

$$Tz = \frac{az + b}{\overline{b}z + \overline{a}} = \frac{a}{\overline{a}} \cdot \frac{z + (b/a)}{\overline{(b/a)}z + 1},$$

and if we define $\lambda \in \mathbb{R}$ by $a = |a|\, e^{i\lambda/2}$ and set $-b/a = z_0 \in \mathbb{D}$, then

$$Tz = e^{i\lambda} \frac{z - z_0}{1 - \overline{z}_0 z} \tag{5}$$

with $\lambda \in \mathbb{R}$ and $z_0 \in \mathbb{D}$. Formula (5) decomposes T into an automorphism that sends z_0 to 0 and a rotation about the origin.

To conclude, we note that every biholomorphic mapping between domains G_1 and G_2 in $\widehat{\mathbb{C}}$ whose boundaries are Möbius circles is a Möbius transformation. In particular, $\operatorname{Aut} G \subset \mathcal{M}$ for every such G. In fact, by Prop. III.4.6, there exist transformations $T_j \in \mathcal{M}$ such that $T_j(\mathbb{H}) = G_j$, where $j = 1, 2$. If $f\colon G_1 \to G_2$ is biholomorphic, then $T_2^{-1} f T_1 \in \operatorname{Aut}\mathbb{H} \subset \mathcal{M}$ and hence $f \in \mathcal{M}$.

Exercises

1. a) Show that, given $z_0 \in \mathbb{D}$ and $z_1, z_2 \in \partial\mathbb{D}$, there is exactly one $T \in \operatorname{Aut}\mathbb{D}$ such that $Tz_0 = z_0$ and $Tz_1 = z_2$.

   b) Given $z_0 \in \mathbb{D}$, describe the group $\{T \in \operatorname{Aut}\mathbb{D}\colon Tz_0 = z_0\}$; given $z_0, w_0 \in \mathbb{D}$, describe the set $\{T \in \operatorname{Aut}\mathbb{D}\colon Tz_0 = w_0\}$.

2. Use the Schwarz lemma to derive statements of a similar type for holomorphic functions $f\colon D_r(a) \to \mathbb{C}$ such that $|f| < M$ or $f\colon D_r(a) \to D_\rho(b)$ such that $f(a) = b$.

3. Let $z_1, \dots, z_r$ be distinct points in $\widehat{\mathbb{C}}$ and $G = \widehat{\mathbb{C}} \setminus \{z_1, \dots, z_r\}$.

   a) Show that every automorphism of G can be extended to a $T \in \mathcal{M}$ that permutes the points $z_1, \dots, z_r$.

   b) Determine $\operatorname{Aut} G$ for $r = 1, 2, 3$.

   c) Investigate the case $r = 4$ ($\operatorname{Aut} G$ depends on $\mathrm{CR}(z_1, z_2, z_3, z_4)$!).

4. Determine $\operatorname{Aut} G$ for $G = \mathbb{D} \setminus \{0\}$, $G = \mathbb{D} \setminus \{0, 1/2\}$, and $G = \mathbb{D} \setminus \{0, 1/2, i/3\}$.

2. The hyperbolic metric

Path lengths in $\mathbb{C}$ are invariant under euclidean movements: $L(S \circ \gamma) = L(\gamma)$ for $S\colon z \mapsto az + b$ with $|a| = 1$. We now derive another way to measure path lengths in $\mathbb{D}$ and $\mathbb{H}$ that is invariant under $\operatorname{Aut}\mathbb{D}$ and $\operatorname{Aut}\mathbb{H}$, respectively. This *hyperbolic metric* is, on the one hand, essential to the theory of Riemann surfaces; it also serves to construct a non-euclidean geometry. We will treat the latter topic in the next section.

In the previous section, we saw that $\operatorname{Aut}\mathbb{D}$ acts transitively on $\mathbb{D}$: for any $z_0, w_0 \in \mathbb{D}$, there exists a $T \in \operatorname{Aut}\mathbb{D}$ such that $Tz_0 = w_0$. In particular,

$$T_{z_0}\colon z \mapsto \frac{z - z_0}{1 - \overline{z}_0 z} \tag{1}$$

sends $z_0 \in \mathbb{D}$ to the origin.

On the other hand, given two pairs of distinct points (z_0, z_1) and (w_0, w_1) in $\mathbb{D}$, there does not always exist a $T \in \operatorname{Aut}\mathbb{D}$ such that $Tz_0 = w_0$ and $Tz_1 = w_1$: If, say, $z_0 = w_0 = 0$ and $Tz_0 = w_0$, then T is a rotation about 0, and $Tz_1 = w_1$ is possible if and only if $|z_1| = |w_1|$. An arbitrary pair of points (z_0, z_1) can be sent to (w_0, w_1) by an automorphism of $\mathbb{D}$ if and only if $(T_{z_0} z_0, T_{z_0} z_1) = (0, T_{z_0} z_1)$ can be sent to $(T_{w_0} w_0, T_{w_0} w_1) = (0, T_{w_0} w_1)$, i.e. if and only if $|T_{z_0} z_1| = |T_{w_0} w_1|$.

Proposition 2.1. *Let (z_0, z_1) and (w_0, w_1) be pairs of distinct points in $\mathbb{D}$. Then there exists a $T \in \operatorname{Aut}\mathbb{D}$ such that $Tz_j = w_j$ $(j = 0, 1)$, if and only if*

$$\left|\frac{z_1 - z_0}{1 - \overline{z}_0 z_1}\right| = \left|\frac{w_1 - w_0}{1 - \overline{w}_0 w_1}\right|.$$

If such a T exists, then it is unique.

In particular, for $T \in \operatorname{Aut}\mathbb{D}$ and $z, z_0 \in \mathbb{D}$, one always has

$$\left|\frac{Tz - Tz_0}{1 - \overline{Tz_0} \cdot Tz}\right| = \left|\frac{z - z_0}{1 - \overline{z}_0 z}\right|.$$

Dividing by $z - z_0$, then letting $z \to z_0$, one obtains the relation

$$\frac{|T'(z_0)|}{1 - |Tz_0|^2} = \frac{1}{1 - |z_0|^2}, \tag{2}$$

which we will need below.

In order to define the hyperbolic length of a path in $\mathbb{D}$, we begin with a general consideration concerning the concept of length. The euclidean length of a path of integration $\gamma\colon [a, b] \to G$ in a domain $G \subset \mathbb{C}$ is the integral over the length of its tangent vectors: $L(\gamma) = \int_a^b |\gamma'(t)|\, dt$. One can now modify this concept by weighting the tangent vectors, i.e. by choosing a continuous, positive function λ on G and setting

$$L_\lambda(\gamma) = \int_a^b \lambda(\gamma(t))|\gamma'(t)|\, dt. \tag{3}$$

An immediate consequence of this definition is:

i. $L_\lambda(\gamma) \geqslant \varepsilon(\gamma) \cdot L(\gamma) > 0$, provided γ is not constant; here

$$\varepsilon(\gamma) = \min\{\lambda(z)\colon z \in \operatorname{Tr}\gamma\}.$$

ii. $L_\lambda(-\gamma) = L_\lambda(\gamma)$.

iii. $L_\lambda(\gamma_1 + \gamma_2) = L_\lambda(\gamma_1) + L_\lambda(\gamma_2)$, provided $\gamma_1 + \gamma_2$ is defined.

Under which condition is the "λ-length" invariant under $\operatorname{Aut} G$, i.e.

iv. $L_\lambda(T \circ \gamma) = L_\lambda(\gamma)$ for all $T \in \operatorname{Aut} G$ and all γ?

With

$$L_\lambda(T \circ \gamma) = \int_a^b \lambda(T\gamma(t)) \cdot |(T \circ \gamma)'(t)|\, dt = \int_a^b \lambda(T\gamma(t)) \cdot |T'(\gamma(t))| \cdot |\gamma'(t)|\, dt,$$

we obtain a sufficient (and necessary) condition for the validity of *iv*, namely

$$\lambda(Tz) \cdot |T'(z)| = \lambda(z) \tag{4}$$

for all $z \in G$ and $T \in \operatorname{Aut} G$.

Let us apply this to $G = \mathbb{D}$. Substituting T_{z_0} from (1) into (4), we obtain for $z = z_0$

$$\lambda(0) \cdot |T'_{z_0}(z_0)| = \lambda(z_0)$$

Computing $|T'_{z_0}(z_0)|$ and writing z for z_0, a necessary condition for the Aut $\mathbb{D}$-invariance of L_λ is seen to be

$$\lambda(z) = \frac{1}{1-|z|^2}\lambda(0).$$

We normalize this by putting $\lambda(0) = 1$, so that

$$\lambda(z) = \frac{1}{1-|z|^2}. \tag{5}$$

Equation (4) then reads

$$\frac{|T'(z)|}{1-|Tz|^2} = \frac{1}{1-|z|^2},$$

and by (2), this is satisfied for all $T \in \operatorname{Aut}\mathbb{D}$, so that L_λ is indeed Aut $\mathbb{D}$-invariant.

Definition 2.1. *Let $\gamma\colon [a,b] \to \mathbb{D}$ be a path of integration. The hyperbolic length of γ is given by*

$$L_h(\gamma) = \int_a^b \frac{|\gamma'(t)|}{1-|\gamma(t)|^2}\, dt. \tag{6}$$

By construction, it has the properties *i* to *iv*; in *i* we have $L_h(\gamma) \geqslant L(\gamma)$.

As an example, let us compute the hyperbolic length of the segment $[0,s]$, where $0 \leqslant s < 1$:

$$L_h([0,s]) = \int_0^s \frac{dx}{1-x^2} = \frac{1}{2}\log\frac{1+s}{1-s}. \tag{7}$$

As $s \to 1$, $L_h([0,s])$ grows to ∞!

A definition of length of the type (3) or (6) leads to a new definition of distance:

Definition 2.2. *The hyperbolic distance (h-distance) between $z_0, z_1 \in \mathbb{D}$ is*

$$\delta(z_0,z_1) = \inf\{L_h(\gamma)\colon \gamma \textit{ is a path of integration in } \mathbb{D} \textit{ from } z_0 \textit{ to } z_1\}.$$

By conditions *i* to *iv*, δ indeed satisfies the axioms of a metric on $\mathbb{D}$ and is invariant under automorphisms of $\mathbb{D}$.

This raises the question whether, for given points $z_0, z_1 \in \mathbb{D}$ (where $z_0 \neq z_1$), there exists a path of minimal h-length connecting them, i.e. that turns the infimum in Def. 2.2 into a minimum. Using a $T \in \operatorname{Aut}\mathbb{D}$, we first send z_0 to 0 and z_1 to the point

$$s = \left|\frac{z_0 - z_1}{1-\overline{z_1}z_0}\right| \in\,]0,1[.$$

Let $\gamma = \gamma_1 + i\gamma_2 \colon [a,b] \to \mathbb{D}$ be an arbitrary path of integration from 0 to s. Then

$$L_h(\gamma) = \int_a^b \frac{|\gamma'(t)|\,dt}{1-|\gamma(t)|^2} \geqslant \int_a^b \frac{\gamma_1'(t)\,dt}{1-|\gamma_1(t)|^2} = \int_0^s \frac{dx}{1-x^2} = L_h([0,s]), \tag{8}$$

so that the segment $[0,s]$ is the hyperbolically shortest path joining 0 and s. The preimage of $[0,s]$ under T is then the shortest path joining z_0 and z_1. Since T^{-1} is angle-preserving, the preimage of the diameter of $\mathbb{D}$ passing through s is a circular arc (or a diameter) that intersects the boundary circle $\mathbb{S} = \partial\mathbb{D}$ perpendicularly. We call such arcs, as well as diameters of $\mathbb{D}$, *orthocircles*. We have:

Proposition 2.2. *Two distinct points $z_0, z_1 \in \mathbb{D}$ lie on a unique orthocircle. The hyperbolically shortest path joining z_0 to z_1 is the corresponding arc of this orthocircle.*

To use a term from differential geometry, orthocircles are the geodesics of the hyperbolic metric on $\mathbb{D}$.

The relationships (7) and (8) yield a formula for the hyperbolic distance:

$$\delta(z_0,z_1) = \frac{1}{2}\log\frac{1+d(z_0,z_1)}{1-d(z_0,z_1)}, \quad \text{where} \quad d(z_0,z_1) = \left|\frac{z_0-z_1}{1-\overline{z}_1 z_0}\right|. \tag{9}$$

Using the hyperbolic tangent

$$\tanh x = \frac{e^{2x}-1}{e^{2x}+1},$$

this formula can be rewritten as

$$\tanh\delta(z_0,z_1) = \left|\frac{z_0-z_1}{1-\overline{z}_1 z_0}\right| = d(z_0,z_1). \tag{10}$$

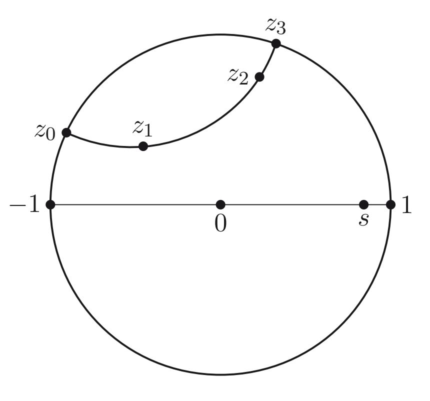


Figure 17. Orthocircles; $\delta(0,s) = \delta(z_1,z_2)$

Using (9) or (10), we can reformulate Prop. 2.1: Two pairs of distinct points (z_0, z_1) and (w_0, w_1) in $\mathbb{D}$ can be mapped to one another by an automorphism of $\mathbb{D}$ if and only if $\delta(z_0, z_1) = \delta(w_0, w_1)$.

The h-distance between two points is not smaller than their euclidean distance: $\delta(z_0, z_1) \geqslant |z_1 - z_0|$. It can, however, be much larger: Keeping z_0 fixed and letting z_1 approach the boundary $\mathbb{S}$, we see that $\delta(z_0, z_1) \to \infty$, since $\delta(0, |z_1|) = \delta(0, z_1) \leqslant \delta(0, z_0) + \delta(z_0, z_1)$ and $\delta(0, |z_1|) \to \infty$ as $|z_1| \to 1$.

Happily, we nevertheless have the following:

Proposition 2.3. *The hyperbolic metric induces the same topology on $\mathbb{D}$ as the euclidean metric.*

Proof: It suffices to show that both metrics lead to the same notion of convergence. Since $\delta(z, w) \geqslant |z - w|$ and the function δ is continuous in both variables, this is clear. □

We now transfer these concepts to the upper half-plane $\mathbb{H}$ via a biholomorphic mapping $S\colon \mathbb{H} \to \mathbb{D}$.

Given a path of integration $\gamma\colon [a, b] \to \mathbb{H}$, we define its hyperbolic length to be

$$L_{\mathbb{H}}(\gamma) = L_h(S \circ \gamma).$$

This is independent of the choice of S: If $\widetilde{S}\colon \mathbb{H} \to \mathbb{D}$ is also biholomorphic, then $\widetilde{S}S^{-1} \in \operatorname{Aut}\mathbb{D}$, so that

$$L_h(\widetilde{S}\gamma) = L_h((\widetilde{S}S^{-1})S\gamma) = L_h(S\gamma).$$

We obtain $\operatorname{Aut}\mathbb{H}$-invariance in similar fashion: For $T \in \operatorname{Aut}\mathbb{H}$,

$$L_{\mathbb{H}}(T\gamma) = L_h((STS^{-1})S\gamma) = L_h(S\gamma) = L_{\mathbb{H}}(\gamma),$$

since $STS^{-1} \in \operatorname{Aut}\mathbb{D}$.

In order to get an explicit expression for $L_{\mathbb{H}}$, we use the special mapping

$$S\colon z \mapsto \frac{i - z}{i + z}, \quad \mathbb{H} \to \mathbb{D}.$$

A short calculation shows that $|S'(z)|(1 - |S(z)|^2)^{-1} = (2\operatorname{Im} z)^{-1}$, so that

$$L_{\mathbb{H}}(\gamma) = \int_a^b \frac{|(S \circ \gamma)'(t)|\, dt}{1 - |S \circ \gamma(t)|^2} = \int_a^b \frac{|\gamma'(t)|\, dt}{2\operatorname{Im} \gamma(t)}.$$

For the corresponding h-distance on $\mathbb{H}$, namely

$$\delta_{\mathbb{H}}(z_0, z_1) = \inf\bigl\{L_{\mathbb{H}}(\gamma)\colon \gamma \text{ is a path of integration in } \mathbb{H} \text{ from } z_0 \text{ to } z_1\bigr\},$$

we have $\delta_{\mathbb{H}}(z_0, z_1) = \delta(Sz_0, Sz_1)$ and the explicit formula

$$\delta_{\mathbb{H}}(z_0, z_1) = \frac{1}{2} \log \frac{1 + d^*(z_0, z_1)}{1 - d^*(z_0, z_1)}, \quad \text{where} \quad d^*(z_0, z_1) = d(Sz_0, Sz_1) = \left| \frac{z_0 - z_1}{z_0 - \overline{z}_1} \right|,$$

or, as before,

$$\tanh \delta_{\mathbb{H}}(z_0, z_1) = d^*(z_0, z_1).$$

Geodesics with respect to the metric $L_{\mathbb{H}}$ are the preimages, under S, of orthocircles in $\mathbb{D}$, i.e. semicircles (whose centres lie on $\mathbb{R}$) or vertical rays $\{z\colon \operatorname{Re} z = c, \operatorname{Im} z > 0\}$ in $\mathbb{H}$. We also call these lines orthocircles in $\mathbb{H}$.

Examples:

i. For $\lambda > 1$, we have

$$\delta_{\mathbb{H}}(i, \lambda i) = \int_1^\lambda \frac{dt}{2t} = \frac{1}{2} \log \lambda.$$

Using $\operatorname{Aut} \mathbb{H}$-invariance, this yields

$$\delta_{\mathbb{H}}(x_1 + \lambda_1 i, x_1 + \lambda_2 i) = \frac{1}{2} \left| \log(\lambda_1 / \lambda_2) \right|.$$

ii. Suppose the points z_1, z_2 lie on the orthocircle with centre a and radius r, e.g. $z_1 - a = r\, e^{i\varphi}$ and $z_2 - a = r\, e^{i\psi}$, where $0 < \varphi \leqslant \psi < \pi$. Then

$$\delta_{\mathbb{H}}(z_1, z_2) = \int_\varphi^\psi \frac{dt}{2 \sin t} = \frac{1}{2} \log \frac{\tan(\psi/2)}{\tan(\varphi/2)}.$$

To conclude, we formulate a generalization of the Schwarz lemma with the help of the hyperbolic metric:

Proposition 2.4 (The Schwarz-Pick lemma). *A holomorphic function $f\colon \mathbb{D} \to \mathbb{D}$ is a contraction for the hyperbolic distance, i.e. for all $z_1, z_2 \in \mathbb{D}$, we have $\delta(f(z_1), f(z_2)) \leqslant \delta(z_1, z_2)$. If equality holds for a pair of distinct points, then f is an automorphism of $\mathbb{D}$.*

Proof: If $z_1 \in \mathbb{D}$ is arbitrary and $w_1 = f(z_1)$, then $T_{w_1} \circ f \circ T_{z_1}^{-1}$ satisfies the conditions of the Schwarz lemma (here T_{z_1} and T_{w_1} are defined by (1)). We thus have either $|T_{w_1} \circ f \circ T_{z_1}^{-1}(\zeta)| < |\zeta|$ for $\zeta \in \mathbb{D} \setminus \{0\}$ or $T_{w_1} \circ f \circ T_{z_1}^{-1}(\zeta) = e^{i\lambda}\zeta$, where $\lambda \in \mathbb{R}$. In the latter case, $f \in \operatorname{Aut} \mathbb{D}$, and in the former case, substituting $\zeta = T_{z_1}(z)$ yields

$$\tanh \delta(f(z), f(z_1)) = \left| \frac{f(z) - f(z_1)}{1 - \overline{f(z_1)} f(z)} \right| < \left| \frac{z - z_1}{1 - \overline{z}_1 z} \right| = \tanh \delta(z, z_1) \tag{11}$$

for $z \neq z_1$. The claim now follows from the fact that tanh is strictly increasing. $\square$

By passing to the limit as $z \to z_1$ in the inequality (11), we obtain an "infinitesimal" version of the proposition that generalizes equation (2):

Corollary 2.5. *Holomorphic functions* $f\colon \mathbb{D} \to \mathbb{D}$ *satisfy*

$$\frac{|f'(z)|}{1-|f(z)|^2} \leqslant \frac{1}{1-|z|^2}.$$

Exercises

1. Prove the formula $\delta(z_1, z_2) = \frac{1}{2}\big|\mathrm{CR}(z_1, z_0, z_2, z_3)\big|$. Here $z_1 \neq z_2$, and z_0 and z_3 are the points of intersection of the orthocircle that passes through z_1 and z_2 with the boundary $\partial\mathbb{D}$.
2. Transfer Prop. 2.4 and its corollary to holomorphic functions $g\colon \mathbb{H} \to \mathbb{H}$.
3. Show that
$$d(z_0, z_1) = \left|\frac{z_0 - z_1}{1 - \overline{z}_1 z_0}\right|$$
defines a metric on $\mathbb{D}$.
4. Let $\gamma\colon [a, b] \to \mathbb{D}$ be continuously differentiable, and for every partition $Z\colon a = t_0 < t_1 < \ldots < t_n = b$ of $[a, b]$, let
$$S_Z(\gamma) = \sum_{\nu=1}^{n} d(\gamma(t_{\nu-1}), \gamma(t_\nu)).$$
Show that $L_h(\gamma) = \sup_Z S_Z(\gamma)$.
5. Show that for $w, z \in \mathbb{H}$, we have $\cosh 2\delta_{\mathbb{H}}(w, z) = 1 + \frac{|w-z|^2}{2(\operatorname{Im} w)(\operatorname{Im} z)}$.

3. Hyperbolic geometry

The concepts of point, line, segment, distance, and so on are fundamental in plane euclidean geometry. We now give these words a new meaning and thereby define a new geometry: hyperbolic geometry. It will be seen that many statements that are valid in euclidean geometry are valid in hyperbolic geometry as well – and also that there exist important differences between these two geometries.

The construction of new "non-euclidean" geometries was essential in clarifying the (axiomatic) foundations of geometry; we will address this point at the end of the section.

The basic idea in our construction of hyperbolic geometry is to regard only points in the interior of the unit disk $\mathbb{D}$ as points of our geometry and to use the hyperbolic metric on $\mathbb{D}$ to measure distances. The results of the previous section suggest the following terminology:

The hyperbolic plane (h-plane) is $\mathbb{D}$. *Straight lines* in the h-plane (h-lines for short) are the orthocircles in $\mathbb{D}$ (including diameters of $\mathbb{D}$). The *h-segment* $[z_1, z_2]_h$ from z_1 to z_2 (where $z_1 \neq z_2$) is the arc between z_1 and z_2 on the orthocircle that passes through z_1 and z_2. The *h-length* of this segment is the hyperbolic distance $\delta(z_1, z_2)$.

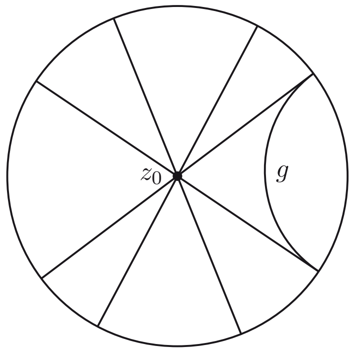


Figure 18. h-parallels through z_0 to g

These h-concepts are invariant under $\operatorname{Aut}\mathbb{D}$; for example, given an arbitrary $T \in \operatorname{Aut}\mathbb{D}$, the image of the h-segment $[z_1, z_2]_h$ under T is the h-segment $[Tz_1, Tz_2]_h$, and the h-length remains the same. We may thus regard $\operatorname{Aut}\mathbb{D}$ as a group of orientation-preserving isometries of the h-plane; later, we will show that $\operatorname{Aut}\mathbb{D}$ is in fact the full group of such isometries (see Prop. 3.6).

Many elementary geometric concepts can also be formulated in terms of h-geometry, e.g. rays, angles (two rays emanating from the same point), and triangles. We leave it to the reader to give precise formulations.

Since biholomorphic mappings are angle-preserving, euclidean angle measure is invariant under $\operatorname{Aut}\mathbb{D}$; this legitimizes measuring angles in the h-plane in the usual way.

By regarding orthocircles in $\mathbb{D}$, the following claims are seen to be valid in both h-geometry and euclidean geometry:

– Two distinct points determine a unique h-line.

– Two distinct h-lines intersect in at most one point.

– For every h-line g_1 and every point $z_0 \in g_1$, there exists a unique perpendicular to g_1 that passes through z_0, i.e. there exists a unique h-line g_2 that intersects g_1 at a right angle at the point z_0. (Using an h-isometry, g_1 can be made to be a diameter of $\mathbb{D}$, whence the claim is evident.)

– For every h-line g_1 and every point $z_0 \notin g_1$, there is a unique perpendicular to g_1 that passes through z_0, i.e. there is a unique h-line g_2 through z_0 that intersects g_1 at a right angle.

The most important difference is this: In the euclidean plane, we have the parallel axiom, namely "for every straight line g and every point $P \notin g$, there exists exactly one straight line that passes through P and does not intersect g". In h-geometry, for every h-line g and every point $z_0 \notin g$, there exist infinitely many h-lines that pass through z_0 and do not intersect g. This is clear if one chooses z_0 to be the origin – h-lines that pass through 0 are diameters of $\mathbb{D}$.

If one founds euclidean geometry on an axiom system that includes the parallel axiom, then one can show the following: In h-geometry, all theorems of euclidean geometry are valid whose proofs do not require the parallel axiom.

The points of the boundary $\mathbb{S}$ of $\mathbb{D}$ do not belong to the hyperbolic plane. But because $\delta(z_0, z) \to \infty$ as $|z| \to 1$, we may regard them as "points at infinity". There are thus two points at infinity associated to every h-line (whereas there is only one point at infinity associated to every line in the projective closure of the euclidean plane).

We would like to measure not only length but also area in the h-plane. Differential geometry tells us that if we have a definition of length in a domain $G \subset \mathbb{C}$, namely $L_\lambda(\gamma) = \int_a^b \lambda\big(\gamma(t)\big)|\gamma'(t)|\,dt$ (for a path $\gamma\colon [a,b] \to G$), then the corresponding area measure is given by $F_\lambda(A) = \int_A \lambda^2(z)\,dx\,dy$. Thus:

Definition 3.1. *The h-area of $A \subset \mathbb{D}$ is given by*

$$F_h(A) = \int\limits_A \frac{dx\,dy}{(1-|z|^2)^2},$$

provided the integral exists.

Area is invariant: $F_h(TA) = F_h(A)$ for any $T \in \operatorname{Aut}\mathbb{D}$.

Since biholomorphic mappings from $\mathbb{D}$ to $\mathbb{H}$ are isometries with respect to the hyperbolic metric, we can transfer the h-geometry of $\mathbb{D}$ to $\mathbb{H}$. We then speak of the half-plane model and also call $\mathbb{D}$ the disk model of h-geometry. Some considerations and computations are more easily carried out in the half-plane model. – More explicitly: h-lines in $\mathbb{H}$ are the orthocircles in $\mathbb{H}$, i.e. semicircles whose centres lie on $\mathbb{R}$, as well as vertical euclidean rays $\{z\colon \operatorname{Re} z = c\} \cap \mathbb{H}$. The points at infinity are now the points on $\mathbb{R} \cup \{\infty\}$. The area of $B \subset \mathbb{H}$ is given by

$$F_{\mathbb{H}}(B) = \int\limits_B \frac{dx\,dy}{(2y)^2}.$$

It is invariant under $\operatorname{Aut}\mathbb{H}$ and satisfies $F_{\mathbb{H}}(B) = F_h(SB)$ for biholomorphic mappings $S\colon \mathbb{H} \to \mathbb{D}$.

We now investigate circles and triangles in the disk model. The *h-circle* of h-radius r centred at $z_0 \in \mathbb{D}$ is naturally defined to be $K_r(z_0) = \{z \in \mathbb{D}\colon \delta(z, z_0) = r\}$. In case $z_0 = 0$, this is a euclidean circle of radius $\rho = \tanh r$ centred at 0, in view of $\delta(z,0) = \delta(|z|,0)$. As automorphisms of $\mathbb{D}$ send circles to circles, $K_r(z_0)$ is always a euclidean circle in $\mathbb{D}$ whose euclidean centre, however, is different from z_0 for $z_0 \neq 0$.

In order to find the h-length of $K_r(z_0)$, we may assume that $z_0 = 0$ and parametrize $K_r(0)$ as $\gamma(t) = \rho\, e^{it}$. Formula (4) below yields

$$L_h\big(K_r(0)\big) = \int_0^{2\pi} \frac{\rho\, dt}{1-\rho^2} = \frac{2\pi\rho}{1-\rho^2} = 2\pi \frac{\tanh r}{1-\tanh^2 r} = \pi \sinh(2r).$$

To find the area of the disk $D_r^h(z_0) = \{z \in \mathbb{D}\colon \delta(z, z_0) \leqslant r\}$, we again assume that $z_0 = 0$ and obtain

$$F_h\big(D_r^h(z_0)\big) = \int_{|z|\leqslant\rho} \frac{dx\, dy}{(1-|z|^2)^2} = 2\pi \int_0^{\rho} \frac{t\, dt}{(1-t^2)^2} = \frac{\pi\rho^2}{1-\rho^2} = \pi \sinh^2 r.$$

Since $\sinh x > x$ for $x > 0$, h-circumferences and h-areas of disks are always larger than their corresponding euclidean values.

An *h-triangle* is determined by its vertices z_a, z_b, and z_c, which of course must not all lie on the same h-line. Its edges are the h-segments $[z_b, z_c]_h$, etc., and we denote their h-lengths by $a = \delta(z_b, z_c)$, etc. and the interior angles by α (for z_a), etc.

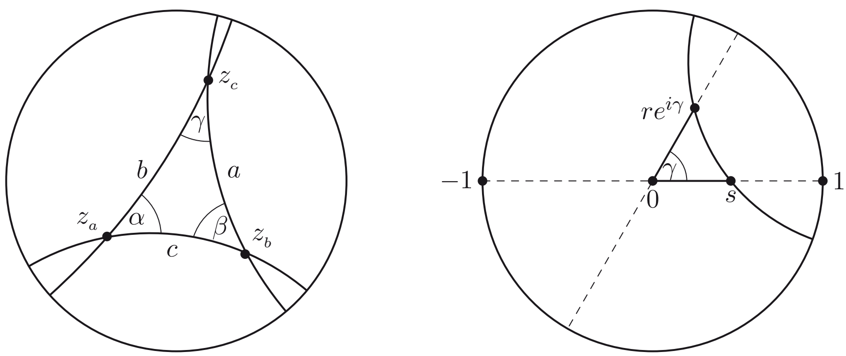


Figure 19. h-triangles

One can transform a given h-triangle with an automorphism $T \in \operatorname{Aut} \mathbb{D}$ in such a way that one of its vertices is the origin; the edges that meet there lie on radii of $\mathbb{D}$. Put in this position, it is evident that the sum $\alpha + \beta + \gamma$ of the angles of an h-triangle is less than π – contrary to euclidean geometry!

The sum of the h-lengths of two sides of a triangle is greater than the h-length of the third: δ satisfies the triangle inequality. Conversely, given three side lengths that satisfy this requirement, one can, following Euclid, construct a triangle. Let

$$\max(a, b) \leqslant c < a + b. \tag{1}$$

Choose an h-segment $[z_a, z_b]_h$ of length c and draw the circles $K_b(z_a)$ and $K_a(z_b)$. By (1), these intersect at two points z_c' and z_c'', and together with z_a and z_b, each of these points yields a triangle with the desired lengths (these two triangles are mirror images of one another with respect to the h-line that passes through z_a and z_b).

It is occasionally useful to consider "improper" triangles as well, namely triangles at least one of whose vertices is at infinity, i.e. belongs to $\mathbb{S}$ (or to $\mathbb{R} \cup \{\infty\}$, in the half-plane model). The interior angle corresponding to a vertex at infinity is 0, and the edges that meet at such a vertex have infinite length. Nevertheless, the area remains finite, as we will soon see.

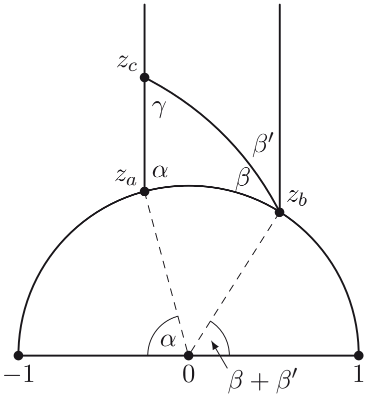


Figure 20. For the proof of Prop. 3.1

In order to compute the area of a triangle Δ with vertices z_a, z_b, and z_c, it is practical to work with the half-plane model. After applying a suitable automorphism, we can assume that z_a and z_b lie on the unit circle, that the h-line through z_a and z_c is a vertical ray, and that $\operatorname{Im} z_c > \operatorname{Im} z_a$ (see Fig. 20). We next compute the area of the improper triangle Δ_1 with vertices z_a, z_b, and ∞ and angles α, $\beta+\beta'$, and 0. We have

$$\Delta_1 = \left\{ z = x + iy \colon \cos(\pi - \alpha) \leqslant x \leqslant \cos(\beta+\beta'),\ y \geqslant \sqrt{1-x^2} \right\},$$

$$\begin{aligned} 4F_{\mathbb{H}}(\Delta_1) &= \int_{\Delta_1} \frac{dx\,dy}{y^2} = \int_{\cos(\pi-\alpha)}^{\cos(\beta+\beta')} \left(\int_{\sqrt{1-x^2}}^{\infty} \frac{dy}{y^2} \right) dx \\ &= \int_{\cos(\pi-\alpha)}^{\cos(\beta+\beta')} \frac{dx}{\sqrt{1-x^2}} = \int_{\beta+\beta'}^{\pi-\alpha} dt = \pi - \alpha - (\beta+\beta'). \end{aligned}$$

Similarly, the area of the improper triangle Δ_2 with vertices z_c, z_b, and ∞ and angles $\pi-\gamma$, β', and 0 is seen to be $4F_{\mathbb{H}}(\Delta_2) = \pi - (\pi-\gamma) - \beta'$. We thus have

$$4F_{\mathbb{H}}(\Delta) = 4F_{\mathbb{H}}(\Delta_1) - 4F_{\mathbb{H}}(\Delta_2) = \pi - (\alpha+\beta+\gamma).$$

Proposition 3.1. *The area of an h-triangle* Δ *with interior angles* α, β, *and* γ *is*

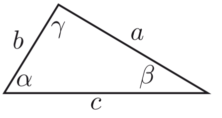
$$F(\Delta) = \frac{1}{4}\big(\pi - (\alpha+\beta+\gamma)\big).$$

This again implies that the sum of angles $\alpha+\beta+\gamma$ is always less than π, because area is positive! Moreover, it can be seen that there do not exist h-triangles of arbitrarily large area: $\pi/4$ is an upper bound – it is assumed by improper triangles all of whose vertices are at infinity.

In euclidean geometry, computations involving triangles are governed by trigonometric functions. The fundamental laws are:

The law of sines: $\dfrac{\sin\alpha}{a} = \dfrac{\sin\beta}{b} = \dfrac{\sin\gamma}{c}.$

The law of cosines: $c^2 = a^2 + b^2 - 2ab\cos\gamma.$



We will derive analogues to these laws for h-geometry, as well as a second law of cosines. In doing so, the h-lengths of edges will appear as arguments of hyperbolic functions – this is one reason for the name "hyperbolic geometry". First, we give several formulas; they are easily derived from the definitions of the hyperbolic functions:

$$\cosh^2 x - \sinh^2 x = 1 \tag{2}$$

$$\cosh(2x) = \cosh^2 x + \sinh^2 x = \frac{1+\tanh^2 x}{1-\tanh^2 x} \tag{3}$$

$$\sinh(2x) = 2\cosh x \cdot \sinh x = \frac{2\tanh x}{1-\tanh^2 x}. \tag{4}$$

We consider an h-triangle in $\mathbb{D}$, using the usual notation.

Proposition 3.2 (The first law of cosines)**.**

$$\cosh(2c) = \cosh(2a)\cosh(2b) - \sinh(2a)\sinh(2b)\cos\gamma.$$

Proof: By applying a suitable transformation in $\operatorname{Aut}\mathbb{D}$, we may assume that our vertices satisfy

$$z_c = 0, \quad z_a = s \in\,]0,1[, \quad z_b = w = r\,e^{i\gamma}, \text{ where } 0<r<1,\ 0<\gamma<\pi.$$

Then the edge lengths $a = \delta(z_b, z_c)$, b, and c satisfy

$$\tanh a = r, \quad \tanh b = s, \quad \tanh c = \left|\frac{w-s}{1-sw}\right| =: t.$$

We now use (3):

$$\cosh(2c) = \frac{1+t^2}{1-t^2} = \frac{|1-sw|^2 + |w-s|^2}{|1-sw|^2 - |w-s|^2}$$
$$= \frac{1+r^2+s^2+r^2s^2-2s(w+\overline{w})}{1-r^2-s^2+r^2s^2},$$

and since $w+\overline{w} = 2r\cos\gamma$,

$$\cosh(2c) = \frac{1+r^2}{1-r^2}\cdot\frac{1+s^2}{1-s^2} - \frac{2r}{1-r^2}\cdot\frac{2s}{1-s^2}\cos\gamma. \tag{5}$$

By (3) and (4), (5) is exactly our claim. □

Proposition 3.3 (The law of sines)**.**

$$\frac{\sin\alpha}{\sinh(2a)} = \frac{\sin\beta}{\sinh(2b)} = \frac{\sin\gamma}{\sinh(2c)}.$$

Proof: We write $A = \cosh 2a$, $B = \cosh 2b$, and $C = \cosh 2c$ for short, so that $\sinh 2a = (A^2-1)^{1/2}$, etc. Prop. 3.2 then gives

$$\cos\gamma = \frac{AB-C}{\sqrt{A^2-1}\sqrt{B^2-1}}, \tag{6}$$

and likewise for the other angles. From this, we have

$$\sin^2\gamma = 1-\cos^2\gamma = \frac{D}{(A^2-1)(B^2-1)}, \tag{7}$$

where $D = 1-A^2-B^2-C^2+2ABC$; thus,

$$\frac{\sin^2\gamma}{\sinh^2 2c} = \frac{D}{(A^2-1)(B^2-1)(C^2-1)}.$$

The latter term is symmetric in A, B, and C and must therefore coincide with $\sin^2\alpha/\sinh^2(2a)$ and $\sin^2\beta/\sinh^2(2b)$. The claim then follows by taking the square root, since all values of sin and sinh in question are positive. □

Proposition 3.4 (The second law of cosines)**.**

$$\cosh(2c)\sin\alpha\sin\beta = \cos\alpha\cos\beta + \cos\gamma.$$

Proof: By (7), the left hand side of this equation is

$$C\frac{D^{1/2}}{\sqrt{(B^2-1)(C^2-1)}}\cdot\frac{D^{1/2}}{\sqrt{(A^2-1)(C^2-1)}} = \frac{CD}{\sqrt{(A^2-1)(B^2-1)}(C^2-1)},$$

and by (6), the right hand side is

$$\frac{BC-A}{\sqrt{(B^2-1)(C^2-1)}}\cdot\frac{AC-B}{\sqrt{(A^2-1)(C^2-1)}}+\frac{AB-C}{\sqrt{(A^2-1)(B^2-1)}}$$
$$=\frac{(BC-A)(AC-B)+(AB-C)(C^2-1)}{\sqrt{(A^2-1)(B^2-1)}(C^2-1)}.$$

The numerator of the last fraction is again CD, which proves the proposition. □

Prop. 3.4 does not have an analogue in euclidean geometry, but it has a remarkable consequence:

Corollary 3.5. *The three angles of an h-triangle uniquely determine the lengths of its edges.*

Any two h-triangles with the same angles and the same orientations can thus be mapped to one another via an isometry $T \in \operatorname{Aut}\mathbb{D}$; contrary to euclidean geometry, in h-geometry there are no "similar triangles". Moreover for given angles α, β, and γ such that $\alpha+\beta+\gamma<\pi$, there is always a triangle whose interior angles are precisely these. This can be seen in the disk model by placing one vertex at the origin.

The holomorphic automorphisms of $\mathbb{D}$ keep h-distances invariant and are thus isometries, in the following sense.

Definition 3.2. *An isometry of the h-plane $\mathbb{D}$ is a bijection $\tau\colon \mathbb{D}\to\mathbb{D}$ such that $\delta(\tau z,\tau w)=\delta(z,w)$ for all $z,w\in\mathbb{D}$.*

The definition of δ immediately implies that $\kappa\colon z\mapsto\overline{z}$, i.e. reflection in the h-line $g_0=\mathbb{R}\cap\mathbb{D}$ is also an isometry; it is clearly not holomorphic. If g is an arbitrary h-line, then $g=Tg_0$ for some $T\in\operatorname{Aut}\mathbb{D}$, and $\sigma_g=T\kappa T^{-1}$ is the *h-reflection* in g, characterized by the properties $\sigma_g^2=\mathrm{id}$ and $\{z\in\mathbb{D}\colon \sigma_g(z)=z\}=g$.

The isometries of $\mathbb{D}$ form a group $\operatorname{Iso}\mathbb{D}$ under composition that contains $\operatorname{Aut}\mathbb{D}$ as a subgroup.

Proposition 3.6. *The group* $\operatorname{Iso}\mathbb{D}$ *is generated by* $\operatorname{Aut}\mathbb{D}$ *and h-reflections.*

Remarks:

a) Since $\sigma_g=T\kappa T^{-1}$, $\operatorname{Iso}\mathbb{D}$ is generated by $\operatorname{Aut}\mathbb{D}$ and κ. Clearly, $\kappa T\kappa^{-1}\in\operatorname{Aut}\mathbb{D}$ for $T\in\operatorname{Aut}\mathbb{D}$; thus, $\operatorname{Aut}\mathbb{D}$ is a normal subgroup of index 2 in $\operatorname{Iso}\mathbb{D}$.

b) Every isometry τ that does not belong to $\operatorname{Aut}\mathbb{D}$ can therefore be written as $\tau=T\kappa$, where $T\in\operatorname{Aut}\mathbb{D}$, i.e. it can be written in the form

$$z\mapsto\frac{a\overline{z}+b}{\overline{b}\overline{z}+\overline{a}},\quad\text{where } a\overline{a}-b\overline{b}=1.$$

Proof of Prop. 3.6**:** Let $\tau \in \operatorname{Iso}\mathbb{D}$ and $\tau \neq \mathrm{id}$.

a) Suppose τ has (at least) two fixed points $z_0, z_1 \in \mathbb{D}$. Then τ is the reflection in the h-line that passes through z_0 and z_1: Since τ is an isometry, g is fixed pointwise (say $z \in [z_0, z_1]_h$ – then $\delta(z_0, z) + \delta(z, z_1) = \delta(z_0, z_1)$, and hence $\delta(z_0, \tau z) + \delta(\tau z, z_1) = \delta(z_0, z_1)$, i.e. $\tau z \in [z_0, z_1]_h$; since $\delta(z_0, \tau z) = \delta(z_0, z)$, we have $\tau z = z$). Every $z \in \mathbb{D}\backslash g$ is a point of intersection of the circles $K_{r_0}(z_0)$ and $K_{r_1}(z_1)$, where $r_0 = \delta(z, z_0)$ and $r_1 = \delta(z, z_1)$. Therefore, τz must coincide with z or the other point of intersection z^* of these circles. Since τ is continuous (by virtue of being an isometry), either $\tau z = z^*$ for all z or $\tau z = z$ for all z. But we excluded the latter possibility.

b) Suppose τ has exactly one fixed point z_0. Taking T_{z_0} as in VII.2, formula (1), we pass to $T_{z_0}\tau T_{z_0}^{-1}$, i.e. we may assume that $z_0 = 0$. We then choose a point $z_1 \neq 0$. Since $\delta(0, \tau z_1) = \delta(0, z_1)$, there exists a rotation $D_\varphi z = e^{i\varphi} z$ such that $D_\varphi(\tau z_1) = z_1$. Then $D_\varphi \tau$ has the fixed points 0 and z_1 and, by part a), is therefore either a reflection or the identity.

c) Suppose τ has no fixed points. Setting $z_0 = \tau(0)$, we see that $T_{z_0}\tau$ has at least the fixed point 0, and we may apply part a) or b). □

The following statement is interesting with regard to the axiomatic development of geometry:

Proposition 3.7. *The group* $\operatorname{Iso}\mathbb{D}$ *is generated by h-reflections.*

Proof: By Prop. 3.6, it suffices to show that every $T \in \operatorname{Aut}\mathbb{D}$ is a product of reflections. We may write $T = D_\varphi T_{z_0}$ and need only concern ourselves with the individual factors.

a) We easily verify that $D_\varphi = (D_{\varphi/2}\kappa D_{-\varphi/2})\kappa$: D_φ is the product of two reflections.

b) Consider T_{z_0}, where $z_0 \neq 0$. Let g_1 and g_2 be the perpendiculars to the h-line connecting 0 and z_0 at the point 0 and at the h-midpoint of $[0, z_0]_h$, respectively. Then $\sigma_{g_1}\sigma_{g_2}$ belongs to $\operatorname{Aut}\mathbb{D}$, leaves the h-line through 0 and z_0 invariant, and maps z_0 to 0. But these properties characterize T_{z_0}, so that $T_{z_0} = \sigma_{g_1}\sigma_{g_2}$. □

In our introduction of h-geometry, we imposed two normalizing conditions: We developed our geometry in the disk of radius $R = 1$, and we required of our metric $\lambda(z)$ that $\lambda(0) = 1$ (see VII.2, preceding (5)). The latter condition guarantees that $\lim_{z\to 0} \delta(z, 0)/|z - 0| = 1$, i.e. in a very small neighbourhood of 0, the hyperbolic and euclidean metrics almost coincide. We will keep this normalization but now regard $D_R = D_R(0)$, where R is arbitrary, as the hyperbolic plane.

Our previous computations easily carry over to this new situation; let us give several results:

$$\operatorname{Aut} D_R = \left\{ T\colon z \mapsto e^{it} R^2 \frac{z - z_0}{R^2 - \overline{z_0} z}, \text{ where } |z_0| < R, t \in \mathbb{R} \right\}.$$

The $\operatorname{Aut} D_R$-invariant metric λ_R on D_R such that $\lambda_R(0) = 1$ is

$$\lambda_R(z) = \frac{R^2}{R^2 - |z|^2},$$

and the h-distance $\delta^R(w, z)$ between $w, z \in D_R$ satisfies

$$\tanh\left(\frac{1}{R}\delta^R(w, z)\right) = R\frac{|w - z|}{|R^2 - \overline{z}w|}.$$

The h-length of an h-circle K_r of h-radius r in D_R is thus

$$L_h^R(K_r) = \pi R \sinh\frac{2r}{R}, \tag{8}$$

and the h-area of the disk B_r bounded by K_r is

$$F_h^R(B_r) = \pi R^2 \sinh^2\frac{r}{R}. \tag{9}$$

In the trigonometric formulas, the edge lengths a, b, and c must be replaced by a/R, b/R, and c/R, e.g. the first law of cosines in D_R is

$$\cosh(2c/R) = \cosh(2a/R)\cosh(2b/R) - \sinh(2a/R)\sinh(2b/R)\cos\gamma. \tag{10}$$

If one lets $R \to \infty$ in these formulas, then by expanding the hyperbolic functions into power series, we see that for r fixed,

$$\lim_{R\to\infty} L_h^R(K_r) = 2\pi r, \quad \lim_{R\to\infty} F_h^R(B_r) = \pi r^2,$$

and as $R \to \infty$, (10) turns into the euclidean law of cosines:

$$c^2 = a^2 + b^2 - 2ab\cos\gamma.$$

Intuitively speaking, for a sufficiently large R, h-geometry is arbitrarily close to euclidean geometry on any fixed compact subset of D_R; plane euclidean geometry is the limiting case of h-geometry of D_R as $R \to \infty$.

A corresponding relation holds for the geometry on a sphere of radius R. Let us briefly investigate this: Let $S_R = \{\mathbf{x} \in \mathbb{R}^3 \colon |\mathbf{x}| = R\}$. The shortest path between two points $A, B \in S_R$ is an arc of the great circle that passes through A and B; its length is the spherical distance $\delta_s^R(A, B) = R \cdot \sphericalangle AOB$, provided $\sphericalangle AOB \leqslant \pi$ (here O is the origin in $\mathbb{R}^3$). The length of the circle $K_r(M) = \{A \in S_R \colon \delta_s^R(A, M) = r\}$ is seen to be

$$L_s^R(K_r) = 2\pi R \sin\frac{r}{R}, \tag{11}$$

and the area of the (smaller) "cap" $B_r = \{A \in S_R \colon \delta_s^R(A, M) \leqslant r\}$ bounded by K_r is

$$F_s^R(B_r) = 2\pi R^2\left(1 - \cos\frac{r}{R}\right) = 4\pi R^2 \sin^2\frac{r}{2R}. \tag{12}$$

Given a triangle ABC on S_R with interior angles α, β, and γ and edge lengths $a = \delta_s^R(B,C)$, b, and c, it is not difficult to derive the first law of cosines of spherical trigonometry:

$$\cos(c/R) = \cos(a/R)\cos(b/R) + \sin(a/R)\sin(b/R)\cos\gamma. \tag{13}$$

If one lets $R \to \infty$ in (11)-(13), then one again obtains the corresponding formulas from euclidean geometry, i.e. the latter is the limiting case of spherical geometry on S_R.

A remarkable connection arises upon substituting iR for R in the "spherical" formulas (11)-(13) (use $\cos ix = \cosh x$ and $\sin ix = i \sinh x$): (11)-(13) then turn into the formulas (8)-(10) for hyperbolic geometry on D_{2R}! It is for this reason that, for instance, Lambert (1766) suggested regarding h-geometry, for which he attained initial – still tentative – results (see below), as geometry on a "sphere with imaginary radius".

The concept of (Gaussian) curvature κ on a surface equipped with a metric sheds further light on the situation. For the euclidean plane, $\kappa \equiv 0$ (and every surface for which $\kappa \equiv 0$ is locally isometric to a subset of the euclidean plane); the sphere S_R has constant curvature $\kappa \equiv 1/R^2$. If in a domain $G \subset \mathbb{R}^2 = \mathbb{C}$ the metric is defined via (3) from VII.2, then

$$\kappa(z) = -\frac{1}{\lambda(z)^2}\Delta \log \lambda(z).$$

For the h-metric λ_R on D_R, one computes $\kappa \equiv -4/R^2$: the h-plane has constant negative curvature! And as $R \to \infty$, the curvatures of D_R and S_R tend to 0. The euclidean plane is thus the common limiting case of surfaces of constant positive or negative curvature.

Historical remark

Euclid's *Elements* begins with definitions, postulates, and axioms. They are all straightforward (the third postulate, for example, requires that one always be able to construct a circle given a centre and a radius), with the exception of the fifth postulate:

> *The following is required: If a straight line h intersects two straight lines g_1 and g_2 lying in the same plane in such a way that the interior angles on one side of h together are less than two right angles, then g_1 and g_2 intersect on this side of h.*

We will denote this (slightly reformulated) postulate as P.5. Euclid then develops geometry in a strictly deductive manner; in Proposition 17, he proves the converse of the implication in P.5, and it is in the proof of Proposition 29 (concerning the equality of alternate angles with respect to a transversal of two parallel straight lines) that P.5 is first used. The converse of Proposition 29 is proved earlier.

This complicated situation led later Hellenistic mathematicians to attempt to prove P.5 as a theorem using Euclid's first 28 propositions. These efforts were continued by Islamic mathematicians (ca. 900-1300), and when European mathematicians began to seriously study Euclid in the late sixteenth century, the topic was revived.

These efforts led to many "proofs", which, however, always used other unproved statements. Speaking in modern terms, one realized that P.5 was, for example, equivalent to:

– For every straight line g and every point P not lying on g, there exists exactly one parallel to g that passes through P.

– The sum of the angles of a triangle is two right angles.

– There exist similar, non-congruent triangles.

– The points that are equidistant from a fixed straight line lie on two straight lines.

– Given three points that do not lie on the same straight line, it is always possible to draw a circle that passes through them.

A more profound analysis was given by Saccheri and Lambert in the eighteenth century. Lambert considers a quadrangle with three right angles. The assumption that the fourth angle is obtuse soon leads to a contradiction; the assumption that the fourth angle is a right angle yields P.5. Assuming that the fourth angle is acute, Lambert (in the hope of finding a contradiction) constructs a substantial building of propositions, showing, for example, that the area of a triangle is proportional to $\pi - (\alpha + \beta + \gamma)$. No contradiction turned up.

The progress in the beginning of the nineteenth century was intimately connected with a change in viewpoint. The protagonists Gauss (who did not publish anything on the matter), N. I. Lobachevsky (beginning in 1826), and J. Bolyai (1832) did not search for a contradiction, but held a new geometry as plausible. They constructed, assuming the negation of P.5, a "non-euclidean" (two- and three-dimensional) geometry, complete with trigonometry, area and volume measure – a comprehensive, "self-contained" theory. A universal positive constant k arises in the quantitative non-euclidean geometry of Gauss; he gives the circumference of a circle to be $2\pi k \sinh(r/k)$, for instance. In our notation, k is thus none other than the radius of the disk D_R on which one considers h-geometry.

Geometry was then to a large extent still understood as a description of physical space. The question of whether this space is "euclidean" or "non-euclidean" could not be answered empirically, since, roughly speaking, measuring differences in sufficiently small regions between the two geometries is beyond the precision of physical instruments. If, for example, one measures the radius r and circumference L of a circle, then one cannot say with certainty whether $L = 2\pi r$ or $L = \pi R \sinh(2r/R)$ for an r that is very small compared to R.

In the second half of the nineteenth century, the mathematical concept of space was broadened substantially, particularly by Riemann; the problem of mathematically modelling physical space reached a new level. Since various "non-euclidean" geometries have by now been discovered, we have referred to the theory of Lobachevsky and Bolyai as "hyperbolic geometry".

Even a well-developed theory based on plausible assumptions may in principle contain a contradiction. Since 1871, however, when Klein, following Beltrami (1868) [Bel], gave a model of hyperbolic geometry in the framework of projective geometry, it is clear that if hyperbolic geometry contains a contradiction, then so must euclidean geometry. In 1882, Poincaré found a simple model of the hyperbolic plane (which had also been discovered by Beltrami) while studying complex differential equations – this is the model that we have treated here. It may readily be generalized to a model of three-dimensional hyperbolic space.

Exercises

1. Investigate when two h-lines have a common perpendicular. Hint: Use the half-plane model.

2. a) Let z_1 and z_2 be distinct points. Show that their perpendicular bisector (i.e. the perpendicular to $[z_1, z_2]_h$ through its midpoint) is defined by
$$\{z \colon \delta(z, z_1) = \delta(z, z_2)\}.$$
   b) Let Δ be an h-triangle. Show that the three perpendicular bisectors of the edges of Δ either intersect at one point (which is then the centre of the h-circle that passes through the vertices of Δ) or do not intersect at all. Show that the latter is the case if and only if the euclidean circle that passes through the vertices of Δ is not contained in $\mathbb{D}$ (or $\mathbb{H}$).

3. Let g be an h-line and $d > 0$. Investigate the set of points whose distance from g is d.
4. Show that all improper triangles whose interior angles are all 0 are congruent, meaning that they can be mapped to one another under isometries.
5. Consider an improper triangle with interior angles $\pi/2$, α, and 0 and whose finite edge has h-length ℓ. Prove the (equivalent) formulas

$$e^{-2\ell} = \tan\frac{\alpha}{2}, \quad \cosh 2\ell = \frac{1}{\sin\alpha}.$$

Warning: The second law of cosines is not proved for improper triangles! Hint: Choose vertices i, z_0, and ∞ in $\mathbb{H}$.
6. Prove that two h-triangles with the same interior angles can be mapped onto one another via an isometry.
7. Show that an h-triangle is equilateral (meaning that $a = b = c$) if and only if all of its interior angles are the same. Show that for every edge length a, there exists an equilateral triangle, and that $2\cosh a \cdot \sin(\alpha/2) = 1$ holds.
8. Show that every isometry of the h-plane is the product of at most three reflections.

4. The Riemann mapping theorem

The Schwarz-Pick lemma establishes a close connection between complex analysis and (plane) hyperbolic geometry: the holomorphic automorphisms of the unit disk or of the upper half-plane are precisely the orientation-preserving hyperbolic isometries. This connection is given significantly more meaning by the following result:

Theorem 4.1 (The Riemann mapping theorem)**.** *Every simply connected subdomain of the sphere $\widehat{\mathbb{C}}$ that has at least two boundary points is conformally equivalent to the unit disk.*

Before proving the theorem, let us make a few comments.

We call a subdomain of $\widehat{\mathbb{C}}$ *simply connected* if it is either $\widehat{\mathbb{C}}$ itself or biholomorphically equivalent to a simply connected subdomain of $\mathbb{C}$. By the Riemann mapping theorem, there are thus exactly three simply connected domains up to biholomorphic equivalence:

– The sphere $\widehat{\mathbb{C}}$
– The plane $\mathbb{C}$
– The unit disk $\mathbb{D}$.

These three domains are clearly not conformally equivalent to one another: $\widehat{\mathbb{C}}$ is not even homeomorphic to $\mathbb{C}$ or $\mathbb{D}$. The unit disk $\mathbb{D}$ is homeomorphic to $\mathbb{C}$, as is shown by the mapping

$$z \mapsto \frac{z}{1-|z|},$$

but certainly not biholomorphically equivalent to $\mathbb{C}$: a holomorphic map from $\mathbb{C}$ to $\mathbb{D}$ is a bounded entire function and hence constant.

The unit disk, with its automorphisms and geometry, gains still further significance via the following fact, known as the *uniformization theorem*:

Theorem 4.2. *Let G be an arbitrary subdomain of $\widehat{\mathbb{C}}$ that has at least three boundary points. Then there exists an unbranched surjective holomorphic map f from $\mathbb{D}$ to G.*

"Unbranched" means that $f' \neq 0$ everywhere, i.e. f is locally biholomorphic. This theorem is most readily proved in the context of the theory of Riemann surfaces – see, for example, [FL2] – but there are also "direct" proofs that only rely on complex analysis in the plane – cf. [RS]. In Sect. 7 we shall prove Thm. 4.2 for the special case $G = \widehat{\mathbb{C}} \setminus \{0, 1, \infty\}$. The consequences of this particular theorem, expanded in Sect. 8, indicate the power of the uniformization theorem.

Proof of Thm. 4.1**:** *Step 1.* We map G conformally onto a subdomain of the unit disk.

By assumption, G has at least two boundary points. Using a Möbius transformation, we send one of them to ∞, so that the image of G, again denoted by G, is a simply connected subdomain of $\mathbb{C}$ with at least one boundary point $a \in \mathbb{C}$. Using a translation, we send a to the origin. Now, if G is a simply connected domain that does not contain the origin, then a holomorphic branch h of the square root exists on G:

$$\begin{aligned} h\colon G &\to G' = h(G), \\ z &\mapsto w = h(z), \quad h(z)^2 \equiv z. \end{aligned}$$

The inverse of h on G' is the function $w \mapsto w^2$. Since this is necessarily injective on G', we have: if $w \in G'$, then $-w \notin G'$. Let $w_0 \in G'$, and let r be small enough that the disk $D_r(w_0)$ belongs to G'. The "mirrored disk"

$$-D_r(w_0) = \{w\colon -w \in D_r(w_0)\}$$

does not intersect G', which means that G' lies in the exterior of a certain disk. Using a translation and a homothety, we map G' onto a subdomain in the exterior of the unit disk. The reflection $z \mapsto 1/z$ exchanges the interior and exterior of the unit disk, so that we can now biholomorphically map G onto a subset of $\mathbb{D}$. By possibly applying an automorphism of $\mathbb{D}$, we may assume that 0 belongs to our domain.

Step 2. It remains to show that every simply connected subdomain $G \subset \mathbb{D}$ with $0 \in G$ can be biholomorphically mapped onto $\mathbb{D}$.

Let $\mathcal{F}$ denote the following family of holomorphic mappings $f\colon G \to \mathbb{D}$:

$$\mathcal{F} = \{f\colon G \to \mathbb{D}, f \text{ is injective}, f(0) = 0, f'(0) > 0\}.$$

The identity certainly belongs to $\mathcal{F}$, so $\mathcal{F}$ is nonempty. Now let

$$\alpha = \sup\{f'(0)\colon f \in \mathcal{F}\}.$$

We have $1 \leqslant \alpha \leqslant \infty$, and by definition there exists a sequence of functions $f_\nu \in \mathcal{F}$ such that $f'_\nu(0) \to \alpha$. Since the sequence f_ν is uniformly bounded, it contains a locally uniformly convergent subsequence by Montel's theorem (Thm. II.5.3), which we also denote by f_ν. By the theorem of Weierstrass (Thm. II.5.2), the limit function f satisfies

$$\lim_{\nu\to\infty} f_\nu(0) = f(0) = 0$$
$$\lim_{\nu\to\infty} f'_\nu(0) = f'(0) = \alpha,$$

which implies that $\alpha < \infty$. Since $f'(0) \neq 0$, the function f is nonconstant and therefore injective by Prop. IV.6.5.

Step 3. Let us now show that the above mapping $f\colon G \to \mathbb{D}$ is also surjective and therefore solves the problem.

We have $|f(z)| \leqslant 1$ for all z; therefore, since f is nonconstant, $|f(z)| < 1$ by the maximum modulus principle. Now $G_0 = f(G)$ is a simply connected subdomain of $\mathbb{D}$ that contains 0. In case this domain is not all of $\mathbb{D}$, then by the following converse (Prop. 4.3) of the Schwarz lemma (Thm. 1.3), choose an injective holomorphic mapping $h\colon G_0 \to \mathbb{D}$ such that $h(0) = 0$ and $h'(0) > 1$. Then we would have $h \circ f \in \mathcal{F}$ with

$$(h \circ f)'(0) > \alpha,$$

contradicting the definition of α.

Step 4. It remains to find the mapping h, i.e. we will show the following:

Proposition 4.3. *Let $G_0 \subset \mathbb{D}$ be a proper subdomain of $\mathbb{D}$ that is simply connected and contains 0. Then there exists an injective holomorphic mapping $h\colon G_0 \to \mathbb{D}$ such that $h(0) = 0$ and $h'(0) > 1$.*

Proof: Let $a \in \mathbb{D} \setminus G_0$. Via the automorphism

$$Tz = \frac{z-a}{1-\overline{a}z},$$

G_0 is biholomorphically mapped to a simply connected subdomain $G_1 \subset \mathbb{D}$ with $Ta = 0$, i.e. $0 \notin G_1$. Via $f(z) = \sqrt{z}$, G_1 is biholomorphically mapped onto a subdomain $G_2 \subset \mathbb{D}$ (here $\sqrt{z}$ is a fixed branch of the square root function) with $f(T0) = f(-a) = \sqrt{-a} = b \in G_2$. Finally, the transformation

$$Sz = \frac{z-b}{1-\overline{b}z}$$

(we again have $S \in \operatorname{Aut}\mathbb{D}$) biholomorphically maps the domain G_2 onto $G_3 \subset \mathbb{D}$ with $Sb = 0 \in G_3$. We now put

$$h_0 = S \circ f \circ T$$

and compute

$$h_0(0) = 0$$
$$|h_0'(0)| = \left|\frac{1+|b|^2}{2b}\right| > 1.$$

A further rotation through an angle ϑ, i.e. $S_\vartheta z = e^{i\vartheta} z$, then yields

$$h = S_\vartheta h_0$$

with $h'(0) > 1$. □

This also completes the proof of the Riemann mapping theorem. □

Remarks:

a) Two conformal mappings $f, g\colon G \to \mathbb{D}$ differ only by an automorphism of $\mathbb{D}$:

$$f \circ g^{-1} \in \operatorname{Aut} \mathbb{D}.$$

We may thus uniquely determine f by requiring, for an arbitrary point $a \in G$, that $f(a) = 0$ and $f'(a) > 0$.

b) Although almost every simply connected domain can be conformally mapped onto $\mathbb{D}$, our proof says nothing about the "effective construction" of such a mapping. The only elementary case is that of "crescents" – see the exercises. The effective construction of conformal mappings for particular domains – such as polygons – is connected with the theory of special functions; cf. [FL2].

c) There is no Riemann mapping theorem in the theory of several complex variables! To conclude, we prove the following theorem:

Theorem 4.4. *For $n > 1$, the polydisk $\mathbb{D}^n = \{\mathbf{z}\colon |z_\nu| < 1\}$ and the unit ball $\mathbb{B}^n = \{\mathbf{z}\colon \sum_{\nu=1}^n |z_\nu|^2 < 1\}$ are not biholomorphically equivalent.*

Proof: Let

$$\mathbb{D}^n = \mathbb{D}^{n-1} \times \mathbb{D} = \{(\mathbf{z}, w)\colon |\mathbf{z}| < 1, |w| < 1\}$$

be the n-dimensional unit polydisk, which we write as a product of the $(n-1)$-dimensional unit polydisk (with coordinates $\mathbf{z} = (z_1, \dots, z_{n-1})$) and the unit disk in the w-plane; we have $n \geqslant 2$. The unit ball $\mathbb{B}^n \subset \mathbb{C}^n$ is given by

$$\mathbb{B}^n = \{\mathbf{u} \in \mathbb{C}^n\colon \|\mathbf{u}\|^2 = \sum_{\nu=1}^n |u_\nu|^2 < 1\}.$$

Suppose that

$$F = (f_1, \dots, f_n)\colon \mathbb{D}^n \to \mathbb{B}^n$$

is a biholomorphic mapping. The functions

$$u_\nu = f_\nu(\mathbf{z}, w), \quad 1 \leqslant \nu \leqslant n$$

are thus holomorphic. For every $\mathbf{z} \in \mathbb{D}^{n-1}$, we denote by

$$F_{\mathbf{z}} : \mathbb{D} \to \mathbb{B}^n$$

the mapping

$$w \mapsto F(\mathbf{z}, w),$$

i.e. $F_{\mathbf{z}} = (f_{1\mathbf{z}}, \ldots, f_{n\mathbf{z}})$; the $f_{\nu\mathbf{z}}$ are thus holomorphic functions in the unit disk of the w-plane that are bounded independently of $\mathbf{z}$.

Let $\mathbf{z}_j \in \mathbb{D}^{n-1}$ be a sequence that converges to a boundary point $\mathbf{z}_0 \in \partial\mathbb{D}^{n-1}$. The sequence $f_{\nu\mathbf{z}_j}$ is then bounded for every $1 \leqslant \nu \leqslant n$ and therefore has a locally uniformly convergent subsequence by Montel's theorem. By passing to a subsequence of $\mathbf{z}_j$, we may assume that

$$\begin{aligned} \mathbf{z}_j &\to \mathbf{z}_0 \in \partial\mathbb{D}^{n-1} \\ f_{\nu\mathbf{z}_j} &\to \varphi_\nu, \quad 1 \leqslant \nu \leqslant n. \end{aligned}$$

The φ_ν are holomorphic on the unit disk $\mathbb{D}$, and the fact that $F(\mathbf{z}_j, w) \in \mathbb{B}^n$, i.e.

$$\sum_{\nu=1}^{n} |f_\nu(\mathbf{z}_j, w)|^2 < 1,$$

implies that in the limit

$$\sum_{\nu=1}^{n} |\varphi_\nu(w)|^2 \leqslant 1.$$

That is, for $w \in \mathbb{D}$, the point

$$\Phi(w) = \big(\varphi_1(w), \ldots, \varphi_n(w)\big)$$

belongs to $\overline{\mathbb{B}^n}$. Suppose $\Phi(w_0) \in \mathbb{B}^n$ for a point $w_0 \in \mathbb{D}$. Then $\Phi(w_0) = F(\mathbf{z}_*, w_*)$, where $\mathbf{z}_* \in \mathbb{D}^{n-1}$, $w_* \in \mathbb{D}$, and

$$(\mathbf{z}_j, w_0) = F^{-1} \circ F(\mathbf{z}_j, w_0) \to F^{-1}\big(\Phi(w_0)\big) = (\mathbf{z}_*, w_*).$$

Hence $\mathbf{z}_j \to \mathbf{z}_* \in \mathbb{D}^{n-1}$, contradicting $\mathbf{z}_j \to \mathbf{z}_0 \in \partial\mathbb{D}^{n-1}$. We thus have $\Phi(\mathbb{D}) \subset \partial\mathbb{B}^n$, i.e.

$$\sum_{\nu=1}^{n} |\varphi_\nu(w)|^2 \equiv 1$$

on $\mathbb{D}$. Denote the derivative of φ_ν by φ'_ν; differentiating with respect to w and $\overline{w}$ yields

$$\sum_{\nu=1}^{n} |\varphi'_\nu(w)|^2 \equiv 0,$$

i.e. $\varphi'_\nu \equiv 0$ on $\mathbb{D}$.

By the construction of φ_ν, this means that

$$\lim_{j\to\infty} \frac{\partial}{\partial w} f_\nu(\mathbf{z}_j, w) = 0.$$

The holomorphic function

$$\mathbf{z} \mapsto \frac{\partial}{\partial w} f_\nu(\mathbf{z}, w), \quad w \in \mathbb{D} \text{ fixed},$$

on $\mathbb{D}^{n-1}$ thus admits a continuous extension by 0 to the boundary of $\mathbb{D}^{n-1}$; by the maximum modulus principle, it is thus identically zero. The function F therefore does not depend on w and, in particular, cannot be injective. □

Exercises

1. Let $W_\alpha = \{z = r\,e^{it} \colon 0 < t < \alpha\}$ be the open angular domain with angle $\alpha \leqslant 2\pi$. Find an elementary function that conformally maps it onto $\mathbb{D}$.
2. Using elementary functions, conformally map a crescent onto $\mathbb{D}$. Do the same for a strip that is parallel to the real axis.
3. Use elliptic functions to conformally map a rectangle onto the unit disk.
4. Let $\mathbb{H}_{n+1} = \{(\mathbf{z}, w) \in \mathbb{C}^{n+1} \colon \operatorname{Im} w > \|\mathbf{z}\|^2\}$, where $\|\mathbf{z}\|$ is the euclidean norm. The space $\mathbb{H}_{n+1}$ is called the Siegel upper half-space. We define a mapping $u \mapsto (\mathbf{z}, w)$ of the ball $\mathbb{B}^{n+1}$ into $\mathbb{H}_{n+1}$ by
$$z_j = \frac{u_j}{1 + u_{n+1}}, \quad \text{for } 1 \leqslant j \leqslant n; \quad w = i\frac{1 - u_{n+1}}{1 + u_{n+1}}.$$
Show that $\mathbb{B}^{n+1}$ is thereby biholomorphically mapped onto $\mathbb{H}_{n+1}$ (for $n = 0$, this is a mapping from $\mathbb{D}$ onto $\mathbb{H} = \mathbb{H}_1$). Investigate the behaviour of this mapping on the boundaries!

5. Harmonic functions

We recall some facts from I.2: The Laplace operator

$$\Delta = 4\,\frac{\partial^2}{\partial z \partial \overline{z}} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

acts on twice differentiable functions in plane regions. A function u is harmonic if $\Delta u = 0$. The Cauchy-Riemann equations show that holomorphic functions as well as their real and imaginary parts are harmonic. Another example is given by $\log|f|$, where f is a holomorphic function without zeros. – Henceforth, we consider only real-valued harmonic functions.

Proposition 5.1. *Let u be a real-valued harmonic function on the convex domain $G \subset \mathbb{C}$. Then there is a holomorphic function F on G such that $u = \operatorname{Re} F$.*

Proof: Since $u_{z\overline{z}} = 0$, the function u_z is holomorphic on G. As G is convex, it has a primitive f on G. Then

$$\big(\operatorname{Re} f\big)_z = \frac{1}{2}(f + \overline{f})_z = \frac{1}{2}f_z = \frac{1}{2}u_z, \quad \big(\operatorname{Re} f\big)_{\overline{z}} = \frac{1}{2}\overline{f_z} = \frac{1}{2}u_{\overline{z}}.$$

Therefore $u = \operatorname{Re}(2f) + c$ with a real constant c. □

We note that the function F is determined uniquely up to addition of a purely imaginary constant.

Corollary 5.2. *A harmonic function is infinitely often (real) differentiable.* □

Corollary 5.3. *Let $f\colon G \to \mathbb{C}$ be a non-constant holomorphic function. If u is harmonic on $f(G)$, then $u \circ f$ is harmonic on G.*

Proof: Locally, we can write $u = \operatorname{Re} g$ with a holomorphic function g.
Then $u \circ f = \operatorname{Re}(g \circ f)$. □

From Prop. 5.1 we derive an identity theorem for harmonic functions:

Proposition 5.4. *If a harmonic function $u\colon G \to \mathbb{R}$, G a domain, vanishes on a non-void open subset $U \subset G$, it vanishes on G.*

The example $u(z) = \operatorname{Re} z$ shows that harmonic functions may vanish on a non-discrete set without vanishing identically.

Proof: Let $M = \{z \in G : u \equiv 0$ in a neighbourhood of $z\}$. Then M is open and by assumption non-void. We show that M is also relatively closed in G, hence $M = G$: Let $z_1 \in \overline{M} \cap G$, choose a convex neighbourhood $V \subset G$ of z_1 and $f \in \mathcal{O}(V)$ with $u = \operatorname{Re} f$ on V. Then $\operatorname{Re} f$ vanishes on the non-void open set $V \cap M$. By the identity theorem for holomorphic functions, f is an imaginary constant on V, i.e. $u \equiv 0$ on V and $z_1 \in M$. □

Let u be a harmonic function. A function v is said to be a *conjugate harmonic* of u, if $u + iv$ is holomorphic. If it exists, it is uniquely determined up to addition of a real constant. The example $u(z) = \log|z|$ on $\mathbb{C}^*$ shows that conjugate harmonic functions need not exist globally; on the other hand, Prop. 5.1 shows they always exist locally, e.g. on convex domains. Note that if v is a conjugate harmonic of u, then v is harmonic, and $-u$ is a conjugate harmonic function of v, because $-i(u+iv) = v - iu$ is holomorphic.

Definition 5.1. *A continuous function $f\colon G \to \mathbb{C}$ has the mean value property at $z_0 \in G$, if there exists $D_R(z_0) \subset\subset G$ such that*

$$f(z_0) = \frac{1}{2\pi} \int_0^{2\pi} f(z_0 + re^{it})\, dt \qquad \text{for } 0 \leqslant r \leqslant R. \tag{1}$$

If this holds for all $z_0 \in G$, the function f has the mean value property on G.

Holomorphic functions have the mean value property on their domain of definition – this is an immediate consequence of Cauchy's integral formula (see the proof of Thm. II.5.5). Prop. 5.1 implies

Proposition 5.5. *Harmonic functions have the mean value property.*

Proof: Represent the harmonic function u locally as $u = \operatorname{Re} f$ with a holomorphic function f and take real parts in (1). □

As in the case of holomorphic functions, we have a maximum principle for functions with the mean value property, in particular for harmonic functions.

Proposition 5.6. *Let u be a real-valued function on the domain G having the mean value property. If u attains a local maximum or minimum at a point z_0 of G, then u is constant in a neighbourhood of z_0. If u attains a global extremum at $z_0 \in G$, then u is constant.*

Proof: A minimum of u is a maximum of $-u$, it suffices therefore to consider the case of a local maximum of u in $z_0 \in G$. Choose $D = D_R(z_0) \subset\subset G$ such that (1) holds (with u in place of f) and that $u(z) \leqslant u(z_0)$ for $z \in \overline{D}$. Then, for every $r \leqslant R$,

$$\frac{1}{2\pi} \int_0^{2\pi} \big(u(z_0 + re^{it}) - u(z_0)\big)\, dt = u(z_0) - u(z_0) = 0.$$

Since the integrand is continuous and non-positive, it must be $\equiv 0$, that is $u\big(z_0 + re^{it}\big) = u(z_0)$ for $0 \leqslant t \leqslant 2\pi$ and $r \leqslant R$. Thus u is constant on D. The set $M = \{z \in G : u(z) = u(z_0)\}$ is closed in G and nonvoid; if $u(z_0)$ is a global maximum of u, M is open by the above, whence $M = G$. □

Corollary 5.7. *Let G be a bounded domain and $u\colon \overline{G} \to \mathbb{R}$ a continuous function satisfying the mean value property on G. Then u attains its maximum and its minimum on ∂G, and $\min\limits_{\partial G} u < u(z) < \max\limits_{\partial G} u$ for $z \in G$ unless u is constant.*

We now consider a function u harmonic in a neighbourhood of a disk $\overline{D_R(z_0)}$. Prop. 5.5 represents the value $u(z_0)$ at the centre as an integral over the boundary $\partial D_R(z_0)$. Applying an automorphism of $D_R(z_0)$, we deduce an integral representation of $u(z)$, $z \in D_R(z_0)$ arbitrary. To simplify the notation, we assume $z_0 = 0$.

Thus, let u be harmonic on a neighbourhood of $\overline{D_R} = \overline{D_R(0)}$. For fixed $z \in D_R$, the Möbius transformation $S\zeta = R^2 \dfrac{\zeta - z}{R^2 - \bar{z}\zeta}$ is in $\operatorname{Aut} D_R$, maps (a neighbourhood of) $\overline{D_R}$ onto (a neighbourhood of) $\overline{D_R}$ and z to 0. The mean value property of the harmonic function $u \circ S^{-1}$ implies

$$u(z) = \big(u \circ S^{-1}\big)(0) = \frac{1}{2\pi} \int_0^{2\pi} \big(u \circ S^{-1}\big)\big(Re^{i\vartheta}\big)\, d\vartheta = \frac{1}{2\pi i} \int_{|\eta|=R} \big(u \circ S^{-1}\big)(\eta)\, \frac{d\eta}{\eta}.$$

Substituting $\zeta = S^{-1}\eta$ we get

$$u(z) = \frac{1}{2\pi i} \int_{|\zeta|=R} u(\zeta)\, \frac{\zeta S'(\zeta)}{S(\zeta)}\, \frac{d\zeta}{\zeta}.$$

An easy computation gives $\dfrac{\zeta S'(\zeta)}{S(\zeta)} = \dfrac{R^2 - |z|^2}{|\zeta - z|^2}$. Thus

$$u(z) = \frac{1}{2\pi i} \int_{|\zeta|=R} u(\zeta)\, \frac{R^2 - |z|^2}{|\zeta - z|^2}\, \frac{d\zeta}{\zeta} = \frac{1}{2\pi} \int_0^{2\pi} u\big(Re^{i\vartheta}\big)\, \frac{R^2 - |z|^2}{|Re^{i\vartheta} - z|^2}\, d\vartheta. \tag{2}$$

Definition 5.2. *The Poisson kernel of the disk $D_R(0)$ is the function*

$$P_R\colon \partial D_R(0) \times D_R(0) \to \mathbb{R}, \quad P_R(\zeta, z) = \frac{1}{2\pi}\, \frac{R^2 - |z|^2}{|\zeta - z|^2}. \tag{3}$$

Sharpening the result (2), we have

Proposition 5.8. *Let $u\colon \overline{D_R(0)} \to \mathbb{R}$ be continuous and harmonic on $D_R(0)$. Then, with $\zeta = Re^{i\vartheta}$,*

$$u(z) = \int_0^{2\pi} u(\zeta)\, P_R(\zeta, z)\, d\vartheta \tag{4}$$

for $z \in D_R(0)$.

Proof: By (2), the claim is true if u is harmonic in a neighbourhood of $\overline{D_R(0)}$. For the general case, we apply (2) to the functions $u_r(z) = u(rz)$, $r < 1$, which are harmonic on a neighbourhood of $\overline{D_R(0)}$:

$$u_r(z) = \int_0^{2\pi} u_r(\zeta)\, P_R(\zeta,\, z)\, d\vartheta. \tag{5}$$

For $r \to 1$, the functions u_r tend to u uniformly on $\overline{D_R(0)}$. Therefore, we can interchange limit and integral in (5) to get (4). □

For later use, we list some properties of the Poisson kernel:

i. $P_R(\zeta,\, z) > 0$,

ii. $P_R(\zeta,\, z) = \dfrac{1}{2\pi} \operatorname{Re} \dfrac{\zeta + z}{\zeta - z}$, hence $P_R(\zeta,\, z)$ is harmonic in $z \in D_R(0)$,

iii. $\int_0^{2\pi} P_R(\zeta,\, z)\, d\vartheta = 1 \quad (\zeta = Re^{i\vartheta})$,

iv. $P_R(Re^{i\vartheta},\, re^{it}) = \dfrac{1}{2\pi} \dfrac{R^2 - r^2}{R^2 - 2Rr\cos(\vartheta - t) + r^2}$.

i is clear, *ii* and *iv* follow from the definition by short computations, for *iii* apply Prop. 5.8 to the function $u \equiv 1$.

Prop. 5.8 shows that the values of a harmonic function u on $D_R(0)$, remaining continuous on the boundary, are completely determined by its boundary values. On the other hand, we can write down the *Poisson integral*

$$\int_0^{2\pi} h(\zeta)\, P_R(\zeta,\, z)\, d\vartheta \tag{6}$$

for any (real-valued) continuous function h on $\partial D_R(0)$. It represents a function $u(z)$ on $D_R(0)$, which is harmonic: In applying the Laplace operator to (6), we may differentiate under the integral sign and then use *ii*. In fact, more is true:

Theorem 5.9. *Let h be a continuous real-valued function on $\partial D_R(0)$. Then the function*

$$u(z) = \begin{cases} \int_0^{2\pi} h(\zeta)\, P_R(\zeta,\, z)\, d\vartheta & \text{for } |z| < R \\ h(z) & \text{for } |z| = R \end{cases} \tag{7}$$

is continuous on $\overline{D_R(0)}$ and harmonic on $D_R(0)$.

Proof: We need only show that $u(z)$ as defined by (7) is continuous at an arbitrary point $z_0 = Re^{i\vartheta_0}$ of the boundary. In view of $u(z_0) = h(z_0)$ and *iii*,

$$u(z) - u(z_0) = \int_{\vartheta_0-\pi}^{\vartheta_0+\pi} \big(h(\zeta) - h(z_0)\big)\, P_R(\zeta,\, z)\, d\vartheta. \tag{8}$$

Since h is continuous at z_0, there is, for every $\varepsilon > 0$, a $\delta \in\,]0, \pi[$ such that $\big|h\big(Re^{i\vartheta}\big) - h\big(Re^{i\vartheta_0}\big)\big| \leqslant \varepsilon$ if $\vartheta \in J_1 = \{\vartheta : |\vartheta - \vartheta_0| \leqslant \delta\}$. Accordingly, we split the integral in (8) into the integrals over J_1 and $J_2 = [\vartheta_0 - \pi,\, \vartheta_0 + \pi] \setminus J_1$. The integral over J_1 is easily estimated using i and iii:

$$\Big|\int_{J_1} \big(h(\zeta) - h(z_0)\big)\, P_R(\zeta,\, z)\, d\vartheta\Big| \leqslant \int_{J_1} \big|h(\zeta) - h(z_0)\big|\, P_R(\zeta,\, z)\, d\vartheta$$

$$\leqslant \varepsilon \int_{J_1} P_R(\zeta,\, z)\, d\vartheta \leqslant \varepsilon \int_0^{2\pi} P_R(\zeta,\, z)\, d\vartheta = \varepsilon$$

for any $z \in D_R(0)$. To estimate the integral over J_2, we note that $h(\zeta) - h(z_0)$ is bounded for $|\zeta| = R$, say $|h(\zeta) - h(z_0)| \leqslant M$. Thus

$$\Big|\int_{J_2} \big(h(\zeta) - h(z_0)\big)\, P_R(\zeta,\, z)\, d\vartheta\Big| \leqslant M \int_{J_2} P_R(\zeta,\, z)\, d\vartheta. \tag{9}$$

Now we restrict z to a small neighbourhood V of z_0, such that $|\zeta - z| \geqslant c > 0$ for $\zeta = Re^{i\vartheta}$, $\vartheta \in J_2$, $z \in V \cap D_R(0)$. Then $P_R(\zeta,\, z) = \dfrac{1}{2\pi}\,\dfrac{R^2 - |z|^2}{|\zeta - z|^2}$ tends to 0 uniformly in $\vartheta \in J_2$ if $z \in V \cap D_R(0)$, $|z| \to R$. Therefore, the integral in (9) tends to 0 if $z \to z_0$, and the proof is complete. □

We place the result within the context of the *Dirichlet problem*: Given a bounded domain G and a continuous real-valued function h on ∂G, is there a function u, continuous on $\overline{G}$ and harmonic on G, such that $u|\partial G = h$? Thm. 5.9 solves the Dirichlet problem for circular domains and arbitrary continuous boundary values. The solution is unique, as follows from

Proposition 5.10. *Given $G \subset\subset \mathbb{C}$ and a continuous function h on ∂G, the Dirichlet problem has at most one solution.*

Proof: Let u and v be continuous on $\overline{G}$, harmonic on G, and satisfy $u|\partial G = v|\partial G = h$. Then $u - v$ is continuous on $\overline{G}$, harmonic on G and $\equiv 0$ on ∂G. By the maximum principle (Cor. 5.7), $u - v \equiv 0$. □

We are now in a position to prove the converse of Prop. 5.5, namely

Proposition 5.11. *A continuous function with the mean value property is harmonic.*

Proof: Let $f\colon G \to \mathbb{R}$ be a continuous function with the mean value property; choose $z_0 \in G$ and $D_R(z_0) \subset\subset G$ such that (1) holds. Let u be the solution of the Dirichlet problem for $D_R(z_0)$ and the boundary values $f|\partial D_R(z_0)$. Then the difference $f - u$ has the mean value property (Prop. 5.5) and is $\equiv 0$ on $\partial D_R(z_0)$. By the maximum principle, f coincides with u in the neighbourhood $D_R(z_0)$ of z_0 and hence is harmonic there. As z_0 was arbitrary, the claim is proved. □

The proposition shows the remarkable strength of the mean value property: it is a property of continuous functions that forces them to be infinitely often differentiable and even real analytic.

Exercises

1. For the Poisson kernel $P_R(\zeta, z)$ of $D_R(0)$ prove the estimate
$$\frac{1}{2\pi}\,\frac{R-|z|}{R+|z|} \leqslant P_R(\zeta, z) \leqslant \frac{1}{2\pi}\,\frac{R+|z|}{R-|z|}$$
for $|z| < R$. Deduce from this: if u is a nonnegative continuous function on $\overline{D_R(0)}$, harmonic on $D_R(0)$, then for $|z| \leqslant r < R$
$$\frac{R-r}{R+r}\,u(0) \leqslant u(z) \leqslant \frac{R+r}{R-r}\,u(0).$$
2. Prove: A locally uniform limit of harmonic functions is harmonic.
3. Let $u_1 \leqslant u_2 \leqslant u_3 \leqslant \dots$ be a monotonic sequence of harmonic functions on a domain G. Assume that for one point $z_0 \in G$, the sequence $u_n(z_0)$ is bounded. Prove: The u_n converge locally uniformly to a harmonic function. Hint: Use the results of Ex. 1 and 2.
4. Give an example of an unbounded region G, for which the solution of the Dirichlet problem is not unique. Hint: Consider a period strip of a periodic function.
5. Let $D = D_R(0)$ and u be a bounded harmonic function on $D \setminus \{0\}$. Show that u can be extended to a harmonic function on D. Hint: One may assume u to be continuous on $\overline{D} \setminus \{0\}$. Let v be the solution of the Dirichlet problem on $\overline{D}$ with boundary values $u|\partial D$. For $\varepsilon > 0$, apply the maximum principle to $h_\varepsilon = v - u + \varepsilon \log|z|$.

6. Schwarz's reflection principle

We consider holomorphic functions f defined in a domain whose boundary contains a line segment or a circular arc, and look for conditions that guarantee the existence of a holomorphic extension of f across the segment respectively arc to a larger domain.

To begin with, we consider a domain $G \subset \mathbb{C}$, symmetric with respect to the real axis, and set $G^+ = G \cap \mathbb{H} = \{z \in G : \operatorname{Im} z > 0\}$, $G^- = \{z \in G : \operatorname{Im} z < 0\}$ and $J = G \cap \mathbb{R}$. If f is holomorphic on G and real-valued on J, then $f(\overline{z}) = \overline{f(z)}$. For $u = \operatorname{Re} f$ and

$v = \operatorname{Im} f$ this means $u(\bar{z}) = u(z)$, $v(\bar{z}) = -v(z)$. If f is only defined and continuous on $G^+ \cup J$, holomorphic on G^+ and real-valued on J, then it is not difficult to show that f can be extended to a holomorphic function on all of G by defining $f(z) = \overline{f(\bar{z})}$ for $z \in G^-$ (cf. Ex. 7 in II.4). In order to weaken the continuity condition, we first state

Lemma 6.1. *Let $G = G^+ \cup J \cup G^-$ as above and let v be a real harmonic function on G^+. Assume $v(z) \to 0$ if z tends to any point of J. Then*

$$\hat{v}\colon z \mapsto \begin{cases} v(z) & \text{for } z \in G^+ \\ 0 & \text{for } z \in J \\ -v(\bar{z}) & \text{for } z \in G^- \end{cases}$$

is a harmonic extension of v to G.

Proof: As $\hat{v}$ is clearly continuous on G and harmonic on the subdomains G^+ and G^-, it suffices to show that $\hat{v}$ has the mean value property at the points $x_0 \in J$. But, if $D_R(x_0) \subset\subset G$, the equality

$$\int_{-\pi}^{\pi} \hat{v}(x_0 + re^{it})\, dt = 0 = \hat{v}(x_0)$$

(for $r \leqslant R$) follows from $\hat{v}(x_0 + re^{it}) = -\hat{v}(x_0 + re^{-it})$. □

Proposition 6.2 (Schwarz reflection principle)**.** *Let $G = G^+ \cup J \cup G^-$ as above, and let f be a holomorphic function on G^+ such that $v(z) = \operatorname{Im} f(z)$ tends to 0 if z tends to any point of J. Then f has a holomorphic extension $\hat{f}$ to G.*

Note that we make no continuity assumptions on the real part of f! – Of course $\operatorname{Re} f$ and $\operatorname{Im} f$ may be exchanged.

Proof: By the lemma, $v = \operatorname{Im} f$ has a harmonic extension to G, we denote it by v, too. Now let $x \in J$ be arbitrary and choose $D_x = D_r(x) \subset G$. On D_x, the function v has a harmonic conjugate $-\tilde{u}$. As $-\operatorname{Re} f$ is a harmonic conjugate of v on $D_x \cap G^+$, we can assume $\tilde{u} = \operatorname{Re} f$ on $D_x \cap G^+$. Then $\tilde{f}^x = \tilde{u} + iv$ is a holomorphic extension of f to D_x.

For $x_1, x_2 \in J$, the intersection $D_{x_1} \cap D_{x_2}$, if not empty, is connected and intersects G^+. As $\tilde{f}^{x_1} = f = \tilde{f}^{x_2}$ on $D_{x_1} \cap D_{x_2} \cap G^+$, the identity theorem implies $\tilde{f}^{x_1} = \tilde{f}^{x_2}$ on $D_{x_1} \cap D_{x_2}$.

Thus, the functions $\tilde{f}^x$ give rise to a holomorphic extension $\tilde{f}$ of f to $G^+ \cup U$ with $U \subset G$ a neighbourhood of J. Since $v|J = 0$, $\tilde{f}$ is real-valued on J and hence $\tilde{f}(z)$ coincides with $\overline{f(\bar{z})}$ on $U \cap G^-$. The latter function is holomorphic on all of G^- and thus extends $\tilde{f}$ holomorphically to G. □

In the next section, we need a generalization of Prop. 6.2. We first introduce some terminology.

We call an open circular arc or line segment C a free boundary arc of the domain G, if each $z_0 \in C$ has a circular neighbourhood $U(z_0)$ such that $U(z_0) \setminus C$ consists of two components, only one of which is in G – i.e. G is on one side of C only. A holomorphic (meromorphic) function f on G is then said to be extendible across C, if there is an open neighbourhood W of C with $W \cap \partial G = C$ such that f extends to a holomorphic (meromorphic) function on $G \cup W$.

For a meromorphic function f on G and a point $z_0 \in \partial G$, the *cluster set* of f in z_0 is

$$C_G(f, z_0) = \{w \in \widehat{\mathbb{C}} : \text{ there is a sequence } z_n \text{ in } G \text{ with } z_n \to z_0,\ f(z_n) \to w\}.$$

For example, the condition on f in Prop. 6.2 means $C_{G^+}(f, x_0) \subset \mathbb{R}$ for all $x_0 \in J$.

Theorem 6.3. *Let G be a domain with free (circular or rectilinear) boundary arc C and f meromorphic on G. If $C_G(f, z_0) \subset \widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ for all $z_0 \in C$, then f can be extended meromorphically across C.*

Remarks:

a) It is natural to work here with meromorphic functions: if $\infty \in C_G(f, z_0)$, an extension of f will have a pole at z_0.

b) If one assumes $C_G(f, z_0) \subset \widehat{\mathbb{R}} \setminus \{w_0\}$ for a fixed $w_0 \in \widehat{\mathbb{R}}$ and all $z_0 \in C$, the claim can be readily deduced from Prop. 6.2 by applying suitable Möbius transformations (see Ex. 2).

Proof: a) We simplify the situation, first by choosing a Möbius transformation T that maps C on an open interval $J \subset \mathbb{R}$. Setting $G_1 = T(G)$, $f_1 = f \circ T^{-1}$, we have $C_{G_1}(f_1, x_0) \subset \widehat{\mathbb{R}}$ for all $x_0 \in J$. Secondly, we replace f_1 by $F = \dfrac{f_1 - i}{f_1 + i}$. Since $w \mapsto (w - i)/(w + i)$ maps $\widehat{\mathbb{R}}$ onto $\partial\mathbb{D}$, the cluster sets of F at points $x_0 \in J$ are contained in $\partial\mathbb{D}$. We shall show that F can be holomorphically extended across J; then the meromorphic extendibility of f_1 and f follows. Since $f_1 = i(1 + F)/(1 - F)$, the points with $F(z) = 1$ are poles of f_1.

b) Now fix $x_0 \in J$ and choose $\varepsilon = \varepsilon(x_0) > 0$ such that

i. $[x_0 - \varepsilon, x_0 + \varepsilon] \subset J$,
ii. $D^+ = D_\varepsilon(x_0) \cap G$ is the upper (or lower) half of $D_\varepsilon(x_0)$,
iii. $1/2 < |F(z)| < 2$ for $z \in D^+$.

There is in fact an $\varepsilon > 0$ such that *iii* is satisfied: Otherwise there would be a sequence z_n in D^+ with $z_n \to x_0$ and $|F(z_n)| \leqslant 1/2$ or $|F(z_n)| \geqslant 2$ for all n, contradicting $C_G(F, x_0) \subset \partial\mathbb{D}$.

In view of *iii*, there exists a holomorphic $\log F$ on the simply connected domain D^+, we have $\operatorname{Re}\log F(z) = \log|F(z)| \to 0$ if z tends to a point of $]x_0 - \varepsilon, x_0 + \varepsilon[$. By Prop. 6.2, $\log F$ extends to a holomorphic function on $G \cup D_\varepsilon(x_0)$, hence so does F.

The "local" extensions of F for all $x_0 \in J$ give rise to a well defined holomorphic extension of F across J, this follows as in the last part of the proof of Prop. 6.2. $\square$

We conclude with some remarks on reflection in Möbius circles. The reflection in $\widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ is complex conjugation $\kappa\colon z \mapsto \bar{z}$. It is uniquely determined by being antiholomorphic and having $\widehat{\mathbb{R}}$ as fixed point set: If $\sigma\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is another such map, $\kappa\sigma\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is holomorphic and the identity map on $\widehat{\mathbb{R}}$, hence everywhere, i.e. $\sigma = \kappa$.

If $K \subset \widehat{\mathbb{C}}$ is any Möbius circle, i.e. a euclidean circle or a straight line (with ∞ added), there is a unique antiholomorphic map $\sigma_K\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ fixing K pointwise: Pick a Möbius transformation S carrying K into $\widehat{\mathbb{R}}$ and define $\sigma_K = S^{-1}\kappa S$. Uniqueness follows as above. We call σ_K the reflection in K.

We visualize σ_K for a euclidean circle $K = \{z \in \mathbb{C} : |z - z_0| = r\}$. With

$$S\colon K \to \widehat{\mathbb{R}}, \quad Sz = i\frac{r - (z - z_0)}{r + (z - z_0)},$$

$\sigma_K = S^{-1}\kappa S$ computes to

$$\sigma_K(z) = \frac{r^2}{\bar{z} - \bar{z}_0} + z_0 \qquad \text{or} \qquad \big(\sigma_K(z) - z_0\big)\big(\overline{z - z_0}\big) = r^2.$$

The last formula shows that, for $z \neq z_0$, z and $\sigma_K(z)$ lie on a ray starting at z_0, and that the distances $|\sigma_K(z) - z_0|$ and $|z - z_0|$ multiply to r^2. In particular, σ_K exchanges z_0 and ∞ as well as the interior and the exterior of K.

Now consider a domain G, a Möbius circle K such that $G \cap K$ is an arc, and a holomorphic function f on G which maps $G \cap K$ into another Möbius circle L. We claim $\sigma_L \circ f = f \circ \sigma_K$ whenever both sides are defined, i.e. on $G \cap \sigma_K G$. Indeed, choose Möbius transformations S with $S(K) = \widehat{\mathbb{R}}$ and T with $T(L) = \widehat{\mathbb{R}}$. Then $\tilde{f} = TfS^{-1}\colon S(G) \to \widehat{\mathbb{C}}$ is real-valued on $S(G \cap K) = S(G) \cap \widehat{\mathbb{R}}$, hence $\tilde{f}(\bar{z}) = \overline{\tilde{f}(z)}$. Inserting the definitions of σ_K and σ_L proves the claim.

Exercises

1. Let f be a conformal map of the annulus $K_1 = \{z : r_1 < |z| < r_2\}$ onto the annulus $K_2 = \{w : \varrho_1 < |w| < \varrho_2\}$, $r_1, \varrho_1 > 0$. Show that f can be extended to a biholomorphic map $\hat{f}\colon \mathbb{C}^* \to \mathbb{C}^*$. Conclude $\varrho_2/\varrho_1 = r_2/r_1$.
2. Let K and L be Möbius circles, $w_0 \in L$ ($w_0 = \infty$ admitted), let G be a domain symmetric with respect to K, i.e. $\sigma_K G = G$, denote G^+ one of the two (!) components of $G \setminus K$. Show that a holomorphic function $f\colon G^+ \to \mathbb{C}$ with $C_{G^+}(f, z) \subset L \setminus \{w_0\}$ can be extended holomorphically to G (use Prop. 6.2 only!).

7. The modular map λ

In this section, we construct a holomorphic function $\lambda\colon \mathbb{H} \to \mathbb{C} \setminus \{0, 1\}$ which exhibits the upper half plane $\mathbb{H}$ as universal covering of the sphere $\widehat{\mathbb{C}}$ punctured in 0, 1, ∞. In the next section, the existence of λ will be essential in proving the Little and Big Picard theorems. – The term universal covering is taken from topology: An (unbranched) covering is a continuous surjective map $f\colon X \to Y$ of topological spaces with the property: Every point y of Y has a neighbourhood V such that the preimage $f^{-1}(V)$ is the disjoint union of sets $U_\iota \subset X$, each of which is mapped homeomorphically onto V by f. A covering is *universal* if X is connected and simply connected. An easy example is the exponential function $\exp\colon \mathbb{C} \to \mathbb{C}^*$, which is a universal covering map.

As a first step, we tessellate $\mathbb{H}$ by hyperbolic triangles, starting from a special one and using iterated reflections. In a second step, we define λ, following the construction of the tessellation and using the reflection principle.

It is convenient to use two terms of hyperbolic geometry (we do not need more of VII.3): An orthocircle (in $\mathbb{H}$) is either a euclidean semicircle with centre on $\mathbb{R}$ or a vertical ray $\{z : \operatorname{Re} z = x_0,\ \operatorname{Im} z > 0\}$. The reflection σ_K in an orthocircle K is defined in the obvious manner, it interchanges the two parts into which K divides $\mathbb{H}$. – A hyperbolic triangle is a curvilinear triangle in $\overline{\mathbb{H}} = \mathbb{H} \cup \mathbb{R} \cup \{\infty\}$ bounded by arcs of orthocircles. We allow the vertices to lie on $\widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$, in fact, only such triangles will occur. – We shall often omit the attribute “hyperbolic”.

We start with the open hyperbolic triangle Δ_0 with vertices 0, 1, ∞, i.e. the set

$$\left\{z \in \mathbb{H} : 0 < \operatorname{Re} z < 1,\ \left|z - \tfrac{1}{2}\right| > \tfrac{1}{2}\right\}.$$

We then reflect Δ_0 in each of its edges to obtain three new triangles Δ_1 (of “first generation”), then reflect each Δ_1 in its two free edges to obtain six new triangles Δ_2 (triangles of “second generation”). Repeating this process infinitely often results in a net of triangles in $\mathbb{H}$ with vertices on $\widehat{\mathbb{R}}$. This net is often called a modular net. Different (open) triangles of the net do not intersect: It suffices to show $\Delta_0 \cap \Delta = \varnothing$ for $\Delta \neq \Delta_0$. This may be seen by induction on the “generation” (cf. Fig. 21) – we defer the details to Ex. 9.

Furthermore, our net covers all of $\mathbb{H}$, more precisely, $\mathbb{H}$ is contained in the union A of the closures of our triangles. To see this, let $\Pi_0 = \overline{\Delta}_0$, and if Π_k is already defined, let Π_{k+1} be the union of Π_k with all its reflections in its edges. Thus Π_1 is a curvilinear hexagon consisting of $\overline{\Delta}_0$ and the three $\overline{\Delta}_1$'s, all of whose vertices lie on $\widehat{\mathbb{R}}$, Π_2 is a 30-gon etc. We have

$$\Pi_0 \subset \Pi_1 \subset \Pi_2 \subset \dots\ , \qquad \bigcup_k \Pi_k = A.$$

The boundary of Π_k consists of two vertical rays L_1, L_2 and a collection of euclidean semicircles. The reflection σ_K in one of these semicircles takes ∞ to the centre M_K

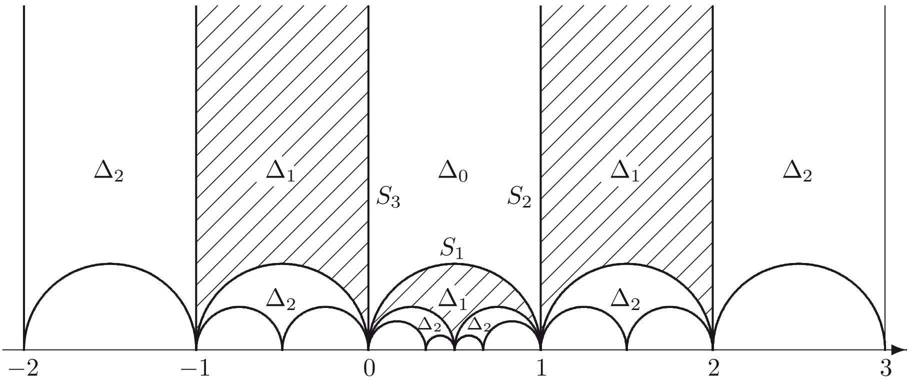


Figure 21. The modular net

of K, hence $\sigma_K L_j$ is a semicircle in the interior of K abutting in M_K, and the radius of $\sigma_K L_j$ is at most half the radius of K. By this we see inductively

$$\Pi_k \supset \{x + iy : 0 \leqslant x \leqslant 1,\, y \geqslant 2^{-k-1}\},$$

and $A \supset \{x + iy : 0 \leqslant x \leqslant 1,\, y > 0\}$. As A is clearly invariant under translations $z \mapsto z + k$, $k \in \mathbb{Z}$, the claim $A \supset \mathbb{H}$ follows.

We proceed to the construction of λ.

We start with a biholomorphic mapping $f\colon \Delta_0 \to \mathbb{H}$. Consider a sequence z_n in Δ_0 tending to a point z_0 of the boundary $\partial\Delta_0$ (with respect to $\widehat{\mathbb{C}}$). As f is bijective, $f(z_n)$ cannot have an accumulation point in $\mathbb{H}$, therefore the cluster sets $C_{\Delta_0}(f, z_0)$ are in $\widehat{\mathbb{R}}$ for all $z_0 \in \partial\Delta_0$.

Denote S_j, $j = 1, 2, 3$, the (open) edges of Δ_0 as in Fig. 21 and let σ_j be the reflection in S_j. For the moment fix j. By Thm. 6.3, f can be meromorphically extended across S_j; by the remark at the end of the last section, this extension – denoted f, too – coincides with $\overline{f \circ \sigma_j}$ on $\sigma_j\Delta_0$, hence maps $\sigma_j\Delta_0$ biholomorphically onto $\mathbb{H}^- = \{w \in \mathbb{C} : \operatorname{Im} w < 0\}$. As f is injective on Δ_0 and $\sigma_j\Delta_0$, f takes all its values on S_j with multiplicity 1 (cf. Prop. III.5.4); of course, f may have a pole on S_j. Hence S_j is mapped bijectively onto an open arc $J_j \subset \widehat{\mathbb{R}}$, and $\Delta_0 \cup S_j \cup \sigma_j\Delta_0$ biholomorphically onto $\mathbb{H} \cup J_j \cup \mathbb{H}^-$.

This works for $j = 1, 2, 3$. We claim moreover, that the arcs J_1, J_2, J_3 are disjoint: Let $z_1, z_2 \in \bigcup S_j$, $z_1 \neq z_2$. Choose disjoint neighbourhoods U_1, U_2 of z_1, z_2, symmetric with respect to the pertinent σ_j. Then $f(U_1 \cap \Delta_0)$ and $f(U_2 \cap \Delta_0)$ are disjoint, by symmetry so are $f(U_1)$ and $f(U_2)$, hence $f(z_1) \neq f(z_2)$.

We prove next that $\widehat{\mathbb{R}} \setminus (J_1 \cup J_2 \cup J_3)$ consists of three points only: Assume there is an open arc $J \subset \widehat{\mathbb{R}}$ disjoint from $J_1 \cup J_2 \cup J_3$; we may suppose $\infty \notin J$. Consider the domain $G = \mathbb{H} \cup J \cup \mathbb{H}^-$ and the function $g = (f|\Delta_0)^{-1}\colon \mathbb{H} \to \Delta_0$. If w_n is a sequence in $\mathbb{H}$ converging to $w_* \in J$, the points $g(w_n)$ can only accumulate at the vertices 0, 1, ∞ of Δ_0, i.e. $C_{\mathbb{H}}(g, w_*) \subset \{0, 1, \infty\} \subset \widehat{\mathbb{R}}$ for $w_* \in J$. By Thm. 6.3, g has a meromorphic extension $\hat{g}$ to G. In particular, $\hat{g}$ is continuous on J with values in $\{0, 1, \infty\}$, hence constant on J, hence constant on all of G – a contradiction.

We denote by w_0 resp. w_1 resp. w_∞ the points of $\widehat{\mathbb{R}}$ between J_3 and J_1 resp. J_1 and J_2 resp. J_2 and J_3. Setting $f(0) = w_0$, $f(1) = w_1$, $f(\infty) = w_\infty$ yields a continuous extension of f to the vertices 0, 1, ∞ of Δ_0. Of course, we then define $f(\sigma_j 0) = w_0$ etc., thus obtaining a continuous extension of f to $\overline{\Delta_0 \cup \bigcup \sigma_j \Delta_0}$.

Up to now, f has been arbitrary under the condition that $f\colon \Delta_0 \to \mathbb{H}$ be biholomorphic. We now normalize and define

$$\lambda(z) = T\big(f(z)\big),$$

where $T \in \operatorname{Aut} \mathbb{H}$ is such that $Tw_0 = 1$, $Tw_1 = \infty$, $Tw_\infty = 0$. Note that the possible pole of f on one of the S_j disappears under this transformation! The properties of f translate to properties of λ, namely:

i. λ defines a continuous bijection of $\overline{\Delta_0}$ (closure in $\widehat{\mathbb{C}}$) onto $\overline{\mathbb{H}} = \mathbb{H} \cup \widehat{\mathbb{R}}$, which maps Δ_0 biholomorphically onto $\mathbb{H}$, the vertices 0, 1, ∞ of Δ_0 to 1, ∞, 0, and the (open) edges of Δ_0 to the real intervals between these points.

ii. λ can be extended by reflection in the edges of Δ_0 to continuous functions $\overline{\Delta_0 \cup \sigma_j \Delta_0} \to \widehat{\mathbb{C}}$, which are locally biholomorphic in the interior and map $\sigma_j \Delta_0$ biholomorphically onto $\mathbb{H}^-$ ($j = 1, 2, 3$).

iii. The interior of $\overline{\Delta_0 \cup \bigcup \sigma_j \Delta_0}$ is mapped onto $\widehat{\mathbb{C}} \setminus \{0, 1, \infty\}$.

We can now extend the domain of definition of λ successively to $\mathbb{H}$: First, extend λ to the triangles Δ_2 of "second generation" by reflection in the free edges of the triangles of first generation $\sigma_j \Delta_0$, $j = 1, 2, 3$. This extension maps each Δ_2 biholomorphically onto $\mathbb{H}$ and its boundary onto $\widehat{\mathbb{R}}$. Then continue in this manner, always using reflections in the free edges of the polygon to which λ has been extended in the previous step. – We summarize the properties of λ:

Theorem 7.1. *Given the modular net in $\mathbb{H}$ described above, there is a unique holomorphic function $\lambda\colon \mathbb{H} \to \mathbb{C} \setminus \{0, 1\}$ with the following properties:*

i. *λ is surjective and locally biholomorphic,*

ii. *λ maps each triangle of the modular net biholomorphically onto $\mathbb{H}$ or $\mathbb{H}^-$,*

iii. *λ maps the open edges of any of these triangles on the intervals $]0, 1[$, $]1, +\infty[$, $]-\infty, 0[$,*

iv. $\lim_{z\to 0} \lambda(z) = 1$, $\lim_{z\to 1} \lambda(z) = \infty$, $\lim_{z\to \infty} \lambda(z) = 0$,

v. *λ is a covering map.*

Property v follows from the construction: For example, the preimage $\lambda^{-1}(U)$ of an open set $U \subset \mathbb{H}$ consists of open sets, one in each triangle of even generation, which are mapped biholomorphically onto U. In fact, λ is a universal covering map because $\mathbb{H}$ is simply connected. We call λ the *modular map*.

A final remark: If S is a Möbius transformation mapping $\mathbb{D}$ onto $\mathbb{H}$, then S transfers the modular net from $\mathbb{H}$ to $\mathbb{D}$ and $\lambda \circ S\colon \mathbb{D} \to \mathbb{C} \setminus \{0, 1\}$ is a universal covering map, too.

Much more can be said about this function λ, which is an example of a modular function. There are connections to group theory (cf. exercises) and, surprisingly, to elliptic functions and algebraic number theory. We do not go into this.

Exercises

Notations: σ_1, σ_2, σ_3 and $\lambda(z)$ as in the text. Γ_0 is the group of mappings $\mathbb{H} \to \mathbb{H}$ the elements of which are products of an even number of factors σ_j. $\Gamma(2)$ is the subgroup of $\operatorname{Aut}\mathbb{H}$ consisting of the transformations $z \mapsto (az+b)/(cz+d)$ with $ad - bc = 1$, a and d odd, b and c even.

1. Γ_0 is generated by $\tau_1 = \sigma_2\sigma_3$ and $\tau_2 = \sigma_3\sigma_1$ (note $\tau_3 = \sigma_1\sigma_2 = (\tau_1\tau_2)^{-1}$!). Find explicit formulas for $\tau_1(z)$ and $\tau_2(z)$, conclude $\Gamma_0 \subset \Gamma(2)$. Moreover, λ is Γ_0-invariant, i.e. $\lambda(\tau z) = \lambda(z)$ for all $\tau \in \Gamma_0$.
2. A fundamental set of Γ_0 is a subset $\mathcal{F} \subset \mathbb{H}$ such that $\mathbb{H} = \bigcup_{\tau\in\Gamma_0} \tau(\mathcal{F})$ and $\tau(\mathcal{F}) \cap \tau'(\mathcal{F}) = \varnothing$ for $\tau, \tau' \in \Gamma_0$, $\tau \neq \tau'$. Show that $\mathcal{F} = \{z \in \mathbb{H} : -1 \leqslant \operatorname{Re} z < 1, |z + \frac{1}{2}| \geqslant \frac{1}{2}, |z - \frac{1}{2}| > \frac{1}{2}\}$ is a fundamental set of Γ_0 (use the mapping properties of λ). Conclude: $\lambda(z_1) = \lambda(z_2)$ if and only if $z_2 = \tau z_1$ with $\tau \in \Gamma_0$.
3. Prove $\Gamma_0 = \Gamma(2)$. This is equivalent to proving: Every $A = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in SL_2(\mathbb{Z})$ with $A \equiv I \mod 2$ is $(\pm 1) \cdot$ product of powers of T_1 and T_2, $T_1 = \left(\begin{smallmatrix} 1 & 2 \\ 0 & 1\end{smallmatrix}\right)$, $T_2 = \left(\begin{smallmatrix} 1 & 0 \\ -2 & 1\end{smallmatrix}\right)$. Hint: The case $b = 0$ is clear. For $b \neq 0$, consider the effect of right multiplication by ${T_1}^k$ or ${T_2}^k$ on the first row of A, combine to reduce to the case $b = 0$.
4. Prove the functional equations
$$\lambda(-1/z) = 1 - \lambda(z),\ \lambda(z+1) = \lambda(z)/(\lambda(z) - 1) \quad \text{for } z \in \mathbb{H}.$$
Hint: Consider the mappings $\overline{\Delta_0} \to \overline{\mathbb{H}}$ induced by the sides of the equations.
5. Show that $\Gamma = SL_2(\mathbb{Z})/\{\pm I\} \subset \mathcal{M}$ is generated by $S\colon z \mapsto -1/z$ and $T\colon z \mapsto z+1$. (Proof similar to, but simpler than proof in Ex. 3).
6. Expressing the functional equations of Ex. 4 in the form $\lambda \circ S = U \circ \lambda$, $\lambda \circ T = W \circ \lambda$ with suitable $U, W \in GL_2(\mathbb{Z})/\{\pm I\}$, conclude the existence of a homomorphism of Γ onto the group $\mathcal{S}$ generated by U and W, the kernel of which is Γ_0.
7. The group $\mathcal{S}$ of Ex. 6 is the group of automorphisms of $\widehat{\mathbb{C}} \setminus \{0, 1, \infty\}$. It is isomorphic to the full permutation group $\mathcal{S}_3$ on three letters. Hence $[\Gamma : \Gamma_0] = 6$.
8. Calculate $\lambda(i)$, $\lambda\big(\frac{1+i}{2}\big)$, $\lambda(1+i)$, $\lambda\big(e^{2\pi i/3}\big)$, $\lambda\big(e^{\pi i/3}\big)$.
9. Show: The triangles Δ_n of n-th generation ($n \geqslant 1$) in the modular net are contained in $\bigcup_{k\in\mathbb{Z}} \{z : \big|z - \big(k + \frac{1}{2}\big)\big| < \frac{1}{2}\}$ and their "upper edge" is not free, excepting the triangles with vertices $(-n, -n+1, \infty)$ and $(n, n+1, \infty)$. Hence they do not intersect Δ_0.

8. Theorems of Picard and Montel

We shall admit of the following topological statement: *Let $p\colon X \to Y$ a covering map, and assume the topological space Z to be path connected, locally path connected, and simply connected. Then for every continuous map $f\colon Z \to Y$ and every choice of $z_0 \in Z$ and $x_0 \in X$ satisfying $f(z_0) = p(x_0)$, there exists a unique continuous map $F\colon Z \to X$ with $p \circ F = f$ and $F(z_0) = x_0$.*

In our context, X, Y, Z will be domains in $\widehat{\mathbb{C}}$. They are path connected by definition and locally path connected as open sets, so the relevant condition is that Z be simply connected.

In this case, the so-called "lift" F of f is automatically holomorphic if p and f are: With $z_1 \in Z$, let $U \subset X$ be a neighbourhood of $F(z_1)$ that is mapped by p biholomorphically onto $p(U)$, and let V be a neighbourhood of z_1 with $F(V) \subset U$. Then $F|V = (p|U)^{-1} \circ (f|V)$ is holomorphic.

In the last section we have constructed a holomorphic universal covering map $\lambda\colon \mathbb{D} \to \mathbb{C} \setminus \{0, 1\} =: \mathbb{C}''$. Together with the above, the "Little Picard Theorem" follows almost immediately:

Theorem 8.1. *Let $f\colon \mathbb{C} \to \mathbb{C}$ be a non-constant entire function. Then f assumes every complex number with at most one exception as a value.*

Proof: Suppose that f is an entire function with values in $\mathbb{C} \setminus \{a, b\}$, $a \neq b$. Then $f_1(z) = \dfrac{f(z) - a}{f(z) - b}$ is an entire function with values in $\mathbb{C} \setminus \{0, 1\}$ and can be lifted to a holomorphic function $F_1\colon \mathbb{C} \to \mathbb{D}$. By Liouville's theorem, F_1 is constant, and so are $f_1 = \lambda \circ F_1$ and f. □

To prove the "Big Theorem" of Picard, we need a considerably sharpened version of the Montel theorem II.5.3, which we now present.

Theorem 8.2 (Montel). *Let f_n be a sequence of holomorphic functions on a domain G, that do not assume the values 0 and 1. Then f_n has a subsequence f_{n_k} converging locally uniformly to a holomorphic function $f\colon G \to \mathbb{C}$ or to the constant map $G \to \{\infty\}$.*

The latter means: for all compact $K \subset G$ and all radii R, $f_{n_k}(K) \cap D_R(0) = \varnothing$ for k sufficiently large.

Proof: a) We assume G to be simply connected and fix $z_0 \in G$. Assume further that the numbers $f_n(z_0)$ have an accumulation point w_0 in $\mathbb{C}''$ (the case that all accumulation points of $f_n(z_0)$ are in $\{0, 1, \infty\}$ will be considered in part b). Replacing f_n by a subsequence, again noted f_n, we may suppose $f_n(z_0) \to w_0$.

Now consider the universal cover $\lambda\colon \mathbb{D} \to \mathbb{C}''$, choose a point $\zeta_0 \in \mathbb{D}$ with $\lambda(\zeta_0) = w_0$ and a connected neighbourhood $V \subset \mathbb{C}''$ of w_0 such that the components of $\lambda^{-1}(V)$ are mapped biholomorphically onto V by λ. Denote U the component of ζ_0 in $\lambda^{-1}(V)$.

The values $f_n(z_0)$ are in V for $n \geqslant n_0$; in the sequel we always suppose $n \geqslant n_0$. Set $\zeta_n = (\lambda|U)^{-1} f_n(z_0)$. Then $\zeta_n \to \zeta_0$. Now let $F_n\colon G \to \mathbb{D}$ be the lift of f_n with $F_n(z_0) = \zeta_n$. As the F_n are uniformly bounded, by Thm. II.5.3 we can choose a subsequence converging locally uniformly to $F\colon G \to \overline{\mathbb{D}}$. We denote this subsequence again by F_n. If $F(G) \cap \partial\mathbb{D} \neq \varnothing$, then F is a constant of modulus 1 by the maximum principle. But $F(z_0) = \lim F_n(z_0) = \zeta_0 \in \mathbb{D}$, so $F\colon G \to \mathbb{D}$. To finish the proof in the case considered, we show that $f_n = \lambda \circ F_n \to \lambda \circ F =: f$ locally uniformly. Indeed, if $K \subset G$ is compact, there is a radius $R < 1$ such that $F(K)$ and the $F_n(K)$ are contained in $\{\zeta : |\zeta| \leqslant R\}$. Then, for $z \in K$

$$\begin{aligned} |f_n(z) - f(z)| &= |\lambda\big(F(z)\big) - \lambda\big(F_n(z)\big)| \\ &= \Big| \int_{[F_n(z),\ F(z)]} \lambda'(\zeta)\, d\zeta \Big| \leqslant \max_{|\zeta|\leqslant R} |\lambda'| \cdot \sup_K \big|F_n - F\big| \to 0. \end{aligned}$$

b) We still assume that G is simply connected and suppose now that $0, 1, \infty$ are the only possible accumulation points of $f_n(z_0)$, for a fixed $z_0 \in G$.

Assume first that, after passing to a subsequence, we have $f_n(z_0) \to 1$. As G is simply connected and f_n without zeros, there are holomorphic square roots $g_n = \sqrt{f_n}\colon G \to \mathbb{C}''$. We choose the branches such that $g_n(z_0) \to -1$. Then, by a), g_n has a subsequence g_{n_k} converging locally uniformly to a holomorphic function $g\colon G \to \mathbb{C}''$, and the $f_{n_k} = g_{n_k}^2$ converge locally uniformly to g^2.

If 1 is not an accumulation point of $f_n(z_0)$, but 0 is, we replace f_n by $1 - f_n$ and apply the above.

If $f_n(z_0) \to \infty$, we replace f_n by $1 - 1/f_n$ and again apply the above. Note that, if $\sqrt{1 - 1/f_n}$ converges to the constant -1, the f_{n_k} will tend to ∞.

c) Parts a) and b) prove the theorem for simply connected G. Now let the domain G be arbitrary. We choose a sequence of disks D_k such that

$$D_k \subset\subset G, \quad G = \bigcup_k D_k, \quad D_k \cap \bigcup_1^{k-1} D_j \neq \varnothing \ \ (k \geqslant 2)$$

(cf. Ex. 1). Let $f_n\colon G \to \mathbb{C}''$ be a sequence of holomorphic functions. Then there is a subsequence $\big(f_{n1}\big)$ converging uniformly on D_1 to a map $\hat{f}^1\colon D_1 \to \widehat{\mathbb{C}}$ which is either a holomorphic function or $\equiv \infty$ (apply a) resp. b) to a disk D_1' with $D_1 \subset\subset D_1' \subset\subset G$). Now assume that for $k > 1$ we have found subsequences $\big(f_{n1}\big), \big(f_{n2}\big), \ldots, \big(f_{n,k-1}\big)$ of $\big(f_n\big)$ such that

$\big(f_{nj}\big)$ is a subsequence of $\big(f_{n,j-1}\big)$,

$$\big(f_{n,k-1}\big) \text{ converges uniformly on } \bigcup_1^{k-1} D_j \text{ to } \hat{f}^{k-1}\colon \bigcup_1^{k-1} D_j \to \widehat{\mathbb{C}},$$
$$\text{which is a holomorphic function or } \equiv \infty.$$

Then $\big(f_{n,k-1}\big)$ has a subsequence $\big(f_{nk}\big)$ converging uniformly on D_k to a map $f^k\colon D_k \to \widehat{\mathbb{C}}$. On $D_k \cap \bigcup_1^{k-1} D_j$ – an open nonvoid set – f^k coincides with $\hat{f}^{k-1}$, hence f^k defines an extension of $\hat{f}^{k-1}$ to $\bigcup_1^{k} D_j$ which is either a holomorphic function or $\equiv \infty$. The diagonal sequence $\big(f_{nn}\big)$ then converges locally uniformly on G either to a holomorphic function or to ∞. □

We are now in a position to prove the "big theorem" of Picard:

Theorem 8.3 (Picard)**.** *Let f be holomorphic on the punctured neighbourhood $U\setminus\{z_0\}$ of z_0 and have an essential singularity at z_0. Then each number $w \in \mathbb{C}$, with one possible exception, is assumed infinitely often as a value of f in $U \setminus \{z_0\}$.*

An exceptional "non-value" may indeed occur, as shown by $\exp(1/z)$ at $z_0 = 0$.

Proof: Suppose that $f \in \mathcal{O}(U \setminus \{z_0\})$ attains two values $a, b \in \mathbb{C}$, $a \neq b$, only finitely often. We will show that, in this case, z_0 is a removable singularity or a pole of f.

By shrinking U, we may assume that f omits the values a and b. As in the proof of Thm. 8.1, we normalize $a = 0$, $b = 1$. To simplify notations, suppose $z_0 = 0$ and $U = D_R(0)$, $R > 1$. For $n = 0, 1, 2, \ldots$ define $f_n(z) = f(2^{-n}z)$ on $U \setminus \{0\}$.

Then f takes on $|z| = 2^{-n}$ the same values as f_n takes on $|z| = 1$, and no f_n takes the values 0 and 1. By Thm. 8.2, there is a subsequence f_{n_k} of f_n converging locally uniformly on $U \setminus \{0\}$ either to a holomorphic function g or to ∞. In the first case, the f_{n_k} are uniformly bounded on $|z| = 1$, say $|f_{n_k}| \leqslant M$ on $\partial\mathbb{D}$.

This implies $|f(z)| \leqslant M$ on $|z| = 2^{-n_k}$, $k = 1, 2, \ldots$. By the maximum modulus principle, $|f(z)| \leqslant M$ for $2^{-n_{k+1}} \leqslant |z| \leqslant 2^{-n_k}$, $k = 1, 2, \ldots$, hence $|f(z)| \leqslant M$ on $0 < |z| < 2^{-n_1}$. The Riemann extension theorem then implies that 0 is a removable singularity of f.

– In the second case, $1/f_{n_k} \to 0$ locally uniformly in $U \setminus \{0\}$. By the above, $1/f$ has a removable singularity at 0, hence $1/f$ has a zero in 0 and f has a pole. □

Exercises

1. a) Prove that any domain $G \subset \mathbb{C}$ may be covered by a denumerable family of disks $D_k \subset\subset G$, $k = 1, 2, 3, \ldots$. Hint: Choose the points of a denumerable dense subset $\{z_1, z_2, z_3, \ldots\} \subset G$ as centres.

   b) Show that by renumbering the D_k from a), one can achieve $D_k \cap \bigcup_1^{k-1} D_j \neq \varnothing$, $k = 2, 3, \ldots$.

Hints and solutions of selected exercises

As a rule, there is more than one way to solve any problem. The way proposed here may be suboptimal, the reader may find a better one.

In the sequel, a symbol like IV.6.2 refers to exercise 2 in section 6 of chapter IV. The notation used in the text of an exercise will be used in its solution without further explanation.

I.1.3. $f(z) = |z|^2$ is complex differentiable only at the origin; the other functions nowhere.

I.2.1. The chain rule for Wirtinger derivatives is

$$\frac{\partial}{\partial z}(f \circ g) = \frac{\partial f}{\partial w} \cdot \frac{\partial g}{\partial z} + \frac{\partial f}{\partial \overline{w}} \cdot \frac{\partial \overline{g}}{\partial z}, \qquad \frac{\partial}{\partial \overline{z}}(f \circ g) = \frac{\partial f}{\partial w} \cdot \frac{\partial g}{\partial \overline{z}} + \frac{\partial f}{\partial \overline{w}} \cdot \frac{\partial \overline{g}}{\partial \overline{z}}.$$

I.2.3. $AJ_f^{\mathbb{R}}A^{-1} = J_f^{\mathbb{C}}$ with $A = \left(\begin{smallmatrix} 1 & i \\ 1 & -i \end{smallmatrix}\right)$.

I.3.1. Let $M \subset \mathbb{R}^n$ be arbitrary. A sequence of functions on M converges compactly on M if it converges locally uniformly. The converse is true if every point of M has a compact neighbourhood in M. With $M = \mathbb{R}\backslash\{1, \frac{1}{2}, \frac{1}{3}, \dots\}$, this condition is not satisfied, and $f_n(x) = n$ when $1/(n+1) < x < 1/n$, $f_n(x) = 0$ else, converges compactly on M, but not uniformly in any neighbourhood of 0.

I.3.4. The radii are 1, ∞, 0, 0, $\frac{1}{4}$.

I.3.6. The nontrivial implication: If the subseries converge, then $\sum|\mathrm{Re}\, a_\nu| < \infty$, $\sum|\mathrm{Im}\, a_\nu| < \infty$, hence $\sum|a_\nu| \leqslant \sum\big(|\mathrm{Re}\, a_\nu| + |\mathrm{Im}\, a_\nu|\big) < \infty$.

I.4.2. Let G be an infinite subgroup of $\mathbb{S}$. As $\mathbb{S}$ is compact, there are $z_\nu \in G$ with $z_\nu \to 1$, $z_\nu \neq 1$. Any arc of $\mathbb{S}$ contains points ${z_\nu}^k$ (ν large), thus G is dense in $\mathbb{S}$.

I.4.4. To study the mapping properties, decompose $\tan z = g(h(z))$ with $\zeta = h(z) = e^{2iz}$ and $g(\zeta) = (\zeta - 1)/i(\zeta + 1)$.

II.2.2. The first integral is $\sqrt{\pi}\exp(-a^2)$, the second $\sqrt{\pi}\exp(-\lambda^2/4)$.

II.2.4. The claims combine to

$$(1 - ia)\int_0^\infty \exp\big(-(1 - ia)^2x^2\big)\, dx = \frac{\sqrt{\pi}}{2}.$$

To prove this, apply the Cauchy integral theorem to the function $\exp(-z^2)$ and the triangle with vertices 0, R, $(1 - ia)R$, then let $R \to +\infty$.

II.3.1.c) The function $f(z) = \exp(1-z)/(1-z)$ is holomorphic in a neighbourhood of $\overline{D_{1/2}(0)}$, hence the value of the integral is $\pi i f''(0) = i\pi e$.

II.4.3. The recursion formula is $E_0 = 1$, $\sum\limits_0^m \binom{2m}{2n} E_{2n} = 0$ $(m \geqslant 1)$. Hint for the addenda: For $f(z) = \tan z$ we have $f^{(n)}(z) = f(z)P_n(\tan z)$ with polynomials $P_n(X)$ satisfying a recursion formula

II.4.6a) The function $\overline{f(\bar{z})}$ is holomorphic on G and coincides with f on $G \cap \mathbb{R}$.

W. Fischer, I. Lieb, *A Course in Complex Analysis*, DOI 10.1007/978-3-8348-8661-3,

II.5.2. For $|z| \leqslant r < 1$ we have $\left|\frac{z^n}{1-z^n}\right| \leqslant \frac{r^n}{1-r^n}$, hence the series converges uniformly on $|z| \leqslant r$. Inserting geometric series for $z^n/(1-z^n)$ yields the power series $\sum a_m z^m$ with a_m equal to the number of divisors of m.

II.5.3a) Write $p(z) = z^n q\left(\frac{1}{z}\right)$. If q is not constant, then $1 = q(0) < M_1(q) = M_1(p)$.

II.5.3b) $r^{-n} M_r(p) = M_{1/r}(q)$; if $q \not\equiv 1$, $M_{1/r}(q)$ is a strictly decreasing function of r. Furthermore, $\lim\limits_{r\to\infty} r^{-n} M_r(p) = 1$.

II.5.6. For $a_0 = 0$ the claim is clear. – Without loss of generality $a_0 = 1$. Write $f(z) = 1 + a_m z^m\big(1 + zh(z)\big)$, choose r_0 smaller than the radius of convergence of the series, then C with $|h(z)| \leqslant C$ for $|z| \leqslant r_0$. For $0 < r \leqslant r_0$ choose z_0 such that $a_m z_0^m = |a_m| r^m > 0$. Then for $r < 1/(2C)$

$$|f(z_0)| \geqslant 1 + |a_m| r^m - |a_m| r^{m+1} C > 1.$$

II.5.7a) For $|z| < \varrho$ by $|a_n| \leqslant \varrho^{-n} M_\varrho(f)$

$$|f(z) - 1| \leqslant \sum_1^\infty \varrho^{-n} M_\varrho(f) |z|^n = M_\varrho(f)|z|(\varrho - |z|)^{-1} < 1,$$

if $|z| < \varrho\big(1 + M_\varrho(f)\big)^{-1}$.

II.6.3. Let z_0 be a pole of f. Then $f\big(D_r(z_0) \setminus \{z_0\}\big)$, r small, contains the complement of a disk, and this contains period strips of exp ... The other cases are simpler.

II.6.5. Without loss of generality $z_0 = 0$, $a_n(0) \neq 0$. Then $z_k \to 0$ and $f(z_k) \to 0$ imply $g(z_k) \to a_n(0) \neq 0$; for a sequence $z_k' \to 0$ such that $f(z_k')$ converges to a zero of $w^n + a_1(0)w^{n-1} + \ldots + a_n(0)$, we have $g(z_k') \to 0$. Hence 0 is an essential singularity of g. The case of meromorphic coefficients can be reduced to the above by multiplication with a suitable power of $z - z_0$.

III.2.2a) Write $q(z) = z^n\big(1 - q_*(z)\big)$. On $]0, +\infty[$, q_* strictly decreases from $+\infty$ to 0, hence there exists a unique $r > 0$ with $q_*(r) = 1$, i.e. $q(r) = 0$.

III.2.2b) $|z_0| \leqslant r$ follows from

$$0 = p(z_0) \geqslant |z_0|^n - \sum_0^{n-1} |a_\nu||z_0|^\nu = q(|z_0|).$$

III.3.3a) The integral formula for a_n yields $|a_n| \leqslant r^{-n} M_r = (1/r) r^{-n+1} M_r$. Hence $r^{-n+1} M_r \to \infty$ if $a_n \neq 0$.

III.3.3b) $r^{-n} M_r \to \infty$ for $r \to \infty$ implies $\log M_r / \log r > n$ for large r. Hence, if f is transcendental, $\log M_r / \log r \to \infty$. For f a polynomial, $\log M_r / \log r \to \deg f$.

III.3.5a) Writing $z = re^{it}$, we have $|\exp z^2| = \exp(r^2 \cos 2t)$. If $|t| \leqslant \alpha < \pi/4$, then $\cos 2t \geqslant \cos 2\alpha > 0$, hence $|\exp z^2| \geqslant \exp(r^2 \cos 2\alpha) \to \infty$. If $|t - \pi/2| \leqslant \alpha < \pi/4$, then $\cos 2t \leqslant -\cos 2\alpha < 0$, hence $|\exp z^2| \leqslant \exp(-r^2 \cos 2\alpha) \to 0$.

III.4.3. Let f have a double pole. Find $T \in \mathcal{M}$ such that $f_1 = fT$ has a double pole at ∞, i.e. f_1 is a polynomial of deg 2. Find S_1, S_2 such that $S_2 f_1 S_1(z) = z^2$. Let f have two simple poles. Find $T \in \mathcal{M}$ such that $f_1 = fT(z) = az + b + c/z$. Then find S_1, S_2 such that $S_2 f_1 S_1(z) = z + 1/z$.

III.5.2a) $2\,\mathrm{Log}(1-z)$ is a holomorphic logarithm of $(1-z)^2$ on $\mathbb{C} \setminus [1, +\infty[$. – Choose the principal branch of the square root. Then $z + \sqrt{z}$ maps the slit plane $\mathbb{C} \setminus \mathbb{R}_{\leqslant 0}$ into itself.

III.5.3. On S_0 we have $f(z) = 1 - \cos z = z^2 \tilde{f}_0(z)$, with $\tilde{f}_0$ without zeros. S_0 being convex, $\tilde{f}_0$ has a holomorphic square root $\tilde{g}_0$, without loss of generality $\tilde{g}_0 > 0$ on $S_0 \cap \mathbb{R}$. Set $g_0(z) = z\tilde{g}_0(z)$. Similarly, $f(z) = (z - 2\pi)^2 \tilde{g}_1(z)^2$ on $S_1 = \{|\mathrm{Re}\, z - 2\pi| < 2\pi\}$, $\tilde{g}_1 > 0$ on $S_1 \cap \mathbb{R}$. Set $g_1(z) = (z - 2\pi)\tilde{g}_1(z)$. Then $g_0 = g_1$ on $S_0 \cap S_1 \cap \mathbb{R}$, hence on $S_0 \cap S_1$, i.e. they define a holomorphic square root of f on $S_0 \cup S_1$. Continue like this.

III.5.5. The mapping properties of the sine function become clearer by writing $\sin z = h(g(z))$ with $\zeta = g(z) = e^{iz}$ and $h(\zeta) = \frac{1}{2i}(\zeta + 1/\zeta)$. Hence

$$\arcsin w = \frac{1}{i}\operatorname{Log}(iw + \sqrt{1-w^2}),$$

where Log is the principal branch, and the chosen branch of $\sqrt{1-w^2}$ on G_1 has value 1 at $w = 0$.

III.6.2. The first formula results from

$$1 = \frac{e^z - 1}{z} \cdot f(z) = \Big(\sum_{\nu=0}^{\infty} \frac{1}{(\nu+1)!} z^{\nu}\Big)\Big(\sum_{\mu=0}^{\infty} \frac{B_\mu}{\mu!} z^{\mu}\Big),$$

the second from

$$\frac{e^z - 1}{z} \cdot \sum_{\mu=1}^{\infty} \frac{B_{2\mu}}{(2\mu)!} z^{2\mu} = 1 - \frac{e^z - 1}{z}\Big(1 - \frac{z}{2}\Big).$$

III.6.4a) By the first formula in Ex. III.6.3,

$$\frac{\pi z}{\sin \pi z} = 1 + 2\sum_{1}^{\infty}(1 - 2^{1-2\mu})\zeta(2\mu) z^{2\mu}.$$

III.6.4b) $\tan z = \sum\limits_{1}^{\infty}(-1)^{\mu-1} 2^{2\mu}(2^{2\mu} - 1)\frac{B_{2\mu}}{(2\mu)!} z^{2\mu-1}$.

III.7.2. $e^{2\pi z} - 1 = 2\pi e^{\pi z} \prod\limits_{n\neq 0}\big(1 - \frac{z}{in}\big) e^{z/in}$.

IV.1.3. Let $F\colon G \to G'$ be biholomorphic, G simply connected. For a cycle $\Gamma' = \sum n_\varrho \gamma_\varrho$ in G' set $\Gamma = \sum n_\varrho (F^{-1} \circ \gamma_\varrho)$. Then

$$\int_{\Gamma'} f(w)\, dw = \int_{\Gamma} f\big(F(z)\big) F'(z)\, dz = 0$$

for every $f \in \mathcal{O}(G')$. Apply this to $f(w) = 1/(w - w_0)$, $w_0 \notin G'$. – The function $F(z) = z^3$ maps $\mathbb{H}$ onto $\mathbb{C}^*$.

IV.3.1. The sets of convergence are

$$\{z : 1/2 < |z| < 2\},\ \{z : 1 < |z-1| < 3\},\ \{z : |z| = 1\},\ \varnothing.$$

IV.3.2. The principal part of $z(z^2 + b^2)^{-2}$ at ib is $\frac{1}{4ib}(z - ib)^{-2}$.

IV.4.3. If γ is a closed path in $G \setminus S$, then $\int\limits_{\gamma} f(z)\, dz = 2\pi i \sum\limits_{z \in G} n(\gamma, z) \operatorname{res}_z f = 0$. Apply Prop. II.1.2.

IV.5.2. Let z^α denote the principal branch. Then $f(z) = z^\alpha/(1 + z^n)$ is holomorphic in a neighbourhood of the sector save for a simple pole at $z_0 = e^{\pi i/n}$ with residue $-{z_0}^{\alpha+1}/n$. The integral of f over $[0, r{z_0}^2]$ is ${z_0}^{2(\alpha+1)} \int\limits_0^r f(z)\, dz$, the integral over the circular arc tends to 0 for $r \to \infty$ (standard estimate), hence $(1 - {z_0}^{2(\alpha+1)}) \int\limits_0^\infty f(z)\, dz = -2\pi i\, {z_0}^{\alpha+1}/n$, i.e.

$$\int_0^\infty f(z)\, dz = \frac{\pi/n}{\sin\big(\pi(\alpha+1)/n\big)}.$$

IV.5.3. a) $\pi(a - \sqrt{a^2 - 1})$, b) $2\pi/(1 - a^2)$ if $|a| < 1$, $2\pi/(a^2 - 1)$ if $|a| > 1$.

IV.5.4. a) $\pi/6$, c) $\pi/2\sqrt{2}$.

IV.5.5b) $I(a) = \pi \exp(-|a|)$ fails to be differentiable at $a = 0$.

IV.5.7. The principal value of the first integral is πt, the second integral is one half the real part of the first one. The third can be reduced to the second by $2\sin^2 x = 1 - \cos 2x$.

IV.5.9. Integrate $R(z)z^\alpha \log z$ over the path γ described in the text.

IV.6.2. The formula for the inverse function: If $f(z) = w$, w fixed, $z \notin \operatorname{Tr}\Gamma$, then

$$\frac{1}{2\pi i}\int_\Gamma \frac{\zeta f'(\zeta)}{f(\zeta) - w}\,d\zeta = n(\Gamma,\, z)\, z.$$

For $\Gamma = \partial D_r(z_0)$ with $D_r(z_0) \subset\subset G$, one obtains for $w \in f\big(D_r(z_0)\big)$

$$f^{-1}(w) = \frac{1}{2\pi i}\int_{|\zeta - z_0| = r} \frac{\zeta f'(\zeta)}{f(\zeta) - w}\,d\zeta.$$

IV.6.4. The number of zeros is 3 resp. 4.

V.1.3. Compare with a suitable definite integral of $1/t$.

V.2.2. To show $\int_1^\infty (x - [x])x^{-2}\,dx = 1 - \gamma$, compute

$$\int_1^{k+1} (x - [x])x^{-2}\,dx = \sum_1^k \int_n^{n+1} (x - n)x^{-2}\,dx = \log(k+1) - \sum_1^{k+1} 1/n + 1.$$

V.2.4. $\mathcal{M}_\psi(s) = \sum \Lambda(n)n^{-s}$ follows by Ex. 3. On the other hand,

$$-\zeta'(s)/\zeta(s) = \sum_p \frac{p^{-s}}{1 - p^{-s}} \log p = \sum_p \sum_{m=1}^\infty \log p (p^{-s})^m = \sum_1^\infty \Lambda(n)n^{-s}.$$

V.3.2. Corollary 3.5 implies $\xi(1-s) = \xi(s)$. As $\zeta(s)$ has only one simple pole at $s = 1$ and $\Gamma(s)$ has (simple) poles at $0, -1, -2, \dots$, $\xi(s)$ can have (simple) poles at most at $s = 1$ and $s = 0, -2, -4, \dots$. But in these points either $s(s-1)$ or $\zeta(s)$ vanish. Hence $\xi(s)$ is an entire function.

V.4.3. By construction, $|\tau| \geqslant 1$. If $\operatorname{Re}\tau > \frac{1}{2}$, then $|\omega_2 - \omega_1| < |\omega_2|$ and if $\operatorname{Re}\tau < \frac{1}{2}$, then $|\omega_2 + \omega_1| < |\omega_2|$, contradicting the choice of ω_2. If $\operatorname{Re}\tau = \frac{1}{2}$, replace (ω_1, ω_2) by $(\omega_1, \omega_2 - \omega_1)$; if $|\tau| = 1$, $\operatorname{Re}\tau > 0$, replace (ω_1, ω_2) by $(-\omega_2, \omega_1)$.

V.4.4. $f(z) = \big(a_0 + a_1\wp(z)\big)/\big(\wp(z) - \wp(z_1)\big)$.

V.4.6. The recursion formula is

$$(k-1)(2k+5)c_{2(k+1)} = 3\sum_{\ell+m=k} c_{2\ell}c_{2m} \qquad \text{for } k \geqslant 2.$$

V.4.7. *(i)* $\Rightarrow$ *(ii)* by $(*)$; *(iii)* $\Rightarrow$ *(iv)*: the poles $\neq 0$ of $\wp$ occur in pairs of complex conjugates; *(iv)* $\Rightarrow$ *(i)*: the series defining g_2 and g_3 are invariant under complex conjugation.

V.4.8. Let $\omega_1 \in \Omega \cap \mathbb{R}$ the minimal positive number and $\omega_2 \in \Omega \cap i\mathbb{R}$ with minimal positive imaginary part. If the parallelogram spanned by ω_1, ω_2 contains no $\omega \in \Omega$ in its interior, then $\Omega = \langle\omega_1, \omega_2\rangle$. In the other case, $\omega = \frac{1}{2}(\omega_1 + \omega_2)$ and $\Omega = \langle\omega, \overline{\omega}\rangle$.

V.4.11. By $(*)$, the Laurent series of $\wp_\Omega$ and $\wp_{\Omega'}$ at 0 coincide, hence $\wp_\Omega = \wp_{\Omega'}$. But $\Omega = \{\text{poles of } \wp_\Omega\}$, $\Omega' = \{\text{poles of } \wp_{\Omega'}\}$.

V.4.12. Both sides have simple poles at $z = 0$ and $z = \pm u$ (if $2u \notin \Omega$) with the same residues, hence their difference is a constant. But at $z = \omega_1/2$, both sides vanish.

V.4.14. If $2u \notin \Omega$, both sides are elliptic functions of z, with only one pole at $z = 0$, where the principal part is $1/z^2$. And both sides vanish at $z = \pm u$.

V.5.1. Differentiate the addition theorem of $\wp$ with respect to z_1 and z_2, then add and simplify.

VII.2.1. The cross ratio as well as $\delta(z_1, z_2)$ are $\operatorname{Aut}\mathbb{D}$-invariant. Hence we may assume $z_0 = 1$, $z_1 = 0$, $z_2 = s \in]0, 1[$, $z_3 = -1$. Then $\mathrm{CR}(z_0, z_1, z_2, z_3) = (1+s)/(1-s) > 1$.

VII.2.2. Corollary 2.5 now reads: If $g\colon \mathbb{H} \to \mathbb{H}$ is holomorphic, then $|g'(z)| \leqslant |\operatorname{Im} g(z)/\operatorname{Im} z|$.

VII.2.5. It suffices to check the equation for $z = i$, $w = \lambda i$ ($\lambda > 1$), as both sides are $\operatorname{Aut}\mathbb{H}$-invariant.

VII.3.3. We work in $\mathbb{H}$ and take $g = i\mathbb{R}_{>0}$. Let z_1, z_2 be the points on $|z| = 1$ with $\delta_{\mathbb{H}}(z_1, i) = d = \delta_{\mathbb{H}}(i, z_2)$. As the homotheties $z \mapsto \lambda z$ ($\lambda > 0$) are in $\operatorname{Aut}\mathbb{H}$, the set $\{z \in \mathbb{H} : \delta_{\mathbb{H}}(z, g) = d\}$ consists of the euclidean rays from 0 through z_1 resp. z_2. The general case follows by application of $T \in \operatorname{Aut}\mathbb{H}$: $\{z \in \mathbb{H} : \delta_{\mathbb{H}}(z, g) = d\}$ consists of two euclidean circular arcs passing through the points at infinity of g.

VII.3.4. Any two triples of distinct points on $\partial\mathbb{D}$ can be mapped onto each other by a $T \in \operatorname{Aut}\mathbb{D}$.

VII.3.5. Example *ii* for $\delta_{\mathbb{H}}$ in VII.2 helps for the h-triangle in $\mathbb{H}$ with vertices i, z_0, ∞.

VII.3.6. By the second law of cosines, corresponding edges have the same length. The construction of a h-triangle from given edges shows uniqueness up to isometry.

VII.3.7. Use both laws of cosines.

VII.3.8. Work in $\mathbb{D}$ and show that $D_\lambda T_{z_0}$ is a product of two reflections. Write $T_{z_0} = \sigma_{g_1}\sigma_{g_2}$ as in the proof of Prop. 3.7. Then g_1 passes through 0 and $D_\lambda = (D_{\lambda/2}\sigma_{g_1}D_{-\lambda/2})\sigma_{g_1} =: \sigma_{g_0}\sigma_{g_1}$, hence $D_\lambda T_{z_0} = \sigma_{g_0}\sigma_{g_2}$.

VII.5.3. Let z_1 be any point of G with $u_n(z_1) \to c < +\infty$, and $D = D_R(z_1) \subset\subset G$. By Ex. 1, for $n > m$ and $|z - z_1| \leqslant r < R$, we have $0 \leqslant u_n(z) - u_m(z) \leqslant \dfrac{R+r}{R-r}\big(u_n(z_1) - u_m(z_1)\big)$, hence u_n converges locally uniformly on D to a function that is harmonic by Ex. 2. Similarly, if $u_n(z_2) \to +\infty$, then $u_n \to +\infty$ on a neighbourhood of z_2. Now use that G is connected.

VII.6.1. Extend f by reflection in $\{|z| = r_2\}$ and $\{|z| = r_1\}$ to a larger annulus, which is then mapped conformally to an annulus containing K_2, etc.

VII.7.1. $\tau_1(z) = z + 2$, $\tau_2(z) = z/(1 - 2z)$.

VII.7.2. By construction, the $\tau\overline{\Delta_0}$ and $\tau\sigma_3\overline{\Delta_0}$ ($\tau \in \Gamma_0$) cover $\mathbb{H}$. Hence, for $z \in \mathbb{H}$, there is $\tau \in \Gamma_0$ such that $\tau z = z_1 \in \overline{\mathcal{F}}$. If $\operatorname{Re} z_1 = 1$, then ${\tau_1}^{-1} z_1 \in \mathcal{F}$; if $|z_1 - \frac{1}{2}| = 1$, then $\tau_2 z_1 \in \mathcal{F}$. As λ is Γ_0-invariant and maps $\mathcal{F}$ injectively, $z_1, z_2 \in \mathcal{F}$ and $z_2 = \tau z_1$, $\tau \in \Gamma_0$, imply $\tau = \mathrm{id}$.

VII.7.4. E.g. $\lambda(-1/z)$ and $1 - \lambda(z)$ both map $\overline{\Delta_0}$ onto $\overline{\mathbb{H}^-}$ with $(0, 1, \infty) \mapsto (0, \infty, 1)$.

VII.7.7. Note that any $f \in \operatorname{Aut}\big(\widehat{\mathbb{C}} \setminus \{0, 1, \infty\}\big)$ can be extended to an $\widehat{f} \in \operatorname{Aut}\widehat{\mathbb{C}}$ which permutes $0, 1, \infty$.

VII.7.8. $\lambda(i) = 1/2$, $\lambda\big((1+i)/2\big) = 2$, $\lambda(1+i) = -1$; $\lambda\big(e^{2\pi i/3}\big) = -e^{2\pi i/3}$, $\lambda\big(e^{\pi i/3}\big) = e^{\pi i/3}$.

Example: Apply a functional equation of λ to $i = (-1)/i$.

Bibliography

[A] AHLFORS, L. V.: *Complex Analysis* 3^{rd} ed. McGraw-Hill (New York) 1979

[Be] BEARDON, A. F.: *The Geometry of Discrete Groups* Springer (New York, Heidelberg, Berlin) 1983

[Bel] BELTRAMI, E.: *Saggio di interpretazione della geometria non euclidea* Giorn. di Mat. **6**, 284–312 (1868)

[Bo] BOCHNER, S.: *Analytic and meromorphic continuation by means of Green's formula.* Ann. math. **44**, 652–673 (1943)

[Ca] CARATHÉODORY, C.: *Funktionentheorie* 2 Vol. Birkhäuser (Basel) 1950

[Car] CARTAN, H.: *Sur le théorème de préparation de Weierstrass.* AG Forschung des Landes NRW. Bd. 33, 155–168 (1966) ≡ Coll. Works Vol II, 875–888. Springer (Berlin) 1979

[D] DIXON, J. D.: *A Brief Proof of Cauchy's Integral Theorem.* Proc. Am. Math. Soc. **29**, 625–626 (1971)

[Eh] EHRENPREIS, L.: *A new proof and extension of Hartogs' theorem.* Bull. AMS **67**, 507–509 (1961)

[FL] FISCHER, W./LIEB, I.: *Einführung in die Komplexe Analysis* Vieweg+Teubner (Wiesbaden) 2010

[FL1] FISCHER, W./LIEB, I.: *Funktionentheorie* 9^{th} ed. Vieweg+Teubner (Wiesbaden) 2005

[FL2] FISCHER, W./LIEB, I.: *Ausgewählte Kapitel aus der Funktionentheorie.* Vieweg (Braunschweig) 1988

[Fo1] FORSTER, O.: *Lokale analytische Geometrie.* Vorlsg. Münster 1975/76

[Fo2] FORSTER, O.: *Lectures on Riemann surfaces.* Springer (New York) 1981

[FG] FRITZSCHE, K./GRAUERT, H.: *From holomorphic functions to complex manifolds* Springer (Berlin) 2002

[GR] GRAUERT, H./REMMERT, R.: *Theory of Stein spaces.* Springer (Berlin) 2004

[Ha] HARTOGS, F.: *Einige Folgerungen aus der Cauchyschen Integralformel bei Funktionen mehrerer komplexer Veränderlicher.* Münchner Berichte **36**, 223–242 (1906)

[Hö] HÖRMANDER, L.: *Complex analysis in several variables.* 2^{nd} ed. North Holland P.C. (Amsterdam) 1973

[HuC] HURWITZ, A./COURANT, R.: *Funktionentheorie* 4^{th} ed. Springer (Berlin) 1964

[Kn] KNESER, H.: *Funktionentheorie.* Vandenhoeck & Ruprecht (Göttingen) 1958

[Ko] KOREVAAR, J.: *On Newman's quick way to the prime number theorem.* The Mathematical Intelligencer, **4**, 108–115 (1982)

[La] LANG, S.: *Introduction to Complex Hyperbolic Spaces.* Springer (New York) 1987

[Le] LENZ, H.: *Nichteuklidische Geometrie.* Bibliographisches Institut (Mannheim) 1967

[Li] LIEB, I.: *Das Levische Problem.* Bonner math. Schr. **387**, 1–34 (2007)

[M] MILNOR, J.: *Hyperbolic Geometry: The first* 150 *years.* Bull. Am. Math. Soc. **6**, 9–24 (1982)

W. Fischer, I. Lieb, *A Course in Complex Analysis*, DOI 10.1007/978-3-8348-8661-3,

[N] Narasimhan, R./Nievergelt, Y.: *Complex Analysis in One Variable* 2nd ed. Birkhäuser (Boston) 2001

[Ne] Newman, D. J.: *Simple analytic proof of the prime number theorem.* Am. Math. Monthly **87**, 693–696 (1980)

[Ok] Oka, K.: *Collected Papers.* Springer (Berlin) 1984

[Ra] Range, R. M.: *Holomorphic functions and integral representations in several complex variables* 2nd ed. Springer (Berlin) 1998

[RS] Remmert, R./Schumacher, G.: *Funktionentheorie* 2 Vol. Springer (Berlin) 2006/2007

[RSt] Remmert, R./Stein, K.: *Eigentliche holomorphe Abbildungen.* Math. Zs. **73**, 159–189 (1960)

[Ri] Riemann, B.: *Über die Anzahl der Primzahlen unter einer gegebenen Größe.* Collected Works, Dover (New York), 145–153 (1953)

[Ru] Rudin, W.: *Principles of Mathematical Analysis.* McGraw-Hill (New York) 1976

[Ru1] Rudin, W.: *Real and Complex Analysis* 2nd ed. McGraw-Hill (New York) 1974

[Si] Siegel, C. L.: *Meromorphe Funktionen auf kompakten analytischen Mannigfaltigkeiten.* Nr. Ak. Wiss. Göttingen 1955, 71–77 ≡ Ges. Abh. III, 216–222 Springer (Berlin) 1966

[St] Stickelberger, L.: *Über einen Satz des Herrn Noether.* Math. Ann. **30**, 401–409 (1887)

[Sto] Stolz, O.: *Grundzüge der Differential- und Integralrechnung.* Teubner (Leipzig) 1893

Index of Symbols

W. Fischer, I. Lieb, *A Course in Complex Analysis*, DOI 10.1007/978-3-8348-8661-3,

Index

W. Fischer, I. Lieb, *A Course in Complex Analysis*, DOI 10.1007/978-3-8348-8661-3,