Chapter 2

Complex Functions

The functions we shall be exploring in this book are complex-valued functions of a single complex variable. When we speak of *complex functions*, we do not necessarily mean that these functions are analytic, although emphasis will be placed on this special and extremely important class.

We begin with a condensed introduction to the arithmetic and geometry of the complex number system. This is not intended as a full course on the subject; we shall give only a brief summary of some fundamental facts and fix some notation. Readers who have never worked with complex numbers before are advised to supplement the study of this text by consulting one of the more comprehensive textbooks, like Krantz [19], Lin [35], Marsden and Hoffmann[41], Needham [44], Palka [52], to mention a few.

The second section reviews some basic material about functions in general, with emphasis on complex functions. This topic is continued in Section 2.3, where the focus moves to geometric aspects of complex functions as mappings between (subsets of) two copies of the complex plane. In particular we investigate the arithmetic operations as geometric transformations.

In Section 2.4 we briefly discuss a classical approach for visualizing complex functions: the analytic landscape or module surface. Although traditionally it was used to depict only the modulus of a function as a graph, the missing information about the argument can be conveniently incorporated by coloring this surface.

In Section 2.5 we lay the groundwork for the central topic of this book: visualization and exploration of complex functions by color representations. We begin by briefly touching upon the method of domain coloring, which depicts a function on its domain by a color representation of its values. A similar idea is utilized in *phase portraits*, which we introduce and discuss next. In contrast to conventional domain coloring, these images represent the color-coded phase (argument) of a function only and neglect its modulus. Nevertheless, the phase portrait contains enough information to reconstruct an *analytic function* uniquely up to a scaling factor.

E. Wegert, *Visual Complex Functions: An Introduction with Phase Portraits*,
DOI 10.1007/978-3-0348-0180-5_2, © Springer Basel 2012

Quite a number of properties of a function can immediately be read off from its phase portrait. Modifications of the underlying color scheme allow us to visualize additional information, which further improves the readability. The section ends with a discussion about the construction of phase portraits on the Riemann sphere by stereographic projection. This part can be skipped on a first reading.

In Section 2.6 we briefly consider sequences of complex numbers and study their convergence in the complex plane and on the Riemann sphere. We also review some elementary properties of continuous functions.

Finally, in Section 2.7, some relevant facts from plane geometry are summarized. Special attention is paid to the concepts of homotopic paths and winding numbers.

We will assume that the reader is familiar with the fundamental facts of arithmetic, linear algebra, real analysis, and plane topology. Readers with some background in complex numbers may skip the first three sections and go directly to Section 2.4.

All notation will be explained where it appears for the first time, here we just mention that $\mathbb{N}$, $\mathbb{Z}_+$, $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{R}_+$ denote the sets of natural numbers (including zero), positive integers, integers, real numbers, and positive real numbers, respectively.

2.1 Complex Numbers

The Euclidean Plane. Our starting point is $\mathbb{R}^2$, the set of ordered pairs (x, y) of real numbers. Interpreting x and y as Cartesian coordinates of a point in the plane, $\mathbb{R}^2$ is said to be the *Euclidean plane*. We shall therefore identify points in the plane and pairs (x, y) of real numbers.

Arithmetic Operations. Next we introduce *arithmetic operations* with points in $\mathbb{R}^2$. The *sum* and the *product* of (x_1, y_1) and (x_2, y_2) are defined as follows:[1]

$$(x_1, y_1) + (x_2, y_2) := (x_1 + x_2, y_1 + y_2), \tag{2.1}$$

$$(x_1, y_1) \cdot (x_2, y_2) := (x_1 x_2 - y_1 y_2, x_1 y_2 + y_1 x_2). \tag{2.2}$$

While the rule for addition is familiar from the corresponding vector operations, the multiplication rule looks a bit odd. It was the Irish mathematician Rowan William Hamilton who discovered that the above definition of multiplication is the only one which is compatible with (vector) addition. Here "compatible" means that both operations together satisfy the ordinary arithmetic rules known from the real numbers, namely the associative, commutative and distributive laws. We shall not verify them here, but readers who are seeing this definition for the first time are encouraged to check them.

[1] In (2.1), (2.2), and throughout the text, the sign $:=$ indicates a definition: the object on the left-hand side is defined by the expression on the right.

The pairs $(0,0)$ and $(1,0)$ are the *neutral elements* of addition and multiplication, respectively, i.e., for all $(x,y) \in \mathbb{R}^2$

$$(x,y) + (0,0) = (x,y), \qquad (x,y) \cdot (1,0) = (x,y).$$

Further, the pairs of type $(x,0)$ obey the simple arithmetic rules

$$(x,0) + (y,0) = (x+y,0), \qquad (x,0) \cdot (y,0) = (xy,0),$$

which allows one to identify $(x,0)$ with the corresponding real number x, i.e., $x = (x,0)$. In particular the pairs $(0,0)$ and $(1,0)$ represent the real numbers 0 and 1, respectively.

The Imaginary Unit. Besides the two special numbers zero and one, the third distinguished element is the *imaginary unit* $\mathrm{i} := (0,1)$. This extraordinary object satisfies the identity

$$\mathrm{i}^2 := \mathrm{i} \cdot \mathrm{i} = (0,1) \cdot (0,1) = (-1,0) = -1,$$

which demonstrates that the square of a complex number can be negative. The idea of denoting the imaginary unit by the symbol i dates back to a paper by Leonhard Euler in 1733.

Using $x = (x,0)$, $y = (y,0)$, and the definition of multiplication, we obtain $(x,y) = (x,0) + (0,1) \cdot (y,0) = x + \mathrm{i} \cdot y$, which is the *complex representation* of the pair (x,y). Omitting the dot denoting multiplication we obtain the conventional form $x + \mathrm{i}\,y$ of writing *complex numbers*. Usually complex numbers are denoted by a single letter, as in

$$z := x + \mathrm{i}\,y.$$

Because the product is commutative, there is no difference in representing a complex number as $x + \mathrm{i}\,y$ or $x + y\,\mathrm{i}$.

The Complex Plane. Any complex number $z = x + \mathrm{i}\,y$ with $x, y \in \mathbb{R}$ can be identified with the point in the Euclidean plane $\mathbb{R}^2$ having Cartesian coordinates x and y. In this context we refer to $\mathbb{R}^2$ as the *complex plane*, or the *Gaussian plane*, and denote it by $\mathbb{C}$.

The coordinates x and y are said to be the *real part* and the *imaginary part* of z, respectively, and written as $x = \operatorname{Re} z$ and $y = \operatorname{Im} z$. Note that the imaginary part of $x + \mathrm{i}\,y$ is the *real* number y, and *not* $\mathrm{i}y$ as its name might suggest.

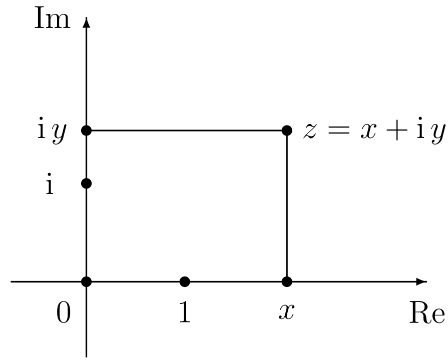

Figure 2.2: The complex plane

We point out that $z = x + \mathrm{i}\, y$ does not necessarily imply that x is the real part and y is the imaginary part of z, this only follows if x and y are real. However, here and also later we shall often automatically assume that x and y are both real without explicit mention.

Complex numbers with $\operatorname{Re} z = 0$ are called *purely imaginary*, the *imaginary axis* is the set of all purely imaginary numbers, the *real line* $\mathbb{R}$ is identified with the *real axis* consisting of all complex numbers with vanishing imaginary part.

The unfortunate attribute "imaginary" enshrouds complex numbers with some kind of mystery. In fact, in the construction of number systems, the transition from real to complex numbers is only a small step which, as we have seen, causes no serious problems once the right approach is found. On the other hand, even after a long time of historical development, the foundation of the real number system still remains rather technical. So the *real numbers* are actually the nontrivial objects which still have their deep secrets. Given real numbers, complex numbers are just pairs of reals, endowed with some additional arithmetic.

More Arithmetic. Indeed, the arithmetic rules are the reason for writing complex numbers in their usual form $x + \mathrm{i}\, y$, and not as pairs (x, y). Remembering that the imaginary unit exhibits the fundamental property $\mathrm{i}^2 = -1$ is all what one needs to compute the sum and the product of complex numbers: just treat i as a variable, apply the arithmetic rules known from the reals, and simplify the result so that it gets the form $x + \mathrm{i}\, y$.

In the next step *subtraction* and *division* are introduced as the *reverse operations* of addition and multiplication, respectively. The *difference* $z_1 - z_2$ of $z_1 = x_1 + \mathrm{i}\, y_1$ and $z_2 = x_2 + \mathrm{i}\, y_2$ is the unique number z which solves the equation $z_2 + z = z_1$, i.e.,

$$z = (x_1 + \mathrm{i}\, y_1) - (x_2 + \mathrm{i}\, y_2) := (x_1 - x_2) + \mathrm{i}\,(y_1 - y_2).$$

Similarly, if $z_2 \neq 0$, the *quotient* z_1/z_2 of z_1 and z_2 is the solution z of $z_2 \cdot z = z_1$. It can easily be verified that this solution is unique, namely

$$z = \frac{x_1 + \mathrm{i}\, y_1}{x_2 + \mathrm{i}\, y_2} := \frac{x_1 x_2 + y_1 y_2}{x_2^2 + y_2^2} + \mathrm{i}\, \frac{y_1 x_2 - x_1 y_2}{x_2^2 + y_2^2}. \tag{2.3}$$

The number $-z := 0 - z$ is termed the *negative* of z, and the *inverse* of $z \neq 0$ is defined by $z^{-1} := 1/z$.

There is a simple trick to remember the division formula (2.3). It utilizes the *conjugate* $\overline{z}$ of a complex number $z = x + \mathrm{i}\, y$, defined by

$$\overline{z} := x - \mathrm{i}\, y.$$

Since $z\,\overline{z} = x^2 + y^2$ is real, multiplying the numerator and denominator of z_1/z_2 by $\overline{z}_2$ allows one to separate the real and imaginary parts of the quotient, which yields (2.3).

The following arithmetic rules for conjugation of complex numbers can easily be verified:

$$\overline{\overline{z}} = z, \quad \overline{z_1 \pm z_2} = \overline{z_1} \pm \overline{z_2}, \quad \overline{z_1\, z_2} = \overline{z_1}\,\overline{z_2}, \quad \overline{z_1/z_2} = \overline{z_1}/\overline{z_2}.$$

We also mention two useful formulas expressing the real and imaginary parts of z in terms of z and $\overline{z}$,

$$\operatorname{Re} z = \frac{z+\overline{z}}{2}, \qquad \operatorname{Im} z = \frac{z-\overline{z}}{2\mathrm{i}}.$$

Integer powers z^n are defined for positive n as the n-fold product of z with itself, and by $z^n := 1/z^{-n}$ if n is negative and $z \neq 0$. Finally, if $z \neq 0$ we set $z^0 := 1$.

Polar Representation. Like points in the plane, complex numbers have an alternative description in terms of *polar coordinates*. If $z = x + \mathrm{i}\,y$ is a complex number different from zero, the real and imaginary parts of z are given by the formulas

$$x = r\,\cos\varphi, \quad y = r\,\sin\varphi, \tag{2.4}$$

where r denotes the distance of the point z from the origin 0, and φ is the oriented angle between the positive real axis and the ray from 0 to z (see Figure 2.3). Inserting (2.4) in $z = x + \mathrm{i}\,y$ yields the *polar representation* of complex numbers

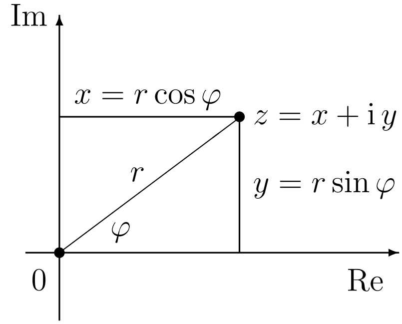

Figure 2.3: Polar representation

$$z = r\,(\cos\varphi + \mathrm{i}\,\sin\varphi). \tag{2.5}$$

If $z = 0$, then (2.5) holds with $r = 0$ and any value of φ. The value of r is called the *modulus* of z and denoted by $|z|$. The notions *absolute value* or *magnitude* are commonly used as synonyms for modulus. The modulus can be expressed in terms of the real and imaginary part using Pythagoras' theorem,

$$r = |z| = \sqrt{x^2 + y^2} = \sqrt{(\operatorname{Re} z)^2 + (\operatorname{Im} z)^2}.$$

For $z \neq 0$, each real number φ for which (2.5) holds is said to be an *argument* of z. The argument of $z = 0$ is undefined. *Polar angle* or just *angle* are alternative names for the argument of a complex number. Any argument φ of $z = x + \mathrm{i}\,y$ with $x \neq 0$ satisfies

$$\tan\varphi = y/x = \operatorname{Im} z / \operatorname{Re} z. \tag{2.6}$$

However, the *converse is not true* since $\tan\varphi = \tan(\varphi + \pi)$. Consequently the relation (2.6) alone is not sufficient to find an appropriate value of φ, one must take into account the location of z in the four quadrants.

Moreover, even (2.4) does not determine φ uniquely since addition of an integer multiple of 2π does not change the values of the sine and cosine functions.

Argument and Phase. In this text the notation $\arg z$ is used to designate an arbitrary argument of z, which means that $\arg z$ is a set rather than a number. In particular, the relation $\arg z_1 = \arg z_2$ is *not an equation*, but expresses equality of two sets.

As a consequence, two non-zero complex numbers $r_1 \big(\cos\varphi_1 + \mathrm{i}\,\sin\varphi_1\big)$ and $r_2 \big(\cos\varphi_2 + \mathrm{i}\,\sin\varphi_2\big)$ are equal if and only if

$$r_1 = r_2, \quad \text{and} \quad \varphi_1 - \varphi_2 \in 2\pi\,\mathbb{Z}, \tag{2.7}$$

the latter meaning that there exists an integer k such that $\varphi_1 = \varphi_2 + 2k\,\pi$.

In order to make the argument of z a well-defined number, it is sometimes restricted to the interval $(-\pi, \pi]$. This special choice is called the *principal value* or the *main branch* of the argument and is written as $\operatorname{Arg} z$. Note that there is no general convention about the definition of the principal value, sometimes its values are supposed to be in the interval $[0, 2\pi)$. This ambiguity is a perpetual source of misunderstandings and errors.

Moreover, restricting the argument to its principle value has serious disadvantages and we shall see applications where this is even impossible. One way out is to avoid the argument and to work with the *phase* $\psi(z) := z/|z|$ of z. We shall discuss this issue in more detail later.

Arithmetic Revisited. Let us now briefly reconsider the arithmetic operations in the polar representation of complex numbers. Suppose that z_1 and z_2 are given in the form $z_1 = r_1 \big(\cos\varphi_1 + \mathrm{i}\,\sin\varphi_1\big)$ and $z_2 = r_2 \big(\cos\varphi_2 + \mathrm{i}\,\sin\varphi_2\big)$. A short computation involving the addition theorems

$$\sin(\alpha+\beta) = \sin\alpha\cdot\cos\beta + \cos\alpha\cdot\sin\beta \tag{2.8}$$

$$\cos(\alpha+\beta) = \cos\alpha\cdot\cos\beta - \sin\alpha\cdot\sin\beta \tag{2.9}$$

yields

$$z_1\,z_2 = r_1\,r_2\,\big(\cos(\varphi_1+\varphi_2) + \mathrm{i}\,\sin(\varphi_1+\varphi_2)\big). \tag{2.10}$$

So the *modulus* of a product is the *product* of the moduli of its factors, while the *argument* of a product is the *sum* of the arguments of its factors (as long as one has the correct interpretation of the latter statement).

The absolute value of $z - a$ is the Euclidean distance between the points z and a, which is often used in representations of geometric objects. For example, a *disk* and a *circle* with center a and radius r can be described as

$$D_r(a) := \{z \in \mathbb{C} :\ |z-a| < r\}, \quad T_r(a) := \{z \in \mathbb{C} :\ |z-a| = r\},$$

respectively. The special notation $\mathbb{D}$ and $\mathbb{T}$ is reserved for the *unit disk* and the *unit circle*

$$\mathbb{D} := \{z \in \mathbb{C} :\ |z| < 1\}, \quad \mathbb{T} := \{z \in \mathbb{C} :\ |z| = 1\}.$$

The following arithmetic properties of the modulus can easily be checked,

$$|z_1 z_2| = |z_1|\cdot|z_2|, \quad |z_1/z_2| = |z_1|/|z_2|, \quad |\overline{z}| = |z|, \quad z\,\overline{z} = |z|^2. \tag{2.11}$$

Moreover, we have the estimates

$$\operatorname{Re} z \le |z|, \quad \operatorname{Im} z \le |z|, \quad |z_1 \pm z_2| \le |z_1| + |z_2|.$$

The first two relations follow immediately from the definition of the absolute value. Observing that $|z_1|$, $|z_2|$ and $|z_1 - z_2|$ are the lengths of the sides of the triangle with vertices $0, z_1$ and z_2 in the complex plane, the third inequality (with minus sign) is the *triangle inequality* of plane geometry.

Complex numbers with $|z| = 1$ are termed *unimodular*. In particular the phase $\psi(z) := z/|z|$ of $z \in \mathbb{C} \setminus \{0\}$ is unimodular. Since the sum of the arguments of two factors is an argument of their product, it follows from (2.11) that the phase is multiplicative: for $z_1, z_2 \in \mathbb{C}$ with $z_1, z_2 \ne 0$,

$$\psi(z_1 z_2) = \psi(z_1) \cdot \psi(z_2), \quad \psi(z_1/z_2) = \psi(z_1)/\psi(z_2). \tag{2.12}$$

In particular we have for $z \ne 0$

$$\psi(-z) = -\psi(z), \qquad \psi(1/z) = 1/\psi(z) = \overline{\psi(z)}. \tag{2.13}$$

Powers and Roots. Applying (2.10) repeatedly with one and the same factor z, we get the celebrated De Moivre's formula for the n-th *power* of $z = r\,\big(\cos\varphi + \mathrm{i}\,\sin\varphi\big)$,

$$z^n = r^n\,\big(\cos n\varphi + \mathrm{i}\,\sin n\varphi\big). \tag{2.14}$$

It is easily seen that this formula remains valid if n is a negative integer. In particular, the *inverse* z^{-1} of a complex number $z = r\,(\cos\varphi + \mathrm{i}\,\sin\varphi) \in \mathbb{C} \setminus \{0\}$ can be expressed as

$$z^{-1} = 1/z = r^{-1}\,\Big(\cos(-\varphi) + \mathrm{i}\,\sin(-\varphi)\Big) = r^{-1}\,(\cos\varphi - \mathrm{i}\,\sin\varphi).$$

One main advantage of complex numbers over the reals is that negative numbers have a square root. More generally, a number z is called a *n-th root* of w if it satisfies $z^n = w$. Writing both numbers in polar form,

$$z = r\,\big(\cos\varphi + \mathrm{i}\,\sin\varphi\big), \qquad w = R\,\big(\cos\Phi + \mathrm{i}\,\sin\Phi\big),$$

by (2.7), $z^n = w$ is equivalent to $r^n = R$ and $n\,\varphi = \Phi + 2k\,\pi$ with $k \in \mathbb{Z}$. Resolving this with respect to r and φ yields

$$r = \sqrt[n]{R}, \qquad \varphi = \frac{\Phi}{n} + \frac{2k\,\pi}{n}.$$

Here k is an arbitrary integer, but since adding a multiple of n to k does not change the value of z, it suffices to take the values $k = 0, 1, \ldots, n-1$. So any complex number w different from zero has exactly n roots of order n. In the complex plane, the n-th roots of a non-zero complex number form the vertices of a regular n-gon centered at the origin, as Figure 2.4 shows for the 7-th roots of 1.

If n is a natural number, the *n-th roots of unity* are the solutions of $z^n = 1$. Introducing the *primitive root* of order n,

$$\omega := \cos\frac{2\pi}{n} + \mathrm{i}\,\sin\frac{2\pi}{n},$$

the n-th roots of unity are just the powers $1 = \omega^0, \omega^1, \omega^2, \ldots, \omega^{n-1}$ of ω.

The Point at Infinity. Having solved the problem of extracting roots in the system of complex numbers, we we are still left with another annoying problem with real arithmetic – that of not being able to divide by zero. In order to resolve this, we extend the complex plane by an ideal element, the *point at infinity*, which is denoted by ∞ and is supposed to satisfy

$$\frac{1}{0} := \infty, \qquad \frac{1}{\infty} := 0.$$

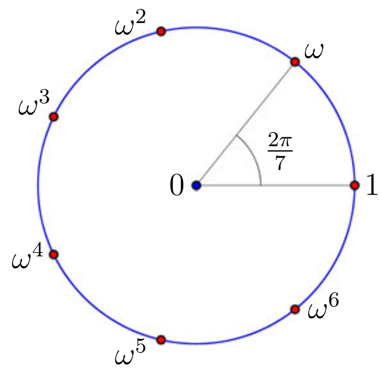

Figure 2.4: The seventh roots of unity

The name is motivated by the observation that the point $1/z$ moves further and further away (to "infinity") as z approaches zero (and vice versa). The union of the complex plane $\mathbb{C}$ with the point at infinity is called the *extended complex plane* and is denoted by $\widehat{\mathbb{C}}$.

The Riemann Sphere. The point at infinity is by no means a mysterious object. It has a simple and beautiful geometric interpretation if complex numbers are considered as points on a *sphere*. The story begins about 2000 years ago, when the Greek mathematician and philosopher Ptolemy invented a method for depicting points from a sphere (in particular from the "celestial sphere") on a flat map. In the 19th century, Bernhard Riemann proposed to utilize *stereographic projection* in the reverse direction for representing complex numbers.

In order to describe this construction, imagine that $\mathbb{C}$ is the XY-plane in three-dimensional XYZ space. To fix the meaning of "up" and "down" we assume that $\mathbb{C}$ is seen in its usual orientation if we look at it "downward from above" (see Figure 2.5).

Let $\mathbb{S}$ be a sphere with radius 1 centered at the common origin of $\mathbb{C}$ and $\mathbb{R}^3$. Taking recourse to concepts of geography, the unit circle $\mathbb{T}$ of the complex plane is the *equator* E of $\mathbb{S}$, and the two points on $\mathbb{S}$ at maximal distance from $\mathbb{C}$ are the *north pole* N (above $\mathbb{C}$) and the *south pole* S (below $\mathbb{C}$).

The *stereographic image* z of a point P on the sphere $\mathbb{S}$ is the intersection of $\mathbb{C}$ with the straight line through P and the north pole N.[2] The point z is well

[2]Note that there is an alternative definition, where the sphere "lies on top" of the complex plane, touching $\mathbb{C}$ with its south pole.

defined for all P on $\mathbb{S}$, with only one exception: the north pole N. If P approaches N, then the distance of the corresponding point z in the plane to the origin gets arbitrarily large. This observation shows that the north pole on $\mathbb{S}$ plays the same role as the point at infinity with respect to the complex plane.

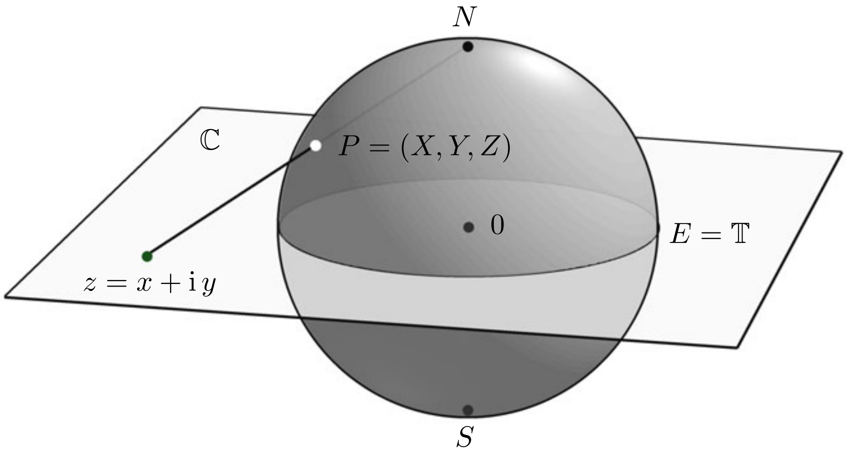

Figure 2.5: Stereographic projection of a sphere onto the complex plane

Extending the stereographic projection to all of $\mathbb{S}$ by assigning the north pole N to the point at infinity results in a bijective correspondence between $\mathbb{S}$ and $\widehat{\mathbb{C}}$. Hence we can label the points on $\mathbb{S}$ with the corresponding complex numbers of $\widehat{\mathbb{C}}$. In what follows we shall therefore identify the sphere $\mathbb{S}$ and the extended complex plane $\widehat{\mathbb{C}}$ and call it the *Riemann sphere*. The *spherical distance* $d(z_1, z_2)$ of two points in $\widehat{\mathbb{C}}$ is the Euclidean length of the straight segment connecting the corresponding points on the sphere $\mathbb{S}$. If $z_1, z_2 \in \mathbb{C}$ then

$$d(z_1, z_2) = \frac{2\,|z_1 - z_2|}{\sqrt{1+|z_1|^2}\sqrt{1+|z_2|^2}}, \qquad d(z_1, \infty) = \frac{2}{\sqrt{1+|z_1|^2}}.$$

Arithmetic on the Sphere. Stereographic projection allows us to transplant the arithmetic operations from $\mathbb{C}$ to $\widehat{\mathbb{C}}$. Additionally we postulate that

$$\begin{aligned}
&z/\infty := 0 && \text{for} \quad z \in \mathbb{C} \\
&z/0 := \infty && \text{for} \quad z \in \mathbb{C} \setminus \{0\} \\
&z \pm \infty = \infty \pm z := \infty && \text{for} \quad z \in \mathbb{C} \\
&z \cdot \infty = \infty \cdot z := \infty && \text{for} \quad z \in \widehat{\mathbb{C}} \setminus \{0\}.
\end{aligned} \tag{2.15}$$

Note that we do not define $\infty \pm \infty$, ∞/∞, $0/0$ and $0 \cdot \infty$. After extending modulus $|z|$ and phase $\psi(z)$ to all points of the Riemann sphere by setting

$$|\infty| := \infty, \quad \psi(0) := 0, \quad \psi(\infty) := \infty,$$

their ranges now become the *extended positive real line* and the *extended unit circle*, respectively, which we define by

$$\widehat{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{0, \infty\}, \qquad \widehat{\mathbb{T}} := \mathbb{T} \cup \{0, \infty\}.$$

Properties of Stereographic Projection. The following properties of stereographic projection are immediate:

(i) The points on the equator remain fixed.

(ii) The lower hemisphere is mapped to the unit disk $\mathbb{D}$.

(iii) The upper hemisphere is mapped to $\mathbb{E}$, the exterior of the unit circle.

Here the exterior of the unit circle is defined such that it includes the point at infinity,

$$\mathbb{E} := \{z \in \mathbb{C} : |z| > 1\} \cup \{\infty\}.$$

Stereographic projection has a remarkable property: it maps circles to circles. More precisely, the image of a circle on $\mathbb{S}$ through N is a line in $\widehat{\mathbb{C}}$, i.e., the union of a line in $\mathbb{C}$ with the point at infinity, while the images of all other circles are proper circles. So the above statement is literally true if we adopt the folkloristic aphorism "lines are circles with infinite radius".

Computing stereographic projections requires explicit formulas. For this end we represent points z in the complex plane by Cartesian coordinates x, y, and points P on the sphere $\mathbb{S}$ by Cartesian coordinates X, Y, Z in three-dimensional space. The systems are positioned such that the spatial X and Y axes coincide with the x and y axes in the complex plane. Then the coordinates (X, Y, Z) of a point P on $\mathbb{S} \setminus N$ and its stereographic projection $z = x + \mathrm{i}\, y$ are related by the equations

$$x = \frac{X}{1 - Z}, \quad y = \frac{Y}{1 - Z}. \tag{2.16}$$

The north pole $(X, Y, Z) = (0, 0, 1)$ corresponds to the point at infinity, which has no representation in the xy-system. Conversely, if $z = x + \mathrm{i}\, y \in \mathbb{C}$, then

$$X = \frac{2x}{x^2 + y^2 + 1}, \quad Y = \frac{2y}{x^2 + y^2 + 1}, \quad Z = \frac{x^2 + y^2 - 1}{x^2 + y^2 + 1}, \tag{2.17}$$

and $(X, Y, Z) = (0, 0, 1)$ for $z = \infty$.

In summary, we have four interpretations of a complex number $z = x + \mathrm{i}\, y$: as a pair (x, y) of real numbers, as a vector with components x and y, as a point in the Gaussian plane with Cartesian coordinates x and y, and as a point on the Riemann sphere with Cartesian coordinates X, Y, Z given by (2.17).

2.2 Functions and Mappings

A function is a "rule of correspondence" which assigns to each input element a well-defined output element. More precisely, if X and Y are arbitrary sets, a function f from X to Y assigns to any *argument* x in X exactly one *value* $f(x)$ in Y. This is expressed symbolically by writing

$$f : X \to Y,\ x \mapsto f(x).$$

The set X is said to be the *domain set* (or simply the *domain*), Y is called the *target set*.

Mappings or *transformations* are just synonyms for functions. Though there is no basic difference between these concepts, they emphasize different aspects of a function, and the context determines which name is preferred.

It is important to distinguish between a function and its values – while the symbol f stands for the (complete) function, $f(x)$ refers to its value at x. Having mentioned this, we shall nevertheless sometimes say "the function $f(x) = \sin x$", whenever this is convenient and causes no confusion. Occasionally we also write $f(\cdot)$, where the dot is a placeholder for the variable x.

Mapping Properties. The *image* $f(A)$ of a subset A of X is the set of values that f attains on A, and the *pre-image* of $B \subset Y$ is the set of all x in X which are mapped into B,

$$f(A) := \{f(x) \in Y : x \in A\}, \qquad f^{-1}(B) := \{x \in X : f(x) \in B\}.$$

The image $f(X)$ of the domain set is called the *range* of f. A function $f : X \to Y$ is said to be *surjective* if $f(X) = Y$, it is called *injective* (or *one-to-one*) if the equality $f(x_1) = f(x_2)$ only holds for $x_1 = x_2$. A function which is both injective and surjective is termed *bijective*. If f is bijective, there is a unique function $f^{-1} : Y \to X$, the *inverse* of f, which is defined by $f^{-1}(y) = x$ when $f(x) = y$.

A function $f : D \subset \mathbb{C} \to \widehat{\mathbb{C}}$ is said to be *periodic* if there exists a complex number $p \neq 0$ such that for all $z \in D$ also $z + p \in D$ and $f(z + p) = f(z)$. Any such p is called a *period* of f. A period is *primitive* if there is no integer $n > 1$ such that p/n is also a period.

Operations with Functions. Two functions f and g are *equal* if they have the same domain set X and $f(x) = g(x)$ for all $x \in X$. If f and g are defined on X and Y, respectively, with $X \subset Y$ and $f(x) = g(x)$ for all $x \in X$, then f is the *restriction* of g to X and g is an *extension* of f to Y. We shall often use the same symbol to denote a function and its restriction to some subset.

The *composition* of two functions $f : X \to Y$ and $g : Y \to Z$ is the mapping

$$g \circ f : X \to Z,\ z \mapsto g(f(z)).$$

In this book we mainly consider complex functions $f : D \subset \mathbb{C} \to \mathbb{C}$, where the domain set D is a subset of the complex plane $\mathbb{C}$ and the target set is the complex plane, but, more generally, we shall also encounter functions defined on (subsets of) the Riemann sphere $\widehat{\mathbb{C}}$ with values in $\widehat{\mathbb{C}}$. For such functions we denote points $z_0 \in \widehat{\mathbb{C}}$ where $f(z_0) = 0$ or $f(z_0) = \infty$ as *zeros* and *poles* of f, respectively.

If $f : D_f \to \mathbb{C}$ and $g : D_g \to \mathbb{C}$ are complex functions, their sum $f + g$ is defined on $D := D_f \cap D_g$ by $(f+g)(z) := f(z)+g(z)$. A similar definition is made for the difference $f - g$ and the product $f\,g$, while the quotient f/g is naturally defined only on $D \setminus \{z \in D_g : g(z) = 0\}$. If the target set includes the point at infinity, these definitions can be modified accordingly, taking into account the arithmetic rules (2.15).

Functions as Mappings. Figure 2.6 visualizes the action of a complex function as a mapping from a subset of the z-plane to the w-plane. The light yellow regions are the domain set and the range of the function, respectively. Any point z of the domain set is mapped to the corresponding point $f(z)$ in the range. In this manner the function maps ("transplants") the colored objects from the domain to the range.

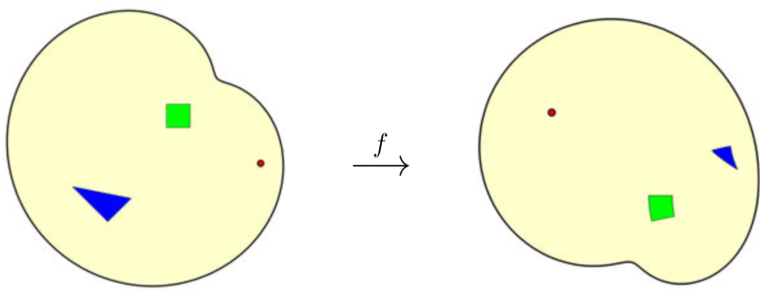

Figure 2.6: A complex function as a mapping which "transplants" sets

In complex analysis the notion of *domain* has two different meanings. The first one alludes to the domain set of a function, while the second pertains to any *open* and *connected* subset of the complex plane or the Riemann sphere (precise definitions of these notions will be given in Section 2.7). Most domain sets of complex functions we shall encounter in this book will indeed be domains in the topological sense.

Decompositions of Complex Functions. A complex function f is composed of its *real part* $u = \operatorname{Re} f$ and its *imaginary part* $v = \operatorname{Im} f$ as $f = u + \mathrm{i}\,v$. Note that $\operatorname{Re} f$ and $\operatorname{Im} f$ are real-valued functions, having the same domain as f. Similarly, f admits a multiplicative decomposition $f = |f| \cdot \psi(f)$ into its *modulus* $|f|$ and its *phase* $\psi(f)$. Unlike the argument $\arg f$, the phase of f is a well-defined function $\psi(f) : D \to \widehat{\mathbb{T}} := \mathbb{T} \cup \{0, \infty\}$.

By a slight abuse of notation, we consider functions f of a complex variable z also as a function of two real variables $x = \operatorname{Re} z$ and $y = \operatorname{Im} z$, thus writing, for example,

$$f(x + \mathrm{i}y) = f(x, y) = u(x, y) + \mathrm{i}\,v(x, y). \tag{2.18}$$

Conversely, if $f : D \subset \mathbb{R}^2 \to \mathbb{R}^2$ is a function of two variables x and y with two real-valued components u and v, then the right-hand side of (2.18) defines a complex-valued function of $z = x + \mathrm{i}\,y$. Using $x = (z + \overline{z})/2$ and $y = (z - \overline{z})/(2\mathrm{i})$, we can rewrite this in the form

$$f(z) = u\left(\frac{z+\overline{z}}{2}, \frac{z-\overline{z}}{2\mathrm{i}}\right) + \mathrm{i}\, v\left(\frac{z+\overline{z}}{2}, \frac{z-\overline{z}}{2\mathrm{i}}\right).$$

Since the right-hand side is an expression in z and $\overline{z}$, some authors prefer to write $f(z, \overline{z})$ instead of $f(z)$. We shall not use this notation since it suggests, rather inappropriately, that the variables $\overline{z}$ and z are independent.

2.3 Arithmetic and Geometry

As we have seen, complex functions can be interpreted as mappings or transformations of (subsets of) the complex plane. In this section we study the arithmetic operations within this framework. In contrast to Section 2.1, here we do not consider operations with two given numbers, instead we fix just one of them and let the second vary through the complex plane.

Addition and Multiplication. Denoting by a and b two given complex numbers with $b \neq 0$, we study the functions

$$f : \mathbb{C} \to \mathbb{C},\ z \mapsto z + a, \qquad g : \mathbb{C} \to \mathbb{C},\ z \mapsto b\, z, \tag{2.19}$$

generated by addition and multiplication. Interpreting a as a vector, it is immediate from the definition of addition that f is a *translation* (or shift) of the complex plane by the vector a. The mapping g can be more conveniently investigated using the polar representation of complex numbers. Here we distinguish several cases.

If b is a *positive number*, then g does not change the argument (or phase) of z, while the modulus of z is multiplied by b. Consequently, g is a *dilation*. If $b > 1$ it stretches the complex plane, and if $b < 1$ it shrinks it, such that each ray emanating from the origin is mapped onto itself and all distances are multiplied by b.

If b is *unimodular*, that is, if $|b| = 1$, then g does not change the modulus of z, while the argument of z is increased by $\beta := \arg b$ (here one can take any fixed argument of b). Thus g is a *rotation* of the plane about the origin by the angle β.

If $b \neq 0$, with $\beta = \arg b$ and $r = |b|$, then g is the composition of the mappings $z \mapsto (\cos\beta + \mathrm{i}\,\sin\beta)\, z$ and $z \mapsto r\, z$, i.e., a rotation by the angle β about the origin followed by a dilation with the same center and stretching factor r. Notice that performing these operations in reverse order yields the same result. We propose to call such transformations *rotostretch* (for the German “Drehstreckung”).

Note that f and g are orientation-preserving similarity transformations in the language of plane Euclidean geometry. In particular, the images of straight lines are straight lines and the images of circles are circles.

Division and Inversion. Finally, we investigate the quotient mapping $z \mapsto b/z$. Since it is the composition of $z \mapsto 1/z$ and $z \mapsto b\,z$, it suffices to study the transformation

$$h : \mathbb{C} \setminus \{0\} \to \mathbb{C} \setminus \{0\}, \; z \mapsto 1/z. \tag{2.20}$$

We first observe that complex conjugation $\mathbb{C} \to \mathbb{C}$, $z \mapsto \overline{z}$ reflects the point z along the real axis, which is an orientation-reversing similarity transformation. We extend this mapping to the Riemann sphere by setting $\overline{\infty} := \infty$.

Next, the modulus and the phase of $1/\overline{z}$ satisfy the relations $|1/\overline{z}| = 1/|z|$ and $\psi(1/\overline{z}) = \psi(z)$, and hence the points z and $1/\overline{z}$ lie on the same ray emanating from the origin and the product of their distances from the origin is equal to one. This characterizes a transformation which is known as *inversion in the unit circle* or simply *inversion*. It sends circles through the origin to straight lines, all other circles are mapped to proper circles. Note that the center of the image circle is not the inverted center of the original circle.

If the inversion is extended to all of the Riemann sphere such that $0 \mapsto \infty$ and $\infty \mapsto 0$, it maps circles on $\widehat{\mathbb{C}}$ to circles on $\widehat{\mathbb{C}}$.
Rewriting the function h in the form $h(z) = 1/z = \overline{1/\overline{z}}$ shows that h is an inversion followed by a reflection across the real axis (or vice versa), which we call an *anti-inversion*.

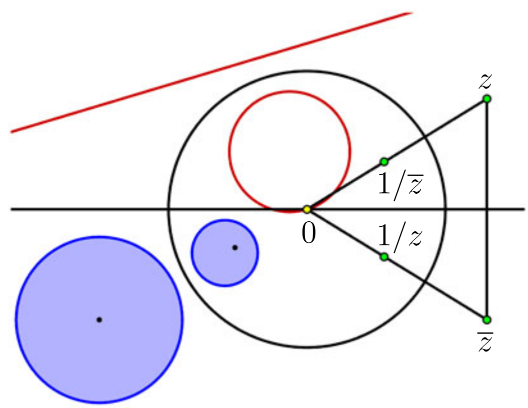

Figure 2.7: Inversion and mapping $z \mapsto 1/z$

After extending the functions f, g, and h, defined in (2.19) and (2.20), respectively, to the Riemann sphere by setting

$$f(\infty) := \infty, \quad g(\infty) := \infty, \quad h(\infty) := 0, \quad h(0) := \infty,$$

all three functions become bijective mappings of $\widehat{\mathbb{C}}$ onto itself.

Möbius Transformations. It is not difficult to see that all possible compositions of functions f (translation), g (rotostretch) and h (anti-inversion), with arbitrarily chosen values of their parameters a and $b \neq 0$, have a specific form, namely

$$F(z) = \frac{Az + B}{Cz + D}, \quad \text{with} \quad AC - BD \neq 0.$$

Functions of this type play a prominent role in complex analysis. In honor of August Ferdinand Möbius they are called *Möbius transformations*. All Möbius transformations are bijections of the Riemann sphere onto itself and form a *group* with respect to composition. In particular the inverse of F is

$$F^{-1}(z) = \frac{Dz - B}{-Cz + A}.$$

Conversely, any Möbius transformation can be composed of at most two translations, one rotostretch, and one anti-inversion. For $C = 0$ this is trivial, if $C \neq 0$, then $w = F(z)$ is a composition of

$$z \mapsto w_1 := \frac{C^2}{BC - AD}\, z \mapsto w_2 := w_1 + \frac{CD}{BC - AD} \mapsto w_3 := \frac{1}{w_2} \mapsto w := w_3 + \frac{A}{C}.$$

The investigation of Möbius transformations will be continued in Section 6.3 which is exclusively devoted to this class of functions.

2.4 The Analytic Landscape

By now we already know some special complex functions, and it might be interesting to *see* them. Real functions can be conveniently depicted by their *graph*. But when we try to do the same for a complex function $f : D \subset \mathbb{C} \to \mathbb{C}$ we are quickly stumped because the graph

$$G_f := \big\{\big(z, f(z)\big) \in \mathbb{C} \times \mathbb{C} : z \in D\big\}$$

of f lives in $\mathbb{C} \times \mathbb{C}$, which has four real dimensions. In order to stay in three spatial dimensions, we need a substitute for the missing fourth dimension.

As has already been discussed in Chapter 1, one option is to start with the traditional analytic landscapes and to incorporate the missing information by color.

The *analytic landscape* A_f (also known as the *relief* or *module surface*) of a function $f : D \subset \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is the graph of its absolute value,

$$A_f := \big\{\big(z, |f(z)|\big) \in \widehat{\mathbb{C}} \times \widehat{\mathbb{R}}_+ : z \in D\big\}.$$

Analytic landscapes involve only the *modulus* of a function and leave out its *argument*. In order to avoid the ambiguity of the latter, we shall work with the *phase* instead. Since the phase of non-zero complex numbers lives on the unit circle $\mathbb{T}$, and points on a circle can naturally be encoded by colors, color is an ideal candidate for visualizing phase.

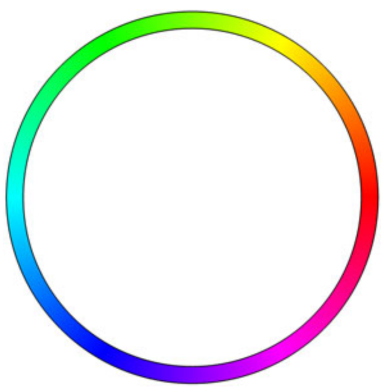

Figure 2.8: The color circle

Colored Analytic Landscapes. The *colored analytic landscape* of a complex function is the graph of its modulus colored according to its phase. Formally, the colored analytic landscape C_f of a function $f : D \to \widehat{\mathbb{C}}$ can be defined as the set

$$C_f := \big\{ \big(z, |f(z)|, \psi(f(z))\big) \in \widehat{\mathbb{C}} \times \widehat{\mathbb{R}}_+ \times \widehat{\mathbb{T}} : z \in D \big\},$$

where z is the position of the base point, $|f(z)|$ is interpreted as the height of the point, and the phase $\psi(f(z))$ determines its color.

The picture on the left in Figure 2.9 shows the colored analytic landscape of the function

$$f(z) = (z-1)/(z^2+z+1) \tag{2.21}$$

in the square $|\mathrm{Re}\, z| \le 2$, $|\mathrm{Im}\, z| \le 2$, which will henceforth serve as a *standard example*. The function has a *zero* at $z_0 = 1$, i.e., $f(z_0) = 0$. At $z_1 := (-1+\sqrt{3}\,\mathrm{i})/2$ and $z_2 := (-1-\sqrt{3}\,\mathrm{i})/2$ the denominator of $f(z)$ vanishes. We set $f\big(z_{1/2}\big) := \infty$ and refer to z_1 and z_2 as the *poles* of f.

If the modulus of a function varies over a wide range, it is better to use a *logarithmic scaling* of the vertical axis. This representation is also more natural since $\log|f|$ and $\arg f$ are conjugate harmonic functions (see Section 4.6). The corresponding *colored logarithmic analytic landscape* of the standard example is depicted in Figure 2.9 (right).

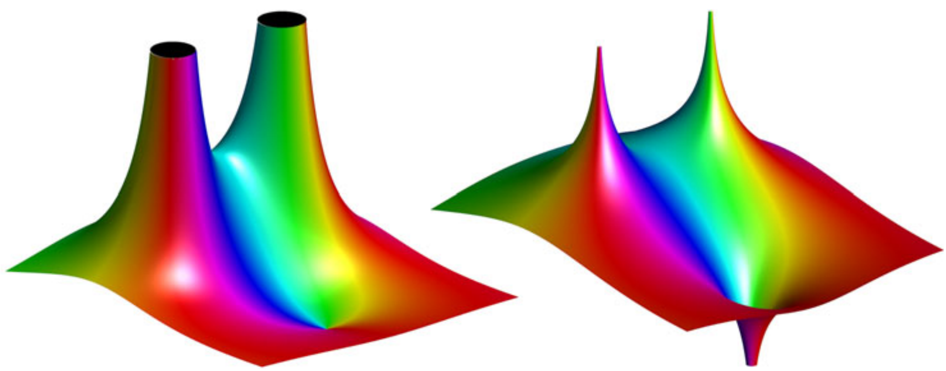

Figure 2.9: Colored analytic landscapes with conventional and logarithmic scaling

The Color Scheme for Phase. Of course the color coding of the phase is by no means unique. It is clear that saturated colors are preferable because they can be identified and distinguished better. This, in some sense, picks out the 'hue' component of the `HSV` (hue, saturation, value) color scheme. The corresponding color wheel is shown in Figure 2.8. For psychological reasons, the standardized color circle has been rotated such that positive values are encoded red. Finally, the exceptional values zero and infinity are associated with black and white, respectively. This color scheme will be used throughout the text. We would like to encourage others to adopt the same color scheme to make phase portraits comparable.

2.5 Color Representations

With colored analytic landscapes the problem of visualizing complex functions could be considered solved. However, the module surface is a three-dimensional object which usually must be projected onto a two-dimensional sheet of paper or a screen for visualization. Often this projection causes problems since interesting elements (like zeros) become invisible or are hard to detect.

Domain Coloring. But there is yet an alternative approach which is not only simpler but also more general. Having introduced colors, it is kind of natural to use them not only for representing the phase of a function, but also to *completely encode its values, using a two-dimensional color scheme.* Then, instead of drawing a graph, one can depict a function directly on its domain by color-coding its values, thus converting it to an *image.*

Coloring techniques have been customary for many decades, for example in depicting altitudes or temperatures on maps, but in most cases they represent *real-valued* functions using a one-dimensional color scheme. Two-dimensional color schemes for visualizing complex valued functions have been in use at least since the late 1980s (Larry Crone [7], see Hans Lundmark [38]), but they became popular only with Frank Farris' review [15] of Tristan Needham's book [44]. Farris also coined the suggestive name *domain coloring.*

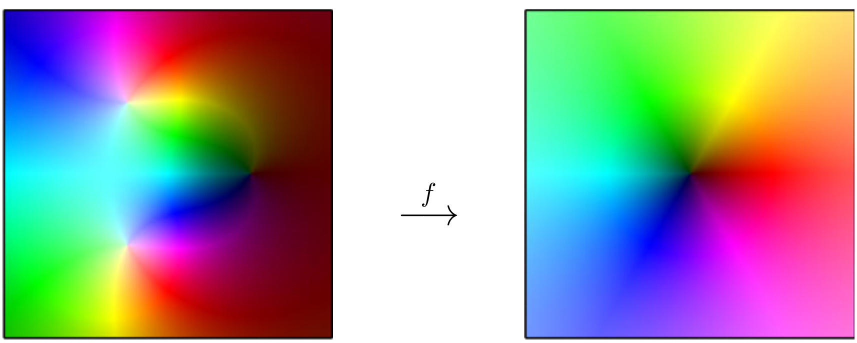

Figure 2.10: Domain coloring of the function $f(z) = (z-1)/(z^2+z+1)$

Figure 2.10 (left) shows a domain coloring for our standard example. The related color scheme for the values in the complex w-plane is shown in the window on the right. As usual, phase is encoded as "color" (in the more precise meaning of *hue*), while "lightness" corresponds to the modulus. Any point z in the domain of f carries the same color like its image $f(z)$ in the w-plane.

Compared to the colored analytic landscape, domain coloring has the advantage that functions are represented in two dimensions, which makes it easier to visualize complicated functions.

Domain coloring pictures are often very beautiful and some are even quite artistic (see Crone [7], or Glaeser and Polthier [18]). On the other hand they may be somewhat fuzzy, which makes it difficult, for example, to locate zeros precisely. Moreover, typically the modulus of a function varies in a wide range, while the human eye normally is not very sensitive to different shades of gray. Thus the color scheme must be appropriately tuned to the function – using one and the same standardized scheme for all functions does not always give satisfactory results.

Phase Portraits. What happens, if we forget about the modulus completely and just depict the color-coded phase? This *phase portrait* is exactly what we see looking at the colored analytic landscape straight from the top in the direction perpendicular to the xy-plane. Figure 2.11 shows the result for the example function defined in (2.21).

To give a formal definition, if f is a complex function on D, then the mapping

$$\Psi_f : D \to \widehat{\mathbb{T}},\ z \mapsto \psi\big(f(z)\big)$$

will be designated as the *phase* of f, and its graph

$$P_f := \{(z, \Psi_f(z)) : z \in D\}$$

is referred to as the *phase portrait* or *phase plot* of f. If $\widehat{\mathbb{T}} := \mathbb{T} \cup \{0, \infty\}$ is identified with the color circle extended by black (corresponding to 0) and white (corresponding to ∞), then the phase portrait of a function can be interpreted as an image.

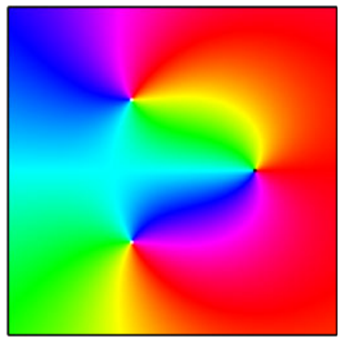

Figure 2.11: A phase portrait

Phase Portraits Versus Analytic Landscapes. In a sense, phase portraits are complementary to the (uncolored) analytic landscape: one neglects modulus, the other one omits phase. So it seems that not much is gained, but indeed there are some essential advantages of phase portraits over analytic landscapes.

First of all, the phase portrait is a two-dimensional image, which does not need to be projected, and our brain is well trained in interpreting such images. Secondly, compared with the range of the modulus of a typical function, the range of the phase is quite small since it is a subset of the extended unit circle. Consequently the visual resolution is much higher for the phase than for the absolute value, which allows us to represent all functions with one and the same color scheme. Thirdly, the reconstruction of missing information is simpler and more accurate for phase plots. We cannot discuss this in detail right now, but, hopefully, it will become clear in due course.

On the other hand, one must also mention that phase portraits are not appropriate to visualize all complex functions because two different functions may have the same phase portrait when they differ only in their modulus.

The situation changes when we restrict ourselves to the important class of *analytic functions* which are of prime importance in this book. As we shall see in Section 3.4 (Corollary 3.4.9), such functions are *completely characterized* (up to a positive scaling factor) by their phase portraits.

Enhanced Phase Portraits. Since phase occupies only one dimension, there is plenty of room in the color space to incorporate additional information. For example, encoding the modulus in a scale of gray would lead us back to conventional domain coloring. But we can make other more interesting modifications as well. Which one we shall choose depends on the properties which we would like to emphasize and perhaps also on the function which we are investigating. In order to understand this better, let us reconsider the phase portrait from a slightly different point of view.

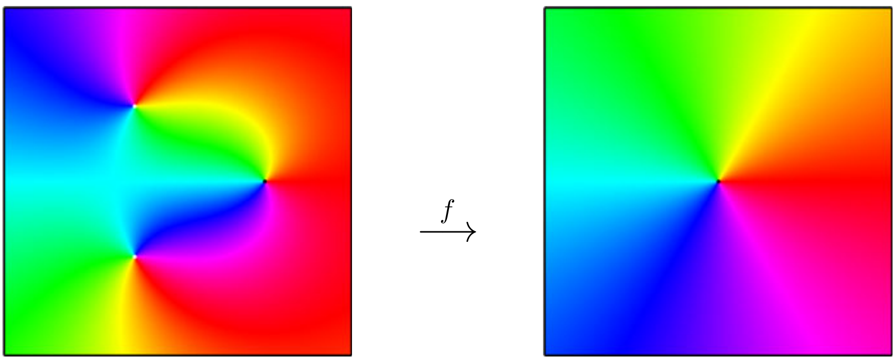

Figure 2.12: The phase portrait as pull back of the colored w-plane

Figure 2.12 shows a function f mapping the complex z-plane to the complex w-plane. In order to generate the phase portrait of f, first the w-plane is colored according to the phase of its points. In the second step, every point z in the domain of definition of f gets the same color as the value $f(z)$ has in the w-plane. In short, the phase portrait on the left is the *pull back* of the picture on the right by the function f.

Let us pause here for a moment and think about what could happen if we would try to transplant an image in the other direction. *Pushing forward* an image from the z-plane to the w-plane via f means that the color of every point z in the domain D is transplanted to the corresponding point $f(z)$ in the range of f. But this causes a conflict whenever two points z_1 and z_2 are colored differently and f attains the same value at z_1 and z_2.

While pushing forward may be problematic, it is clear that *any* picture in the w-plane can be pulled back to the z-plane by the function f. This provides many options to modify and enhance the color scheme of phase portraits.

In this book we shall mainly use three variations of phase portraits which are depicted on page 32 for the standard example $f(z) := (z-1)/(z^2+z+1)$.

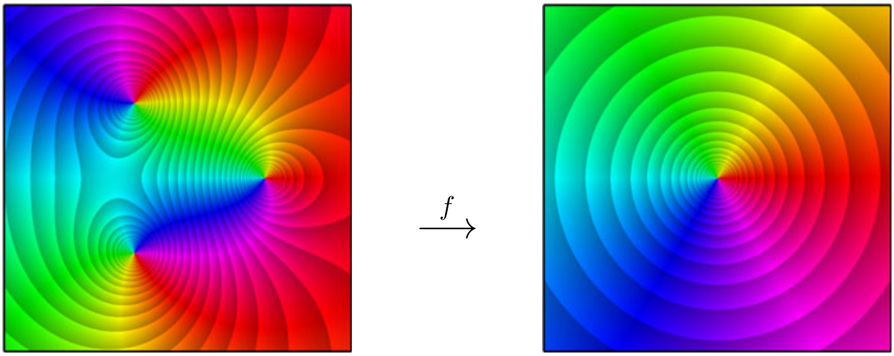

Figure 2.13: Generation of a phase portrait with modulus contour lines

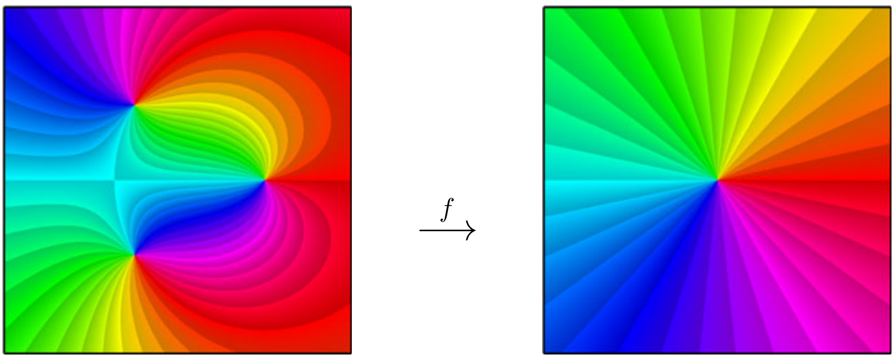

Figure 2.14: Generation of a phase portrait with phase contour lines

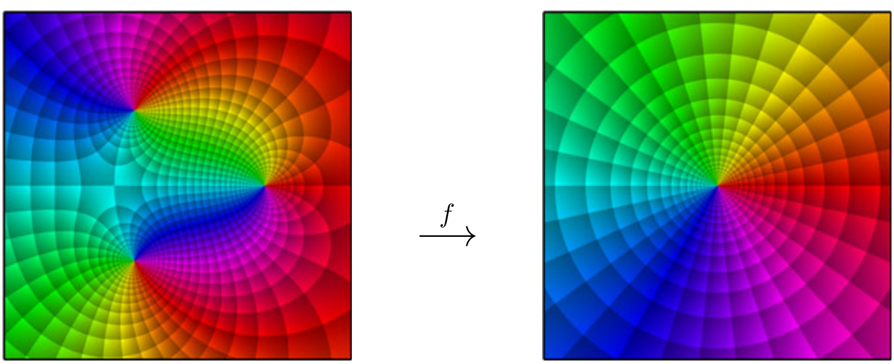

Figure 2.15: A phase portrait with contour lines of modulus and phase

The visible domain is the square given by $|\mathrm{Re}\, z| \le 2$, $|\mathrm{Im}\, z| \le 2$. Recall that the function has a zero at $z_0 = 1$ and two poles at $z_{1/2} = (-1 \pm \sqrt{3}\,\mathrm{i})/2$. The right-hand windows of all three figures show the color scheme of the w-plane, on the left we see the corresponding enhanced phase plots generated by pulling back the right image by the function f.

The pictures in Figure 2.13 involve a gray component g which is a sawtooth function of $\log|f|$, like

$$g = \lceil \log|f| \rceil - \log|f|.$$

Here $x \mapsto \lceil x \rceil$ is the *ceiling function*, which determines the smallest integer not less than x. Figure 2.16 shows the function $x \mapsto \lceil x \rceil - x$ (left) and a typical gray intensity as function of $|f|$ (right).

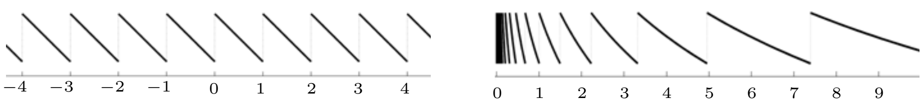

Figure 2.16: The function $x \mapsto \lceil x \rceil - x$ and the gray value as function of $|f|$

The jumps in the gray component generate *contour lines* of $|f|$, i.e., lines of constant modulus. In between two such lines darker colors correspond to smaller values of $|f|$. Though this coloring is relatively insensitive to the range of the function depicted, the distance between adjacent contour lines can be adjusted by modifying the frequency of the sawtooth function g.

It is worth noticing that the contour lines arise from a visual effect and need not really be computed. The proposed shading methods are stable, they do not require sophisticated numerical algorithms, and work with almost no additional computational effort.

Figure 2.14 demonstrates an analogous effect for the phase. In the plain phase portrait the sets of constant phase are *isochromatic*. Here some of these lines are enhanced by the discontinuities of the shading.

In Figure 2.15, the gray value is the product of two sawtooth functions depending on the logarithm of the modulus and the phase of w, respectively. In the w-plane the discontinuities of this shading generate a (logarithmically scaled) polar *tiling*.

Notice that the frequencies of the sawtooth functions encoding modulus and phase are not independent of each other, but chosen such that the tiles are "almost squares". This vague statement will be made precise in Section 6.1, it suffices here to say that all tiles have the property that they have four right-angled corners and four sides of approximately the same lengths.

The corresponding tiles in the z-plane, generated by the pull-back via the function f, are shown in Figure 2.15 (left). Somewhat surprisingly, these tiles also seem to possess the same property, as far as visual inspection allows us to decide and provided that we ignore some exceptions. The reason behind this observation is a specific property of the mapping $z \mapsto f(z)$: the transplantation via f preserves the angle of intersection between curves. Mappings with this property are called *angle-preserving* or *conformal* and will be studied in Section 6.1.

Other Color Schemes. While investigating conformal maps, functions will be visualized using black-and-white coloring, because it depicts conformality more clearly and suppresses irrelevant information. Two such schemes can be seen at the top and in the middle of page 35.

The starting point of Figure 2.17 is a polar chessboard-like tiling of the w-plane. The corresponding coloring of the domain reflects the phase and the (logarithm of the) modulus of f. This figure is a black-and-white version of Figure 2.15. Observe the high resolution at the three points where f is zero or infinity.

In Figure 2.18, the w-plane is colored in the conventional Cartesian chessboard style. In its pull-back to the z-plane we see an extremely fine structure near the poles of f, where a large number of squares from the w-plane is compressed into a small region. On the other hand, the zero of f cannot be seen at all, as its neighborhood does not have any special structure.

The last figure neglects phase information completely. The ring-shaped alternating black and white stripes in Figure 2.19 are the sets

$$\{z \in D : \; k\,d < \log|f(z)| \le (k+1)\,d\}, \quad \{w \in \mathbb{C} : \; k\,d < \log|w| \le (k+1)\,d\}$$

with $k \in \mathbb{Z}$ and some positive d. In the left picture, their boundaries are contour lines of f. This type of coloring is particularly useful in producing equipotential lines in plane electrostatics, which will be demonstrated in Section 4.6.

The color schemes which are based on a polar grid in the w-plane yield a high resolution near zeros (and poles) of f in the z-plane. This 'microscope effect' can be utilized to explore the structure of a function f in a neighborhood of any other point of its domain as well. To enhance the resolution at a point a in the z-plane, just shift the origin of the polar grid in the w-plane to the point $b = f(a)$. The same effect is achieved by considering the function $f(z) - f(a)$ in the usual scheme. One should be aware that this transformation changes phase and modulus of the function, and that there is no simple relation between the phases of the functions f and $f - a$.

Finally, we mention that the images on the left side of all figures are constructed as the pull-back of the *entire* colored complex w-plane, while the windows on the right-hand side show only a section illustrating the color scheme.

Since it is the declared goal of this book to promote phase portraits as a tool for visualizing and exploring complex functions, we refrain from using conventional domain coloring. Though, in principle, we will prefer to explain ideas and concepts using plain phase portraits as much as possible, we shall often resort to modifications if they demonstrate relevant facts more clearly. Another motive for making modifications is aesthetics; for example, when the plain phase portraits lack an interesting structure.

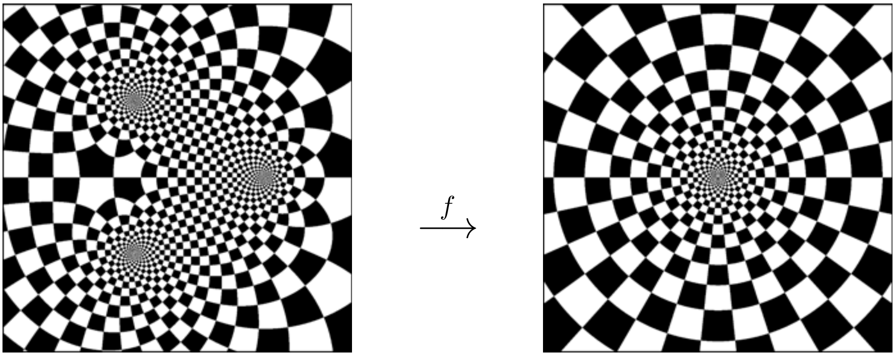

Figure 2.17: The pull back of a polar chessboard visualizes conformality

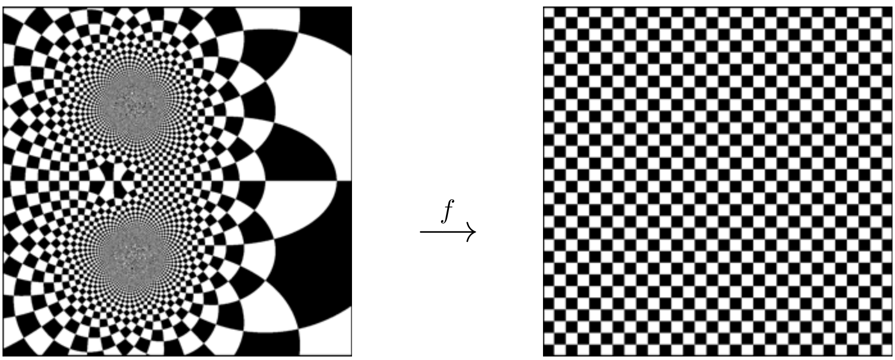

Figure 2.18: The pull back of a Cartesian chessboard hides zeros

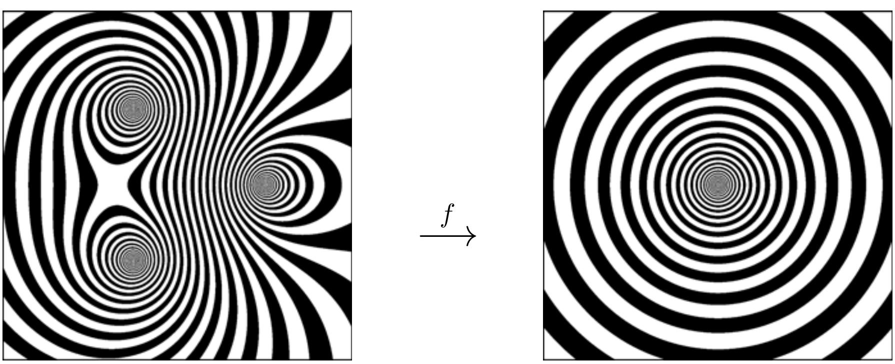

Figure 2.19: The pull back of circular rings generates modulus contour lines

Practical Excursion. With this much preparation, we can now try to get some practice in working with phase portraits. We shall introduce some more basic notions at a somewhat informal level as we go along.

Figure 2.20 depicts again the familiar example $f(z) = (z-1)/(z^2+z+1)$. We have already mentioned that the three exceptional points where all colors come together correspond to the zero and the two poles of the function. Notice that zeros and poles can be distinguished by the ordering of colors in their neighborhood.

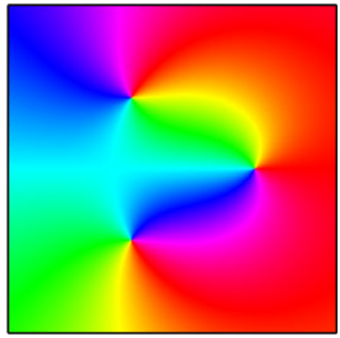

Figure 2.20: A phase portrait

A more careful inspection reveals that there might be yet another special point in the light blue region on the negative real axis. Since it is located in a diffused spot, we modify the phase coloring (by overlaying a discontinuous gray component) to get a sharp contrast in the color of interest. Adjusting the jump in the color scheme correctly, we get the result shown on the left in Figure 2.21. Now the exceptional point is clearly visible, it is the crossing point of (two) smooth curves having the same color. We shall call such points *saddles*.

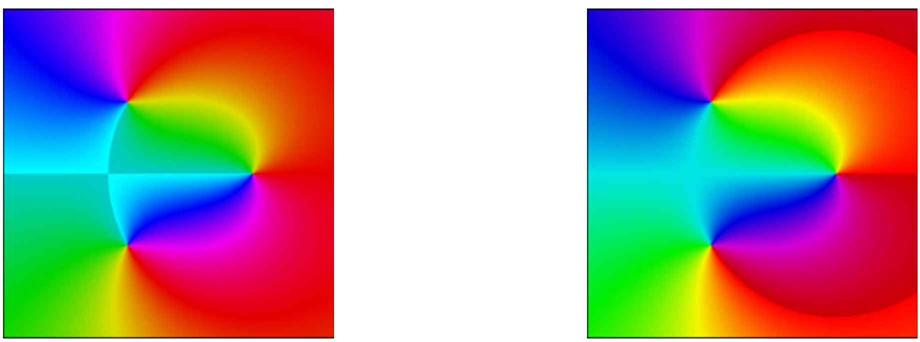

Figure 2.21: Two phase portraits with enhanced isochromatic lines

Isochromatic Sets. In order to describe this more precisely and to explore the structure of phase portraits in some detail, we introduce *isochromatic sets*. If f is a function on a domain D these are defined by

$$S(c) := \{z \in D : \psi(f(z)) = c\}, \quad c \in \widehat{\mathbb{T}}.$$

If the point c on the unit circle is identified with the corresponding color on the color wheel, then $S(c)$ is the set of all points in D carrying the color c.

In Figure 2.21 (left) the blue line and the crossing arc constitute the isochromatic set $S(-1)$. Similarly, in the window on the right the color changes abruptly across three red arcs belonging to the isochromatic set $S(1)$.

Complex and Analytic Functions. If we do not impose additional restrictions, like continuity or differentiability, the isochromatic sets of complex functions can be arbitrary – but this is not so for "analytic" functions, which are the objects of prime interest in this text. Further investigation of this question is deferred to the next chapter and will be continued in more detail in Volume 2 where we shall see that any isochromatic set of an analytic function is the union of smooth arcs which can be linked only at saddle points in a very specific way.[3]

Warning. *Most statements and results in the following chapters are about analytic functions and do not hold for complex functions in general.*

In Section 3.4, we shall see that *analytic functions* are (almost) uniquely determined by their (pure) phase portraits (compare Theorem 3.4.10), but this is not so for general functions. For example, the functions f (analytic) and g (not analytic) defined by

$$f(z) = (z-1)/(z^2+z-1), \qquad g(z) = (z-1) \cdot (\overline{z}^2 + \overline{z} - 1), \tag{2.22}$$

have the same phase (except at their zeros and poles) though they are completely different.

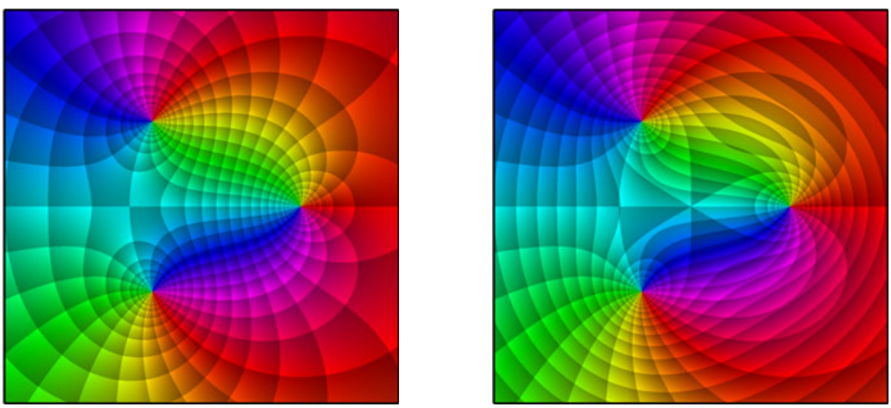

Figure 2.22: Enhanced phase portraits of f (left) and g (right)

Since pure phase portraits do not always display enough information for exploring general complex functions, we recommend use of their enhanced versions with contour lines of modulus and phase in such cases. Figure 2.22 shows two such portraits of the functions f (left) and g (right) defined in (2.22).

[3]See also Wegert [69].

A notable distinction between the two portraits is the shape of the tiles. In the left picture most of them are almost squares and have right-angled corners. In contrast, many tiles in the portrait of g are prolate and their angles differ significantly from $\pi/2$ – at some points the contour lines of modulus and phase are even mutually tangent. We shall explore these observations in Section 6.1.

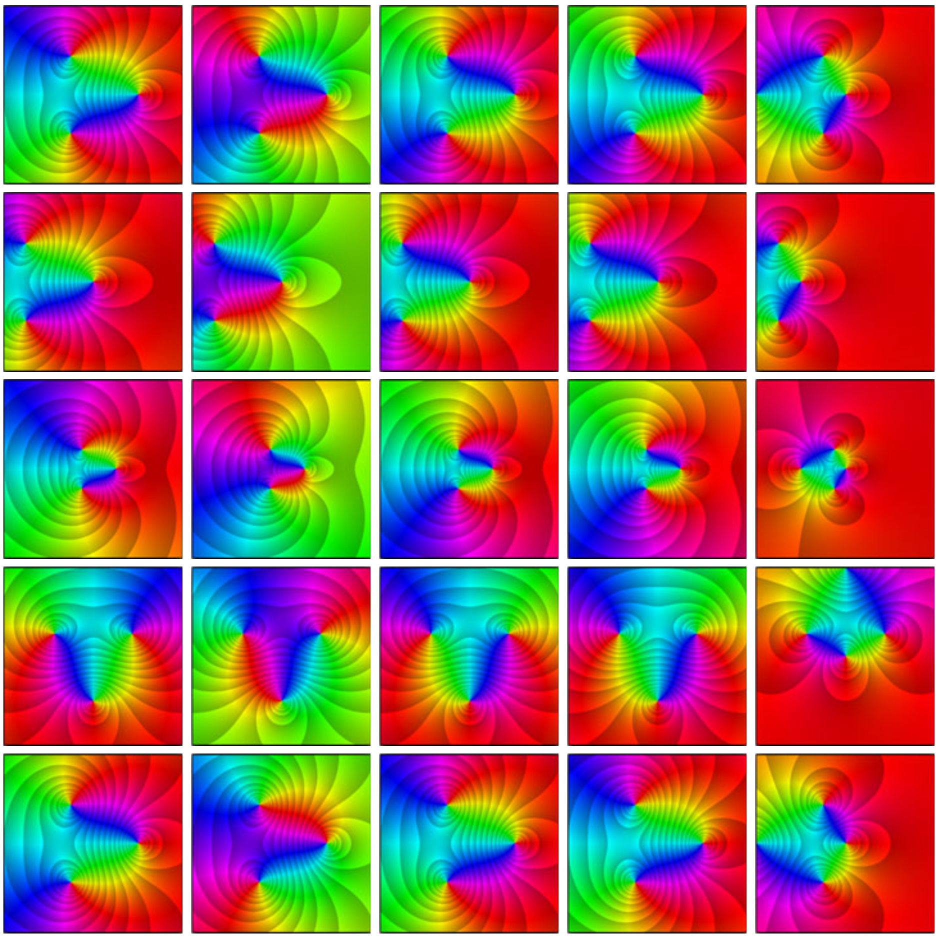

Figure 2.23: Phase portraits with contour lines of compositions $h \circ f \circ g$

The Effect of Transformations. Figure 2.23 demonstrates what happens with the phase portrait of a function $w = f(z)$ when the variables z and w are transformed. It displays the phase portraits of the compositions $h \circ f \circ g$, where g corresponds to the rows and h is associated with the columns, and, in this order,

$$g : z \mapsto z, \quad z+1, \quad 2z, \quad \mathrm{i}\, z, \quad \overline{z}, \qquad h : w \mapsto w, \quad \mathrm{i}\, w, \quad 1/w, \quad \overline{w}, \quad w+1.$$

The domain depicted is the square $-2 \leq \operatorname{Re} z, \operatorname{Im} z \leq 2$. Notice that, compared to the first row, the plots are shifted to the left in the second row, shrunk by a factor of two in the third row, rotated clockwise by $\pi/2$ in the fourth row, and reflected at the real axis in the fifth row.

Though the pure phase portraits of the third and the fourth columns would be identical, there is a difference in the gray shading which represents the modulus. The fifth column seems to have no relation to the others.

Phase Portraits on the Sphere. Phase portraits of functions defined on the extended complex plane can be drawn directly on the Riemann sphere. Though this idea is natural, and results in quite nice pictures (and I cannot resist the temptation of showing you the transplantations of the images on page 35 to the sphere), it has several disadvantages.

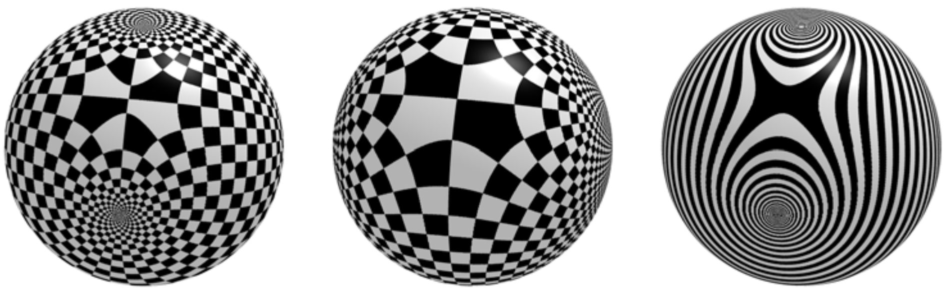

Figure 2.24: A function on the sphere represented by black-and-white schemes

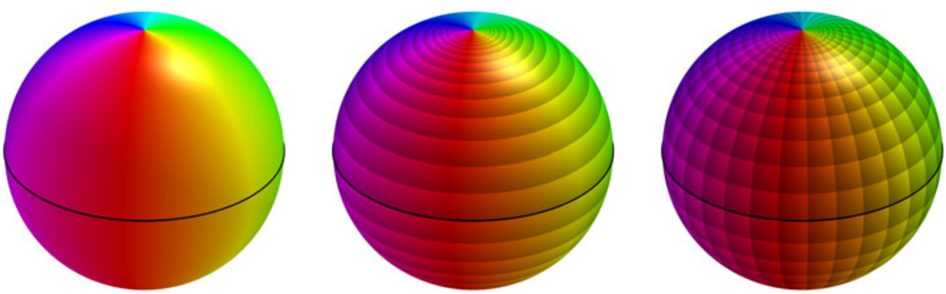

Figure 2.25: Three color schemes on the Riemann sphere

Figure 2.25 shows the coloring of $\mathbb{S}$ according to the three color schemes on page 32, which can be interpreted as the (modified) phase portraits of the function $f(z) = z$.

Since only part of the sphere can be seen from a fixed position, at least two pictures are needed to represent the complete sphere, and due to perspective projection, precise localization of points is difficult. Also, the lighting and shading needed to make the sphere look round distorts the colors. But worst of all, when we inspect a neighborhood of the point at infinity, where the value of the function $f(z) = z$ is infinite, we see a coloring which resembles a zero. Conversely, the neighborhood of a zero is colored in a manner similar to that for poles. Why does this happen?

The reason is that stereographic projection acts between the *upper side* of the plane and the *interior side* of the sphere. This becomes obvious when we consider the projection of the lower hemisphere onto the unit disk, but it remains true for the upper hemisphere as well. Looking at a sphere from the outside, as we usually do, *reverses the orientation*, which makes zeros look like poles and vice versa.

Of course, once we are aware of this phenomenon, we could live with it and need not change anything, except mentally altering orientation whenever we switch between the plane and the sphere. However, since orientation is crucial in the interpretation of phase portraits, this may be rather confusing.

A simple alternative is to represent the sphere $\mathbb{S}$ by *two charts* which depict what we see looking at $\mathbb{S}$ from the inside. The first one is the stereographic projection P_N of the sphere $\mathbb{S}$ from the north pole N (identified with the point at infinity) to a complex plane $\mathbb{C}$ attached to the *upper side* of the horizontal plane P. The projection P_N maps $\mathbb{S} \setminus \{N\}$ onto $\mathbb{C}$ and will be used to represent the lower hemisphere.

The second chart P_S is the stereographic projection of $\mathbb{S}$ from the south pole S to a complex plane $\mathbb{C}$ attached to the *lower side* of P. It maps $\mathbb{S} \setminus \{S\}$ onto $\mathbb{C}$ and will be used to represent the upper hemisphere.

The result of transplanting the phase portrait from $\mathbb{S}$ to the two copies of $\mathbb{C}$ via the charts P_N and P_S is shown in Figure 2.27. In the left window the image of the lower hemisphere is highlighted, while the saturated colors in the right window emphasize the image of the upper hemisphere. In both pictures the images of the points $1, \mathrm{i}, -1, -\mathrm{i}$ are marked.

All points on $\mathbb{S}$, except the two poles N and S, are represented in both charts. This defines a *transition map* between the images of the intersection of the domains of both charts,

$$P_S \circ P_N^{-1} : \ \mathbb{C} \setminus \{0\} \to \mathbb{C} \setminus \{0\}, \ z \mapsto 1/z.$$

Intuitively, the colored sphere can be modelled from the two pictures in Figure 2.27 by the following construction: cut off the two highlighted disks, put the pieces together face-to-face (colors inside) so that the marked points (and then all points on the unit circle) fit and glue them along the unit circle. Now blow it up to a spherical balloon!

After rotating the right window by an angle of π about the origin, as shown in Figure 2.28, this procedure can be simplified: fold at the dashed line, glue along the black circle, cut off the outer part, and inflate it.

Since this interpretation of Figure 2.28 is intuitive, we shall mainly use these pictures for representing colorings of the Riemann sphere.

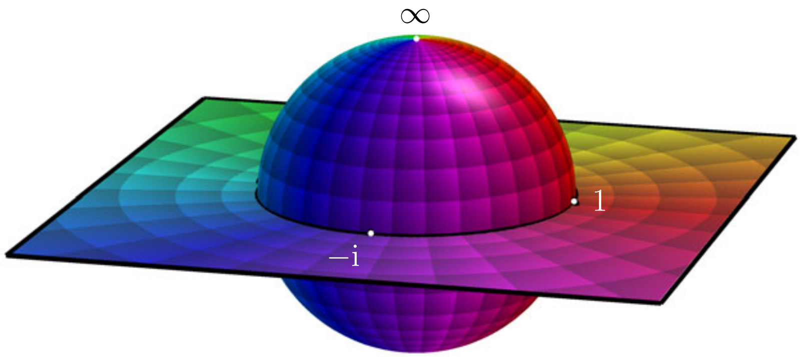

Figure 2.26: Stereographic projection of phase portraits

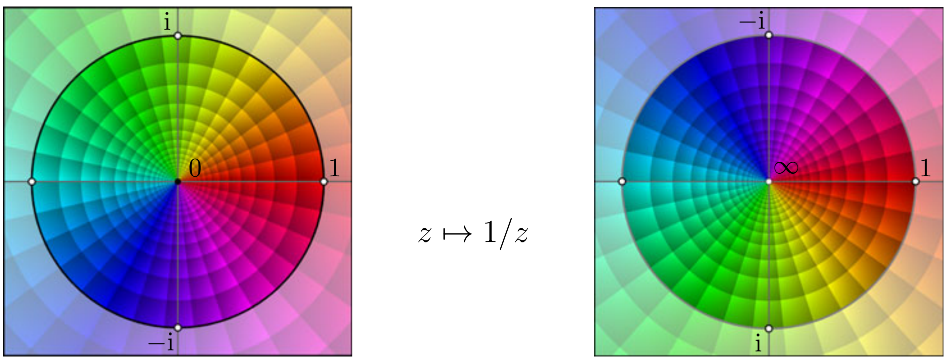

Figure 2.27: Two charts of the sphere corresponding to the hemispheres

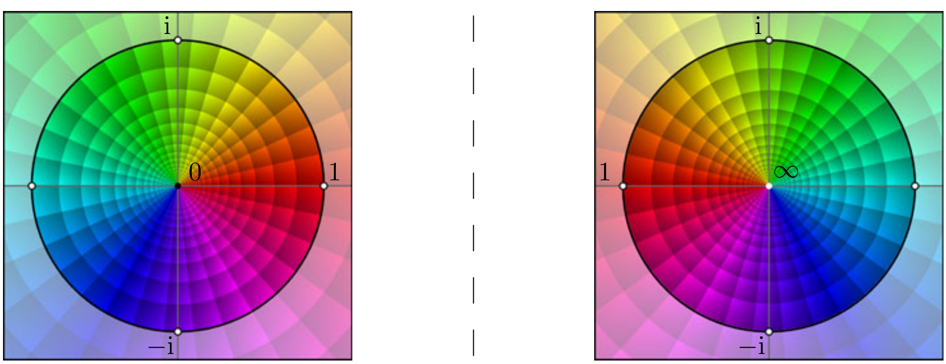

Figure 2.28: The standard representation of the sphere by two charts

An Example. Figure 2.29 depicts an enhanced phase portrait with modulus and phase contour lines of the standard example $f(z) = (z-1)/(z^2+z+1)$ on the Riemann sphere. The picture on the left shows a region which we have already seen before – the domain in the right is the exterior of the unit disk with the point at infinity in the center.

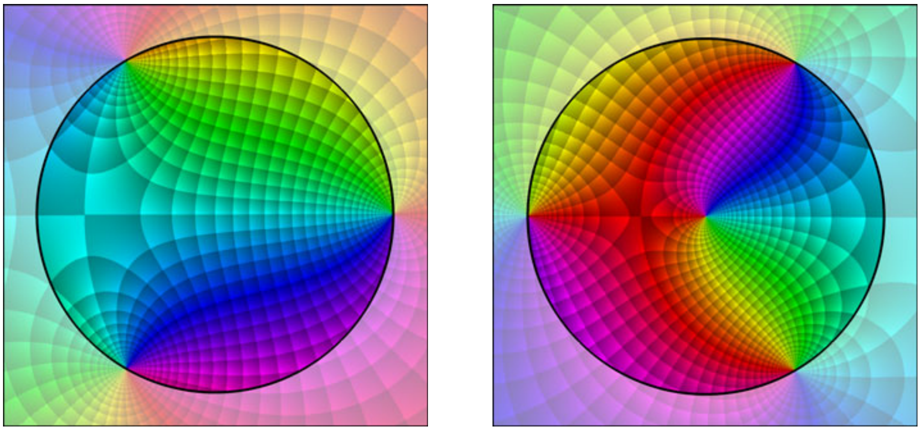

Figure 2.29: Enhanced phase portrait of the standard example on the sphere

It seems that something special happens at this point: it is one of the four distinctive points on the sphere where all colors of the chromatic circle meet. The other three points of this type are the zero and the two poles on the unit circle. Though the function is not yet defined at $z = \infty$, we observe that its phase portrait *in a neighborhood of this point* resembles the typical behavior at a zero. The function somehow seems to *request* a zero at infinity! So we satisfy its desire by setting $f(\infty) := 0$. Why this is indeed the right choice will be investigated in the next section.

2.6 Convergence and Continuity

Though we assume that the reader has some familiarity with these notions, we review some basic facts without proofs.

Sequences. A *sequence* $(z_n)_{n=1}^{\infty}$ in a set A is a function on the set $\mathbb{Z}_+$ of positive integers with values in A. For brevity we shall often use the simplified notation (z_n), but sometimes we also use the longish but more intuitive form $z_1, z_2, z_3, \ldots$. In order to indicate that (z_n) is a sequence in A we write $(z_n) \subset A$. Most sequences we shall meet in this book are sequences in $\mathbb{C}$, or more generally, in $\widehat{\mathbb{C}}$.

Note that the indexing of a sequence need not necessarily start with $n = 1$, the point is that the elements of a sequence are strictly ordered, which *in principle* allows indexing by the natural numbers.

The concept of convergent sequences in $\mathbb{C}$ is analogous to that for sequences of real numbers: a sequence $(z_n) \subset \mathbb{C}$ *converges* to $z_0 \in \mathbb{C}$ if [4] for every $\varepsilon > 0$, there is an integer N such that $n \geq N$ implies $|z_n - z_0| < \varepsilon$. If such a number z_0 exists it is unique and is said to be the *limit* of the sequence (z_n). Sequences which do not converge are termed *divergent*. Convergence of (z_n) to z_0 is written as

$$\lim_{n\to\infty} z_n = z_0, \quad \text{or} \quad z_n \to z_0.$$

Less formally, a sequence (z_n) converges to z_0 if the points z_n get arbitrarily close to z_0 as n gets sufficiently large.

Limits of complex sequences have the same arithmetic properties as limits of real sequences, for example $z_n \to z_0$ and $w_n \to w_0$ imply that $z_n + w_n \to z_0 + w_0$.

Convergence of sequences in $\widehat{\mathbb{C}}$ is defined in a similar manner, with only a single modification: the Euclidean distance $|z - w|$ between two points z and w in the plane is replaced by the spherical distance $d(z, w)$. Convergence in $\mathbb{C}$ is compatible with this more general concept; this implies that if $z_0 \in \mathbb{C}$, then the sequence $(z_n) \subset \mathbb{C}$ converges to z_0 in $\mathbb{C}$, if and only if it converges to z_0 in $\widehat{\mathbb{C}}$. Convergence of z_n to ∞ can be rephrased as $1/z_n \to 0$, that is, for every positive ε there exists an N such that $|z_n| > 1/\varepsilon$ for all $n \geq N$.

Continuity. Another fundamental concept of analysis is continuity of functions. Roughly speaking, a function $f : X \to Y$ is continuous at a point $x_0 \in X$ if the distance of $f(x)$ and $f(x_0)$ becomes arbitrarily small whenever x in X is sufficiently close to x_0. Expressed in ε-δ-language, and restricted to complex functions, this reads as follows:

A function $f : D \subset \mathbb{C} \to \mathbb{C}$ is *continuous* at $z_0 \in D$ if for every positive ε there exists a positive δ such that $z \in D$ and $|z - z_0| < \delta$ imply $|f(z) - f(z_0)| < \varepsilon$.

We say that f is continuous on D, if it is continuous at each point of D. Then, in general, the value of δ depends on ε and on the chosen point $z_0 \in D$. If, for any $\varepsilon > 0$, there exists a $\delta > 0$ which does the job for all $z_0 \in D$, then f is called *uniformly continuous* on D.

The relation between continuity and convergent sequences is established in the following basic result.

Theorem 2.6.1. *A function f on D is continuous at a point $z_0 \in D$ if and only if, for any sequence $(z_n) \subset D$, $z_n \to z_0$ implies $f(z_n) \to f(z_0)$.*

Function Sequences. Next, we consider sequences (f_n) of functions $f_n : D \to \mathbb{C}$. The sequence (f_n) *converges at $z \in D$* if the number sequence $(f_n(z))$ converges. If this happens for all points z in D, we say that (f_n) *converges pointwise* on D.

Pointwise convergence is often too weak to get nice results, for example the limit function of a pointwise convergent sequence of continuous functions is not necessarily continuous. Therefore we need a stronger version of convergence.

[4] By convention, the "if" in definitions has always the meaning "if and only if".

The sequence (f_n) *converges uniformly* on D to a function $f : D \to \mathbb{C}$ if for every $\varepsilon > 0$ there exists an integer N such that $|f_n(z) - f(z)| < \varepsilon$ for all $n \geq N$ and $z \in D$.

Theorem 2.6.2. *If a sequence of continuous functions f_n converges uniformly on D, then its limit function is continuous on D.*

We remark that the definitions made above also make sense when the domain set D is replaced by a subset A. For functions $f : D \subset \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ the definitions have to be modified accordingly, replacing the Euclidean distance $|z - w|$ with the spherical distance $d(z, w)$.

Number Series. A *series* $\sum_{k=1}^{\infty} a_k$ of complex numbers *converges*, if the sequence (s_n) of its *partial sums* $s_n := \sum_{k=1}^{n} a_k$ converges. The limit $s := \lim_{n\to\infty} s_n$ is said to be the *sum* of the series. The following result is a consequence of the *completeness* of $\mathbb{C}$.

Theorem 2.6.3 (Cauchy Criterion). *A complex series $\sum_{k=1}^{\infty} a_k$ converges if and only if for each $\varepsilon > 0$ there is an N such that $n \geq m \geq N$ implies that*

$$\left| \sum_{k=m}^{n} a_k \right| < \varepsilon.$$

Some manipulations with series (for example rearrangements of the summands) require an even stronger concept of convergence.

A series $\sum_{k=1}^{\infty} a_k$ is said to *converge absolutely*, if $\sum_{k=1}^{\infty} |a_k|$ converges. It follows from Theorem 2.6.3 and the triangle inequality that absolute convergence of a series implies convergence.

The following result on changing the order of summation will be of special importance in later chapters. Here we consider a double sequence a_{jk} of complex numbers with $j, k \in \mathbb{Z}_+$ and form the *iterated sums*

$$\sum_{k=1}^{\infty} \left(\sum_{j=1}^{\infty} a_{jk} \right), \qquad \sum_{j=1}^{\infty} \left(\sum_{k=1}^{\infty} a_{jk} \right). \tag{2.23}$$

In general, the behavior of both sums can be completely different, but the situation improves if one of the series converges absolutely, i.e., it converges if the a_{jk} are replaced by their absolute values $|a_{jk}|$.

Theorem 2.6.4 (Weierstrass Double Series Theorem). *Let $a_{jk} \in \mathbb{C}$ for $j, k \in \mathbb{Z}_+$. If one of the iterated sums (2.23)* converges absolutely, *then both series converge and*

$$\sum_{k=1}^{\infty} \left(\sum_{j=1}^{\infty} a_{jk} \right) = \sum_{j=1}^{\infty} \left(\sum_{k=1}^{\infty} a_{jk} \right).$$

Function Series. Pointwise and uniform convergence of *function series* $\sum_{k=1}^{\infty} f_k(z)$ on a set $D \subset \widehat{\mathbb{C}}$ are defined by the corresponding properties of their partial sums. The next theorem is a convenient criterion for proving convergence of such series.

Theorem 2.6.5 (Weierstrass M-Test)**.** *Let (f_k) be a sequence of functions defined on $D \subset \widehat{\mathbb{C}}$. If there is a sequence of real numbers M_k such that $|f_k(z)| \le M_k$ for all $z \in A$ and $k \in \mathbb{Z}_+$, and $\sum_{k=1}^{\infty} M_k$ converges, then $\sum_{k=1}^{\infty} f_k(z)$ converges absolutely and uniformly on D.*

Landau Notation. A convenient notation for comparing the asymptotic behavior of two functions in a neighborhood of a point $z_0 \in \widehat{\mathbb{C}}$ is provided by the *Landau symbols* O, o, and $\sim$. Following the usual conventions, we write

(i) $f(z) = O\big(g(z)\big)$ if for some C and all z in a neighborhood of z_0

$$|f(z)| \le C\,|g(z)|,$$

(ii) $f(z) = o\big(g(z)\big)$ if $f(z)/g(z) \to 0$ as $z \to z_0$,

(iii) $f(z) \sim g(z)$ if $f(z)/g(z) \to 1$ as $z \to z_0$.

It goes without saying that the point z_0 where the functions f and g are compared must be specified. The Landau symbols are also used without explicitly mentioning the variable z, i.e., $f = O(g)$, $f = o(g)$, $f \sim g$.

2.7 Some Plane Geometry

The theory of complex functions is a fascinating blend of analysis and geometry. For the convenience of the reader, we assemble some related definitions and results from plane topology in this section. It is not necessary to work through the whole material at once, rather we recommend that the reader consults this section as and when the corresponding concepts become relevant in the ongoing course.

Since the focus of this book is on complex functions, we do not aim to prove *en passant* theorems in plane geometry. Instead we just quote these facts and refer the interested reader to textbooks like Munkres [42] or Henle [25]. For an introduction to geometric concepts in the context of complex analysis we also recommend Chapter II of Palka [52].

Domains. When we shall study analytic functions in the next chapter, it will become clear that their domain sets are necessarily *open* subsets of the complex plane or the Riemann sphere (Lemma 3.3.3). Such sets can be decomposed into even simpler pieces which will be introduced in Definition 2.7.1.

Recall that a set S (in a topological space) is *connected* if whenever S is the union of two non-empty disjoint sets A and B, then (at least) one of these sets contains a point which belongs to the closure of the other.

Definition 2.7.1. A *domain* is a non-void, open and connected subset of the complex plane or the Riemann sphere.

Note that this definition is consistent: any subset of $\mathbb{C}$ which is a domain in $\mathbb{C}$ is also a domain in $\widehat{\mathbb{C}}$. An alternative commonly used term for domain is *"region"*.

The Riemann sphere $\widehat{\mathbb{C}}$, the complex plane $\mathbb{C}$, the unit disk $\mathbb{D}$, the upper half-plane $\mathbb{H} := \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$, and the ring $R := \{z \in \mathbb{C} : 1/2 < |z| < 2\}$ are domains. The next result shows that any open subset of the Riemann sphere is composed of domains.

Proposition 2.7.2. *Any open set in* $\widehat{\mathbb{C}}$ *is the disjoint union of at most countably many domains.*

The domains which are the building blocks of an open set D according to Proposition 2.7.2 are said to be the *components* of D.

For some purposes even the concept of a domain is too general, and we need a finer classification. For example, a disk and a ring-shaped domain are topologically different. In order to make this distinction precise we need some further preparation.

Paths and Loops. Contrary to a common sense interpretation, a path is a *continuous function* from some closed interval into a topological space. The following definition gives a more detailed description.

Definition 2.7.3. A *path* from a to b in D is a continuous map $\gamma : [\alpha, \beta] \to D$ with $\gamma(\alpha) = a$, $\gamma(\beta) = b$, and $\alpha < \beta$. The points a and b are called the *initial point* and the *terminal point* of γ, respectively. If the initial and the terminal points coincide we speak of a *closed path* or a *loop*. The image set $[\gamma] := \gamma([\alpha, \beta])$ is said to be the *trace* (or *trajectory*) of γ.

The function γ is also referred to as a *parametrization* of the trace $[\gamma]$. Being the continuous image of an interval, the trace of a path is a compact connected set.

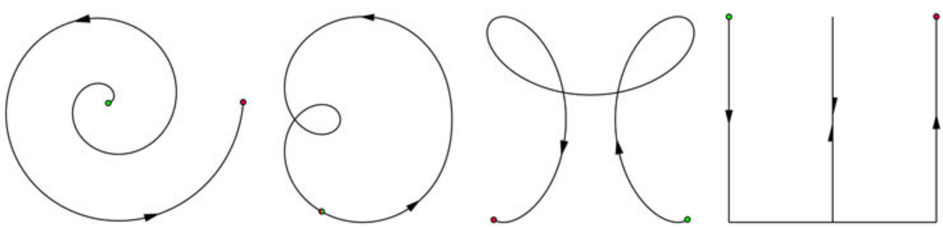

Figure 2.30: Traces of four paths with direction of increasing parameter indicated

A path is an *oriented* object. When we depict the trace of a path, the direction in which the point $\gamma(t)$ traverses $[\gamma]$ is often indicated by an arrow. Figure 2.30

shows a few examples. The green and the red dots are the initial point a and the terminal point b, respectively. These two points are also known simply as the *endpoints* of the path, and we say that γ *joins* (or *connects*) a with b. If γ is a loop, its common initial and terminal point is referred to as the *base point* of γ.

Remark 2.7.4. We do not introduce the concept of a *curve* here, because there is no need for it, at least for the moment. One should be aware that there is no general convention about the meaning of this notion. Some authors consider "curve" and "path" as synonyms, some call the trace of a path a curve, and others define a curve as an equivalence class of paths. In order to avoid confusion, we shall try to avoid using this term, except when we talk about *Jordan curves*, a notion which will be made precise below.

There are some standard paths which will come up often in our discussion. If z_1 and z_2 are two (not necessarily distinct) points in the plane, we denote by $[z_1, z_2]$ the path $\gamma(t) = z_1 + t(z_2 - z_1)$ with $t \in [0, 1]$. So $[z_1, z_2]$ denotes a path with initial point z_1 and terminal point z_2, but we use the same notation also for its (oriented) trace, the *segment* $[z_1, z_2]$.

Speaking of the *standard parametrization of a circle* with center a and radius r we usually mean the path $\gamma(t) = a + r\,\exp(2\pi i t)$ with $t \in [0, 1]$. Sometimes we shall also use $\gamma(t) = a + r\,\exp(it)$ with $t \in [0, 2\pi]$.

Simple Paths. Without additional assumptions, paths may have unexpected properties, which might be inconvenient under certain circumstances. For example, the space filling *"Peano curve"* (which is a path in our terminology) is a continuous surjective mapping of $[0, 1]$ onto the unit square $[0, 1] \times [0, 1]$. It is constructed as a (uniform) limit of a sequence (γ_n) of paths three of which are shown in Figure 2.31.[5] Such weird objects can be excluded by additional assumptions.

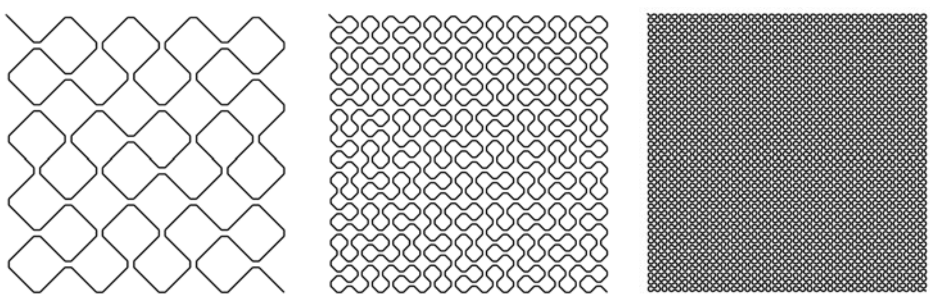

Figure 2.31: Three stages in the construction of the Peano curve

Definition 2.7.5. A path $\gamma : [\alpha, \beta] \to D$ is called *simple*if $\gamma(s) = \gamma(t)$ with $s < t$ implies that $s = \alpha$ and $t = \beta$.

[5]The curves depicted are computed using the MATLAB routine `peano.m` by Andreas Klimke.

This somewhat artificial definition allows for two possibilities: either $\gamma(\alpha) \neq \gamma(\beta)$, (in which case γ is not closed) and γ is injective, or $\gamma(\alpha) = \gamma(\beta)$ (i.e., γ is a loop) and the restriction of γ to $[\alpha, \beta)$ is injective.

Change of Parameter. As it turns out, the choice of the parameter interval is not that essential for the definition of a path. If $\gamma_1 : [\alpha_1, \beta_1] \to D$ is a path in D and $\tau : [\alpha_2, \beta_2] \to [\alpha_1, \beta_1]$ is a linear mapping with $\tau(\alpha_2) = \alpha_1$ and $\tau(\beta_2) = \beta_1$, then the path $\gamma_2 := \gamma_1 \circ \tau$ is a path in D with parameter interval $[\alpha_2, \beta_2]$. We call γ_2 a *linear reparametrization* of γ_1.

In fact there are no "essential differences" between γ_1 and γ_2, and none of the results which we shall encounter is affected by linear reparametrization – the sceptical reader is invited to formulate and prove the corresponding results whenever he or she has doubts.

Any path γ admits a (unique) linear reparametrization with the parameter interval $[0, 1]$. In what follows we shall often assume that paths are given in this *normalized form.*

Path Manipulations. Next we introduce some manipulations with paths. The *reversed path* (or *negative* path) γ^- of $\gamma : [\alpha, \beta] \to \mathbb{C}$ is defined on $[\alpha, \beta]$ by $\gamma^-(t) := \gamma(\alpha + \beta - t)$.

If the terminal point of $\gamma_1 : [\alpha_1, \beta_1] \to \mathbb{C}$ coincides with the initial point of $\gamma_2 : [\alpha_2, \beta_2] \to \mathbb{C}$, the *concatenation* (also denoted as *sum*) $\gamma := \gamma_1 \oplus \gamma_2$ of γ_1 and γ_2 is defined on $[\alpha_1, \beta_1 + \beta_2 - \alpha_2]$ by

$$\gamma(t) := \begin{cases} \gamma_1(t) & \text{if} \quad t \in [\alpha_1, \beta_1] \\ \gamma_2(t + \alpha_2 - \beta_1) & \text{if} \quad t \in (\beta_1, \beta_1 + \beta_2 - \alpha_2]. \end{cases}$$

This "sum" is associative, but not commutative. Further we set $\gamma_1 \ominus \gamma_2 := \gamma_1 \oplus \gamma_2^-$, provided that the right-hand side makes sense.

When working with normalized paths γ_1 and γ_2 defined on $[0, 1]$, we use a modified definition of concatenation, where the parameter interval of $\gamma_1 \oplus \gamma_2$ is rescaled to $[0, 1]$ so that the resultant path is also a normalized one.

Further Properties. Sometimes we shall need paths with specific properties. Three such classes are described in the next definition.

Definition 2.7.6. A path γ is said to be *smooth*, if $\operatorname{Re} \gamma$ and $\operatorname{Im} \gamma$ are continuously differentiable functions. A *polygonal path* is the sum $\gamma = \gamma_1 \oplus \ldots \oplus \gamma_n$ of segments $\gamma_k = [z_{k-1}, z_k]$. A polygonal path is called *paraxial* if each of its segments is parallel to the real or the imaginary axis.

Note that smoothness of a path does not necessarily imply that its trace looks like a smooth curve.

Figure 2.32 shows the traces of three (polygonal) paraxial paths. They are members of a sequence (γ_n) which converges (uniformly) to the so-called *Hilbert curve*, another continuous surjective mapping of an interval onto a square.[6]

---

[6]The depicted curves are computed using the MATLAB routine `hilbert.m` by Federico Forte.

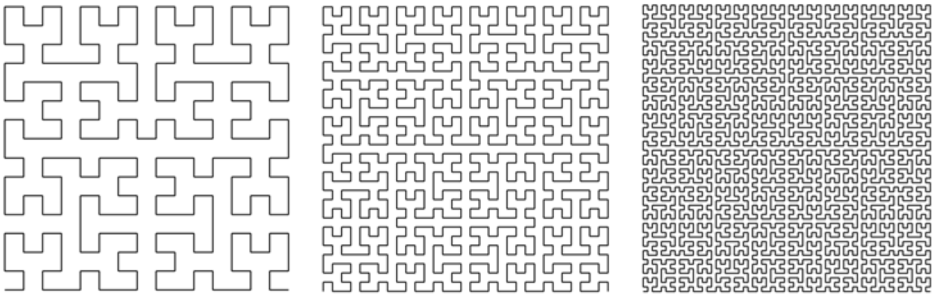

Figure 2.32: Three paraxial paths (stages in the construction of the Hilbert curve)

Path Covering. The following technical lemma will be useful in several proofs below. More importantly, it will be crucial for our discussion on analytic continuation of complex functions in Section 3.6. Roughly speaking the result says that the trace of a path in a domain D can be covered by a finite collection of well-overlapping disks in D as is illustrated in Figure 2.33. For later applications we need some additional requirements on the location of the centers and the covering properties of the disks; they are made precise in the following definition.

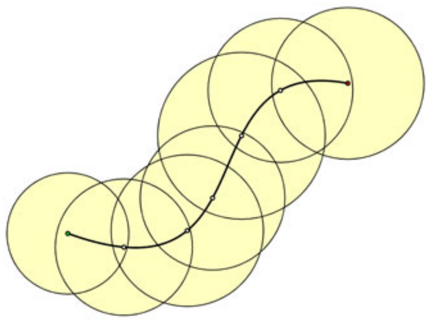

Figure 2.33: Covering a path by disks

Definition 2.7.7. Let $\gamma : [0,1] \to \mathbb{C}$ be a path in the complex plane. A *chain of disks covering* γ is a finite sequence $(D_0, D_1, \ldots, D_n)$ of open disks D_k with the following properties:

(i) There exists a partition $0 = t_0 < t_1 < \ldots < t_n = 1$ of the interval $[0,1]$ such that $\gamma(t_k)$ is the center of D_k for $k = 0, 1, \ldots, n$.

(ii) The section of γ between $\gamma(t_{k-1})$ and $\gamma(t_{k+1})$ is contained in D_k, more precisely,

$$\begin{aligned} &\gamma(t) \subset D_0, \quad t_0 \le t \le t_1, \\ &\gamma(t) \subset D_k, \quad t_{k-1} \le t \le t_{k+1}, \qquad (k = 1, \ldots, n-1), \\ &\gamma(t) \subset D_n, \quad t_{n-1} \le t \le t_n. \end{aligned}$$

Lemma 2.7.8 (Path Covering Lemma)**.** *Let γ be a path in the domain $D \subset \mathbb{C}$. Then there exists a chain of disks which is contained in D and covers γ. Moreover, the radii of all disks can be chosen to be of the same size and arbitrarily small.*

Proof. Since γ is continuous on the compact interval $[0,1]$, its trace $[\gamma]$ is a compact subset of D. The complement of D in $\mathbb{C}$ is closed, and hence the distance d between $[\gamma]$ and $\mathbb{C} \setminus D$ is positive. If $0 < r < d$, then all disks with radius r and centers on $[\gamma]$ are contained in D. Because γ is uniformly continuous, there exists a positive number δ such that $s,t \in [0,1]$ and $|s-t| < \delta$ imply that $|\gamma(s) - \gamma(t)| < r$. So all requirements are satisfied if the partition $0 = t_0 < t_1 < \ldots < t_n = 1$ is chosen such that $t_k - t_{k-1} < \delta$. $\square$

Homotopy. Next we sketch a powerful technique, which formalizes the intuitive concept of a *continuous deformation* of a path γ_0 into a path γ_1. The idea is to embed γ_0 and γ_1 into a family of paths $\gamma_s : t \mapsto h(s,t)$, which depend continuously on the parameter s. The technical term for this procedure is *homotopy.* We shall consider two situations: deformations of paths with fixed endpoints and, somewhat later, free deformations of *closed* paths (loops) without this restriction. To simplify notation we assume that all paths are defined on $[0,1]$, which can always be achieved by a linear reparametrization.

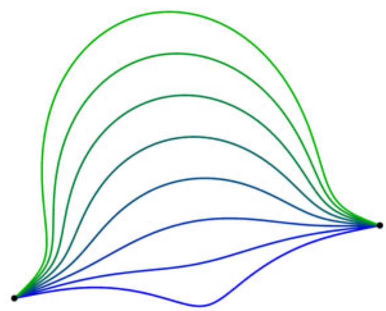

Figure 2.34: Homotopic paths

Definition 2.7.9. Two paths $\gamma_0 : [0,1] \to D$ and $\gamma_1 : [0,1] \to D$ from a to b are said to be *homotopic in D with fixed endpoints* if there exists a continuous function $h : [0,1] \times [0,1] \to D$ which satisfies the following conditions (i)-(iv):

$$\begin{array}{lll} \text{(i)} & h(0,t) = \gamma_0(t), & 0 \le t \le 1 \\ \text{(ii)} & h(1,t) = \gamma_1(t), & 0 \le t \le 1 \\ \text{(iii)} & h(s,0) = a, & 0 \le s \le 1 \\ \text{(iv)} & h(s,1) = b, & 0 \le s \le 1. \end{array}$$

The function h will be referred to as a *homotopy* from γ_0 to γ_1.

If there is no danger of confusion we shall just speak of homotopic paths without mentioning that the endpoints are fixed.

Figure 2.34 illustrates that the family (γ_s) of intermediate paths $\gamma_s := h(s,.)$ with $s \in [0,1]$ is a continuous deformation of γ_0 into γ_1. The traces of γ_s are colored according to the homotopy parameter s.

Note that the coincidence of the traces $[\gamma_0]$ and $[\gamma_1]$ does *not* guarantee that γ_0 and γ_1 are homotopic. For example the paths γ_0 and γ_1 in the domain $\mathbb{C} \setminus \{0\}$, defined by

$$\gamma_0(t) := \mathrm{e}^{2\pi \mathrm{i} t}, \quad \gamma_1(t) := \mathrm{e}^{4\pi \mathrm{i} t}, \qquad t \in [0,1],$$

have the same trace (the unit circle) though they are not homotopic in $\mathbb{C} \setminus \{0\}$.

It is not difficult to see that "being homotopic" is an equivalence relation. So the set of all paths in D with fixed endpoints a and b splits into classes of homotopic paths. Note also that reparametrization of a path does not change its homotopy class, no matter what the domain D is.

Definition 2.7.10. A path $\gamma_1 : [0,1] \to D$ is said to be a *reparametrization* of the path $\gamma_0 : [0,1] \to D$ if there exists a continuous function $\varphi : [0,1] \to [0,1]$ with $\varphi(0) = 0$ and $\varphi(1) = 1$ such that $\gamma_1 = \gamma_0 \circ \varphi$.

Lemma 2.7.11. *Let γ be a path in D. Then any reparametrization $\gamma \circ \varphi$ is homotopic to γ in D.*

Proof. The function h defined by $h(s,t) := \gamma\big((1-s)\,t + s\,\varphi(t)\big)$ maps $[0,1] \times [0,1]$ into D and is a homotopy from γ to $\gamma \circ \varphi$. $\square$

Homotopic Paths with Specific Properties. Technically it is of great importance that any path in D can be approximated by homotopic paths with specific properties. In particular, the next lemma shows that any class of homotopic paths contains a smooth and a paraxial path (see Definition 2.7.6).

Lemma 2.7.12. *Let $\gamma : [0,1] \to D$ be a path in an open set $D \subset \mathbb{C}$. Then there exist a smooth path $\widetilde{\gamma} : [0,1] \to D$ and a paraxial path $\widehat{\gamma} : [0,1] \to D$ which are homotopic to γ in D. For each positive ε both paths can be chosen such that*

$$\max_{t \in [0,1]} |\gamma(t) - \widetilde{\gamma}(t)| < \varepsilon, \qquad \max_{t \in [0,1]} |\gamma(t) - \widehat{\gamma}(t)| < \varepsilon.$$

Proof. By the path covering lemma (Lemma 2.7.8, see page 49 for the notation used in the following), γ can be covered by a sequence of disks D_k with radii less than $\varepsilon/2$. Let $0 = t_0 < t_1 < \ldots < t_n = 1$ be the subdivision of the parameter interval $[0,1]$, and denote by $z_k = \gamma(t_k)$ the centers of the covering disks D_k. Then the restriction γ_k of γ to $[t_{k-1}, t_k]$ is homotopic in D_k (and hence in D) to the line segment $[z_{k-1}, z_k]$ for all $k = 1, \ldots, n$. This induces a homotopy between γ and the polygonal path $\widehat{\gamma} := [z_0, z_1] \oplus \ldots \oplus [z_{n-1}, z_n]$. Smoothing the function $\widehat{\gamma}$ at the points t_k appropriately, we obtain a smooth path $\widetilde{\gamma}$ which is homotopic to $\widehat{\gamma}$ and hence to γ. Finally, the segments $[z_{k-1}, z_k]$ are homotopic in D_k to the sum $[z_{k-1}, \operatorname{Re} z_k + \mathrm{i} \operatorname{Im} z_{k-1}] \oplus [\operatorname{Re} z_k + \mathrm{i} \operatorname{Im} z_{k-1}, z_k]$ of two segments which are parallel to the real and imaginary axis, respectively. $\square$

Connectedness Revisited. The following result is sometimes used as a definition of connectedness for open sets. Note that the "only if" conclusion fails when D is not open.

Proposition 2.7.13. *An open set $D \subset \widehat{\mathbb{C}}$ is connected if and only if any two points a and b in D can be joined by a path γ in D.*

In conjunction with Lemma 2.7.12 we conclude that γ can be chosen to be smooth or paraxial.

Simply Connected Domains. The definition of homotopy implies that the initial points and the terminal points of two homotopic paths must coincide. In general the converse it not true, but the next definition singles out a class of domains for which the endpoint condition is also sufficient for the homotopy of paths.

Definition 2.7.14. A domain D is called *simply connected* if any two paths $\gamma_0 \subset D$ and $\gamma_1 \subset D$ with $\gamma_0(0) = \gamma_1(0)$ and $\gamma_0(1) = \gamma_1(1)$ are homotopic in D. Domains which are not simply connected are said to be *multiply connected*.

The Riemann sphere $\widehat{\mathbb{C}}$, the complex plane $\mathbb{C}$, and the unit disk $\mathbb{D}$ are simply connected, while the *punctured unit disk*

$$\dot{\mathbb{D}} := \{z \in \mathbb{C} : 0 < |z| < 1\},$$

and the ring domain R

$$R := \{z \in \mathbb{C} : 1/2 < |z| < 2\}$$

are not simply connected. The two paths γ_0 and γ_1 depicted in Figure 2.35 have the same initial and the same terminal point, but they are not homotopic in R.

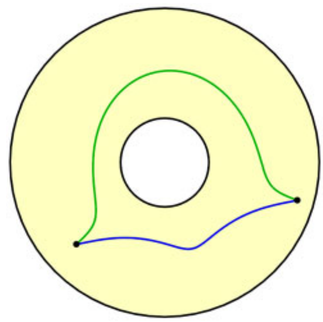

Figure 2.35: Non-homotopic paths

Homotopy of Loops. For closed paths there is a second notion of homotopy which is more general because it does not require that the endpoints be kept fixed.

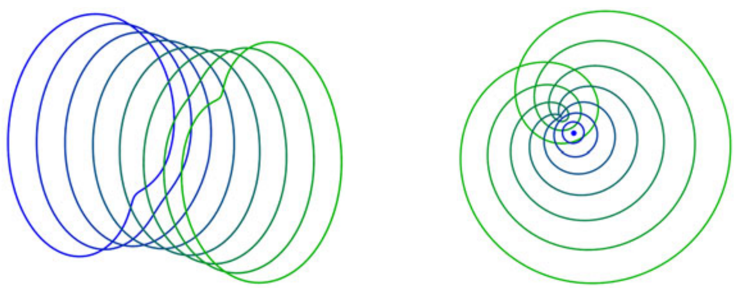

Figure 2.36: Traces of two families of freely homotopic closed paths

Definition 2.7.15. Two loops γ_0 and γ_1 in a set D are called *freely homotopic in D* if there exists a continuous function $h : [0,1] \times [0,1] \to D$ such that

(i) $h(0,t) = \gamma_0(t), \quad 0 \le t \le 1$

(ii) $h(1,t) = \gamma_1(t), \quad 0 \le t \le 1$

(iii) $h(s,0) = h(s,1), \quad 0 \le s \le 1.$

A loop which is freely homotopic in D to a constant path is said to be *null-homotopic* in D.

Figure 2.36 illustrates the definition. Both pictures visualize a freely homotopic family of closed path. The homotopy depicted on the right-hand side contracts the green curve to the blue point.

Lemma 2.7.16. *For any path γ in D the loop $\gamma \oplus \gamma^-$ is null-homotopic in D to its base point.*

Proof. We contract $\gamma \oplus \gamma^-$ along $[\gamma]$ to its base point $\gamma(0)$ as illustrated on the left of Figure 2.37. More formally, the function h defined for $s, t \in [0,1]$ by

$$h(s,t) = \begin{cases} \gamma(2st) & \text{if} \quad 0 \le t \le 1/2 \\ \gamma\big(2s(1-t)\big) & \text{if} \quad 1/2 < t \le 1 \end{cases}$$

is continuous on $[0,1] \times [0,1]$, its range is contained in $[\gamma]$, and hence in D, and it satisfies $h(0,.) = \gamma(0)$ and $h(1,.) = \gamma \oplus \gamma^-$. $\square$

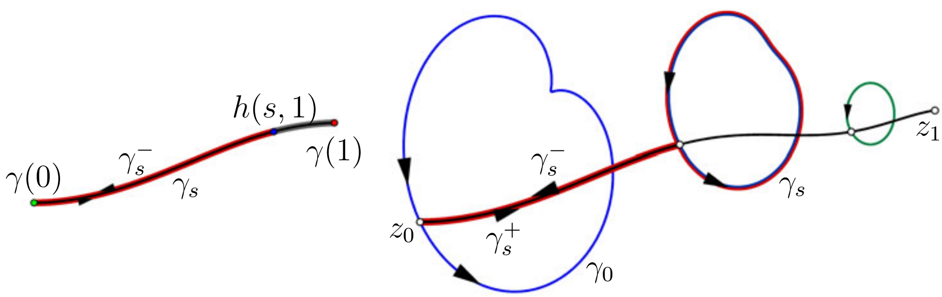

Figure 2.37: Illustrations to the proofs of Lemma 2.7.16 and Lemma 2.7.17

The following result shows that a null-homotopic loop can always be contracted to its base point.

Lemma 2.7.17. *If a loop with base point z_0 is null-homotopic in D, then it is also homotopic with fixed endpoints to the constant path z_0.*

Proof. Let h be a homotopy from the given path γ_0 to a point z_1. We define γ_s and γ_s^+ by

$$\gamma_s(t) := h(s,t), \quad \gamma_s^+(t) = h(st, 0), \qquad s, t \in [0,1],$$

and set $\gamma_s^- := (\gamma_s^+)^-$. Then the path γ_s^+ lies in D and connects z_0 with the moving base point $z_s := h(s,0) = h(s,1)$ of the loop γ_s, as shown in Figure 2.37 (right). The family of loops

$$\gamma_s^* := \gamma_s^+ \oplus \gamma_s \oplus \gamma_s^-$$

has fixed base point z_0 and all paths in this family are homotopic in D (Figure 2.37 shows the trace of one such path γ_s^* in red). Now γ_0 is homotopic to γ_0^*, γ_0^* is homotopic to γ_1^*, and γ_1^* is homotopic to $\gamma_1^+ \oplus \gamma_1^-$, and by Lemma 2.7.16 the latter is homotopic to the base point z_0. □

Lemma 2.7.18. *A domain is simply connected if and only if any loop in D is null-homotopic in D.*

Proof. 1. Assume that D is simply connected and $\gamma : [0,1] \to D$ is a loop. If γ_1 and γ_2 denote the restrictions of γ to the intervals $[0,1/2]$ and $[1/2,1]$, respectively, then $\gamma = \gamma_1 \oplus \gamma_2$. The paths γ_2 and γ_1^- have the same initial and the same terminal points and hence they are homotopic in D with fixed endpoints. This implies that γ and $\gamma_1 \oplus \gamma_1^-$ are homotopic in D (with fixed endpoints and also freely). Now apply Lemma 2.7.16.

2. Let D be a domain and suppose that any loop in D is null-homotopic. If γ_0 and γ_1 are paths with initial point a and terminal point b, then $\gamma := \gamma_0 \oplus \gamma_1^-$ is a loop with base point a and $\gamma_0 = \gamma \oplus \gamma_1$. By assumption and Lemma 2.7.16 γ is homotopic with fixed endpoints to the constant path a via a family of paths γ^s. This induces the homotopic family of paths $\gamma_s := \gamma^s \oplus \gamma_1$ from γ_0 to γ_1. □

Winding Numbers. We now introduce a geometric characteristic of loops which describes how many times they "wind around" some point in the plane. Among other applications, this concept will be essential for understanding the complex logarithm function. We begin with an auxiliary result.

Lemma 2.7.19. *Let $\gamma : [0,1] \to \mathbb{C} \setminus \{0\}$ be a path. Then there exist continuous functions $a : [0,1] \to \mathbb{R}$ and $r : [0,1] \to \mathbb{R}_+$, such that*

$$\gamma(t) = r(t)\,\mathrm{e}^{\mathrm{i}\,a(t)}. \tag{2.24}$$

Proof. The function $r(t) := |\gamma(t)|$ is continuous and positive. So the proof amounts to finding an appropriate argument $a(t)$ of $\gamma(t)$ such that $t \mapsto a(t)$ is continuous. For this purpose we use the path covering lemma with $D := \mathbb{C} \setminus \{0\}$ (see page 49 for notation).

At the initial point of γ we choose the principal branch of the argument, $a(0) := \operatorname{Arg}\gamma(0)$. If $t \in [t_0, t_1]$, all points $\gamma(t)$ lie in the disk D_0. Since D_0 does not contain the origin, it is contained in a sector with vertex at 0 and opening angle less than π. Consequently the argument $a(t) = \arg\gamma(t)$ can be chosen such that $|a(t) - a(0)| < \pi/2$, which yields a continuous function a on $[0, t_1]$.

Suppose that such a function has already been constructed on some interval $[0, t_k]$. Then it can be prolongated to $[0, t_{k+1}]$ by choosing $a(t) = \arg\gamma(t)$ on $[t_k, t_{k+1}]$ such that $|a(t) - a(t_k)| < \pi/2$, which is possible since $\gamma(t) \in D_k$ and $0 \notin D_k$. By induction, a can be extended to all of $[0,1]$. □

Any continuous function a satisfying (2.24) is called a *continuous branch* of the argument along the path γ. The difference of two such functions a_1 and a_2 on $[0,1]$ is a constant integral multiple of 2π.

If a is a continuous branch of the argument along a loop, then $a(1) - a(0)$ is an integral multiple of 2π which does not depend on the special choice of the branch a.

Definition 2.7.20. Let γ be a loop in $\mathbb{C} \setminus \{0\}$ and denote by a a continuous branch of the argument along γ. Then the integer

$$\operatorname{wind} \gamma := \frac{1}{2\pi} \big(a(1) - a(0) \big)$$

is called the *winding number* (or *index*) of γ. If $z_0 \in \mathbb{C}$ and γ is a loop in $\mathbb{C} \setminus \{z_0\}$, the *winding number of γ about z_0* is defined by

$$\operatorname{wind}(\gamma, z_0) := \operatorname{wind}(\gamma - z_0).$$

Figure 2.38 shows the trace of a loop with winding number 1 about the origin (marked by a black dot). The coloring visualizes the winding number of γ about points z_0 in the different components of $\mathbb{C} \setminus [\gamma]$.

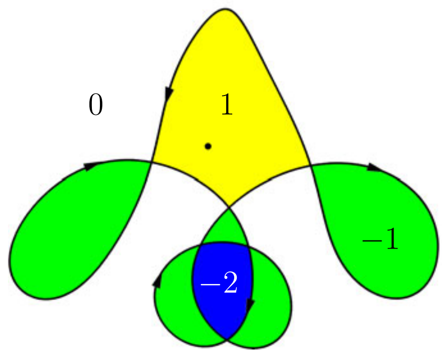

Figure 2.38: Winding numbers $\operatorname{wind}(\gamma, z_0)$

Stability of Winding Numbers. Next we prove the intuitive fact that small perturbations of a loop do not change its winding number. In order to obtain a quantitative version of this result, we measure the distance between two paths γ_0 and γ defined on $[0,1]$ by

$$\|\gamma - \gamma_0\| := \max_{t \in [0,1]} |\gamma(t) - \gamma_0(t)|.$$

Lemma 2.7.21. *Let $\gamma_0 : [0,1] \to\in \mathbb{C} \setminus \{0\}$ be a loop, and denote by d the distance of its trace $[\gamma_0]$ from the origin. Then for all loops $\gamma : [0,1] \to \mathbb{C}$ with $\|\gamma - \gamma_0\| < d$*

$$\operatorname{wind} \gamma = \operatorname{wind} \gamma_0.$$

Proof. Since $[\gamma_0]$ is a compact subset of $\mathbb{C} \setminus \{0\}$, its distance d from the origin is positive. Then $|\gamma(t) - \gamma_0(t)| < d$ implies that $\gamma(t)/\gamma_0(t)$ lies in the right half-plane.

Let a_0 be a continuous branch of the argument along γ_0. If we chose a continuous branch of the argument along γ such that $|a(0) - a_0(0)| < \pi/2$, then $|a(t) - a_0(t)| < \pi/2$ for all $t \in [0,1]$. Invoking the triangle inequality we see that $\big|\big(a(1) - a(0)\big) - \big(a_0(1) - a_0(0)\big)\big| < \pi$, and since this number is an integral multiple of 2π it must be zero. $\square$

Homotopic Loops in Punctured Domains. The next result describes a situation where the winding number alone allows one to decide if two loops are homotopic.

Lemma 2.7.22. *Let $D \subset \mathbb{C}$ be a simply connected domain and $z_0 \in D$. Then two loops γ_0 and γ_1 are homotopic in the punctured domain $D \setminus \{z_0\}$ if and only if they have the same winding number about z_0.*

Proof. That homotopic loops have the same winding number follows easily from Lemma 2.7.21 and a compactness argument. Though the converse is plausible, its proof is non-trivial and we refer to the literature (for instance Hatcher [24]). □

Jordan Arcs and Jordan Curves. Next we turn our attention to geometric properties of traces of simple paths. This is one of the rare occasion where we resort to the notion of a curve.

Definition 2.7.23. A *Jordan curve* is the trace of a simple loop, the trace of a simple non-closed path is called a *Jordan arc*.

In the definition above we have avoided specifying the target set of the loop γ. Usually we assume that γ maps into $\mathbb{C}$, but occasionally we also admit more generally that $\gamma : [\alpha, \beta] \to \widehat{\mathbb{C}}$, in which case we speak of a Jordan curve on the Riemann sphere. The minor modifications which are needed in the formulation of the corresponding results are left to the reader.

As geometric objects, Jordan curves are not that tame as one might expect from their definition. Indeed their menagerie hosts such badly behaved entities like the fractal *Koch curve* (the boundary of the Koch snowflake) and the *Osgood curve* which has a positive area (see Sagan [60]). So it is not surprising that the following celebrated theorem is not easy to prove.

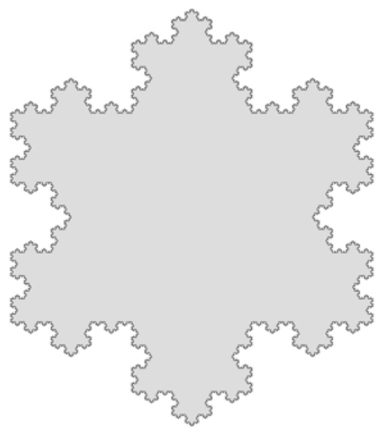

Figure 2.39: The Koch snowflake

Proposition 2.7.24 (Jordan Curve Theorem)**.** *Let J be a Jordan curve in $\mathbb{C}$. Then its complement $\mathbb{C} \setminus J$ is the union of two domains, one of which is bounded and simply connected (the* interior *of J), and the other one is unbounded (the* exterior *of J). The Jordan curve J is the common boundary of its interior and its exterior.*

A relatively short and elementary proof is given by Maehara [39], another instructive proof can be found in Henle [25]. For an account of the history of the theorem we refer to Hales [23].

We denote the interior and the exterior of J by $\operatorname{int} J$ and $\operatorname{ext} J$, respectively. If γ is a simple loop which parameterizes J, then its winding number $\operatorname{wind}(\gamma, z_0)$ about z_0 is well defined for every point $z_0 \in \mathbb{C} \setminus J$ and we have

$$\operatorname{wind}(\gamma, z_0) = \pm 1 \quad \text{if} \quad z_0 \in \operatorname{int} J, \qquad \operatorname{wind}(\gamma, z_0) = 0 \quad \text{if} \quad z_0 \in \operatorname{ext} J, \tag{2.25}$$

with one and the same sign in the first equality for all $z_0 \in \operatorname{int} J$.

Orientation. We say that γ is *positively oriented*, if $\mathrm{wind}\,(\gamma, z_0) = 1$ for (one and then for all) $z_0 \in \mathrm{int}\, J$. In this case we shall also call the Jordan curve J parameterized by γ *positively oriented.*

Example 2.7.1 (Triangles). Let z_1, z_2, z_3 be three (not necessarily distinct) points in the complex plane. The *triangle* $\triangle := \triangle(z_1, z_2, z_3)$ is the (closed) convex hull of z_1, z_2, z_3, and

$$\gamma := [z_1, z_2] \oplus [z_2, z_3] \oplus [z_3, z_1]$$

is a path along its boundary $\partial\triangle$. If the interior $\mathrm{int}\,\triangle$ of $\triangle$ is not empty, we pick a point $z_0 \in \mathrm{int}\,\triangle$ and define the *standard parametrization* of $\partial\triangle$ as

$$\delta := \begin{cases} \gamma & \text{if} \quad \mathrm{wind}\,(\gamma, z_0) > 0 \\ \gamma^- & \text{if} \quad \mathrm{wind}\,(\gamma, z_0) < 0. \end{cases}$$

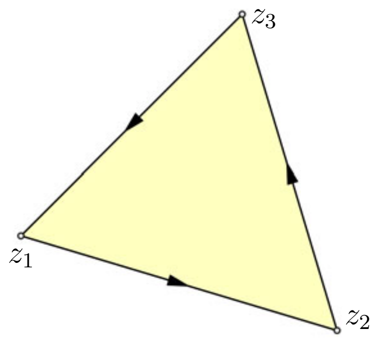

Figure 2.40: A Triangle $\triangle(z_1, z_2, z_3)$

Note that δ is a positively oriented parametrization of $\partial\triangle$.

Jordan Domains. A *Jordan domain* is the interior of a Jordan curve. It is clear that the closure of a Jordan domain is a compact set, but in fact it has a much nicer property.

Proposition 2.7.25. *The closure of a Jordan domain is homeomorphic to the closed unit disk.*

Spelled out explicitly, this means that there is a continuous bijective mapping φ of $\overline{\mathbb{D}}$ onto $K := \mathrm{int}\, J \cup J$ with continuous inverse. Since such a mapping sends interior points to interior points, and boundary points to boundary points, the images $\varphi(\mathbb{D})$ and $\varphi(\mathbb{T})$ of the unit disk and the unit circle are the Jordan domain $\mathrm{int}\, J$ and the Jordan curve J, respectively. In fact Jordan curves are precisely the homeomorphic images of circles.

Contractability. The last result in this section tells us that every Jordan curve is “contractible” to any point in its interior. The precise meaning of this statement is made clear in the following result.

Lemma 2.7.26. *Let γ be a parametrization of a Jordan curve J and let $z_0 \in \mathrm{int}\, J$. Then γ is null-homotopic in $K := \mathrm{int}\, J \cup J$ to the constant path z_0.*

Proof. Let $\varphi : \overline{\mathbb{D}} \to K$ be a homeomorphism between the closed unit disk and K according to Proposition 2.7.25. Then $\gamma_0 := \varphi^{-1} \circ \gamma$ is a loop which parameterizes the unit circle $\mathbb{T}$. The homotopy

$$h(s, t) := (1 - s)\, \gamma_0(t) + s\, w_0, \qquad s, t \in [0, 1]$$

contracts γ_0 in the (convex) closed unit disk $\overline{\mathbb{D}}$ to the point $w_0 := \varphi^{-1}(z_0)$. Since φ is a homeomorphism, $\varphi \circ h$ is a homotopy of γ to the point z_0 in K. $\square$

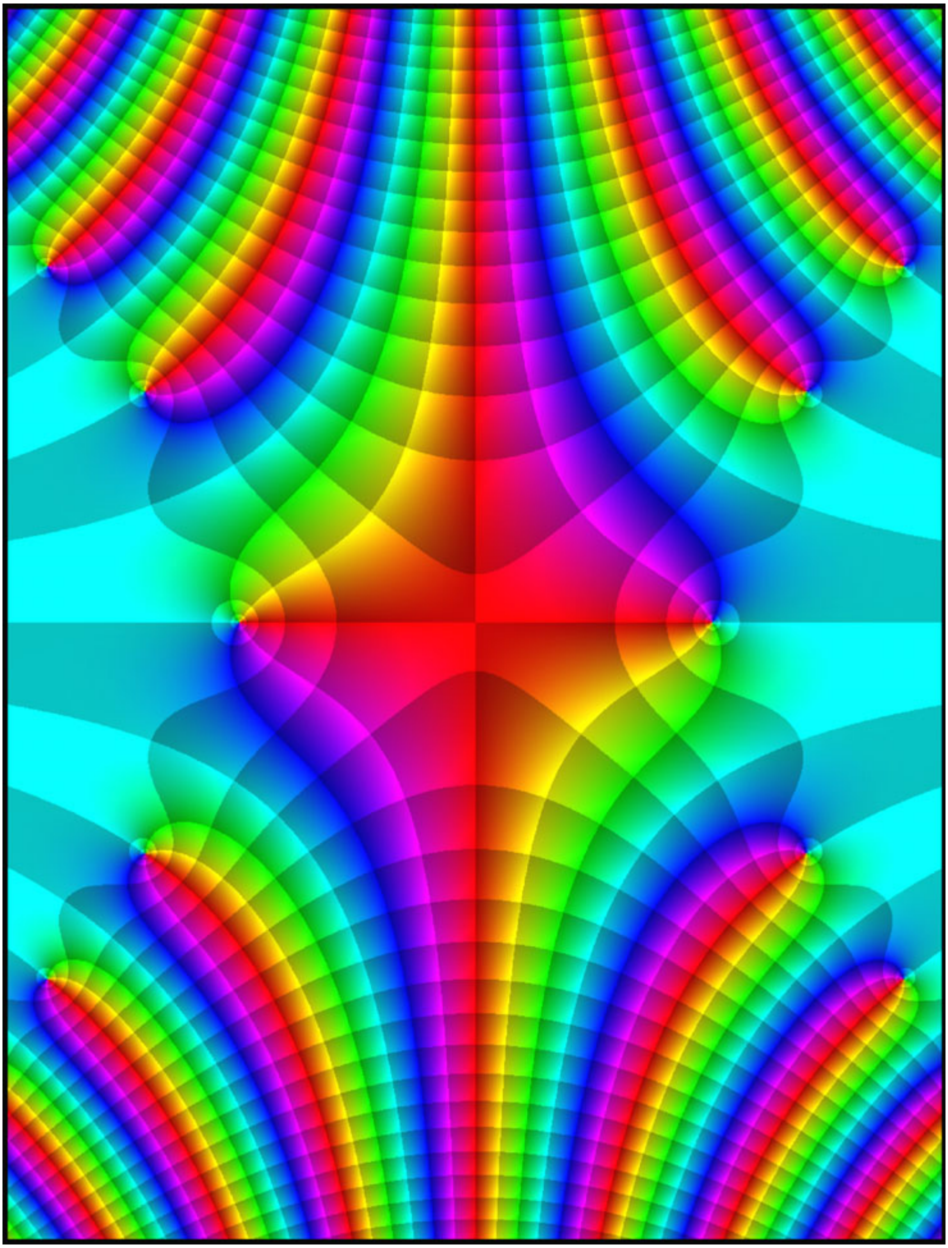

Figure 3.1: Enhanced phase portrait of the function $f(z) = \mathrm{e}^{-z^2} - 1$

http://www.springer.com/978-3-0348-0179-9

Visual Complex Functions
An Introduction with Phase Portraits
Wegert, E.
2012, XIV, 360 p. 231 illus., 202 illus. in color.,
Softcover
ISBN: 978-3-0348-0179-9
A product of Birkhäuser Basel