1. Course outline

- Möbius transformations
- Elementary functions of complex variable
- Differential 1-forms on the plane and their integrals
- Complex differentiability, holomorphic functions
- The Cauchy integral formula
- General properties of holomorphic functions
- Harmonic functions and the Poisson formula
- Normal families and Montel's theorem
- The Riemann mapping theorem
- Multivalued analytic functions
- Riemann surfaces and uniformization

2. Background

To make the these lecture notes self-contained, I will provide all necessary terminology and notation here. I will also state (without proof) some basic facts from algebra, analysis and point-set topology that we will use.

2.1. Sets and maps.

We will write $x \in A$ if x is an element of a set A. Let $\varnothing$ denote the empty set that contains no elements. For two sets A and B, the *intersection* $A \cap B$ is by definition the set consisting of elements x such that $x \in A$ AND $x \in B$; the *union* $A \cup B$ is by definition the set consisting of elements x such that $x \in A$ OR $x \in B$. The *difference* $A \setminus B$ is the set of elements x such that $x \in A$ but $x \notin B$. The *Cartesian product* $A \times B$ is defined as the set of all ordered pairs (a, b), where $a \in A$ and $b \in B$ (by definition, two ordered pairs (x, y) and (x', y') are regarded the same if and only if $x = x'$ and $y = y'$).

A *map* $f : A \to B$ is a subset of $A \times B$ that contains exactly one pair (x, y) for every $x \in A$. Then we write $y = f(x)$. The map f is said to be *bijective* if, for every $y \in B$, there is a unique $x \in A$ such that $y = f(x)$. A bijective map of a set onto itself is called a *transformation* of this set. The map f is said to be *injective* if $f(x) = f(x')$ implies $x = x'$; and f is said to be *surjective* if, for every $y \in B$, there is some $x \in A$ with $y = f(x)$. Clearly, bijective maps are those, which are simultaneously injective and surjective. If $y = f(x)$, then we say that f *takes* x to y, or that f *maps* x to y. If $f : A \to B$ and $g : B \to C$ are two maps, then we define the *composition* $g \circ f$ as the map that takes every element $x \in A$ to $g(f(x))$. The expression $f^{\circ 2}$ means $f \circ f$, the expression $f^{\circ 3}$ means $f \circ f \circ f$, etc. For a subset $X \subset A$, we let $f(X)$ denote the set of points $f(x)$, where x runs through X. The set $f(A)$

is called the *image* of f. If f is a transformation of a set X, then we let f^{-1} denote the *inverse transformation* that takes element $y \in X$ to the unique element $x \in X$ such that $f(x) = y$. It is easy to see that

$$f \circ f^{-1} = f^{-1} \circ f = id_X,$$

where id_X denotes the *identity transformation* of X that maps any point of X to itself.

Let $f : A \to A$ be a self-map of some set A. A point $a \in A$ is *fixed* under f is $f(a) = a$ (we also say that f *fixes* the point a).

2.2. **Complex numbers.** We will write $\mathbb{R}$ for the set of real numbers and $\mathbb{C}$ for the set of complex numbers. Recall that complex numbers are formal expressions $x + iy$, where x and y are reals. Of course, the number $x+iy$ can be identified with the ordered pair (x, y), hence also with the point of the plane $\mathbb{R}^2$, whose Cartesian coordinates are x and y. The addition and the multiplication of complex numbers are defined by the formulas

$$(x + iy) + (x' + iy') = (x + x') + i(y + y')$$
$$(x + iy)(x' + iy') = (xx' - yy') + i(xy' + x'y).$$

We identify complex numbers of the form $x + i0$ with real numbers x. Having this identification in mind and using the rules for the addition and the multiplication of complex numbers, we can interpret any complex number $x + iy$ as the sum of a real number x and a certain complex number i multiplied by a real number y. This number i has the property that $i^2 = -1$. In practice, computations with complex numbers are performed by using standard rules of algebra and the relation $i^2 = -1$. For example, to multiply two complex numbers $x + iy$ and $x'+iy'$, it suffices to open the parenthesis as usual and then employ the relation $i^2 = -1$. The *modulus* of a complex number $z = x + iy$ is defined as $|z| = \sqrt{x^2 + y^2}$. If z and w are two complex numbers, then we have $|zw| = |z| \cdot |w|$. The number x is called the *real part* of the complex number z, and is denoted by $\mathrm{Re}(z)$; the number y is called the *imaginary part* of z and is denoted by $\mathrm{Im}(z)$. The number $\overline{z} = x - iy$ is said to be *complex conjugate* to z. We have

$$z\overline{z} = |z|^2, \quad \mathrm{Re}(z) = \frac{1}{2}\left(z + \overline{z}\right), \quad \mathrm{Im}(z) = \frac{1}{2i}\left(z - \overline{z}\right).$$

Division by any nonzero complex number is well-defined. To divide complex numbers, it suffices to know how to find $z^{-1} = 1/z$ for any $z \neq 0$, and the latter is easy: $z^{-1} = \overline{z}/|z|^2$ (in the right-hand side, we divide by the real number $|z|^2$, i.e., multiply by the real number $|z|^{-2}$). We will discuss further properties of complex numbers later.

2.3. **Groups.** A *transformation group* G of a set X is a set of transformations of X with the following properties: $f \circ g \in G$ whenever f, $g \in G$, and $f^{-1} \in G$ for all $f \in G$. It follows that $id_X = f^{-1} \circ f \in G$. The notion of a (abstract) *group* generalizes that of a transformation group. A group is a set G endowed with a binary operation, i.e., a map $G \times G \to G$, often denoted multiplicatively (in the same way as the multiplication), with the following properties:

- the group operation is *associative*: $(fg)h = f(gh)$ for all f, g, $h \in G$;
- there exists an *identity element* $1 \in G$ such that $1f = f1 = f$ for all $f \in G$;
- for every $f \in G$, there is an element f^{-1}, called an *inverse* of f, with the property $ff^{-1} = f^{-1}f = 1$.

It easily follows that the identity element is unique, and that any element of G has a unique inverse.

Let us give one important example of a group. A 2×2 *matrix* is a table

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

consisting of four numbers a_{ij}, where $i = 1, 2$ and $j = 1, 2$. The *product of* 2×2 *matrices* is defined by the formula

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{12} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}.$$

The *determinant* of a matrix is defined by the formula

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

It is easy to check that $\det(AB) = \det(A)\det(B)$ for any two matrices A, B. We let $\mathrm{GL}(2, \mathbb{C})$ denote the set of all 2×2 matrices A with complex entries such that $\det(A) \neq 0$ (here GL stands for "General Linear group"). This is a group under matrix multiplication. The identity element of this group is the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

called the *identity matrix*. The inverse matrix is given by the formula

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

2.4. **Vector spaces.** A *vector space* over $\mathbb{R}$ is defined as a set V, whose elements are called *vectors*, equipped with two operations: addition of vectors, and multiplication of a vector by a real number. We will write $u+v$ for the sum of two vectors u and v, and λv for the product of a number λ and a vector v. The two operations must satisfy the following properties:

$$u+v=v+u, \quad (u+v)+w=u+(v+w), \quad \lambda(u+v)=\lambda u+\lambda v,$$
$$(\lambda+\mu)v=\lambda v+\mu v, \quad \lambda(\mu v)=(\lambda\mu)v, \quad 1v=v, \quad 0u=0v$$

for all u, v, $w \in V$ and all λ, $\mu \in \mathbb{R}$. The vector $0u$ (which is independent of u) is called the *zero vector*, and we will simply write 0 for the zero vector. For every $v \in V$, we have $v+0=(1+0)v=v$. We write $-v$ for $(-1)v$, and we have $v+(-v)=(1+(-1))v=0$. Vector spaces over $\mathbb{C}$ (also called *complex vector spaces*) are defined in the same way, except that we use complex numbers instead of real numbers. The numbers (real numbers if we deal with vector spaces over $\mathbb{R}$ and complex numbers if we deal with vector spaces over $\mathbb{C}$) are also called *scalars*. Thus vectors can be added and multiplied by scalars. Vectors $v_1, \ldots, v_n$ in a real or complex vector space V are said to be *linearly independent* if $\lambda_1 v_1+\cdots+\lambda_n v_n=0$ implies $\lambda_1=\cdots=\lambda_n=0$. We say that V *has dimension* d if there d linearly independent vectors in V but no more than d.

As an example, consider a two-dimensional vector space V over $\mathbb{R}$. Choose a basis e_1, e_2 in V. Then any vector $x \in V$ can be written as $x=x_1e_1+x_2e_2$ for some coefficients x_1, $x_2 \in \mathbb{R}$ called the *coordinates* of x in the basis e_1, e_2. Thus V can be identified with the set $\mathbb{R}^2$ of pairs (x_1, x_2). Note that the vector addition and multiplication by scalars are performed coordinate-wise. Of course, this generalizes to vector spaces of any dimension: if $e_1, \ldots, e_n$ is a basis of a vector space V, then any vector $x \in V$ can be written in the form

$$x=\sum_{j=1}^{n} x_j e_j$$

for a unique n-tuple of numbers $(x_1, \ldots, x_n)$ called the *coordinates* of the vector x in the basis $e_1, \ldots, e_n$.

2.5. **Linear maps.** Let V be a vector space over $\mathbb{R}$ or over $\mathbb{C}$. A map $f: V \to V$ is said to be *linear* if, for any pair of vectors u, $v \in V$ and for any pair of scalars λ, μ, we have

$$f(\lambda u+\mu v)=\lambda f(u)+\mu f(v).$$

Suppose that $e_1, \ldots, e_n$ is a basis of V. Then f can be written in the form

$$f(x) = \sum_{j,k=1}^{n} a_{jk} x_k e_j,$$

where $x_1, \ldots, x_n$ are the coordinates. The table consisting of numbers a_{jk} (in which the number a_{jk} is places at the intersection of the j-th row and the k-th column) is called the *matrix* of the linear map f in the basis $e_1, \ldots, e_n$.

We have discussed 2×2 matrices, in particular, we defined the product of two such matrices. In general, the product AB of two $n \times n$ matrices A and B is defined as the matrix of the linear map obtained as the composition of the linear maps with matrices A and B. It is straightforward to check that this definition is consistent with our earlier formula for the product of two 2×2 matrices.

A matrix A is called *invertible* if there exists a matrix A such that $AA^{-1} = A^{-1}A = E$, where E is the *identity matrix*, the matrix of the identity transformation. E.g. we have seen that a 2 by 2 matrix A is invertible if and only if $\det(A) \neq 0$. This statement can be generalized for square matrices of arbitrary size. A matrix B is said to be *conjugate* to a matrix A if B can be represented in the form CAC^{-1} for some invertible matrix C. Conjugate matrices are matrices of the same linear map with respect to different bases.

The *Jordan decomposition theorem* allows one to reduce a matrix to some rather simple form by a conjugation. We will state this theorem in the case of 2×2 matrices: any complex 2×2 matrix A is conjugate to a matrix of one of the following two forms:

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \quad \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_1 \end{pmatrix}.$$

Here λ_1 (and also λ_2 in the former case) is called an *eigenvalue* of A.

2.6. **Metric spaces.** We will identify *the plane* $\mathbb{R}^2$ with the set $\mathbb{R} \times \mathbb{R}$, and *the 3-space* $\mathbb{R}^3$ with $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$. A point (x, y) of $\mathbb{R}^2$ is said to *have coordinates* x and y, similarly for $\mathbb{R}^3$. A *metric* on a set A is a function $d : A \times A \to \mathbb{R}$ such that

$$d(x, y) = d(x, y), \quad d(x, z) \leqslant d(x, y) + d(y, z), \quad d(x, x) = 0$$

for all x, $y \in A$ and such that $d(x, y) > 0$ for $x \neq y$. For example, the metric on $\mathbb{R}$ is defined by the formula $d(x, y) = |x - y|$, and the metric on $\mathbb{R}^2$ is defined by the formula

$$d((x, y), (x', y')) = \sqrt{(x' - x)^2 + (y' - y)^2}.$$

A set A equipped with a particular metric is called a *metric space*.

The following will be main examples of metric spaces for us.

1. The plane $\mathbb{C}$ of complex numbers with the Euclidean metric $d(z, w) = |z - w|$.

2. Any subset $A \subset \mathbb{C}$ is equipped with the so called *induced Euclidean metric*, which is just the restriction of the Euclidean metric to $A \times A$.

3. The extended plane $\overline{\mathbb{C}}$ with the *spherical metric* (the definition will be given later). Similarly, to example 2, we can equip any subset $A \subset \overline{\mathbb{C}}$ with the induced spherical metric.

A *ball* $B_\varepsilon(a)$ of radius $r > 0$ in a metric space A centered at $a \in A$ is defined as the set of all points $x \in A$ such that $d(x, a) < r$. A subset $U \subset A$ is said to be *open* if U is a union of balls. Equivalently, U is open if and only if, for every $a \in U$, there is $\varepsilon > 0$ with the property $B_\varepsilon(a) \subset U$. Clearly, the intersection of finitely many open sets, and the union of any collection of open sets are also open sets. A *neighborhood* of a point $x \in A$ is defined as any subset of A containing some ball $B_r(a)$ with $r > 0$. An open set can alternatively be defined as a set that, together with any its point, contains some neighborhood of it. Note that, if $A \subset \mathbb{C}$ is equipped with the induced metric, then open subsets of A may not be open in $\mathbb{C}$. A *closed subset* of A is defined as a subset $X \subset A$ such that $X \setminus A$ is open.

Given a subset $X \subset A$, we define

- the *interior of* X in A as the set of points $x \in X$ such that some neighborhood of x is contained in X;
- the *closure* $\overline{X}$ *of* X as the intersection of all closed sets containing X; alternatively, $\overline{X}$ can be defined as the complement of the interior of $A \setminus X$;
- the *boundary* ∂X *of* X as the set of points $x \in A$ with the following property: arbitrary neighborhood of x intersects both X and $A \setminus X$; alternatively, ∂X is the complement of the interior of X in $\overline{X}$.

Let X and Y be metric spaces, $U \subset X$ be an open subset, $a \in U$ be a point, and $f : U \setminus \{a\} \to Y$ be a function. We say that the point $b \in Y$ is the *limit of the function* f at the point a and write $\lim_{x \to a} f(x) = b$ if for every $\varepsilon > 0$ there exists $\delta > 0$ with the property that $0 < d_X(x, a) < \delta$ implies $d_Y(f(x), b) < \varepsilon$. Here the metrics in X and Y are denoted respectively by d_X and d_Y. The function f (defined on some neighborhood of a) is said to be *continuous* at the point a if $f(a) = \lim_{x \to a} f(x)$.

A metric space X is said to be *connected* if X cannot be represented as a union of disjoint nonempty open subsets. The following statements will be used without proofs (proofs can be found in any standard introductory topology textbook): any connected space X can be uniquely represented as a union of connected spaces called *components* of X (there may be finitely many or infinitely many components). Any component of any open subset of $\mathbb{C}$ is open. More generally, a metric space X is said to be *locally connected* if any point $x \in X$ admits a nested sequence of connected neighborhoods such that any neighborhood of x contains some element of this sequence. Any component of any open subset of a locally connected space is open. An open connected subset of $\mathbb{C}$ is said to be a *domain* in $\mathbb{C}$.

3. Möbius transformations

Consider the three-dimensional Euclidean space $\mathbb{R}^3$. We will fix a Cartesian coordinate system in $\mathbb{R}^3$, in which the distance between two points $\mathbf{x} = (x, y, z)$ and $\mathbf{x}' = (x', y', z')$ is given by the formula

$$|\mathbf{x} - \mathbf{x}'| = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}.$$

A *sphere* $S_R(\mathbf{a})$ *in* $\mathbb{R}^3$ *of radius* $R > 0$ *centered at a point* $\mathbf{a} \in \mathbb{R}^3$ is by definition the locus of points $\mathbf{x}$ such that $|\mathbf{x}-\mathbf{a}| = R$. If $\mathbf{a} = (a, b, c)$, then the equation of this sphere in coordinates (x, y, z) is

$$(x - a)^2 + (y - b)^2 + (z - c)^2 = R.$$

Opening the parentheses, we see that the equation has the form

$$x^2 + y^2 + z^2 + p_1x + p_2y + p_3z + p_4 = 0.$$

In particular, the equation of any sphere contains the part $x^2+y^2+z^2$. If p_1, p_2, p_3 and p_4 are arbitrary real numbers, then the equation displayed above may describe a sphere, or a point, or the empty set. It is often convenient to include the last two cases into the definition of a sphere (thus a sphere can also be a single point or the empty set). Equivalently, we allow the radius to be zero or even negative. Although all spheres of negative radius are the same (they are all empty), it makes sense to distinguish spheres given by different equations. E.g. a sphere of radius -1 should be thought of being different from a sphere of radius -2. The next step is to rewrite the equation of a sphere in the homogeneous form:

$$p_0(x^2 + y^2 + z^2) + p_1x + p_2y + p_3z + p_4 = 0.$$

We now have 5 parameters in the equation of a sphere: $p_0, \ldots, p_4$. However, these parameters are only defined up to proportionality: if we multiply all of them by the same number, then the new equation will describe the same sphere as the old equation. Note that if we set $p_0 = 0$, then we obtain an equation of an affine plane. It will also be convenient to think of planes as being spheres (of infinite radius). To summarize, we give the following definition:

Definition 3.1 (Spheres). We define a sphere as a subset of $\mathbb{R}^3$ given by the equation

$$p_0(x^2 + y^2 + z^2) + p_1 x + p_2 y + p_3 z + p_4 = 0$$

together with a particular choice of the equation. Thus, a sphere is determined by 5 numbers $p_0, \ldots, p_5$ called the *homogeneous coordinates* of the sphere. We always assume that homogeneous coordinates do not vanish simultaneously. Two spheres are thought to be the same if and only if the homogeneous coordinates of one sphere are obtained from the corresponding homogeneous coordinates of the other sphere by multiplication by the same factor.

We define a *circle* as the intersection of two different spheres. Thus a line, a point and the empty set are examples of circles. It is easy to see that any circle can be obtained as the intersection of a sphere and a plane. Indeed, a system of equations

$$x^2 + y^2 + z^2 + p_1 x + p_2 y + p_3 z + p_4 = 0,$$
$$q_0(x^2 + y^2 + z^2) + q_1 x + q_2 y + q_3 z + q_4 = 0$$

is equivalent to the system

$$x^2 + y^2 + z^2 + p_1 x + p_2 y + p_3 z + p_4 = 0,$$
$$(q_1 - q_0 p_1)x + (q_2 - q_0 p_2)y + (q_3 - q_0 p_3)z + (q_4 - q_0 p_4) = 0.$$

We now consider the *unit sphere* S in $\mathbb{R}^3$, i.e., the sphere of radius 1 centered at the origin. The following definition is due to Möbius:

Definition 3.2 (Möbius transformation). Define a *Möbius transformation* of S as a bijective map $f : S \to S$ that takes circles to circles.

The problem we address in this section is to give an explicit description of all Möbius transformations. Most textbooks in complex analysis reverse the history and introduce the solution of this challenging problem as a definition of Möbius transformations. However, since I personally like the problem, I would like not to make this shortcut. I will sketch a solution leaving some details as (useful) exercises. Another advantage of discussing this problem of Möbius is that complex numbers appear naturally in the solution, although there are no complex numbers in the statement.

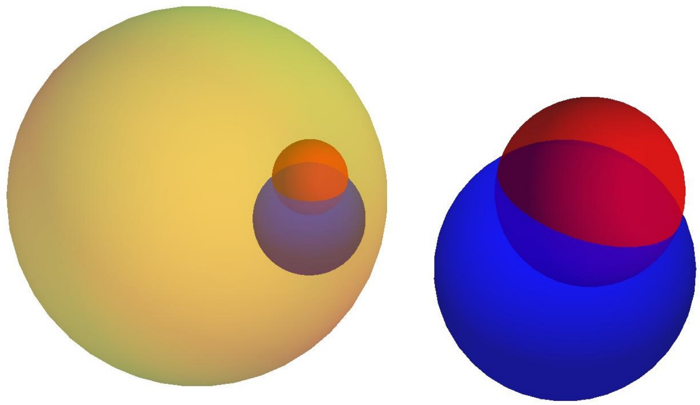


FIGURE 1. The blue and the red spheres are reflected in the surface of the yellow sphere. The reflection is the same as the inversion.

We start with examples of Möbius transformations. An obvious example is provided by rotations of the sphere. Clearly, rotations are bijective maps, and they map circles to circles. Are there Möbius transformations different from rotations? The answer is YES.

To give examples, we discuss a nice geometric construction. Fix a point $\mathbf{a} \in \mathbb{R}^3$ and a sphere $S_R(\mathbf{a})$ around $\mathbf{a}$ of radius R. For any point $\mathbf{x} \in \mathbb{R}^3$ different from $\mathbf{a}$, we let $\mathbf{x}'$ denote the unique point with the following properties: $\mathbf{x}'$ lies on the straight ray emanating from $\mathbf{a}$ and containing $\mathbf{x}$, and we have

$$|\mathbf{x}' - \mathbf{a}'| \cdot |\mathbf{x} - \mathbf{a}| = R^2.$$

The map $\mathbf{x} \mapsto \mathbf{x}'$ is defined on $\mathbb{R}^3 \backslash \{\mathbf{a}\}$. This map is called the *inversion* in the sphere $S_R(\mathbf{a})$. In fact, if we add to $\mathbb{R}^3$ one point at infinity, then we can extend the inversion to this point by saying that the infinity maps to $\mathbf{a}$. This is justified by the fact that far away points map close to $\mathbf{a}$ under the inversion.

As an example, let us discuss the inversion in the unit sphere S. If $\mathbf{x} = (x, y, z)$ is any point different from the origin, and $\mathbf{x}' = (x', y', z')$ is the image of this point under the inversion, then we have

$$x' = \frac{x}{x^2 + y^2 + z^2}, \quad y' = \frac{y}{x^2 + y^2 + z^2}, \quad z' = \frac{z}{x^2 + y^2 + z^2}.$$

Suppose that a point $\mathbf{x}$ satisfies the equation

$$p_0(x^2 + y^2 + z^2) + p_1 x + p_2 y + p_3 z + p_4 = 0.$$

Then the point $\mathbf{x}'$ satisfies the equation

$$p_4(x^2 + y^2 + z^2) + p_1 x + p_2 y + p_3 z + p_0 = 0.$$

We see that the image of any sphere under an inversion is a subset of a sphere. Since a circle can be represented as the intersection of two spheres, it follows that an inversion maps circles to subsets of circles. We would like to make a stronger statement that an inversion maps spheres *onto* spheres. The problem with this statement is that, for example, the image of a plane under an inversion is not an entire sphere but rather a sphere with one point removed. The removed point is $\mathbf{a}$. This issue can be rectified by assuming that the point at infinity is included into any plane (this makes sense since a plane is unbounded, thus it "extends to infinity" in a sense). An inversion is now though of as a bijective map from the set $\mathbb{R}^3 \cup \{\infty\}$ to itself (the point $\mathbf{a}$ is mapped to infinity, and the infinity is mapped to $\mathbf{a}$). With this understanding, we can state the following theorem.

t:invers **Theorem 3.3.** *An inversion maps spheres onto spheres.*

This gives us an idea of further examples of Möbius transformations. Consider an inversion that takes the sphere S to itself. The restriction of this inversion to S is a Möbius transformation! We need only find inversions that take S to itself. Suppose that $S_R(\mathbf{a})$ is a sphere such that the inversion in $S_R(\mathbf{a})$ takes S to itself. Suppose that the intersection $S \cap S_R(\mathbf{a})$ is a nonempty circle. Then every point of this circle is mapped to itself by the inversion. It can be shown any line interval connecting $\mathbf{a}$ with a point in $S \cap S_R(\mathbf{a})$ is tangent to S. (This and some other statements are left as exercises). In this case, we say that the spheres S and $S_R(\mathbf{a})$ are *orthogonal*. Thus, an inversion that takes S to itself is an inversion in a sphere orthogonal to S. The converse is also true: the inversion in any sphere orthogonal to S takes S to itself. Any circle in S is the intersection of S with some sphere orthogonal to S. However, the latter sphere can sometimes degenerate into a plane containing the origin. In this case, the circle of intersection is called a *great circle* in S. The inversion in a plane containing the origin can be defined as simply the reflection in this plane. Clearly, reflections in the planes containing the origin take S to itself, and their restrictions to S are Möbius transformations.

It can also be proved that an inversion preserves tangency: it maps pairs of tangent spheres to pairs of tangent spheres. This property in fact holds for a much larger class of maps: it can be shown that any differentiable map from an open subset of $\mathbb{R}^3$ to an open subset of $\mathbb{R}^3$ takes pairs of tangent surfaces to pairs of tangent surfaces. Again, in case of an inversion, this can be verified by a simple explicit computation.

We now choose $\mathbf{a}$ to be the north pole of the sphere S, i.e., we set $\mathbf{a} = (0, 0, 1)$. Let F be the inversion in the sphere $S_2(\mathbf{a})$. By definition, F maps the south pole $\mathbf{b} = (0, 0, -1)$ to itself. Let P be the plane tangent to S at the point $\mathbf{b}$. We claim that $F(P) = S$. First of all, $F(P)$ must be tangent to S, P and $S_2(\mathbf{b})$ at $\mathbf{b}$. Indeed, F takes the sphere $S_2(\mathbf{a})$ to itself, the sphere $S_2(\mathbf{a})$ is tangent to P and to S at $\mathbf{b}$, and F takes a pair of tangent spheres to a pair of tangent spheres. Next, the sphere $F(P)$ must pass through $\mathbf{a}$, since P contains the infinity, and the infinity is mapped to $\mathbf{a}$. There is only one sphere tangent to S at $\mathbf{b}$ that passes through $\mathbf{a}$. This sphere is S. We must conclude that $F(P) = S$. Since $F \circ F$ is the identity transformation, the map F can be thought either as a map from P to S or as a map from S to P. In particular, it establishes a one-to-one correspondence between P and S.

The restriction of F to S is called the *stereographic projection* of the sphere S into the plane P. It is characterized by the property that a point $\mathbf{x} \in S$, its image $F(\mathbf{x})$, and the north pole $\mathbf{a}$ always lie on the same line. This property is used as a standard definition of the stereographic projection. Note that the stereographic projection takes the north pole $\mathbf{a}$ to infinity. It follows from Theorem 3.3 that the stereographic projection maps circles in S to circles in P.

We will use the stereographic projection to express whatever happens in S in terms of plane geometry. In particular, let $f : S \to S$ be a Möbius transformation. Then we can define a transformation $\widehat{f}$ of P given by the formula $\widehat{f} = F \circ f \circ F$. Here the stereographic projection F takes P to S, the map f takes S to S, and then F takes S to P. Thus $\widehat{f}$ is a transformation of P. It is important to remember that, according to our convention, the plane P contains the infinity, and the infinity is mapped by $\widehat{f}$ to the point $F(f(\mathbf{a}))$. Since f takes circles to circles, and F takes circles to circles, the map $\widehat{f}$ also takes circles to circles. Thus we can restate our problem as follows: *describe all transformations* $\widehat{f} : P \to P$ *that take circles to circles.* This reformulation has an advantage: instead of dealing with the sphere S, which is an object of 3-dimensional geometry, we can deal only with the plane.

The next step is to identify the plane P with the plane $\mathbb{C}$ of complex numbers, or rather with the *extended plane* $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ of complex numbers since P contains the infinity by our convention. This will allow us to write a map $\widehat{f}$ by a formula! In fact, it can be shown that

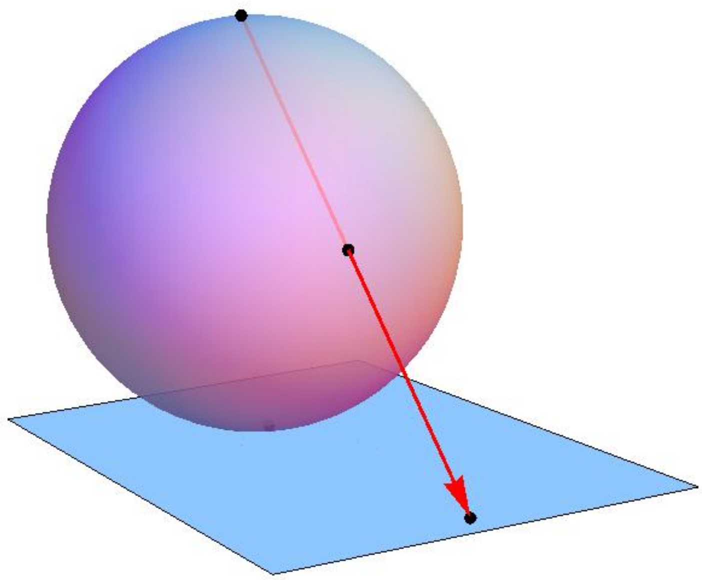


FIGURE 2. Stereographic projection.

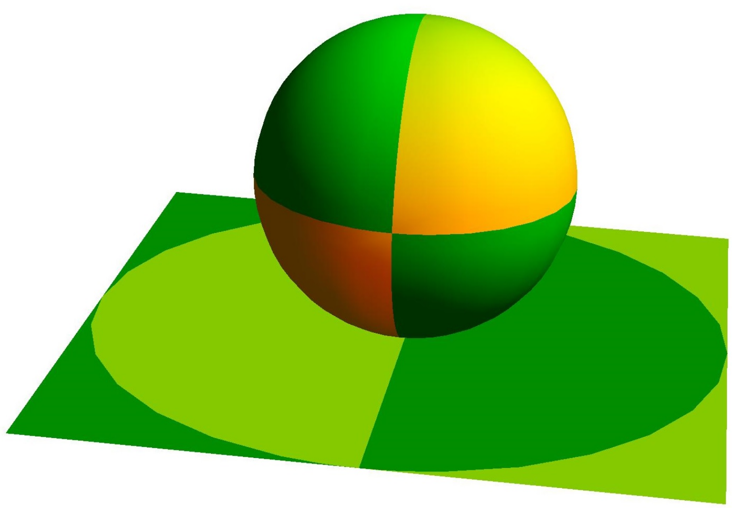


FIGURE 3. Stereographic projection of a sphere.

if f is a rotation of the sphere S, then $\widehat{f}$ is given by the formula

$$\widehat{f}(\zeta) = \frac{a\zeta + b}{c\zeta + d}$$

for some complex numbers a, b, c, $d \in \mathbb{C}$. Here ζ is a complex variable (which can take value ∞) representing a point in P. Maps of this type are called *fractional linear transformations* (provided that $c\zeta + d$ is not identically equal to zero, and $a\zeta + b$ is not a multiple of $c\zeta + d$). We see that rotations of S correspond to some fractional linear transformations of $\overline{\mathbb{C}}$.

For example, the rotation of S by angle θ around the z-axis is represented by the linear transformation $\zeta \mapsto (\cos\theta + i \sin\theta)\zeta$. This map can be though of as fractional linear, the denominator being equal to $0 \cdot \zeta + 1$. Another example is the reflection in the equator of S given by the equation $z = 0$. In terms of the coordinate ζ, the corresponding transformation is $\zeta \mapsto \overline{\zeta}^{-1}$. On the other hand, the reflection in the equator of S given by the equation $x = 0$, corresponds to the map $\zeta \mapsto \overline{\zeta}$. Neither the map $\zeta \mapsto \overline{\zeta}^{-1}$, nor the map $\zeta \mapsto \overline{\zeta}$, is a fractional linear transformation. The reason is that these maps *reverse orientation*, i.e., transform clockwise rotations to counterclockwise rotations, and the other way around, whereas fractional linear transformations always preserve orientation. Note however, that the composition of the map $\zeta \mapsto \overline{\zeta}^{-1}$ with the map $\zeta \mapsto \overline{\zeta}$ is the fractional linear transformation $\zeta \mapsto \zeta^{-1}$.

Any *linear transformation*, i.e., any transformation of the form $\zeta \mapsto a\zeta + b$, where $a \neq 0$, takes circles to circles. The map $\zeta \mapsto \zeta^{-1}$ also takes circles to circles as this map corresponds to the composition of two inversions. We can now deduce from these two observations that any fractional linear map

$$\widehat{f}(\zeta) = \frac{a\zeta + b}{c\zeta + d}$$

takes circles to circles. Indeed, the map $\widehat{f}$ can be written in the form

$$\widehat{f}(\zeta) = c^{-1}a - c^{-1}\frac{ad - bc}{c\zeta + d}.$$

This means that $\widehat{f}$ is the composition of three maps. The first map is linear, it takes ζ to $\zeta_1 = c\zeta + d$. The second map takes ζ_1 to $\zeta_2 = 1/\zeta_1$; it corresponds to the composition of two inversions. The third map is also linear, it takes ζ_2 to $\widehat{f}(\zeta) = c^{-1}(a - (ad - bc)\zeta_2)$. Thus we have proved the following theorem:

t:fraclin **Theorem 3.4.** *Any fractional linear transformation takes circles to circles.*

It is the time when we can guess the answer to the problem of Möbis. We state it as a theorem. Although we defined Möbius transformations as self-maps of the sphere, we can also talk about Möbius transformations of $\overline{\mathbb{C}}$: these are bijective maps that take circles to circles.

t:moeb

Theorem 3.5 (Möbius). *Any Möbius transformation of the extended complex plane $\overline{\mathbb{C}}$ is of one of the following two forms:*

$$\widehat{f}(\zeta) = \frac{a\zeta + b}{c\zeta + d}, \quad \widehat{f}(z) = \frac{a\overline{\zeta} + b}{c\overline{\zeta} + d},$$

i.e., it is either fractional linear, or conjugate to a fractional linear transformation.

Proof. Suppose that $\widehat{f}$ is a Möbius transformation of $\overline{\mathbb{C}}$. Then the map

$$\widehat{g}(\zeta) = \frac{1}{f(\zeta) - \widehat{f}(\infty)}$$

is also a Möbius transformation, and we have $\widehat{g}(\infty) = \infty$. Thus it suffices to prove the following: if $\widehat{g}$ is a Möbius transformation of $\overline{\mathbb{C}}$ that fixes the point ∞, then $\widehat{g}$ is either a linear transformation or a conjugate linear transformation.

Consider the map

$$\widehat{h}(\zeta) = \frac{\widehat{g}(\zeta) - \widehat{g}(0)}{\widehat{g}(1) - \widehat{g}(0)}.$$

This map takes 0 to 0 and 1 to 1, and this map is also a Möbius transformation. Thus it suffices to prove the following: if $\widehat{h}$ is a Möbius transformation of $\overline{\mathbb{C}}$ that fixes the points ∞, 0 and 1, then we have $\widehat{h}(\zeta) = \zeta$ or $\widehat{h}(\zeta) = \overline{\zeta}$.

Note that the *real line* $\mathrm{Im}(\zeta) = 0$ is the only circle passing through the three points 0, 1 and ∞. It follows that $\widehat{h}$ takes this line to itself. We now discuss several properties of the map $\widehat{h}$ that will eventually lead to the conclusion that $\widehat{h}$ is either the identify transformation, or the complex conjugation.

First note that the map $\widehat{h}$ takes lines to lines. Indeed, lines are precisely circles that pass through ∞. Note also that parallel lines (i.e., lines that intersect at infinity only) are mapped to parallel lines. We can now draw some consequences. Consider a line $0a$ passing through 0 and different from the real line. We use two points (in our case 0 and a) on a line to denote this line. Thus, "the line $0a$" means the line passing through 0 and a. Then the real line can be denoted 01. Consider a line ab parallel to the line 01. We may assume that the point b on this line is chosen so that the line $1b$ is parallel to the line $0a$. Draw the line parallel to $a1$ through the point b. The intersection point of this line with the line 01 is the point 2. Thus, the point 2 can be constructed using only the line 01, and the possibility of drawing lines and parallel lines. Since $\widehat{h}$ maps lines to lines and parallel lines to parallel lines, it follows that $\widehat{h}(2) = 2$.

Performing the same construction repeatedly, it is easy to show that in fact every integer is fixed under the map $\widehat{h}$. We now claim that the point $\frac{1}{2}$ is also fixed. Indeed, using the constructions performed above, let c denote the intersection point of the two diagonals of the parallelogram $0ab1$. Through the point c, draw the line parallel to $1a$. This line intersects the line 01 at point $\frac{1}{2}$. Thus the point $\frac{1}{2}$ can also be constructed only by means of drawing parallel lines and taking their intersection points, hence this point is also fixed under the map $\widehat{h}$. This observation can be generalized as follows: if x and y are two fixed points of $\widehat{h}$ on the line 01, then the point $\frac{x+y}{2}$ is also fixed under $\widehat{h}$. It follows that all binary rational points, i.e., all rational numbers, whose denominators are powers of 2, are fixed.

If the map $\widehat{h}$ were known to be continuous, then we would immediately conclude that $\widehat{h}$ fixes every point of the line 01. However, we cannot use the continuity, thus we need a different argument. Note that, for a circle C, the *interior* of C is characterized by the following property: if z is inside C, then any line passing through z intersects C in two points. Since this description involves only lines, circles, and intersection points, we can deduce that $\widehat{h}$ takes the interior of C to the interior of $\widehat{h}(C)$. In particular, it follows that $\widehat{h}$ preserves the order of points in the line 01. It now follows that all points of the line 01 are fixed.

Take any point ζ not in the real line. We claim that $\widehat{h}(\zeta)$ can be either ζ or $\overline{\zeta}$. □

Problems.

Problem 3.6. A surface in $\mathbb{R}^3$ is given by the equation

$$x^2 + y^2 + 2z^2 + 2x - y + 3z = 10.$$

Is this surface a sphere? Rigorously justify your answer.

Problem 3.7. A sphere S' is orthogonal to the unit sphere S. Express this relation as an equation involving the homogeneous coordinates of S'.

Problem 3.8. Express the inversion in a sphere S' in terms of homogeneous coordinates of S'.

Problem 3.9. Prove that the inversion in a sphere orthogonal to S takes S to itself.

Problem 3.10. Prove that an inversion maps pairs of tangent spheres to pairs of tangent spheres. Prove also that an inversion maps pairs of tangent circles to pairs of tangent circles.

Problem 3.11. Suppose that f is a rotation of the sphere S. Prove that $\widehat{f}$ is given by the formula

$$\widehat{f}(z) = \frac{az+b}{cz+d}.$$

4. Fractional linear transformations

In this section, we discuss fractional linear transformations in more detail. The coefficients of a fractional linear transformation

$$\varphi_A(z) = \frac{az+b}{cz+d}$$

can be conveniently organized into a 2 by 2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Thus, every matrix A with complex entries, whose determinant $ad-bc$ is nonzero, gives rise to a unique fractional linear transformation φ_A. If the determinant of A is zero, then the numerator and the denominator of φ_A are proportional, hence φ_A is either a constant (perhaps infinity) or the expression of the form $\frac{0}{0}$. In all these cases, φ_A is not a transformation. The transformation φ_A corresponds to infinitely many matrices but all these matrices differ by a nonzero scalar factor only.

The product of matrices corresponds to the composition of fractional linear transformations. In other words, we have $\varphi_{AB} = \varphi_A \circ \varphi_B$ for any two matrices A, B with nonzero determinants. This can be verified by a straightforward computation. There is a more conceptual explanation, involving homogeneous coordinates on $\overline{\mathbb{C}}$ and the interpretation of $GL(2,\mathbb{C})$ as the group of all linear transformations of $\mathbb{C}^2$.

p:3pts **Proposition 4.1.** *Let z_0, z_1 and z_∞ be three different points of $\overline{\mathbb{C}}$. Then there is a fractional linear transformation that takes the points z_0, z_1 and z_∞ to 0, 1 and ∞.*

Proof. The map $f(z) = \frac{1}{z-z_\infty}$ takes z_∞ to ∞. The map $g(z) = z - f(z_0)$ takes $f(z_0)$ to 0 and fixes ∞, hence $g \circ f$ takes z_0 and z_∞ to 0 and ∞, respectively. Note that $g(f(z_1))$ is different from 0 and 1. The map $h(z) = z/g(f(z_1))$ fixes 0 and ∞ and takes $g(f(z_1))$ to 1. Hence the map $h \circ g \circ f$ is as desired. □

Fixed points of the fractional linear transformation

$$\varphi_A(z) = \frac{az+b}{cz+d}$$

are given by the following quadratic equation: $cz^2 + (d-a)z - b = 0$. The discriminant of this quadratic equation is equal to

$$(d-a)^2 + 4bc = (a+d)^2 - 4(ad-bc) = \mathrm{tr}(A)^2 - 4\det(A).$$

If the discriminant is zero, then there is only one fixed point (possibly at infinity: this case corresponds to $c = 0$). A fractional linear transformation with only one fixed point is said to be *parabolic*.

Suppose now that a given fractional linear transformation φ_A has two different fixed points z_0 and z_∞. By Proposition 4.1, there exists a fractional linear transformation f that maps these points to 0 and ∞, respectively. Then the map $\varphi_B = f \circ \varphi_A \circ f^{-1}$ fixes 0 and ∞, hence it has the form $z \mapsto \lambda z$. The number λ does not depend on the choice of f, and it changes to λ^{-1} if the points z_0 and z_∞ are interchanged. In fact, λ is equal to the ratio of the eigenvalues of A (this becomes clear from the observation that B has the same eigenvalues as A, up to a common factor; on the other hand, B is diagonal). If $|\lambda| = 1$, then the transformation φ_A is said to be *elliptic*. If $|\lambda| \neq 1$ and λ is not real, then the transformation φ_A is said to be *loxodromic*. Finally, if λ is real (and different from ± 1), then the transformation φ_A is said to be *hyperbolic*.

Let us first consider a loxodromic transformation f. If we map its fixed points to 0 and ∞, then f is given by the formula $f(z) = \lambda z$, where λ is a complex number, whose modulus is different from 1. We may assume without loss of generality that $|\lambda| < 1$ (otherwise, it suffices to interchange the fixed points of f by the fractional linear map $z \mapsto \frac{1}{z}$). The point 0 is *attracting* for f in the sense that, for every point $z \neq \infty$, we have $f^{\circ n}(z) \to 0$ as $n \to \infty$. The set of points $f^{\circ n}(z)$ (where n runs through all integers, positive and negative) is called the *orbit* of z under the map f.[1] The point ∞ is *repelling* for f in the sense that all points sufficiently close to infinity map farther away from infinity.

It follows that any loxodromic transformation f has one attracting and one repelling fixed points. The orbit $f^{\circ n}(z)$ of any point z different from the repelling fixed point converges to the attracting fixed point of f as $n \to \infty$. On the other hand, if z is different from the attracting fixed point, then this orbit converges as $n \to -\infty$ to the repelling fixed point.

Let us again suppose that the fixed points of f are 0 and ∞, i.e., that f is the multiplication by some complex number λ. We now consider

[1] For any map g defined on a neighborhood U of a point z_0, the point z_0 is said to be an *attracting fixed point* of g if there exists a neighborhood $V \subset U$ of z_0 such that $g^{\circ n}(z)$ is defined for all $z \in V$ and positive integers n, and $g^{\circ n}(z) \to z_0$. In other words, the forward orbits of all nearby points converge to z_0.

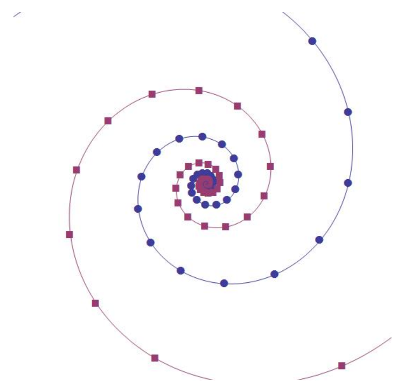


FIGURE 4. Two orbits of the loxodromic map $z \mapsto (1+\frac{i}{2})z$ and the spirals containing them.

not only the loxodromic case but also the elliptic and the hyperbolic cases. Thus λ can have modulus 1 and can be real. Note that f maps any circle centered at the origin to another (or the same) circle centered at the origin. Thus we have an *invariant family of circles*.[2] In fact, there is another invariant family of circles consisting of lines passing through the origin (recall that, according to our convention, a line is also a circle). We see that every loxodromic, hyperbolic or elliptic fractional linear transformation has two invariant families of circles. Although, in our argument, we have used that f fixes the points 0 and ∞, the statement remains true for a completely arbitrary hyperbolic or elliptic transformation f, since a fractional linear change of coordinates maps circles to circles. These two invariant families of circles form the so called *Steiner net*, cf. Problem 4.11. Elliptic and hyperbolic transformations can be characterized by the property that they fix every circle of one of the two families.

It remains to consider the parabolic case. A parabolic fractional linear transformation f has only one fixed point. If we assume that this fixed point is ∞, then f has the form $f(z) = z + a$ for some complex number $a \neq 0$. A further change of variables reduces the map f to the form $f(z) = z+1$. This map has one invariant family of circles, namely, the lines $\mathrm{Re}(z) = const$. All these circles are tangent at infinity (as can be seen by means of the stereographic projection to the sphere). We conclude that any parabolic fractional linear transformation as an invariant family of circles, all of which are tangent to a given direction at the fixed point.

[2]A family of curves is said to be *invariant* under a map f if f takes any curve of this family to a curve of the same family.

Suppose that we are given a matrix A, and want to determine the type of the corresponding fractional linear transformation φ_A, i.e. whether φ_A is loxodromic, hyperbolic, elliptic or parabolic. To this end, we can find the eigenvalues λ_1, λ_2 of the matrix A, and consider the ratio $\lambda = \frac{\lambda_1}{\lambda_2}$ of these eigenvalues. If the eigenvalues λ_1, λ_2 are interchanged, then λ is replaced with λ^{-1}. The type of the transformation f can be determined just by looking at λ and the Jordan type of the matrix A. If $|\lambda| \neq 1$ and $\mathrm{Im}\lambda \neq 0$, then f is loxodromic. If $|\lambda| \neq 1$ and $\mathrm{Im}\lambda = 0$, then f is hyperbolic. If $|\lambda| = 1$ and A is diagonalizable, then f is elliptic. Finally, if $|\lambda| = 1$ and A is a Jordan block, then f is parabolic.

It is not even necessary to compute the eigenvalues of A in order to determine the type of the transformation φ_A. Consider the number

$$\lambda + \frac{1}{\lambda} = \frac{\lambda_1^2 + \lambda_2^2}{\lambda_1 \lambda_2}.$$

This number is a symmetric function of the eigenvalues of A. Therefore, it can be expressed as a function of the entries of A. Suppose that

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Then λ_1 and λ_2 are roots of the equation

$$t^2 - \mathrm{tr}(A)t + \det(A) = 0.$$

The symmetric polynomial $\lambda_1\lambda_2$ is just equal to $\det(A)$. The symmetric polynomial $\lambda_1^2 + \lambda_2^2$ is equal to

$$(\lambda_1 + \lambda_2)^2 - 2\lambda_1\lambda_2 = \mathrm{tr}(A)^2 - 2\det(A).$$

It follows that

$$\lambda + \frac{1}{\lambda} = \frac{\mathrm{tr}(A)^2 - 2\det(A)}{\det(A)} = \frac{\mathrm{tr}(A)^2}{\det(A)} - 2.$$

We can now restate our classification of fractional linear transformations in terms of the numbers $\mathrm{tr}(A)$ and $\det(A)$. Usually, people assume that $\det(A) = 1$, and this can be always arrange by multiplying the numerator and the denominator of the map φ_A by the same suitably chosen factor. We will not make this assumption. Loxodromic maps correspond to matrices A such that $\frac{\mathrm{tr}(A)^2}{\det(A)}$ is not real. Hyperbolic maps correspond to matrices A such that $\frac{\mathrm{tr}(A)^2}{\det(A)}$ is real but not in the interval $[0, 4]$. Elliptic and parabolic maps correspond to matrices A such that $\frac{\mathrm{tr}(A)^2}{\det(A)}$ is in the interval $[0, 4]$. Parabolic maps can be characterized by the property that $\frac{\mathrm{tr}(A)^2}{\det(A)}$ is equal to 4 but A is not the identity matrix.

Problems.

Problem 4.2. Prove that $\varphi_{AB} = \varphi_A \circ \varphi_B$ for any two matrices A, $B \in \mathrm{GL}(2, \mathbb{C})$.

Solution. We will use the so called *homogeneous coordinates* on $\overline{\mathbb{C}}$. A point z of $\overline{\mathbb{C}}$ has homogeneous coordinates $[z_1 : z_2]$ (where the complex numbers z_1 and z_2 do not vanish simultaneously) if $z = \frac{z_1}{z_2}$. In particular, if $z = \infty$, then $z_2 = 0$. Homogeneous coordinates of a point are defined only up to a common nonzero factor. Observe that, if we apply a linear transformation with a matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

to the pair of homogeneous coordinates $[z_1 : z_2]$, i.e. consider the point w, whose homogeneous coordinates $[w_1 : w_2]$ are

$$w_1 = az_1 + bz_2,$$
$$w_2 = cz_1 + dz_2,$$

then we have $w = \varphi_A(z)$. Now set $z = \varphi_B(t)$. Then $w = \varphi_A \circ \varphi_B(t)$. On the other hand, the homogeneous coordinates of w are obtained from the homogeneous coordinates of t by the composition of two linear maps, one with matrix A, and the other with matrix B. This composition is the linear map with matrix AB. We obtain that $\varphi_A \circ \varphi_B = \varphi_{AB}$, as desired.

Problem 4.3. Consider the fractional linear transformation

$$f(z) = \frac{i - z}{i + z}.$$

Find the images under f of the following sets: $\mathrm{Re}(z) \geqslant 0$, $\mathrm{Im}(z) \geqslant 0$, $|z| < 1$.

Problem 4.4. Prove that fractional linear transformations preserve cross-ratios.

Problem 4.5. Find all fractional linear transformations with fixed points 1 and -1.

Problem 4.6. * Find all pairs of commuting fractional linear transformations.

Problem 4.7. * Find the general form of a fractional linear transformation corresponding to a rotation of a sphere under the stereographic projection onto $\overline{\mathbb{C}}$.

Problem 4.8. Determine the type (elliptic, parabolic, loxodromic, hyperbolic) of the following fractional linear transformations:

$$\frac{z}{2z-1}, \quad \frac{2z}{3z-1}, \quad \frac{3z-4}{z-1}, \quad \frac{z}{3-z}, \quad \frac{iz}{2-z}.$$

Problem 4.9. Suppose that a fractional linear transformation f satisfies the identity $f(f(z)) = z$. Prove that f is elliptic.

pb:Apol *Problem* 4.10. **The Apollonius circle theorem.** Let two different points $a, b \in \mathbb{C}$ and a positive real number $r > 0$ be given. Prove that the locus of points $z \in \mathbb{C}$ such that

$$\frac{|z-a|}{|z-b|} = r$$

is a circle.

pic:steiner

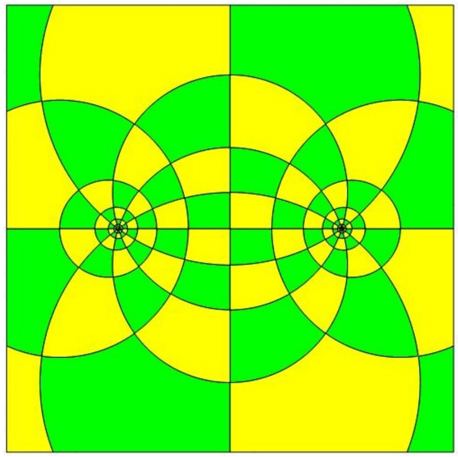


FIGURE 5. The Steiner net

pic:loxo1

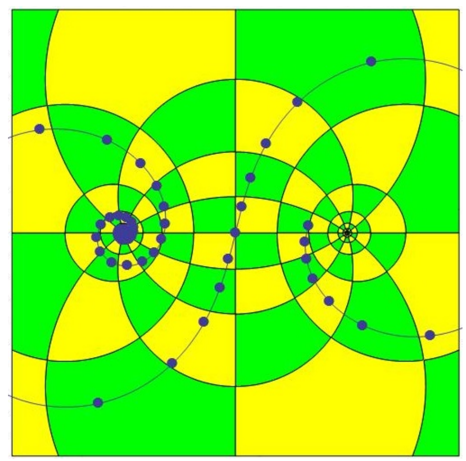


FIGURE 6. An orbit of a loxodromic map with attracting fixed point -1 and repelling fixed point 1.

pb:Steiner

Problem 4.11. **The Steiner net.** Find two families of circles that are invariant under any fractional linear transformation with fixed points ± 1 (in other words, every fractional linear transformation f such that $f(\pm 1) = \pm 1$ takes every circle of one of the two families to a circle of the same family).

5. Power series

In this section, we discuss some particularly important functions of a complex variable and their series expansions. A *polynomial function* (or simply a polynomial) of degree d is a function of the form

$$f(z) = a_0 + a_1 z + \cdots + a_d z^d,$$

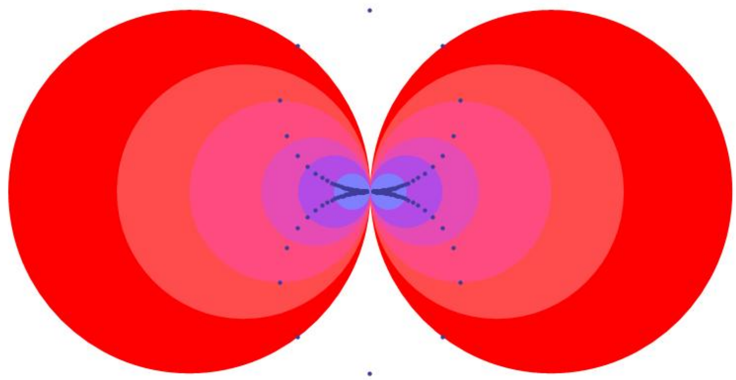


FIGURE 7. A parabolic fractional linear transformation: Steiner circles and orbits.

where the *coefficients* $a_0, \ldots, a_d$ are complex numbers, and $a_d \neq 0$. Polynomials of degree 2 are also called *quadratic* polynomials, and polynomials of degree 3 are also called *cubic* polynomials. Recall the following theorem (due to Bezout): if polynomial f of degree d and a complex number α are such that $f(\alpha) = 0$ (such α is called a *root*, or a *zero*, of f), then $f(z) = (z - \alpha)g(z)$ for some polynomial g of degree $d - 1$. The proof of this theorem is based on the division of $f(z)$ by $z - \alpha$ with a remainder. It accounts for showing that the remainder is zero. It follows from the theorem of Bezout that a polynomial of degree d can have at most d roots. Another important corollary is the following: if $\alpha_1, \ldots, \alpha_d$ are different zeros of a degree d polynomial f, then f has the form

$$f(z) = const \cdot (z - \alpha_1) \cdots (z - \alpha_d).$$

The set of all polynomial functions of z with complex coefficients is usually denoted by $\mathbb{C}[z]$. The degree of a polynomial f is written as $\deg(f)$.

A *rational function* is a function of the form

$$f(z) = \frac{P(z)}{Q(z)},$$

where P and Q are two polynomials that can be arbitrary except for only one forbidden situation, namely, that both polynomials identically equal to zero. A rational function can be thought of as a function from the *Riemann sphere* $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ to itself. We say that *the degree of a rational function* f is at most d if the degrees of P and Q are at most d. Note one subtlety here: even if $\max(\deg(P), \deg(Q)) = d$, the degree of the rational function $f = P/Q$ may be less than d. This happens if P

and Q have a nonconstant common polynomial factor. Fractional linear transformations, which we discussed earlier, are examples of rational functions.

We now state some important analytic property of polynomials and rational functions.

Definition 5.1 (Complex differentiability). Consider an open subset $U \subset \mathbb{C}$, a point $a \in U$, and a function $f : U \to \mathbb{C}$ defined on U. We say that f *has complex derivative* $f'(a)$ at the point a if the limit

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

exists and is equal to $f'(a)$.

For example, a constant function on an open set $U \subset \mathbb{C}$ has complex derivative at every point of U, and this derivative is equal to 0. If a function f has a complex derivative at a point a, then it is also continuous at a. This can be proved in the same way as the corresponding statement from real 1-dimensional analysis. The derivative $f'(a)$ can also be written as $\frac{d}{dz} f(z)|_{z=a}$

p:arith-diff **Proposition 5.2.** *Suppose that functions f and g are defined in some neighborhood of a point $a \in \mathbb{C}$. If f and g have complex derivatives at a, then so do the functions $f + g$ and fg. Moreover, we have*

$$(f + g)'(a) = f'(a) + g'(a), \quad (fg)'(a) = f'(a)g(a) + f(a)g'(a).$$

Suppose additionally that $g(a) \neq 0$. Then the function f/g is also differentiable at the point a, and we have

$$\frac{d}{dz}\frac{f(z)}{g(z)}|_{z=a} = \frac{f'(a)g(a) - f(a)g'(a)}{g(a)^2}.$$

This is also proved in the same way as in real 1-dimensional analysis.

Definition 5.3 (Holomorphic function). A function $f : U \to \mathbb{C}$ defined on some open subset $U \subset \mathbb{C}$ is said to be *holomorphic* if it has complex derivatives at all points of U.

Corollary 5.4. *Every polynomial is holomorphic on $\mathbb{C}$.*

Proof. Indeed, every polynomial can be obtained from constant functions and the function z by the operations of multiplication and addition. Clearly, constant functions and the function z are holomorphic (the derivative of the latter is equal to 1). It follows from Proposition 5.2 that every polynomial is differentiable. □

We can write an explicit formula for the derivative of a complex polynomial

$$f(z) = a_0 + a_1 z + \cdots + a_d z^d$$

similar to the well-known formula for the derivative of a real polynomials:

$$f'(z) = a_1 + 2a_2 z + \cdots + d a_d z^{d-1} = \sum_{k=1}^{d} k a_k z^{k-1}.$$

The property of being a holomorphic function is much stronger than it might seem. To illustrate this, we state a theorem, whose prove will be postponed until we develop a necessary technique:

Theorem 5.5. *Let $f : \mathbb{C} \to \mathbb{C}$ be a holomorphic function with the property that $|f(z)| < M(|z|^d + 1)$ for some positive integer d and some positive real number M. Then f is a polynomial of degree at most d.*

Note that the condition $|f(z)| < (M|z|^d + 1)$ just says that f does not grow too fast at infinity. Complex derivatives satisfy the *chain rule*, which is also very similar to that in real analysis:

Proposition 5.6 (Chain rule)**.** *Suppose that a function g has complex derivative at a point a, and a function f has complex derivative at the point $g(a)$. Then the composition $f \circ g$ has complex derivative at the point a, and we have*

$$(f \circ g)'(a) = f'(g(a))g'(a).$$

The Riemann sphere $\overline{\mathbb{C}}$ can be equipped with the following metric called the *spherical metric*. As we have seen, the stereographic projection establishes a one-to-one correspondence between the points of the unit sphere S in the Euclidean 3-space and the points of $\overline{\mathbb{C}}$. The distance between two points of S is by definition equal to the length of the shortest arc of a great circle connecting these two points. The distance between two points of $\overline{\mathbb{C}}$ is by definition equal to the distance between the corresponding points of S. Since $\overline{\mathbb{C}}$ can be regarded as a metric space, we can talk about neighborhoods of points in $\overline{\mathbb{C}}$, in particular, we can talk about neighborhoods of infinity.

We now define what the existence of a complex derivative means for a function defined on a subset of $\overline{\mathbb{C}}$ and taking values in $\overline{\mathbb{C}}$ in the cases, where either the point, at which we differentiate, or the value of the function at this point, is equal to infinity. The trick is to use the fractional linear transformation $z \mapsto 1/z$ that maps a neighborhood of ∞ to a neighborhood of 0.

Definition 5.7 (Complex derivative at infinity). Suppose that $a \in \overline{\mathbb{C}}$, an open set $U \subset \overline{\mathbb{C}}$ contains a, and $f : U \to \overline{\mathbb{C}}$ is a function. If $f(a) = \infty$, then we set $f_1(z) = 1/f(z)$, otherwise we set $f_1(z) = f(z) - f(a)$. If $a = \infty$, then we set $f_2(z) = f_1(1/z)$, otherwise we set $f_2(z) = f_1(z-a)$. The function f_2 is defined on a neighborhood of 0, and we have $f_2(0) = 0$. We say that f has complex derivative at the point a if f_2 has complex derivative at 0.

Suppose that $U \subset \overline{\mathbb{C}}$ is an open subset, and $f : U \to \overline{\mathbb{C}}$ is a function. If f has complex derivative at every point of U, then we say that f is a *holomorphic map of U to $\overline{\mathbb{C}}$* or that f is a *meromorphic function* on U. According to the classical terminology, *holomorphic functions* on U are only those that take finite values at all points of U. Thus it is important to distinguish between holomorphic maps to $\overline{\mathbb{C}}$ and holomorphic functions.

The following theorem will be proved later.

Theorem 5.8. *Suppose that $f : \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ is a holomorphic map. Then f is a rational function.*

This purely analytic characterization of rational functions plays a very important role in holomorphic dynamics.

To discuss other elementary functions, we first need to discuss convergence of series consisting of complex numbers. We say that a series

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + \cdots, \quad a_n \in \mathbb{C}$$

converges to a complex number a if the sequence of partial sums

$$S_N = \sum_{n=1}^{N} a_n$$

converges to a. The number a is called the *sum* of the series $a_1 + a_2 + \ldots$. We say that this series is *absolutely convergent* if the series

$$\sum_{n=1}^{\infty} |a_n|$$

converges.

Proposition 5.9. *An absolutely convergent series converges.*

Proof. First note that the sequence S_N of partial sums is a *Cauchy sequence*: for every $\varepsilon > 0$ there exists N_0 such that for all N, $N' >$

N_0, we have $|S_N - S_{N'}| < \varepsilon$. Indeed, this follows from the absolute convergence and the triangle inequality: if, say, $N > N'$, then

$$|S_N - S_{N'}| = |a_{N'+1} + \cdots + a_N| \leqslant |a_{N'+1}| + \cdots + |a_N|.$$

The fact that any Cauchy sequence converges is called the *completeness* of $\mathbb{C}$ (in general, a metric space is said to be *complete* if any Cauchy sequence in this space converges). The completeness of $\mathbb{C}$ can be seen as follows. If S_N is a Cauchy sequence, then the sequences $\mathrm{Re}(S_N)$ and $\mathrm{Im}(S_N)$ are both Cauchy. Therefore, these two sequences converge. It now follows that S_N converges. □

p:ser-comm **Proposition 5.10.** *Suppose that a series*

$$\sum_{n=1}^{\infty} a_n$$

is absolutely convergent, and let a be its sum. Consider any sequence of finite subsets $I_k \subset \mathbb{N}$ such that $I_k \subset I_{k+1}$ and $\bigcup_k I_k = \mathbb{N}$ (here $\mathbb{N}$ is the set of natural numbers 1, 2, ...). Set $s_k = \sum_{\ell \in I_k} a_\ell$. Then the sequence s_k converges to a.

Proof. Note that, for every N, there is k such that I_k contains the set $\{1, \ldots, N\}$. Therefore, we have

$$|a - s_k| \leqslant |a_{N+1}| + |a_{N+2}| + \ldots.$$

The right-hand side of this inequality tends to 0 as $k \to \infty$. □

We now mention some important consequences of Proposition 5.10.

c:ser-comm1 **Corollary 5.11.** *If a series A is absolutely convergent, and we form a different series B, whose elements are the same as elements of A but taken in a different order, then B is also absolutely convergent, and the sum of B is equal to the sum of A.*

c:ser-comm2 **Corollary 5.12.** *Consider an absolutely convergent series $\sum_{n=1}^{\infty} a_n$, whose sum is equal to a. Suppose that I_k, $k \in \mathbb{N}$, are disjoint sets (finite or infinite) such that $\mathbb{N} = \bigcup I_k$. Set $A_k = \sum_{\ell \in I_k} a_\ell$ (this series is absolutely convergent). Then the series $\sum_{k=1}^{\infty} A_k$ converges to a.*

Consider the following series

$$\sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2!} + \cdots.$$

This series converges absolutely for every $z \in \mathbb{C}$ since the corresponding series of the absolute values is the series for $e^{|z|}$ well known from real calculus.

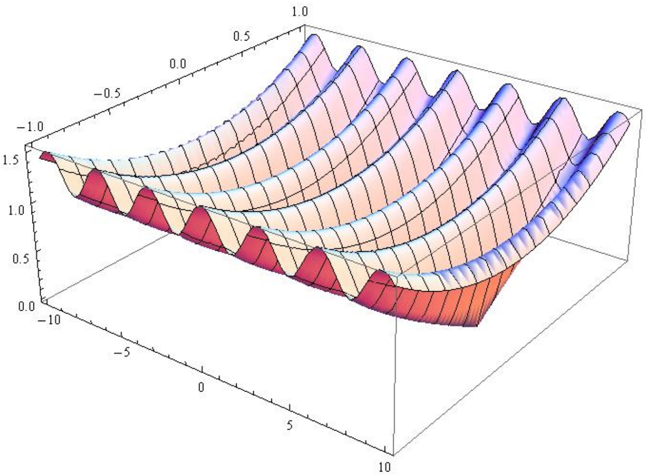


FIGURE 8. The graph of the function $z \mapsto |\cos(z)|$.

p:exp-hom **Proposition 5.13.** *The function e^z satisfies the following functional equation:*

$$e^{z+w} = e^z e^w.$$

Proof. Using Corollaries 5.11 and 5.12, we can rewrite the series for $e^{z_1} e^{z_2}$ as follows:

$$e^z e^w = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{z^n w^m}{n! m!} = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{n=0}^{k} \binom{k}{n} z^n w^{k-n} = \sum_{k=0}^{\infty} \frac{(z+w)^k}{k!},$$

and the latter series converges to e^{z+w}. □

Suppose that ϕ is real, and consider $e^{i\phi}$:

$$e^{i\phi} = \sum_{n=0}^{\infty} (-1)^n \frac{\phi^{2n}}{(2n)!} + i \sum_{n=0}^{\infty} (-1)^n \frac{\phi^{2n+1}}{(2n+1)!}.$$

We see that the power series for $\mathrm{Re}(e^{i\phi})$ coincides with the well-known power series for $\cos\phi$. The power series for $\mathrm{Im}(e^{i\phi})$ coincides with the well-known power series for $\sin\phi$. Therefore, we obtain that

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

In particular, setting $\phi = \pi$, we obtain the famous formula of Euler relating three important constants: e, π and i:

$$e^{i\pi} = -1.$$

For any complex number $z = x + iy$, we obtain using Proposition 5.13 that

$$e^z = e^x(\cos y + i \sin y).$$

On the other hand, we can define $\cos z$ and $\sin z$ for all complex values of z by the same power series expansions that we know from real calculus. Then we have

$$e^{iz} = \cos z + i \sin z, \quad e^{-iz} = \cos z - i \sin z.$$

This system can be solved for $\cos z$ and $\sin z$, which gives the following expressions:

$$\cos z = \frac{e^{iz} + e^{-iz}}{2}, \quad \sin z = \frac{e^{iz} - e^{-iz}}{2i}.$$

The first application of the complex exponential is to solving algebraic equations of the form $z^n = a$, where z is a complex unknown, and a is a given complex number. Represent the number a in the form $re^{i\phi}$ Here the number r is equal to $|a|$, and ϕ is the angle that the vector a makes with the positive direction of the real axis. In other words, (r, ϕ) are *polar coordinates* of the point a. The angle ϕ is also called the *argument* of a. Then all solutions z of the equation $z^n = a$ are given by the formula $z = \sqrt[n]{r} e^{i\frac{\phi+k}{n}}$, where k runs through all integers from 0 to $n-1$. Thus there are exactly n solutions, and they are located at vertices of a regular n-gon inscribed into the circle of radius $\sqrt[n]{r}$ centered at 0.

We now discuss radii of convergence for power series.

p:rad-conv **Proposition 5.14.** *Suppose that a power series*

$$\sum_{n=0}^{\infty} a_n z^n$$

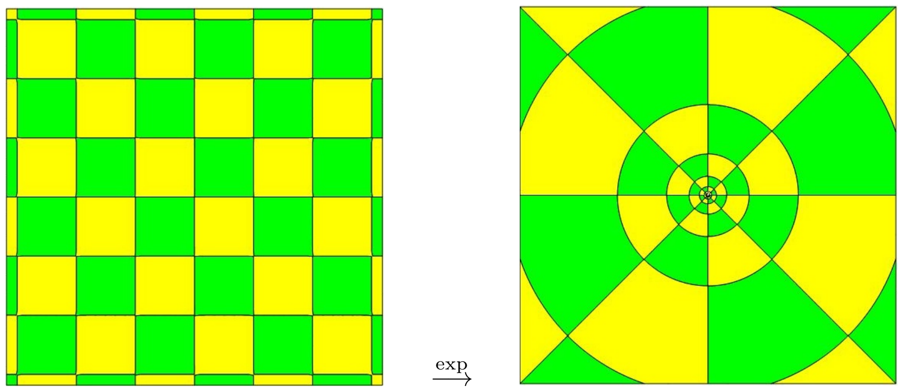
converges for $z = u$. *Then this series is absolutely convergent for all* z *such that* $|z| < |u|$.

Proof. Take any z such that $|z| < |u|$. Since the series $\sum_{n=0}^{\infty} a_n u^n$ converges, we have $|a_n||u|^n < C$ for some positive real number C independent of n. We conclude that $|a_n| < C|u|^{-n}$, hence

$$|a_n z^n| < Cq^n, \quad q = \frac{|z|}{|u|} < 1.$$

Since the series $\sum Cq^n$ converges, the series $\sum a_n z^n$ is absolutely convergent. □

Definition 5.15 (Radius of convergence). It follows from Proposition 5.14 that, for any given power series $A(z) = \sum a_n z^n$ (when we speak of power series, we always assume that the summation index n runs through all nonnegative integers from 0 to infinity), there exists a positive real number R_0 with the following properties:



- if $|z| < R_0$, then the power series $A(z)$ converges;
- if $|z| > R_0$, then the power series $A(z)$ diverges.

The number R_0 with these properties is called the *radius of convergence* of the power series $A(z)$.

Recall that, if α_n is a sequence of real numbers, then the *upper limit* of α_n is defined as

$$\limsup_{n\to\infty} \alpha_n = \lim_{N\to\infty} \sup\{\alpha_N, \alpha_{N+1}, \dots\}.$$

The upper limit always exists, since, in the right-hand side, the limit is taken of an (non-strictly) decreasing sequence. The upper limit can be either a real number or $+\infty$.

Theorem 5.16. *The radius of convergence of a power series* $\sum a_n z^n$ *is equal to*

$$R = \frac{1}{\limsup_{n\to\infty} |a_n|^{1/n}}.$$

Problems.

Problem 5.17. Find the images of the upper half-plane $\mathrm{Im}(z) > 0$ and of circles centered at the origin under the map

$$f(z) = z + \frac{1}{z}.$$

exp-img

Problem 5.18. Find the images of horizontal and vertical lines under the map $z \mapsto e^z$.

Problem 5.19. Find the images of horizontal and vertical lines under the map $z \mapsto \cos z$.

Problem 5.20. Draw the image of the circle $|z - 1| = 1$ under the map $z \mapsto z^2$ (this curve is called a *cardioid*).

Problem 5.21. Find $\cos \frac{\pi}{16}$.

Answer:

$$\cos\frac{\pi}{16} = \frac{1}{2}\sqrt{2+\sqrt{2+\sqrt{2}}}$$

Problem 5.22. Find a rational function R such that $\tan(n\alpha) = R(\tan(\alpha))$ for every $\alpha \in \mathbb{R}$.

Answer:

$$R(t) = \frac{\sum_{k=0}^{[(n-1)/2]}(-1)^k\binom{n}{2k+1}t^{2k+1}}{\sum_{k=0}^{[n/2]}(-1)^k\binom{n}{2k}t^{2k}}$$

Problem 5.23. Find the vertices of the regular pentagon inscribed into the unit circle $\{|z| = 1\}$ such that -1 is one of the vertices. Find the lengths of all diagonals of this pentagon.

Problem 5.24. Does the series

$$\sum_{n=1}^{\infty}\frac{n\sin(ni)}{3^n}$$

converge?

Problem 5.25. Find all complex numbers z such that $e^z = i$.

Problem 5.26. Find all complex numbers z such that $\sin(z) = 3$.

Problem 5.27. For every linear polynomial $u(x, y) = ax+by+c$, find a holomorphic function $f(z)$ such that the real part of $f(x + iy)$ coincides with u.

Problem 5.28. Prove that

$$\sum_{n=1}^{\infty}\frac{u^{2n}}{(n!)^2} = \frac{1}{2\pi}\int_0^{2\pi} e^{2u\cos\theta}d\theta.$$

Problem 5.29. Find the radius of convergence of the following power series:

$$\sum_{n=1}^{\infty}\frac{z^n}{n^a} \quad (a \in \mathbb{R})$$

Problem 5.30. Find the radius of convergence of the following power series:

$$\sum_{n=1}^{\infty}\frac{z^n}{\log n^2}, \quad \sum_{n=1}^{\infty}(3+(-1)^n)z^n, \quad \sum_{n=1}^{\infty}\frac{n!}{n^n}z^n.$$

Problem 5.31. Let the radius of convergence of a series $\sum_{n=1}^{\infty} u_n z^n$ be R. Find the radii of convergence of the following series:

$$\sum_{n=1}^{\infty}(2^n-1)u_n z^n, \quad \sum_{n=1}^{\infty}u_n\frac{z^n}{n!}, \quad \sum_{n=1}^{\infty}n^n u_n z^n.$$

Problem 5.32. Find a nonzero power series $y(x)$ that solves the following differential equation:

$$x^2y'' + xy' + (x^2-1)y = 0.$$

What is its radius of convergence?

Problem 5.33. Find power series expansions of the following functions:

$$\int_0^z e^{x^2}\,dx, \quad \int_0^z \frac{\sin x}{x}\,dx.$$

Find the radii of convergence of these power series.

Problem 5.34. Suppose that the power series $f(z) = \sum_{n=0}^{\infty} a_n z^n$ converges in the disk $|z| < 1 + \varepsilon$, where $\varepsilon > 0$. Find the area of the image of the unit disk $|z| < 1$ under f in terms of the coefficients a_n.

6. Differential 1-forms on the plane

In this section, we give a very down-to-earth outline of differential 1-forms on the plane and their integrals. Note that our assumptions are less restrictive than in most textbooks dealing with calculus on manifolds.

We let $\mathbb{R}^2$ denote the real two-dimensional plane, and use bold letters to denote points in $\mathbb{R}^2$. Define a *vector* at a point $\mathbf{a} \in \mathbb{R}^2$ as a pair of points $(\mathbf{a}, \mathbf{b})$ (the standard way to visualize this vector is to think of an arrow originating at $\mathbf{a}$ and terminating at $\mathbf{b}$). We will usually write $\mathbf{b} - \mathbf{a}$ for the vector $(\mathbf{a}, \mathbf{b})$. Fix a Cartesian coordinate system. Suppose that (a_x, a_y) are coordinates of $\mathbf{a}$, and (b_x, b_y) are coordinates of $\mathbf{b}$. Then the numbers $(b_x - a_x, b_y - a_y)$ are called *coordinates of the vector* $\mathbf{b} - \mathbf{a}$. The *addition of vectors* and the *multiplication of a vector by a real number* are performed coordinate-wise. Vector addition can be also visualized with the parallelogram rule familiar from the high-school. The set $T_{\mathbf{a}}\mathbb{R}^2$ of all vectors originating at a point $\mathbf{a} \in \mathbb{R}^2$ is called the *tangent plane* to $\mathbb{R}^2$ at $\mathbf{a}$. Since vectors can be added and multiplied by real numbers, the tangent plane $T_{\mathbf{a}}\mathbb{R}^2$ has a natural structure of a real vector space of dimension 2. We will write $T^*_{\mathbf{a}}\mathbb{R}^2$ for the *cotangent plane* to $\mathbb{R}^2$ at $\mathbf{a}$ consisting of all linear functionals on $T_{\mathbf{a}}\mathbb{R}^2$. The set $T^*_{\mathbf{a}}\mathbb{R}^2$ is also a vector space of dimension 2. A linear functional α that maps a vector with coordinates (dx, dy) to the number $Adx + Bdy$ is determined by the pair of numbers (A, B) that are called *coordinates* (or *coefficients*) of α. We will sometimes write α in the form $Adx+Bdy$.

Definition 6.1 (Differential 1-form). Let $U \subset \mathbb{R}^2$ be an open subset. A *differential 1-form* α on U is a correspondence assigning to every point $\mathbf{x} \in U$ an element $\alpha_{\mathbf{x}} \in T^*_{\mathbf{x}}\mathbb{R}^2$. We will always assume that $\alpha_{\mathbf{x}}$ depends continuously on $\mathbf{x}$. This means that

$$\alpha_{\mathbf{x}} = A(\mathbf{x})dx + B(\mathbf{x})dy$$

for some continuous functions A and B on U. We will sometimes need the stronger assumption that the functions A and B are *smooth*, i.e., sufficiently many times differentiable. In this case, the differential 1-form α is said to be smooth. For example, we will sometimes talk about continuously differentiable differential 1-forms meaning differential 1-forms, whose coefficients are continuously differentiable. We will sometimes refer to differential 1-forms as just 1-forms skipping the adjective "differential".

A nice thing about 1-forms is that they can (sometimes) be integrated over curves. Let $U \subset \mathbb{R}^2$ be an open subset. Recall that a *path* in U is a continuous map $\gamma : [0,1] \to U$. A path γ is said to be *simple* if γ is injective. In this case, the set $C = \gamma[0,1]$ (the image of γ) is called a *simple curve*. We will assume that a simple curve $C = \gamma[0,1]$ is equipped with the *orientation*, i.e., among the two *endpoints* of it, $\gamma(0)$ and $\gamma(1)$, one point, namely, $\gamma(0)$, is chosen as the *initial point*, and the other point $\gamma(1)$ is chosen as the *terminal point*. We also say that C *originates* at $\gamma(0)$ and *terminates* at $\gamma(1)$. The path γ is called a *parameterization* of C. In the sequel, by an *arc* we always mean a simple oriented curve.

Let us stress again that 1-forms, paths and curves are a priori continuous but may not be differentiable unless some kind of smoothness assumption is explicitly imposed.

Definition 6.2 (Integrating 1-forms over arcs). Let $\sigma \subset U$ be an arc in U, and α be a 1-form on U. We will use a parameterization $\gamma : [0,1] \to U$ of σ, however, nothing will depend on a particular choice of a parameterization. A *partition* τ of σ is a choice of points $\gamma(t_0)$, $\dots$, $\gamma(t_n)$ on the arc σ, where

$$0 = t_0 \leqslant t_1 \leqslant \dots \leqslant t_n = 1$$

is a partition of $[0,1]$. The *diameter* of τ is by definition the maximum of the numbers $|\gamma(t_{k+1}) - \gamma(t_k)|$, where k runs from 0 to $n-1$. Define the *integral sum* of α over σ corresponding to a partition τ as the number

$$S_\sigma(\alpha,\tau) = \sum_{k=0}^{n-1} \alpha_{\gamma(t_k)}(\gamma(t_{k+1}) - \gamma(t_k)).$$

Here the expression $\alpha_{\gamma(t_k)}(\gamma(t_{k+1}) - \gamma(t_k))$ means the value of the linear functional $\alpha_{\gamma(t_k)} \in T^*_{\gamma(t_k)}\mathbb{R}^2$ on the vector $\gamma(t_{k+1}) - \gamma(t_k) \in T_{\gamma(t_k)}\mathbb{R}^2$. We say that α is *integrable* over σ, and that the integral of α over σ is equal to I if, for every $\varepsilon > 0$ there exists $\delta > 0$ with the following property: for every partition τ, whose diameter is less than δ, we have $|S_\sigma(\alpha,\tau) - I| < \varepsilon$. This definition mimics the classical definition of a

Riemann integral. If α is integrable over σ, then we will write

$$\int_\sigma \alpha$$

for the integral of α over σ. By saying that this *integral exists*, we mean the same thing as by saying that α is integrable over σ.

Example 6.3. The 1-form dx is integrable over any arc σ in $\mathbb{R}^2$. Indeed, suppose that σ originates at $\mathbf{a}$ and terminates at $\mathbf{b}$. Then, for every partition τ, we have $S_\sigma(dx, \tau) = dx(\mathbf{b} - \mathbf{a})$. Here $dx(\mathbf{b} - \mathbf{a})$ means the value of the linear functional dx on the vector $\mathbf{b} - \mathbf{a}$, i.e., the difference of the x-coordinates of the points $\mathbf{b}$ and $\mathbf{a}$. More generally, consider a 1-form $\alpha = Adx + Bdy$ with constant coefficients, i.e., we assume that A and B are constant functions on $\mathbb{R}^2$. Then any integral sum of α over any arc σ is equal to the value of α at the vector connecting the endpoints of σ. Thus we have

$$\int_\gamma \alpha = \alpha(\mathbf{b} - \mathbf{a}),$$

where $\mathbf{a}$ is the initial point of σ, and $\mathbf{b}$ is the terminal point of σ.

We now discuss the dependence of integrals over arcs on the orientation of arcs. Let σ be an arc. We will write $-\sigma$ for the same arc, equipped with the opposite orientation, i.e. the initial and the terminal points interchanged. If $\gamma : [0, 1] \to \mathbb{R}^2$ is a parameterization of σ, then the arc $-\sigma$ can be parameterized by the path γ^* given by the formula $\gamma^*(t) = \gamma(1 - t)$. It is clear from the definition of the integral that

$$\int_{-\sigma} \alpha = -\int_\sigma \alpha.$$

Definition 6.4 (The group of 1-chains). Let $U \subset \mathbb{R}^2$ be an open subset. We define a *1-chain* in U as a formal sum of arcs. Thus any chain in U has the form

$$\Sigma = \sigma_1 + \cdots + \sigma_k,$$

where σ_1, …, σ_k are arcs in U, and $+$ is just a formal symbol that allows to combine different objects. The only properties we impose for the operation $+$ is that $+$ is commutative and satisfies the relation $\sigma + (-\sigma) = 0$. We will use standard abbreviations: 2σ for $\sigma + \sigma$, etc. Suppose that a 1-form α is integrable over each of the arcs $\sigma_1, \ldots, \sigma_k$. Then we define the integral of α over the chain Σ by the formula

$$\int_\Sigma \alpha = \int_{\sigma_1} \alpha + \cdots + \int_{\sigma_k} \alpha.$$

Note that the set $Z_1(U)$ of all 1-chains in U has a natural structure of an Abelian group. With the definition given above, we have

$$\int_{\Sigma+\Sigma'} \alpha = \int_{\Sigma} \alpha + \int_{\Sigma'} \alpha$$

provided that both integrals in the right-hand side exist.

We can now discuss some special types of arcs and chains. Suppose that a path $\gamma : [0,1] \to U$ is *continuously differentiable*, i.e., for every $t_0 \in (0,1)$, there exists the *derivative*

$$\dot{\gamma}(t_0) = \lim_{t \to t_0} \frac{1}{t - t_0} (\gamma(t) - \gamma(t_0))$$

(note that $\gamma(t) - \gamma(t_0)$ is a vector from $T_{\gamma(t_0)}\mathbb{R}^2$, and so is the derivative $\dot{\gamma}(t_0)$), which depends continuously on t_0. Here t_0 may be equal to 0 and 1, and t is always taken from $[0,1]$ (thus, for $t_0 = 0$ or 1, the limit in the right-hand side is a one-sided limit). Moreover, suppose that $\dot{\gamma}(t) \neq 0$ for all $t \in [0,1]$. Then the arc $\sigma = \gamma[0,1]$ is said to be *smooth*. A *smooth chain* is a chain consisting of smooth arcs. The subgroup of $Z_1(U)$ consisting of smooth chains is denoted by $Z_1^{sm}(U)$.

p:int_sm **Proposition 6.5.** *Any 1-form on U is integrable over any smooth chain in U.*

Proof. Let α be a 1-form on U. It suffices to prove that α is integrable over any smooth arc in U. Consider a smooth arc $\sigma = \gamma[0,1]$, where $\gamma : [0,1] \to U$ is a continuously differentiable path such that γ' never vanishes on $[0,1]$. If τ is a partition of σ consisting of points $\gamma(t_0) = \gamma(0), \ldots, \gamma(t_n) = \gamma(1)$, then the corresponding integral sum is equal to

$$S(\sigma,\tau) = \sum_{k=0}^{n-1} \alpha_{\gamma(t_k)}(\gamma(t_{k+1}) - \gamma(t_k)) = \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} \alpha_{\gamma(t_k)}(\dot{\gamma}(t))dt.$$

The coefficients of the 1-form α are continuous. Therefore, their restrictions to the arc σ are absolutely continuous. It follows that, for every $\varepsilon > 0$ there exists $\delta > 0$ with the following property: $|t - t'| < \delta$ implies that

$$|\alpha_{\gamma(t)}(\vec{v}) - \alpha_{\gamma(t')}(\vec{v})| < \varepsilon|\vec{v}|$$

for every vector $\vec{v}$. In particular, if the diameter of the partition τ is less than δ, then the number $\alpha_{\gamma(t_k)}(\dot{\gamma}(t))$ for $t \in [t_k, t_{k+1}]$ differs from the number $\alpha_{\gamma(t)}(\dot{\gamma}(t))$ at most by εM, where M is the maximum of $|\dot{\gamma}(t)|$ over all $t \in [0,1]$. Therefore, the integral sum $S(\sigma,\tau)$ differs from the integral

$$\int_0^1 \alpha_{\gamma(t)}(\dot\gamma(t))dt$$

at most by εM. Since ε, hence also εM, can be made arbitrarily small, the 1-form α is integrable over σ, and the integral is equal to

$$\int_\sigma \alpha = \int_0^1 \alpha_{\gamma(t)}(\dot\gamma(t))dt.$$

□

A *horizontal arc* is defined as a line interval of the form $[a, b] \times \{c\}$ for some real numbers $a < b$ and c, equipped with some orientation. A *vertical arc* is defined as a line interval of the form $\{a\} \times [c, d]$ for some real numbers $c < d$ and a, equipped with some orientation. The group of *coordinate 1-chains* $Z_1^{xy}(U)$ in U is defined as the subgroup of $Z_1(U)$ generated by all horizontal and vertical arcs. Coordinate chains are just a technical tool. They have no self-importance since the very notion of a coordinate chain depends on a specific choice of a coordinate system. Note that $Z_1^{xy}(U) \subset Z_1^{sm}(U)$.

Definition 6.6 (Exact 1-forms). Suppose that $U \subset \mathbb{R}^2$ is an open subset, and f is a continuously differentiable function on U (this is equivalent to saying that f has continuous partial derivatives everywhere in U). The differential $d_{\mathbf{x}}f$ of the function f at a point $\mathbf{x}$ is the linear functional

$$d_{\mathbf{x}}f = \frac{\partial f}{\partial x}(\mathbf{x})dx + \frac{\partial f}{\partial y}(\mathbf{x})dy$$

depending continuously on $\mathbf{x}$. Thus the correspondence $\mathbf{x} \mapsto d_{\mathbf{x}}f$ is a differential 1-form, which will be denoted by df and will be called the *differential* of f. A 1-form that can be written as df for some continuously differentiable function f is said to be *exact.*

The following theorem is a 2-dimensional analog of the Fundamental Theorem of Calculus:

Theorem 6.7. *The integral of an exact 1-form df on an open subset $U \subset \mathbb{R}^2$ over a smooth arc $\sigma \subset U$ is equal to*

$$\int_\sigma df = f(\mathbf{b}) - f(\mathbf{a}),$$

where $\mathbf{a}$ is the initial point of σ and $\mathbf{b}$ is the terminal point of σ.

Proof. Let $\gamma : [0, 1] \to U$ be a parameterization of σ. We may assume that γ is continuously differentiable. Set $\varphi(t) = f(\gamma(t))$. By the chain rule, we have $\frac{d}{dt}\varphi(t) = d_{\gamma(t)}f(\dot\gamma(t))$. Therefore, we have

$$f(\gamma(1)) - f(\gamma(0)) = \int_0^1 \frac{d}{dt}\varphi(t)\,dt = \int_0^1 d_{\gamma(t)}f(\dot\gamma(t))\,dt = \int_\sigma \alpha.$$

The last equality follows from Proposition 6.5. □

In particular, we see that the integral of an exact form over a smooth arc depends only on the endpoints of the arc and on the orientation but not on the arc itself.

Problems.

Problem 6.8. Prove that the boundary of the square $0 \leqslant x, y \leqslant 1$ can be represented in the form $\gamma[0,1]$, where γ is a continuously differentiable path.

Problem 6.9. Prove that any broken line can be represented in the form $\gamma[0,1]$, where γ is a continuously differentiable path.

Problem 6.10. Suppose that $U \subset \mathbb{R}^2$ is an open connected subset. Prove that, for any pair of points $\mathbf{a}$, $\mathbf{b} \in U$, there exists a coordinate path γ in U such that $\gamma(0) = \mathbf{a}$ and $\gamma(1) = \mathbf{b}$. *Hint:* consider the subset $V \subset U$ consisting of all points $\gamma(1)$, where γ is an arbitrary coordinate path originating at $\mathbf{a}$. Prove that V is open and that $U \setminus V$ is also open.

Problem 6.11. Compute

$$\int_\sigma x\,dz,$$

where σ is a directed line segment connecting 0 with $1 + i$.

Problem 6.12. Compute

$$\int_\sigma x\,dz,$$

where $\sigma = C_r(0)$ is the circle $\{|z| = r\}$ oriented counterclockwise (it can be represented as a smooth 1-cycle, see Section 8).

7. Homology

In this section, we discuss topology of open subsets $U \subset \mathbb{R}^2$.

By a *coordinate rectangle* $[a,b] \times [c,d]$, we mean the set of points $(x,y) \in \mathbb{R}^2$ such that $a \leqslant x \leqslant b$ and $c \leqslant y \leqslant d$. Consider a coordinate rectangle $\Delta = [a,b] \times [c,d]$. We define the *boundary* $\partial\Delta$ of Δ as the coordinate chain

$$\sigma_S + \sigma_E + \sigma_N + \sigma_W,$$

where σ_S (S from “South”) is the horizontal arc $[a,b] \times \{c\}$ oriented from left to right, σ_E (E from “East”) is the vertical arc $\{b\} \times [c,d]$ oriented from the bottom to the top, σ_N (N from “North”) is the horizontal arc $[a,b] \times \{d\}$ oriented from right to left, and σ_W (W from “West”) is the vertical arc $\{a\} \times [c,d]$ oriented from the top to the bottom.

Definition 7.1 (Boundaries and cycles). Boundaries of all coordinate rectangles are elements of $Z_1^{xy}(U)$. Let $B_1^{xy}(U)$ be the subgroup of $Z_1^{xy}(U)$ generated by the boundaries of all coordinate rectangles. All elements of the group $B_1^{xy}(U)$ are called *boundaries*. Define 0-*chains* in U as formal linear combinations of points in U with integer coefficients. Thus any 0-chain in U has the form

$$m_1\mathbf{a}_1 + \cdots + m_k\mathbf{a}_k,$$

where $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are points in U, and + is a formal operation that just serves to combine points. We assume that this operation is commutative and satisfies the property $m\mathbf{a}+m'\mathbf{a} = (m+m')\mathbf{a}$. We also identify all 0-chains of the form $0\mathbf{a}$. If all points $\mathbf{a}_j$ are different, then the corresponding coefficients m_j are called *multiplicities* of points $\mathbf{a}_j$. Clearly, the set $Z_0(U)$ of all 0-chains thus defined is an Abelian group. There is a natural homomorphism $\partial : Z_1(U) \to Z_0(U)$ called the *boundary map*. It suffices to define the homomorphism ∂ on the set of generators of $Z_1(U)$, i.e., on arcs. For every arc σ with the initial point $\mathbf{a}$ and the terminal point $\mathbf{b}$, we set $\partial\sigma = \mathbf{b} - \mathbf{a}$. Note that, for every coordinate rectangle Δ in U, the boundary of Δ is in the kernel of the boundary homomorphism $\partial(\partial\Delta) = 0$. Although here two instances of ∂ mean different maps defined on different sets, we keep the same letter since in both cases it means taking the boundary. Define the *group of 1-cycles* $C_1(U)$ as the kernel of the boundary map $\partial : Z_1(U) \to Z_0(U)$. We set $C_1^{xy}(U) = C_1(U) \cap Z_1^{xy}(U)$. The elements of the group $C_1^{xy}(U)$ are called *coordinate 1-cycles*.

Note that $B_1^{xy}(U) \subset C_1^{xy}(U)$ since $\partial(\partial\Delta) = 0$ for every coordinate rectangle Δ in U.

Definition 7.2 (the first homology group). Although the groups $C_1^{xy}(U)$ and $B_1^{xy}(U)$ are always infinite, the quotient

$$H_1^{xy}(U) = C_1^{xy}(U)/B_1^{xy}(U)$$

is very often a finite group (in fact, it is finite whenever U has only finitely many "holes"). This group is called the *first homology group* of U (with integer coefficients). A 1-cycle is said to be *homologous to zero* if it represents the zero element of $H_1^{xy}(U)$. This is equivalent to say that the given 1-cycle is a boundary. A connected open set U is said to be *simply connected* if the group $H_1^{xy}(U)$ is trivial.

ex:collin *Example* 7.3. Consider the following chain:

$$\sigma = \sigma_1 + \sigma_2 - \sigma_3,$$

where $\sigma_1 = [a,b] \times \{d\}$ and $\sigma_2 = [b,c] \times \{d\}$ are two horizontal intervals sharing an endpoint (σ_1 being oriented from (a,d) to (b,d), and σ_2 being oriented from (b,d) to (c,d)), and σ_3 is the horizontal interval $[a,c] \times \{d\}$. We claim that σ is a boundary. Indeed, choose some $e > d$ and consider the following coordinate rectangles

$$\Delta_1 = [a,b] \times [d,e], \quad \Delta_2 = [b,c] \times [d,e], \quad \Delta_2 = [a,c] \times [d,e].$$

We claim that $\sigma = \partial(\Delta_1) + \partial(\Delta_2) - \partial(\Delta_3)$ (this is clear from the definition of the homomorphism ∂).

ex:rect *Example* 7.4. Let U be an open coordinate rectangle $(A,B) \times (C,D)$, where $A < B$ and $C < D$. We claim that U is simply connected, i.e., every coordinate 1-cycle σ is a boundary. Suppose that

$$\sigma = \sigma_1 + \cdots + \sigma_m,$$

where each σ_k is a horizontal or a vertical arc. We will use induction by m. If $m \leqslant 4$, then either all σ_k lie on the same line, or there is a coordinate rectangle, whose boundary coincides with σ. In the former case, σ is also a boundary by example 7.3. Suppose now that $m > 4$. It suffices to prove that σ is homologous to some chain consisting of less than m arcs.

Without loss of generality, we may assume that σ_1 is horizontal. There must be some k such that σ_k shares an endpoint with σ_1. Permuting the summands of σ if necessary, we can arrange that $k = 2$. If σ_2 is horizontal, then the sum $\sigma_1 + \sigma_2$ is homologous to one horizontal arc by Example 7.3. Therefore, σ is homologous to a chain consisting of at most $m-1$ terms. Thus we can assume that σ_2 is vertical. Let $\mathbf{y}$ be the endpoint of σ_2 that is not an endpoint of σ_1. There is some k such that $\mathbf{y}$ is an endpoint of σ_k. Without loss of generality, we may assume that $k = 3$. If σ_3 is vertical, then we can reduce the number of summands in σ by Example 7.3. Thus we may assume that σ_3 is horizontal. Let $\mathbf{x}$ be the endpoint of σ_1 that is not an endpoint of σ_2, and let $\mathbf{z}$ be the endpoint of σ_3 that is not an endpoint of σ_2. We must have $\partial(\sigma_1 + \sigma_2 + \sigma_3) = \pm(\mathbf{z} - \mathbf{x})$. Assume that the sign is plus (otherwise, we just change the orientations of σ_1, σ_2 and σ_3 simultaneously)

Let Π be the coordinate rectangle, whose diagonal is the line interval connecting $\mathbf{x}$ and $\mathbf{z}$. There are two coordinate arcs σ_1' and σ_2' from the boundary of Π such that $\partial(\sigma_1' + \sigma_2') = \mathbf{z} - \mathbf{x}$. We claim that the chain

$$\sigma_1' + \sigma_2' - \sigma_1 - \sigma_2 - \sigma_3$$

is a boundary. This can be easily shown with the help of Example 7.3. Then we can reduce the number of summands of σ by replacing

$\sigma_1+\sigma_2+\sigma_3$ with $\sigma_1'+\sigma_2'$. (There are several cases to consider but they are straightforward).

Example 7.5. Suppose that U is a round disk. Then U is simply connected. This can be shown in almost the same way as in Example 7.4. The difference is that the rectangle Π is not necessarily contained in U. On the other hand, there is at least one vertex of Π different from $\mathbf{x}$ and $\mathbf{z}$ that is contained in U. We now choose σ_1' and σ_2' to be the edges of Π that contain this vertex. These edges are contained in U. The rest of the proof is the same.

Suppose now that $\alpha = Adx + Bdy$ is a differential 1-form on an open subset $U \subset \mathbb{R}^2$.

Definition 7.6 (Closed 1-forms). We say that α is *closed* if the following two assumptions are fulfilled:

(1) the functions A and B are differentiable in U;
(2) the partial derivatives of these functions satisfy the relation

$$\frac{\partial A}{\partial y} = \frac{\partial B}{\partial x}.$$

Note that we do not assume that the partial derivatives of the functions A and B are continuous.

We will see that closed 1-forms have some remarkable properties. In particular, we will prove that a closed form on a simply-connected domain is exact. Consider an open subset $U \subset \mathbb{R}^2$. Suppose that $\alpha = Adx + Bdy$ is a closed 1-form on U.

p:cl-rect1 **Proposition 7.7.** *If the partial derivatives* $\frac{\partial A}{\partial y}$ *and* $\frac{\partial B}{\partial x}$ *are continuous, then the integral of* α *over the boundary of any coordinate rectangle contained in* U *is zero.*

This proposition is easy to prove. The main technical difficulty that we would need to overcome later is to avoid the assumption of continuity of $\frac{\partial A}{\partial y}$ and $\frac{\partial B}{\partial x}$.

Proof. Let $\Delta = [a,b]\times[c,d]$ be a coordinate rectangle contained in U. By definition, the integral of α over the boundary of Δ is equal to

$$\int_a^b (A(x,c)-A(x,d))dx + \int_c^d (B(b,y)-B(a,y))dy.$$

By the Fundamental Theorem of Calculus,

$$\int_a^b (A(x,c)-A(x,d))dx = -\iint_\Delta \frac{\partial A}{\partial y}\,dx\,dy,$$

$$\int_c^d (B(b,y) - B(a,y))dy = \iint_\Delta \frac{\partial B}{\partial x}\, dx\, dy.$$

Substituting these expressions, we obtain the desired. □

We now want to prove the same statement without assuming that the partial derivatives $\frac{\partial A}{\partial y}$ and $\frac{\partial B}{\partial x}$ are continuous. This is significantly more complicated.

l:rectangle **Lemma 7.8.** *Suppose that a 1-from α on an open subset $U \subset \mathbb{R}^2$ is closed. Then, for every point $\mathbf{a} \in U$ and every $\varepsilon > 0$, there exists $\delta > 0$ with the following property. For every coordinate rectangle Δ containing the point $\mathbf{a}$, whose diagonal has length $\lambda < \delta$, we have*

$$\left| \int_{\partial\Delta} \alpha \right| < \varepsilon\lambda^2.$$

Intuitively, Lemma 7.8 says that the integral of α over the boundary of a small rectangle is much smaller than the square of the diagonal of this rectangle.

Proof. Suppose that $\mathbf{a} = (a, b)$. Fix $\varepsilon > 0$. Since the function A is differentiable, there exists $\delta_A > 0$ such that the function

$$A^*(x,y) = A(x,y) - A(a,b) - \frac{\partial A}{\partial x}(a,b)(x-a) - \frac{\partial A}{\partial y}(a,b)(y-b)$$

satisfies the inequality $|A^*(\mathbf{x})| < \varepsilon|\mathbf{x} - \mathbf{a}|$ for every point $\mathbf{x} = (x, y)$ with $|\mathbf{x} - \mathbf{a}| < \delta_A$. Similarly, there exists $\delta_B > 0$ such that the function

$$B^*(x,y) = B(x,y) - B(a,b) - \frac{\partial B}{\partial x}(a,b)(x-a) - \frac{\partial B}{\partial y}(a,b)(y-b)$$

satisfies the inequality $|B^*(\mathbf{x})| < \varepsilon|\mathbf{x} - \mathbf{a}|$ for every point $\mathbf{x} = (x, y)$ with $|\mathbf{x} - \mathbf{a}| < \delta_B$. Set δ to be the minimum of δ_A and δ_B.

Choose any coordinate rectangle $\Delta = [a_0, a_1] \times [b_0, b_1]$ such that $a_0 \leqslant a \leqslant a_1$ and $b_0 \leqslant b \leqslant b_1$ and such that the distance λ between the points $\mathbf{a}_0 = (a_0, b_0)$ and $\mathbf{a}_1 = (a_1, b_1)$ satisfies the inequality $\lambda < \delta$.

We first estimate the integral of the form $\alpha^* = A^*dx + B^*dy$. Note that $|\mathbf{x} - \mathbf{a}| < \lambda$ for all $\mathbf{x} \in \Delta$. It follows that $|A^*| < \varepsilon\lambda$ in Δ. Therefore, we have

$$\left| \int_{\partial\Delta} A^* dx \right| \leqslant 2 \int_{a_0}^{a_1} |A^*| dx < 2\varepsilon\lambda(a_1 - a_0) \leqslant 2\varepsilon\lambda^2.$$

Similarly, the integral of B^* over $\partial\Delta$ has modulus $< 2\varepsilon\lambda^2$, therefore,

$$\left| \int_{\partial\Delta} \alpha^* \right| < 4\varepsilon\lambda^2.$$

On the other hand, we can compare the integral of α over $\partial\Delta$ with the integral of α^*:

$$\int_{\partial\Delta} \alpha - \alpha^* = \int_{\partial\Delta} \frac{\partial A}{\partial y}(a,b)(y-b)dx + \frac{\partial B}{\partial x}(a,b)(x-a)dy.$$

We used here that the integrals of the forms dx, $(x-a)dx$, dy, $(y-b)dy$ over $\partial\Delta$ are equal to zero. Let $|\Delta|$ denote the area of Δ, i.e., the number $(a_1-a_0)(b_1-b_0)$. Note that

$$\int_{\partial\Delta} \frac{\partial A}{\partial y}(a,b)(y-b)dx = -\frac{\partial A}{\partial y}(a,b)|\Delta|,$$
$$\int_{\partial\Delta} \frac{\partial B}{\partial x}(a,c)(x-a)dy = \frac{\partial B}{\partial x}(a,c)|\Delta|.$$

These two numbers sum up to zero! Therefore, the forms α and α^* have the same integrals over $\partial\Delta$. It follows that the modulus of the integral of the form α over $\partial\Delta$ is bounded above by the number $4\varepsilon\lambda^2$. Since ε was arbitrary, we can replace 4ε with ε. ☐

We can now get rid of the continuity assumption in Proposition 7.7. The following Proposition is due to Goursat.

p:cl-rect2

Proposition 7.9. *Suppose that α is a closed 1-form on an open subset $U \subset \mathbb{R}^2$. Then the integral of α over the boundary of any coordinate rectangle contained in U is equal to zero.*

Proof. We will use bisection. Let Δ be any coordinate rectangle contained in U. Draw the two line intervals connecting the midpoints of opposite edges of Δ. These two line intervals divide Δ into 4 smaller rectangles. Note that the integral of α over $\partial\Delta$ is equal to the sum of integrals of α over the boundaries of these four smaller rectangles. Suppose that the integral of α over $\partial\Delta$ has nonzero modulus $I(\Delta)$. Then there is at least one smaller rectangle $\Delta_1 \subset \Delta$ (among the four rectangles we considered) such that $I(\Delta_1) \geqslant I(\Delta)/4$. Continuing this bisection process, we obtain a nested sequence of coordinate rectangles $\Delta \supset \Delta_1 \supset \Delta_2 \supset \dots$ with the following properties. The rectangle Δ_n is 2^n times smaller than the rectangle Δ. In particular, the length λ_n of a diagonal of Δ_n is related with the length λ of a diagonal of Δ as follows: $\lambda_n = 2^{-n}\lambda$. By construction, we have $I(\Delta_n) \geqslant 4^{-n}I(\Delta)$. It follows that the numbers $I(\Delta_n)/\lambda_n^2$ are bounded below by $I(\Delta)/\lambda^2$.

On the other hand, we have $\lambda_n \to 0$, in particular, there is exactly one point $\mathbf{a}$ in all rectangles Δ_n. Choose $\varepsilon = I(\Delta)/\lambda^2$. By Lemma 7.8, for sufficiently large n, we must have $I(\Delta_n) < \varepsilon\lambda_n^2$, a contradiction. ☐

cor:homol0 **Corollary 7.10.** *Let α be a closed 1-form on an open subset $U \subset \mathbb{R}^2$. Suppose that a coordinate 1-chain σ in U is homologous to zero. Then the integral of α over σ is equal to zero.*

Proof. Indeed, σ is a linear combination of boundaries of coordinate rectangles, and the integral of α over any such boundary is equal to zero. □

t:cl-ex **Theorem 7.11.** *Suppose that $U \subset \mathbb{R}^2$ is a simply connected open set, and α is a closed 1-form on U. Then α is exact, i.e., we have $\alpha = df$ for some continuously differentiable function $f : U \to \mathbb{R}$.*

Proof. Fix any point $\mathbf{a} \in U$. For a point $\mathbf{x} \in U$, let $\sigma(\mathbf{x})$ be a coordinate 1-chain in U, whose boundary coincides with $\mathbf{x} - \mathbf{a}$. Set

$$f(\mathbf{x}) = \int_\sigma (\mathbf{x})\alpha.$$

We first need to show that $f(\mathbf{x})$ is well defined, that is, it depends only on $\mathbf{x}$ rather than on a particular choice of the 1-chain $\sigma(\mathbf{x})$. Indeed, let $\sigma'(\mathbf{x})$ be a different 1-chain with the property $\partial\sigma'(\mathbf{x}) = \mathbf{x} - \mathbf{a}$. Then $\partial(\sigma(\mathbf{x}) - \sigma'(\mathbf{x})) = 0$. It follows that the 1-chain $\sigma(\mathbf{x}) - \sigma'(\mathbf{x})$ is a 1-cycle. On the other hand, since U is simply connected, every 1-cycle in U is also a boundary. It follows that the integral of α over the 1-cycle $\sigma(\mathbf{x}) - \sigma'(\mathbf{x})$ is equal to zero; therefore, the integral of α over $\sigma(\mathbf{x})$ coincides with the integral of α over $\sigma'(\mathbf{x})$. We have thus proved that $f(\mathbf{x})$ depends only on $\mathbf{x}$.

We now compute partial derivatives of the function f at a point $\mathbf{x}$. Take a point $\mathbf{x} = (x, y)$ and a point $\mathbf{x}' = (x', y)$. The partial derivative $\frac{\partial f}{\partial x}$ is by definition the limit of the expression

$$\frac{f(\mathbf{x}') - f(\mathbf{x})}{x' - x}$$

as $x' \to x$. We want to show that this limit exists and, simultaneously, compute this limit. Fix a 1-chain $\sigma(\mathbf{x})$ with the property $\partial\sigma(\mathbf{x}) = \mathbf{x} - \mathbf{a}$. Let $[\mathbf{x}, \mathbf{x}']$ be the horizontal arc connecting $\mathbf{x}$ with $\mathbf{x}'$ and oriented from $\mathbf{x}$ to $\mathbf{x}'$. We can set $\sigma(\mathbf{x}') = \sigma(\mathbf{x}) + [\mathbf{x}, \mathbf{x}']$. Then we have, by the definition of the function f, that

$$f(\mathbf{x}') = f(\mathbf{x}) + \int_{[\mathbf{x},\mathbf{x}']} \alpha.$$

Therefore,

$$\lim_{x' \to x} \frac{f(\mathbf{x}') - f(\mathbf{x})}{x' - x} = \lim_{x' \to x} \frac{1}{x' - x} \int_x^{x'} A(t, y)dt = A(x, y).$$

Thus we obtained that $\frac{\partial f}{\partial x} = A$. Similarly, we obtain that $\frac{\partial f}{\partial y} = B$. Since the first order partial derivatives of f are defined and are continuous in U, the function f is differentiable in U, and its differential is given by

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy = \alpha.$$

Thus we have proved the desired claim: any closed 1-from in a simply connected domain has the form df for some continuously differentiable function f. □

s:cauchy-int

8. The Cauchy integral theorem

Let $U \subset \mathbb{C}$ be an open set, and $f : U \to \mathbb{C}$ be a complex-valued function. Then there are two real-valued functions $u : U \to \mathbb{R}$ and $v : U \to \mathbb{R}$ such that $f = u + iv$. We can define the *differential* of the function f by the formula $df = du + idv$. The following proposition is straightforward.

Proposition 8.1. *Suppose that f has complex derivative $f'(a)$ at a point $a \in U$. Then the differential $d_a f$ is the multiplication by the complex number $f'(a)$, i.e. we have $d_a f(dz) = f'(a)dz$ for every $dz \in T_a\mathbb{C}$.*

Recall that the multiplication by $f'(a)$ is the composition of a rotation by angle $\arg f'(a)$ and the dilation with the coefficient $|f'(a)|$. In particular, this transformation preserves angles. Suppose that $f'(a) = p + iq$. Then the matrix of the linear map $d_a f : dz \mapsto f'(a)dz$ is

$$\begin{pmatrix} p & q \\ -q & p \end{pmatrix}.$$

On the other hand, the same matrix is equal to the Jacobi matrix of f at the point a, i.e., it can be written in terms of partial derivatives of u and v as follows:

$$\begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}.$$

Comparing the two matrices, we conclude that

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

These differential relations on the functions u and v are called the *Cauchy–Riemann relations*. We have proved that the real and the imaginary parts of a holomorphic function always satisfy the Cauchy–Riemann relations.

A function $u(x, y)$ of two variables x and y is said to be *harmonic* if it satisfies the *Laplace equation*

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

Here $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the so called *Laplace operator*. It follows from the Cauchy–Riemann relations that if a function $f = u+iv$ is holomorphic, then u and v are harmonic functions. The Laplace equation appears in Mathematical Physics. E.g. it describes a stationary distribution of temperature in a medium, it provides a useful approximation (a linearization) of the equation describing minimal surfaces.

Theorem 8.2. *Suppose that f is a holomorphic function on an open subset $U \subset \mathbb{C}$. Then the differential 1-form $\alpha = f(z)dz$ is closed.*

Proof. We have

$$\alpha = (u + iv)(dx + idy) = Adx + Bdy,$$

where $A = u+iv$ and $B = -v+iu$. By definition of a holomorphic function, the functions A and B are differentiable. The Cauchy–Riemann relations imply that $\frac{\partial A}{\partial y} = \frac{\partial B}{\partial x}$. Thus α has both properties imposed in the definition of a closed 1-form. □

Theorem 7.11 now implies the following:

Corollary 8.3. *If $U \subset \mathbb{C}$ is a simply connected open subset, and $f : U \to \mathbb{C}$ is a holomorphic function, then the integral of the form $f(z)dz$ over any 1-cycle is equal to zero. Moreover, we have $f(z)dz = dF$ for some holomorphic function $F : U \to \mathbb{C}$.*

Proof. We need only prove that the function F is holomorphic. We know that it is continuously differentiable. Since the differential of F at $z \in U$ is the multiplication by the complex number $f(z)$, the function F has complex derivative $f(z)$. □

The function F is called the *antiderivative* (or the *primitive*) of the function f.

Consider an open subset $U \subset \mathbb{C}$, a point a. Suppose that the disk of radius R around a is contained in U. Let us now define some chain $C_R(a)$ in U. Intuitively, this is just the circle $\{|z - a| = R\}$ oriented counterclockwise. Formally, we need to define $C_R(a)$ as a linear combination of arcs. Let σ_+ and σ_- be the upper and the lower semicircles of $\{|z - a| = R\}$ oriented counterclockwise. We define $C_R(a)$ as $\sigma_+ + \sigma_-$. It is clear that $\partial C_R(a) = 0$.

The following is a particular case of the Cauchy integral formula:

t:cauchy-int **Theorem 8.4.** *Let $U \subset \mathbb{C}$ be an open subset, and $f : U \to \mathbb{C}$ be a holomorphic function. Suppose that $a \in U$ and that $C_R(a)$ is a chain in U for some $r > 0$. Then we have*

$$f(a) = \frac{1}{2\pi i} \int_{C_R(a)} \frac{f(z)dz}{z-a}.$$

Note that the form $\frac{f(z)dz}{z-a}$ is defined on the set $U \setminus \{a\}$ but not on the set U. To prove Theorem 8.4, we need the following lemma.

l:rR **Lemma 8.5.** *Suppose that both $C_r(a)$ and $C_R(a)$ (for some $0 < r < R$) are chains in U, and α is a closed 1-form on $U \setminus \{a\}$. Then we have*

$$\int_{C_r(a)} \alpha = \int_{C_R(a)} \alpha.$$

We give a proof modulo two topological facts that are intuitively obvious. They are also rather straightforward, but the proofs of these facts are left to the reader.

Sketch of a proof. Let A be the closed annulus $\{r \leqslant |z-a| \leqslant R\}$. Consider a sufficiently small arc σ of the circle $\{|z-a| = r\}$ with endpoints b and c. Let Δ be the coordinate rectangle such that b and c are its opposite vertices. Since σ is sufficiently small, two edges of Δ are contained in A. We will write $\tilde{\sigma}$ for the sum of these two edges oriented from b to c. There is a simply connected neighborhood V of σ in $U \setminus \{a\}$ that contains both σ and $\tilde{\sigma}$ (this is the first topological fact we are using without detailed justification). Since α is exact on V, the integral of α over σ coincides with the integral of α over $\tilde{\sigma}$. Subdividing the circle $\{|z-a| = r\}$ into sufficiently small arcs, we can replace the integral of α over $C_r(a)$ with the integral of α over some coordinate 1-circle $\tilde{C}_r(a)$ contained in a small neighborhood of $C_r(a)$. These two integrals are equal. Similarly, there is a coordinate 1-chain $\tilde{C}_R(a)$ contained in a small neighborhood of $C_R(a)$ such that the integral of α over $\tilde{C}_R(a)$ is equal to the integral of α over $C_R(a)$.

The second topological fact we are using without detailed justification is the following: $\tilde{C}_R(a) - \tilde{C}_r(a) \in B_1^{xy}(U \setminus \{a\})$. It follows that the integrals of α over $\tilde{C}_R(a)$ and over $\tilde{C}_r(a)$ are the same. □

Proof of Theorem 8.4. Let $\varepsilon > 0$ be a small positive number. It follows from Lemma 8.5 that we may assume the radius R to be sufficiently small. In particular, we may assume that $|f(z) - f(a)| < \varepsilon$ for all z such that $|z-a| = R$. Note that

$$\left| \frac{1}{2\pi i} \int_{C_R(a)} \frac{f(z) - f(a)}{z-a} dz \right| < \varepsilon.$$

Indeed, the absolute value of the function $(f(z) - f(a))/(z - a)$ is bounded above by εR^{-1}, and the length of $C_R(a)$ is $2\pi R$. Therefore, it suffices to compute the integral

$$\frac{1}{2\pi i}\int_{C_R(a)}\frac{f(a)dz}{z-a}$$

and show that this integral is equal to $f(a)$. We will use the parameterization $z = a + Re^{it}$ of $C_R(a)$.

A straightforward computation yields:

$$\int_{C_R(a)}\frac{dz}{z-a} = \int_0^{2\pi} R^{-1}e^{-it}d(Re^{it}) = \int_0^{2\pi} i dt = 2\pi i.$$

The theorem follows. □

We will now derive several consequences of Theorem 8.4.

Theorem 8.6 (The Mean Value Theorem). *Suppose that $f : U \to \mathbb{C}$ is a holomorphic function, and $B_r(a) \subset U$ is a disk contained in U. Then the value of f at a equal the mean value of f on the circle $\partial B_r(a)$:*

$$f(a) = \frac{1}{2\pi}\int_0^{2\pi} f(a + re^{it})dt.$$

Proof. This follows directly from Theorem 8.4 and from the observation that

$$\frac{dz}{z-a} = i dt.$$

(Both expressions can be regarded as differential 1-forms on $\mathbb{R}$ by setting $z = a + re^{it}$). □

c:upper **Corollary 8.7.** *Suppose that a function f is defined and holomorphic on some open set containing the disk $\{|z - a| \leqslant R\}$. Then we have*

$$|f(a)| \leqslant \max_{|z-a|=R} |f(z)|.$$

In particular, the maximal value of $|f|$ on the disk $\{|z - a| \leqslant R\}$ coincides with the maximal value of $|f|$ on the circle $\{|z - a| = R\}$.

In fact, a stronger statement is true:

t:maxmod **Theorem 8.8** (Maximum Modulus Principle). *Let $U \subset \mathbb{C}$ be a bounded connected open set, and let $f : \overline{U} \to \mathbb{C}$ be a continuous function, whose restriction to U is holomorphic. Then we have*

$$\max_{z\in\overline{U}} |f(z)| = \max_{z\in\partial U} |f(z)|.$$

Here $\overline{U}$ stands for the closure of U, and ∂U stands for the boundary of U.

In other words, the maximum of $|f|$ on $\overline{U}$ is attained on the boundary.

Proof. Assume the contrary: there is some $a \in U$ such that $|f(a)|$ is bigger than $m = \max_{z\in\partial U} |f(z)|$. We may assume that $|f|$ attains its maximum at the point a. Set $M = |f(a)|$. Let $r > 0$ be a positive real number with the property $B_r(a) \subset$ By the mean value property, $f(a)$ is equal to the mean value of f on any circle C centered at a, whose radius is less that r. Since all values of $|f|$ on C do not exceed M, the only possibility is that $|f|$ is constant on C and is equal to M. It follows that $|f|$ is constant on some neighborhood of a point a. Consider the set $X \subset U$ consisting of all points $x \in U$ with $|f(x)| = M$. The argument explained above implies that the set X is open. On the other hand, it is obviously closed in U (i.e. $\overline{X} \cap U = X \cap U$) since X is the full preimage of the set $\{M\}$ under the map $|f|$. It follows from the continuity that $X = U$, a contradiction with our assumption. □

As an illustration of the Maximum Modulus Principle, we prove the *Fundamental Theorem of Algebra*, which asserts that any complex polynomial P of degree $d > 0$ has at least one complex root, i.e. there is a point $a \in \mathbb{C}$ such that $P(a) = 0$. Assume the contrary: $P \neq 0$ at all points of $\mathbb{C}$. Note that the values of $|P(z)|$ are large for large z (in fact, we have $|P(z)| > c|z|^d$ for some $c > 0$ and all sufficiently large z). It follows that the minimum of the function $|P|$ is attained in the finite part of the plane. This minimum is nonzero by our assumption. But then the function $1/P$ is holomorphic on $\mathbb{C}$ and attains its maximum in an interior point, which contradicts the Maximum Modulus Principle.

We will now state a more general version of the Cauchy integral formula:

t:cauchy-int2

Theorem 8.9. *Let $U \subset \mathbb{C}$ be an open subset and $f : U \to \mathbb{C}$ be a holomorphic function. Suppose that a disk $\{|z - a| \leqslant r\}$ is contained in U, and that a point z lies in this disk. Then*

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

Proof. It suffices to show that the integral over the 1-cycle $C_r(a)$ coincides with the integral over the 1-cycle $C_\varepsilon(z)$ for a sufficiently small $\varepsilon > 0$. □

8.1. **Problems.**

harm-vs-hol

Problem 8.10. Let $U \subset \mathbb{C}$ be a simply connected domain, and $u : U \to \mathbb{R}$ be a harmonic function on U. Prove that u is the real part of some holomorphic function on U.

Solution. We want to find a function v such that the function $f = u + iv$ is holomorphic. Functions u and v are subject to the Cauchy–Riemann equations, which allow to express both partial derivatives of v in terms of u:

$$v_x = u_y, \quad v_y = -u_x.$$

Therefore, the differential of the function v must coincide with the 1-form $\omega = u_y\,dx - u_x\,dy$.

We now consider the form ω thus defined and show that it is indeed the differential of some function. Since the domain U is simply connected, it suffices to verify that the 1-form ω is closed. The coefficients $A = u_y$ and $B = -u_x$ of ω are clearly differentiable. The identity $\frac{\partial A}{\partial y} = \frac{\partial B}{\partial x}$ is equivalent to the identity $u_{xx} + u_{yy} = 0$, which expresses the fact that u is harmonic.

We have proved that $\omega = dv$ for some twice differentiable function v. It follows that u and v satisfy the Cauchy–Riemann relations in U. This and the fact that u and v are continuously differentiable implies that the function $f = u + iv$ is holomorphic on U.

Problem 8.11. Which of the following functions are real parts of holomorphic functions in the unit disk $\{|x + iy| < 1\}$:

(1) $u(x, y) = x^2 - axy + y^2$,
(2) $u(x, y) = \frac{x^2-y^2}{(x^2+y^2)^2}$,
(3) $u(x, y) = x^3$,
(4) $u(x, y) = \log((x - 2)^2 + y^2)$?

Rigorously justify your answer.

Problem 8.12. Compute

$$\int_{C_1(0)} \frac{e^z}{z} dz.$$

Problem 8.13. Compute

$$\int_{C_2(0)} \frac{dz}{z^2 - 1}.$$

Problem 8.14. Let $f : U \to \mathbb{C}$ be a holomorphic function, and let $\sigma \in C_1^{sm}(U)$ be a smooth cycle. Prove that the integral

$$\int_\sigma \overline{f(z)} f'(z)\,dz$$

is purely imaginary (i.e., has zero real part).

Problem 8.15. Let $f : U \to \mathbb{C}$ be a holomorphic function such that $|f(z) - 1| < 1$ for all $z \in U$, and let $\sigma \in C_1^{sm}(U)$ be a smooth cycle. Prove that

$$\int_\sigma \frac{f'(z)}{f(z)} dz = 0$$

Problem 8.16. Let $P : \mathbb{C} \to \mathbb{C}$ be a polynomial, whose coefficients are known. Evaluate the integral

$$\int_{C_R(a)} P(z)\,d\overline{z}.$$

Problem 8.17. Let $a_1, \dots, a_n$ be points on the unit circle. Prove that there is a point z on the unit circle such that the product of distances from z to $a_1, \dots, a_n$ is at least one.

Problem 8.18. Suppose that a function f is defined and holomorphic in the annulus $\{1-\varepsilon \leqslant |z| \leqslant 2+\varepsilon\}$ for some $\varepsilon > 0$. Prove that if $|f(z)| \leqslant 1$ for $|z| = 1$ and $|f(z)| \leqslant 4$ for $|z| = 2$, then $|f(z)| \leqslant |z|^2$ for $1 \leqslant |z| \leqslant 2$.

Problem 8.19. Suppose that a function f is defined and holomorphic on the disk $\{|z| < 2\}$. Given that $|f| \leqslant 10$ and that $f(1) = 0$ find the best upper bound for $|f(\frac{1}{2})|$.

9. Uniform convergence

Let $X \subset \mathbb{C}$ be any subset, and $f_n : X \to \mathbb{C}$ a sequence of functions. We say that f_n converges *uniformly* to a function $f : X \to \mathbb{C}$ if, for every ε, there is n_0 such that $|f_n(x) - f(x)| < \varepsilon$ for all $n > n_0$ and all $x \in X$. A series is said to be uniformly convergent if the associated sequence of partial sums is uniformly convergent.

p:conv **Proposition 9.1.** *Let $U \subset \mathbb{C}$ be an open subset, and $f_n : U \to \mathbb{C}$ a sequence of continuous functions converging uniformly to a function $f : U \to \mathbb{C}$. For any smooth chain $\sigma \in Z_1^{sm}(U)$, we have*

$$\int_\sigma f_n(z)dz \to \int_\sigma f(z)dz.$$

Proof. Set $\delta_n = \sup_{x\in U} |f_n(x) - f(x)|$. Then $\delta_n \to 0$ as $n \to \infty$. Then

$$\left| \int_\sigma f_n(z)dz - \int_\sigma f(z)dz \right| \leqslant \delta_n \cdot Length(\sigma).$$

□

t:hol-ser **Theorem 9.2.** *Suppose that a $\mathbb{C}$-valued function f is defined and holomorphic on a neighborhood of a disk $\{|z-a| \leqslant r\}$. Then f can be represented as the sum of a convergent power series in the disk $\{|z-a| < r\}$:*

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

Proof. We will use the Cauchy integral formula:

$$f(z) = \frac{1}{2\pi i} \int_{C_r(z)} \frac{f(\zeta)dz}{\zeta - z}$$

as stated in Theorem 8.9. We have

$$\frac{f(\zeta)}{\zeta - z} = \frac{f(\zeta)}{\zeta - a}\left(1 + \frac{z-a}{\zeta - a} + \left(\frac{z-a}{\zeta-a}\right)^2 + \cdots\right).$$

The series in the right-hand side is uniformly convergent on $C_r(a)$ (with respect to ζ) since we have

$$\left|\frac{z-a}{\zeta-a}\right| \leqslant \frac{|z-a|}{r} < 1$$

(i.e., the terms of the series are bounded above by the terms of a convergent geometric series, whose quotient is independent of ζ). From the uniform convergence of the series, we deduce the convergence of the integrals:

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

This is a power series, whose coefficients are given by the formula

$$c_n = \frac{1}{2\pi i}\int_{C_r(a)} \frac{f(\zeta)d\zeta}{(\zeta-a)^{n-1}}.$$

□

Thus all holomorphic functions can be expanded locally into power series. We now discuss the differentiation and the integration of power series (due to the result stated above, this discussion will have immediate applications to the differentiation and the integration of holomorphic functions).

t:int-antider **Theorem 9.3.** *Suppose that a power series $f(z) = \sum c_n(z-a)^n$ converges for $|z-a| < R$. Then the power series $F(z) = \sum \frac{c_n}{n+1}(z-a)^{n+1}$ also converges for $|z-a| < R$. Moreover, F is an antiderivative of $f(z)$.*

Proof. Note that, for any $r < R$, the series $f(z)$ converges uniformly in the disk $|z-a| < r$. Then the integral $F(z)$ of $f(\zeta)d\zeta$ over a smooth arc connecting a to z is equal to the sum of integrals of $c_n(\zeta-a)^n d\zeta$, for every z with $|z-a| < r$. We obtain the desired. □

t:pow-der **Theorem 9.4.** *Suppose that a power series $f(z) = \sum c_n(z-a)^n$ converges for $|z-a| < R$. Then the power series $g(z) = \sum_{n=1}^{\infty} nc_n(z-a)^{n-1}$ also converges for $|z-a| < R$. Moreover, g is the derivative of $f(z)$.*

Proof. The convergence of the series g follows from Hadamard's theorem. By Theorem 9.3, the function f is then an antiderivative of g, hence g is the derivative of f. □

It follows from Theorems 9.2 and 9.4 that any holomorphic function is infinitely many times differentiable.

t:der-formula **Theorem 9.5.** *Let $U \subset \mathbb{C}$ be an open subset and $f : U \to \mathbb{C}$ be a holomorphic function. Suppose that a disk $\{|z - a| \leqslant r\}$ is contained in U. Then*

$$f^{(n)}(a) = \frac{1}{2\pi i} \int_{C_r(a)} \frac{n! f(z) dz}{(z - a)^{n+1}}.$$

Proof. We know that $f(z)$ can be represented by a convergent power series $\sum c_n(z - a)^n$, where

$$c_n = \frac{1}{2\pi i} \int_{C_r(a)} \frac{f(\zeta) d\zeta}{(\zeta - a)^{n-1}}.$$

On the other hand, by differentiating the series for f and substituting $z = a$, we obtain that the series $\sum c_n(z - a)^n$ is the Taylor series for f, i.e., we have $c_n = \frac{f^{(n)}(a)}{n!}$. □

Theorem 9.6 (The Weierstrass Convergence Theorem)**.** *Suppose that $U \subset \mathbb{C}$ is a bounded open domain, and the functions $f_n : \overline{U} \to \mathbb{C}$ are defined and continuous in $\overline{U}$, and are holomorphic on U. If f_n converge uniformly on ∂U, then f_n converge uniformly on $\overline{U}$. Moreover, the limit f of f_n is holomorphic on U.*

Proof. We know that $|f_n - f_m| \to 0$ on ∂U as n, $m \to \infty$. It follows from the Maximum Modulus Principle that $|f_n - f_m| \to 0$ on $\overline{U}$ as n, $m \to \infty$, i.e., the sequence f_n is uniformly fundamental on $\overline{U}$. It follows that f_n converges uniformly on $\overline{U}$ to some function f. It remains to prove that f is holomorphic on U.

Let F_n be an antiderivative of f_n. The function F_n is defined up to an additive constant (a "constant of integration"). We may choose the constants of integration so that the sequence $F_n(a) = 0$ for some $a \in U$ and for all n. Then $F_n(z)$ is the integral of f_n over some arc connecting a with z in U. We know from Proposition 9.1 that the functions F_n converge uniformly in U to an antiderivative F of the function f. Since $F' = f$, the function F is holomorphic in U, therefore, it is infinitely many times differentiable. It follows that the complex derivative $f' = F''$ exists, hence f is holomorphic. □

A function $f : \mathbb{C} \to \mathbb{C}$ that is defined and holomorphic everywhere on $\mathbb{C}$ is called an *entire function*. Recall that a function $f : U \to \mathbb{C}$ is said to be *bounded* if there exists some real number $M > 0$ such that $|f(z)| \leqslant M$ for all $z \in U$.

Theorem 9.7 (The Liouville Theorem)**.** *Let f be a bounded entire function. Then f is constant.*

Proof. We will use the formula for the derivative of a holomorphic function, see Theorem 9.5:

$$f'(z) = \frac{1}{2\pi i}\int_{C_R(a)} \frac{2f(\zeta)d\zeta}{(z-\zeta)^2}.$$

We will now estimate the modulus of the right-hand side above:

$$\left|\int_{C_R(a)} \frac{f(\zeta)d\zeta}{(z-\zeta)^2}\right| \leqslant \frac{(2M)(2\pi R)}{(R-|z|)^2}.$$

Clearly, this estimate converges to zero as $R \to \infty$. It follows that $f'(z) = 0$ at all points $z \in \mathbb{C}$, hence f is constant. □

We will later discuss a local version of the Liouville theorem. The following theorem can be proved in the same way as the Liouville theorem:

t:poly **Theorem 9.8.** *Let f be an entire function such that $|f(z)| \leqslant C|z|^d$ for sufficiently large z. Then f is a polynomial, whose degree is at most d.*

9.1. **Problems.**

Problem 9.9. Prove that

$$\int_0^{2\pi} e^{\cos t}\cos(\sin t - nt) = \frac{2\pi}{n!}$$

Solution. Set $f(z) = e^z$. We will use the formula from Theorem 9.5:

$$1 = e^0 = f^{(n)}(0) = \frac{n!}{2\pi i}\int_{C_1(0)} \frac{e^\zeta d\zeta}{\zeta^{n+1}}.$$

Parameterize the circle $C_1(0)$ by $z = e^{it}$. Then the integral in the right-hand side becomes

$$\frac{n!}{2\pi}\int_0^{2\pi} e^{e^{it}-int}dt.$$

Taking the real part of this integral, we obtain that

$$1 = \frac{n!}{2\pi}\int_0^{2\pi} e^{\cos t}\cos(\sin t - nt)\,dt.$$

We used that

$$\mathrm{Re}(e^z) = e^{\mathrm{Re}z}\cos(\mathrm{Im}z)$$

for any complex number $z \in \mathbb{C}$.

Problem 9.10. Compute

$$\int_{C_2(0)} z^n(1-z)^m dz$$

for all integer values of n and m.

Problem 9.11. Find $\sup|f^{\circ n}(0)|$ over all entire functions such that $|f(z)| \leqslant 1$ for all z with $|z| = 1$.

Problem 9.12. Suppose that a function f is defined and holomorphic in some neighborhood of 0 Prove that the higher derivatives of f at 0 cannot satisfy the inequalities

$$|f^{(n)}(0)| > n!n^n.$$

10. Local behavior of holomorphic functions

We start with the following lemma, which is yet another version of the Cauchy integral formula. We will write $D_R(a)$ for the disk $\{|z-a| < R\}$ and $\overline{D}_R(a)$ for its closure $\{|z-a| \leqslant R\}$.

l:stokes-mult **Lemma 10.1.** *Suppose that the closed disks $\overline{D}_{r_1}(a_1)$, …, $\overline{D}_{r_n}(a_n)$ are disjoint subsets of an open disk $D_R(a)$. Define σ as the 1-cycle $C_R(a) - C_{r_1}(a_1) - \cdots - C_{r_n}(a_n)$. Finally, assume that a function f is defined and holomorphic on some open set containing the set $Z = \overline{D}_R(a) \setminus (D_{r_1}(a_1) \cup \cdots \cup D_{r_n}(a_n))$. Then the integral of $f(z)dz$ over σ is equal to zero.*

Sketch of a proof. Intuitively, the 1-cycle σ is the boundary of Z, hence the integral of the closed form $f(z)dz$ over σ must be zero. However, we only defined boundaries of coordinate rectangles and their linear combinations, thus we need to approximate σ with a coordinate cycle, and prove that this approximating cycle is a boundary. □

l:cauchy-mult **Lemma 10.2.** *Under the assumptions of Lemma 10.1, for every point $z \in Z$, we have*

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

Proof. Let $\varepsilon > 0$ be a small positive number such that the disk $\overline{D}_\varepsilon(z)$ is contained in the interior of Z. The boundary σ' of $Z \setminus \overline{D}_\varepsilon(z)$ (properly oriented) is a 1-cycle consisting of circles. Applying Lemma 10.1 to σ' and $Z \setminus \overline{D}_\varepsilon(z)$ instead of σ and Z, we conclude that

$$\int_\sigma \frac{f(\zeta)d\zeta}{\zeta - z} = \int_{C_\varepsilon(z)} \frac{f(\zeta)d\zeta}{\zeta - z}.$$

The result now follows from the usual Cauchy integral formula. □

A *Laurent series* centered at a point a is a series of the form

$$L(z) = \sum_{n=-\infty}^{\infty} c_n(z-a)^n.$$

It is similar to a power series with the only difference that a Laurent series may contain negative powers of $z - a$. The series $L(z)$ can be naturally represented as the sum of two series $L_+(z)$ and $L_-(z)$. By definition, the series $L_+(z)$ consists of all terms in $L(z)$ with nonnegative powers of $z-a$. Thus $L_+(z)$ is a usual power series centered at a. Let R_+ be the radius of convergence of this series. The series $L_-(z)$ consists of all terms in $L(z)$ with negative powers of $z-a$. It can be regarded as a power series in $\frac{1}{z-a}$. Let $1/R_-$ be its radius of convergence. Then the series $L_-(z)$ converges in the region $\{|z| > R_-\}$. We see that if $R_- < R_+$, then the series $L(z)$ converges in the annulus $\{R_- < |z| < R_+\}$. This annulus is called the *annulus of convergence of* $L(z)$.

t:laurent **Theorem 10.3** (Laurent series expansion). *Suppose that a function f is defined and holomorphic in some annulus $A = \{r < |z| < R\}$ (it is possible that $r = 0$ and/or $R = \infty$). Then f can be represented in A as the sum of some convergent Laurent series.*

Proof. Take a point $z \in A$. Choose a sufficiently small $\varepsilon > 0$ so that z belongs to the annulus $\{r + \varepsilon < |z| < R - \varepsilon\}$ (if $R = \infty$, then we just take a sufficiently large number instead of $R - \varepsilon$). By Lemma 10.2, we have

$$f(z) = \int_{C_{R-\varepsilon}(a)} \frac{f(\zeta)d\zeta}{\zeta - z} - \int_{C_{r+\varepsilon}(a)} \frac{f(\zeta)d\zeta}{\zeta - z}.$$

Consider the first integral in the right-hand side. The integrand can be expanded into a convergent (uniformly with respect to ζ) geometric series

$$\frac{f(\zeta)}{\zeta - z} = \frac{f(\zeta)}{\zeta - a}\left(1 + \frac{z-a}{\zeta - a} + \left(\frac{z-a}{\zeta-a}\right)^2 + \cdots\right).$$

The convergence follows from the estimate

$$\left|\frac{z-a}{\zeta-a}\right| < \frac{|z-a|}{R-\varepsilon} < 1,$$

which holds for all $\zeta \in C_{R-\varepsilon}(a)$. Integrating this geometric series termwise over $C_{R-\varepsilon}(a)$, we obtain a power series $L_+(z) = \sum_{n\geqslant 0} c_n(z-a)^n$, whose coefficients c_n $(n \geqslant 0)$ are given by the formula

$$c_n = \frac{1}{2\pi i}\int_{C_{R-\varepsilon}(a)} \frac{f(\zeta)}{(\zeta-a)^{n+1}}.$$

Consider now the second integral, the integral over the circle $C_{r+\varepsilon}(a)$. The integrand can also be expanded into a convergent geometric series but differently:

$$\frac{f(\zeta)}{\zeta - z} = \frac{-f(\zeta)}{z-a}\left(1 + \frac{\zeta - a}{z-a} + \left(\frac{\zeta-a}{z-a}\right)^2 + \cdots\right).$$

The convergence (uniform in ζ) follows from the estimate

$$\left|\frac{\zeta - a}{z - a}\right| < \frac{r + \varepsilon}{|z - a|} < 1.$$

Integrating this geometric series term-wise over $C_{r+\varepsilon}(a)$, we obtain a series $L_-(z) = \sum_{n<0} \tilde{c}_n (z-a)^n$. The coefficients c_n $(n < 0)$ are given by the formula

$$c_n = \frac{1}{2\pi i} \int_{C_{r+\varepsilon}(a)} \frac{f(\zeta)d\zeta}{(\zeta - a)^{n+1}}$$

□

Consider an open subset $U \subset \mathbb{C}$, a point $a \in U$, and a holomorphic function $f : U \setminus \{a\} \to \mathbb{C}$. The point a is then called an *isolated singularity* of f, or an *isolated singular point*. Being singular means nothing more than just the fact that f is not defined on a. An isolated singularity a is said to be *removable* if there is a holomorphic function $\widetilde{f} : U \to \mathbb{C}$ such that $\widetilde{f} = f$ on $U \setminus \{a\}$.

Theorem 10.4 (The Removable Singularity Theorem)**.** *Suppose that f is bounded on $U \setminus \{a\}$. Then a is a removable singularity for f.*

Proof. Suppose that $|f(z)| \leqslant M$ for some $M > 0$ and all $z \in U \setminus \{a\}$. Let $R > 0$ be some positive real number such that $\overline{D}_R(a) \subset U$. By Theorem 10.3, the function f has a Laurent series expansion in the annulus $\{0 < |z| < R\}$. By the proof of Theorem 10.3, for $n < 0$, the corresponding coefficients of the Laurent series for f are

$$c_n = \frac{1}{2\pi i} \int_{C_\varepsilon(a)} \frac{f(\zeta)d\zeta}{(\zeta - a)^{n+1}}.$$

Note that, since n is negative, we have

$$\frac{1}{(\zeta - a)^{n+1}} = (\zeta - a)^{|n|-1}.$$

Therefore, we have the following upper estimate:

$$|c_n| \leqslant \varepsilon^{|n|} M.$$

Since ε can be chosen arbitrarily small, we have $c_n = 0$ for all negative n. Then a holomorphic extension of $f(z)$ to a neighborhood of a is the sum of the convergent power series $L_+(z)$ centered at a. The theorem follows. □

Another type of a singular point is a *pole*. We say that a holomorphic function $f : U \setminus \{a\} \to \mathbb{C}$ has a pole at a if the function $\widetilde{f} : U \to \overline{\mathbb{C}}$ defined by setting $\widetilde{f}(z) = f(z)$ for $z \neq a$ and $\widetilde{f}(a) = \infty$ is continuous at

a and has a complex derivative at a. In fact, it suffices to assume that $\widetilde{f}$ is continuous at a, i.e., that we have $\lim_{z\to a} f(z) = \infty$. Indeed, then the function $1/f$ has a removable singularity at a, and we can apply the Removable Singularity Theorem. If a singular point a of a holomorphic function f is neither a removable singularity nor a pole, then a is called an *essential singularity* of f. If a is an essential singularity for f, then f does not have (finite or infinite) a limit at a.

t:pole **Theorem 10.5.** *Let $U \subset \mathbb{C}$ be an open subset. A point $a \in U$ is not an essential singularity of a holomorphic function $f : U \setminus \{a\} \to \mathbb{C}$ if and only if the Laurent expansion of f in an annulus $\{0 < |z-a| < r\}$ contains finitely many terms with negative powers of $z - a$.*

The minimal integer m such that $(z-a)^m$ enters with a nonzero coefficient into the Laurent expansion of f is called the *order* of f at a (notation: $\mathrm{ord}_a(f)$).

Proof. Suppose that a is not an essential singularity of f. If a is a removable singularity, then there are no negative powers of $z-a$ at all in the Laurent expansion of f. Suppose now that a is a pole. Consider the function $g(z) = 1/f(z)$, which we also extend to a by setting $g(a) = 0$. The function g can be represented as a convergent power series in some neighborhood of a:

$$g(z) = a_k(z-a)^k + a_{k+1}(z-a)^{k+1} + \cdots = (z-a)^k h(z).$$

Suppose that a_k, $k > 0$, is the first nonzero coefficient. The function $h(z) = a_k + a_{k+1}(z-a) + \cdots$ is holomorphic in a neighborhood of a, and $h(a) \neq 0$. The holomorphic function $1/h$ can be expanded into a convergent power series in some open disk around a. Then the Laurent series expansion of f can be obtained from the power series expansion of $1/h$ by dividing it by $(z-a)^k$. Obviously, this Laurent series has finitely many terms with negative powers of $z-a$.

Let us now assume the contrary: the Laurent series for f has finitely many terms with negative powers of $z-a$. We will prove that f has a pole or a removable singularity at a. If there are no terms with negative powers of $z-a$, then a is of course a removable singularity. If k is the maximal positive integer with the property that $(z-a)^{-k}$ enters with a nonzero coefficient into the Laurent series for f, then $(z-a)^k f$ extends to a holomorphic function on a neighborhood of a, hence f has a pole at a. □

Recall that a map $f : U \to \overline{\mathbb{C}}$ is called a *holomorphic map* if it has complex derivatives at all points of U. Such a holomorphic map is also called a *meromorphic function*. According to a classical terminology, a holomorphic function cannot take infinite values. Any holomorphic function with finitely many non-essential isolated singular points defines an extension which is a meromorphic function. The proof of Theorem 10.5 also implies the following corollary.

c:hol-loc **Corollary 10.6.** *Suppose that $U \subset \mathbb{C}$ is an open subset and $f : U \to \overline{\mathbb{C}}$ is a holomorphic map (=meromorphic function). Then, for every $a \in U$, we have $f(z) = (z-a)^{\mathrm{ord}_a(f)} h_a(z)$, where $h_a : U \to \overline{\mathbb{C}}$ is a holomorphic function such that $h_a(a) \neq 0$.*

We will now discuss singularities of functions at infinity. Suppose that a function f is defined and holomorphic on $\{|z| > R\} \subset \mathbb{C}$ for some positive radius R. Then we can define a new function $g(z) = f(1/z)$ that is defined and holomorphic on the annulus $\{0 < |z| < 1/R\}$. We say that f has an essential singularity (a removable singularity, a pole) at ∞ if g has an essential singularity (a removable singularity, a pole) at 0. Moreover, we set $\mathrm{ord}_\infty(f) = \mathrm{ord}_0(g)$.

As an example, we can now describe all holomorphic maps of the Riemann sphere $\overline{\mathbb{C}}$ to itself. Let $f : \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ be such a holomorphic map, and let P be the subset of $\overline{\mathbb{C}}$ consisting of all points a with the property $f(a) = \infty$. It follows from Corollary 10.6 that all points in P are isolated (for every point of P, there is a neighborhood of this point in $\overline{\mathbb{C}}$ that does not contain any other points of P). Recall that the sphere $\overline{\mathbb{C}}$ (equipped with the spherical metric) is compact. We will use the following topological statement: a subset of a compact space, all of whose points are isolated, is finite. Thus P is a finite set. Consider the product

$$\widetilde{F}(z) = f(z) \prod_{a \in P \cap \mathbb{C}} (z-a)^{-\mathrm{ord}_a(f)}$$

The function $\widetilde{F}$ has isolated singularities at all points of $P \cap \mathbb{C}$ (it is defined and holomorphic elsewhere on $\mathbb{C}$). Thus the function $\widetilde{F}$ extends to an entire function $F : \mathbb{C} \to \mathbb{C}$.

Clearly, the function F has a finite order d at infinity. Then $|F(z)| \leqslant M|z|^d$ for some $M > 0$ and all sufficiently large z. It follows that F is a polynomial of degree d. Then f is a ratio of two polynomials, i.e., a rational function.

Let $U \subset \mathbb{C}$ be an open subset. A *holomorphic 1-form* on U is by definition a form $f(z)dz$, where $f : U \to \mathbb{C}$ is a holomorphic function. Suppose now that $a \in U$ and α is a holomorphic 1-form on the set

$U \setminus \{a\}$. The *residue* of α at a is defined as the complex number

$$\mathrm{res}_a(\alpha) = \frac{1}{2\pi i}\int_{C_\varepsilon(a)} \alpha,$$

where $\varepsilon > 0$ is a sufficiently small number (so that the set $\{|z-a| \leqslant \varepsilon\}$ is a subset of U.) The integral in the right-hand side does not depend on a particular choice of ε by Lemma 10.1.

The following theorem is often used for computing integrals:

Theorem 10.7 (The Residue Theorem)**.** *Suppose that $U \subset \mathbb{C}$ is a bounded domain, whose boundary can be represented by a smooth cycle σ. Let $a_1, \dots, a_k$ be finitely many points in U, and α be a holomorphic 1-form defined on $V \setminus \{a_1, \dots, a_k\}$, where V is some open set containing $\overline{U}$. Then*

$$\int_\sigma \alpha = 2\pi i \left(\sum_{j=1}^{k} \mathrm{res}_{a_j}(\alpha) \right).$$

If U is a round disk, then the Residue Theorem follows from Lemma 10.1. If U is not a round disk, the proof uses the same ideas as the proof of Lemma 10.1. Note that many classical authors write about residues of meromorphic functions $f(z)$ rather than residues of holomorphic 1-forms $f(z)dz$. This is justified by the fact that the function f can be uniquely recovered by α as α/dz. On the other hand, if we talk about residues of functions, then we should always keep in mind that this notion depends not only on the function but also on a specific choice of a coordinate z. If we change a coordinate in a function, then the residues will change. However, if we perform a coordinate change in the corresponding 1-form, then the residue will not change. In particular, if $f(z)$ is a holomorphic function defined for all sufficiently large values of z, then $\mathrm{res}_\infty f(z)dz$ is not equal to $\mathrm{res}_0 f(1/z)dz$ but it is equal to $\mathrm{res}_0 f(1/z)d(1/z)$.

10.1. **Problems.**

Problem 10.8. Compute the order of the function $f(z) = \sin(z) - z$ at point $z = 0$.

Problem 10.9. Which of the following functions have isolated singularities at point $z = 0$:

$$\frac{\sin z}{z}, \quad e^{1/z}, \quad \cos\left(\frac{1}{z}\right), \quad \frac{1}{e^{\frac{1}{z^2}} + 1}?$$

Determine which of these isolated singularities are essential. For non-essential singularities, find the corresponding orders.

Problem 10.10. Let a function f be defined and holomorphic on some neighborhood of the disk $\{|z| \leqslant R\}$. Suppose that $|f(z)| \neq 0$ for all z with $|z| = R$. Prove that

$$\int_{C_R(0)} \frac{f'(z)}{f(z)} dz = \sum_{|a|<R} \mathrm{ord}_a(f).$$

In the right-hand side, the summation is performed over all points a in the disk $\{|z| \leqslant R\}$ such that $f(a) = 0$ (at all other points, we have $\mathrm{ord}_a(f) = 0$).

Problem 10.11. Prove that, for sufficiently large n, the equation

$$1 + z + \frac{z^2}{2!} + \cdots + \frac{z^n}{n!} = 0$$

has no solutions in the disk of radius 100 around the origin.

Problem 10.12. Find a Laurent series that converges to the function

$$f(z) = \frac{1}{z(z-1)}$$

in the annulus $\{0 < |z| < 1\}$.

Problem 10.13. Let b_n be any sequence of different complex numbers such that $b_n \to \infty$ as $n \to \infty$. Prove that there exists an entire function that vanishes at points b_n and at no other points.

Problem 10.14. Suppose that a function f is defined and holomorphic in the annulus $\{0 < |z - a| < \varepsilon\}$, and that a is an essential singularity of f. Prove that the values of f are dense in $\mathbb{C}$.

11. Geometric properties of conformal mappings

In this section, we discuss an important principle. Roughly speaking, it says that complex solutions of analytic equations depend continuously on parameters and, in particular, cannot disappear. This is the *preservation of number principle.* We will discuss several implementations of this principle.

We start with the following theorem, which is a simple consequence of the Residue Theorem:

t:numb **Theorem 11.1.** *Let a function f be defined and holomorphic on some neighborhood of the disk $\{|z| \leqslant R\}$. Suppose that $|f(z)| \neq 0$ for all z with $|z| = R$. Then*

$$\int_{C_R(0)} \frac{f'(z)}{f(z)} dz = 2\pi i \left(\sum_{|a|<R} \mathrm{ord}_a(f) \right).$$

In the right-hand side, the summation is performed over all points a in the disk $\{|z| \leqslant R\}$ such that $f(a) = 0$ (at all other points, we have $\mathrm{ord}_a(f) = 0$*).*

The right-hand side of the formula given in Theorem 11.1 is called the *number of zeros of f in the disk $\{|z| < R\}$ (counted with multiplicities)*. We will always count multiplicities even if we skip mentioning this. The number of zeros of f in a domain U (counting multiplicities) is denoted by $\#\{f(z) = 0 \mid z \in U\}$

Note that the differential 1-form $\frac{f'(z)dz}{f(z)}$ is equal to $d \log f(z)$, where $\log f(z)$ is not a globally defined function but rather a local branch, i.e., a continuous function $g(z)$ that is defined on some neighborhood of $z_0 \in C_R(0)$ and satisfies the equation $e^{g(z)} = f(z)$ in this neighborhood. Note that, for $w \neq 0$, we have

$$\log w = \log |w| + i \, \mathrm{arg} w$$

for any branch of the logarithm, where arg$w$ is the *argument* of w, an angle of a rotation that takes the positive real direction to the direction of the vector w. Note that the real part of $\log w$ is well-defined on the punctured plane $\{0 < |w|\}$. However, the imaginary part, arg$w$, is not. Namely, as w makes a counterclockwise loop around the origin, the number 2π is added to arg$w$.

The formula in Theorem 11.1 can be now rewritten as follows:

$$\#\{f(z) = 0 \mid |z| < R\} = \frac{1}{2\pi} \int_{C_R(0)} d \, \mathrm{arg} f(z).$$

We used the fact that the integral of the exact form $d \log |f(z)|$ over the cycle $C_R(0)$ is equal to zero. Note that $d\mathrm{arg} f(z)$ is well defined as a differential 1-form on the complement of the set of zeros of f (indeed, different choices of local branches of the argument give rise to the same differential). Geometrically, the integral on the right-hand side can be interpreted as follows: *it is the number of full turns the vector $f(z)$ makes around the origin*. This number must be counted with sign: counterclockwise turns are counted with sign plus, and clockwise turns are counted with sign minus.

Thus we arrive at the following principle, sometimes called the **argument principle**: *Suppose that a function f is defined and holomorphic on a neighborhood of some closed disk D. Suppose also that $f \neq 0$ on the boundary of D. Then the number $\#\{f(z) = 0 \mid z \in D\}$ is equal to the number of full turns the vector $f(z)$ makes around the origin as z makes one full turn around the boundary of D.* In this statement, a disk can be replaced with some other bounded domain with sufficiently smooth boundary.

The statement of the argument principle given above is rather informal. It can be formalized using the notion of a mapping degree.

The following theorem is a straightforward consequence of the argument principle:

Theorem 11.2 (Rouché). *Let D be a closed disk in $\mathbb{C}$, and f, g be functions defined and holomorphic on some neighborhood of D. Suppose that $|g| < |f|$ everywhere on the boundary of D. Then we have*

$$\{f(z) + g(z) = 0 \,|\, z \in D\} = \{f(z) = 0 \,|\, z \in D\}.$$

Intuitively, this theorem follows from the observation that the vector $f(z)+g(z)$ makes the same number of turns around the origin as $f(z)$. The following is the main corollary of the Rouche theorem:

Corollary 11.3 (Preservation of Number). *Suppose that a function f is holomorphic in a neighborhood of a closed disk D and that $f \neq 0$ on the boundary of D. Then any holomorphic function defined on a neighborhood of D and sufficiently close to f has the same number of zeros in D, counting multiplicities.*

Indeed, any holomorphic function sufficiently close to f can be represented in the form $f + g$, where $|g| < |f|$ on the boundary of D.

c:k-to-1 **Corollary 11.4.** *Suppose that a function f is defined and holomorphic on a neighborhood of a point $a \in \mathbb{C}$, and let $k = \mathrm{ord}_a(f)$. Then, for every sufficiently small complex number c, the equation $f(z) = c$ has exactly k solutions, counting multiplicities.*

Proof. We know that the equation $f(z) = 0$ has k solutions near a counting multiplicities (all these solutions coincide with a). Then the result follows from the preservation of number. □

c:domain **Corollary 11.5** (Preservation of domains). *Suppose that $U \subset \mathbb{C}$ is open, and $f : U \to \mathbb{C}$ is a nonconstant holomorphic function. Then the subset $f(U) \subset \mathbb{C}$ is also open.*

Proof. Suppose that $w \in f(U)$. Then, by the preservation of number, for all w' in some neighborhood V of w, the equation $f(z) = w'$ has at least one solution in U. It follows that $V \subset f(U)$. □

Let U and V be two open subsets of the Riemann sphere. A *conformal isomorphism* $\varphi : U \to V$ is by definition a conformal bijective mapping.

p:conf-iso **Proposition 11.6.** *If $\varphi : U \to V$ is a conformal isomorphism, then $\varphi^{-1} : V \to U$ is also a conformal isomorphism.*

Proof. We need only prove that φ^{-1} is holomorphic. Since the differential of the inverse map φ^{-1} at a point $w = \varphi(z)$ is the inverse of the differential of φ at z provided that the latter is invertible, it suffices to prove that $\varphi'(z) \neq 0$ for all $z \in U$. Since this statement is local, we may assume that U and V are subsets of $\mathbb{C}$. Assume the contrary: $\varphi'(a) = 0$ for some $a \in U$. Then the order k of $\varphi(z) - \varphi(a)$ at $z = a$ is bigger than one. By Corollary 11.4, the map $\varphi(z)$ takes all values sufficiently close to $\varphi(a)$ at k points near a, counting multiplicities. A contradiction with our assumption that φ is a bijection. □

A conformal isomorphism of U with itself is called a *conformal automorphism* of U. It follows from Proposition 11.6 that the set $\mathrm{Aut}(U)$ of all conformal automorphisms of U is a group under composition.

The following is a very important property of holomorphic functions.

Theorem 11.7 (The Uniqueness Theorem)**.** *Suppose that $U \subset \mathbb{C}$ is an open connected subset, and $f : U \to \mathbb{C}$ is a holomorphic function. If there is a sequence of different points $z_n \in U$ that converges to some point in U such that $f(z_n) = 0$ for all n, then f is identically zero.*

Proof. Let z be the limit of z_n. Since all z_ns are different, the function f cannot have finite order at z. It follows that f vanishes in some neighborhood of z. Let $Z \subset U$ be the set of non-isolated zeros of f. Clearly, this set is closed. By the argument explained above, it is also open. Since U is connected, we must have $Z = U$. □

Theorem 11.8 (The Schwarz lemma)**.** *Let $f : \mathbb{D} \to \mathbb{D}$ be a holomorphic map such that $f(0) = 0$. Then $|f'(0)| \leqslant 1$. Moreover, if $|f'(0)| = 1$, then $f(z) = e^{i\theta} z$ for some $\theta \in \mathbb{R}$.*

Proof. Consider the holomorphic map $g(z) = f(z)/z$ defined on $\mathbb{D}\backslash\{0\}$. This map has a removable singularity at 0 by the Removable Singularity Theorem. Therefore, g extends to a holomorphic map of $\mathbb{D}$ to $\mathbb{C}$, for which we will keep the same notation. Since $|g(z)| \leqslant 1 + \varepsilon$ for z near the boundary of $\mathbb{D}$, we conclude by the Maximum Modulus Principle that $|g(z)| \leqslant 1 + \varepsilon$ on $\mathbb{D}$. Since ε is arbitrary, it follows that $|g(z)| \leqslant 1$ on $\mathbb{D}$. Then $|f(z)| \leqslant |z|$. It follows that

$$|f'(0)| = \lim_{z \to 0} \frac{|f(z)|}{|z|} \leqslant 1.$$

Suppose that $f'(0) = 1$, then $|g(0)| = 1$. Since this is the maximal value of $|g|$, we must conclude by the Maximum Modulus Principle that $|g| = 1$ everywhere on $\mathbb{D}$. We claim that g is constant. Indeed, otherwise by Corollary 11.5 the function $|g|$ would take an open set of values. The claim now follows. □

We can now describe all conformal automorphisms of $\mathbb{D}$. Let $f : \mathbb{D} \to \mathbb{D}$ be any conformal automorphism. Suppose first that $f(0) = 0$. By the Schwarz lemma, we have $|f'(0)| \leqslant 1$. If g is the inverse map, then we also have $|g'(0)| \leqslant 1$. However, $g'(0) = 1/f'(0)$. Therefore, we must have $|f'(0)| = |g'(0)| = 1$. Using the Schwarz lemma again, we conclude that $f(z) = e^{i\theta} z$. Suppose now that a conformal automorphism $f : \mathbb{D} \to \mathbb{D}$ is arbitrary. Set $a = f(0)$, and consider the fractional linear map

$$h(z) = \frac{z - a}{1 - \overline{a}z}.$$

We know that h is automorphism of $\mathbb{D}$ and that $h(a) = 0$. Therefore, we have $h \circ f(0) = 0$. It follows by the preceding argument that $h \circ f(z) = e^{i\theta} z$. Therefore,

$$f^{-1}(z) = e^{-i\theta} \frac{z - a}{1 - \overline{a}z}.$$

Since f was an arbitrary automorphism of $\mathbb{D}$, the map f^{-1} is also an arbitrary automorphism of $\mathbb{D}$. In other words, an arbitrary automorphism of $\mathbb{D}$ has the form

$$f(z) = \frac{az + b}{-\overline{b}z + \overline{a}}$$

for some complex numbers a and b.

11.1. **Problems.**

Problem 11.9. Find all conformal isomorphisms of the upper half-plane $H = \{\mathrm{Im}(z) > 0\}$.

Answer:

$$f(z) = \frac{az + b}{cz + d}, \quad a, b, c, d \in \mathbb{R}, \quad ad - bc > 0.$$

12. Normal families of holomorphic functions

Let X and Y be metric spaces. A *family of mappings* $\mathcal{F} : X \to Y$ is the same as a set of mappings from X to Y. A family $\mathcal{F}$ of mappings is *equicontinuous* if, in the ε-δ definition of continuity, we can choose δ depending only on ε, not on a particular function from f. To be more precise: $\mathcal{F}$ is equicontinuous at a point $x \in X$ if, for every $\varepsilon > 0$, there exists $\delta > 0$ with the property that $d_X(x, x') < \delta$ implies $d_Y(f(x), f(x')) < \varepsilon$ for all $f \in \mathcal{F}$ and all $x' \in X$. The family $\mathcal{F}$ is equicontinuous if it is equicontinuous at all points of X. The following theorem explains the meaning of equicontinuity.

t:Ascoli **Theorem 12.1** (Ascoli)**.** *Suppose that X has a dense countable set, and Y is compact. If a family of mappings $\mathcal{F} : X \to Y$ is equicontinuous, then every sequence in $\mathcal{F}$ has a pointwise convergent subsequence.*

Proof. Consider a sequence $f_n \in \mathcal{F}$. Let A be a dense countable subset of X. Passing to subsequences and using the diagonal argument, we may assume that, for every $a \in A$, the sequence $f_n(a)$ converges in Y. Take any point $x \in X$ and choose ε. Let δ be such that $d_X(x, x') < \delta$ implies $d_Y(f_n(x), f_n(x')) < \varepsilon/3$ for all n and x'. Choose $a \in A$ such that $d_X(x, a) < \delta$ and choose N such that $d_Y(f_n(a), f_m(a)) < \varepsilon/3$ provided that n, $m > N$. Then we have, for n, $m > N$:

$$d_Y(f_n(x), f_m(x)) \leqslant$$
$$d_Y(f_n(x), f_n(a)) + d_Y(f_n(a), f_m(a)) + d_Y(f_m(a), f_m(x)) < \varepsilon.$$

It follows that the sequence $f_n(x)$ converges to some point $f(x) \in Y$. Thus the sequence f_n converges to a function f pointwise. □

Another general topological fact that we need is the following:

t:unif **Theorem 12.2.** *Suppose that X is compact, and a pointwise convergent sequence of continuous mappings $f_n : X \to Y$ is equicontinuous. Then the convergence is uniform.*

Note that the equicontinuity assumption is important. There is a pointwise convergent sequence of continuous functions on the interval $[0, 1]$ that does not converge uniformly.

Proof. Let $f : X \to Y$ denote the pointwise limit of f_n. Take any $\varepsilon > 0$. For every point $x \in X$, there is some positive integer N_x with the property that $n > N_x$ implies $d_Y(f_n(x), f(x)) < \varepsilon/3$. By equicontinuity, there is some $\delta_x > 0$ with the property that $d_X(x, x') < \delta_x$ implies $d_Y(f_n(x), f_n(x')) < \varepsilon/3$ for all n. Passing to the limit as $n \to \infty$, we also conclude that $d_Y(f(x), f(x')) \leqslant \varepsilon/3$. It follows that, for all $n > N_x$ and all x' with $d_X(x, x') < \delta_x$, we have

$$d_Y(f_n(x'), f(x')) \leqslant$$
$$d_Y(f_n(x'), f_n(x)) + d_Y(f_n(x), f(x)) + d_Y(f(x), f(x')) < \varepsilon.$$

The open covering of X by balls of radius δ_x centered at x has a finite subcovering. Let x_1, …, x_k be the corresponding centers. Choose N to be the maximum of N_{x_1}, …, N_{x_k}. Then f_n are uniformly ε-close to f for all $n > N$. □

Theorem 12.3. *Let $U \subset \mathbb{C}$ be a bounded open subset, and $\mathcal{F} : U \to \{|z| < 1\}$ be a family of holomorphic mappings. Then $\mathcal{F}$ is equicontinuous.*

Proof. If $f \in \mathcal{F}$, $a \in U$, and r is a positive number such that the disk $\{|z-a|<r\}$ is contained in U, then we have

$$|f'(a)| \leqslant \frac{4\pi}{r},$$

see the proof of the Liouville theorem. We now fix two points a and b in U. Let σ be an smooth arc connecting a with b in U. Set r to be the smallest distance from σ to the boundary of U. Then $|f'(z)| < 4\pi/r$ for all $f \in \mathcal{F}$ and for all $z \in \sigma$. If L is the length of σ, then, for every $f \in \mathcal{F}$, we have

$$|f(b) - f(a)| \leqslant 4\pi L/r.$$

Since the upper bound does not depend on f and we may assume L to be sufficiently small provided that b is sufficiently close to a, the family $\mathcal{F}$ is equicontinuous. □

Recall that $\mathbb{D}$ is the unit disk $\{|z|<1\}$.

Theorem 12.4 (The Riemann Mapping Theorem)**.** *Let $U \subset \overline{\mathbb{C}}$ be a connected simply connected open set, whose complement contains more than one point. Then there is a holomorphic bijective map from U to $\mathbb{D}$.*

We will need the following lemma.

l:hol-branch **Lemma 12.5.** *Let $U \subset \overline{\mathbb{C}}$ be a simply connected domain, and $f : U \to \mathbb{C}$ a holomorphic map such that $0 \notin f(U)$. Then there is a holomorphic function $g : U \to \mathbb{C}$ with the property $g(z)^2 = f(z)$ for all $z \in U$.*

The function g will be referred to as a *holomorphic branch* of a multivalued function $\sqrt{f}$ over U.

Sketch of a proof of Lemma 12.5. Consider a subset Γ of $U \times \mathbb{C}$ consisting of all pairs (z, w) such that $w^2 = f(z)$ (this set Γ is called the *graph* of the multivalued function $\sqrt{f}$). Let Γ_1 be any connected component of Γ. Let us first prove that the image $\pi(\Gamma_1)$ of Γ_1 under the projection $\pi : (z, w) \mapsto z$ coincides with U. Indeed, it is not hard to see that the set $\pi(\Gamma_1)$ is open and closed in U. Now the statement follows from the connectivity of U.

We now prove that Γ_1 is a graph of some function g. Assuming the contrary, there is some point $a \in U$ with the property that there are two different points (a, b) and (a, c) in Γ_1. A smooth path on Γ_1 connecting (a, b) with (a, c) projects to a smooth loop σ in U based at a. Then $f(\sigma)$ must make an odd number of turns around 0, in other words, the integer

$$N = \frac{1}{2\pi i} \int_\sigma \frac{df}{f}$$

is odd.

Perturbing σ slightly, we may assume that σ is defined by coordinate chain (abusing notation, we use the same letter σ to denote this chain). Then σ is a boundary, since U is simply connected. It follows that the integral over σ of a closed 1-form $\frac{df}{f}$ is zero. Hence $N = 0$, a contradiction.

It remains only to prove that the function g is holomorphic. But this is clear from the fact that a holomorphic branch of the square root exists on every open disk not containing 0. $\square$

Proof of the Riemann mapping theorem. Fix some point $a \in U \cap \mathbb{C}$. Let $\mathcal{F}$ be the set of all holomorphic injective maps from U to $\mathbb{D}$ that take a to 0.

Let us first prove that the set $\mathcal{F}$ is non-empty. We may assume that $0, \infty \notin U$, since the complement of U contains more than one point. By Lemma 12.5, there is a holomorphic branch f of the function $z \mapsto \sqrt{z}$ on U. We claim that the complement of $f(U)$ contains a disk of nonzero radius. If not, then the open sets $f(U)$ and $-f(U)$ must intersect. Let z be an intersection point, i.e., $z = f(w)$ and $-z = f(w')$ for some w, $w' \in U$. However, it follows that both w and w' are equal to z^2. A contradiction.

Intuitively, we want to maximize the value $|f'(a)|$ over all $f \in \mathcal{F}$. Set $M = \sup_{f\in\mathcal{F}} |f'(a)|$.

We first prove that there is a mapping $F \in \mathcal{F}$ with $|F'(a)| = M$. Indeed, there is a sequence $f_n \in \mathcal{F}$ such that $|f_n'(a)| \to M$. Since $\mathcal{F}$ is equicontinuous on U, it contains a convergent subsequence. Passing to this subsequence, we may assume that the sequence f_n converges on U to some function F. By Theorem 12.2, the convergence is uniform on any compact subset of U. By the Weierstrass convergence theorem, the map F is also holomorphic, and the derivative of F is the limit of the derivatives of f_n. In particular, $|F'(a)| = M$.

We now want to prove that F is a bijective holomorphic map between U and $\mathbb{D}$. To prove injectivity, we assume by the way of contradiction that $F(z) \neq F(z')$ for some pair of points z, $z' \in U$. By the preservation of number, f_n takes the value $F(z)$ at a point near z and takes the value $F(z')$ at a point near z', for sufficiently large n. It follows that f_n is not injective, a contradiction.

We now prove that F is surjective. Suppose not. Then there is a point $b \in \mathbb{D}$ that is not in $F(U)$. Let A be an automorphism of $\mathbb{D}$ that maps b to 0, let c be any complex number such that $c^2 = A(0)$, and, finally, let B be an automorphism of $\mathbb{D}$ that maps c to 0. By Lemma 12.5, there is a holomorphic branch g of the function $\sqrt{A \circ f}$ on U such that $g(0) = c$ (we may need to replace g with $-g$ to arrange for the last property). Consider the map $h = B \circ g$. It is not hard to see that $h \in \mathcal{F}$, in particular, $h(a) = 0$.

We claim that $|h'(a)| > M$ (this leads to a contradiction with our assumption that F is not surjective and concludes the proof of the Riemann Mapping Theorem). Indeed, we have $F(z) = A^{-1}((B^{-1} \circ h(z))^2) = \psi(h(z))$, where $\psi : \mathbb{D} \to \mathbb{D}$ is a holomorphic map taking 0 to 0 and different from a rotation. By the Schwarz lemma, we obtain that $|\psi'(0)| < 1$, hence $M = |F'(a)| = |\psi'(0)||h'(a)| < |h'(a)|$. □

Note that the Riemann Mapping Theorem has a purely topological consequence: any connected simply connected domain in the sphere is homeomorphic to the disk $\mathbb{D}$. The argument given above and making use of complex analysis is probably the simplest way of proving this topological statement.

p:no-line **Proposition 12.6.** *Suppose that a function f is continuous on a disk $\{|z| < r\}$ and holomorphic everywhere in the disk but on the real line. Then f must also be holomorphic at all real points of the disk.*

Proof. Let $\varepsilon > 0$ be a small positive integer. Consider the function

$$g(z) = \frac{1}{2\pi i} \int_{C_r(0)} \frac{f(\zeta)}{\zeta - z}$$

defined on the disk $\{|z| < r - \varepsilon\}$. This function is holomorphic (and is represented by a convergent power series; the proof is similar to that of Theorem 9.2). We want to prove that g coincides with f at all non-real points of the disk $\{|z| < r - \varepsilon\}$. Then it will follow from continuity that $f(z) = g(z)$ for all z with $|z| < r - \varepsilon$. Since any real point x with $|x| < r$ satisfies also $|x| < r - \varepsilon$ for a sufficiently small $\varepsilon > 0$, the function f is holomorphic at all such points x.

Note that the cycle $C_{r-\varepsilon}(0)$ is the sum of the following two cycles: the cycle σ_+, which is the sum of the real line interval $[r - \varepsilon, r + \varepsilon]$ oriented from left to right and the upper semi-circle of $C_{r-\varepsilon}(0)$ oriented counterclockwise, and the cycle σ_-, which is the sum of the real line interval $[r-\varepsilon, r+\varepsilon]$ oriented from right to left and the lower semi-circle of $C_{r-\varepsilon}(0)$ oriented counterclockwise. By definition, we have

$$g(z) = \frac{1}{2\pi i} \left(\int_{\sigma_+} \alpha(z) + \int_{\sigma_-} \alpha(z) \right).$$

Here $\alpha(z)$ denotes the differential 1-form $\frac{f(\zeta)d\zeta}{\zeta - z}$.

Suppose that $|z| < r - \varepsilon$ and that $\mathrm{Im}(z) > 0$. Then, by the Residue Theorem,

$$\int_{\sigma_+} \alpha(z) = f(z), \quad \int_{\sigma_-} \alpha(z) = 0.$$

Therefore, we have $g(z) = f(z)$ for all points z such that $|z| < r - \varepsilon$ and $\mathrm{Im}(z) > 0$. Suppose now that $|z| < r - \varepsilon$ and that $\mathrm{Im}(z) < 0$. Then, by the Residue Theorem,

$$\int_{\sigma_+} \alpha(z) = 0, \quad \int_{\sigma_-} \alpha(z) = f(z).$$

It follows that $g(z) = f(z)$ in this case as well. □

For a subset $E \subset \mathbb{C}$, we let E^* denote the set of all points $\overline{z}$, where $z \in E$. The following principle allows to extend holomorphic functions to larger domains:

t:refl **Theorem 12.7** (The Schwarz reflection principle). *Suppose that f is a continuous function on a closed set E bounded by a real line interval $[a, b]$ and a smooth curve, whose endpoints are a and b, and which lies in the open upper-half plane otherwise. Suppose also that f is holomorphic in the interior of E and that it takes real values on $[a, b]$. Then the function F defined on the set $E \cup \overline{E}$ by the formula*

$$F(z) = \begin{cases} f(z), & z \in E \\ \overline{f(\overline{z})}, & z \in \overline{E} \end{cases}$$

is holomorphic in the interior of $E \cup \overline{E}$.

Note that, in Theorem 12.7, the real line can be replaced with any circle, since any circle is the image of the real line under a fractional linear transformation. Then, instead of the complex conjugation, we need to use the inversion in the circle.

Proof. We need only prove that F is holomorphic at all points of (a, b). This follows from Proposition 12.6. □

12.1. **Digon.** We now discuss Riemann mappings of some simple domains. First consider a domain U defined by the inequalities

$$0 < \arg(z) < \alpha\pi,$$

where $\alpha \in (0, 2)$. This is a solid angle with the apex at 0. Note that the map $f(z) = z^{\frac{1}{\alpha}}$ takes U to the upper half-plane $\mathbb{H} = \{\mathrm{Im}(z) > 0\}$. Then the map

$$z \mapsto \frac{f(z) - i}{f(z) + i}$$

is a conformal isomorphism between U and $\mathbb{D}$, i.e., a Riemann map of U. Thus, in this case, a Riemann map can be found explicitly.

Let us now discuss a more general case of a digon. A digon U is a region bounded by two arcs of circles. Let $\alpha\pi$ be the angle between the two circles, and let a, b be the two vertices of U. Note that the map

$$z \mapsto e^{i\phi}\frac{z-a}{z-b},$$

where ϕ is a suitable angle, takes U to the solid angle $0 < \arg(w) < \alpha\pi$. Thus the map

$$z \mapsto \left(e^{i\phi}\frac{z-a}{z-b}\right)^{\frac{1}{\alpha}}$$

is a conformal isomorphism between U and $\mathbb{H}$. It is easy to obtain a conformal isomorphism between U and $\mathbb{D}$.

12.2. **Semi-strip.** Consider the domain

$$U = \{z \in \mathbb{C} \,|\, 0 < \mathrm{Re}(z) < \frac{\pi}{2}, \mathrm{Im}(z) > 0\}.$$

This is a lower half of a vertical strip. Note that the map $z \mapsto e^{2iz}$ takes U to the part V of the lower half-plane given by the inequality $|z| > 1$. Note that V is a digon with angle $\frac{\pi}{2}$. The map

$$z \mapsto \frac{z-1}{z+1}$$

takes V to the coordinate quadrant $\mathrm{Re}(w) > 0$, $\mathrm{Im}(w) < 0$. Finally, the map $z \mapsto -z^2$ takes this coordinate quadrant to the upper half-plane. We see that the map

$$z \mapsto -\left(\frac{e^{2iz}-1}{e^{2iz}+1}\right)^2 = \tan(z)^2$$

takes U to the upper half-plane.

12.3. **Rectangle.** Let us look for a conformal isomorphism f between the upper half-plane and the rectangle R with vertices $\pm\omega_1$, $\pm\omega_1 + i\omega_2$, where ω_1, ω_2 are given positive numbers. We assume that f maps the first quadrant $\mathrm{Re}(z) > 0$, $\mathrm{Im}(z) > 0$ to the rectangle with vertices 0, ω_1, $\omega_1 + i\omega_2$ and $i\omega_2$, and that f maps the second quadrant $\mathrm{Re}(z) < 0$, $\mathrm{Im}(z) > 0$ to the rectangle with vertices 0, $-\omega_1$, $-\omega_1 + i\omega_2$ and $i\omega_2$. We will also assume that f extends by continuity to the real line and that $f(0) = 0$, $f(\pm 1) = \pm\omega_1$ and $f(\infty) = i\omega_2$. We will see later than all our assumptions can indeed be fulfilled. Let $\frac{1}{k}$ denote the real number with the property $f(\frac{1}{k}) = \omega_1 + i\omega_2$. Then, by symmetry, we must have $f(-\frac{1}{k}) = -\omega_1 + i\omega_2$.

Let a be any of the four points ± 1, $\pm\frac{1}{k}$. Suppose that $f(a) = b$. The map $z \mapsto (f(z) - b)^2$ takes a neighborhood of a in $\mathbb{H}$ to a neighborhood of 0 in $\mathbb{H}$. By the Schwarz reflection principle, this map extends to a holomorphic map from a neighborhood of a in $\mathbb{C}$ to a neighborhood of 0 in $\mathbb{C}$ preserving 0 and the real axis. Thus we have $(f(z) - b)^2 = \mu(z - a) + \dots$, where μ is a positive real number, and dots denote higher order terms with respect to $(z - a)$. It follows that

$$f(z) = b + \sqrt{\mu}(z-a)^{\frac{1}{2}}(1 + \dots).$$

Here dots denote a holomorphic function of $z-a$. Differentiating, we obtain that

$$f'(z) = \frac{\sqrt{\mu}}{2}(z-a)^{-\frac{1}{2}}(1+\dots),$$
$$f''(z) = -\frac{\sqrt{\mu}}{4}(z-a)^{-\frac{3}{2}}(1+\dots).$$

It follows that

$$\frac{f''(z)}{f'(z)} = -\frac{1}{2}\frac{1}{z-a} + \dots.$$

It follows from the Liouville theorem that

$$\frac{f''(z)}{f'(z)} = -\frac{1}{2}\left(\frac{1}{z-1} + \frac{1}{z+1} + \frac{1}{z-k^{-1}} + \frac{1}{z+k^{-1}}\right),$$

where C is some complex constant. Indeed, the difference of the left-hand side and the right-hand side is a holomorphic map from $\mathbb{C}$ to $\mathbb{C}$. Thus it suffices to prove that the left-hand side tends to zero as $z \to \infty$. Near infinity, we have $f(z) = P(z^{-1})$ is a power series of z^{-1}. It follows that $f'(z) = -z^{-2}P'(z^{-1})$ and

$$f''(z) = 2z^{-3}P'(z^{-1}) + z^{-4}P''(z^{-1}).$$

Therefore, the function $f''(z)/f'(z)$ tends to 0 as $z \to \infty$.

Integrating the expression for $f''(z)/f'(z)$, we obtain that

$$\log f'(z) = -\frac{1}{2}\log(z^2-1)(z^2-k^{-2}) + const.$$

Exponentiating and integrating once more, we obtain that

$$f(z) = A' \int_{z_0}^{z} \frac{dz}{(1-z^2)(k^{-2}-z^2)} + B.$$

Since $f(0) = 0$, we may set $z_0 = 0$ and $B = 0$:

$$f(z) = A \int_0^z \frac{dz}{(1-z^2)(1-k^2z^2)}.$$

Here $A = kA'$. The integral that appears in the right-hand side is called an *elliptic integral* in the form of Legendre.

The interval $(-1,1)$ of the real axis must be mapped by f to the interval $(-\omega_1, \omega_1)$. It follows that $A > 0$. In particular, we have

$$\omega_1 = A \int_0^1 \frac{dt}{(1-t^2)(1-k^2t^2)}.$$

The interval $(1, k^{-1})$ is mapped to a vertical interval. This vertical interval must be $(\omega_1, \omega_1 + i\omega_2)$. It follows that

$$\omega_2 = \int_1^{\frac{1}{k}} \frac{dt}{(t^2-1)(1-k^2t^2)}.$$

The expressions for ω_1 and ω_2 provide a system of two equations with two unknowns A and k. This system can be solved for A and k. Supposing that A and k are found, it is easy to see that the elliptic integral maps the real line to the boundary of the rectangle R. It is not hard to see that it must map the upper half-plane to the interior of R.

13. Multivalued analytic functions

We start with the following important theorem.

Theorem 13.1 (The uniqueness theorem)**.** *Suppose that $U \subset \overline{\mathbb{C}}$ is a connected open set, and $f : U \to \mathbb{C}$ is a holomorphic function. If there is a sequence of pairwise different points $a_n \in U$ such that $f(a_n) = 0$ and $a_n \to a \in U$, then $f = 0$ everywhere on U.*

Proof. Let k be the order of f at a. Then $f(z) = (z - a)^k h(z)$, where $h(a) \neq 0$. Since h is continuous, we must have $h(a_n) \neq 0$ for all sufficiently large n. Then $f(a_n) \neq 0$, a contradiction. □

The theory we now describe is due to Weierstrass. This theory provides a rigorous description of multivalued analytic functions. An (analytic) *element* centered at $a \in \mathbb{C}$ is defined as a power series $f_a(z) = \sum c_k(z - a)^k$ with a nonzero radius of convergence. An *element centered at infinity* is by definition a Laurent series of the form $\sum c_k z^{-k}$, where $c_k = 0$ for all negative k. Two elements f_a and f_b are said to be *neighboring* if their disks of convergence intersect, and $f_a = f_b$ on the intersection. Two elements f_a and f_b are said to be *equivalent* if there is a finite sequence of elements $f_{a_0} = f_a, \dots, f_{a_m} = f_b$ such that f_{a_k} and $f_{a_{k+1}}$ are neighboring for all $k = 0, \dots, m - 1$.

Definition 13.2. A *multivalued analytic function* is defined as an equivalence class of analytic elements.

We say that $w\overline{\mathbb{C}}$ is a value of a multivalued analytic function f at a point $a \in \overline{\mathbb{C}}$ if there is an analytic element f_a of f centered at a such that $f_a(a) = w$. Note that there may be different analytic elements of the same function f centered at the same point. Thus a multivalued analytic function can have many values at the same point, as is suggested by the terminology.

Example 13.3. The following power series is an analytic element centered at 1:

$$f_1(z) = \sum_{k=0}^{\infty} \binom{1/2}{k} (z-1)^k, \qquad \binom{\alpha}{k} = \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!}.$$

The equivalence class of f_1 is a multivalued analytic function called the square root $z \mapsto \sqrt{z}$. We claim that the values of $\sqrt{1}$ are exactly ± 1.

Example 13.4. The following power series is an analytic element centered at 1:

$$g_1(z) = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{(z-1)^k}{k}.$$

The equivalence class of g_1 is a multivalued analytic function called the logarithm $z \mapsto \log z$. We claim that all values of the logarithm at the point 1 are numbers $2\pi i m$, where m is an arbitrary integer.

We now define a *Riemann surface*. The definition will follow roughly the same logic as the definition of a multivalued analytic function. If X is a set, $U \subset X$ is a subset of X, and $\varphi : U \to \mathbb{C}$ is an injective map, then we call the pair (U, φ) a *chart* of X provided that $\varphi(U) \subset \mathbb{C}$ is an open subset. Two charts (U, φ) and (V, ψ) are said to be *compatible* if $U \cap V \neq$, and the *transition function*

$$\tau = \psi \circ \varphi^{-1} : \varphi(U \cap V) \to \psi(U \cap V)$$

is holomorphic. An *atlas* of X is any collection of pair-wise compatible charts $(U_\alpha, \varphi_\alpha)$ such that $\bigcup_\alpha U_\alpha = X$.

Definition 13.5. A *Riemann surface* is a set equipped with an atlas. For technical reasons, we will always assume that either the atlas is at most countable, or there is another at most countable atlas, whose charts are compatible with charts of the first atlas.

We now give two examples of Riemann surfaces. The first example is trivial. Let $U \subset \mathbb{C}$ be an open subset, and $\varphi : U \to U$ be the identity mapping. Then the set U equipped with the atlas consisting of the single chart (U, φ) is a Riemann surface. Thus all open subsets of $\mathbb{C}$ can be regarded as Riemann surfaces. It is also easy to see that any open subset of a Riemann surface carries a natural structure of a Riemann surface.

Let f be a multivalued analytic function. We will defined a *punctured Riemann surface* X of f as the set of all elements of f. To fix a Riemann surface structure on X, we need to define an atlas on X. In fact, every element $f_a \in X$ defines a coordinate chart (U, φ), where U consists of all elements f_b of f such that b is in the disk of convergence of f_a, we have $f_b = f_a$ on the intersection of their disks of convergence, and φ takes f_b to b. It is easy to see that all these charts are pairwise compatible, so that they form an atlas of X. Punctured Riemann surfaces of multivalued analytic functions provide examples of Riemann surfaces. Note also that X comes with a natural map $\pi : X \to \overline{\mathbb{C}}$ assigning the point a to every element $f_a \in X$.

The multivalued analytic function f gives rise to a holomorphic function F on its punctured Riemann surface X. A point in X is an element f_a, and the function F takes this element to its value at a.

We now discuss some general concepts related to Riemann surfaces. Let X and Y be Riemann surfaces, and $F : X \to Y$ a mapping. We say that the mapping F is *holomorphic* if, for every point $x \in X$ and every pair of charts (U, φ) with $x \in U \subset X$ and (V, ψ) with $f(x) \in V \subset Y$, the function

$$\psi \circ f \circ \varphi^{-1} : \varphi(U \cap f^{-1}(V)) \to \psi(f(U) \cap V)$$

is holomorphic. In practice, it suffices to check that this condition is fulfilled for one particular pair of charts (U, φ) and (V, ψ). An *isomorphism* of Riemann surfaces is a bijective holomorphic map. Then the inverse map is also holomorphic, as follows from Proposition 11.6.

We now define the operation of *filling a puncture*. Let us first consider an open subset $U \subset \mathbb{C}$. We say that U has a puncture at a point $a \in \mathbb{C} \setminus U$ if $U \cup \{a\}$ is still open. Now let X be a Riemann surface, and (U, φ) some chart of it. We say that X has a puncture (associated with the chart (U, φ) and a point $a \in \mathbb{C} \setminus \varphi(U)$) if $\varphi(U)$ has a puncture at a, and φ^{-1} does not extend to a neighborhood of a holomorphically. To keep it short, we will talk about a puncture (U, φ, a). Define the set $\widehat{X}$ as the set-theoretic union of X and a point x_0 not lying in X. The set $\widehat{X}$ is endowed with a Riemann surface structure as follows. We keep all the charts of X but the chart (U, φ). The latter is modified by setting $\widehat{U} = U \cup \{x_0\}$ and $\widehat{\varphi}(x_0) = a$ (the values of $\widehat{\varphi}$ at all points of U are the same as the corresponding values of φ). It is easy to see that the charts of the new atlas are compatible. We say that $\widehat{X}$ is obtain from X by *filling the puncture* (U, φ, a).

We have just described a procedure of filling one puncture. However, the same procedure (with obvious modifications) can be applied to fill all punctures simultaneously or any given set of punctures.

13.1. **Problems.**

Problem 13.6. Find the set of all values of the function $z \mapsto \sqrt{z}$ at $z = 1$.

Solution. Let f_a be any element of the function $z \mapsto \sqrt{z}$. We claim that $f_a(z)^2 = z$. Indeed, this is true for one of the elements (which we used to define the function), and, by the uniqueness theorem, if this is true for some element, this is also true for any neighboring element. It follows that the function "square root" can take only two values at $z = 1$, namely, ± 1. To obtain the value -1, we consider a sequence of points ζ^n, where $\zeta = e^{2\pi i/N}$ for a sufficiently large N. Set $\eta = e^{\pi i/N}$, and consider the following analytic elements f_{ζ^n} centered at ζ^n:

$$f_{\zeta^n}(z) = \eta^n \sum_{k=0}^{\infty} \binom{1/2}{k} \left(\frac{z}{\zeta^n} - 1 \right)^k .$$

We claim that the value of f_{ζ^n} at ζ^{n+1} is η^{n+1} rather than $-\eta^{n+1}$, and hence $f_{\zeta^{n+1}}$ and f_{ζ^n} are neighboring. Note finally that the value of the element f_{ζ^N} at 1 is -1.

14. Algebraic singularities and Puiseux series

In this section, we apply the operation of filling punctures to algebraic singularities of a multivalued analytic function.

The following definition carries over standard topological notions (open and closed sets, connectivity, etc.) to Riemann surfaces.

Definition 14.1 (Open subsets of a Riemann surface). Suppose that $W \subset X$ is a subset of a Riemann surface X. The set W is said to be *open* if, for every chart (U, φ) of X, the set $\varphi(W \cap U) \subset \mathbb{C}$ is open. Clearly, the union of any family of open sets is open, and the intersection of finitely many open sets is open. Closed sets are defined as the complements of open sets. All topological notions that can be defined in terms of open and closed sets (such as continuity, connectivity, compactness, etc.) carry over to Riemann surfaces.

Let f be a multivalued analytic function defined by some analytic element f_a centered at a point a. Consider a continuous path $\gamma : [0,1] \to \mathbb{C}$ such that $\gamma(0) = a$, and set $b = a$. We say that an element f_b is obtained by *analytic continuation* of f_a along γ if there is a partition $0 = t_0 \leqslant t_1 \leqslant \ldots \leqslant t_m = 1$ of $[0,1]$ and a sequence of analytic elements $f_{\gamma(t_i)}$ such that the elements $f_{\gamma(t_i)}$ and $f_{\gamma(t_{i+1})}$ are neighboring. An analytic continuation along a given path may or may not exist. If f_b is obtained from f_a by analytic continuation along some path connecting a with b, then f_b is of course equivalent to f_a, i.e., both f_a and f_b belong to the same multivalued analytic function f.

Definition 14.2 (Singular point). Suppose that f is a multivalued analytic function defined by some analytic element f_a, where $a \in \mathbb{C}$. A point $b \in \overline{\mathbb{C}}$ is called a *singular point* of f if there is a path $\gamma : [0,1] \to \overline{\mathbb{C}}$ with the following properties:

1. we have $\gamma(0) = a$ and $\gamma(1) = b$,
2. the element f_a admits analytic continuation along any truncated path $\gamma : [0, \tau] \to \overline{\mathbb{C}}$, $\tau < 1$;
3. the element f_a does not admit analytic continuation along the path γ.

Clearly, this property of b does not depend on a particular choice of an analytic element f_a. Note that even if b is a singular point of f, there may be another path $\widetilde{\gamma}$ connecting a and b such that there is an analytic continuation of f_a along $\widetilde{\gamma}$.

Example 14.3. The point 1 is a singular point of the function $\sqrt{\sqrt{z} + \sqrt[4]{z}}$, however, there is a holomorphic branch of this function over 1.

Example 14.4. The function $\log(\log z + i\log(z-1) + (\sqrt{2}+i\sqrt{3})\log(z+1))$ has a dense set of singular points.

We now discuss the structure of punctures of X, the punctured Riemann surface of f. We start with the following lemma.

l:precomp **Lemma 14.5.** *Let g be a holomorphic function on a neighborhood of 0 such that $\mathrm{ord}_0 g = k$. Then there is a holomorphic injective map ψ on a neighborhood of 0 such that $\psi(0) = 0$ and $g \circ \psi(z) = z^k$.*

Proof. We know that $g(z) = cz^k(1 + zh(z))$, where $c \neq 0$ and h is holomorphic near 0. Choose any complex number a such that $a^k = c$, and consider the series

$$\xi(z) = az \sum_{n=0}^{\infty} \binom{1/k}{n} (zh(z))^n.$$

This series converges in a neighborhood of 0 and defines a holomorphic function on a neighborhood of 0 with the property $\xi(z)^k = g(z)$. Since the derivative of ξ at 0 (which is equal to a) is nonzero, ξ is injective on some neighborhood of 0 (this follows from the Preservation of Number principle). A local inverse function ψ of ξ is as desired. □

Definition 14.6 (Algebraic singularity). Let f be a multivalued analytic function and X be its punctured Riemann surface. We will write $F : X \to \mathbb{C}$ for the holomorphic function on the Riemann surface X associated with f. Consider a puncture of X associated with a chart (U, φ) (recall that there is a point $a \in \mathbb{C}$ such that $\varphi(U) \cup \{a\}$ is an open subset of $\overline{\mathbb{C}}$, the Riemann sphere). Suppose that the functions $\pi \circ \varphi^{-1} : \varphi(U) \to \overline{\mathbb{C}}$ AND $F \circ \varphi^{-1} : \varphi(U)$ have non-essential singularities at the point a. The point $b = \pi \circ \varphi^{-1}(a)$ (which is a finite point or the infinity) is then a singular point of f, which is called an *algebraic singular point*.

Definition 14.7. Let f be a multivalued analytic function with the punctured Riemann surface X. The *Riemann surface of f* is defined as the result of the filling punctures operation applied to the punctured of X corresponding to algebraic singularities.

The Riemann surface $\widehat{X}$ of f comes with a holomorphic mapping $\pi : \widehat{X} \to \overline{\mathbb{C}}$ (which is the holomorphic extension of the previously constructed mapping $\pi : X \to \overline{\mathbb{C}}$). Recall that f gives rise to a single-valued holomorphic function F on its punctured Riemann surface X. By definition of algebraic singularities and the Removable Singularity Theorem, this function extends to a holomorphic mapping F of the Riemann surface $\widehat{X}$ of f to $\overline{\mathbb{C}}$ (we will keep the same notation F for the extension; note that F can take value ∞). Thus a multivalued analytic function is well-defined and single-valued on its Riemann surface. The image $\pi(\widehat{X})$ is called the *domain* of f.

We now study the behavior of a multivalued analytic function near its algebraic singularity. Let f be a multivalued analytic function, F be the corresponding holomorphic function on the Riemann surface $\widehat{X}$ of f, a be an algebraic singular point of f, and x_0 be a point in $\widehat{X} \setminus X$ such that $\pi(x_0) = a$. By definition, there is a chart (U, φ) of $\widehat{X}$ such that U is a neighborhood of x_0, and $\varphi(x_0) = 0$. By Lemma 14.5, possibly replacing φ with a different holomorphic embedding, we may assume that $\pi(\varphi^{-1}(u)) = (a+u)^k$. The function $F \circ \varphi^{-1}$ is meromorphic near zero, therefore, it has a power series expansion

$$F(\varphi^{-1}(u)) = c_{-m}u^{-m} + \cdots + c_{-1}u^{-1} + c_0 + c_1 u + c_2 u^2 + \cdots .$$

If we write $z = \pi(\varphi^{-1}(u))$, then the values of $f(z)$ are equal to $F(\sqrt[k]{z-a})$, i.e., we can formally write the following series expansion

$$f(z) = c_{-m}(z-a)^{-\frac{m}{k}} + \cdots + c_{-1}(z-a)^{-\frac{1}{k}} + c_0 + c_1(z-a)^{\frac{1}{k}} + c_2(z-a)^{\frac{2}{k}} + \cdots .$$

A series of this form is called a *Puiseux series* for f.

If $k > 1$, then the point a is called a *branch point* of f (or a *ramification point* of f). If $k = 1$ but the series for f contains negative powers of $z - a$, then a is called a *pole* of f. If a is neither a pole nor a ramification point, then it is not a singular point of f.

Definition 14.8 (Proper maps). Let X and Y be Riemann surfaces, and let $f : X \to Y$ be a holomorphic map. The map f is said to be *proper* if the preimage of every compact subset of Y is a compact subset of X.

t:deg **Theorem 14.9.** *Consider Riemann surfaces X and Y. Assume that Y is connected. Let $f : X \to Y$ be a proper holomorphic map between Riemann surfaces. If there is some point $y_0 \in Y$ such that the fiber $f^{-1}(y_0)$ consists of $k < \infty$ points, counting multiplicities, then every fiber of f consists of k points counting multiplicities.*

The number k is called the *degree* of f. Thus any proper holomorphic map has a well-defined degree.

Proof. Let us first prove that there is a neighborhood V of y_0 such that, for every $y \in V$, the set $f^{-1}(y)$ consists of exactly k points, counting multiplicities. Let U be a small open neighborhood of $f^{-1}(y_0)$ (which in coordinate charts looks like the union of several disks). By the Preservation of Number Principle, for all y sufficiently close to y_0, the set U contains exactly k points of $f^{-1}(y)$ counting multiplicities. Suppose now that there is no neighborhood V of y_0 with the property

$f^{-1}(y) \subset U$ for all $y \in V$. Then there is a sequence $y_n \to y_0$ and a sequence $x_n \in f^{-1}(y_n) \setminus U$. Since f is proper, we can arrange that $x_n \to x_0 \in X$ by passing to a suitable subsequence. But then $x_0 \notin U$ and $f(x_0) = y_0$, a contradiction.

Since the function assigning the cardinality of $f^{-1}(y)$ to $y \in Y$ is locally constant on Y, and Y is connected, this function is constant on Y. □

14.1. **Proper holomorphic endomorphisms of the disk.** Recall that $\mathbb{D}$ denotes the unit disk. We now describe all proper holomorphic maps from $\mathbb{D}$ to $\mathbb{D}$ (these are also called proper holomorphic *endomorphisms* of $\mathbb{D}$). Let $f : \mathbb{D} \to \mathbb{D}$ be a proper holomorphic map. Set a be any point in $\mathbb{D}$ such that $f(a) = 0$. Consider a holomorphic automorphism A of $\mathbb{D}$ such that $A(a) = 0$, and set $g = f/A$. Since $f(a) = 0$ and A has a simple zero at a, the function g has a removable singularity at a. We have $|g(z)| \to 1$ as $|z| \to 1$, since this is true both for f and A.

Suppose first that $f^{-1}(0)$ consists of only one point (counting multiplicities). Then f has degree one, and it follows from Theorem 14.9 that f is an automorphism of $\mathbb{D}$. In this case, g is a constant function of modulus one.

Suppose now that the degree of f is bigger than one. Then there is some point b with the property $g(b) = 0$ (we may have $b = a$ if $\mathrm{ord}_a(f) > 1$). Since $|g(z)| \to 1$ as $|z| \to 1$ and $g(b) = 0$, we must have $g(\mathbb{D}) = \mathbb{D}$. We claim that g is proper. Saying that g is proper is equivalent to saying that $|g(z)| \to 1$ as $|z| \to 1$.

By induction, we conclude that every proper holomorphic endomorphism of $\mathbb{D}$ is a product (in the sense of multiplication of complex numbers) of automorphisms. Such products are called *Blaschke products*.

14.2. **Puiseux series and Newton diagrams.** Consider a polynomial $f(x, y)$ of two variables such that $f(0, 0) = 0$. Suppose that

$$f(x, y) = \sum_{(k,\ell)} c_{k,\ell} x^k y^\ell.$$

The set of points $(k, \ell) \in \mathbb{Z}^2$ such that $c_{k,\ell} \neq 0$ is called the *support* of f. Let C be the positive real quadrant. A subset Δ of $\mathbb{R}^2$ is called *C-convex* if Δ is convex and $\Delta + C = \Delta$. The minimal C-convex set $N(f)$ containing the support of f is called the *Newton diagram* of f. We will assume that the complement of the Newton diagram in C is bounded.

The equation $f(x, y) = 0$ defines a multivalued function $y(x)$ (or several multivalued functions). The Newton diagram of f plays an important role in finding a Puiseux expansion of $y(x)$, as was first shown by Newton.