A Guide to Complex Variables

Steven G. Krantz

October 14, 2007

To Paul Painlevé (1863–1933).

Table of Contents

Preface

Most every mathematics Ph.D. student must take a qualifying exam in complex variables. The task is a bit daunting. This is one of the oldest areas in mathematics, it is beautiful and compelling, and there is a plethora of material. The literature in complex variables is vast and diverse. There are a great many textbooks in the subject, but each has a different point of view and places different emphases according to the tastes of the author.

Thus it is a bit difficult for the student to focus on what are the essential parts of this subject. What must one absolutely know for the qualifying exam? What will be asked? What techniques will be stressed? What are the key facts?

The purpose of this book is to answer these questions. This is definitely *not* a comprehensive textbook like [GRK]. It is rather an entree to the discipline. It will tell you the key ideas in a first-semester graduate course in the subject, map out the important theorems, and indicate most of the proofs. Here by "indicate" we mean that **(i)** if the proof is short then we include it, **(ii)** if the proof is of medium length then we outline it, and bf (iii) if the proof is long then we sketch it.

This book has plenty of figures, plenty of examples, copious commentary, and even in-text exercises for the students. But, since it is *not* a formal textbook, it does not have exercise sets. It does not have a Glossary or a Table of Notation.

This is meant to be a breezy book that you could read at one or two sittings, just to get the sense of what this subject is about and how it fits together. In that wise it is quite different from a typical mathematics text or monograph. After reading this book (or even *while* reading this book), you will want to pick up a more traditional and comprehensive tome and work your way through it. The present book will get you started on your journey.

This volume is part of a comprehensive series by the Mathematical Association of America that is intended to augment graduate education in this country. We hope that the present volume is a positive contribution to that effort.

Palo Alto, California Steven G. Krantz

Chapter 1

The Complex Plane

1.1 Complex Arithmetic

1.1.1 The Real Numbers

We assume the reader to be familiar with the real number system $\mathbb{R}$. We let $\mathbb{R}^2 = \{(x, y) : x \in \mathbb{R}\ ,\ y \in \mathbb{R}\}$ (Figure 1.1). These are ordered pairs of real numbers.

As we shall see, the complex numbers are nothing other than $\mathbb{R}^2$ equipped with a special algebraic structure.

1.1.2 The Complex Numbers

The complex numbers $\mathbb{C}$ consist of $\mathbb{R}^2$ equipped with some binary algebraic operations. One defines

$$\begin{aligned}(x, y) + (x', y') &= (x + x', y + y')\,,\\ (x, y) \cdot (x', y') &= (xx' - yy', xy' + yx').\end{aligned}$$

These operations of $+$ and $\cdot$ are commutative and associative.

We denote $(1, 0)$ by 1 and (0,1) by i. If $\alpha \in \mathbb{R}$, then we identify α with the complex number $(\alpha, 0)$. Using this notation, we see that

$$\alpha \cdot (x, y) = (\alpha, 0) \cdot (x, y) = (\alpha x, \alpha y). \tag{1.1.2.1}$$

As a result, if (x, y) is *any* complex number, then

$$(x, y) = (x, 0) + (0, y) = x \cdot (1, 0) + y \cdot (0, 1) = x \cdot 1 + y \cdot i \equiv x + iy\,.$$

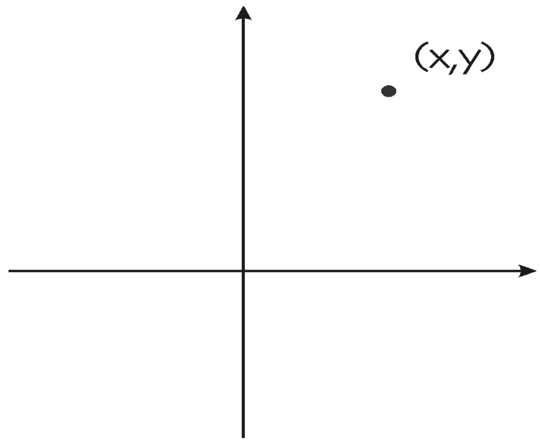

Figure 1.1: The plane $\mathbb{R}^2$.

Thus every complex number (x, y) can be written in one and only one fashion in the form $x \cdot 1 + y \cdot i$ with $x, y \in \mathbb{R}$. As indicated, we usually write the number even more succinctly as $x + iy$. The laws of addition and multiplication become

$$\begin{aligned} (x + iy) + (x' + iy') &= (x + x') + i(y + y'), \\ (x + iy) \cdot (x' + iy') &= (xx' - yy') + i(xy' + yx'). \end{aligned}$$

Observe that $i \cdot i = -1$. Finally, the multiplication law is consistent with the scalar multiplication introduced in line (1.1.2.1).

The symbols z, w, ζ are frequently used to denote complex numbers. We usually take $z = x + iy$, $w = u + iv$, $\zeta = \xi + i\eta$. The real number x is called the *real part* of z and is written $x = \mathrm{Re}\, z$. The real number y is called the *imaginary part* of z and is written $y = \mathrm{Im}\, z$.

The complex number $x - iy$ is by definition the complex *conjugate* of the complex number $x + iy$. If $z = x + iy$, then we denote the conjugate of z with the symbol $\overline{z}$; thus $\overline{z} = x - iy$. The complex conjugate is initially of interest because if p is a quadratic polynomial with real coefficients and if z is a root of p then so is $\overline{z}$.

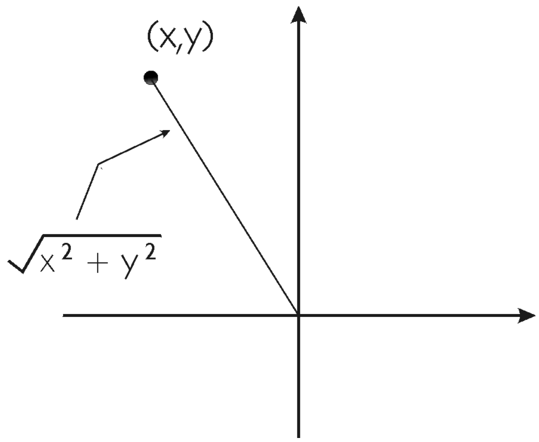

Figure 1.2: Euclidean distance (modulus) in the plane.

1.1.3 Complex Conjugate

Note that $z + \overline{z} = 2x$, $z - \overline{z} = 2iy$. Also

$$\overline{z+w} = \overline{z} + \overline{w}\,, \tag{1.1.3.1}$$

$$\overline{z \cdot w} = \overline{z} \cdot \overline{w}\,. \tag{1.1.3.2}$$

A complex number is real (has no imaginary part) if and only if $z = \overline{z}$. It is imaginary (has no real part) if and only if $z = -\overline{z}$.

1.1.4 Modulus of a Complex Number

The ordinary Euclidean distance of (x, y) to $(0, 0)$ is $\sqrt{x^2 + y^2}$ (Figure 1.2). We also call this number the *modulus* of the complex number $z = x + iy$ and we write $|z| = \sqrt{x^2 + y^2}$. Note that

$$z \cdot \overline{z} = x^2 + y^2 = |z|^2.$$

The distance from z to w is $|z-w|$. We also have the formulas $|z \cdot w| = |z| \cdot |w|$ and $|\mathrm{Re}\, z| \le |z|$ and $|\mathrm{Im}\, z| \le |z|$.

1.1.5 The Topology of the Complex Plane

If P is a complex number and $r > 0$, then we set

$$D(P, r) = \{z \in \mathbb{C} : |z - P| < r\} \tag{1.1.5.1}$$

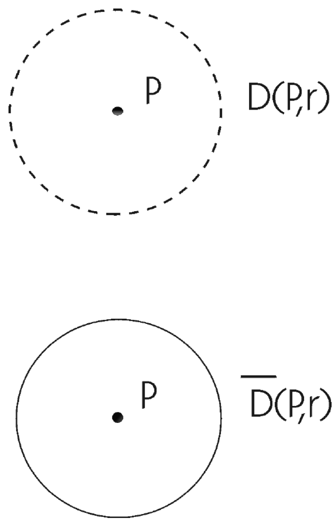

Figure 1.3: Open and closed discs.

and

$$\overline{D}(P,r) = \{z \in \mathbb{C} : |z - P| \leq r\}. \tag{1.1.5.2}$$

The first of these is the *open disc with center P and radius r*; the second is the *closed disc with center P and radius r* (Figure 1.3). We often use the simpler symbols D and $\overline{D}$ to denote, respectively, the discs $D(0,1)$ and $\overline{D}(0,1)$.

We say that a subset $U \subseteq \mathbb{C}$ is *open* if, for each $P \in \mathbb{C}$, there is an $r > 0$ such that $D(P,r) \subseteq U$. Thus an open set is one with the property that each point P of the set is surrounded by neighboring points that are still in the set (that is, the points of distance less than r from P)—see Figure 1.4. Of course the number r will depend on P. As examples, $U = \{z \in \mathbb{C} : \operatorname{Re} z > 1\}$ is open, but $F = \{z \in \mathbb{C} : \operatorname{Re} z \leq 1\}$ is not (Figure 1.5).

A set $E \subseteq \mathbb{C}$ is said to be *closed* if $\mathbb{C} \setminus E \equiv \{z \in \mathbb{C} : z \notin E\}$ (the complement of E in $\mathbb{C}$) is open. The set F in the last paragraph is closed.

It is *not* the case that any given set is either open or closed. For example, the set $W = \{z \in \mathbb{C} : 1 < \operatorname{Re} z \leq 2\}$ is *neither* open *nor* closed (Figure 1.6).

We say that a set $E \subset \mathbb{C}$ is *connected* if there do not exist non-empty 不相交的 disjoint open sets U and V such that $E = (U \cap E) \cup (V \cap E)$. Refer to Figure 1.7 for these ideas. It is a useful fact that if $E \subseteq \mathbb{C}$ is an open set, then E is connected if and only if it is path-connected; this last means that any two points of E can be connected by a continuous path or curve. See Figure 1.8.

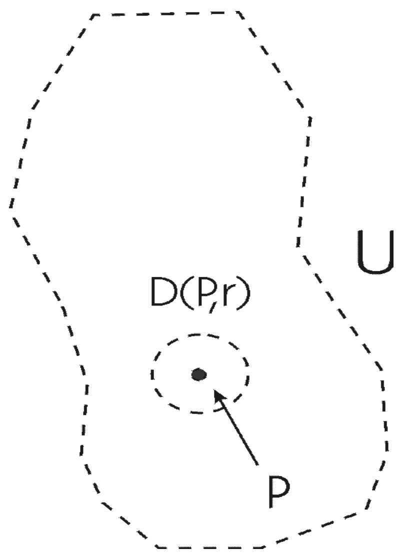

Figure 1.4: An open set.

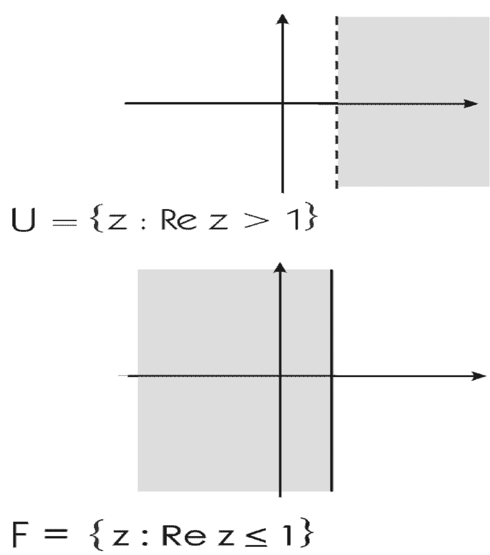

Figure 1.5: Open and non-open sets.

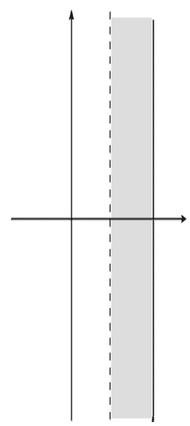

Figure 1.6: A set that is neither open nor closed.

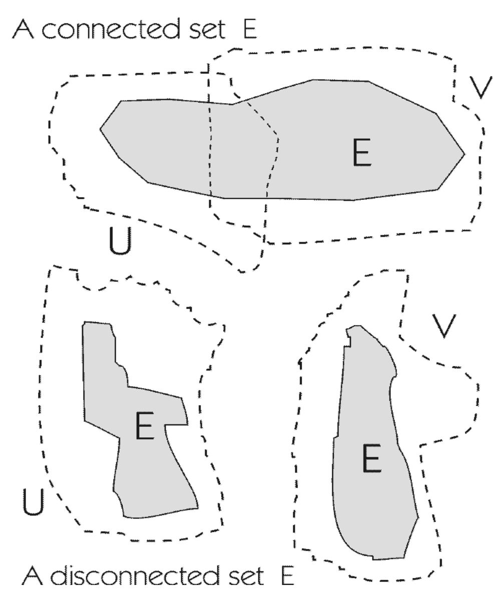

Figure 1.7: The concept of connectivity.

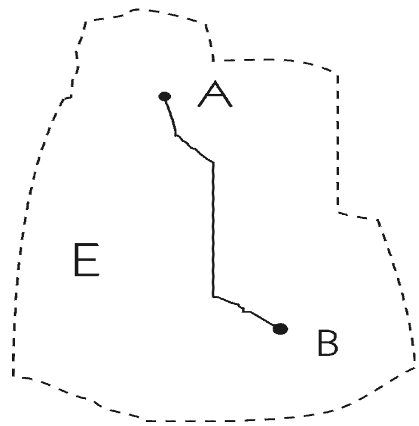

Figure 1.8: Path connectedness.

1.1.6 The Complex Numbers as a Field

Let 0 denote the number $0 + i0$. If $z \in \mathbb{C}$, then $z + 0 = z$. Also, letting $-z = -x - iy$, we have $z + (-z) = 0$. So every complex number has an additive inverse, and that inverse is unique.

Since $1 = 1 + i0$, it follows that $1 \cdot z = z \cdot 1 = z$ for every complex number z. If $z \neq 0$, then $|z|^2 \neq 0$ and

$$z \cdot \left(\frac{\overline{z}}{|z|^2} \right) = \frac{|z|^2}{|z|^2} = 1. \tag{1.1.6.1}$$

So every non-zero complex number has a multiplicative inverse, and that inverse is unique. It is natural to define $1/z$ to be the multiplicative inverse $\overline{z}/|z|^2$ of z and, more generally, to define

$$\frac{z}{w} = z \cdot \frac{1}{w} = \frac{z\overline{w}}{|w|^2} \qquad \text{for } w \neq 0. \tag{1.1.6.2}$$

We also have $\overline{z/w} = \overline{z}/\overline{w}$.

Multiplication and addition satisfy the usual distributive, associative, and commutative laws. Therefore $\mathbb{C}$ is a *field* (see [HER]). The field $\mathbb{C}$ contains a copy of the real numbers in an obvious way:

$$\mathbb{R} \ni x \mapsto x + i0 \in \mathbb{C}. \tag{1.1.6.3}$$

This identification respects addition and multiplication. So we can think of $\mathbb{C}$ as a field extension of $\mathbb{R}$: it is a larger field which contains the field $\mathbb{R}$.

1.1.7 The Fundamental Theorem of Algebra

It is not true that every non-constant polynomial with real coefficients has a real root. For instance, $p(x) = x^2 + 1$ has no real roots. The Fundamental Theorem of Algebra states that every polynomial with complex coefficients has a complex root (see the treatment in §§3.1.4 below). The complex field $\mathbb{C}$ is the *smallest* field that contains $\mathbb{R}$ and has this so-called algebraic closure property. One of the first powerful and elegant applications of complex variable theory is to provide a proof of the Fundamental Theorem of Algebra.

1.2 The Exponential and Applications

1.2.1 The Exponential Function

We define the complex exponential as follows:

(1.2.1.1) If $z = x$ is real, then

$$e^z = e^x \equiv \sum_{j=0}^{\infty} \frac{x^j}{j!}$$

as in calculus. Here ! denotes "factorial": $j! = j\cdot(j-1)\cdot(j-2)\cdots 3\cdot 2\cdot 1$.

(1.2.1.2) If $z = iy$ is pure imaginary, then

$$e^z = e^{iy} \equiv \cos y + i \sin y.$$

(1.2.1.3) If $z = x + iy$, then

$$e^z = e^{x+iy} \equiv e^x \cdot e^{iy} = e^x \cdot (\cos y + i \sin y).$$

Part and parcel of the last definition of the exponential is the following complex-analytic definition of the sine and cosine functions:

$$\cos z = \frac{e^{iz} + e^{-iz}}{2}, \tag{1.2.1.4}$$

$$\sin z = \frac{e^{iz} - e^{-iz}}{2i}. \tag{1.2.1.5}$$

Note that when $z = x + i0$ is real this new definition coincides with the familiar Euler formula from calculus:

$$e^{it} = \cos t + i \sin t \, . \tag{1.2.1.6}$$

1.2.2 The Exponential Using Power Series

It is also possible to define the exponential using power series:

$$e^z = \sum_{j=0}^{\infty} \frac{z^j}{j!}. \tag{1.2.2.1}$$

Either definition (that in §§1.2.1 or in §§1.2.2) is correct for any z, and they are logically equivalent.

1.2.3 Laws of Exponentiation

The complex exponential satisfies familiar rules of exponentiation:

$$e^{z+w} = e^z \cdot e^w \quad \text{and} \quad (e^z)^w = e^{zw}. \tag{1.2.3.1}$$

Also

$$\left(e^z\right)^n = \underbrace{e^z \cdots e^z}_{n \text{ times}} = e^{nz}. \tag{1.2.3.2}$$

One may verify these properties directly from the power series definition, or else use the more explicit definitions in **(1.2.1.1)–(1.2.1.3)**.

1.2.4 Polar Form of a Complex Number

A consequence of our first definition of the complex exponential —see (1.2.1.2)— is that if $\zeta \in \mathbb{C}$, $|\zeta| = 1$, then there is a unique number θ, $0 \le \theta < 2\pi$, such that $\zeta = e^{i\theta}$ (see Figure 1.9). Here θ is the (signed) angle between the positive x axis and the ray $\overrightarrow{0\zeta}$.

Now, if z is any non-zero complex number, then

$$z = |z| \cdot \left(\frac{z}{|z|} \right) \equiv |z| \cdot \zeta \, , \tag{1.2.4.1}$$

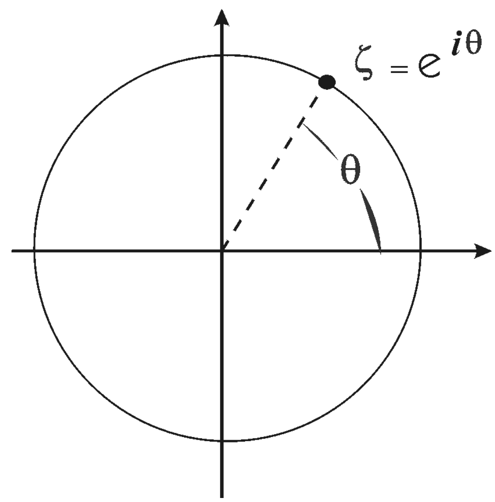

Figure 1.9: Polar representation of a complex number of modulus 1.

where $\zeta = z/|z|$ has modulus 1. Again letting θ be the angle between the real axis and $\overrightarrow{0\zeta}$, we see that

$$\begin{aligned} z &= |z| \cdot \zeta \\ &= |z| e^{i\theta} \\ &= r e^{i\theta}, \end{aligned} \tag{1.2.4.2}$$

where $r = |z|$. This form is called the *polar* representation for the complex number z. (Note that some classical books write the expression $z = re^{i\theta} = r(\cos\theta + i\sin\theta)$ as $z = r\mathrm{cis}\,\theta$. The reader should be aware of this notation, though we shall not use it in this book.) Engineers like the cis notation.

EXAMPLE 1.2.4.1 Let $z = 1 + \sqrt{3}i$. Then $|z| = \sqrt{1^2 + (\sqrt{3})^2} = 2$. Hence

$$z = 2 \cdot \left(\frac{1}{2} + i\,\frac{\sqrt{3}}{2} \right).$$

The unit-modulus number in parenthesis subtends an angle of $\pi/3$ with the positive x-axis. Therefore

$$1 + \sqrt{3}i = z = 2 \cdot e^{i\pi/3}. \qquad \square$$

It is often convenient to allow angles that are greater than or equal to 2π in the polar representation; when we do so, the polar representation is no longer unique. For if k is an integer, then

$$\begin{aligned} e^{i\theta} &= \cos\theta + i\sin\theta \\ &= \cos(\theta + 2k\pi) + i\sin(\theta + 2k\pi) \\ &= e^{i(\theta+2k\pi)} . \end{aligned} \tag{1.2.4.3}$$

1.2.5 Roots of Complex Numbers

The properties of the exponential operation can be used to find the n^{th} roots of a complex number.

EXAMPLE 1.2.5.1 To find all sixth roots of 2, we let $re^{i\theta}$ be an arbitrary sixth root of 2 and solve for r and θ. If

$$\left(re^{i\theta}\right)^6 = 2 = 2 \cdot e^{i0} \tag{1.2.5.1.1}$$

or

$$r^6 e^{i6\theta} = 2 \cdot e^{i0} , \tag{1.2.5.1.2}$$

then it follows that $r = 2^{1/6} \in \mathbb{R}$ and $\theta = 0$ solve this equation. So the real number $2^{1/6} \cdot e^{i0} = 2^{1/6}$ is a sixth root of two. This is not terribly surprising, but we are not finished.

We may also solve

$$r^6 e^{i6\theta} = 2 = 2 \cdot e^{2\pi i}. \tag{1.2.5.1.3}$$

Hence

$$r = 2^{1/6} \ , \ \theta = 2\pi/6 = \pi/3. \tag{1.2.5.1.4}$$

This gives us the number

$$2^{1/6} e^{i\pi/3} = 2^{1/6}\big(\cos\pi/3 + i\sin\pi/3\big) = 2^{1/6}\left(\frac{1}{2} + i\frac{\sqrt{3}}{2}\right) \tag{1.2.5.1.5}$$

as a sixth root of two. Similarly, we can solve

$$\begin{aligned} r^6 e^{i6\theta} &= 2 \cdot e^{4\pi i} \\ r^6 e^{i6\theta} &= 2 \cdot e^{6\pi i} \\ r^6 e^{i6\theta} &= 2 \cdot e^{8\pi i} \\ r^6 e^{i6\theta} &= 2 \cdot e^{10\pi i} \end{aligned}$$

to obtain the other four sixth roots of 2:

$$2^{1/6}\left(-\frac{1}{2}+i\frac{\sqrt{3}}{2}\right) \tag{1.2.5.1.6}$$

$$-2^{1/6} \tag{1.2.5.1.7}$$

$$2^{1/6}\left(-\frac{1}{2}-i\frac{\sqrt{3}}{2}\right) \tag{1.2.5.1.8}$$

$$2^{1/6}\left(\frac{1}{2}-i\frac{\sqrt{3}}{2}\right). \tag{1.2.5.1.9}$$

These are in fact all the sixth roots of 2. □

Remark: One could of course continue the procedure in the last example, solving $r^6 e^{i6\theta} = 2 \cdot e^{12\pi i}$, etc.. But this would simply result in repetition of the roots we have already found.

EXAMPLE 1.2.5.2 Let us find all third roots of i. We begin by writing i as

$$i = e^{i\pi/2}. \tag{1.2.5.2.1}$$

Solving the equation

$$(re^{i\theta})^3 = i = e^{i\pi/2} \tag{1.2.5.2.2}$$

then yields $r = 1$ and $\theta = \pi/6$.

Next, we write $i = e^{i5\pi/2}$ and solve

$$(re^{i\theta})^3 = e^{i5\pi/2} \tag{1.2.5.2.3}$$

to obtain that $r = 1$ and $\theta = 5\pi/6$.

Lastly, we write $i = e^{i9\pi/2}$ and solve

$$(re^{i\theta})^3 = e^{i9\pi/2} \tag{1.2.5.2.4}$$

to obtain that $r = 1$ and $\theta = 9\pi/6 = 3\pi/2$.

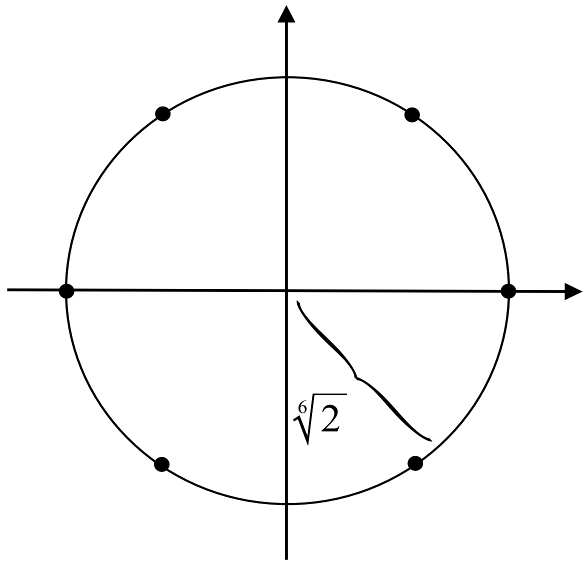

Figure 1.10: The sixth roots of 2.

In summary, the three cube roots of i are

$$e^{i\pi/6} = \frac{\sqrt{3}}{2} + i\frac{1}{2}\,, \tag{1.2.5.2.5}$$

$$e^{i5\pi/6} = -\frac{\sqrt{3}}{2} + i\frac{1}{2}\,, \tag{1.2.5.2.6}$$

$$e^{i3\pi/2} = -i\,. \tag{1.2.5.2.7}$$

□

It is worth noting that, in both Examples 1.2.5.1 and 1.2.5.2, the roots of the given complex number are equally spaced about a circle centered at the origin. See Figures 1.10 and 1.11.

1.2.6 The Argument of a Complex Number

The (non-unique) angle θ associated to a complex number $z \neq 0$ is called its *argument*, and is written $\arg z$. For instance, $\arg(1+i) = \pi/4$. But it is also correct to write $\arg(1+i) = 9\pi/4, 17\pi/4, -7\pi/4$, etc. We generally choose the argument θ to satisfy $0 \leq \theta < 2\pi$. This is the principal branch of the argument—see §§9.1.2, §§9.4.2.

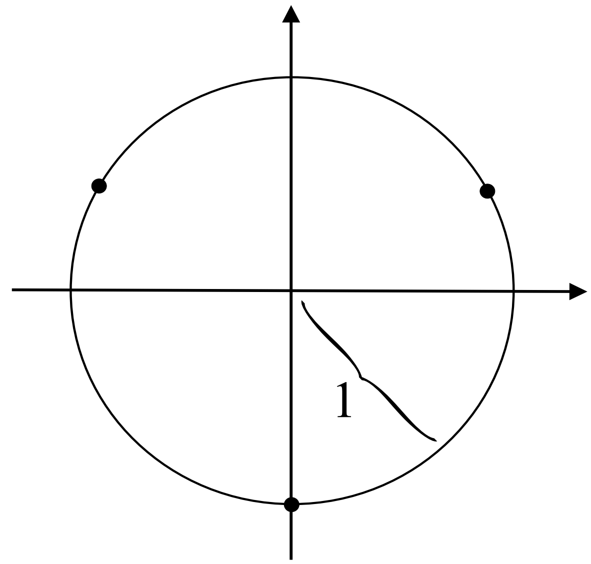

Figure 1.11: The third roots of i.

Under multiplication of complex numbers, arguments are additive and moduli multiply. That is, if $z = re^{i\theta}$ and $w = se^{i\psi}$ then

$$z \cdot w = re^{i\theta} \cdot se^{i\psi} = (rs) \cdot e^{i(\theta+\psi)}. \qquad (1.2.6.1)$$

1.2.7 Fundamental Inequalities

We next record a few important inequalities.

The Triangle Inequality: If $z, w \in \mathbb{C}$, then

$$|z + w| \leq |z| + |w|. \qquad (1.2.7.1)$$

More generally,

$$\left| \sum_{j=1}^{n} z_j \right| \leq \sum_{j=1}^{n} |z_j|. \qquad (1.2.7.2)$$

For $n = 2$, this basic fact can be seen immediately from a picture: any side of a triangle has length not exceeding the sum of the other two sides. The general case follows by induction on n. The rigorous proof involves solving an extremal problem using calculus—see [KRA3].

The Cauchy-Schwarz Inequality: If $z_1, \ldots, z_n$ and $w_1, \ldots, w_n$ are complex numbers, then

$$\left| \sum_{j=1}^{n} z_j w_j \right|^2 \leq \left[\sum_{j=1}^{n} |z_j|^2 \right] \cdot \left[\sum_{j=1}^{n} |w_j|^2 \right]. \tag{1.2.7.3}$$

This result is immediate from the Triangle Inequality: Just square both sides and multiply everything out.

1.3 Holomorphic Functions

1.3.1 Continuously Differentiable and C^k Functions

In this book we will frequently refer to a *domain* or a *region* $U \subseteq \mathbb{C}$. Usually this will mean that U is an open set and that U is connected (see §1.1.5).

Holomorphic functions are a generalization of complex polynomials. But they are more flexible objects than polynomials. The collection of all polynomials is closed under addition and multiplication. However, the collection of all holomorphic functions is closed under reciprocals, inverses, exponentiation, logarithms, square roots, and many other operations as well.

There are several different ways to introduce the concept of holomorphic function. It can be defined by way of power series, or using the complex derivative, or using partial differential equations. We shall touch on all these approaches; but our initial definition will be by way of partial differential equations. First ewe need some preliminary concepts from real analysis.

If $U \subseteq \mathbb{R}^2$ is open and $f : U \to \mathbb{R}$ is a continuous function, then f is called C^1 (or *continuously differentiable*) on U if $\partial f/\partial x$ and $\partial f/\partial y$ exist and are *continuous* on U. We write $f \in C^1(U)$ for short.

More generally, if $k \in \{0, 1, 2, ...\}$, then a real-valued function f on U is called C^k (k times continuously differentiable) if all partial derivatives of f up to and including order k exist and are continuous on U. We write in this case $f \in C^k(U)$. In particular, a C^0 function is just a continuous function.

A function $f = u + iv : U \to \mathbb{C}$ is called C^k if both u and v are C^k.

1.3.2 The Cauchy-Riemann Equations

If f is *any* complex-valued function, then we may write $f = u + iv$, where u and v are real-valued functions.

EXAMPLE 1.3.2.1 Consider

$$f(z) = z^2 = (x^2 - y^2) + i(2xy);$$

in this example $u = x^2 - y^2$ and $v = 2xy$. We refer to u as the *real part* of f and denote it by $\mathrm{Re}\, f$; we refer to v as the *imaginary part* of f and denote it by $\mathrm{Im}\, f$. □

Now we formulate the notion of "holomorphic function" in terms of the real and imaginary parts of f :

Let $U \subseteq \mathbb{C}$ be an open set and $f : U \to \mathbb{C}$ a C^1 function. Write

$$f(z) = f(x + iy) \equiv \widetilde{f}(x, y) = u(x, y) + iv(x, y),$$

with $z = x + iy$ and u and v real-valued functions. If u and v satisfy the equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad\qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \tag{1.3.2.2}$$

at every point of U, then the function f is said to be holomorphic (see §§1.3.4, where a formal definition of "holomorphic" is provided). The first order, linear partial differential equations in (1.3.2.2) are called the *Cauchy-Riemann equations*. A practical method for checking whether a given function is holomorphic is to check whether it satisfies the Cauchy-Riemann equations. Another intuitively appealing method, which we develop in the next subsection, is to verify that the function in question depends on z only and *not* on $\overline{z}$.

1.3.3 Derivatives

We define, for $f = u + iv : U \to \mathbb{C}$ a C^1 function,

$$\frac{\partial}{\partial z} f \equiv \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right) f = \frac{1}{2}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + \frac{i}{2}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) \tag{1.3.3.1}$$

and

$$\frac{\partial}{\partial \overline{z}} f \equiv \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) f = \frac{1}{2}\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) + \frac{i}{2}\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right). \tag{1.3.3.2}$$

If $z = x + iy$, $\overline{z} = x - iy$, then one can check directly that

$$\begin{aligned} \frac{\partial}{\partial z} z = 1 \quad &, \quad \frac{\partial}{\partial z} \overline{z} = 0 \,, \\ \frac{\partial}{\partial \overline{z}} z = 0 \quad &, \quad \frac{\partial}{\partial \overline{z}} \overline{z} = 1 \,. \end{aligned} \tag{1.3.3.3}$$

If a C^1 function f satisfies $\partial f/\partial z \equiv 0$ on an open set U, then f does not depend on z (but it *does* depend on $\overline{z}$). If instead f satisfies $\partial f/\partial \overline{z} \equiv 0$ on an open set U, then f does not depend on $\overline{z}$ (but it *does* depend on z). The condition $\partial f/\partial \overline{z} = 0$ is a reformulation of the Cauchy-Riemann equations—see §§1.3.4.

1.3.4 Definition of Holomorphic Function

Functions f that satisfy $(\partial/\partial \overline{z}) f \equiv 0$ are the main concern of complex analysis. A continuously differentiable (C^1) function $f : U \to \mathbb{C}$ defined on an open subset U of $\mathbb{C}$ is said to be *holomorphic* if

$$\frac{\partial f}{\partial \overline{z}} = 0 \tag{1.3.4.1}$$

at every point of U. Note that this last equation is just a reformulation of the Cauchy-Riemann equations (§§1.3.2). To see this, we calculate:

$$\begin{aligned} 0 &= \frac{\partial}{\partial \overline{z}} f(z) \\ &= \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)[u(z) + iv(z)] \\ &= \left[\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right] + \ i \ \left[\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right] \,. \end{aligned} \tag{1.3.4.2}$$

Of course the far right-hand side cannot be identically zero unless each of its real and imaginary parts is identically zero. It follows that

$$\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} = 0$$

A **holomorphic** function f:C→C is differentiable at each point in an open set A⊂C.

An **analytic** function f:C→C can be expanded as power series.

and

$$\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} = 0.$$

These are the Cauchy-Riemann equations (1.3.2.2).

1.3.5 The Complex Derivative

Let $U \subseteq \mathbb{C}$ be open, $P \in U$, and $g : U \setminus \{P\} \to \mathbb{C}$ a function. We say that

$$\lim_{z \to P} g(z) = \ell \; , \quad \ell \in \mathbb{C} \; , \tag{1.3.5.1}$$

if for any $\epsilon > 0$ there is a $\delta > 0$ such that when $z \in U$ and $0 < |z - P| < \delta$ then $|g(z) - \ell| < \epsilon$. This is similar to the calculus definition of limit, but it allows z to approach P from *any* direction.

We say that f possesses the *complex derivative* at P if

$$\lim_{z \to P} \frac{f(z) - f(P)}{z - P} \tag{1.3.5.2}$$

exists. In that case we denote the limit by $f'(P)$ or sometimes by

$$\frac{df}{dz}(P) \qquad \text{or} \qquad \frac{\partial f}{\partial z}(P). \tag{1.3.5.3}$$

This notation is consistent with that introduced in §§1.3.3: for a *holomorphic function*, the complex derivative calculated according to formula (1.3.5.2) or according to formula (1.3.3.1) is just the same (use the Cauchy-Riemann equations). We shall say more about the complex derivative in §2.2.3 and §2.2.4.

It should be noted that, in calculating the limit in (1.3.5.2), z must be allowed to approach P from *any* direction (see Figure 1.12). As an example, the function $g(x, y) = \overline{z} = x - iy$—equivalently, $g(z) = \overline{z}$—does *not* possess the complex derivative at 0. To see this, calculate the limit

$$\lim_{z \to P} \frac{g(z) - g(P)}{z - P}$$

with z approaching $P = 0$ through values $z = x + i0$. The answer is

$$\lim_{x \to 0} \frac{x - 0}{x - 0} = 1.$$

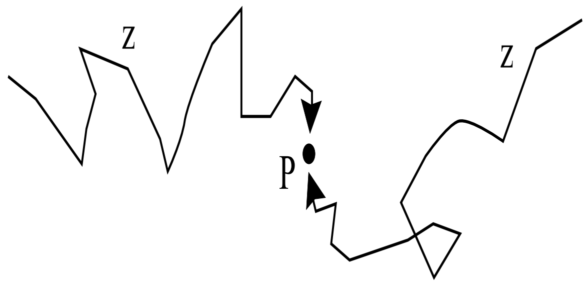

Figure 1.12: The limit from any direction.

If instead z is allowed to approach $P = 0$ through values $z = iy$, then the value is

$$\lim_{y \to 0} \frac{-iy - 0}{iy - 0} = -1.$$

Observe that the two answers do not agree. *In order for the complex derivative to exist, the limit must exist and assume only one value no matter how z approaches P.* Therefore this example g does not possess the complex derivative at $P = 0$. In fact a similar calculation shows that this function g does not possess the complex derivative at any point.

If a function f possesses the complex derivative at every point of its open domain U, then f is holomorphic. This definition is equivalent to definitions given in §§1.3.2, §§1.3.4. We repeat some of these ideas in §2.2.

1.3.6 Alternative Terminology for Holomorphic Functions

Some books use the word "analytic" instead of "holomorphic." Still others say "differentiable" or "complex differentiable" instead of "holomorphic." The use of the term "analytic" derives from the fact that a holomorphic function has a local power series expansion about each point of its domain (see §§3.1.6). In fact this power series property is a complete characterization of holomorphic functions; we shall discuss it in detail below. The use of "differentiable" derives from properties related to the complex derivative. These pieces of terminology and their significance will all be sorted out as the book develops. Somewhat archaic terminology for holomorphic functions, which may be found in older texts, are "regular" and "monogenic."

Another piece of terminology that is applied to holomorphic functions is "conformal" or "conformal mapping." "Conformality" is an important geometric property of holomorphic functions that make these functions useful for modeling incompressible fluid flow and other physical phenomena. We shall discuss conformality in §§2.2.5. See also [KRA6].

1.4 Holomorphic and Harmonic Functions

1.4.1 Harmonic Functions

A C^2 function u is said to be *harmonic* if it satisfies the equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) u = 0. \tag{1.4.1.1}$$

This equation is called *Laplace's equation*, and is frequently abbreviated as

$$\triangle u = 0. \tag{1.4.1.2}$$

1.4.2 How They are Related

If f is a holomorphic function and $f = u + iv$ is the expression of f in terms of its real and imaginary parts, then both u and v are harmonic. An elegant way to see this is to observe that

$$\frac{\partial}{\partial \overline{z}} f = 0$$

hence

$$\frac{\partial}{\partial z}\frac{\partial}{\partial \overline{z}} f = 0\,.$$

But we may write out the lefthand side of the last equation to find that

$$\frac{1}{4} \triangle f = 0$$

or

$$(\triangle u) + i(\triangle v) = 0\,.$$

It is important to note here that the Laplacian $\triangle$ is a real operator. Thus the only way that the last identity can be true is if

$$\triangle u = 0 \quad \text{and} \quad \triangle v = 0\,.$$

This is what we have asserted.

A converse is true provided the functions involved are defined on a domain with no holes:

Theorem: If $\mathcal{R}$ is an open rectangle (or open disc) and if u is a real-valued harmonic function on $\mathcal{R}$, then there is a holomorphic function F on $\mathcal{R}$ such that $\mathrm{Re}\, F = u$. In other words, for such a function u there exists a harmonic function v defined on $\mathcal{R}$ such that $f \equiv u + iv$ is holomorphic on $\mathcal{R}$. Any two such functions v must differ by a real constant.

More generally, if U is a region with no holes (a *simply connected* region—see §§2.3.3), and if u is harmonic on U, then there is a holomorphic function F on U with $\mathrm{Re}\, F = u$. In other words, for such a function u there exists a harmonic function v defined on U such that $f \equiv u + iv$ is holomorphic on U. Any two such functions v must differ by a constant. We call the function v a *harmonic conjugate* for u.

To give an indication of why these statements are true we note that, given u harmonic, we seek v such that

$$\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y} \equiv \alpha(x, y)$$

and

$$\frac{\partial v}{\partial y} = \frac{\partial u}{\partial x} \equiv \beta(x, y)$$

(these are the Cauchy-Riemann equations). We know from calculus that a pair of equations like this is solvable on a region with no holes precisely when

$$\frac{\partial \alpha}{\partial y} = \frac{\partial \beta}{\partial x} .$$

But this last is just the condition that u be harmonic. This explains why v, and hence $F = u + iv$, exists.

The displayed theorem is false on a domain with a hole, such as an annulus. For example, the harmonic function $u = \log(x^2 + y^2)$, defined on the annulus $U = \{z : 1 < |z| < 2\}$, has no harmonic conjugate on U. See also §§7.1.4.

Chapter 2

Complex Line Integrals

2.1 Real and Complex Line Integrals

In this section we shall recast the line integral from calculus in complex notation. The result will be the complex line integral. The complex line integral is essential to the Cauchy theory, which we develop below, and that in turn is key to the argument principle and many of the other central ideas of the subject.

2.1.1 Curves

It is convenient to think of a *curve* as a continuous function γ from a closed interval $[a,b] \subseteq \mathbb{R}$ into $\mathbb{R}^2 \approx \mathbb{C}$. We sometimes let $\widetilde{\gamma}$ denote the *image* of the mapping. Thus

$$\widetilde{\gamma} = \{\gamma(t) : t \in [a,b]\}\,.$$

Often we follow the custom of referring to *either* the function or the image with the single symbol γ. It will be clear from context what is meant. Refer to Figure 2.1.

It is often convenient to write

$$\gamma(t) = (\gamma_1(t), \gamma_2(t)) \qquad \text{or} \qquad \gamma(t) = \gamma_1(t) + i\gamma_2(t). \tag{2.1.1.1}$$

For example, $\gamma(t) = (\cos t, \sin t) = \cos t + i \sin t,\ t \in [0, 2\pi]$, describes the unit circle in the plane. The circle is traversed in a counterclockwise manner as t increases from 0 to 2π. Again see Figure 2.1.

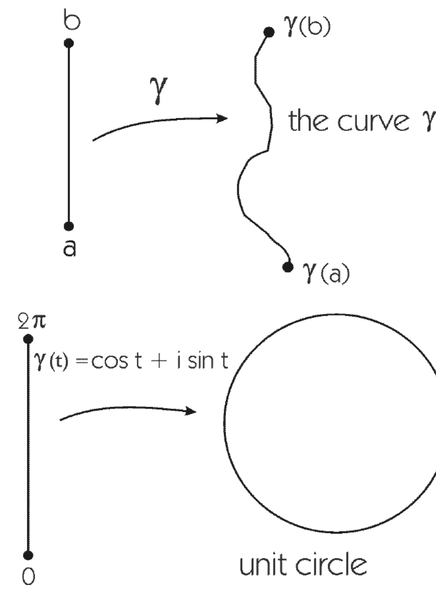

Figure 2.1: Curves in the plane.

2.1.2 Closed Curves

The curve $\gamma : [a, b] \to \mathbb{C}$ is called *closed* if $\gamma(a) = \gamma(b)$. It is called *simple, closed* (or Jordan) if the restriction of γ to the interval $[a, b)$ (which is commonly written $\gamma\big|_{[a,b)}$) is one-to-one *and* $\gamma(a) = \gamma(b)$ (Figure 2.2). Intuitively, a simple, closed curve is a curve with no self-intersections, except of course for the closing up at $t = a, b$.

In order to work effectively with γ we need to impose on it some differentiability properties.

2.1.3 Differentiable and C^k Curves

A function $\varphi : [a, b] \to \mathbb{R}$ is called *continuously differentiable* (or C^1), and we write $\varphi \in C^1([a, b])$, if

(2.1.3.1) φ is continuous on $[a, b]$;
(2.1.3.2) φ' exists on (a, b);
(2.1.3.3) φ' has a continuous extension to $[a, b]$.

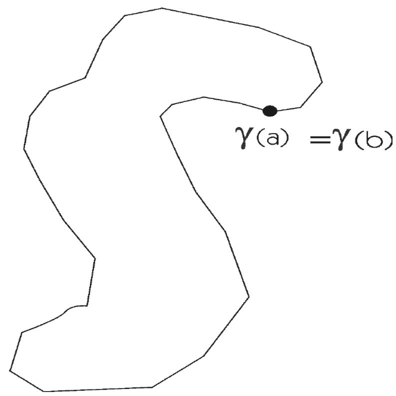

Figure 2.2: A simple, closed curve.

In other words, we require that

$$\lim_{t\to a^+} \varphi'(t) \quad \text{and} \quad \lim_{t\to b^-} \varphi'(t)$$

both exist.

Note that

$$\varphi(b) - \varphi(a) = \int_a^b \varphi'(t)\,dt, \tag{2.1.3.4}$$

so that the Fundamental Theorem of Calculus holds for $\varphi \in C^1([a,b])$.

A curve $\gamma : [a,b] \to \mathbb{C}$, with $\gamma(t) = \gamma_1(t) + i\gamma_2(t)$ is said to be *continuous* on $[a,b]$ if both γ_1 and γ_2 are. We write $\gamma \in C^0([a,b])$. The curve is *continuously differentiable* (or C^1) on $[a,b]$, and we write

$$\gamma \in C^1([a,b]), \tag{2.1.3.5}$$

if γ_1, γ_2 are continuously differentiable on $[a,b]$. Under these circumstances we will write

$$\frac{d\gamma}{dt} = \frac{d\gamma_1}{dt} + i\frac{d\gamma_2}{dt}. \tag{2.1.3.6}$$

We also write $\gamma'(t)$ or even $\dot\gamma(t)$ for $d\gamma/dt$.

2.1.4 Integrals on Curves

Let $\psi : [a,b] \to \mathbb{C}$ be continuous on $[a,b]$. Write $\psi(t) = \psi_1(t) + i\psi_2(t)$. Then we define

$$\int_a^b \psi(t)\,dt \equiv \int_a^b \psi_1(t)\,dt + i\int_a^b \psi_2(t)\,dt. \tag{2.1.4.1}$$

We summarize the ideas presented thus far by noting that if $\gamma \in C^1([a,b])$ is complex-valued, then

$$\gamma(b) - \gamma(a) = \int_a^b \gamma'(t)\, dt. \tag{2.1.4.2}$$

2.1.5 The Fundamental Theorem of Calculus along Curves

Now we state the Fundamental Theorem of Calculus (see [BKR]) adapted to curves.

Let $U \subseteq \mathbb{C}$ be a domain and let $\gamma : [a,b] \to U$ be a C^1 curve. If $f \in C^1(U)$, then

$$f(\gamma(b)) - f(\gamma(a)) = \int_a^b \left(\frac{\partial f}{\partial x}(\gamma(t)) \cdot \frac{d\gamma_1}{dt} \ + \frac{\partial f}{\partial y}(\gamma(t)) \cdot \frac{d\gamma_2}{dt} \right) dt. \tag{2.1.5.1}$$

For the proof, simply reduce the assertion (2.1.5.1) to the analogous classical assertion from the calculus.

2.1.6 The Complex Line Integral

When f is holomorphic, then formula (2.1.5.1) may be rewritten (using the Cauchy-Riemann equations) as

$$f(\gamma(b)) - f(\gamma(a)) = \int_a^b \frac{\partial f}{\partial z}(\gamma(t)) \cdot \frac{d\gamma}{dt}(t)\, dt, \tag{2.1.6.1}$$

where, as earlier, we have taken $d\gamma/dt$ to be $d\gamma_1/dt + id\gamma_2/dt$.

This latter result plays much the same role for holomorphic functions as does the Fundamental Theorem of Calculus for functions from $\mathbb{R}$ to $\mathbb{R}$. The expression on the right of (2.1.6.1) is called the *complex line integral* and is denoted

$$\oint_\gamma \frac{\partial f}{\partial z}(z)\, dz\,. \tag{2.1.6.2}$$

More generally, if g is *any* continuous function whose domain contains the curve γ, then the complex line integral of g along γ is defined to be

$$\oint_\gamma g(z)\, dz \equiv \int_a^b g(\gamma(t)) \cdot \frac{d\gamma}{dt}(t)\, dt. \tag{2.1.6.3}$$

The main point here is that $\oint dz$ entails an expression of the form $\gamma'(t)\,dt$ in the integrand. Thus the trajectory and orientation of the curve will play a decisive role in the calculation, interpretation, and meaning of the complex line integral.

The whole concept of complex line integral is central to our further considerations in later sections. We shall use integrals like the one on the right of (2.1.6.3) even when f is not holomorphic; but we can be sure that the equality (2.1.6.1) holds only when f *is* holomorphic.

Note that when $\gamma(a) = \gamma(b) = A$ (and the domain U is simply connected) then the lefthand side of (2.1.6.1) is automatically equal to 0; and the right-hand side is simply the complex line integral of f around a closed curve. So we have a preview of the Cauchy integral theorem (see §§2.3.1) in this context.

2.1.7 Properties of Integrals

We conclude this section with some easy but useful facts about integrals.

(2.1.7.1) If $\varphi : [a, b] \to \mathbb{C}$ is continuous, then

$$\left| \int_a^b \varphi(t)\,dt \right| \leq \int_a^b |\varphi(t)|\,dt. \qquad (2.1.7.1.1)$$

(2.1.7.2) If $\gamma : [a, b] \to \mathbb{C}$ is a C^1 curve and φ is a continuous function on the curve γ, then

$$\left| \oint_\gamma \varphi(z)\,dz \right| \leq \left[\max_{t\in[a,b]} |\varphi(t)| \right] \cdot \ell(\gamma)\,, \qquad (2.1.7.2.1)$$

where

$$\ell(\gamma) \equiv \int_a^b |\varphi'(t)|\,dt$$

is the *length* of γ. Note that (2.1.7.2.1) follows from (2.1.7.1.1), and (2.1.7.1.1) is just calculus.

(2.1.7.3) The calculation of a complex line integral is independent of the way in which we parametrize the path:

Let $U \subseteq \mathbb{C}$ be an open set and $f : U \to \mathbb{C}$ a continuous function. Let $\gamma : [a, b] \to U$ be a C^1 curve. Suppose that $\varphi : [c, d] \to [a, b]$ is a one-to-one, onto, increasing C^1 function with a C^1 inverse. Let $\widetilde{\gamma} = \gamma \circ \varphi$. Then

$$\oint_{\widetilde{\gamma}} f\,dz = \oint_{\gamma} f\,dz. \tag{2.1.7.3.1}$$

The result follows from the change of variables formula in calculus.

This last statement implies that one can use the idea of the integral of a function f along a curve γ when the curve γ is described geometrically but without reference to a specific parametrization. For instance, "the integral of $\overline{z}$ *counterclockwise* around the unit circle $\{z \in \mathbb{C} : |z| = 1\}$" is now a phrase that makes sense, even though we have not indicated a specific parametrization of the unit circle. Note, however, that the direction counts: The integral of $\overline{z}$ counterclockwise around the unit circle is $2\pi i$. If the direction is reversed, then the integral changes sign: The integral of $\overline{z}$ *clockwise* around the unit circle is $-2\pi i$.

2.2 Complex Differentiability and Conformality

2.2.1 Limits

Until now we have developed a complex differential and integral calculus. We now unify the notions of partial derivative and total derivative in the complex context. For convenience, we shall repeat some ideas from §1.3.

2.2.2 Holomorphicity and the Complex Derivative

Let $U \subseteq \mathbb{C}$ be an open set and let f be holomorphic on U. Then f' exists at each point of U and

$$f'(z) = \frac{\partial f}{\partial z} \tag{2.2.4.1}$$

for all $z \in U$ (where $\partial f/\partial z$ is defined as in §§1.3.3). This is because we see that $\partial f/\partial x$ (according to the definition) coincides with df/dz and $\partial f/\partial y$

coincides with idf/dz. Hence

$$\frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right) = \frac{df}{dz}\,.$$

Note that, as a consequence, we can (and often will) write f' for $\partial f/\partial z$ when f is holomorphic. The following result is a converse:

If $f \in C^1(U)$ and f has a complex derivative f' at each point of U, then f is holomorphic on U. In particular, if a continuous, complex-valued function f on U has a complex derivative at each point and, if f' is continuous on U, then f is holomorphic on U. Such a function satisfies the Cauchy-Riemann equations (1.3.2.2).

It is perfectly logical to consider an f that possesses a complex derivative at each point of U *without* the additional assumption that $f \in C^1(U)$. It turns out that, under these circumstances, u and v still satisfy the Cauchy-Riemann equations. It is a deeper result, due to Goursat, that if f has a complex derivative at each point of U, then $f \in C^1(U)$ and hence f is holomorphic. See [GRK], especially the Appendix on Goursat's theorem, for details.

2.2.3 Conformality

Now we make some remarks about "conformality." Stated loosely, a function is *conformal* at a point $P \in \mathbb{C}$ if the function "preserves angles" at P and "stretches equally in all directions" at P. Holomorphic functions enjoy both properties. Now we shall discuss them in detail.

Let f be holomorphic in a neighborhood of $P \in \mathbb{C}$. Let w_1, w_2 be complex numbers of unit modulus. Consider the directional derivatives

$$D_{w_1} f(P) \equiv \lim_{t\to 0} \frac{f(P + tw_1) - f(P)}{t} \qquad (2.2.5.1)$$

and

$$D_{w_2} f(P) \equiv \lim_{t\to 0} \frac{f(P + tw_2) - f(P)}{t}. \qquad (2.2.5.2)$$

Then

(2.2.5.3) $\quad |D_{w_1} f(P)| = |D_{w_2} f(P)|\,.$

(2.2.5.4) If $|f'(P)| \neq 0$, then the directed angle from w_1 to w_2 equals the directed angle from $D_{w_1} f(P)$ to $D_{w_2} f(P)$.

In fact the last statement has an important converse: If **(2.2.5.4)** holds at P, then f has a complex derivative at P. If **(2.2.5.3)** holds at P, then either f or $\overline{f}$ has a complex derivative at P. Thus a function that is conformal (in either sense) at all points of an open set U must possess the complex derivative at each point of U. By the discussion in §§2.2.4, f is therefore holomorphic if it is C^1. Or, by Goursat's theorem, it would then follow that the function is holomorphic on U, with the C^1 condition being automatic.

Proof of Conformality: Notice that

$$\begin{aligned} D_{w_j} f(P) &= \lim_{t \to 0} \frac{f(P + tw_j) - f(P)}{tw_j} \cdot \frac{tw_j}{t} \\ &= f'(P) \cdot w_j \ , \quad j = 1, 2. \end{aligned}$$

The first assertion is now immediate and the second follows from the usual geometric interpretation of multiplication by a nonzero complex number, namely, that multiplication by $re^{i\theta}, r \neq 0$, multiplies lengths by r and rotates (around the origin) by the angle θ. □

The converse to this theorem asserts in effect that if *either* of statements (2.2.4.2) or (2.2.4.3) holds at P, then f has a complex derivative at P. Thus a C^1 function that is conformal (in either sense) at all points of an open set U must possess the complex derivative at each point of U. Of course then f is holomorphic if it is C^1. We leave these assertions to the reader.

It is worthwhile to consider the theorem expressed in terms of real functions. That is, we write $f = u + iv$, where u, v are real-valued functions. Also we consider $f(x + iy)$, and hence u and v, as functions of the real variables x and y. Thus f, as a function from an open subset of $\mathbb{C}$ into $\mathbb{C}$, can be regarded as a function from an open subset of $\mathbb{R}^2$ into $\mathbb{R}^2$. With f viewed in these real-variable terms, the first derivative behavior of f is described by its Jacobian matrix:

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

Recall that this matrix, evaluated at a point (x_0, y_0), is the matrix of the linear transformation that best approximates $f(x, y) - f(x_0, y_0)$ at (x_0, y_0). Now the Cauchy-Riemann equations for f mean exactly that this matrix has the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} .$$

Such a matrix is either the zero matrix or it can be written as the product of two matrices:

$$\begin{pmatrix} \sqrt{a^2 + b^2} & 0 \\ 0 & \sqrt{a^2 + b^2} \end{pmatrix} \cdot \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

for some choice of $\theta \in \mathbb{R}$. One chooses θ so that

$$\begin{aligned} \cos\theta &= \frac{a}{\sqrt{a^2 + b^2}} , \\ \sin\theta &= \frac{b}{\sqrt{a^2 + b^2}} . \end{aligned}$$

Such a choice of θ is possible because

$$\left(\frac{a}{\sqrt{a^2 + b^2}} \right)^2 + \left(\frac{b}{\sqrt{a^2 + b^2}} \right)^2 = 1.$$

Thus the Cauchy-Riemann equations imply that the (real) Jacobian of f has the form

$$\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \cdot \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

for some $\lambda \in \mathbb{R}, \lambda > 0$, and some $\theta \in \mathbb{R}$.

Geometrically, these two matrices have simple meanings. The matrix

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

is the representation of a rotation around the origin by the angle θ. The matrix

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

is multiplication of all vectors in $\mathbb{R}^2$ by λ. Therefore the product

$$\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \cdot \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

represents the same operation on $\mathbb{R}^2$ as does multiplication on $\mathbb{C}$ by the complex number $\lambda e^{i\theta}$.

Notice that, for our particular (Jacobian) matrix

$$\begin{pmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{pmatrix} ,$$

we have

$$\lambda = \sqrt{\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2} = |f'(z)|,$$

in agreement with the theorem.

2.3 The Cauchy Integral Formula and Theorem

2.3.1 The Cauchy Integral Theorem, Basic Form

If f is a holomorphic function on an open disc U in the complex plane, and if $\gamma : [a, b] \to U$ is a C^1 curve in U with $\gamma(a) = \gamma(b)$, then

$$\oint_\gamma f(z)\, dz = 0. \tag{2.3.1.1}$$

There are a number of different ways to prove the Cauchy integral theorem. One of the most natural is by way of a complex-analytic form of Stokes's theorem: If γ is a simple, closed curve surrounding a region U in the plane then

$$\oint_\gamma f(z)\, dz = \iint\limits_U \frac{\partial f}{\partial \overline{z}}\, d\overline{z} \wedge dz = \iint\limits_U 0 d\overline{z} \wedge dz = 0\,.$$

An important converse of Cauchy's theorem is called *Morera's theorem*:

Let f be a continuous function on a connected open set $U \subseteq \mathbb{C}$. If

$$\oint_\gamma f(z)\, dz = 0 \tag{2.3.1.2}$$

for every simple, closed curve γ in U, then f is holomorphic on U.

In the statement of Morera's theorem, the phrase "every simple, closed curve" may be replaced by "every triangle" or "every square" or "every circle." Morera's theorem may also be proved using Stokes's theorem (as above). We leave the details to the reader, or see [GRK].

2.3.2 The Cauchy Integral Formula

Suppose that U is an open set in $\mathbb{C}$ and that f is a holomorphic function on U. Let $z_0 \in U$ and let $r > 0$ be such that $\overline{D}(P, r) \subseteq U$. Let $\gamma : [0, 1] \to \mathbb{C}$ be the C^1 curve $\gamma(t) = P + r\cos(2\pi t) + ir\sin(2\pi t)$. Then, for each $z \in D(P, r)$,

$$f(z) = \frac{1}{2\pi i} \oint_\gamma \frac{f(\zeta)}{\zeta - z}\, d\zeta. \tag{2.3.2.1}$$

One may derive this result directly from Stokes's theorem (see [KRA5] and also our Subsection 2.3.1).

2.3.3 More General Forms of the Cauchy Theorems

Now we present the very useful general statements of the Cauchy integral theorem and formula. First we need a piece of terminology. A curve $\gamma : [a, b] \to \mathbb{C}$ is said to be *piecewise* C^k if

$$[a, b] = [a_0, a_1] \cup [a_1, a_2] \cup \cdots \cup [a_{m-1}, a_m] \tag{2.3.3.1}$$

with $a = a_0 < a_1 < \cdots a_m = b$ and $\gamma\big|_{[a_{j-1}, a_j]}$ is C^k for $1 \le j \le m$. In other words, γ is piecewise C^k if it consists of finitely many C^k curves chained end to end.

Cauchy Integral Theorem: Let $f : U \to \mathbb{C}$ be holomorphic with $U \subseteq \mathbb{C}$ an open set. Then

$$\oint_\gamma f(z)\, dz = 0 \tag{2.3.3.2}$$

for each piecewise C^1 closed curve γ in U that can be deformed in U through closed curves to a point in U—see Figure 2.3.

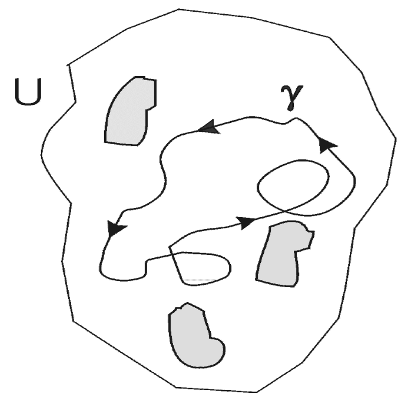

Figure 2.3: General form of the Cauchy theorem.

Cauchy Integral Formula: Suppose that $\overline{D}(z,r) \subseteq U$. Then

$$\frac{1}{2\pi i}\oint_\gamma \frac{f(\zeta)}{\zeta - z}\, d\zeta = f(z) \tag{2.3.3.3}$$

for any piecewise C^1 closed curve γ in $U \setminus \{z\}$ that can be continuously deformed in $U \setminus \{z\}$ to $\partial D(z,r)$ equipped with counterclockwise orientation. Refer to Figure 2.4. Of course one derives this more general version of Cauchy's formula with the standard device of deformation of curves.

A topological notion that is special to complex analysis is simple connectivity. We say that a domain $U \subseteq \mathbb{C}$ is *simply connected* if any closed curve in U can be continuously deformed to a point. Simple connectivity is a mathematically rigorous condition that corresponds to the intuitive notion that the region U has no holes. If U is simply connected, and γ is a closed curve in U, then it follows that γ can be continuously deformed to lie inside a disc in U. It follows that Cauchy's theorem applies to γ. To summarize: on a simply connected region, Cauchy's theorem applies (without any further hypotheses) to any closed curve in γ. Likewise, in a simply connected U, Cauchy's integral formula applied to any simple, closed curve that is oriented counterclockwise and to any point z that is inside that curve.

2.3.4 Deformability of Curves

A central fact about the complex line integral is the deformability of curves. Let $\gamma : [a,b] \to U$ be a piecewise C^1 curve in a region U of the complex

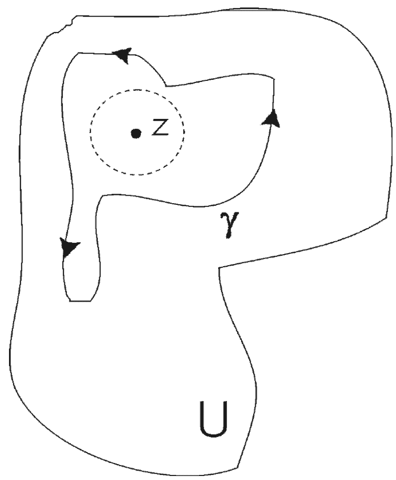

Figure 2.4: General form of the Cauchy formula.

plane. Let f be a holomorphic function on U. The value of the complex line integral

$$\oint_\gamma f(z)\,dz \qquad (2.3.4.1)$$

does not change if the curve γ is smoothly deformed within the region U. Note that, in order for this statement to be valid, the curve γ must remain inside the region of holomorphicity U of f while it is being deformed, and it must remain a closed curve while it is being deformed. Figure 2.5 shows curves γ_1, γ_2 that *can* be deformed to one another, and a curve γ_3 that can be deformed to neither of the first two (because of the hole inside γ_3).

2.4 A Coda on the Limitations of The Cauchy Integral Formula

If f is any continuous function on the boundary of the unit disc $D = D(0,1)$, then the Cauchy integral

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

defines a holomorphic function $F(z)$ on D (use Morera's theorem, for example, to confirm this assertion). What does the new function F have to do

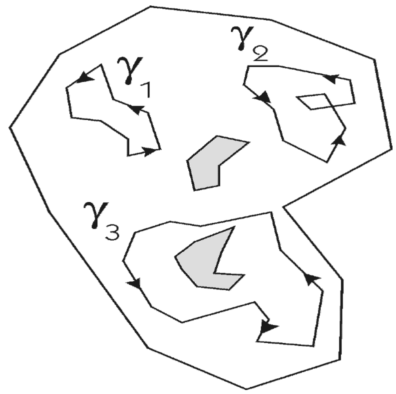

Figure 2.5: Deformation of curves.

with the original function f? In general, not much.

For example, if $f(\zeta) = \overline{\zeta}$, then $F(z) \equiv 0$ (exercise). In no sense is the original function f any kind of "boundary limit" of the new function F. The question of which functions f are "natural boundary functions" for holomorphic functions F (in the sense that F is a continuous extension of F to the closed disc) is rather subtle. Its answer is well understood, but is best formulated in terms of Fourier series and the so-called Hilbert transform. The complete story is given in [KRA1]. See also [GAR] for a discussion of the F. and M. Riesz theorem.

Contrast this situation for holomorphic function with the much more succinct and clean situation for harmonic functions (§7.3).

Chapter 3

Applications of the Cauchy Theory

3.1 The Derivatives of a Holomorphic Function

3.1.1 A Formula for the Derivative

Let $U \subseteq \mathbb{C}$ be an open set and let f be holomorphic on U. Then $f \in C^\infty(U)$. Moreover, if $\overline{D}(P,r) \subseteq U$ and $z \in D(P,r)$, then

$$\left(\frac{\partial}{\partial z}\right)^k f(z) = \frac{k!}{2\pi i}\oint_{|\zeta-P|=r} \frac{f(\zeta)}{(\zeta-z)^{k+1}}\, d\zeta, \quad k = 0,1,2,\ldots. \tag{3.1.1.1}$$

This formula is obtained simply by differentiating the standard Cauchy formula (2.3.2.1) under the integral sign.

3.1.2 The Cauchy Estimates

If f is a holomorphic on a region containing the closed disc $\overline{D}(P,r)$ and if $|f| \leq M$ on $\overline{D}(P,r)$, then

$$\left|\frac{\partial^k}{\partial z^k} f(P)\right| \leq \frac{M \cdot k!}{r^k}. \tag{3.1.2.1}$$

This is proved by direct estimation of the Cauchy formula (3.1.1.1).

3.1.3 Entire Functions and Liouville's Theorem

A function f is said to be *entire* if it is defined and holomorphic on all of $\mathbb{C}$, i.e., $f : \mathbb{C} \to \mathbb{C}$ is holomorphic. For instance, any holomorphic polynomial is entire, e^z is entire, and $\sin z, \cos z$ are entire. The function $f(z) = 1/z$ is not entire because it is undefined at $z = 0$. [In a sense that we shall make precise later (§4.1, ff.), this last function has a "singularity" at 0.] The question we wish to consider is: "Which entire functions are bounded?" This question has a very elegant and complete answer as follows:

Liouville's Theorem A bounded entire function is constant.

Proof: Let f be entire and assume that $|f(z)| \leq M$ for all $z \in \mathbb{C}$. Fix a $P \in \mathbb{C}$ and let $r > 0$. We apply the Cauchy estimate (3.1.2.1) for $k = 1$ on $\overline{D}(P, r)$. So

$$\left|\frac{\partial}{\partial z} f(P)\right| \leq \frac{M \cdot 1!}{r}. \tag{3.1.3.1}$$

Since this inequality is true for every $r > 0$, we conclude that

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

Since P was arbitrary, we conclude that

$$\frac{\partial f}{\partial z} \equiv 0.$$

Therefore f is constant. □

The end of the last proof bears some commentary. We prove that $\partial f/\partial z \equiv 0$. But we know, since f is holomorphic, that $\partial f/\partial \overline{z} \equiv 0$. It follows from linear algebra that $\partial f/\partial x \equiv 0$ and $\partial f/\partial y \equiv 0$. Then calculus tells us that f is constant.

The reasoning that establishes Liouville's theorem can also be used to prove this more general fact: If $f : \mathbb{C} \to \mathbb{C}$ is an entire function and if for some real number C and some positive integer k, it holds that

$$|f(z)| \leq C \cdot (1 + |z|)^k$$

for all z, then f is a polynomial in z of degree at most k.

3.1.4 The Fundamental Theorem of Algebra

One of the most elegant applications of Liouville's Theorem is a proof of what is known as the Fundamental Theorem of Algebra (see also §§1.1.7):

> **The Fundamental Theorem of Algebra:** Let $p(z)$ be a non-constant (holomorphic) polynomial. Then p has a root. That is, there exists an $\alpha \in \mathbb{C}$ such that $p(\alpha) = 0$.

Proof: Suppose not. Then $g(z) = 1/p(z)$ is entire. Also when $|z| \to \infty$, then $|p(z)| \to +\infty$. Thus $1/|p(z)| \to 0$ as $|z| \to \infty$; hence g is bounded. By Liouville's Theorem, g is constant, hence p is constant. Contradiction.

□

The polynomial p has degree $k \geq 1$, then let α_1 denote the root provided by the Fundamental Theorem. By the Euclidean algorithm (see [HUN]), we may divide $z - \alpha_1$ into p with no remainder to obtain

$$p(z) = (z - \alpha_1) \cdot p_1(z). \tag{3.1.4.1}$$

Here p_1 is a polynomial of degree $k - 1$. If $k - 1 \geq 1$, then, by the theorem, p_1 has a root α_2 . Thus p_1 is divisible by $(z - \alpha_2)$ and we have

$$p(z) = (z - \alpha_1) \cdot (z - \alpha_2) \cdot p_2(z) \tag{3.1.4.2}$$

for some polynomial $p_2(z)$ of degree $k - 2$. This process can be continued until we arrive at a polynomial p_k of degree 0; that is, p_k is constant. We have derived the following fact: If $p(z)$ is a holomorphic polynomial of degree k, then there are k complex numbers $\alpha_1, \ldots \alpha_k$ (not necessarily distinct) and a non-zero constant C such that

$$p(z) = C \cdot (z - \alpha_1) \cdots (z - \alpha_k). \tag{3.1.4.3}$$

If some of the roots of p coincide, then we say that p has *multiple roots*. To be specific, if m of the values $\alpha_{j_1}, \ldots, \alpha_{j_m}$ are equal to some complex number α, then we say that p has a root of order m at α (or that p has a root of *multiplicity* m at α). It is an easily verified fact that the polynomial p has a root of order m at α if $p(\alpha) = 0$, $p'(\alpha) = 0$, $\ldots p^{(m-1)}(\alpha) = 0$ (where the parenthetical exponent denotes a derivative).

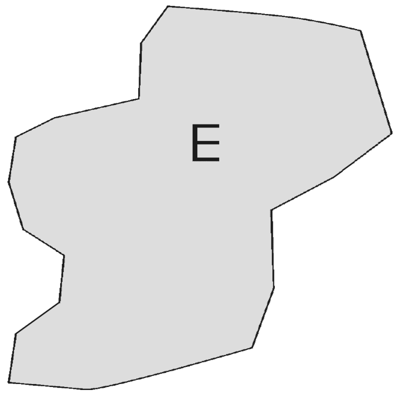

Figure 3.1: A compact set.

An example will make the idea clear: Let

$$p(z) = (z-5)^3 \cdot (z+2)^8 \cdot (z-3i) \cdot (z+6).$$

Then we say that p has a root of order 3 at 5, a root of order 8 at -2, and it has roots of order 1 at 3i and at -6. We also say that p has *simple roots* at 1 and -6.

3.1.5 Sequences of Holomorphic Functions and their Derivatives

A sequence of functions g_j defined on a common domain E is said to *converge uniformly* to a limit function g if, for each $\epsilon > 0$, there is a number $N > 0$ such that for all $j > N$ it holds that $|g_j(x) - g(x)| < \epsilon$ for every $x \in E$. The key point is that the degree of closeness of $g_j(x)$ to $g(x)$ is independent of $x \in E$.

Let $f_j : U \to \mathbb{C}$, $j = 1, 2, 3 \ldots$, be a sequence of holomorphic functions on an open set U in $\mathbb{C}$. Suppose that there is a function $f : U \to \mathbb{C}$ such that, for each compact subset E (a compact set is one that is closed and bounded—see Figure 3.1) of U, the restricted sequence $f_j|_E$ converges uniformly to $f|_E$. Then f is holomorphic on U. [In particular, $f \in C^\infty(U)$.]

If f_j, f, U are as in the preceding paragraph, then, for any $k \in \{0, 1, 2, \ldots\}$, we have

$$\left(\frac{\partial}{\partial z}\right)^k f_j(z) \to \left(\frac{\partial}{\partial z}\right)^k f(z) \qquad (3.1.5.1)$$

uniformly on compact sets.[1] The proof is immediate from (3.1.1.1), which we derived from the Cauchy integral formula, for the derivative of a holomorphic function.

3.1.6 The Power Series Representation of a Holomorphic Function

The ideas being considered in this section can be used to develop our understanding of power series. A *power series*

$$\sum_{j=0}^{\infty} a_j (z-P)^j \tag{3.1.6.1}$$

is defined to be the limit of its *partial sums*

$$S_N(z) = \sum_{j=0}^{N} a_j (z-P)^j. \tag{3.1.6.2}$$

We say that the partial sums *converge* to the sum of the entire series.

Any given power series has a *disc of convergence*. More precisely, let

$$r = \frac{1}{\limsup_{j\to\infty} |a_j|^{1/j}}. \tag{3.1.6.3}$$

The power series (3.1.6.2) will then certainly converge on the disc $D(P, r)$; the convergence will be absolute and uniform on any disc $\overline{D}(P, r')$ with $r' < r$.

For clarity, we should point out that in many examples the sequence $|a_j|^{1/j}$ actually converges as $j \to \infty$. Then we may take r to be equal to $1/\lim_{j\to\infty} |a_j|^{1/j}$. The reader should be aware, however, that in case the sequence $\{|a_j|^{1/j}\}$ does not converge, then one must use the more formal definition (3.1.6.3) of r. See [KRA3], [RUD1].

Of course the partial sums, being polynomials, are holomorphic on *any* disc $D(P, r)$. If the disc of convergence of the power series is $D(P, r)$, then let f denote the function to which the power series converges. Then for any $0 < r' < r$ we have that

$$S_N(z) \to f(z),$$

[1]It is also common to say that the functions converge *normally*.

uniformly on $\overline{D}(P, r')$. We can conclude immediately that $f(z)$ is holomorphic on $D(P, r)$. Moreover, we know that

$$\left(\frac{\partial}{\partial z}\right)^k S_N(z) \to \left(\frac{\partial}{\partial z}\right)^k f(z). \tag{3.1.6.4}$$

This shows that a differentiated power series has a disc of convergence at least as large as the disc of convergence (with the same center) of the original series, and that the differentiated power series converges on that disc to the derivative of the sum of the original series.

The most important fact about power series for complex function theory is this: If f is a holomorphic function on a domain $U \subseteq \mathbb{C}$, if $P \in U$, and if the disc $D(P, r)$ lies in U, then f may be represented as a convergent power series on $D(P, r)$. Explicitly, we have

$$f(z) = \sum_{j=0}^{\infty} a_j (z - P)^j, \tag{3.1.6.5}$$

where

$$a_j = \frac{f^{(j)}(P)}{j!}. \tag{3.1.6.6}$$

[Here the exponent $^{(j)}$ on f denotes the j^{th} derivative.] The provenance of this formula will be explained below. Thus we have an explicit way of calculating the power series expansion of any holomorphic function f about a point P of its domain, and we have an a priori knowledge of the disc on which the power series representation will converge.

The matter bears further consideration. We know that every smooth function $f(x)$ of a real variable has a Taylor series expansion about any point p in the interior of its domain. But it is a fact that this Taylor expansion generically does *not* converge; even when it does converge, it generically does *not* converge back to f. The situation for a holomorphic function of a complex variable is markedly different: in that circumstance, the Taylor or power series expansion *always* converges. The proof is simplicity itself. Take the center of the disc in the Cauchy formula to be the origin 0. We write the

Cauchy formula as

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\oint_{\partial D(0,r)} \frac{f(\zeta)}{\zeta - z}\, d\zeta \\ &= \oint_{\partial D(0,r)} f(\zeta)\cdot\left[\frac{1}{\zeta}\cdot\frac{1}{1-z/\zeta}\right] d\zeta \\ &= \oint_{\partial D(0,r)} f(\zeta)\cdot\left[\frac{1}{\zeta}\cdot\sum_{j=0}^{\infty}(z/\zeta)^j\right] d\zeta \\ &= \oint_{\partial D(0,r)} f(\zeta)\cdot\frac{1}{\zeta^{j+1}}\, d\zeta\cdot z^j \\ &= \sum_j a_j\cdot z^j\,. \end{aligned}$$

We see explicitly that

$$a_j = \oint_{\partial D(0,r)} f(\zeta)\cdot\frac{1}{\zeta^{j+1}}\, d\zeta\,,$$

and this corresponds, by the Cauchy formula, to derivatives of f.

Of course the series converges absolutely and uniformly for $|z| < r = |\zeta|$. The key point here is that holomorphic functions are analytic because the Cauchy kernel is analytic. We know from our formula for the derivatives of a holomorphic function that the j^{th} coefficent of the power series in the last expansion is $f^{(j)}(0)/j!$.

3.2 The Zeros of a Holomorphic Function

3.2.1 The Zero Set of a Holomorphic Function

Let f be a holomorphic function. If f is not identically zero, then it turns out that f cannot vanish at too many points. This once again bears out the dictum that holomorphic functions are a lot like polynomials. To give this concept a precise formulation, we need to recall the topological notion of connectedness (§1.1.5).

3.2.2 Discreteness of the Zeros of a Holomorphic Function

> Let $U \subseteq \mathbb{C}$ be a connected (§§1.1.5) open set and let $f : U \to \mathbb{C}$ be holomorphic. Let the zero set of f be $\mathcal{Z} = \{z \in U : f(z) = 0\}$. If there are a $z_0 \in \mathcal{Z}$ and $\{z_j\}_{j=1}^{\infty} \subseteq \mathcal{Z} \setminus \{z_0\}$ such that $z_j \to z_0$, then $f \equiv 0$.

Let us formulate the result in topological terms. We recall that a point z_0 is said to be an *accumulation point* of a set $\mathcal{Z}$ if there is a sequence $\{z_j\} \subseteq \mathcal{Z} \setminus \{z_0\}$ with $\lim_{j\to\infty} z_j = z_0$. Then the theorem is equivalent to the statement: If $f : U \to \mathbb{C}$ is a holomorphic function on a connected (§§1.1.5) open set U and if $\mathcal{Z} = \{z \in U : f(z) = 0\}$ has an accumulation point *in* U, then $f \equiv 0$.

For the proof, suppose that the point 0 is an interior accumulation point of zeros $\{z_j\}$ of the holomorphic function f. Thus $f(0) = 0$. We may write $f(z) = z \cdot f^*(z)$. But f^* vanishes at $\{z_j\}$ and 0 is still an accumulation point of $\{z_j\}$. It follows that $f^*(0) = 0$. Hence f itself has a zero of order 2 at 0. Continuing in this fashion, we see that f has a zero of infinite order at 0. So the power series expansion of f about 0 is identically 0. It then follows from an easy connectedness argument (more on this below) that $f \equiv 0$.

3.2.3 Discrete Sets and Zero Sets

There is still more terminology concerning the zero set of a holomorphic function in §§3.2.1. A set S is said to be *discrete* if for each $s \in S$ there is an $\epsilon > 0$ such that $D(s, \epsilon) \cap S = \{s\}$. See Figure 3.2. People also say, in a slight abuse of language, that a discrete set has points that are "isolated" or that S contains only "isolated points." The result in §§3.2.2 thus asserts that if f is a non-constant holomorphic function on a connected open set, then its zero set is discrete or, less formally, the zeros of f are isolated. It is important to realize that the result in §§3.2.2 does *not* rule out the possibility that the zero set of f can have accumulation points in $\mathbb{C} \setminus U$; in particular, a non-constant holomorphic function on an open set U can indeed have zeros accumulating at a point of ∂U. Consider, for instance, the function $f(z) = \sin(1/[1-z])$ on the unit disc. The zeros of this f include $\{1 - 1/[j\pi]\}$, and these accumulate at the boundary point 1. See Figure 3.3.

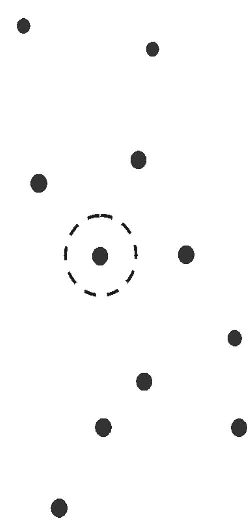

Figure 3.2: A discrete set.

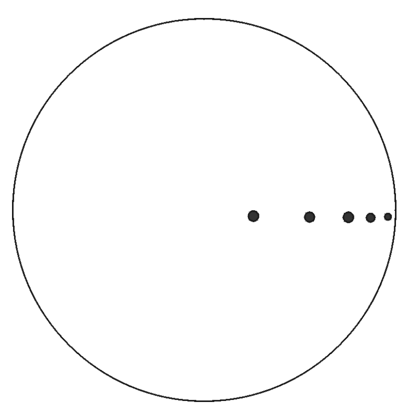

Figure 3.3: Zeros accumulating at a boundary point.

3.2.4 Uniqueness of Analytic Continuation

A consequence of the preceding basic fact (§§3.2.2) about the zeros of a holomorphic function is this: Let $U \subseteq \mathbb{C}$ be a connected open set and $D(P,r) \subseteq U$. If f is holomorphic on U and $f\big|_{D(P,r)} \equiv 0$, then $f \equiv 0$ on U. In fact if $f \equiv 0$ on a *segment* then it must follows that $f \equiv 0$.

Here are some further corollaries:

(3.2.4.1) Let $U \subseteq \mathbb{C}$ be a connected open set. Let f, g be holomorphic on U. If $\{z \in U : f(z) = g(z)\}$ has an accumulation point in U, then $f \equiv g$.

(3.2.4.2) Let $U \subseteq \mathbb{C}$ be a connected open set and let f, g be holomorphic on U. If $f \cdot g \equiv 0$ on U, then either $f \equiv 0$ on U or $g \equiv 0$ on U.

(3.2.4.3) Let $U \subseteq \mathbb{C}$ be connected and open and let f be holomorphic on U. If there is a $P \in U$ such that

$$\left(\frac{\partial}{\partial z}\right)^j f(P) = 0 \qquad (3.2.4.3.1)$$

for every $j \in \{0, 1, 2, \dots\}$, then $f \equiv 0$.

(3.2.4.4) If f and g are entire holomorphic functions and if $f(x) = g(x)$ for all $x \in \mathbb{R} \subseteq \mathbb{C}$, then $f \equiv g$. It also holds that functional identities that are true for all real values of the variable are also true for complex values of the variable (Figure 3.4). For instance,

$$\sin^2 z + \cos^2 z = 1 \qquad \text{for all} \quad z \in \mathbb{C} \qquad (3.2.4.4.1)$$

because the identity is true for all $z = x \in \mathbb{R}$. This is an instance of the "principle of persistence of functional relations"—see [GRK].

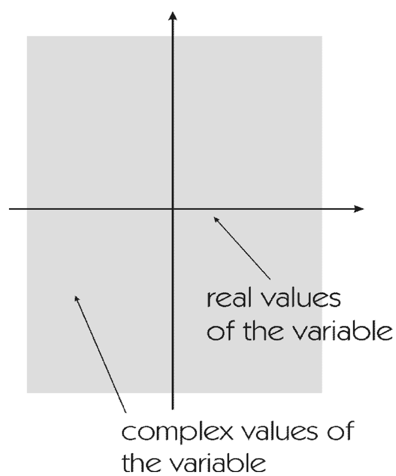

Figure 3.4: Principle of persistence of functional relations.

Chapter 4

Isolated Singularities and Laurent Series

4.1 The Behavior of a Holomorphic Function near an Isolated Singularity

4.1.1 Isolated Singularities

It is often important to consider a function that is holomorphic on a punctured open set $U \setminus \{P\} \subset \mathbb{C}$. Refer to Figure 4.1.

In this chapter we shall obtain a new kind of infinite series expansion which generalizes the idea of the power series expansion of a holomorphic function about a (nonsingular) point—see §§3.1.6. We shall in the process completely classify the behavior of holomorphic functions near an isolated singular point (§§4.1.3).

4.1.2 A Holomorphic Function on a Punctured Domain

Let $U \subseteq \mathbb{C}$ be an open set and $P \in U$. We call the domain $U \setminus \{P\}$ a *punctured domain*. Suppose that $f : U \setminus \{P\} \to \mathbb{C}$ is holomorphic. In this situation we say that f has an *isolated singular point* (or *isolated singularity*) at P. The implication of the phrase is usually just that f is defined and holomorphic on some such "deleted neighborhood" of P. The specification of the set U is of secondary interest; we wish to consider the behavior of f

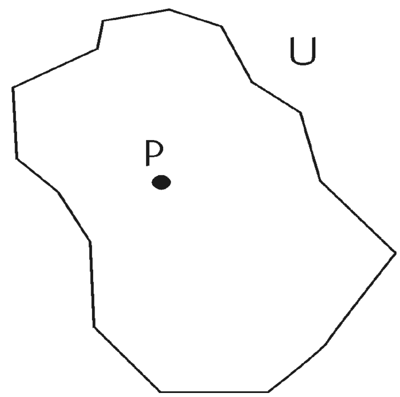

Figure 4.1: An isolated singularity.

"near P."

4.1.3 Classification of Singularities

There are three possibilities for the behavior of f near P that are worth distinguishing:

(4.1.3.1) $|f(z)|$ is bounded on $D(P,r) \setminus \{P\}$ for some $r > 0$ with $D(P,r) \subseteq U$; i.e., there is some $r > 0$ and some $M > 0$ such that $|f(z)| \leq M$ for all $z \in U \cap D(P,r) \setminus \{P\}$.

(4.1.3.2) $\lim_{z \to P} |f(z)| = +\infty$.

(4.1.3.3) Neither **(i)** nor **(ii)**.

Of course elementary logic tells us that these three conditions cover all possibilities. The description of **(4.1.3.3)** is not very satisfying, but it turns out that that is the most subtle situation; there is no simple description of what goes on there. We shall say more about each of these three conditions in the ensuing discussion.

4.1.4 Removable Singularities, Poles, and Essential Singularities

We shall see momentarily that, if case **(4.1.3.1)** holds, then f has a limit at P that extends f so that it is holomorphic on all of U (this is not at all obvious; it is a theorem of Riemann). It is commonly said in this circumstance that f has a *removable singularity* at P. In case **(4.1.3.2)**, we will say that f has a *pole* at P. In case **(4.1.3.3)**, f will be said to have an *essential singularity* at P. Our goal in this and the next two subsections is to understand **(4.1.3.1)**–**(4.1.3.3)** in some further detail.

4.1.5 The Riemann Removable Singularities Theorem

Let $f : D(P, r) \setminus \{P\} \to \mathbb{C}$ be holomorphic and bounded. Then

(4.1.5.1) $\lim_{z \to P} f(z)$ exists.

(4.1.5.2) The function $\widehat{f} : D(P, r) \to \mathbb{C}$ defined by

$$\widehat{f}(z) = \begin{cases} f(z) & \text{if} \quad z \neq P \\ \lim\limits_{\zeta \to P} f(\zeta) & \text{if} \quad z = P \end{cases}$$

is holomorphic.

For the proof, take $P = 0$ and consider the function $g(z) = z^2 \cdot f(z)$. One may verify directly that g is C^1 and satisfies the Cauchy-Riemann equations on all of $D(P, r)$ (the boundedness hypothesis is used to check that both g and its first derivative have limits at 0). Thus g is holomorphic on the disc, and it vanishes to second order at 0. It follows then that $f(z) = g(z)/z^2$ is a *bona fide* holomorphic function on all of $D(P, r)$.

4.1.6 The Casorati-Weierstrass Theorem

> If $f : D(P, r_0) \setminus \{P\} \to \mathbb{C}$ is holomorphic and P is an essential singularity of f, then $f(D(P, r) \setminus \{P\})$ is dense in $\mathbb{C}$ for any $0 < r < r_0$.

For the reason, suppose that the assertion is not true. So there is a complex value μ and a positive number ϵ so that the image of $D(P, r) \setminus \{P\}$

under f does *not* contain the disc $\overline{D}(\mu, \epsilon)$. But then the function $g(z) = 1/[f(z) - \mu]$ is bounded and non-vanishing near P, hence has a removable singularity. We see then that f is bounded near P, and that contradicts that P is an essential singularity.

Now we have seen that, at a removable singularity P, a holomorphic function f on $D(P, r_0) \setminus \{P\}$ can be continued to be holomorphic on all of $D(P, r_0)$. And, near an essential singularity at P, a holomorphic function g on $D(P, r_0) \setminus \{P\}$ has image that is dense in $\mathbb{C}$. The third possibility, that h has a *pole* at P, has yet to be described. This case will be examined further in the coming sections.

We next develop a new type of doubly infinite series that will serve as a tool for understanding isolated singularities—especially poles.

4.2 Expansion around Singular Points

4.2.1 Laurent Series

A *Laurent series* on $D(P, r)$ is a (formal) expression of the form

$$\sum_{j=-\infty}^{+\infty} a_j (z - P)^j. \tag{4.2.1.1}$$

Note that the individual terms are each defined for all $z \in D(P, r) \setminus \{P\}$. The series sums from $j = -\infty$ to $j = +\infty$.

4.2.2 Convergence of a Doubly Infinite Series

To discuss convergence of Laurent series, we must first make a general agreement as to the meaning of the convergence of a "doubly infinite" series $\sum_{j=-\infty}^{+\infty} \alpha_j$. We say that such a series *converges* if $\sum_{j=0}^{+\infty} \alpha_j$ and $\sum_{j=1}^{+\infty} \alpha_{-j} = \sum_{j=-\infty}^{-1} \alpha_j$ converge in the usual sense. In this case, we set

$$\sum_{-\infty}^{+\infty} \alpha_j = \left(\sum_{j=0}^{+\infty} \alpha_j \right) + \left(\sum_{j=1}^{+\infty} \alpha_{-j} \right). \tag{4.2.2.1}$$

In other words, the question of convergence for a bi-infinite series devolves to two separate questions about two sub-series.

We can now present the analogues for Laurent series of our basic results for power series.

4.2.3 Annulus of Convergence

The set of convergence of a Laurent series is either an open set of the form $\{z : 0 \le r_1 < |z-P| < r_2\}$, together with perhaps some or all of the boundary points of the set, *or* a set of the form $\{z : 0 \le r_1 < |z-P| < +\infty\}$, together with perhaps some or all of the boundary points of the set. Such an open set is called an (generalized) *annulus* centered at P. We shall let

$$D(P, +\infty) = \{z : |z-P| < +\infty\} = \mathbb{C}, \tag{4.2.3.1}$$

$$D(P, 0) = \{z : |z-P| < 0\} = \emptyset, \tag{4.2.3.2}$$

and

$$\overline{D}(P, 0) = \{P\}. \tag{4.2.3.3}$$

As a result, using this extended notation, all (open) annuli (plural of "annulus") can be written in the form

$$D(P, r_2) \setminus \overline{D}(P, r_1)\,, \quad 0 \le r_1 \le r_2 \le +\infty. \tag{4.2.3.4}$$

In precise terms, the "domain of convergence" of a Laurent series is given as follows:

Let

$$\sum_{j=-\infty}^{+\infty} a_j (z-P)^j \tag{4.2.3.5}$$

be a doubly infinite series. There are unique nonnegative extended real numbers r_1 and r_2 (r_1 or r_2 may be 0 or $+\infty$) such that the series converges absolutely for all z with $r_1 < |z-P| < r_2$ and diverges for z with $|z-P| < r_1$ or $|z-P| > r_2$ (see (4.2.3.4)). Also, if $r_1 < s_1 \le s_2 < r_2$, then $\sum_{j=-\infty}^{+\infty} |a_j(z-P)^j|$ converges uniformly on $\{z : s_1 \le |z-P| \le s_2\}$ and, consequently, $\sum_{j=-\infty}^{+\infty} a_j(z-P)^j$ converges absolutely and uniformly there.

The reason that the domain of convergence takes this form is that we may rewrite the series (4.2.3.5) as

$$\sum_{j=0}^{\infty} a_j(z-P)^j + \sum_{j=1}^{\infty} a_{-j}\left[(z-P)^{-1}\right]^j .$$

From what we know about power series, the domain of convergence of the first of these two series will have the form $|z-P| < r_2$ and the domain of convergence of the second series will have the form $|(z-P)^{-1}| < 1/r_1$. Putting these two conditions together gives $r_1 < |z-P| < r_2$.

4.2.4 Uniqueness of the Laurent Expansion

Let $0 \leq r_1 < r_2 \leq \infty$. If the Laurent series $\sum_{j=-\infty}^{+\infty} a_j (z-P)^j$ converges on $D(P, r_2) \setminus \overline{D}(P, r_1)$ to a function f, then, for any r satisfying $r_1 < r < r_2$, and each $j \in \mathbb{Z}$,

$$a_j = \frac{1}{2\pi i} \oint_{|\zeta - P| = r} \frac{f(\zeta)}{(\zeta - P)^{j+1}}\, d\zeta. \tag{4.2.4.1}$$

This claim follows from integrating the series term-by-term (most of the terms integrate to zero of course). In particular, the a_j's are uniquely determined by f.

We turn now to establishing that convergent Laurent expansions of functions holomorphic on an annulus do in fact exist.

4.2.5 The Cauchy Integral Formula for an Annulus

Suppose that $0 \leq r_1 < r_2 \leq +\infty$ and that $f : D(P, r_2) \setminus \overline{D}(P, r_1) \to \mathbb{C}$ is holomorphic. Then, for each s_1, s_2 such that $r_1 < s_1 < s_2 < r_2$ and each $z \in D(P, s_2) \setminus \overline{D}(P, s_1)$, it holds that

$$f(z) = \frac{1}{2\pi i} \oint_{|\zeta - P| = s_2} \frac{f(\zeta)}{\zeta - z}\, d\zeta - \frac{1}{2\pi i} \oint_{|\zeta - P| = s_1} \frac{f(\zeta)}{\zeta - z}\, d\zeta. \tag{4.2.5.1}$$

Figure 4.2 shows why this is true. The integral along the two segments (which actually coincide, but with opposite orientations) vanishes. What is left is the integrals along the two circles—with opposite orientations, as indicated in (4.2.5.1).

4.2.6 Existence of Laurent Expansions

Now we may summarize with our main result:

> **Theorem:** If $0 \leq r_1 < r_2 \leq \infty$ and $f : D(P, r_2) \setminus \overline{D}(P, r_1) \to \mathbb{C}$ is holomorphic, then there exist complex numbers a_j such that
>
> $$\sum_{j=-\infty}^{+\infty} a_j (z-P)^j \tag{4.2.6.1}$$

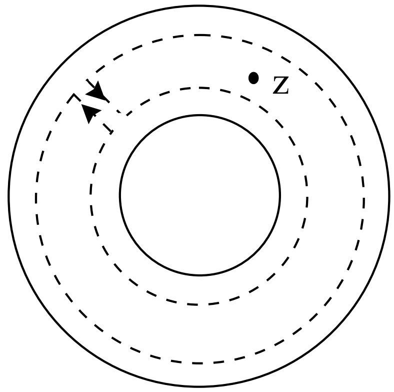

Figure 4.2: The Cauchy integral on an annulus.

converges on $D(P, r_2) \setminus \overline{D}(P, r_1)$ to f. If $r_1 < s_1 < s_2 < r_2$, then the series converges absolutely and uniformly on $\overline{D}(P, s_2) \setminus D(P, s_1)$. this below) that $f \equiv 0$. ,

The series expansion is independent of s_1 and s_2. In fact, for each fixed $j = 0, \pm 1, \pm 2, \ldots$, the value of

$$a_j = \frac{1}{2\pi i} \oint_{|\zeta - P| = r} \frac{f(\zeta)}{(\zeta - P)^{j+1}} \, d\zeta \tag{4.2.6.2}$$

is independent of r provided that $r_1 < r < r_2$.

4.2.7 Holomorphic Functions with Isolated Singularities

Now let us specialize what we have learned about Laurent series expansions to the case of $f : D(P, r) \setminus \{P\} \to \mathbb{C}$ holomorphic, that is, to a holomorphic function with an isolated singularity:

If $f : D(P, r) \setminus \{P\} \to \mathbb{C}$ is holomorphic, then f has a unique Laurent series expansion

$$f(z) = \sum_{j=-\infty}^{\infty} a_j (z - P)^j \tag{4.2.7.1}$$

that converges absolutely for $z \in D(P, r) \setminus \{P\}$. The convergence is uniform on compact subsets of $D(P, r) \setminus \{P\}$. The coefficients are given by

$$a_j = \frac{1}{2\pi i} \oint_{\partial D(P,s)} \frac{f(\zeta)}{(\zeta - P)^{j+1}} \, d\zeta, \quad \text{any} \quad 0 < s < r. \tag{4.2.7.2}$$

4.2.8 Classification of Singularities in Terms of Laurent Series

There are three mutually exclusive possibilities for the Laurent series

$$\sum_{j=-\infty}^{\infty} a_j (z - P)^j$$

about an isolated singularity P:

(4.2.8.1) $a_j = 0$ for all $j < 0$.

(4.2.8.2) For some $k \geq 1, a_j = 0$ for all $-\infty < j < -k$, but $a_k \neq 0$.

(4.2.8.3) Neither **(i)** nor **(ii)** applies.

These three cases correspond exactly to the three types of isolated singularities that we discussed in §4.1.3: case **(4.2.8.1)** occurs if and only if P is a removable singularity; case **(4.2.8.2)** occurs if and only if P is a pole (of order k, meaning that the term a_{-k} in the Laurent expansion in nonzero—more on this below); and case **(4.2.8.3)** occurs if and only if P is an essential singularity.

To put this matter in other words: In case **(4.2.8.1)**, we have a power series that converges, of course, to a holomorphic function. In case **(4.2.8.2)**, our Laurent series has the form

$$\sum_{j=-k}^{\infty} a_j (z - P)^j = (z - P)^{-k} \sum_{j=-k}^{\infty} a_j (z - P)^{j+k} = (z - P)^{-k} \sum_{j=0}^{\infty} a_{j-k} (z - P)^j.$$

Since $a_{-k} \neq 0$, we see that, for z near P, the function defined by the series behaves like $a_{-k} \cdot (z - P)^{-k}$ near P. In short, the function (in absolute value) blows up like $|z - P|^{-k}$ as $z \to P$. A graph in $(|z|, |f(z)|)$-space would exhibit a "pole"-like singularity. This is the source of the terminology "pole." See Figure 4.3. Case **(4.2.8.3)**, corresponding to an essential singularity, is much more complicated; in this case there are infinitely many negative terms in

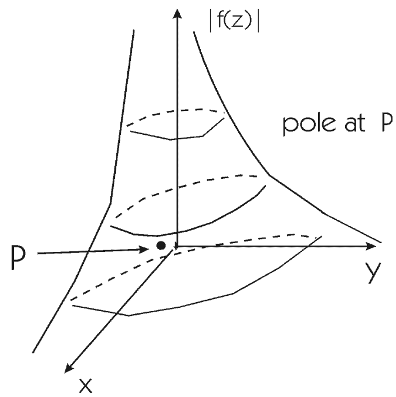

Figure 4.3: A pole.

the Laurent expansion and, by Casorati-Weierstrass (§§4.1.6), they interact in a complicated fashion.

Picard's Great Theorem (§§9.5.2) will tell us more about the behavior of a holomorphic function near an essential singularity.

4.3 Examples of Laurent Expansions

4.3.1 Principal Part of a Function

When f has a pole or essential singularity at P, it is customary to call the negative power part of the Laurent expansion of f around P the *principal part* of f at P. (Occasionally we shall also use the terminology "Laurent polynomial.") That is, if

$$f(z) = \sum_{j=-k}^{\infty} a_j (z-P)^j \tag{4.3.1.1}$$

for z near P, then the *principal part* of f at P is

$$\sum_{j=-k}^{-1} a_j (z-P)^j . \tag{4.3.1.2}$$

As an example, the Laurent expansion about 0 of the function $f(z) =$

$(z^2+1)/\sin(z^3)$ is

$$\begin{aligned} f(z) &= (z^2+1)\cdot\frac{1}{\sin(z^3)} \\ &= (z^2+1)\cdot\frac{1}{z^3}\cdot\frac{1}{1-z^6/3!+\cdots} \\ &= \frac{1}{z^3}+\frac{1}{z}+ \text{ (a holomorphic function).} \end{aligned}$$

The principal part of f is $1/z^3+1/z$.

For a second example, consider the function $f(z)=(z^2+2z+2)\sin(1/(z+1))$. Its Laurent expansion about the point -1 is

$$\begin{aligned} f(z) &= ((z+1)^2+1)\cdot\left[\frac{1}{z+1}-\frac{1}{6(z+1)^3}+\frac{1}{120(z+1)^5}\right. \\ &\qquad \left.-\frac{1}{5040(z+1)^7}+-\cdots\right] \\ &= (z+1)+\frac{5}{6}\frac{1}{(z+1)}-\frac{19}{120}\frac{1}{(z+1)^3}+\frac{41}{5040}\frac{1}{(z+1)^5}-+\cdots . \end{aligned}$$

The principal part of f at the point -1 is

$$\frac{5}{6}\frac{1}{(z+1)}-\frac{19}{120}\frac{1}{(z+1)^3}+\frac{41}{5040}\frac{1}{(z+1)^5}-+\cdots .$$

4.3.2 Algorithm for Calculating the Coefficients of the Laurent Expansion

Let f be holomorphic on $D(P,r)\setminus\{P\}$ and suppose that f has a pole of order k at P. Then the Laurent series coefficients a_j of f expanded about the point P, for $j=-k,-k+1,-k+2,\ldots$, are given by the formula

$$a_j=\frac{1}{(k+j)!}\left(\frac{\partial}{\partial z}\right)^{k+j}\left((z-P)^k\cdot f\right)\Big|_{z=P}. \qquad (4.3.2.1)$$

This formula is easily derived by considering the standard power series coefficients of $(z-P)^k\cdot f(z)$.

4.4 The Calculus of Residues

4.4.1 Functions with Multiple Singularities

It turns out to be useful, especially in evaluating various types of integrals, to consider functions that have more than one "singularity." We want to consider the following general question:

> Suppose that $f : U \setminus \{P_1, P_2, \ldots, P_n\} \to \mathbb{C}$ is a holomorphic function on an open set $U \subseteq \mathbb{C}$ with finitely many distinct points $P_1, P_2, \ldots, P_n$ removed. Suppose further that
>
> $$\gamma : [0,1] \to U \setminus \{P_1, P_2, \ldots, P_n\} \tag{4.4.1.1}$$
>
> is a piecewise C^1 closed curve (§2.3.3) that (typically) "surrounds" some (but perhaps not all) of the points $P_1, \ldots, P_n$. Then how is $\oint_\gamma f$ related to the behavior of f near the points $P_1, P_2, \ldots, P_n$?

The first step is to restrict our attention to open sets U for which $\oint_\gamma f$ is necessarily 0 if $P_1, P_2, \ldots, P_n$ are removable singularities of f. See the next subsection.

4.4.2 The Residue Theorem

Suppose that $U \subseteq \mathbb{C}$ is a simply connected open set in $\mathbb{C}$, and that $P_1, \ldots, P_n$ are distinct points of U. Suppose that $f : U \setminus \{P_1, \ldots, P_n\} \to \mathbb{C}$ is a holomorphic function and γ is a piecewise C^1 curve in $U \setminus \{P_1, \ldots, P_n\}$. Set

$$\begin{aligned} R_j = \text{the coefficient of\&}(z - P_j)^{-1} \\ \text{in the Laurent expansion of } f \text{ about} P_j \,. \end{aligned} \tag{4.4.2.1}$$

Then

$$\oint_\gamma f = \sum_{j=1}^{n} R_j \cdot \left(\oint_\gamma \frac{1}{\zeta - P_j} d\zeta \right) . \tag{4.4.2.2}$$

To see this, first note that the integral over γ may be broken up into integrals over "smaller curves," each of which surrounds just one pole. See Figure 4.4. Then each such integral reduces, by deformation of curves, to an integral around a circle. Thus the result is a straightforward calculation.

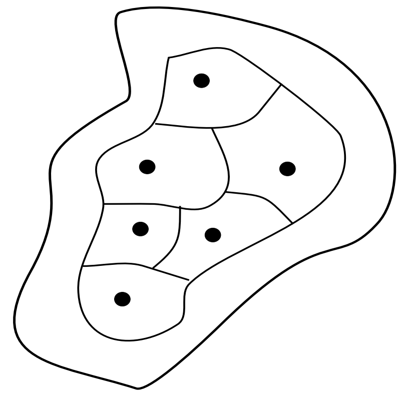

Figure 4.4: Reduction to simpler curves.

4.4.3 Residues

The result just stated is used so often that some special terminology is commonly used to simplify its statement. First, the number R_j is usually called the *residue* of f at P_j , written $\mathrm{Res}_f(P_j)$. Note that this terminology of considering the number R_j attached to the point P_j makes sense because $\mathrm{Res}_f(P_j)$ is completely determined by knowing f in a small neighborhood of P_j. In particular, the value of the residue does not depend on what the other points $P_k,\ k \neq j$, might be, or on how f behaves near those points.

4.4.4 The Index or Winding Number of a Curve about a Point

The second piece of terminology associated to our result deals with the integrals that appear on the right-hand side of equation (4.4.2.2).

If $\gamma : [a, b] \to \mathbb{C}$ is a piecewise C^1 closed curve and if $P \notin \widetilde{\gamma} \equiv \gamma([a, b])$, then the *index of* γ *with respect to* P, written $\mathrm{Ind}_\gamma(P)$, is defined to be the number

$$\frac{1}{2\pi i} \oint_\gamma \frac{1}{\zeta - P}\, d\zeta. \tag{4.4.4.1}$$

The index is also sometimes called the "winding number of the curve γ about the point P." It is a fact that $\mathrm{Ind}_\gamma(P)$ is always an integer. This may be verified by examining a particular differential equation that the curve will satisfy—see [GRK]. Figure 4.5 illustrates the index of various curves γ with respect to different points P. Intuitively, the index measures the number of

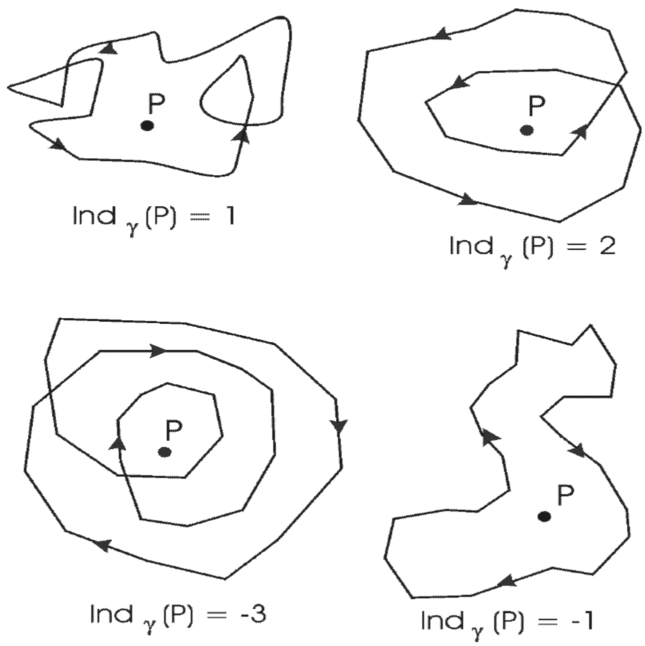

Figure 4.5: The concept of index.

times the curve wraps around P, with counterclockwise being the positive direction of wrapping and clockwise being the negative.

The fact that the index is an integer-valued function suggests that the index counts the topological winding of the curve γ. Note in particular that a curve that traces a circle about the origin k times in a counterclockwise direction has index k with respect to the origin; a curve that traces a circle about the origin k times in a clockwise direction has index $-k$ with respect to the origin.

4.4.5 Restatement of the Residue Theorem

Using the notation of residue and index, the Residue Theorem's formula becomes

$$\oint_\gamma f = 2\pi i \cdot \sum_{j=1}^{n} \mathrm{Res}_f(P_j) \cdot \mathrm{Ind}_\gamma(P_j). \tag{4.4.5.1}$$

People sometimes state this formula informally as "the integral of f around γ equals $2\pi i$ times the sum of the residues counted according to the index of γ about the singularities."

4.4.6 Method for Calculating Residues

We need a method for calculating residues.

Let f be a function with a pole of order k at P. Then

$$\operatorname{Res}_f(P) = \frac{1}{(k-1)!}\left(\frac{\partial}{\partial z}\right)^{k-1}\bigl((z-P)^k f(z)\bigr)\bigg|_{z=P}. \tag{4.4.6.1}$$

This is just a special case of the formula (4.3.2.1).

Poles and Laurent Coefficients

Item	Formula
j^{th} Laurent coefficient of f with pole of order k at P	$\displaystyle \frac{1}{(k+j)!}\frac{d^{k+j}}{dz^{k+j}}\left[(z-P)^k\cdot f\right]\Big\vert_{z=P}$
residue of f with a pole of order k at P	$\displaystyle \frac{1}{(k-1)!}\frac{d^{k-1}}{dz^{k-1}}\left[(z-P)^k\cdot f\right]\Big\vert_{z=P}$
order of pole of f at P	least integer $k \geq 0$ such that $(z-P)^k \cdot f$ is bounded near P
order of pole of f at P	$\displaystyle \lim_{z\to P}\left\vert\frac{\log\vert f(z)\vert}{\log\vert z-P\vert}\right\vert$

4.4.7 Summary Charts of Laurent Series and Residues

We provide two charts, the first of which summarizes key ideas about Laurent coefficients and the second of which contains key ideas about residues.

4.5 Applications to the Calculation of Definite Integrals and Sums

4.5.1 The Evaluation of Definite Integrals

One of the most classical and fascinating applications of the calculus of residues is the calculation of definite (usually improper) real integrals. It is an over-simplification to call these calculations, taken together, a "technique": it is more like a *collection* of techniques. We present several instances of the method.

Techniques for Finding the Residue at P

Function	Type of Pole	Residue Calculation
$f(z)$	simple	$\lim_{z\to P}(z-P)\cdot f(z)$
$f(z)$	pole of order k k is the least integer such that $\lim_{z\to P}\mu(z)$ exists, where $\mu(z)=(z-P)^k f(z)$	$\displaystyle\lim_{z\to P}\frac{\mu^{(k-1)}(z)}{(k-1)!}$
$\dfrac{m(z)}{n(z)}$	$m(P)\neq 0$, $n(z)=0$, $n'(P)\neq 0$	$\dfrac{m(P)}{n'(P)}$
$\dfrac{m(z)}{n(z)}$	m has zero of order k at P n has zero of order $(k+1)$ at P	$(k+1)\cdot\dfrac{m^{(k)}(P)}{n^{(k+1)}(P)}$
$\dfrac{m(z)}{n(z)}$	m has zero of order r at P n has zero of order $(\ell+r)$ at P	$\displaystyle\lim_{z\to P}\frac{\mu^{(\ell-1)}(z)}{(\ell-1)!}$, $\mu(z)=(z-P)^\ell\dfrac{m(z)}{n(z)}$

4.5.2 A Basic Example of the Indefinite Integral

To evaluate

$$\int_{-\infty}^{\infty}\frac{1}{1+x^4}dx, \tag{4.5.2.1}$$

we "complexify" the integrand to $f(z)=1/(1+z^4)$ and consider the integral

$$\oint_{\gamma_R}\frac{1}{1+z^4}dx.$$

See Figure 4.6.

Now part of the game here is to choose the right piecewise C^1 curve or "contour" γ_R. The appropriateness of our choice is justified (after the fact)

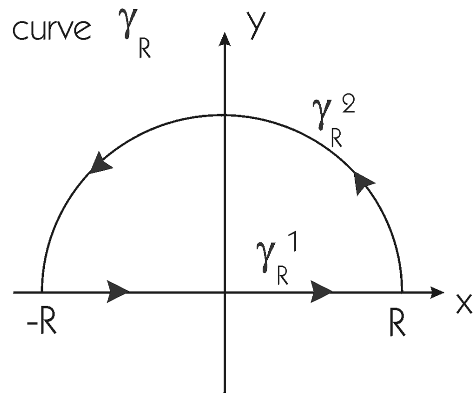

Figure 4.6: The integral in Subsection 4.5.2.

by the calculation that we are about to do. Assume that $R > 1$. Define

$$\begin{aligned} \gamma_R^1(t) &= t + i0 \quad \text{if} \quad -R \le t \le R\,, \\ \gamma_R^2(t) &= Re^{it} \quad \text{if} \quad 0 \le t \le \pi. \end{aligned}$$

Call these two curves, taken together, γ or γ_R.

Now we set $U = \mathbb{C}$, $P_1 = 1/\sqrt{2} + i/\sqrt{2}$, $P_2 = -1/\sqrt{2} + i/\sqrt{2}$, $P_3 = -1/\sqrt{2} - i/\sqrt{2}$, $P_4 = 1/\sqrt{2} - i/\sqrt{2}$; the points P_1, P_2, P_3, P_4 are the poles of $1/[1 + z^4]$. Thus $f(z) = 1/(1 + z^4)$ is holomorphic on $U \setminus \{P_1, \dots, P_4\}$ and the Residue Theorem applies.

On the one hand,

$$\oint_\gamma \frac{1}{1 + z^4}\, dz = 2\pi i \sum_{j=1,2} \mathrm{Ind}_\gamma(P_j) \cdot \mathrm{Res}_f(P_j)\,,$$

where we sum only over the poles of f that lie inside γ. These are P_1 and P_2. An easy calculation shows that

$$\mathrm{Res}_f(P_1) = \frac{1}{4(1/\sqrt{2} + i/\sqrt{2})^3} = -\frac{1}{4}\left(\frac{1}{\sqrt{2}} + i\frac{1}{\sqrt{2}}\right)$$

and

$$\mathrm{Res}_f(P_2) = \frac{1}{4(-1/\sqrt{2} + i/\sqrt{2})^3} = -\frac{1}{4}\left(-\frac{1}{\sqrt{2}} + i\frac{1}{\sqrt{2}}\right).$$

Of course the index at each point is 1. So

$$\begin{aligned}\oint_\gamma \frac{1}{1+z^4}\,dz &= 2\pi i\left(-\frac{1}{4}\right)\left[\left(\frac{1}{\sqrt{2}}+i\frac{1}{\sqrt{2}}\right)+\left(-\frac{1}{\sqrt{2}}+i\frac{1}{\sqrt{2}}\right)\right]\\ &= \frac{\pi}{\sqrt{2}}\,. \end{aligned} \tag{4.5.2.2}$$

On the other hand,

$$\oint_\gamma \frac{1}{1+z^4}\,dz = \oint_{\gamma_R^1} \frac{1}{1+z^4}\,dz + \oint_{\gamma_R^2} \frac{1}{1+z^4}\,dz.$$

Trivially,

$$\oint_{\gamma_R^1} \frac{1}{1+z^4}\,dz = \int_{-R}^{R} \frac{1}{1+t^4}\cdot 1\cdot dt \to \int_{-\infty}^{\infty} \frac{1}{1+t^4}dt \tag{4.5.2.3}$$

as $R \to +\infty$. That is good, because this last is the integral that we wish to evaluate. Better still,

$$\left|\oint_{\gamma_R^2} \frac{1}{1+z^4}dx\right| \le \{\text{length}(\gamma_R^2)\}\cdot \max_{\gamma_R^2}\left|\frac{1}{1+z^4}\right| \le \pi R\cdot \frac{1}{R^4-1}.$$

[Here we use the inequality $|1+z^4| \ge |z|^4 - 1$, as well as (2.1.7.2).] Thus

$$\left|\oint_{\gamma_R^2} \frac{1}{1+z^4}\,dz\right| \to 0 \qquad \text{as} \qquad R\to\infty. \tag{4.5.2.4}$$

Finally, (4.5.2.2)–(4.5.2.4) taken together yield

$$\begin{aligned}\frac{\pi}{\sqrt{2}} &= \lim_{R\to\infty}\oint_\gamma \frac{1}{1+z^4}\,dz\\ &= \lim_{R\to\infty}\oint_{\gamma_R^1} \frac{1}{1+z^4}\,dz + \lim_{R\to\infty}\oint_{\gamma_R^2} \frac{1}{1+z^4}\,dz\\ &= \int_{-\infty}^{\infty}\frac{1}{1+t^4}dt + 0.\end{aligned}$$

This solves the problem: the value of the integral is $\pi/\sqrt{2}$.

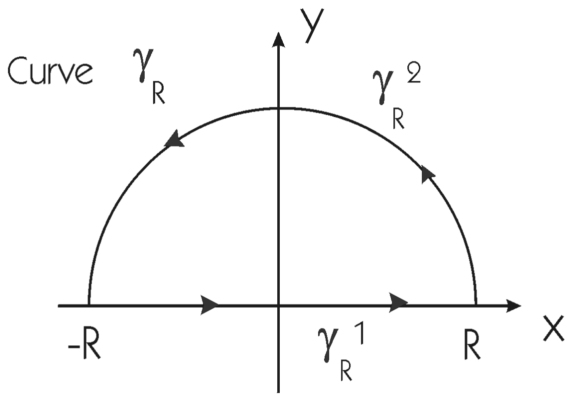

Figure 4.7: The integral in Subsection 4.5.3.

In other problems, it will not be so easy to pick the contour so that the superfluous parts (in the above example, this would be the integral over γ_R^2) tend to zero, nor is it always so easy to prove that they *do* tend to zero. Sometimes, it is not even obvious how to complexify the integrand.

4.5.3 Complexification of the Integrand

We evaluate

$$\int_{-\infty}^{\infty} \frac{\cos x}{1+x^2} dx \tag{4.5.3.1}$$

by using the contour γ_R as in Figure 4.7 (which is the same as Figure 4.6 from the previous example). The obvious choice for the complexification of the integrand is

$$f(z) = \frac{\cos z}{1+z^2} = \frac{[e^{iz} + e^{-iz}]/2}{1+z^2} = \frac{[e^{ix}e^{-y} + e^{-ix}e^{y}]/2}{1+z^2}. \tag{4.5.3.2}$$

Now $|e^{ix}e^{-y}| = |e^{-y}| \leq 1$ on γ_R but $|e^{-ix}e^{y}| = |e^{y}|$ becomes quite large on γ_R when R is large and positive. There is no evident way to alter the contour so that good estimates result. Instead, we alter the function! Let $g(z) = e^{iz}/(1+z^2)$.

On the one hand (for $R > 1$),

$$\begin{aligned}\oint_{\gamma_R} g(z) &= 2\pi i \cdot \mathrm{Res}_g(i) \cdot \mathrm{Ind}_{\gamma_R}(i) \\ &= 2\pi i \left(\frac{1}{2ei}\right) \cdot 1 = \frac{\pi}{e}.\end{aligned}$$

On the other hand, with $\gamma_R^1(t) = t, -R \le t \le R$, and $\gamma_R^2(t) = Re^{it}, 0 \le t \le \pi$, we have

$$\oint_{\gamma_R} g(z)\,dz = \oint_{\gamma_R^1} g(z)\,dz + \oint_{\gamma_R^2} g(z)\,dz.$$

Of course

$$\oint_{\gamma_R^1} g(z)\,dz \to \int_{-\infty}^{\infty} \frac{e^{ix}}{1+x^2}dx \quad \text{as} \quad R \to \infty.$$

And

$$\left|\oint_{\gamma_R^2} g(z)\,dz\right| \le \mathrm{length}(\gamma_R^2) \cdot \max_{\gamma_R^2} |g| \le \pi R \cdot \frac{1}{R^2 - 1} \to 0 \quad \text{as} \quad R \to \infty.$$

Here we have again used (2.1.7.2).

Thus

$$\int_{-\infty}^{\infty} \frac{\cos x}{1+x^2}dx = \mathrm{Re} \int_{-\infty}^{\infty} \frac{e^{ix}}{1+x^2}dx = \mathrm{Re}\left(\frac{\pi}{e}\right) = \frac{\pi}{e}.$$

4.5.4 An Example with a More Subtle Choice of Contour

Let us evaluate

$$\int_{-\infty}^{\infty} \frac{\sin x}{x} dx. \tag{4.5.4.1}$$

Before we begin, we remark that $\sin x/x$ is bounded near zero; also, the integral converges at ∞ (as an improper Riemann integral) by integration by parts. So the problem makes sense. Using the lesson learned from the last example, we consider the function $g(z) = e^{iz}/z$. However, the pole of e^{iz}/z is at $z = 0$ and that lies *on the contour* in Figure 4.8. Thus *that* contour may not be used. We instead use the contour $\mu = \mu_R$ that is depicted in Figure 4.9.

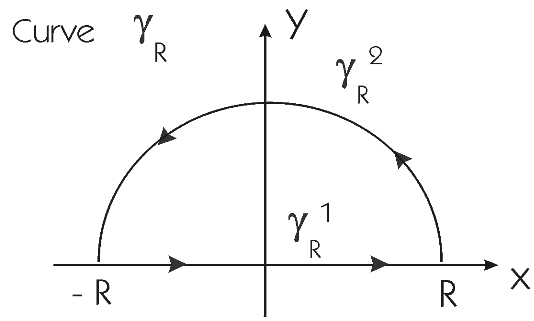

Figure 4.8: The integral in Subsection 4.5.4.

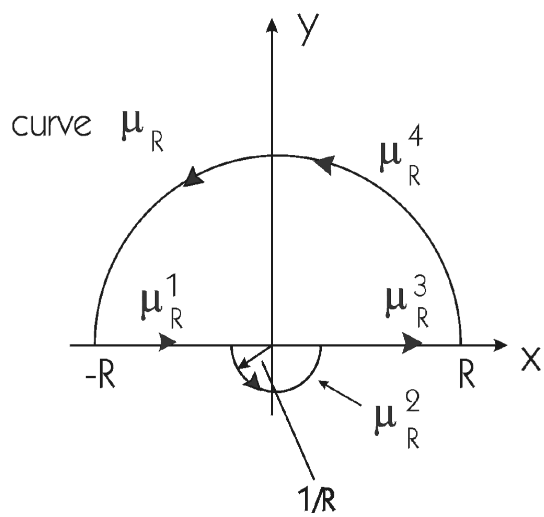

Figure 4.9: More on the integral in Subsection 4.5.4.

Define

$$\begin{array}{rcl} \mu_R^1(t) & = & t, \qquad -R \le t \le -1/R, \\ \mu_R^2(t) & = & e^{it}/R, \ \pi \le t \le 2\pi, \\ \mu_R^3(t) & = & t, \qquad 1/R \le t \le R, \\ \mu_R^4(t) & = & Re^{it}, \ 0 \le t \le \pi. \end{array}$$

Clearly

$$\oint_\mu g(z)\,dz = \sum_{j=1}^{4} \oint_{\mu_R^j} g(z)\,dz.$$

On the one hand, for $R > 0$,

$$\oint_\mu g(z)\,dz = 2\pi i \mathrm{Res}_g(0) \cdot \mathrm{Ind}_\mu(0) = 2\pi i \cdot 1 \cdot 1 = 2\pi i. \tag{4.5.4.2}$$

On the other hand,

$$\oint_{\mu_R^1} g(z)\,dz + \oint_{\mu_R^3} g(z)\,dz \to \int_{-\infty}^{\infty} \frac{e^{ix}}{x} dx \quad \text{as} \quad R \to \infty. \tag{4.5.4.3}$$

Furthermore,

$$\begin{aligned} \left| \oint_{\mu_R^4} g(z)\,dz \right| &\le \left| \oint_{\substack{\mu_R^4 \\ \mathrm{Im}\ y < \sqrt{R}}} g(z)\,dz \right| + \left| \oint_{\substack{\mu_R^4 \\ \mathrm{Im}\ y \ge \sqrt{R}}} g(z)\,dz \right| \\ &\equiv A + B. \end{aligned}$$

Now

$$\begin{array}{rcl} A & \le & \mathrm{length}(\mu_R^4 \cap \{z : \mathrm{Im}\ z < \sqrt{R}\}) \cdot \max\{|g(z)| : z \in \mu_R^4, y < \sqrt{R}\} \\ & \le & 4\sqrt{R} \cdot \left(\dfrac{1}{R}\right) \to 0 \quad \text{as} \quad R \to \infty. \end{array}$$

Also

$$\begin{array}{rcl} B & \le & \mathrm{length}(\mu_R^4 \cap \{z : \mathrm{Im}\ z \ge \sqrt{R}\}) \cdot \max\{|g(z)| : z \in \mu_R^4, y \ge \sqrt{R}\} \\ & \le & \pi R \cdot \left(\dfrac{e^{-\sqrt{R}}}{R}\right) \to 0 \quad \text{as} \quad R \to \infty. \end{array}$$

So

$$\left| \oint_{\mu_R^4} g(z)\, dz \right| \to 0 \quad \text{as} \quad R \to \infty. \tag{4.5.4.4}$$

Finally,

$$\begin{aligned}
\oint_{\mu_R^2} g(z)\, dz &= \int_{\pi}^{2\pi} \frac{e^{i(e^{it}/R)}}{e^{it}/R} \cdot \left(\frac{i}{R} e^{it} \right) dt \\
&= i \int_{\pi}^{2\pi} e^{i(e^{it}/R)} dt.
\end{aligned}$$

As $R \to \infty$ this tends to

$$\begin{aligned}
&= i \int_{\pi}^{2\pi} 1\, dt \\
&= \pi i \quad \text{as} \quad R \to \infty\,.
\end{aligned} \tag{4.5.4.5}$$

In summary, (4.5.4.2) – (4.5.4.5) yield

$$2\pi i = \oint_{\mu} g(z)\, dz = \sum_{j=1}^{4} \oint_{\mu_R^j} g(z)\, dz$$

$$\to \int_{-\infty}^{\infty} \frac{e^{ix}}{x} dx + \pi i \quad \text{as} \quad R \to \infty.$$

Taking imaginary parts yields

$$\pi = \int_{-\infty}^{\infty} \frac{\sin x}{x} dx.$$

4.5.5 Making the Spurious Part of the Integral Disappear

Consider the integral

$$\int_0^{\infty} \frac{x^{1/3}}{1+x^2} dx. \tag{4.5.5.1}$$

We complexify the integrand by setting $f(z) = z^{1/3}/(1+z^2)$. Note that, on the simply connected set $U = \mathbb{C} \setminus \{iy : y < 0\}$, the expression $z^{1/3}$ is

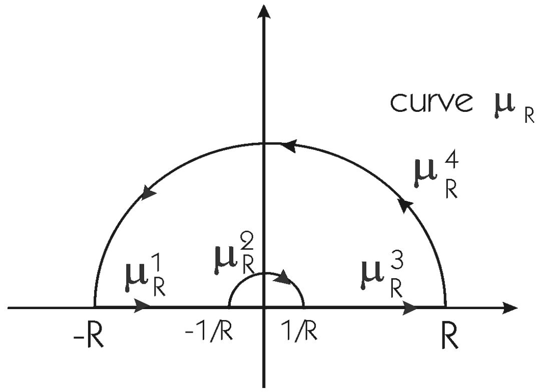

Figure 4.10: The integral in Subsection 4.5.5.

unambiguously defined as a holomorphic function by setting $z^{1/3} = r^{1/3}e^{i\theta/3}$ when $z = re^{i\theta}, -\pi/2 < \theta < 3\pi/2$. We use the contour displayed in Figure 4.10.

We must do this since $z^{1/3}$ is not a well-defined holomorphic function in any neighborhood of 0. Let us use the notation from the figure. We refer to the preceding examples for some of the parametrizations that we now use.

Clearly

$$\oint_{\mu_R^3} f(z)\,dz \to \int_0^\infty \frac{t^{1/3}}{1+t^2}dt.$$

Of course that is good, but what will become of the integral over μ_R^1? We have

$$\begin{aligned}\oint_{\mu_R^1} &= \int_{-R}^{-1/R} \frac{t^{1/3}}{1+t^2}dt \\ &= \int_{1/R}^{R} \frac{(-t)^{1/3}}{1+t^2}dt \\ &= \int_{1/R}^{R} \frac{e^{i\pi/3}t^{1/3}}{1+t^2}dt.\end{aligned}$$

(by our definition of $z^{1/3}$!). Thus

$$\oint_{\mu_R^3} f(z)\,dz + \oint_{\mu_R^1} f(z)\,dz \to \left(1 + \left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)\right) \int_0^\infty \frac{t^{1/3}}{1+t^2}dt \quad \text{as} \quad R \to \infty.$$

On the other hand,

$$\left|\oint_{\mu_R^4} f(z)\,dz\right| \le \pi R \cdot \frac{R^{1/3}}{R^2-1} \to 0 \quad \text{as} \quad R \to \infty$$

and

$$\begin{aligned}\oint_{\mu_R^2} f(z)\,dz &= \int_{-\pi}^{-2\pi} \frac{(e^{it}/R)^{1/3}}{1+e^{2it}/R^2}(i)e^{it}/R\,dt \\ &= R^{-4/3}\int_{-\pi}^{-2\pi} \frac{e^{i4t/3}}{1+e^{2it}/R^2}\,dt \to 0 \quad \text{as} \quad R \to \infty.\end{aligned}$$

So

$$\oint_{\mu_R} f(z)\,dz \to \left(\frac{3}{2}+\frac{\sqrt{3}}{2}i\right)\int_0^\infty \frac{t^{1/3}}{1+t^2}\,dt \quad \text{as} \quad R \to \infty. \tag{4.5.5.2}$$

The calculus of residues tells us that, for $R > 1$,

$$\begin{aligned}\oint_{\mu_R} f(z)\,dz &= 2\pi i \mathrm{Res}_f(i)\cdot \mathrm{Ind}_{\mu_R}(i) \\ &= 2\pi i\left(\frac{e^{i\pi/6}}{2i}\right)\cdot 1 \\ &= \pi\left(\frac{\sqrt{3}}{2}+\frac{i}{2}\right). \end{aligned} \tag{4.5.5.3}$$

Finally, (4.5.5.2) and (4.5.5.3) taken together yield

$$\int_0^\infty \frac{t^{1/3}}{1+t^2}\,dt = \frac{\pi}{\sqrt{3}}. \qquad \square$$

4.5.6 The Use of the Logarithm

While the integral

$$\int_0^\infty \frac{dx}{x^2+6x+8} \tag{4.5.6.1}$$

can be calculated using methods of calculus, it is enlightening to perform the integration by complex variable methods. Note that if we endeavor to

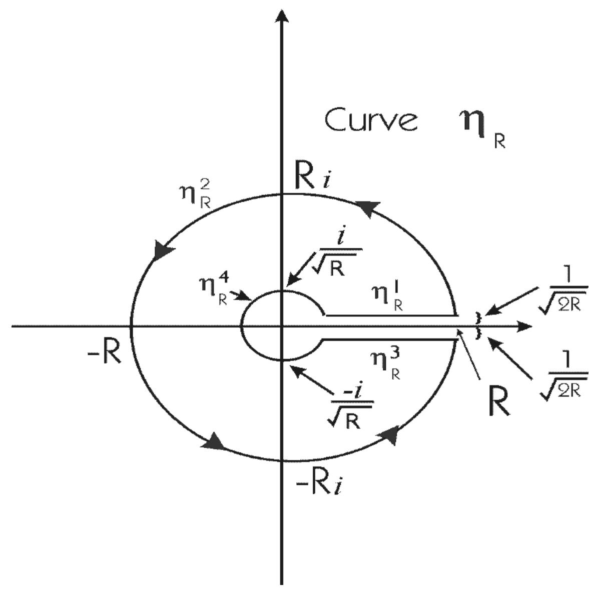

Figure 4.11: The integral in Subsection 4.5.6.

use the integrand $f(z) = 1/(z^2 + 6z + 8)$ together with the idea of the last example, then there is no "auxiliary radius" that helps. More precisely, $((re^{i\theta})^2 + 6re^{i\theta} + 8)$ is a constant multiple of $r^2 + 6r + 8$ only if θ is an integer multiple of 2π. The following non-obvious device is often of great utility in problems of this kind. Define $\log z$ on $U \equiv \mathbb{C} \setminus \{x : x \geq 0\}$ by $\log(re^{i\theta}) = (\log r) + i\theta$ when $0 < \theta < 2\pi, r > 0$. Here $\log r$ is understood to be the standard real logarithm. Then, on U, log is a well-defined holomorphic function. [Observe here that there are infinitely many ways to define the logarithm function on U. One could set $\log(re^{i\theta}) = (\log r) + i(\theta + 2k\pi)$ for any integer choice of k. What we have done here is called "choosing a branch" of the logarithm.]

We use the contour η_R displayed in Figure 4.11 and integrate the function $g(z) = \log z/(z^2 + 6z + 8)$. Let

$$\begin{aligned} \eta_R^1(t) &= t + i/\sqrt{2R}, \quad 1/\sqrt{2R} \leq t \leq R, \\ \eta_R^2(t) &= Re^{it}, \quad \theta_0 \leq t \leq 2\pi - \theta_0, \end{aligned}$$

where $\theta_0(R) = \tan^{-1}(1/(R\sqrt{2R}))$

$$\begin{aligned} \eta_R^3(t) &= R - t - i/\sqrt{2R}, \quad 0 \leq t \leq R - 1/\sqrt{2R}, \\ \eta_R^4(t) &= e^{-it}/\sqrt{R}, \quad \pi/4 \leq t \leq 7\pi/4. \end{aligned}$$

Now

$$\begin{aligned}\oint_{\eta_R} g(z)\,dz &= 2\pi i(\mathrm{Res}_{\eta_R}(-2)\cdot 1 + \mathrm{Res}_{\eta_R}(-4)\cdot 1)\\ &= 2\pi i\left(\frac{\log(-2)}{2} + \frac{\log(-4)}{-2}\right)\\ &= 2\pi i\left(\frac{\log 2 + \pi i}{2} + \frac{\log 4 + \pi i}{-2}\right)\\ &= -\pi i \log 2\,. \end{aligned} \tag{4.5.6.2}$$

Also, it is straightforward to check that

$$\left|\oint_{\eta_R^2} g(z)\,dz\right| \to 0, \tag{4.5.6.3}$$

$$\left|\oint_{\eta_R^4} g(z)\,dz\right| \to 0\,, \tag{4.5.6.4}$$

as $R \to \infty$. The device that makes this technique work is that, as $R \to \infty$,

$$\log(x + i/\sqrt{2R}) - \log(x - i/\sqrt{2R}) \to -2\pi i\,.$$

So

$$\oint_{\eta_R^1} g(z)\,dz + \oint_{\eta_R^3} g(z)\,dz \to -2\pi i \int_0^\infty \frac{dt}{t^2 + 6t + 8}. \tag{4.5.6.5}$$

Now (4.5.6.2)–(4.5.6.5) taken together yield

$$\int_0^\infty \frac{dt}{t^2 + 6t + 8} = \frac{1}{2}\log 2.$$

4.5.7 Summing a Series Using Residues

We sum the series

$$\sum_{j=1}^{\infty} \frac{x}{j^2\pi^2 - x^2} \tag{4.5.7.1}$$

using contour integration. Define $\cot z = \cos z/\sin z$. For $n = 1, 2, \ldots$ let Γ_n be the contour (shown in Figure 4.12) consisting of the counterclockwise oriented square with corners $\{(\pm 1 \pm i) \cdot (n + \frac{1}{2}) \cdot \pi\}$. For z fixed and $n > |z|$ we calculate using residues that

$$\begin{aligned}\frac{1}{2\pi i}\oint_{\Gamma_n} \frac{\cot\zeta}{\zeta(\zeta - z)} d\zeta \quad &= \quad \sum_{j=1}^{n} \frac{1}{j\pi(j\pi - z)} + \sum_{j=1}^{n} \frac{1}{j\pi(j\pi + z)} \\ &\quad + \frac{\cot z}{z} - \frac{1}{z^2}.\end{aligned}$$

When $n \gg |z|$, it is easy to estimate the left-hand side in modulus by

$$\left(\frac{1}{2\pi}\right) \cdot [4(2n+1)\pi] \cdot \left(\frac{C}{n(n - |z|)}\right) \to 0 \quad \text{as} \quad n \to \infty.$$

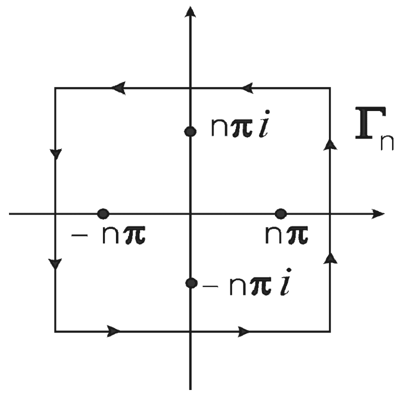

Figure 4.12: The integral in Subsection 4.5.7.

Thus we see that

$$\sum_{j=1}^{\infty} \frac{1}{j\pi(j\pi - z)} + \sum_{j=1}^{\infty} \frac{1}{j\pi(j\pi + z)} = -\frac{\cot z}{z} + \frac{1}{z^2}.$$

We conclude that

$$\sum_{j=1}^{\infty} \frac{2}{j^2\pi^2 - z^2} = -\frac{\cot z}{z} + \frac{1}{z^2}$$

or

$$\sum_{j=1}^{\infty} \frac{z}{j^2\pi^2 - z^2} = -\frac{1}{2}\cot z + \frac{1}{2z}.$$

This is the desired result . □

4.6 Singularities at Infinity

4.6.1 Meromorphic Functions

We have considered carefully those functions that are holomorphic on sets of the form $D(P, r) \setminus \{P\}$ or, more generally, of the form $U \setminus \{P\}$, where U is an open set in $\mathbb{C}$ and $P \in U$. As we have seen in our discussion of the calculus of residues, sometimes it is important to consider the possibility that a function could be "singular" at more than just one point. The appropriate, precise definition requires a little preliminary consideration of what kinds of sets might be appropriate as "sets of singularities."

4.6.2 Definition of Meromorphic Function

Now fix an open set U; we next define the central concept of meromorphic function on U.

A *meromorphic function* f on U *with singular set* S is a function $f : U \setminus S \to \mathbb{C}$ such that

(4.6.3.1) S is discrete;

(4.6.3.2) f is holomorphic on $U \setminus S$ (note that $U \setminus S$ is necessarily open in $\mathbb{C}$);

(4.6.3.3) for each $P \in S$ and $r > 0$ such that $D(P, r) \subseteq U$ and $S \cap D(P, r) = \{P\}$, the function $f\big|_{D(P,r)\setminus\{P\}}$ has a (finite order) pole at P.

For convenience, one often suppresses explicit consideration of the set S and just says that f is a meromorphic function on U. Sometimes we say, informally, that a meromorphic function on U is a function on U that is holomorphic "except for poles." Implicit in this description is the idea that a pole is an "isolated singularity." In other words, a point P is a pole of f if and only if there is a disc $D(P, r)$ around P such that f is holomorphic on $D(P, r) \setminus \{P\}$ and has a pole at P. Back on the level of precise language, we see that our definition of a meromorphic function on U implies that, for each $P \in U$, either there is a disc $D(P, r) \subseteq U$ such that f is holomorphic on $D(P, r)$ *or* there is a disc $D(P, r) \subseteq U$ such that f is holomorphic on $D(P, r) \setminus \{P\}$ and has a pole at P.

4.6.3 Examples of Meromorphic Functions

Meromorphic functions are very natural objects to consider, primarily because they result from considering the (algebraic) reciprocals—or more generally the quotients—of holomorphic functions:

If U is a connected, open set in $\mathbb{C}$ and if $f : U \to \mathbb{C}$ is a holomorphic function having at least some zeros but with $f \not\equiv 0$, then the function

$$F : U \setminus \{z : f(z) = 0\} \to \mathbb{C} \qquad (4.6.4.1)$$

defined by $F(z) = 1/f(z)$ is a meromorphic function on U with singular set (or pole set) equal to $\{z \in U : f(z) = 0\}$. More generally, meromorphic functions locally have the form $g(z)/f(z)$ for f, g holomorphic. In a sense that can be made precise, all meromorphic functions arise as *quotients* of holomorphic functions.

4.6.4 Meromorphic Functions with Infinitely Many Poles

It is quite possible for a meromorphic function on an open set U to have infinitely many poles in U. The function $1/\sin(1/z)$ is an obvious example on $U = D \setminus \{0\}$.

4.6.5 Singularities at Infinity

Our discussion so far of singularities of holomorphic functions can be generalized to include the limit behavior of holomorphic functions as $|z| \to +\infty$. This is a powerful method with many important consequences. Suppose for example that $f : \mathbb{C} \to \mathbb{C}$ is an entire function. We can associate to f a new function $G : \mathbb{C} \setminus \{0\} \to \mathbb{C}$ by setting $G(z) = f(1/z)$. The behavior of the function G near 0 reflects, in an obvious sense, the behavior of f as $|z| \to +\infty$. For instance

$$\lim_{|z| \to +\infty} |f(z)| = +\infty \tag{4.6.6.1}$$

if and only if G has a pole at 0.

Suppose that $f : U \to \mathbb{C}$ is a holomorphic function on an open set $U \subseteq \mathbb{C}$ and that, for some $R > 0, U \supseteq \{z : |z| > R\}$. Define $G : \{z : 0 < |z| < 1/R\} \to \mathbb{C}$ by $G(z) = f(1/z)$. Then we say that

(4.6.6.2) f has a *removable singularity* at ∞ if G has a removable singularity at 0.

(4.6.6.3) f has a *pole at* ∞ if G has a pole at 0.

(4.6.6.4) f has an *essential singularity* at ∞ if G has an essential singularity at 0.

4.6.6 The Laurent Expansion at Infinity

The Laurent expansion of G around 0, $G(z) = \sum_{-\infty}^{+\infty} a_n z^n$, yields immediately a series expansion for f which converges for $|z| > R$, namely,

$$f(z) \equiv G(1/z) = \sum_{-\infty}^{+\infty} a_n z^{-n} = \sum_{-\infty}^{+\infty} a_{-n} z^n. \tag{4.6.7.1}$$

The series $\sum_{-\infty}^{+\infty} a_{-n} z^n$ is called the *Laurent expansion of f around* ∞. It follows from our definitions and from our earlier discussions that f has a removable singularity at ∞ if and only if the Laurent series of f at ∞ has no *positive* powers of z with non-zero coefficients. Also f has a pole at ∞ if and only if the series has only a finite number of positive powers of z with non-zero coefficients. Finally, f has an essential singularity at ∞ if and only if the series has infinitely many positive powers.

4.6.7 Meromorphic at Infinity

Suppose that $f : \mathbb{C} \to \mathbb{C}$ is an entire function. Then $\lim_{|z|\to+\infty} |f(z)| = +\infty$ if and only if f is a nonconstant polynomial. In other words, an entire function that is not a polynomial will have an essential singularity at infinity.

To see this last assertion, supposed that f has a pole of order k at ∞. Subtracting a polynomial p from f if necessary, we may arrange that $f - p$ vanishes to order k at the origin. Of course p has degree at most k. Then the function

$$g(z) = \frac{f(z) - p(z)}{z^k}$$

is entire and is bounded. By Liouville's theorem, g is constant. But then it follows that f is a polynomial.

The entire function f has a removable singularity at ∞ if and only if f is a constant. This claim if obvious because f will be bounded.

Suppose that f is a meromorphic function defined on an open set $U \subseteq \mathbb{C}$ such that, for some $R > 0$, we have $U \supseteq \{z : |z| > R\}$. We say that f is *meromorphic* at ∞ if the function $G(z) \equiv f(1/z)$ is meromorphic in the usual sense on $\{z : |z| < 1/R\}$.

4.6.8 Meromorphic Functions in the Extended Plane

The definition of "meromorphic at ∞" as given is equivalent to requiring that, for some $R' > R$, f has no poles in $\{z \in \mathbb{C} : R' < |z| < \infty\}$ *and* that f has a pole at ∞. The point is that a pole should not be an accumulation point of other poles.

A meromorphic function f on $\mathbb{C}$ which is also meromorphic at ∞ must be a rational function (that is, a quotient of polynomials in z). For we can arrange for one of the poles to be at ∞. Multiplying f by a polynomial p, we may arrange for $p \cdot f$ to have no poles. So it must be a polynomial. It follows that f is a quotient of polynomials. Conversely, every rational function is meromorphic on $\mathbb{C}$ and at ∞.

Remark: It is conventional to rephrase the ideas just presented by saying that the only functions that are meromorphic in the "extended plane" are rational functions. We will say more about the extended plane in §§6.3.1–6.3.3.

Chapter 5

The Argument Principle

5.1 Counting Zeros and Poles

5.1.1 Local Geometric Behavior of a Holomorphic Function

In this chapter, we shall be concerned with questions that have a geometric, qualitative nature rather than an analytical, quantitative one. These questions center around the issue of the local geometric behavior of a holomorphic function.

5.1.2 Locating the Zeros of a Holomorphic Function

Suppose that $f : U \to \mathbb{C}$ is a holomorphic function on a connected, open set $U \subseteq \mathbb{C}$ and that $\overline{D}(P, r) \subseteq U$. We know from the Cauchy integral formula that the values of f on $D(P, r)$ are completely determined by the values of f on $\partial D(P, r)$. In particular, the number and even the location of the zeros of f in $D(P, r)$ are determined in principle by f on $\partial D(P, r)$. But it is nonetheless a pleasant surprise that there is a *simple formula* for the number of zeros of f in $D(P, r)$ in terms of f (and f') on $\partial D(P, r)$. In order to obtain a precise formula, we shall have to agree to count zeros according to multiplicity (see §§3.1.4). We now explain the precise idea.

Let $f : U \to \mathbb{C}$ be holomorphic as before, and assume that f has *some* zeros in U but that f is not identically zero. Fix $z_0 \in U$ such that $f(z_0) = 0$. Since the zeros of f are isolated, there is an $r > 0$ such that $\overline{D}(z_0, r) \subseteq U$ and such that f does not vanish on $\overline{D}(z_0, r) \setminus \{z_0\}$.

Now the power series expansion of f about z_0 has a first non-zero term determined by the least positive integer n such that $f^{(n)}(z_0) \neq 0$. (Note that $n \geq 1$ since $f(z_0) = 0$ by hypothesis.) Thus the power series expansion of f about z_0 *begins* with the n^{th} term:

$$f(z) = \sum_{j=n}^{\infty} \frac{1}{j!} \frac{\partial^j f}{\partial z^j}(z_0)(z - z_0)^j. \tag{5.1.2.1}$$

Under these circumstances we say that f has a zero of *order* n (or *multiplicity* n) at z_0. When $n = 1$, then we also say that z_0 is a *simple* zero of f.

5.1.3 Zero of Order n

The concept of zero of "order n," or "multiplicity n," for a function f is so important that a variety of terminology has grown up around it (see also §§3.1.4). It has already been noted that, when the multiplicity $n = 1$, then the zero is sometimes called *simple*. For arbitrary n, we sometimes say that "n is the order of z_0 as a zero of f." More generally if $f(z_0) = \beta$ so that, for some $n \geq 1$, the function $f(\,\cdot\,) - \beta$ has a zero of order n at z_0, then we say either that "f assumes the value β at z_0 to order n" or that "the order of the value β at z_0 is n." When $n > 1$, then we call z_0 a *multiple point* (and β a multiple value) of the function f.

The next result provides a method for computing the multiplicity n of the zero at z_0 from the values of f, f' on the boundary of a disc centered at z_0.

5.1.4 Counting the Zeros of a Holomorphic Function

If f is holomorphic on a neighborhood of a disc $\overline{D}(P, r)$ and has a zero of order n at P and no other zeros in the closed disc, then

$$\frac{1}{2\pi i} \oint_{\partial D(P,r)} \frac{f'(\zeta)}{f(\zeta)}\, d\zeta = n. \tag{5.1.4.1}$$

More generally, we consider the case that f has several zeros—with different locations and different multiplicities—inside a disc: Suppose that $f : U \to \mathbb{C}$ is holomorphic on an open set $U \subseteq \mathbb{C}$ and that $\overline{D}(P, r) \subseteq U$. Suppose that f is non-vanishing on $\partial D(P, r)$ and that $z_1, z_2, \ldots, z_k$ are the

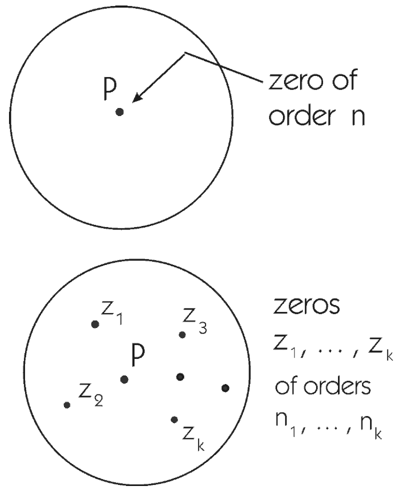

Figure 5.1: Counting the zeros of a holomorphic function.

zeros of f in the interior of the disc. Let n_ℓ be the order of the zero of f at z_ℓ, $\ell = 1, \ldots, k$. Then

$$\frac{1}{2\pi} \oint_{|\zeta - P| = r} \frac{f'(\zeta)}{f(\zeta)}\, d\zeta = \sum_{\ell=1}^{k} n_\ell. \tag{5.1.4.2}$$

Refer to Figure 5.1 for illustrations of both these situations.

The reasons for these formulas is actually quite simple. The primordial situation is when $P = 0$ and $f(z) = z^k$. In that case we may compute the integral (5.1.4.1) directly and the result is immediate. We may write a more general f as $f(z) = \widetilde{f}(z) \cdot z^k$ and then the integral (5.1.4.1) reduces, by simple algebra, to the simpler situation just treated. Of course the integral over a more general curve can be reduced to the integral over the boundary of a disc by our usual device of deformation of curves. Finally, the situation of several different zeros may be reduced to the situation of one zero by breaking up the curve of integration into smaller curves, each having just one zero in its interior.

5.1.5 The Argument Principle

This last formula (5.1.4.2), which is often called the *argument principle*, is both useful and important. For one thing, there is no obvious reason why

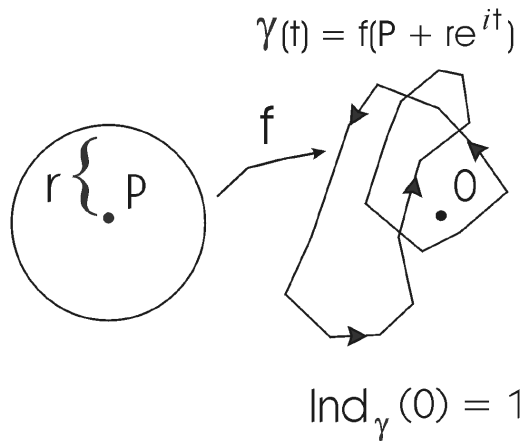

Figure 5.2: The argument principle.

the integral in the formula should be an integer, much less the crucial integer that it is. Since it is an integer, it is a counting function; and we need to learn more about it.

The integral

$$\frac{1}{2\pi}\oint_{|\zeta-P|=r}\frac{f'(\zeta)}{f(\zeta)}\,d\zeta \tag{5.1.5.1}$$

can be reinterpreted as follows: Consider the C^1 closed curve

$$\gamma(t)=f(P+re^{it}),\quad t\in[0,2\pi]. \tag{5.1.5.2}$$

Then

$$\frac{1}{2\pi}\oint_{|\zeta-P|=r}\frac{f'(\zeta)}{f(\zeta)}\,d\zeta=\frac{1}{2\pi}\int_0^{2\pi}\frac{\gamma'(t)}{\gamma(t)}dt, \tag{5.1.5.3}$$

as you can check by direct calculation. The expression on the right is just the index of the curve γ with respect to 0 (with the notion of index that we defined earlier—§§4.4.4). See Figure 5.2. Thus the number of zeros of f (counting multiplicity) inside the circle $\{\zeta : |\zeta - P| = r\}$ is equal to the index of γ with respect to the origin. This, intuitively speaking, is equal to the number of times that the f-image of the boundary circle winds around 0 in $\mathbb{C}$.

The argument principle can be extended to yield information about meromorphic functions, too. We can see that there is hope for this notion by investigating the analog of the argument principle for a pole.

5.1.6 Location of Poles

If $f : U \setminus \{Q\} \to \mathbb{C}$ is a nowhere-zero holomorphic function on $U \setminus \{Q\}$ with a pole of order n at Q and if $\overline{D}(Q, r) \subseteq U$, then

$$\frac{1}{2\pi} \oint_{\partial D(Q,r)} \frac{f'(\zeta)}{f(\zeta)}\, d\zeta = -n. \tag{5.1.6.1}$$

Just as we argued for zeros, the verification of (5.1.6.1) can be reduced to checking the identity for the function $f(z) = z^{-k}$ when $Q = 0$.

5.1.7 The Argument Principle for Meromorphic Functions

Just as with the argument principle for holomorphic functions, this new argument principle gives a counting principle for zeros and poles of meromorphic functions:

Suppose that f is a meromorphic function on an open set $U \subseteq \mathbb{C}$, that $\overline{D}(P, r) \subseteq U$, and that f has neither poles nor zeros on $\partial D(P, r)$. Assume that $n_1, n_2, \ldots, n_p$ are the multiplicities of the zeros $z_1, z_2, \ldots, z_p$ of f in $D(P, r)$ and $m_1, m_2, \ldots, m_q$ are the orders of the poles $w_1, w_2, \ldots, w_q$ of f in $D(P, r)$.

Then

$$\frac{1}{2\pi} \oint_{\partial D(P,r)} \frac{f'(\zeta)}{f(\zeta)}\, d\zeta = \sum_{j=1}^{p} n_j - \sum_{k=1}^{q} m_k. \tag{5.1.7.1}$$

5.2 The Local Geometry of Holomorphic Functions

5.2.1 The Open Mapping Theorem

The argument principle for holomorphic functions has a consequence that is one of the most important facts about holomorphic functions considered as geometric mappings:

> **Theorem:** If $f : U \to \mathbb{C}$ is a non-constant holomorphic function on a connected open set U, then $f(U)$ is an open set in $\mathbb{C}$.

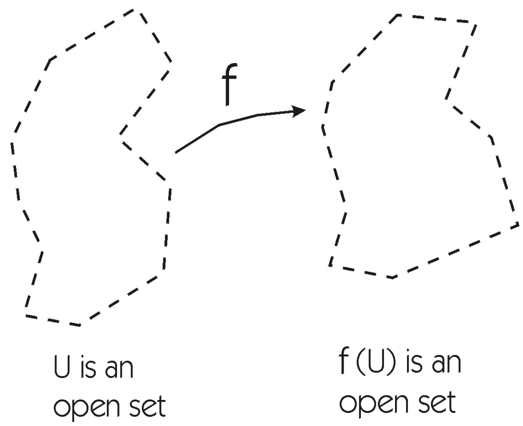

Figure 5.3: The open mapping principle.

(5.2.1.1)

See Figure 5.3. The result says, in particular, that if $U \subseteq \mathbb{C}$ is connected and open and if $f : U \to \mathbb{C}$ is holomorphic, then either $f(U)$ is a connected open set (the non-constant case) or $f(U)$ is a single point.

To see why the open mapping principle is true, let β be a value of the holomorphic function f. Say that $f(b) = \beta$, and let $D(b, r)$ be a small disc as usual in the domain of f. Suppose for simplicity that β is a simple value of f. Then of course

$$1 = \frac{1}{2\pi i} \oint_{\partial D(b,r)} \frac{f'(\zeta)}{f(\zeta) - \beta} \, d\zeta \, . \tag{5.2.1.2}$$

If we perturb β a little bit on the righthand side of this integral (a perturbation *much smaller* than r), then the integral will still be integer-valued, and its value will be very close to 1. So in fact it must be 1. Thus we conclude that f also takes on that small perturbed value of β. We have argued then that f assumes all values near β. But this means that the image of f is open.

In the subject of topology, a function f is defined to be continuous if the inverse image of any open set under f is also open. In contexts where the $\epsilon - \delta$ definition makes sense, the $\epsilon - \delta$ definition (§§2.2.1, 2.2.2) is equivalent to the inverse-image-of-open-sets definition. By contrast, functions for which the direct image of any open set is open are called "open mappings."

Here is a quantitative, or counting, statement that comes from the proof of the open mapping principle: Suppose that $f : U \to \mathbb{C}$ is a non-constant holomorphic function on a connected open set U such that $P \in U$ and $f(P) = Q$ with order k. Then there are numbers $\delta, \epsilon > 0$ such that each $q \in D(Q, \epsilon) \setminus \{Q\}$ has exactly k distinct pre-images in $D(P, \delta)$ and each pre-image is a simple point of f. Of course the justification is again a simple application of (5.2.1.2).

The considerations that establish the open mapping principle can also be used to establish the fact that if $f : U \to V$ is a one-to-one and onto holomorphic function, then $f^{-1} : V \to U$ is also holomorphic.

5.3 Further Results on the Zeros of Holomorphic Functions

5.3.1 Rouché's Theorem

Now we consider global aspects of the argument principle.

Suppose that $f, g : U \to \mathbb{C}$ are holomorphic functions on an open set $U \subseteq \mathbb{C}$. Suppose also that $\overline{D}(P, r) \subseteq U$ and that, for each $\zeta \in \partial D(P, r)$,

$$|f(\zeta) - g(\zeta)| < |f(\zeta)| + |g(\zeta)|. \tag{5.3.1.1}$$

Then

$$\frac{1}{2\pi} \oint_{\partial D(P,r)} \frac{f'(\zeta)}{f(\zeta)}\, d\zeta = \frac{1}{2\pi} \oint_{\partial D(P,r)} \frac{g'(\zeta)}{g(\zeta)}\, d\zeta. \tag{5.3.1.2}$$

That is, the number of zeros of f in $D(P, r)$ counting multiplicities equals the number of zeros of g in $D(P, r)$ counting multiplicities.

Remark: Rouché's theorem is often stated with the stronger hypothesis that

$$|f(\zeta) - g(\zeta)| < |g(\zeta)| \tag{5.3.1.3}$$

for $\zeta \in \partial D(P, r)$. Rewriting this hypothesis as

$$\left| \frac{f(\zeta)}{g(\zeta)} - 1 \right| < 1, \tag{5.3.1.4}$$

we see that it says that the image γ under f/g of the circle $\partial D(P, r)$ lies in the disc $D(1, 1)$. See Figure 5.4. Our weaker hypothesis that $|f(\zeta) - g(\zeta)| < |f(\zeta)| + |g(\zeta)|$ has the geometric interpretation that $f(\zeta)/g(\zeta)$ lies in the set

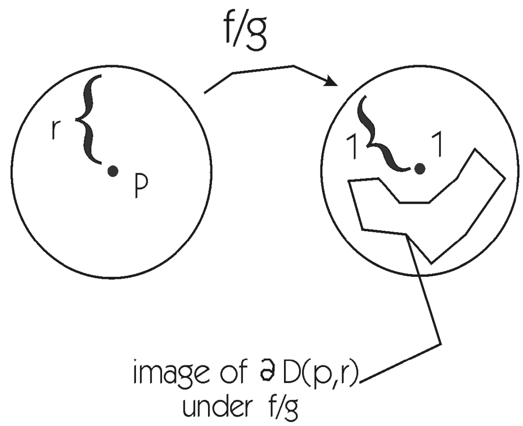

Figure 5.4: Rouché's theorem.

$\mathbb{C} \setminus \{x + i0 : x \leq 0\}$. Either hypothesis implies that the image of the circle $\partial D(P, r)$ under f has the same "winding number" around 0 as does the image under g of that circle.

5.3.2 Typical Application of Rouché's Theorem

EXAMPLE 5.3.2.1 Let us determine the number of roots of the polynomial $f(z) = z^7 + 5z^3 - z - 2$ in the unit disc. We do so by comparing the function f to the holomorphic function $g(z) = 5z^3$ on the unit circle. For $|z| = 1$ we have

$$|f(z) - g(z)| = |z^7 - z - 2| \leq 4 < |g(\zeta)| \leq |f(\zeta)| + |g(\zeta)|.$$

By Rouché's theorem, f and g have the same number of zeros, counting multiplicity, in the unit disc. Since g has three zeros, so does f. $\square$

5.3.3 Rouché's Theorem and the Fundamental Theorem of Algebra

Rouché's theorem provides a useful way to locate approximately the zeros of a holomorphic function that is too complicated for the zeros to be obtained explicitly. As an illustration, we analyze the zeros of a non-constant polynomial

$$P(z) = z^n + a_{n-1}z^{n-1} + a_{n-2}z^{n-2} + \cdots + a_1 z + a_0. \tag{5.3.3.1}$$

If R is sufficiently large (say $R > \max\{1, n \cdot \max_{0 \le j \le n-1} |a_j|\}$) and $|z| = R$, then

$$\frac{|a_{n-1}z^{n-1} + a_{n-2}z^{n-2} + \cdots + a_0|}{|z^n|} < 1. \tag{5.3.3.2}$$

Thus Rouché's theorem applies on $\overline{D}(0, R)$ with $f(z) = z^n$ and $g(z) = P(z)$. We conclude that the number of zeros of $P(z)$ inside $D(0, R)$, counting multiplicities, is the same as the number of zeros of z^n inside $D(0, R)$, counting multiplicities—namely n. Thus we recover the Fundamental Theorem of Algebra. Incidentally, this example underlines the importance of counting zeros with multiplicities: the function z^n has only one root in the naïve sense of counting the number of points where it is zero; but it has n roots when they are counted with multiplicity.

5.3.4 Hurwitz's Theorem

A second useful consequence of the argument principle is the following result about the limit of a sequence of zero-free holomorphic functions:

> **Hurwitz's theorem** Suppose that $U \subseteq \mathbb{C}$ is a connected open set and that $\{f_j\}$ is a sequence of nowhere-vanishing holomorphic functions on U. If the sequence $\{f_j\}$ converges uniformly on compact subsets of U to a (necessarily holomorphic) limit function f_0, then either f_0 is nowhere-vanishing or $f_0 \equiv 0$.

We leave the proof to the reader: Examine the integral

$$\frac{1}{2\pi i} \oint_{\partial D(P,r)} \frac{f_j'(\zeta)}{f_j(\zeta)}\, d\zeta$$

for a suitable disc in the common domain of the functions in question.

5.4 The Maximum Principle

5.4.1 The Maximum Modulus Principle

We repeat that a *domain* in $\mathbb{C}$ is a connected open set (§§1.3.1). A *bounded domain* is a connected open set U such that there is an $R > 0$ with $|z| < R$ for all $z \in U$—or $U \subseteq D(0, R)$.

The Maximum Modulus Principle

> **Theorem:** Let $U \subseteq \mathbb{C}$ be a domain. Let f be a holomorphic function on U. If there is a point $P \in U$ such that $|f(P)| \geq |f(z)|$ for all $z \in U$, then f is constant.

Here is a sharper variant of the theorem:

> **Theorem:** Let $U \subseteq \mathbb{C}$ be a domain and let f be a holomorphic function on U. If there is a point $P \in U$ at which $|f|$ has a *local maximum*, then f is constant.

There are a variety of ways to prove the maximum principle. A standard method is to first establish this mean value property: If f is holomorphic in a neighborhood of $\overline{D}(P, r)$ then

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

One establishes this formula by first taking $P = 0$ and checking the result for the case $f(z) = z^k$, any $k \geq 0$. Then one extends to an arbitrary holomorphic function by using power series.

Now to establish the maximum principle, assume that f is not constant as asserted. Let E be the set on which $|f|$ assumes its maximum value $|f(P)| \equiv \lambda$. Let w be a point of E that is nearest to ∂U. If $r > 0$ is a small number then

$$\lambda = |f(w)| = \frac{1}{2\pi} \left| \int_0^{2\pi} f(w + re^{it})\, dt \right| \leq \frac{1}{2\pi} \int_0^{2\pi} |f(w + re^{it})|\, dt\,.$$

Now because w is an extreme point of E, it is *not* the case that $w + re^{it}$ lies in E for all values of t. As a result, $|f(w + re^{it})| < \lambda$ for t in an open arc of $[0, 2\pi]$. Thus we may conclude that the last line is

$$< \frac{1}{2\pi} \int_0^{2\pi} \lambda\, dt = \lambda\,.$$

We conclude that $\lambda < \lambda$, and that is a contradiction.

We leave the details of the sharper version of the maximum principle for the interested reader.

5.4.2 Boundary Maximum Modulus Theorem

The following version of the maximum principle is intuitively appealing, and is frequently useful.

> **Theorem:** Let $U \subseteq \mathbb{C}$ be a bounded domain. Let f be a continuous function on $\overline{U}$ that is holomorphic on U. Then the maximum value of $|f|$ on $\overline{U}$ (which must occur, since $\overline{U}$ is closed and bounded—see [RUD1], [KRA3]) must in fact occur on ∂U.
>
> In other words,
> $$\max_{\overline{U}} |f| = \max_{\partial U} |f|.$$

The proof is straightforward and we omit it.

5.4.3 The Minimum Principle

Holomorphic functions (or, more precisely, their moduli) *can* have interior minima. The function $f(z) = z^2$ on $D(0,1)$ has the property that $z = 0$ is a global minimum for $|f|$. However, it is not accidental that this minimum value is 0:

> **Theorem:** Let f be holomorphic on a domain $U \subseteq \mathbb{C}$. Assume that f never vanishes. If there is a point $P \in U$ such that $|f(P)| \leq |f(z)|$ for all $z \in U$, then f is constant. This result is proved by applying the maximum principle to the function $1/f$.

There is also a boundary minimum principle:

> **Theorem:** Let $U \subseteq \mathbb{C}$ be a bounded domain. Let f be a continuous function on $\overline{U}$ that is holomorphic on U. Assume that f never vanishes on $\overline{U}$. Then the minimum value of $|f|$ on $\overline{U}$ (which must occur, since $\overline{U}$ is closed and bounded—see [RUD1], [KRA3]) must occur on ∂U.
>
> In other words,
> $$\min_{\overline{U}} |f| = \min_{\partial U} |f|.$$

5.4.4 The Maximum Principle on an Unbounded Domain

It should be noted that the boundary maximum modulus theorem is not always true on an unbounded domain. The standard example is the function $f(z) = \exp(\exp(z))$ on the domain $U = \{z = x + iy : -\pi/2 < y < \pi/2\}$. Check for yourself that $|f| = 1$ on the boundary of U. But the restriction of f to the real number line is unbounded at infinity. The theorem does, however, remain true with some additional restrictions. The result, known as the Phragmen-Lindelöf theorem,[1] is one method of treating maximum modulus theorems on unbounded domains (see [RUD2]).

5.5 The Schwarz Lemma

This section treats certain estimates that bounded holomorphic functions on the unit disc necessarily satisfy. We present the classical, analytic viewpoint in the subject (instead of the geometric viewpoint—see [KRA2]).

5.5.1 Schwarz's Lemma

Theorem: Let f be holomorphic on the unit disc. Assume that

(5.5.1.1) $|f(z)| \leq 1$ for all z.

(5.5.1.2) $f(0) = 0$.

Then $|f(z)| \leq |z|$ and $|f'(0)| \leq 1$.

If either $|f(z)| = |z|$ for some $z \neq 0$ or if $|f'(0)| = 1$, then f is a rotation: $f(z) \equiv \alpha z$ for some complex constant α of unit modulus.

For the proof, consider the function $g(z) = f(z)/z$. This function is holomorphic and is still bounded by 1 as $|z| \to 1$. The result then follows from the maximum principle.

Schwarz's lemma is a profound geometric fact that has exerted considerable influence in the subject. We cannot explore these avenues here, but

[1] This theorem imposes a Tauberian hypothesis on the function to make up for the fact that the domain is unbounded.

see [KRA2]. One nice application is that the lemma enables one to classify the invertible holomorphic self-maps of the unit disc (see [GK]). (Here a *self-map* of a domain U is a mapping $F : U \to U$ of the domain to itself.) These are commonly referred to as the "conformal self-maps" of the disc. The classification is as follows: If $0 \leq \theta < 2\pi$, then define the *rotation through angle* θ to be the function $\rho_\theta(z) = e^{i\theta}z$; if a is a complex number of modulus less than one, then define the associated *Möbius transformation* to be $\varphi_a(z) = [z - a]/[1 - \overline{a}z]$. Any conformal self-map of the disc is the composition of some rotation ρ_θ with some Möbius transformation φ_a. This topic is treated in detail in §6.2.

The classification works as follows. Let ψ be a conformal self-map of the disc D. Suppose that $\psi(0) = a$. Then consider $h = \varphi_a \circ \psi$. We see that $h : D \to D$ and $h(0) = 0$. Thus the Schwarz lemma applies and $|h(z)| \leq |z|$. The same reasoning applies to h^{-1} so that $|h(z)| \geq |z|$. We conclude that $|h(z)| = |z|$ so that h is a rotation. The result follows.

We conclude this section by presenting a generalization of the Schwarz lemma, in which we consider holomorphic mappings $f : D \to D$, but we discard the hypothesis that $f(0) = 0$. This result is known as the Schwarz-Pick lemma.

5.5.2 The Schwarz-Pick Lemma

Theorem: Let f be holomorphic on the unit disc. Assume that

(5.5.2.1) $|f(z)| \leq 1$ for all z.

(5.5.2.2) $f(a) = b$ for some $a, b \in D(0, 1)$.

Then

$$|f'(a)| \leq \frac{1 - |b|^2}{1 - |a|^2}. \tag{5.5.2.3}$$

Moreover, if $f(a_1) = b_1$ and $f(a_2) = b_2$, then

$$\left|\frac{b_2 - b_1}{1 - \overline{b}_1 b_2}\right| \leq \left|\frac{a_2 - a_1}{1 - \overline{a}_1 a_2}\right|. \tag{5.5.2.4}$$

There is a "uniqueness" result in the Schwarz-Pick Lemma. If either

$$|f'(a)| = \frac{1-|b|^2}{1-|a|^2} \qquad \text{or} \qquad \left|\frac{b_2-b_1}{1-\overline{b}_1 b_2}\right| = \left|\frac{a_2-a_1}{1-\overline{a}_1 a_2}\right| \quad \text{with } a_1 \neq a_2\,, \tag{5.5.2.5}$$

then the function f is a conformal self-mapping (one-to-one, onto holomorphic function) of $D(0,1)$ to itself.

The proof of Schwarz-Pick is nearly obvious. Consider $h = \varphi_b \circ f \circ \varphi_{-a}$. Then $h : D \to D$, $h(0) = 0$, and the Schwarz lemma applies.

Chapter 6

The Geometric Theory of Holomorphic Functions

6.1 The Idea of a Conformal Mapping

6.1.1 Conformal Mappings

The main objects of study in this chapter are holomorphic functions $h : U \to V$, with U and V open in $\mathbb{C}$, that are one-to-one and onto. Such a holomorphic function is called a *conformal* (or *biholomorphic*) mapping. The fact that h is supposed to be one-to-one implies that h' is nowhere zero on U [remember that if h' vanishes to order $k \geq 0$ at a point $P \in U$, then h is $(k+1)$-to-1 in a small neighborhood of P—see §§5.2.1]. As a result, $h^{-1} : V \to U$ is also holomorphic—as we discussed in §§5.2.1. A conformal map $h : U \to V$ from one open set to another can be used to transfer holomorphic functions on U to V and vice versa: that is, $f : V \to \mathbb{C}$ is holomorphic if and only if $f \circ h$ is holomorphic on U; and $g : U \to \mathbb{C}$ is holomorphic if and only if $g \circ h^{-1}$ is holomorphic on V.

Thus, if there is a conformal mapping from U to V, then U and V are essentially indistinguishable from the viewpoint of complex function theory. On a practical level, one can often study holomorphic functions on a rather complicated open set by first mapping that open set to some simpler open set, then transferring the holomorphic functions as indicated.

6.1.2 Conformal Self-Maps of the Plane

The simplest open subset of $\mathbb{C}$ is $\mathbb{C}$ itself. Thus it is natural to begin our study of conformal mappings by considering the conformal mappings of $\mathbb{C}$ to itself. In fact the conformal mappings from $\mathbb{C}$ to $\mathbb{C}$ can be explicitly described as follows:

> **Theorem:** A function $f : \mathbb{C} \to \mathbb{C}$ is a conformal mapping if and only if there are complex numbers a, b with $a \neq 0$ such that

$$f(z) = az + b \quad , \quad z \in \mathbb{C}. \tag{6.1.2.1}$$

One aspect of the result is fairly obvious: If $a, b \in \mathbb{C}$ and $a \neq 0$, then the map $z \mapsto az + b$ is certainly a conformal mapping of $\mathbb{C}$ to $\mathbb{C}$. In fact one checks easily that $z \mapsto (z - b)/a$ is the inverse mapping. The interesting part of the assertion is that these are in fact the only conformal maps of $\mathbb{C}$ to $\mathbb{C}$.

For the proof, note that the holomorphic function f satisfies

$$\lim_{|z| \to +\infty} |f(z)| = +\infty.$$

That is, given $\epsilon > 0$, there is a number $C > 0$ such that if $|z| > C$ then $|f(z)| > 1/\epsilon$. The set $\{z : |z| \leq 1/\epsilon\}$ is a compact subset of $\mathbb{C}$. Since $f^{-1} : \mathbb{C} \to \mathbb{C}$ is holomorphic, it is continuous. And the continuous image of a compact set is compact. Therefore $S = f^{-1}\big(\{z : |z| \leq 1/\epsilon\}\big)$ is compact. By the Heine-Borel theorem, S must be bounded. Thus there is a positive number C such that $S \subseteq \{z : |z| \leq C\}$.

Taking contrapositives, we see that if $|w| > C$ then w is not an element of $f^{-1}\big(\{z : |z| \leq 1/\epsilon\}\big)$. Therefore $f(w)$ is not an element of $\{z : |z| \leq 1/\epsilon\}$. In other words, $|f(w)| > 1/\epsilon$. That is the desired result.

Further understanding of the behavior of $f(z)$ when z has large absolute value may be obtained by applying the technique already used in Chapter 4 to talk about singularities at ∞. Define, for all $z \in \mathbb{C}$ such that $z \neq 0$ and $f(1/z) \neq 0$, a function $g(z) = 1/f(1/z)$. By Lemma x.y.z, there is a number C such that $|f(z)| > 1$ if $|z| > C$. Clearly g is defined on $\{z : 0 < |z| < 1/C\}$. Furthermore, g is bounded on this set by 1. By the Riemann removable singularities theorem, g extends to be holomorphic on the full disc $D(0, 1/C) = \{z : |z| < 1/C\}$. Lemma x.y.z tells us that in fact $g(0) = 0$.

Now, because $f : \mathbb{C} \to \mathbb{C}$ is one-to-one, it follows that g is one-to-one on the disc $D(0, 1/C)$. In particular, $g'(0)$ cannot be 0. Since

$$0 \neq |g'(0)| = \lim_{|z|\to 0^+} \left| \frac{g(z) - g(0)}{z} \right| = \lim_{|z|\to 0^+} \left| \frac{g(z)}{z} \right| ,$$

we see that there is a constant $A > 0$ such that

$$|g(z)| \geq A|z|$$

for z sufficiently small. We next translate this to information about the original function f.

Now we claim that there are numbers $B, D > 0$ such that, if $|z| > D$, then

$$|f(z)| < B|z|\,.$$

To see this, as noted above, there is a number $\delta > 0$ such that if $|z| < \delta$ then $|g(z)| \geq A|z|$. If $|z| > 1/\delta$ then

$$|f(z)| = \frac{1}{|g(1/z)|} \leq \frac{1}{A|1/z|} = \frac{1}{A}|z|.$$

Thus the lemma holds with $B = 1/A$ and $D = 1/\delta$.

The proof of our main result is now easily given. By (6.1.2.1), f is a polynomial of degree at most 1, i.e. $f(z) = az + b$ for some $a, b \in \mathbb{C}$. Clearly f is one-to-one and onto if and only if $a \neq 0$. That proves the result.

One part of the proof just given is worth considering in the more general context of singularities at ∞, as discussed in Section 4.7. Suppose now that h is holomorphic on a set $\{z : |z| > \alpha\}$, for some positive α, and that

$$\lim_{|z|\to+\infty} |h(z)| = +\infty.$$

Then it remains true that $g(z) \equiv 1/h(1/z)$ is defined and holomorphic on $\{z : 0 < |z| < \eta\}$, some $\eta > 0$. Also, by the same reasoning as above, g extends holomorphically to $D(0, \eta)$ with $g(0) = 0$. If we do not assume in advance that h is one-to-one then we may not say (as we did before) that $g'(0) \neq 0$. But g is not constant, since h is not, so there is a positive integer n and a positive number A such that

$$|g(z)| \geq A|z|^n$$

for all z with $|z|$ sufficiently small. It then follows (as in the proof of Lemma 6.1.2) that

$$|h(z)| \leq \frac{1}{A}|z|^n$$

for $|z|$ sufficiently large. This line of reasoning, combined with Theorem 3.4.4, recovers Theorem 4.7.6: If $h : \mathbb{C} \to \mathbb{C}$ is a holomorphic function such that

$$\lim_{|z|\to+\infty} |h(z)| = +\infty,$$

then h is a polynomial.

A generalization of this result about conformal maps of the plane is the following (consult §§4.6.8 as well as the detailed explanation in [GRK]):

> If $h : \mathbb{C} \to \mathbb{C}$ is a holomorphic function such that
>
> $$\lim_{|z|\to+\infty} |h(z)| = +\infty,$$
>
> then h is a polynomial.

We in fact treated this result, using slightly different terminology, when we discussed isolated singularities at infinity.

6.2 Linear Fractional Transformations

6.2.1 Linear Fractional Mappings

The automorphisms (that is, conformal self-mappings) of the unit disc D are special cases of functions of the form

$$z \mapsto \frac{az+b}{cz+d} \quad , \quad a, b, c, d \in \mathbb{C}. \tag{6.3.1.1}$$

It is worthwhile to consider functions of this form in generality. One restriction on this generality needs to be imposed, however; if $ad - bc = 0$, then the numerator is a constant multiple of the denominator provided that the denominator is not identically zero. So if $ad - bc = 0$, then the function is either constant or has zero denominator and is nowhere defined. Thus only the case $ad - bc \neq 0$ is worth considering in detail.

A function of the form

$$z \mapsto \frac{az+b}{cz+d} \ , \quad ad-bc \neq 0, \tag{6.3.1.2}$$

is called a *linear fractional transformation.*

Note that $(az+b)/(cz+d)$ is not necessarily defined for all $z \in \mathbb{C}$. Specifically, if $c \neq 0$, then it is undefined at $z = -d/c$. In case $c \neq 0$,

$$\lim_{z \to -d/c} \left| \frac{az+b}{cz+d} \right| = +\infty. \tag{6.3.1.3}$$

This observation suggests that one might well, for linguistic convenience, adjoin formally a "point at ∞" to $\mathbb{C}$ and consider the value of $(az+b)/(cz+d)$ to be ∞ when $z = -d/c \, (c \neq 0)$. Thus we will think of both the domain and the range of our linear fractional transformation to be $\mathbb{C} \cup \{\infty\}$ (we sometimes also use the notation $\widehat{\mathbb{C}}$ instead of $\mathbb{C} \cup \{\infty\}$). Specifically, we are led to the following alternative method for describing a linear fractional transformation.

A function $f : \mathbb{C} \cup \{\infty\} \to \mathbb{C} \cup \{\infty\}$ is a *linear fractional transformation* if there exists $a, b, c, d \in \mathbb{C}, ad - bc \neq 0$, such that either

(6.3.1.4) $c = 0, d \neq 0, f(\infty) = \infty$, and $f(z) = (a/d)z + (b/d)$ for all $z \in \mathbb{C}$;

or

(6.3.1.5) $c \neq 0, f(\infty) = a/c, f(-d/c) = \infty$, and $f(z) = (az+b)/(cz+d)$ for all $z \in \mathbb{C}, z \neq -d/c$.

It is important to realize that, as before, the status of the point ∞ is entirely formal: we are just using it as a linguistic convenience, to keep track of the behavior of $f(z)$ both where it is not defined as a map on $\mathbb{C}$ and to keep track of its behavior when $|z| \to +\infty$. The justification for the particular devices used is the fact that

(6.3.1.6) $\lim_{|z| \to +\infty} f(z) = f(\infty)$ $\quad$ [$c = 0$; case **(6.3.1.4)** of the definition]

(6.3.1.7) $\lim_{z \to -d/c} |f(z)| = +\infty$ $\quad$ [$c \neq 0$; case **(6.3.1.5)** of the definition].

6.2.2 The Topology of the Extended Plane

The limit properties of f that we described in §§6.3.1 can be considered as continuity properties of f from $\mathbb{C} \cup \{\infty\}$ to $\mathbb{C} \cup \{\infty\}$ using the definition of continuity that comes from the topology on $\mathbb{C} \cup \{\infty\}$. It is easy to formulate that topology in terms of open sets. But it is also convenient to formulate that same topological structure in terms of convergence of sequences:

A sequence $\{p_i\}$ in $\mathbb{C} \cup \{\infty\}$ *converges to* $p_0 \in \mathbb{C} \cup \{\infty\}$ (notation $\lim_{i\to\infty} p_i = p_0$) if either

(6.3.2.1) $p_0 = \infty$ and $\lim_{i\to+\infty} |p_i| = +\infty$ where the limit is taken for all i such that $p_i \in \mathbb{C}$;

or

(6.3.2.2) $p_0 \in \mathbb{C}$, all but a finite number of the p_i are in $\mathbb{C}$ and $\lim_{i\to\infty} p_i = p_0$ in the usual sense of convergence in $\mathbb{C}$.

6.2.3 The Riemann Sphere

Stereographic projection puts $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ into one-to-one correspondence with the two-dimensional sphere S in $\mathbb{R}^3, S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$ in such a way that topology is preserved in both directions of the correspondence.

In detail, begin by imagining the unit sphere bisected by the complex plane with the center of the sphere $(0, 0, 0)$ coinciding with the origin in the plane—see Figure 6.1. We define the stereographic projection as follows: If $P = (x, y) \in \mathbb{C}$, then connect P to the "north pole" N of the sphere with a line segment. The point $\pi(P)$ of intersection of this segment with the sphere is called the *stereographic projection* of P. Note that, under stereographic projection, the "point at infinity" in the plane corresponds to the north pole N of the sphere. For this reason, $\mathbb{C} \cup \{\infty\}$ is often thought of as "being" a sphere, and is then called, for historical reasons, the *Riemann sphere*.

The construction we have just described is another way to think about the "extended complex plane"—see §§6.3.2. In these terms, linear fractional transformations become homeomorphisms of $\mathbb{C} \cup \{\infty\}$ to itself. (Recall that a *homeomorphism* is, by definition, a one-to-one, onto, continuous mapping with a continuous inverse.)

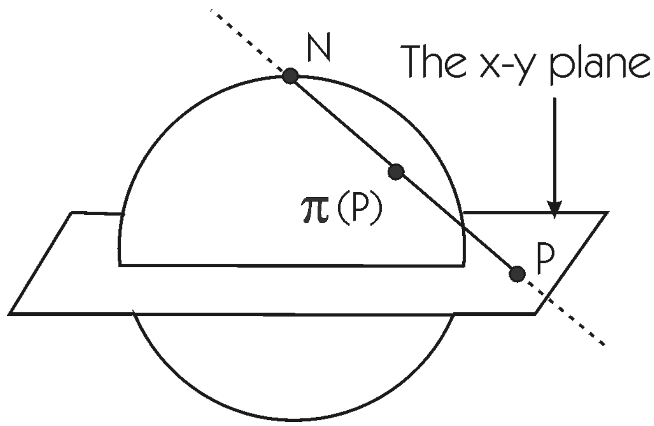

Figure 6.1: Stereographic projection.

> **Proposition:** If $f : \mathbb{C} \cup \{\infty\} \to \mathbb{C} \cup \{\infty\}$ is a linear fractional transformation, then f is a one-to-one, onto, continuous function. Also, $f^{-1} : \mathbb{C} \cup \{\infty\} \to \mathbb{C} \cup \{\infty\}$ is a linear fractional transformation, and is thus a one-to-one, onto, continuous function.

> **Proposition:** If $g : \mathbb{C} \cup \{\infty\} \to \mathbb{C} \cup \{\infty\}$ is also a linear fractional transformation, then $f \circ g$ is a linear fractional transformation.

The simplicity of language obtained by adjoining ∞ to $\mathbb{C}$ (so that the composition and inverse properties of linear fractional transformations obviously hold) is well worth the trouble. Certainly one does not wish to consider the multiplicity of special possibilities when composing $(Az + B)/(Cz + D)$ with $(az + b)/(cz + d)$ (namely $c = 0, c \neq 0, aC + cD \neq 0, aC + cD = 0$, etc.) that arise every time composition is considered.

In fact, it is worth summarizing what we have learned in a theorem (see §§6.3.4). First note that it makes sense now to talk about a homeomorphism from $\mathbb{C} \cup \{\infty\}$ to $\mathbb{C} \cup \{\infty\}$ being conformal: this just means that it (and hence its inverse) are holomorphic in our extended sense. If φ is a conformal map of $\mathbb{C} \cup \{\infty\}$ to itself, then, after composing with a linear fractional transformation, we may suppose that φ maps ∞ to itself. Thus φ, after composition with a linear fraction transformation, is linear. It follows that φ itself is linear fractional. The following result summarizes the situation:

6.2.4 Conformal Self-Maps of the Riemann Sphere

> **Theorem:** A function φ is a conformal self-mapping of $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ to itself if and only if φ is linear fractional.

We turn now to the actual utility of linear fractional transformations (beyond their having been the form of automorphisms of D—see §§6.2.1–6.2.3—and the form of all conformal self maps of $\mathbb{C} \cup \{\infty\}$ to itself in the present section). One of the most frequently occurring uses is the following:

6.2.5 The Cayley Transform

Theorem (The Cayley Transform): The linear fractional transformation $z \mapsto (i - z)/(i + z)$ maps the upper half plane $\{z : \mathrm{Im} z > 0\}$ conformally onto the unit disc $D = \{z : |z| < 1\}$.

6.2.6 Generalized Circles and Lines

Calculations of the type that we have been discussing are straightforward but tedious. It is thus worthwhile to seek a simpler way to understand what the image under a linear fractional transformation of a given region is. For regions bounded by line segments and arcs of circles the following result gives a method for addressing this issue:

Let $\mathcal{C}$ be the set of subsets of $\mathbb{C} \cup \{\infty\}$ consisting of (i) circles and (ii) sets of the form $L \cup \{\infty\}$ where L is a line in $\mathbb{C}$. We call the elements of $\mathcal{C}$ "generalized circles." Then every linear fractional transformation φ takes elements of $\mathcal{C}$ to elements of $\mathcal{C}$. One verifies this last assertion by noting that any linear fractional transformation is the composition of dilations, translations, and the inversion map $z \mapsto 1/z$; and each of these component maps clearly sends generalized circles to generalized circles.

6.2.7 The Cayley Transform Revisited

To illustrate the utility of this last result, we return to the Cayley transformation

$$z \mapsto \frac{i - z}{i + z}. \tag{6.3.7.1}$$

Under this mapping the point ∞ is sent to -1, the point 1 is sent to $(i - 1)/(i + 1) = i$, and the point -1 is sent to $(i - (-1))/(i + (-1)) = -i$.

Thus the image under the Cayley transform (a linear fractional transformation) of three points on $\mathbb{R} \cup \{\infty\}$ contains three points on the unit circle. Since three points determine a (generalized) circle, and since linear fractional transformations send generalized circles to generalized circles, we may conclude that the Cayley transform sends the real line to the unit circle. Now the Cayley transform is one-to-one and onto from $\mathbb{C} \cup \{\infty\}$ to $\mathbb{C} \cup \{\infty\}$. By continuity, it either sends the upper half plane to the (open) unit disc or to the complement of the closed unit disc. The image of i is 0, so in fact the Cayley transform sends the upper half plane to the unit disc.

6.2.8 Summary Chart of Linear Fractional Transformations

The next chart summarizes the properties of some important linear fractional transformations. Note that $U = \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$ is the upper half-plane and $D = \{z \in \mathbb{C} : |z| < 1\}$ is the unit disc; the domain variable is z and the range variable is w.

Linear Fractional Transformations

Domain	**Image**	**Conditions**	**Formula**
$z \in \widehat{\mathbb{C}}$	$w \in \widehat{\mathbb{C}}$		$w = \frac{az+b}{cz+d}$
$z \in D$	$w \in U$		$w = i \cdot \frac{1-z}{1+z}$
$z \in U$	$w \in D$		$w = \frac{i-z}{i+z}$
$z \in D$	$w \in D$		$w = \frac{z-a}{1-\overline{a}z}$, $\|a\| < 1$
$\mathbb{C}$	$\mathbb{C}$	$L(z_1) = w_1$ $L(z_2) = w_2$ $L(z_3) = w_3$	$L(z) = S^{-1} \circ T$ $T(z) = \frac{z-z_1}{z-z_3} \cdot \frac{z_2-z_3}{z_2-z_1}$ $S(m) = \frac{m-w_1}{m-w_3} \cdot \frac{w_2-w_3}{w_2-w_1}$

6.3 The Riemann Mapping Theorem

6.3.1 The Concept of Homeomorphism

Two open sets U and V in $\mathbb{C}$ are *homeomorphic* if there is a one-to-one, onto, continuous function $f : U \to V$ with $f^{-1} : V \to U$ also continuous. Such a function f is called a *homeomorphism* from U to V (see also §§6.3.3).

6.3.2 The Riemann Mapping Theorem

The Riemann mapping theorem, sometimes called the greatest theorem of the nineteenth century, asserts in effect that any planar domain (other than the entire plane itself) that has the topology of the unit disc also has the conformal structure of the unit disc. Even though this theorem has been subsumed by the great uniformization theorem of Köbe (see [FAK]), it is still striking in its elegance and simplicity:

> If U is an open subset of $\mathbb{C}, U \neq \mathbb{C}$, and if U is homeomorphic to D, then U is conformally equivalent to D. That is, there is a holomorphic mapping $\psi : U \to D$ which is one-to-one and onto.

6.3.3 The Riemann Mapping Theorem: Second Formulation

An alternative formulation of this theorem uses the concept of "simply connected" (see also §§2.3.3.). We say that a connected open set U in the complex plane is simply connected if any closed curve in U can be continuously deformed to a point. (This is just a precise way of saying that U has no holes. Yet another formulation of the notion is that the complement of U has only one connected component.)

> **Theorem:** If U is an open subset of $\mathbb{C}, U \neq \mathbb{C}$, and if U is simply connected, then U is conformally equivalent to D.

The proof of the Riemann mapping theorem is long and complex and introduces many fundamentally new ideas and techniques. We cannot treat it in detail here, but see [GRK]. Certainly the ideas introduced in this proof have been profoundly influential.

KEY STEPS IN THE PROOF OF THE RIEMANN MAPPING THEOREM:

Fix a point $P \in U$. Assume for simplicity that U is bounded.

1. Consider
$$\mathcal{S} = \{f : f \text{ maps } D \text{ to } U, f(0) = P\};$$

2. Define
$$\alpha = \sup_{f \in \mathcal{S}} |f(0)| .$$

3. Use a normal families argument (Montel's theorem) to show that there is a function $f_0 \in \mathcal{S}$ so that $|f_0'(0)| = \alpha$.

4. Show, using the argument principle, that f_0 must be univalent.

5. Show, with a clever proof by contradiction, that if f_0 is not onto then it cannot be the solution of the extremal problem enunciated in 3.

6. The function f_0 is the conformal map that we seek.

The full details of the proof of the Riemann mapping theorem appear in [GRK].

6.4 Conformal Mappings of Annuli

6.4.1 A Riemann Mapping Theorem for Annuli

The Riemann mapping theorem tells us that, from the point of view of complex analysis, there are only two simply connected planar domains: the disc and the plane. Any other simply connected region is biholomorphic to one of these. It is natural then to ask about domains with holes. Take, for example, a domain U with precisely one hole. Is it conformally equivalent to an annulus?

Note that if $c > 0$ is a constant, then for any $R_1 < R_2$ the annuli

$$A_1 \equiv \{z : R_1 < |z| < R_2\} \quad \text{and} \quad A_2 \equiv \{z : cR_1 < |z| < cR_2\} \tag{6.5.1.1}$$

are biholomorphically equivalent under the mapping $z \mapsto cz$. The surprising fact that we shall learn is that these are the *only* circumstances under which two annuli are equivalent:

6.4.2 Conformal Equivalence of Annuli

Let

$$A_1 = \{z \in \mathbb{C} : 1 < |z| < R_1\} \tag{6.5.2.1}$$

and

$$A_2 = \{z \in \mathbb{C} : 1 < |z| < R_2\}. \tag{6.5.2.2}$$

Then A_1 is conformally equivalent to A_2 if and only if $R_1 = R_2$.

A perhaps more striking result, and more difficult to prove, is this:

> Let $U \subseteq \mathbb{C}$ be any bounded domain with *one hole*—this means that the complement of U has two connected components, one bounded and one not. Then U is conformally equivalent to some annulus.

See [AHL] as well as [KRA4] for a discursive discussion of this result.

6.4.3 Classification of Planar Domains

The classification of planar domains up to biholomorphic equivalence is a part of the theory of Riemann surfaces. For now, we comment that one of the startling classification theorems (a generalization of the Riemann mapping theorem) is that any bounded planar domain with finitely many "holes" is conformally equivalent to the unit disc with finitely many closed circular arcs, coming from circles centered at the origin, removed. (Here a "hole" in the present context means a bounded, connected component of the complement of the domain in $\mathbb{C}$, a concept which coincides with the intuitive idea of a hole.) An alternative equivalent statement is that any bounded planar domain with finitely many holes is conformally equivalent to the plane with finitely many vertical slits (see [AHL]). The analogous result for domains with infinitely many holes is known to be true when the number of holes is countable (see [HES]).

Chapter 7

Harmonic Functions

7.1 Basic Properties of Harmonic Functions

7.1.1 The Laplace Equation

We reiterate the definition of "harmonic". Let F be a holomorphic function on an open set $U \subseteq \mathbb{C}$. Write $F = u + iv$, where u and v are real-valued. The real part u satisfies a certain partial differential equation known as *Laplace's equation*:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) u = 0. \tag{7.1.1.1}$$

(Of course the imaginary part v satisfies the same equation.) In this chapter we shall study systematically those C^2 functions that satisfy this equation. They are called *harmonic* functions. (Note that we encountered some of these ideas already in §1.4.)

7.1.2 Definition of Harmonic Function

Recall the precise definition of harmonic function:

A real-valued function $u : U \to \mathbb{R}$ on an open set $U \subseteq \mathbb{C}$ is *harmonic* if it is C^2 on U and

$$\Delta u \equiv 0, \tag{7.1.2.1}$$

where the Laplacian Δu is defined by

$$\Delta u = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) u. \tag{7.1.2.2}$$

7.1.3 Real- and Complex-Valued Harmonic Functions

The definition of harmonic function just given applies as well to complex-valued functions. A complex-valued function is harmonic if and only if its real and imaginary parts are each harmonic.

The first thing that we need to check is that real-valued harmonic functions are just those functions that arise as the real parts of holomorphic functions—at least locally.

7.1.4 Harmonic Functions as the Real Parts of Holomorphic Functions

If $u : D(P, r) \to \mathbb{R}$ is a harmonic function on a disc $D(P, r)$, then there is a holomorphic function $F : D(P, r) \to \mathbb{C}$ such that $\operatorname{Re} F \equiv u$ on $D(P, r)$. We already indicated in Section x.y that this result reduces to solving a coupled system of partial differential equations—a situation that can be handled with multi-variable calculus.

Note that v is uniquely determined by u except for an additive constant: the Cauchy-Riemann equations determine the partial derivatives of v and hence determine v up to an additive constant. One can also think of the determination, up to a constant, of v by u in another way: If $\widetilde{v}$ is another function such that $u+i\widetilde{v}$ is holomorphic, then $H \equiv i(v-\widetilde{v}) = (u+iv)-(u+i\widetilde{v})$ is a holomorphic function with zero real part; hence its image is not open. Thus H must be a constant, and v and $\widetilde{v}$ differ by a constant. Any (harmonic) function v such that $u + iv$ is holomorphic is called a *harmonic conjugate* of u (again see §§1.4.2).

> **Theorem:** If U is a simply connected open set (see §§6.4.3) and if $u : U \to \mathbb{R}$ is a harmonic function, then there is a C^2 (indeed a C^∞) harmonic function v such that $u+iv : U \to \mathbb{C}$ is holomorphic.

Another important relationship between harmonic and holomorphic functions is this:

> If $u : U \to \mathbb{R}$ is harmonic and if $H : V \to U$ is holomorphic, then $u \circ H$ is harmonic on V.

This result is proved by direct calculation (i.e., differentiation, using the chain rule).

7.1.5 Smoothness of Harmonic Functions

If $u : U \to \mathbb{R}$ is a harmonic function on an open set $U \subseteq \mathbb{C}$, then $u \in C^\infty$. In fact a harmonic function is always real analytic (has a local power series expansion in powers of x and y). This follows, for instance, because a harmonic function is locally the real part of a holomorphic function (see §§1.4.2, §§7.1.4). A holomorphic function has a power series expansion about each point, so is certainly infinitely differentiable.

7.2 The Maximum Principle and the Mean Value Property

7.2.1 The Maximum Principle for Harmonic Functions

Theorem: If $u : U \to \mathbb{R}$ is harmonic on a connected open set U and if there is a point $P_0 \in U$ with the property that $u(P_0) = \max_{z \in U} u(z)$, then u is constant on U. Compare the maximum modulus principle for holomorphic functions in §§5.4.1. The proof, using (7.2.4.1) below, is essentially the same as that for the maximum principle for holomorphic functions.

7.2.2 The Minimum Principle for Harmonic Functions

Theorem: If $u : U \to \mathbb{R}$ is a harmonic function on a connected open set $U \subseteq \mathbb{C}$ and if there is a point $P_0 \in U$ such that $u(P_0) = \min_{Q \in U} u(Q)$, then u is constant on U. Compare the minimum principle for holomorphic functions in §§5.4.3.

The reader may note that the minimum principle for holomorphic functions requires an extra hypothesis (i.e., non-vanishing of the function) while that for harmonic functions does not. The difference may be explained by noting that with harmonic functions we are considering the real-valued function u, while with holomorphic functions we must restrict attention to the modulus function $|f|$.

7.2.3 The Boundary Maximum and Minimum Principles

An important and intuitively appealing consequence of the maximum principle is the following result (which is sometimes called the "boundary maximum principle"). Recall that a continuous function on a compact set assumes a maximum value. When the function is harmonic, the maximum occurs at the boundary in the following precise sense:

> **Theorem:** Let $U \subseteq \mathbb{C}$ be a bounded domain. Let u be a continuous, real-valued function on the closure $\overline{U}$ of U that is harmonic on U. Then

$$\max_{\overline{U}} u = \max_{\partial U} u. \tag{7.2.3.1}$$

The analogous result for the minimum is:

> **Theorem:** Let $U \subseteq \mathbb{C}$ be a domain and let u be a continuous function on the closure $\overline{U}$ of U that is harmonic on U. Then

$$\min_{\overline{U}} u = \min_{\partial U} u. \tag{7.2.3.2}$$

Compare the analogous results for holomorphic functions in §§5.4.2, 5.4.3.

7.2.4 The Mean Value Property

Suppose that $u : U \to \mathbb{R}$ is a harmonic function on an open set $U \subseteq \mathbb{C}$ and that $\overline{D}(P, r) \subseteq U$ for some $r > 0$. Then

$$u(P) = \frac{1}{2\pi} \int_0^{2\pi} u(P + re^{i\theta})\, d\theta. \tag{7.2.4.1}$$

We will see in §§7.4.1 that the mean value property characterizes harmonic functions.

The mean value property for harmonic functions can be proved on the unit disc by first verifying it for the monomials z^j and $\overline{z}^j$ and then noting that any harmonic function is in the closed linear span of these. It can also be checked because a harmonic function is (locally) the real part of a holomorphic function and so the result can be derived (by taking the real part) from that for a holomorphic function. A very interesting proof may also be gotten from Green's theorem (see [KRA5]).

We conclude this subsection with two alternative formulations of the mean value property (MVP). Either one may be derived with simple changes of variable in the integral. In both cases, u, U, P, r are as above.

First Alternative Formulation of MVP

$$u(P) = \frac{1}{\pi r^2} \iint\limits_{D(P,r)} u(x,y)\, dx dy.$$

Second Alternative Formulation of MVP

$$u(P) = \frac{1}{2\pi r} \int_{\partial D(P,r)} u(\zeta)\, d\sigma(\zeta),$$

where $d\sigma$ is arc-length measure on $\partial D(P,r)$.

7.2.5 Boundary Uniqueness for Harmonic Functions

If $u_1 : \overline{D}(0,1) \to \mathbb{R}$ and $u_2 : \overline{D}(0,1) \to \mathbb{R}$ are two continuous functions, each of which is harmonic on $D(0,1)$ and if $u_1 = u_2$ on $\partial D(0,1) = \{z : |z| = 1\}$, then $u_1 \equiv u_2$. This assertion follows from the boundary maximum principle (§§7.2.3.1) applied to $u_1 - u_2$. Thus, in effect, a harmonic function u on $D(0,1)$ that extends continuously to $\overline{D}(0,1)$ is completely determined by its values on $\overline{D}(0,1) \setminus D(0,1) = \partial D(0,1)$. Of course an analogous result holds on any domain in $\mathbb{C}$.

7.3 The Poisson Integral Formula

7.3.1 The Poisson Integral

The next result builds on the boundary uniqueness idea. After all, if a harmonic function on the interior of a disc is completely determined by its boundary values, then we ought to be able to calculate the interior values from the boundary values. This is in fact what the Poisson integral formula does for us.

Let $u : U \to \mathbb{R}$ be a harmonic function on a neighborhood of $\overline{D}(0,1)$. Then, for any point $a \in D(0,1)$,

$$u(a) = \frac{1}{2\pi}\int_0^{2\pi} u(e^{i\psi}) \cdot \frac{1-|a|^2}{|a-e^{i\psi}|^2}\,d\psi. \tag{7.3.1.1}$$

7.3.2 The Poisson Kernel

The expression

$$\frac{1}{2\pi}\frac{1-|a|^2}{|a-e^{i\psi}|^2} \tag{7.3.2.1}$$

is called the *Poisson kernel* for the unit disc. It is often convenient to rewrite the formula we have just enunciated by setting $a = |a|e^{i\theta} = re^{i\theta}$. Then the result says that

$$u(re^{i\theta}) = \frac{1}{2\pi}\int_0^{2\pi} u(e^{i\psi})\frac{1-r^2}{1-2r\cos(\theta-\psi)+r^2}\,d\psi. \tag{7.3.2.2}$$

In other words

$$u(re^{i\theta}) = \int_0^{2\pi} u(e^{i\theta})P_r(\theta-\psi)\,d\psi, \tag{7.3.2.3}$$

where

$$P_r(\theta-\psi) = \frac{1}{2\pi}\frac{1-r^2}{1-2r\cos(\theta-\psi)+r^2}. \tag{7.3.2.4}$$

There are a number of ways to verify the Poisson integral formula. First, one could use Green's theorem (for which see again [KRA5]). Alternatively, one could first verify the formula with barehand calculation in the case that $f(z) = z^k$ or $f(z) = \overline{z}^k$. A third possibility is to note that formula (7.3.2.2) is plainly true at the origin (i.e., when $r = 0$). Then spread the result to the rest of the disc using conformal self-maps of the disc and the conformal invariance of harmonic functions. We leave the details to the reader.

7.3.3 The Dirichlet Problem

The Poisson integral formula both reproduces and creates harmonic functions. But, in contrast to the holomorphic case (§2.4), there is a simple connection between a continuous function f on $\partial D(0,1)$ and the created harmonic function u on D. The following theorem states this connection precisely. The theorem is usually called "the solution of the Dirichlet problem on the disc":

7.3.4 The Solution of the Dirichlet Problem on the Disc

Theorem: Let f be a continuous function on $\partial D(0,1)$. Define

$$u(z) = \begin{cases} \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi} f(e^{i\psi}) \cdot \dfrac{1-|z|^2}{|z-e^{i\psi}|^2}\, d\psi & \text{if} \quad z \in D(0,1) \\ f(z) & \text{if} \quad z \in \partial D(0,1). \end{cases} \tag{7.3.4.1}$$

Then u is continuous on $\overline{D}(0,1)$ and harmonic on $D(0,1)$. The proof of this theorem is rather technical, and we refer the reader to [GRK] or [KRA5].

Closely related to this result is the *reproducing property* of the Poisson kernel:

Theorem: Let u be harmonic on a neighborhood of $\overline{D}(0,1)$. Then, for $z \in D(0,1)$,

$$u(z) = \frac{1}{2\pi}\int_0^{2\pi} u(e^{i\psi}) \cdot \frac{1-|z|^2}{|z-e^{i\psi}|^2}\, d\psi\,. \tag{7.3.4.2}$$

See (7.3.1.1). One way to understand this last formula is to verify by hand that the Poisson kernel is harmonic in the variable z. Then (7.3.4.2) creates a function that agrees with u on the boundary of the disc and is harmonic inside, so it must be u itself.

7.3.5 The Dirichlet Problem on a General Disc

A change of variables shows that the results of §§7.3.4 remain true on a general disc. To wit, let f be a continuous function on $\partial D(P,r)$. Define

$$u(z) = \begin{cases} \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi} f(e^{i\psi}) \cdot \dfrac{R-|z-P|^2}{|(z-P)-Re^{i\psi}|^2}\, d\psi & \text{if} \quad z \in D(P,R) \\ f(z) & \text{if} \quad z \in \partial D(P,R). \end{cases} \tag{7.3.5.1}$$

Then u is continuous on $\overline{D}(P,R)$ and harmonic on $D(P,R)$.

If instead u is harmonic on a neighborhood of $\overline{D}(P,R)$, then, for $z \in D(P,R)$,

$$u(z) = \frac{1}{2\pi}\int_0^{2\pi} u(P+Re^{i\psi}) \cdot \frac{R^2 - |z-P|^2}{|(z-P) - Re^{i\psi}|^2}\, d\psi. \qquad (7.3.5.2)$$

7.4 Regularity of Harmonic Functions

7.4.1 The Mean Value Property on Circles

A continuous function $h : U \to \mathbb{R}$ on an open set $U \subseteq \mathbb{C}$ has the ϵ_P-*mean value property* if, for each point $P \in U$, there is an $\epsilon_P > 0$ such that $\overline{D}(P,\epsilon_P) \subseteq U$ and, for every $0 < \epsilon < \epsilon_P$,

$$h(P) = \frac{1}{2\pi}\int_0^{2\pi} h(P+\epsilon e^{i\theta})d\theta. \qquad (7.4.1.1)$$

The ϵ_P-mean value property allows the size of ϵ_P to vary arbitrarily with P.

Theorem: If $h : U \to \mathbb{R}$ is a continuous function on an open set U with the ϵ_P-mean value property, then h is harmonic.

Again, the proof of this result is too technical to treat here. See [GRK].

7.4.2 The Limit of a Sequence of Harmonic Functions

If $\{h_j\}$ is a sequence of real-valued harmonic functions that converges uniformly on compact subsets of U to a function $h : U \to \mathbb{R}$, then h is harmonic on U. This is immediate from the Poisson integral formula, since we now know that the Poisson kernel is harmonic.

7.5 The Schwarz Reflection Principle

7.5.1 Reflection of Harmonic Functions

We present in this section an application of what we have learned so far to a question of extension of a harmonic function to a larger domain. This will already illustrate the importance and power of harmonic function theory and will provide us with a striking result about holomorphic functions as well.

7.5.2 Schwarz Reflection Principle for Harmonic Functions

Let V be a connected open set in $\mathbb{C}$. Suppose that

$$V \cap (\text{real axis}) = \{x \in \mathbb{R} : a < x < b\}.$$

Set $U = \{z \in V : \operatorname{Im} z > 0\}$. Assume that $v : U \to \mathbb{R}$ is harmonic and that, for each $\zeta \in V \cap (\text{real axis})$,

$$\lim_{U \ni z \to \zeta} v(z) = 0. \tag{7.5.2.1}$$

Set $\widetilde{U} = \{\overline{z} : z \in U\}$. Define

$$\widehat{v}(z) = \begin{cases} v(z) & \text{if } \ z \in U \\ 0 & \text{if } \ z \in V \cap (\text{real axis}) \\ -v(\overline{z}) & \text{if } \ z \in \widetilde{U}. \end{cases} \tag{7.5.2.2}$$

Then $\widehat{v}$ is harmonic on $U^* \equiv U \cup \widetilde{U} \cup \{x \in \mathbb{R} : a < x < b\}$.

This result provides a way of extending a harmonic function from a given open set to a larger (reflected) open set. The method is known as the *Schwarz Reflection Principle*. One can think of $\widehat{U}$ as the reflection of U in the real axis, and the definition of $\widehat{v}$ on $\widehat{U}$ as the correspondingly appropriate idea of reflecting the function v. See Figure 7.1.

The proof of Schwarz reflection is a clever argument involving the symmetries of the Poisson integral formula. We refer the reader to [GRK] for the details.

7.5.3 The Schwarz Reflection Principle for Holomorphic Functions

Theorem: Let V be a connected open set in $\mathbb{C}$ such that $V \cap$ (the real axis) $= \{x \in \mathbb{R} : a < x < b\}$ for some $a, b \in \mathbb{R}$. Set $U = \{z \in V : \operatorname{Im} z > 0\}$. Suppose that $F : U \to \mathbb{C}$ is holomorphic and that

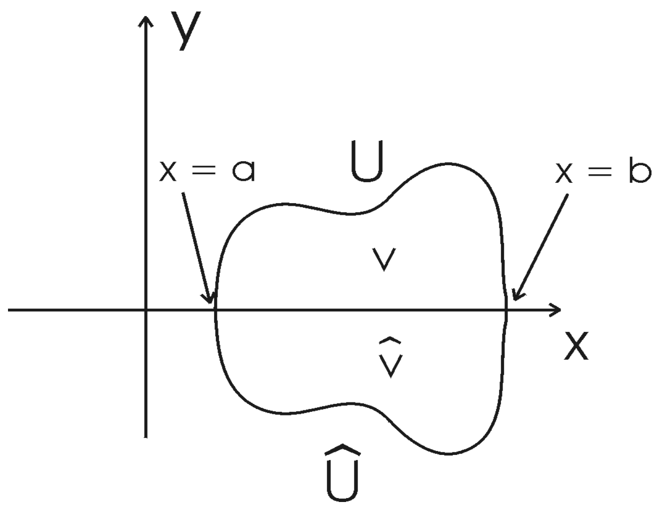

Figure 7.1: Schwarz reflection.

$$\lim_{U \ni z \to x} \operatorname{Im} F(z) = 0 \qquad (7.5.3.1)$$

for each $\zeta \in V \cap$ (real axis). Define $\widetilde{U} = \{z \in \mathbb{C} : \overline{z} \in U\}$. Then there is a holomorphic function G on $U \cup \widetilde{U} \cup \{x \in \mathbb{R} : a < x < b\}$ such that $G\big|_U = F$. In fact $\varphi(\zeta) \equiv \lim_{U \ni z \to \zeta} \operatorname{Re} F(z)$ exists for each $\zeta \in V \cap$ (real axis) and

$$G(z) = \begin{cases} F(z) & \text{if} \quad z \in U \\ \varphi(z) + i0 & \text{if} \quad z \in V \cap \text{ (real axis)} \\ \overline{F(\overline{z})} & \text{if} \quad z \in \widetilde{U}. \end{cases} \qquad (7.5.3.2)$$

7.5.4 More General Versions of the Schwarz Reflection Principle

We take this opportunity to note that Schwarz reflection is not simply a fact about reflection in lines. Since lines are conformally equivalent (by way of linear fractional transformations) to circles, it is also possible to perform Schwarz reflection in a circle (with suitably modified hypotheses). More is true: in fact one can conformally map a neighborhood of any real analytic curve to a line segment; so, with some extra effort, Schwarz reflection may be performed in any real analytic arc.

7.6 Harnack's Principle

7.6.1 The Harnack Inequality

Theorem: Let u be a non-negative, harmonic function on $\overline{D}(0, R)$. Then, for any $z \in D(0, R)$,

$$\frac{R - |z|}{R + |z|} \cdot u(0) \leq u(z) \leq \frac{R + |z|}{R - |z|} \cdot u(0). \tag{7.6.1.1}$$

More generally:

Let u be a non-negative, harmonic function on $\overline{D}(P, R)$. Then, for any $z \in D(P, R)$,

$$\frac{R - |z - P|}{R + |z - P|} \cdot u(P) \leq u(z) \leq \frac{R + |z - P|}{R - |z - P|} \cdot u(P). \tag{7.6.1.2}$$

In fact these extremely useful estimates are a direct reflection of the size of the Poisson kernel. The reader may provide the details.

7.6.2 Harnack's Principle

Theorem: Let $u_1 \leq u_2 \leq \ldots$ be harmonic functions on a connected open set $U \subseteq \mathbb{C}$. Then either $u_j \to \infty$ uniformly on compact sets or there is a (finite-valued) harmonic function u on U such that $u_j \to u$ uniformly on compact sets.

The reader may prove this assertion as a simple application of the Harnack inequalities and a little logic.

7.7 The Dirichlet Problem and Subharmonic Functions

7.7.1 The Dirichlet Problem

Let $U \subseteq \mathbb{C}$ be an open set, $U \neq \mathbb{C}$. Let f be a given continuous function on ∂U. Does there exist a continuous function u on $\overline{U}$ such that $u\big|_{\partial U} = f$ and u is harmonic on U? If u exists, is it unique? These two questions taken together are called the *Dirichlet problem* for the domain U. [Note that we have already solved the Dirichlet problem when U is the unit disc—see §§7.3.1, §§7.3.4.] It has many motivations from physics (see [COH], [LOG]). For instance, suppose that a flat, thin film of heat-conducting material is in thermal equilibrium. That is, the temperature at each point of the film is constant with passing time (§§14.2.2). Then its temperature at various points is a harmonic function (see [KRA1]). Physical intuition suggests that if the boundary ∂U of the film has a given temperature distribution $f : \partial U \to \mathbb{R}$, then the temperatures at interior points are uniquely determined. Historically, physicists have found this intuition strongly compelling.

From the viewpoint of mathematical proof, as opposed to physical intuition, the situation is more complicated. The result of §§7.3.4 asserts in effect that the Dirichlet problem on the unit disc always has a solution. And, on any bounded domain U, it has only one solution corresponding to any given boundary function f, because of the (boundary) maximum principle: If u_1 and u_2 are both solutions, then $u \equiv u_1 - u_2$ is harmonic and is zero on the boundary, so that $u_1 - u_2 \equiv 0$, hence $u_1 \equiv u_2$. While this reasoning demonstrates that the Dirichlet problem on a bounded open set U can have *at most* one solution, it is also the case that on more complicated domains the Dirichlet problem may not have *any* solution.

7.7.2 Conditions for Solving the Dirichlet Problem

The Dirichlet problem is not always solvable on the domain $U = D(0,1) \setminus \{0\}$; in fact the data $f(z) = 1$ when $|z| = 1$ and $f(z) = 0$ when $z = 0$ have no solution—see [GK, p. 229]. Thus *some* conditions on ∂U are necessary in order that the Dirichlet problem be solvable for U. It will turn out that if ∂U consists of "smooth" curves, then the Dirichlet problem is always solvable. The best possible general result is that if each connected component of the boundary of U contains more than one point, then the Dirichlet problem

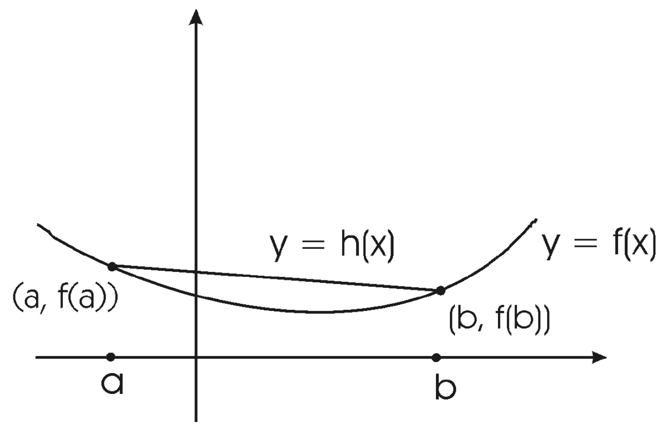

Figure 7.2: A convex function.

can always be solved. Later, we shall enunciate a classical condition for solvability of the Dirichlet formulated in the language of barriers.

7.7.3 Motivation for Subharmonic Functions

We first consider the concept of subharmonicity. This is a complex-analytic analogue of the notion of convexity that we motivate by considering convexity on the real line. For the moment, fix attention on functions $F : \mathbb{R} \to \mathbb{R}$.

On the real line, the analogue of the Laplacian is the operator d^2/dx^2. The analogue of real-valued harmonic functions (that is, the functions u with $\triangle u = 0$) are therefore the functions $h(x)$ such that $[d^2/dx^2][h(x)] \equiv 0$; these are the linear ones. Let $\mathcal{S}$ be the set of continuous functions $f : \mathbb{R} \to \mathbb{R}$ such that whenever $I = [a, b] \subseteq \mathbb{R}$ and h is a real-valued harmonic function on $\mathbb{R}$ with $f(a) \leq h(a)$ and $f(b) \leq h(b)$, then $f(x) \leq h(x)$ for all $x \in I$. (Put simply, if a harmonic function h is at least as large as f at the endpoints of an interval, then it is at least as large as f on the entire interval.) Which functions are in $\mathcal{S}$? The answer is the collection of all convex functions (in the usual sense). Refer to Figure 7.2. [Recall here that a function $f : [a, b] \to \mathbb{R}$ is said to be convex if, whenever $c, d \in [a, b]$ and $0 \leq \lambda \leq 1$ then $f((1 - \lambda)c + \lambda d) \leq (1 - \lambda)f(c) + \lambda f(d)$.] These considerations give us a geometric way to think about convex functions (without resorting to differentiation). See [HOR] for more on this view of subharmonic functions.

7.7.4 Definition of Subharmonic Function

Our definition of subharmonic function on a domain in $\mathbb{C}$ (or $\mathbb{R}^2$) is motivated by the discussion of convexity in the preceding subsection.

> **Definition:** Let $U \subseteq \mathbb{C}$ be an open set and f a real-valued, continuous function on U. Suppose that for each $\overline{D}(P,r) \subseteq U$ and every real-valued harmonic function h defined on a neighborhood of $\overline{D}(P,r)$ which satisfies $f \leq h$ on $\partial D(P,r)$, it holds that $f \leq h$ on $D(P,r)$. Then f is said to be *subharmonic* on U.

7.7.5 Other Characterizations of Subharmonic Functions

A function $f : U \to \mathbb{R}$ that is C^2 is subharmonic if and only if $\triangle f \geq 0$ everywhere. This is analogous to the fact that a C^2 function on (an open set in) $\mathbb{R}$ is convex if and only if it has non-negative second derivative everywhere—and the proof is quite similar. The next result will allow us to identify many subharmonic functions which are only continuous, not C^2, so that the $\triangle f \geq 0$ criterion is not applicable.

Let $f : U \to \mathbb{R}$ be continuous. Suppose that, for each $\overline{D}(P,r) \subseteq U$,

$$f(P) \leq \frac{1}{2\pi} \int_0^{2\pi} f(P + re^{i\theta})d\theta \qquad (7.7.5.1)$$

(this is called the *sub-mean value property*). Then f is subharmonic. This result is derived directly from the mean value property for harmonic functions.

Conversely, if $f : U \to \mathbb{R}$ is a continuous, subharmonic function and if $\overline{D}(P,r) \subseteq U$, then the inequality (7.7.5.1) holds.

7.7.6 The Maximum Principle

A consequence of the sub-mean value property (7.7.5.1) is the maximum principle for subharmonic functions:

> If f is subharmonic on U and if there is a $P \in U$ such that $f(P) \geq f(z)$ for all $z \in U$, then f is constant.

It may be noted that if f is holomorphic then $|f|$ is subharmonic; this explains why a holomorphic function satisfies the maximum principle. The proof of this new maximum principle is identical to proofs of this principle that we have seen in other contexts—see, for example, §§7.2.1, 7.2.2, 7.2.3.

7.7.7 Lack of A Minimum Principle

We note in passing that there is no "minimum principle" for subharmonic functions. Subharmonicity is a "one-sided" property. Put in other words, the negative of a subharmonic function is *not* subharmonic.

7.7.8 Basic Properties of Subharmonic Functions

Here are some properties of subharmonic functions that are worth noting. The third of these explains why subharmonic functions are a much more flexible tool than holomorphic or even harmonic functions. The proofs are immediate from the definitions and the properties of harmonic funcitons discussed thus far.

1. If f_1, f_2 are subharmonic functions on U, then so is $f_1 + f_2$.
2. If f_1 is subharmonic on U and $\alpha > 0$ is a constant, then αf_1 is subharmonic on U.
3. If f_1, f_2 are subharmonic on U then $g(z) \equiv \max\{f_1(z), f_2(z)\}$ is also subharmonic on U.

7.7.9 The Concept of a Barrier

The next notion that we need to introduce is that of a *barrier*. Namely, we want to put a geometric-analytic condition on the boundary of a domain that will rule out examples like the punctured disc in §§7.7.2 (in which the Dirichlet problem could not be solved). The definition of a barrier at a point $P \in \partial U$ is a bit technical, but the existence of a barrier will turn out to be exactly the hypothesis needed for the construction of the solution of the Dirichlet problem.

> **Definition:** Let $U \subseteq \mathbb{C}$ be an open set and $P \in \partial U$. We call a function $b : \overline{U} \to \mathbb{R}$ a *barrier* for U at P if

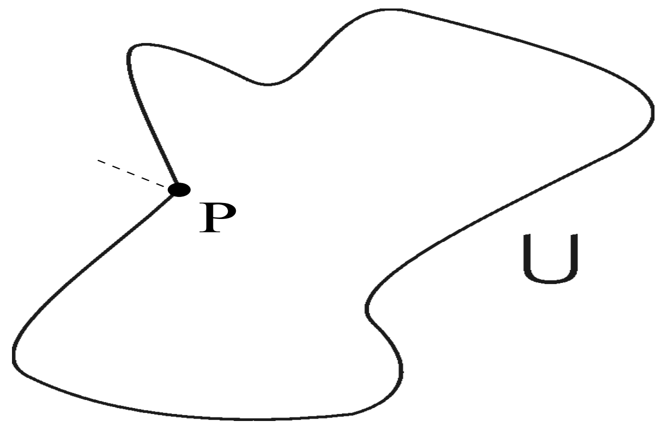

Figure 7.3: The concept of a barrier.

(7.7.9.1) b is continuous;

(7.7.9.2) b is subharmonic on U;

(7.7.9.3) $b\big|_{\partial U} \leq 0$;

(7.7.9.4) $\{z \in \partial U : b(z) = 0\} = \{P\}$.

Thus the barrier b singles out P in a special function-theoretic fashion.

If U is bounded by a C^1 smooth curve (no corners present), then every point of ∂U has a barrier (just conformally map U to a disc). See Figure 7.3.

7.8 The General Solution of the Dirichlet Problem

7.8.1 Enunciation of the Solution of the Dirichlet Problem

Let U be a bounded, connected open subset of $\mathbb{C}$ such that U has a barrier b_P for each $P \in \partial U$. Then the Dirichlet problem can always be solved on U. That is, if f is a continuous function on ∂U, then there is a function u continuous on $\overline{U}$, harmonic on U, such that $u\big|_{\partial U} = f$. The function u is uniquely determined by these conditions.

The result in the preceding paragraph is the standard textbook result about regularity for the Dirichlet problem. More advanced techniques establish that if each connected component of ∂U has at least two points then the Dirichlet problem is solvable on U.

The solution of the Dirichlet problem in the generality that we have been discussing here is a technical *tour de force*. See [GRK] for all the details. The basic idea is that **(i)** there exists some subharmonic function whose boundary limits lie *below* the given boundary data function f; **(ii)** we define a new function u to be the pointwise supremum of all such subharmonic functions; and **(iii)** then the Harnack principle and the maximum principle (along with the barriers) can be used to show that the function u constructed in **(ii)** is harmonic and agrees with f at each boundary point.

Chapter 8

Infinite Series and Products

8.1 Basic Concepts Concerning Infinite Sums and Products

8.1.1 Uniform Convergence of a Sequence

Let $U \subseteq \mathbb{C}$ be an open set and $g_j : U \to \mathbb{C}$ functions. Recall (§3.1.5) that the sequence $\{g_j\}$ is said to *converge uniformly on compact subsets of* U to a function g if the following condition holds: For each compact $K \subseteq U$ (see §§3.1.5) and each $\epsilon > 0$ there is an $N > 0$ such that if $j > N$, then $|g(z) - g_j(z)| < \epsilon$ for all $z \in K$. It should be noted that, in general, the choice of N depends on ϵ *and* on K, but not on the particular point $z \in K$.

8.1.2 The Cauchy Condition for a Sequence of Functions

Because of the completeness (see [RUD1], [KRA2]) of the complex numbers, a sequence of functions is uniformly convergent on compact sets if and only if it is (what is called) *uniformly Cauchy* on compact sets. Here a sequence of functions is said to be *uniformly Cauchy on compact sets* if, for each K compact in U and each $\epsilon > 0$, there is an $N > 0$ such that: for all $j, k > N$ and all $z \in K$ we have $|g_j(z) - g_k(z)| < \epsilon$. The Cauchy condition is useful because it does not make explicit reference to the limit function g. One can thereby verify uniform convergence to some limit without previously determining the limit function explicitly.

8.1.3 Normal Convergence of a Sequence

For the purposes of complex analysis, the basic fact about uniform convergence on compact sets is our result which states that if $\{g_j\}$ are holomorphic functions on U and if the g_j converge uniformly on compact subsets of U to a function g, then g is holomorphic (§§3.1.5). In these circumstances we will say that $\{g_j\}$ converges *normally* to g.

8.1.4 Normal Convergence of a Series

If $f_1, f_2, \ldots$ are functions on U, then we may study the convergence properties of

$$\sum_{j=1}^{\infty} f_j. \tag{8.1.4.1}$$

The series converges *normally* if its sequence of partial sums

$$S_N(z) = \sum_{j=1}^{N} f_j(z) \ , \qquad N = 1, 2, \ldots \tag{8.1.4.2}$$

converges normally in U. The function

$$f(z) = \sum_{j=1}^{\infty} f_j(z) \tag{8.1.4.3}$$

will then be holomorphic because it is the normal limit of the partial sums S_N (each of which is holomorphic).

8.1.5 The Cauchy Condition for a Series

There is a Cauchy condition for normal convergence of a series: the series

$$\sum_{j=1}^{\infty} f_j(z) \tag{8.1.5.1}$$

is said to be *uniformly Cauchy* on compact sets if, for each compact $K \subseteq U$ and each $\epsilon > 0$, there is an $N > 0$ such that for all $M \geq L > N$ it holds that

$$\left| \sum_{j=L}^{M} f_j(z) \right| < \epsilon. \tag{8.1.5.2}$$

[Note that this is just a reformulation of the Cauchy condition for the sequence of partial sums $S_N(z)$.] It is easy to see that a series that is uniformly Cauchy converges normally to its limit function.

8.1.6 The Concept of an Infinite Product

Now we turn to products. One of the principal activities in complex analysis is to construct holomorphic or meromorphic functions with certain prescribed behavior. For some problems of this type, it frequently turns out that infinite products are more useful than infinite sums. The reason is that, for instance, if we want to construct a function that will vanish on a certain infinite set $\{a_j\}$, then we could hope to find individual functions f_j that vanish at a_j and then multiply the f_j's together. This process requires that we make sense of the notion of "infinite product."

8.1.7 Infinite Products of Scalars

We begin with infinite products of complex numbers, and then adapt the ideas to infinite products of functions. For reasons that will become apparent momentarily, it is convenient to write products in the form

$$\prod_{j=1}^{\infty}(1+a_j), \tag{8.1.7.1}$$

where $a_j \in \mathbb{C}$. The symbol $\prod$ stands for multiplication. We want to define what it means for a product such as (8.1.7.1) to converge.

8.1.8 Partial Products

First define the *partial products* P_N of (8.1.7.1) to be

$$P_N = \prod_{j=1}^{N}(1+a_j) \equiv (1+a_1)\cdot(1+a_2)\cdots(1+a_N). \tag{8.1.8.1}$$

We might be tempted to say that the infinite product

$$\prod_{j=1}^{\infty}(1+a_j) \tag{8.1.8.2}$$

converges if the sequence of partial products $\{P_N\}$ converges. But, for technical reasons, a different definition is more useful.

8.1.9 Convergence of an Infinite Product

An infinite product

$$\prod_{j=1}^{\infty}(1+a_j) \tag{8.1.9.1}$$

is said to *converge* if

(8.1.8.2) Only a finite number $a_{j_1}, \ldots, a_{j_k}$ of the a_j's are equal to -1.

(8.1.9.3) If $N_0 > 0$ is so large that $a_j \neq -1$ for $j > N_0$, then

$$\lim_{N\to+\infty} \prod_{j=N_0+1}^{N}(1+a_j) \tag{8.1.9.3.1}$$

exists and is non-zero.

8.1.10 The Value of an Infinite Product

If $\prod_{j=1}^{\infty}(1+a_j)$ converges, then we define its *value* to be (with N_0 as in (8.1.9.3))

$$\left[\prod_{j=1}^{N_0}(1+a_j)\right] \cdot \lim_{N\to+\infty} \prod_{N_0+1}^{N}(1+a_j). \tag{8.1.10.1}$$

8.1.11 Products That Are Disallowed

As the exposition develops, it will become clear why we wish to disallow products with

$$\lim_{N\to+\infty} \prod_{j=N_0+1}^{N}(1+a_j) = 0. \tag{8.1.11.1}$$

8.1.12 Condition for Convergence of an Infinite Product

If

$$\sum_{j=1}^{\infty} |a_j| < \infty\,, \tag{8.1.12.1}$$

then

$$\prod_{j=1}^{\infty} (1 + |a_j|) \tag{8.1.12.2}$$

converges. [See (8.1.12.6), (8.1.12.7) for part of the mathematical reason as to why these assertions are true.] The facts all hinge on the basic identity $e^{\alpha} = 1 + \alpha + \cdots$.

If

$$\prod_{j=1}^{\infty} (1 + |a_j|) \tag{8.1.12.3}$$

converges, then

$$\sum_{j=1}^{\infty} |a_j| \tag{8.1.12.4}$$

converges.

Let $a_j \in \mathbb{C}$. Set

$$P_N = \prod_{j=1}^{N} (1 + a_j)\,, \qquad \widetilde{P}_N = \prod_{j=1}^{N} (1 + |a_j|). \tag{8.1.12.5}$$

Then

(8.1.12.6) $\widetilde{P}_N \leq \exp\left(|a_1| + \cdots + |a_N|\right)$.

(8.1.12.7) $|P_N - 1| \leq \widetilde{P}_N - 1$.

If the infinite product

$$\prod_{j=1}^{\infty} (1 + |a_j|) \tag{8.1.12.8}$$

converges, then so does

$$\prod_{j=1}^{\infty} (1 + a_j). \tag{8.1.12.9}$$

If

$$\sum_{j=1}^{\infty} |a_j| < \infty \,, \tag{8.1.12.10}$$

then

$$\prod_{j=1}^{\infty} (1 + a_j) \tag{8.1.12.11}$$

converges.

This last is our most useful convergence result for infinite products. It is so important that it is worth restating in a standard alternative form: If

$$\sum_{j=1}^{\infty} |1 - a_j| < \infty \,, \tag{8.1.12.12}$$

then

$$\prod_{j=1}^{\infty} a_j \tag{8.1.12.13}$$

converges. The proof is just a change of notation.

8.1.13 Infinite Products of Holomorphic Functions

We now apply these considerations to infinite products of holomorphic functions.

Let $U \subseteq \mathbb{C}$ be open. Suppose that $f_j : U \to \mathbb{C}$ are holomorphic and that

$$\sum_{j=1}^{\infty} |f_j| \tag{8.1.13.1}$$

converges uniformly on compact subsets of U. Then the sequence of partial products

$$F_N(z) = \prod_{j=1}^{N} (1 + f_j(z)) \tag{8.1.13.2}$$

converges uniformly on compact subsets of U. In particular, the limit of these partial products defines a holomorphic function on U.

8.1.14 Vanishing of an Infinite Product

The function f defined on a domain U by the product

$$f(z) = \prod_{j=1}^{\infty}(1 + f_j(z)) \tag{8.1.14.1}$$

vanishes at a point $z_0 \in U$ if and only if $f_j(z_0) = -1$ for some j. The multiplicity of the zero at z_0 is the sum of the multiplicities of the zeros of the functions $1 + f_j$ at z_0.

8.1.15 Uniform Convergence of an Infinite Product of Functions

Remark: For convenience, one says that the product $\prod_1^\infty(1 + f_j(z))$ converges uniformly on a set E if

(8.1.15.1) it converges for each z in E

and

(8.1.15.2) the sequence $\{\prod_1^N(1+f_j(z))\}$ converges uniformly on E to $\prod_1^\infty(1+f_j(z))$.

Then our main convergence result can be summarized as follows:

8.1.16 Condition for the Uniform Convergence of an Infinite Product of Functions

Theorem: If $\sum_{j=1}^\infty |f_j|$ converges uniformly on compact sets, then the product $\prod_1^\infty(1 + f_j(z))$ converges uniformly on compact sets.

It should be noted that the convergence conditions in §§8.1.9 are satisfied automatically in the situation of this last theorem.

8.2 The Weierstrass Factorization Theorem

8.2.1 Prologue

One of the most significant facts about a polynomial function $p(z)$ of $z \in \mathbb{C}$ is that it can be factored (see §§3.1.4):

$$p(z) = c \cdot \prod_{j=1}^{k} (z - a_j). \tag{8.2.1.1}$$

Among other things, such a factorization facilitates the study of the zeros of p. In this section we shall show that in fact any entire function can be factored in such a way that each multiplicative factor possesses precisely one zero (of first order). Since a function holomorphic on all of $\mathbb{C}$ (called an *entire function*) can have infinitely many zeros, the factorization must be an infinite product in at least some cases. We consider such a factorization in this section.

8.2.2 Weierstrass Factors

To obtain the Weierstrass factors, we define

$$E_0(z) = 1 - z \tag{8.2.2.1}$$

and for $1 \leq p \in \mathbb{Z}$ we let

$$E_p(z) = (1 - z) \exp\left(z + \frac{z^2}{2} + \cdots + \frac{z^p}{p}\right). \tag{8.2.2.2}$$

Of course each E_p is holomorphic on all of $\mathbb{C}$. The factorization theory hinges on a technical calculation that says that, in some sense, E_p is close to 1 if $|z|$ is small. This assertion is not surprising since

$$\left(z + \frac{z^2}{2} + \cdots + \frac{z^p}{p}\right) \tag{8.2.2.3}$$

is the initial part of the power series of $-\log(1 - z)$. Thus

$$(1 - z) \exp\left(z + \frac{z^2}{2} + \cdots + \frac{z^p}{p}\right) \tag{8.2.2.4}$$

might be expected to be close to 1 for z small (and p large).

8.2.3 Convergence of the Weierstrass Product

Theorem: Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of non-zero complex numbers with no accumulation point in the complex plane (note, however, that the a_ns need not be distinct). If $\{p_n\}$ are positive integers that satisfy

$$\sum_{n=1}^{\infty} \left(\frac{r}{|a_n|} \right)^{p_n+1} < \infty \tag{8.2.3.1}$$

for every $r > 0$, then the infinite product

$$\prod_{n=1}^{\infty} E_{p_n} \left(\frac{z}{a_n} \right) \tag{8.2.3.2}$$

(called a *Weierstrass product*) converges uniformly on compact subsets of $\mathbb{C}$ to an entire function F. The zeros of F are precisely the points $\{a_n\}$, counted with multiplicity.

8.2.4 Existence of an Entire Function with Prescribed Zeros

Let $\{a_n\}_{n=1}^{\infty}$ be any sequence in the plane with no finite accumulation point. Then there exists an entire function f with zero set precisely equal to $\{a_n\}_{n=1}^{\infty}$ (counting multiplicities). The function f is given by a Weierstrass product.

8.2.5 The Weierstrass Factorization Theorem

Theorem: Let f be an entire function. Suppose that f vanishes to order m at 0, $m \geq 0$. Let $\{a_n\}$ be the other zeros of f, listed with multiplicities. Then there is an entire function g such that

$$f(z) = z^m \cdot e^{g(z)} \prod_{n=1}^{\infty} E_{n-1} \left(\frac{z}{a_n} \right) . \tag{8.2.5.1}$$

8.3 The Theorems of Weierstrass and Mittag-Leffler

8.3.1 The Concept of Weierstrass's Theorem

Let $U \subset \mathbb{C}$ be a domain. The only necessary condition that we know for a set $\{a_j\} \subseteq U$ to be the zero set of a function f holomorphic on U is that $\{a_j\}$ have no accumulation point in U. It is remarkable that this condition is also sufficient: that is the content of Weierstrass's theorem.

8.3.2 Weierstrass's Theorem

> **Theorem:** Let $U \subseteq \mathbb{C}$ be any open set. Let $a_1, a_2, \ldots$ be a finite or infinite sequence in U (possibly with repetitions) that has no accumulation point *in* U. Then there exists a holomorphic function f on U whose zero set is precisely $\{a_j\}$. (The function f is constructed by taking an infinite product.)

The proof converts the problem to a situation on the entire plane, and then uses the Weierstrass product.

We next want to formulate a result about maximal domains of existence (or domains of definition) of holomorphic functions. But first we need a geometric fact about open subsets of the plane.

8.3.3 Construction of a Discrete Set

Let $U \underset{\neq}{\subset} \mathbb{C}$ be any open set. Then there exists a countably infinite set $A \subseteq U$ such that

(8.3.3.1) A has no accumulation point in U.

(8.3.3.2) Every $P \in \partial U$ is an accumulation point of A.

See Figure 8.1. Details of this construction appear in [GRK, p. 268].

8.3.4 Domains of Existence for Holomorphic Functions

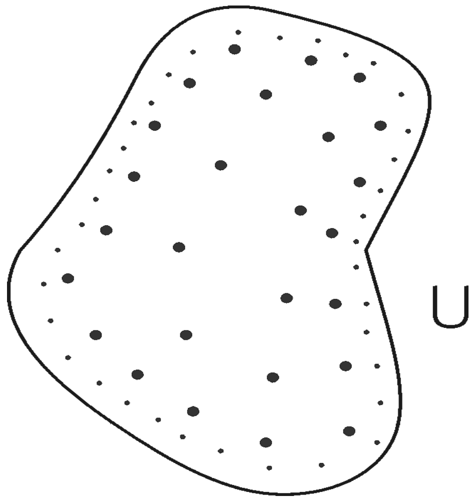

Figure 8.1: Any domain in $\mathbb{C}$ is a domain of holomorphy.

Theorem: Let $U \subseteq \mathbb{C}$ be any proper, connected, open subset. There is a function f holomorphic on U such that f cannot be analytically continued past any $P \in \partial U$ (see also §§9.2.2 on the concept of "regular boundary point").

The verification of this last result is just to apply Weierstrass's theorem to the discrete set described in §§8.3.3. This yields a non-constant holomorphic function f on U whose zero set accumulates at every boundary point of U. If f were to analytically continue to any strictly larger open set $\widetilde{U}$, then $\widetilde{U}$ would contain a point of ∂U, hence would have an interior accumulation point of the zeros of f. Thus f would be identically zero, and that would be a contradiction.

8.3.5 The Field Generated by the Ring of Holomorphic Functions

Another important corollary of Weierstrass's theorem is that, for any open U, the field generated by the ring of holomorphic functions on U is the field of all meromorphic functions on U. In simpler language:

Let $U \subseteq \mathbb{C}$ be open. Let m be meromorphic on U. Then there are holomorphic functions f, g on U such that

$$m(z) = \frac{f(z)}{g(z)}. \tag{8.3.5.1}$$

8.3.6 The Mittag-Leffler Theorem

Since it is possible to prescribe zeros of a holomorphic function on any open U, then of course we can (in principle) prescribe poles—since $1/f$ has poles exactly where f has zeros. But we can do better: with a little extra work we can prescribe the negative power portion of the Laurent series on any discrete subset of U.

We now formulate the basic result on prescribing pole behavior, known as the Mittag-Leffler theorem, in two different (but equivalent) ways: one qualitative and the other quantitative.

Mittag-Leffler Theorem: First Version

> Let $U \subseteq \mathbb{C}$ be any open set. Let $\alpha_1, \alpha_2, \ldots$ be a finite or countably infinite set of *distinct elements of* U with no accumulation point in U. Suppose, for each j, that U_j is a neighborhood of α_j. Further assume, for each j, that m_j is a meromorphic function defined on U_j with a pole at α_j and no other poles. Then there exists a meromorphic m on U such that $m - m_j$ is holomorphic on U_j for every j.

The Mittag-Leffler Theorem: Alternative Formulation

> Let $U \subseteq \mathbb{C}$ be any open set. Let $\alpha_1, \alpha_2, \ldots$ be a finite or countably infinite set of distinct elements of U, having no accumulation point in U. Let s_j be a sequence of Laurent polynomials (or "principal parts"),
>
> $$s_j(z) = \sum_{\ell=-p(j)}^{-1} a_\ell^j \cdot (z - \alpha_j)^\ell \tag{8.3.6.1}$$
>
> (see §§4.3.1). Then there is a meromorphic function on U whose principal part at each α_j is s_j.

8.3.7 Prescribing Principal Parts

The theorems of Weierstrass and Mittag-Leffler can be combined to allow specification of a finite part of the Laurent series at a discrete set of points:

> Let $U \subseteq \mathbb{C}$ be an open set and let $\alpha_1, \alpha_2, \ldots$ be a finite or countably infinite set of distinct points of U having no interior accumulation point in U. For each j let there be given an expression
>
> $$t_j(z) = \sum_{\ell=-M(j)}^{N(j)} a_\ell^j \cdot (z - \alpha_j)^\ell, \tag{8.3.7.1}$$
>
> with $M(j), N(j) \geq 0$. Then there is a meromorphic function m on U, holomorphic on $U \setminus \{\alpha_j\}$, such that: if $-M(j) \leq \ell \leq N(j)$, then the ℓ^{th} Laurent coefficient of m at α_j is a_j^ℓ.

8.4 Normal Families

8.4.1 Normal Convergence

A sequence of functions f_j on an open set $U \subseteq \mathbb{C}$ is said to *converge normally* to a limit function f_0 on U (see §§8.1.3) if $\{f_j\}$ converges to f_0 uniformly on compact subsets of U. That is, the convergence is normal if, for each compact set $K \subseteq U$ and each $\epsilon > 0$, there is an $N > 0$ (depending on K and ϵ) such that when $j > N$ and $z \in K$, then $|f_j(z) - f_0(z)| < \epsilon$ (see §§8.1.3, 3.1.5).

The functions $f_j(z) = z^j$ converge normally on the unit disc D to the function $f_0(z) \equiv 0$. The sequence does *not* converge uniformly on all of D to f_0, but does converge uniformly on each compact subset of D.

8.4.2 Normal Families

Let $\mathcal{F}$ be a family of (holomorphic) functions with common domain U. We say that $\mathcal{F}$ is a *normal family* if every sequence in $\mathcal{F}$ has a subsequence that converges uniformly on compact subsets of U, i.e., *converges normally* on U.

Let $\mathcal{F}$ be a family of functions on an open set $U \subseteq \mathbb{C}$. We say that $\mathcal{F}$ is *bounded* if there is a constant $N > 0$ such that $|f(z)| \leq N$ for all $z \in U$ and all $f \in \mathcal{F}$. We say that $\mathcal{F}$ is *bounded on compact sets* if for each compact set $K \subseteq U$ there is a constant $M = M_K$ such that for all $f \in \mathcal{F}$ and all $z \in K$ we have

$$|f(z)| \leq M. \tag{8.4.2.1}$$

8.4.3 Montel's Theorem, First Version

Theorem (Montel's Theorem, First Version): Let $\mathcal{F} = \{f_\alpha\}_{\alpha \in A}$ be a bounded family of holomorphic functions on an open set $U \subseteq \mathbb{C}$. Then there is a sequence $\{f_j\} \subseteq \mathcal{F}$ such that f_j converges normally on U to a limit (holomorphic) function f_0.

Thus a bounded family of holomorphic functions is normal.

8.4.4 Montel's Theorem, Second Version

Theorem (Montel's Theorem, Second Version): Let $U \subseteq \mathbb{C}$ be an open set and let $\mathcal{F}$ be a family of holomorphic functions on U that is bounded on compact sets. Then there is a sequence $\{f_j\} \subseteq \mathcal{F}$ that converges normally on U to a limit (necessarily holomorphic) function f_0.

Thus a family of holomorphic functions that is bounded on compact sets is normal. Montel's theorem is proved with a judicious application of the Ascoli-Arzela theorem. The hypotheses of equiboundededness and equicontinuity are derived from the Cauchy estimates.

8.4.5 Examples of Normal Families

(8.4.5.1) Consider the family $\mathcal{F} = \{z^j\}_{j=1}^{\infty}$ of holomorphic functions. If we take U to be any subset of the unit disc, then $\mathcal{F}$ is bounded (by 1) so Montel's theorem (first version) guarantees that there is a subsequence that converges uniformly on compact subsets. Of course in this case it is plain by inspection that *any* subsequence will converge uniformly on compact sets to the identically zero function.

The family $\mathcal{F}$ fails to be bounded on compact sets for any U that contains points of modulus greater than one. Thus neither version of Montel's theorem would apply on such a U. And there is no convergent sequence in $\mathcal{F}$ for such a U.

(8.4.5.2) Let $\mathcal{F} = \{z/j\}_{j=1}^{\infty}$ on $\mathbb{C}$. Then there is no bound M such that $|z/j| \leq M$ for all j and all $z \in \mathbb{C}$. But for each fixed compact subset $K \subseteq \mathbb{C}$ there is a constant M_K such that $|z/j| < M_K$ for all j and all $z \in K$. (For instance, $M_K = \max\{|z| : z \in K\}$ would do.) Therefore the second version of Montel's theorem applies. And indeed the sequence $\{z/j\}_{j=1}^{\infty}$ converges normally to 0 on $\mathbb{C}$.

(8.4.5.3) Let $\mathcal{S}$ be the *Schicht functions* on the unit disc. These are the holomorphic functions that are univalent, take 0 to 0, and have derivative 1 at the origin. Then $\mathcal{S}$ is a normal family.

Chapter 9

Analytic Continuation

9.1 Definition of an Analytic Function Element

9.1.1 Continuation of Holomorphic Functions

Suppose that V is a connected, open subset of $\mathbb{C}$ and that $f_1 : V \to \mathbb{C}$ and $f_2 : V \to \mathbb{C}$ are holomorphic functions. If there is an open, non-empty subset U of V such that $f_1 \equiv f_2$ on U, then $f_1 \equiv f_2$ on all of V (see §§3.2.3). Put another way, if we are given an f holomorphic on U, then there is at most one way to extend f to V so that the extended function is holomorphic. [Of course there might not even be one such extension: if V is the unit disc and U the disc $D(3/4, 1/4)$, then the function $f(z) = 1/z$ does not extend. Or if U is the plane with the non-positive real axis removed, $V = \mathbb{C}$, and $f(re^{i\theta}) = r^{1/2}e^{i\theta/2}$, $-\pi < \theta < \pi$, then again no extension from U to V is feasible.]

This chapter deals with the question of when this extension process can be carried out, and in particular what precise meaning can be given to extending f from U to as large a set V as possible. Since there are potentially many different ways to carry out this analytic continuation process, there are questions of ambiguity and redundancy. These all will be addressed here.

9.1.2 Examples of Analytic Continuation

We introduce the basic issues of "analytic continuation" by way of three examples:

EXAMPLE 9.1.2.1 Define

$$f(z) = \sum_{j=0}^{\infty} z^j. \tag{9.1.2.1.1}$$

This series converges normally on the disc $D = \{z \in \mathbb{C} : |z| < 1\}$. It diverges for $|z| > 1$. Is it safe to say that D is the natural domain of definition for f (refer to §§8.3.2, §§8.3.4 for this terminology)? Or can we "continue" f to a larger open set?

We cannot discern easily the answer to this question simply by examining the power series. Instead, we should sum the series and observe that

$$f(z) = \frac{1}{1-z}. \tag{9.1.2.1.2}$$

This formula for f agrees with the original definition of f as a series; however, the formula (9.1.2.1.2) makes sense for all $z \in \mathbb{C}\setminus\{1\}$. In our new terminology, to be made more precise later, f has an analytic continuation to $\mathbb{C} \setminus \{1\}$.

Thus we see that the natural domain of definition for f is the rather large set $\mathbb{C} \setminus \{1\}$. However, the original definition, by way of a series, gave little hint of this fact. □

EXAMPLE 9.1.2.2 Consider the function

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t}\, dt. \tag{9.1.2.2.1}$$

This function is known as the gamma function of Euler. Let us make the following quick observations:

(9.1.2.2.2) The term t^{z-1} has size $|t^{z-1}| = t^{\mathrm{Re}\, z-1}$. Thus the singularity at the origin will be integrable when $\mathrm{Re}\, z > 0$.

(9.1.2.2.3) Because of the presence of the exponential factor, the integrand will certainly be integrable at infinity.

(9.1.2.2.4) The function Γ is holomorphic on the domain $U_0 \equiv \{z : \operatorname{Re} z > 0\}$: the functions

$$\int_a^{1/a} t^{z-1} e^{-t}\, dt, \tag{9.1.2.2.5}$$

with $a > 0$, are holomorphic by differentiation under the integral sign (or use Morera's theorem—§§2.3.2), and $\Gamma(z)$ is the normal limit of these integrals as $a \to 0^+$.

The given definition (9.1.2.2.1) of $\Gamma(z)$ makes no sense when $\operatorname{Re} z \leq 0$ because the improper integral diverges at 0. Can we conclude from this observation that the natural domain of definition of Γ is U_0?

Let us examine this question by integrating by parts:

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt = \frac{1}{z} t^z e^{-t} \Big|_0^\infty + \frac{1}{z} \int_0^\infty t^z e^{-t}\, dt. \tag{9.1.2.2.6}$$

An elementary analysis shows that, as long as $\operatorname{Re} z \neq 0$, the boundary terms vanish (in the limit). Thus we see that

$$\Gamma(z) = \frac{1}{z} \int_0^\infty t^z e^{-t}\, dt. \tag{9.1.2.2.7}$$

Now, whereas the original definition (9.1.2.2.1) of the gamma function made sense on U_0, this new formula (which agrees with the old one on U_0) actually makes sense on $U_1 \equiv \{z : \operatorname{Re} z > -1\} \setminus \{0\}$. No difficulty about the convergence of the integral as the lower limit tends to 0^+ occurs if $z \in \{z : \operatorname{Re} z > -1\}$.

We can integrate by parts once again, and find that

$$\Gamma(z) = \frac{1}{z(z+1)} \int_0^\infty t^{z+1} e^{-t}\, dt. \tag{9.1.2.2.8}$$

This last formula makes sense on $U_2 = \{z : \operatorname{Re} z > -2\} \setminus \{0, -1\}$.

Continuing this process, we may verify that the gamma function, originally defined only on U_0, can be "analytically continued" to $U = \{z \in \mathbb{C} : z \neq 0, -1, -2, \ldots\}$. $\square$

In the first two examples, the functions are given by a formula that only makes sense on a certain open set; yet there is in each case a device for

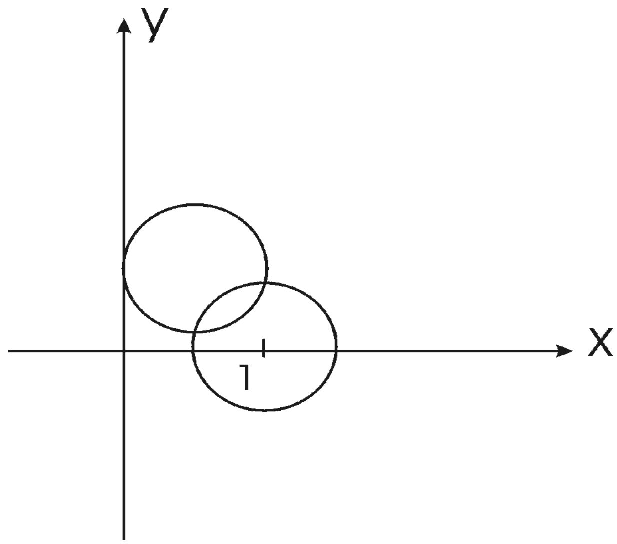

Figure 9.1: Analytic continuation to a second disc.

extending the function to a larger open set. Recall that, by our uniqueness results for analytic functions, there can be at most one way to effect this "analytic continuation" process to a fixed, larger (connected) open set. In the next example, we learn about possible ambiguities in the process when one attempts continuation along two different paths.

EXAMPLE 9.1.2.3 Consider the function $f(z)$, initially defined on the disc $D((1+i0), 1/2)$ by $f(re^{i\theta}) = r^{1/2}e^{i\theta/2}$. Here it is understood that $-\pi/4 < \theta < \pi/4$. This function is well-defined and holomorphic; in fact it is the function usually called the *principal branch* of $\sqrt{z}$. Note that $[f(z)]^2 = z$.

Imagine analytically continuing f to a second disc, as shown in Figure 9.1. This is easily done, using the same definition $f(re^{i\theta}) = r^{1/2}e^{i\theta/2}$. If we continue to a third disc (Figure 9.2), and so on, we end up defining the square root function at $z = -1$. See Figure 9.3. Indeed, we find that $f(-1)$ has the value i.

However, we might have begun our analytic continuation process as shown in Figure 9.4. We begin at 1, and iterate the continuation process in a clockwise direction so that process ends at $z = -1$. Doing so, we would have found that $f(-1) = -i$.

Thus we see that the process of analytic continuation can be ambiguous.

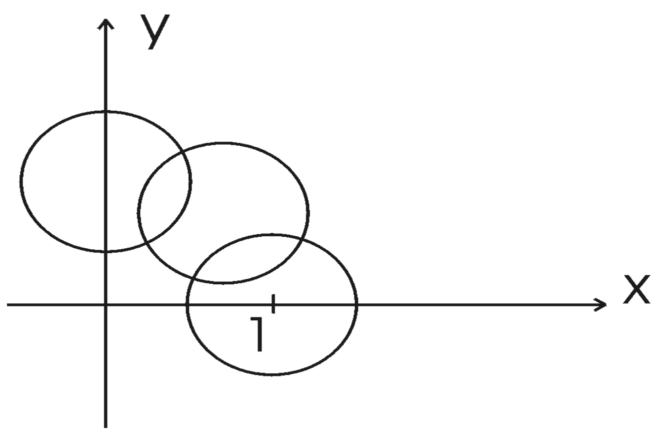

Figure 9.2: Continuation to a third disc.

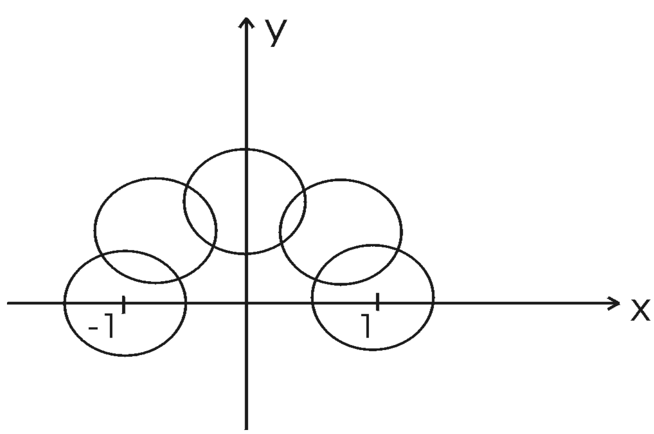

Figure 9.3: Analytic continuation of the square root function.

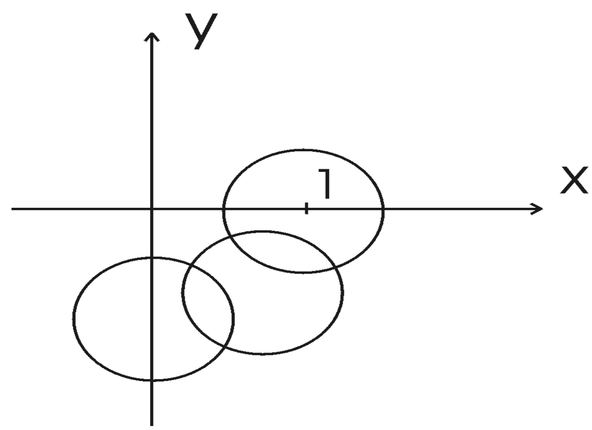

Figure 9.4: Analytic continuation in the other direction.

In the present example, the ambiguity is connected to the fact that a holomorphic square root function cannot be defined in any neighborhood of the origin; yet our two paths of analytic continuation encircle the origin. □

It is because of the phenomena illustrated in these three examples—and many others like them—that we must take a detailed and technical approach to the process of analytic continuation. Even making the questions themselves precise takes some thought.

9.1.3 Function Elements

A *function element* is an ordered pair (f, U), where U is a disc $D(P, r)$ and f is a holomorphic function defined on U. If W is an open set, then a function element in W is a pair (f, U) such that $U \subseteq W$.

9.1.4 Direct Analytic Continuation

Let (f, U) and (g, V) be function elements. We say that (g, V) is a *direct analytic continuation* of (f, U) if $U \cap V \neq \emptyset$ and f and g are equal on $U \cap V$ (Figure 9.5). Obviously (g, V) is a direct analytic continuation of (f, U) if and only if (f, U) is a direct analytic continuation of (g, V).

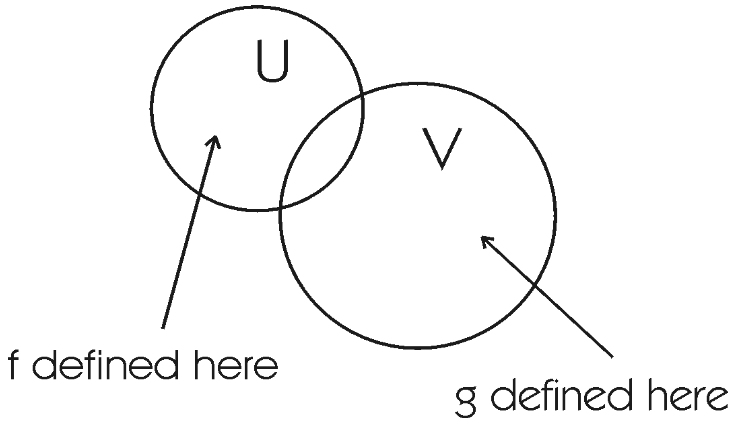

Figure 9.5: Direct analytic continuation.

9.1.5 Analytic Continuation of a Function

If $(f_1, U_1), \ldots, (f_k, U_k)$ are function elements and if each (f_j, U_j) is a direct analytic continuation of (f_{j-1}, U_{j-1}), $j = 2, \ldots, k$, then we say that (f_k, U_k) is an *analytic continuation* of (f_1, U_1) (Figure 9.6).

Clearly (f_k, U_k) is an analytic continuation of (f_1, U_1) if and only if (f_1, U_1) is an analytic continuation of (f_k, U_k). Also if (f_k, U_k) is an analytic continuation of (f_1, U_1) via a chain $(f_1, U_1), \ldots, (f_k, U_k)$ and if $(f_{k+\ell}, U_{k+\ell})$ is an analytic continuation of (f_k, U_k) via a chain $(f_k, U_k), (f_{k+1}, U_{k+1}), \ldots$ $(f_{k+\ell}, U_{k+\ell})$, then stringing the two chains together into $(f_1, U_1), \ldots, (f_{k+\ell}, U_{k+\ell})$ exhibits $(f_{k+\ell}, U_{k+\ell})$ as an analytic continuation of (f_1, U_1). Obviously any function element (f, U) is an analytic continuation of itself.

9.1.6 Global Analytic Functions

Thus we have an equivalence relation (see [KRA3, p. 52] for this terminology) by way of analytic continuation on the set of function elements: namely, two function elements are equivalent if one is the analytic continuation of the other. The equivalence classes ([KRA3, p. 53]) induced by this relation are called *(global) analytic functions*. However, a caution is in order: global analytic functions are not yet functions in the usual sense, and they are not analytic in any sense that we have defined as yet. Justification for the terminology will appear in due course.

Note that the initial element $(f, U) = (f_1, U_1)$ uniquely determines the

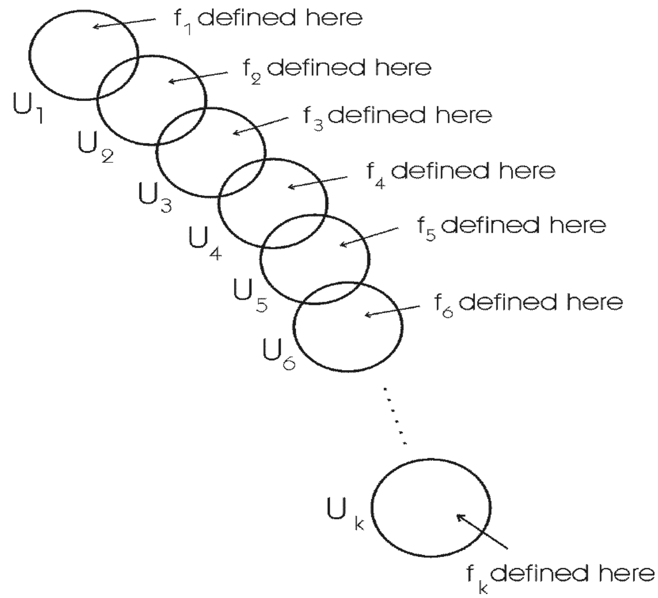

Figure 9.6: Analytic continuation of a function element.

global analytic function, or equivalence class, that contains it. But a global analytic function may include more than one function element of the form (f, U) *for a fixed disc* U (see Example 9.1.2.3). Indeed, a global analytic function f may have in effect more than one value at a point: two function elements (f_1, U) and (f_2, U) can be equivalent even though $f_1(P) \neq f_2(P)$, where P is the center of the disc U. If $\mathbf{f}$ denotes the global analytic function corresponding to (f, U), then we call (f, U) a *branch* of $\mathbf{f}$.

9.1.7 An Example of Analytic Continuation

EXAMPLE 9.1.7.1 Let $U = D(1 + i0, 1/2)$ and let f be the holomorphic function $\log z$. Here $\log z$ is understood to be defined to be $\log |z| + i \arg z$, and $-\pi/4 < \arg z < \pi/4$. As in Example 9.1.2.3, the function element (f, U) can be analytically continued to the point $-1 + i0$ in (at least) two different ways, depending on whether the continuation is along a curve proceeding clockwise about the origin or counterclockwise about the origin.

In fact, all the "branches" of $\log z$ can be obtained by analytic continuation of the $\log |z| + i\arg z$ branch on $D(1 + 0i, 1/2)$ (by continuing several times around the origin, either clockwise or counterclockwise). Thus the idea

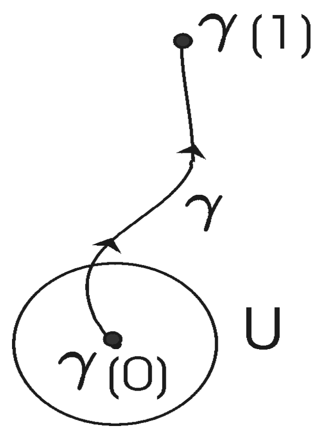

Figure 9.7: Analytic continuation along a curve.

of "branches" of $\log z$ (§§4.5.6, §§9.1.2) is consistent with the general analytic continuation terminology just introduced. $\square$

In some situations, it is convenient to think of a function element as a convergent power series. Then the role of the open disc U is played by the domain of convergence of the power series. This is a useful heuristic idea for the reader to bear in mind. From this viewpoint, two function elements (f_1, U) and (f_2, V) at a point P (such that U and V are discs centered at the same point P) should be regarded as equal if $f_1 \equiv f_2$ on $U \cap V$.

9.2 Analytic Continuation along a Curve

9.2.1 Continuation on a Curve

Let $\gamma : [0,1] \to \mathbb{C}$ be a curve and let (f, U) be a function element with $\gamma(0)$ the center of the disc U (Figure 9.7). An *analytic continuation* of (f, U) along the curve γ is a collection of function elements (f_t, U_t), $t \in [0,1]$, such that

(9.2.1.1) $(f_0, U_0) = (f, U)$.

(9.2.1.2) For each $t \in [0,1]$, the center of the disc U_t is $\gamma(t)$, $0 \leq t \leq 1$.

(9.2.1.3) For each $t \in [0,1]$, there is an $\epsilon > 0$ such that, for each $t' \in [0,1]$ with $|t' - t| < \epsilon$, it holds that

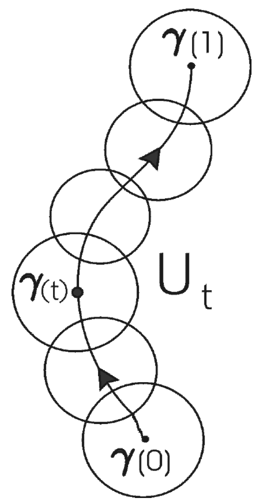

Figure 9.8: Direct analytic continuation.

(a) $\gamma(t') \in U_t$ and hence $U_{t'} \cap U_t \neq \emptyset$;

(b) $f_t \equiv f_{t'}$ on $U_{t'} \cap U_t$ (so that (f_t, U_t) is a direct analytic continuation of $(f_{t'}, U_{t'})$). Refer to Figure 9.8.

Let (f, U) be a function element with U a disc having center P. Let γ be a curve such that $\gamma(0) = P$. Any two analytic continuations of (f, U) along γ agree in the following sense: if (f_m, U_m) is the terminal element of one analytic continuation (f_t, U_t) and if $(\widetilde{f}_{\widetilde{m}}, \widetilde{U}_{\widetilde{m}})$ is the terminal element of another analytic continuation $(\widetilde{f}_{\widetilde{t}}, \widetilde{U}_{\widetilde{t}})$, then f_m and $\widetilde{f}_{\widetilde{m}}$ are equal on $U_m \cap \widetilde{U}_{\widetilde{m}}$.

9.2.2 Uniqueness of Continuation along a Curve

Thus we see that the analytic continuation of a given function element along a given curve is essentially unique, if it exists. From here on, to avoid being pedantic, we shall regard two analytic continuations (f_t, U_t) and $(\widetilde{f}_t, \widetilde{U}_t)$ as "equal," or equivalent, if $f_t = \widetilde{f}_t$ on $U_t \cap \widetilde{U}_t$ for all t. With this terminological convention (which will cause no trouble), the result of §9.2.1 says exactly that analytic continuation of a given function element along a curve is unique.

9.3 The Monodromy Theorem

The fundamental issue to be addressed in the present section is this:

9.3.1 Unambiguity of Analytic Continuation

> Let P and Q be points in the complex plane. Let (f, U) be a function element such that U is a disc centered at P. If γ_1, γ_2 are two curves that begin at P and terminate at Q then does the terminal element of the analytic continuation of (f, U) along γ_1 equal the terminal element of the analytic continuation of (f, U) along γ_2 (on their common domain of definition)?

We shall begin to answer this question in §§9.3.2. The culmination of our discussion will be the monodromy theorem in §§9.3.5.

9.3.2 The Concept of Homotopy

Let W be a domain in $\mathbb{C}$. Let $\gamma_0 : [0,1] \to W$ and $\gamma_1 : [0,1] \to W$ be curves. Assume that $\gamma_0(0) = \gamma_1(0) = P$ and that $\gamma_0(1) = \gamma_1(1) = Q$. We say that γ_0 and γ_1 are *homotopic in* W (with fixed endpoints) if there is a continuous function

$$H : [0,1] \times [0,1] \to W \tag{9.3.2.1}$$

such that

(9.3.2.2) $H(0,t) = \gamma_0(t) \quad \text{for all } t \in [0,1];$

(9.3.2.3) $H(1,t) = \gamma_1(t) \quad \text{for all } t \in [0,1];$

(9.3.2.4) $H(s,0) = P \quad \text{for all } s \in [0,1];$

(9.3.2.5) $H(s,1) = Q \quad \text{for all } s \in [0,1].$

Then H is called a *homotopy* (with fixed endpoints) of the curve γ_0 to the curve γ_1. Refer to Figure 9.9 to view two curves that are homotopic and two that are not.

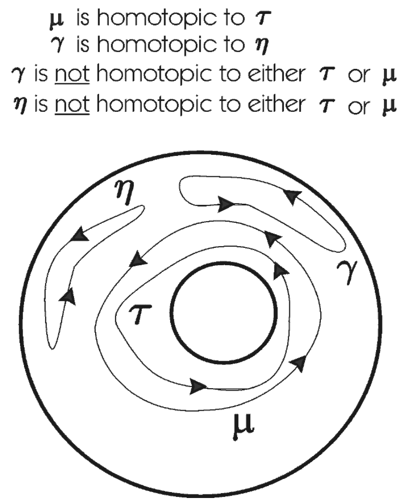

Figure 9.9: The concept of homotopy.

9.3.3 Fixed Endpoint Homotopy

Note: Since we are only interested in homotopies with fixed endpoints, we shall omit the phrase "with fixed endpoints" in the remainder of our discussion.

Intuitively, we think of a homotopy H as follows. Let $H_s(t) = H(s, t)$. Then condition (9.3.2.2) says that H_0 is the curve γ_0. Condition (9.3.2.3) says that H_1 is the curve γ_1. Condition (9.3.2.4) says that all the curves H_s begin at P. Condition (9.3.2.5) says that all the curves H_s terminate at Q. The homotopy amounts to a continuous deformation of γ_0 to γ_1 with *all curves in the process restricted to lie in W.*

We introduce one last piece of terminology:

9.3.4 Unrestricted Continuation

Let W be a domain and let (f, U) be a function element in W. We say (f, U) admits *unrestricted continuation* in W if there is an analytic continuation (f_t, U_t) of (f, U) along every curve γ that begins at P and lies in W.

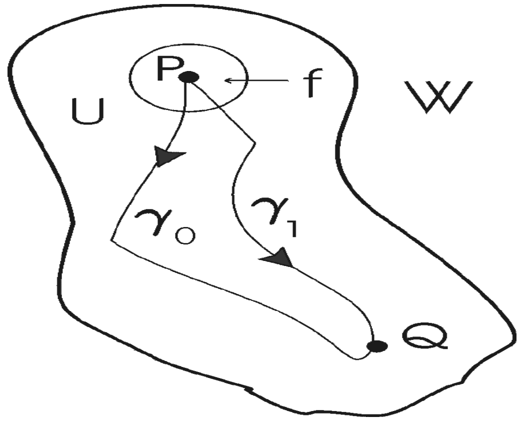

Figure 9.10: The concept of monodromy.

9.3.5 The Monodromy Theorem

One situation, in practice the primary situation, in which the question raised in §§9.3.1 always has an affirmative answer is given by the following theorem:

> **The Monodromy Theorem:** Let $W \subseteq \mathbb{C}$ be a connected open set. Let (f, U) be a function element with $U \subseteq W$. Let P denote the center of the disc U. Assume that (f, U) admits unrestricted continuation in W. If γ_0, γ_1 are each curves that begin at P, terminate at some point Q, and are homotopic in W, then the analytic continuation of (f, U) to Q along γ_0 equals the analytic continuation of (f, U) to Q along γ_1. Refer to Figure 9.10.

9.3.6 Monodromy and Globally Defined Analytic Functions

Let $W \subseteq \mathbb{C}$ be a connected open set. Assume further that W is *topologically simply connected*, in the sense that any two curves that begin at the same point and end at the same point (possibly different from the initial point) are homotopic—see the related discussion of simple connectivity in §§2.3.3. Assume that (f, U) admits unrestricted continuation in W. Then there is a globally defined holomorphic function F on W that equals f on U.

In view of the monodromy theorem, we now can understand specifically how it can be that the function $\sqrt{z}$, and more generally the function $\log z$, cannot be analytically continued in a well-defined fashion to all of $\mathbb{C}\setminus\{0\}$. The difficulty is with the two curves specified in Example 9.1.2.3, or in Example of 9.1.7.1: they are not homotopic *in the region* $\mathbb{C}\setminus\{0\}$ (which is the region of definition of the analytic function being considered).

9.4 The Idea of a Riemann Surface

9.4.1 What is a Riemann Surface?

In this section we give an intuitive description of the concept of what is called a Riemann surface.

The idea of a Riemann surface is that one can visualize geometrically the behavior of function elements and their analytic continuations. At the moment, a global analytic function is an analytic object. A global analytic function is the set of all function elements obtained by analytic continuation along curves (from a base point $P \in \mathbb{C}$) of a function element (f, U) at P. Such a set, which amounts to a collection of convergent power series at different points of the plane $\mathbb{C}$, does not seem very geometric in any sense. But in fact it can be given the structure of a surface, in the intuitive sense of that word, quite easily. (The precise definition and detailed definition of what a "surface" is would take us too far afield: we shall be content here with the informal idea that a surface is a two-dimensional object that locally "looks like" an open set in the plane. A more precise definition would be that a surface is a topological space that is locally homeomorphic to $\mathbb{C}$. See [LST], [ONE] for a more detailed discussion of surfaces.)

9.4.2 Examples of Riemann Surfaces

The idea that we need is most easily appreciated by first working with a few examples. Consider the function element (f, U) defined on $U = D(1+0i, 1/2)$ by

$$z = re^{i\theta} \mapsto r^{1/2}e^{i\theta/2}, \tag{9.4.2.1}$$

where $r > 0$ and $-\pi/4 < \theta < \pi/4$ makes the $re^{i\theta}$ representation of $z \in D(1,1)$ unique. This function element is the "principal branch of $\sqrt{z}$ " at $z = 1$ that we have already discussed in Example 9.1.2.3. The functional element (f, U) can be analytically continued along every curve γ emanating from 1 and lying in $\mathbb{C}\setminus\{0\}$. Let us denote by $\mathcal{R}$ (for "Riemann surface") the totality of all function elements obtained by such analytic continuations. Of course, in set-theoretic terms, $\mathcal{R}$ is just the global analytic function $\sqrt{z}$, just as we defined this concept earlier in §§9.1.6. All we are trying to do now is to "visualize" $\mathcal{R}$ in some sense.

Note that every point of $\mathcal{R}$ "lies over" a unique point of $\mathbb{C}\setminus\{0\}$. A function element $(f, U) \in \mathcal{R}$ is associated to the center of U, that is to say, (f, U) is a function element at a point of $\mathbb{C} \setminus \{0\}$. So we can define a "projection" $\pi : \mathcal{R} \to \mathbb{C} \setminus \{0\}$ by

$$\pi((f, U)) = \text{the center of the disc } U. \tag{9.4.2.2}$$

This is just new terminology for a situation that we have already discussed.

The projection π of $\mathcal{R}$ is two-to-one onto $\mathbb{C} \setminus \{0\}$. In a neighborhood of a given $z \in \mathbb{C} \setminus \{0\}$, there are exactly two holomorphic branches of $\sqrt{z}$. [If one of these is (f, U), then the other is $(-f, U)$. But there is no way to decide which of (f, U) and $(-f, U)$ is *the* square root in any sense that can be made to vary continuously over all of $\mathbb{C} \setminus \{0\}$.] We can think of $\mathcal{R}$ as a "surface" in the following manner:

Let us define neighborhoods of "points" (f, U) in $\mathcal{R}$ by declaring a neighborhood of (f, U) to be

$$\Big\{(f_p, U_p) : p \in U \text{ and } (f_p, U_p) \text{ is a direct analytic continuation of } (f, U) \text{ to } p\Big\}. \tag{9.4.2.3}$$

This new definition may seem formalistic and awkward. But it has the attractive property that it makes $\pi : \mathcal{R} \to \mathbb{C} \setminus \{0\}$ locally one-to-one. Every (f, U) has a neighborhood that maps, under π, one-to-one onto an open subset of $\mathbb{C} \setminus \{0\}$. This gives a way to think of $\mathcal{R}$ as being locally like an open set in the plane.

Let us try to visualize $\mathcal{R}$ still further. Let $W = \mathbb{C} \setminus \{z = x + i0 : x \leq 0\}$. Then $\pi^{-1}(W)$ decomposes naturally into two components, each of which is an open set in $\mathcal{R}$. (Since we have defined neighborhoods of points in $\mathcal{R}$, we naturally have a concept of open set in $\mathcal{R}$ as well.) These two components are "glued together" in $\mathcal{R}$ itself: $\mathcal{R}$ is connected, while $\pi^{-1}(W)$ is not. Note that, on each of the connected components of $\pi^{-1}(W)$, the projection π is one-to-one. All of this language is just a formalization of the fact that, on

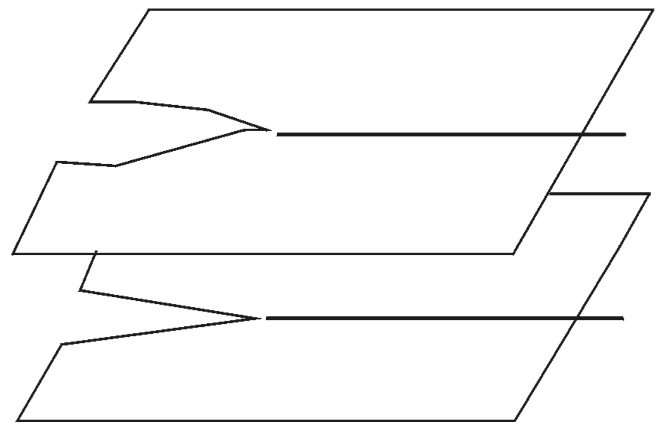

Figure 9.11: Forming a Riemann surface.

W, there are two holomorphic branches of $\sqrt{z}$—namely $re^{i\theta} \mapsto r^{1/2}e^{i\theta/2}$ and $re^{i\theta} \mapsto -r^{1/2}e^{i\theta/2}$, $-\pi < \theta < \pi$.

Each of the components of $\pi^{-1}(W)$ can be thought of as a "copy" of W, since π maps a given component one-to-one onto W. See Figure 9.11. How are these "copies," say $\mathcal{Q}_1$ and $\mathcal{Q}_2$, glued together to form $\mathcal{R}$? We join the second quadrant edge of $\mathcal{Q}_1$ to the third quadrant edge of $\mathcal{Q}_2$ and the second quadrant edge of $\mathcal{Q}_2$ to the third quadrant edge of $\mathcal{Q}_1$. Of course these joins cannot be simultaneously performed in three-dimensional space. So our picture is idealized. See Figure 9.12. Tacitly, in our construction of $\mathcal{R}$, we have restored the negative real axis.

9.4.3 The Riemann Surface for the Square Root Function

We have now constructed a surface, known as the "Riemann surface for the function $\sqrt{z}$." This surface that we have obtained by gluing together the two copies of W is in fact homeomorphic to the topological space that we made from $\mathcal{R}$ (the set of function elements) when we defined neighborhoods in $\mathcal{R}$. So we can regard our geometric surface, built from gluing the two copies of W together, and the function element space $\mathcal{R}$, as being the same thing, that is the same surface.

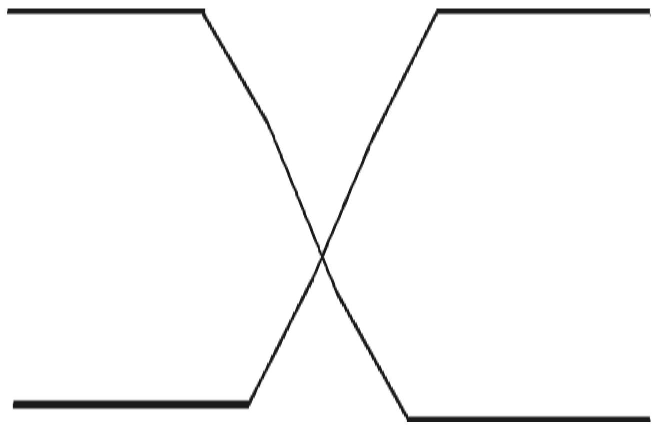

Figure 9.12: Idealized picture of a Riemann surface.

9.4.4 Holomorphic Functions on a Riemann Surface

Since $\pi : \mathcal{R} \to \mathbb{C} \setminus \{0\}$ is locally one-to-one, we can even use this projection to describe what it means for a function $F : \mathcal{R} \to \mathbb{C}$ to be holomorphic. Namely, F is holomorphic if $F \circ \pi^{-1} : \pi(U) \to \mathbb{C}$ is holomorphic for each open set U in $\mathcal{R}$ with π one-to-one on U. With this definition in mind, $f(z) = \sqrt{z}$ becomes a well-defined, "single-valued" holomorphic function on $\mathcal{R}$. To wit, if (f, U) is a function element in $\mathcal{R}$, located at $P \in \mathbb{C} \setminus \{0\}$, i.e., with $\pi(f, U) = P$, then we set

$$F((f, U)) = f(P). \qquad (9.4.4.1)$$

In this setup, $F^2((f, U)) = \pi(f^2, U) = P$ [since $f^2(z) = z$]. Therefore F is the square root function, in the sense described.

There are similar pictures for $\sqrt[n]{z}$—see Figure 9.13. Note that the Riemann surface for $\sqrt[n]{z}$ has n sheets, joined together in sequence.

9.4.5 The Riemann Surface for the Logarithm

It requires some time, and some practice, to become familiar with the construction of a Riemann surface from a given function. To get accustomed to it further, let us now discuss briefly the Riemann surface for "$\log z$" (see also Example 9.1.7.1). More precisely, we begin with the "principal branch" $re^{i\theta} \mapsto \log r + i\theta$ defined on $D(1 + 0i, 1/2)$ by requiring that $-\pi/4 < \theta < \pi/4$ and we consider all its analytic continuations along curves emanating from

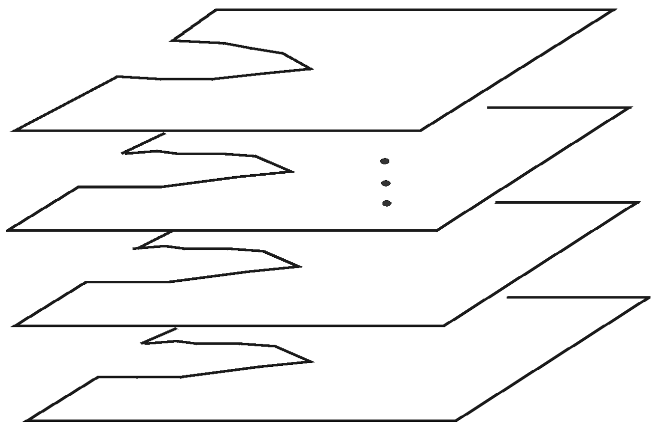

Figure 9.13: The Riemann surface for $\sqrt[n]{z}$.

1. We can visualize the "branches" here by noting that, again with $W = \mathbb{C} \setminus (\{0\} \cup$ the negative real axis), $\pi^{-1}(W)$ has infinitely many components—each a copy of W—on which π maps one-to-one onto W. Namely, these components are the "branches" of $\log z$ on W:

$$re^{i\theta} \mapsto \log r + i\theta + 2\pi i k, \qquad k \in \mathbb{Z}, \tag{9.4.5.1}$$

where $-\pi < \theta < \pi$. Picture each of these (infinitely many) images stacked one above the other (Figure 9.14). We join them in an infinite spiral, or screw, with the upper edge of the k^{th} surface being joined to the lower edge of the $(k-1)^{\text{st}}$ surface. Observe that going around the origin (counted clockwise in $\mathbb{C} \setminus \{0\}$) corresponds to going around and up one level on the spiral surface. This is the geometric representation of the fact that, when we analytically continue a branch of $\log r + i\theta + 2\pi i k$ once around the origin counterclockwise then k increases by 1. This time there is no joining of the first and last "sheets." The spiral goes on without limit in both direction.

9.4.6 Riemann Surfaces in General

The idea that we have been discussing, of building surfaces from function elements, can be carried out in complete generality: Consider the set of all analytic function elements that can be obtained by analytic continuation (along some curve in $\mathbb{C}$) of a given function element (f, U). This is what we called earlier a global analytic function. Then this set of function elements

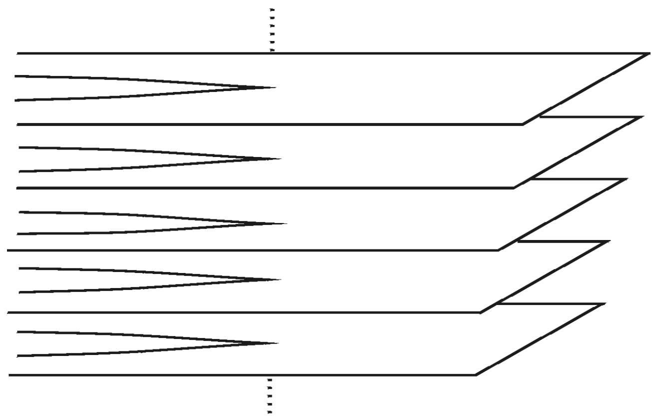

Figure 9.14: The Riemann surface for logarithm.

can actually (and always) be regarded as a connected surface: There is a projection onto an open set in $\mathbb{C}$ obtained by sending each function element to the point of $\mathbb{C}$ at which it is located. This projection is a local identification of the set of function elements with part of $\mathbb{C}$, and so it in effect exhibits the set of function elements as being two-dimensional, i.e., a surface; after this observation, everything proceeds as in the examples. The reader is invited to experiment with these new ideas—see particularly the discussion and exercises in [GK].

9.5 Picard's Theorems

9.5.1 Value Distribution for Entire Functions

The image, or set of values, of an entire function must be quite large. This statement is true in a variety of technical senses, and Nevanlinna theory gives a detailed development of the concept. Here we simply enunciate the theorems of Picard which give some sense of the robustness of entire functions.

9.5.2 Picard's Little Theorem

Theorem: If the range of a holomorphic function $f : \mathbb{C} \to \mathbb{C}$ omits two points of $\mathbb{C}$, then f is constant.

The entire function $f(z) = e^z$ shows that an entire function *can* omit one value (in this case, the value 0). But the theorem says that the only way that it can omit two values is if the function in question is constant.

9.5.3 Picard's Great Theorem

The following theorem, known as the great theorem of Picard, strengthens the Little Theorem (see §§9.5.4 for the explication of this connection).

> **Theorem:** Let U be a region in the plane, $P \in U$, and suppose that f is holomorphic on $U \setminus \{P\}$ and has an essential singularity at P. If $\epsilon > 0$, then the restriction of f to $U \cap [D(P, \epsilon) \setminus \{P\}]$ assumes all complex values except possibly one.

9.5.4 The Little Theorem, the Great Theorem, and the Casorati-Weierstrass Theorem

Compare Picard's great theorem with the Casorati/Weierstrass theorem (§§4.1.6). The Casorati/Weierstrass theorem says that, in a deleted neighborhood of an essential singularity, a holomorphic function assumes a dense set of values. Picard's theorem refines this to "all values except possibly one."

What is the connection between the great theorem and the little theorem? And how do these two theorems relate to the results about entire functions that we have already seen?

A non-constant entire function cannot be bounded near infinity, or else it would be bounded on $\mathbb{C}$ and hence (by Liouville's theorem—§§3.1.3) be constant. So it has either a pole or an essential singularity at infinity. In the first instance, the function is a polynomial (see §§4.6.6). But then the Fundamental Theorem of Algebra (§§1.1.7) tells us that the function assumes *all* complex values. In the second instance, the great theorem applies at the point ∞ and implies the little theorem.

Nevanlinna theory is an analytic refinement of the ideas we have been discussing here. There is a delicate interplay between the rate of growth (at ∞) of an entire function and the distribution of its zeros.

Bibliography

[BKR] B. E. Blank and S. G. Krantz, *Calculus*, Key College Press, Emeryville, CA, 2006.

[OH] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Wiley-Interscience, New York, 1953.

[GAR] J. B. Garnett, *Bounded Analytic Functions*, Academic Press, New York, 1981.

[GRK] R. E. Greene, and S. G. Krantz, *Function Theory of One Complex Variable*, 2^{nd} ed., American Mathematical Society, Providence, RI, 2002.

[HER] I. Herstein, *Topics in Algebra*, Blaisdell, New York, 1964.

[HES] Z. He and O. Schramm, Koebe uniformization and circle packing, *Ann. of Math.* 137(1993), 369–4013.

[HUN] T. W. Hungerford, *Algebra*, Springer-Verlag, New York, 1998.

[KRA1] S. G. Krantz, *A Panorama of Harmonic Analysis*, Mathematical Association of America, Washington, D.C., 1999.

[KRA2] S. G. Krantz, *Complex Analysis: The Geometric Viewpoint*, 2^{nd} ed., Mathematical Association of America, Washington, D.C., 2004.

[KRA3] S. G. Krantz, *Real Analysis and Foundations*, 2^{nd} ed., CRC Press, Boca Raton, FL, 2005.

[KRA4] S. G. Krantz, *Cornerstones of Geometric Function Theory: Explorations in Complex Analysis*, Birkhäuser Publishing, Boston, 2006.

[KRA5] S. G. Krantz, *Function Theory of Several Complex Variables*, 2^{nd} ed., American Mathematical Society, Providence, RI, 2001.

[KRA6] S. G. Krantz, *Complex Variables: A Physical Approach with Applications and* `MatLab`, Taylor & Francis, Boca Raton, FL, 2008.

[LOG] J. D. Logan, *Applied Mathematics*, 2^{nd} ed., John Wiley and Sons, New York, 1997.

[RUD1] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, New York, 1953.

Index