Chapter IX

Conformal Mapping

Introduction

Both from a geometric outlook and in terms of its plentiful applications one of the most appealing areas of complex analysis is the part devoted to the study of univalent analytic functions. Functions of this type are also known as "conformal mappings," for they are "angle-preserving" in a sense to be explained presently. The epoch-making event in the history of conformal mapping was the announcement by Riemann in his 1851 Göttingen dissertation of a remarkable fact: any simply connected domain that is properly contained in the complex plane can be transformed by a conformal mapping to any other domain of the same description. It is this stunning discovery of Riemann's that sits at the core of the present chapter. Before taking up the Riemann mapping theorem we spend a few words in an effort to make precise the notion of a conformal mapping, and we examine the relationship between conformality and analyticity. After this we discuss in some detail an important elementary class of conformal mappings, the Möbius transformations. We follow up our treatment of the mapping theorem itself with an introduction to the boundary behavior of conformal mappings. Finally, we determine the structure of a conformal mapping of a half-plane onto a general polygonal region.

Needless to say, there are many aspects of the theory of conformal mappings that we shall not even touch upon in this book. For instance, we restrict ourselves almost exclusively to the conformal mapping of simply connected domains, thereby avoiding the topological and analytical complications that arise in the multiply connected case. An excellent supplementary text on the subject matter of this chapter is the classic *Conformal Mapping* by Zeev Nehari (McGraw-Hill, New York, 1952). At a more advanced level a fine reference for the topic is *Univalent Functions* by Christian Pommerenke (Vandenhoeck-Ruprecht, Göttingen, 1975), as is Peter Duren's book with the same title (Springer-Verlag, New York, Berlin, 1983).

1 Conformal Mappings

1.1 Curvilinear Angles

Since a typical complex function is unlikely to map straight line segments to straight line segments — hence, to transform angles, understood in the strict sense, to other angles — in speaking of "angle-preserving transformations" we cannot take the expression literally without at the same time drastically limiting the body of admissible transformations. To arrive at an exceedingly rich class of angle-preserving mappings, however, we need only recast the concept in a curvilinear setting. The reader is no doubt aware that the standard way of measuring the angle at which two smooth curves intersect is simply to measure the angle (for definiteness, use the non-obtuse angle) formed by the tangent lines to the respective curves at their point of intersection. A suitably well-behaved function will transform one smooth curve to another, which suggests an informal working definition of the term "angle-preserving" that captures the essential spirit of the later technical definition: a complex-valued function of a complex variable is angle-preserving if it transforms each intersecting pair of smooth curves in its domain-set to a pair of smooth curves whose angle of intersection is the same as that of the original curves. In fact, we wish to be somewhat more careful than this, for we want to consider mappings that preserve angles not just in size, but also in orientation. As a prelude to that refinement, we establish some convenient notation and terminology.

Let z and w be non-zero complex numbers. We refer to the quantity $\theta(z,w) = \text{Arg}(w/z)$ as the *oriented angle from z to w*. Geometrically $\theta(z,w)$ is a measurement in $(-\pi,\pi]$ of the smaller of the two angles formed at the origin by the vectors representing z and w. It is positive if a vector that sweeps out this angle starting at "side" z and proceeding to "side" w moves in a counter-clockwise direction; if that motion is clockwise, $\theta(z,w)$ is negative. (See Figure 1. The preceding remarks do not apply in the situation where $\theta(z,w) = \pi$, for then the vectors in question form a straight line

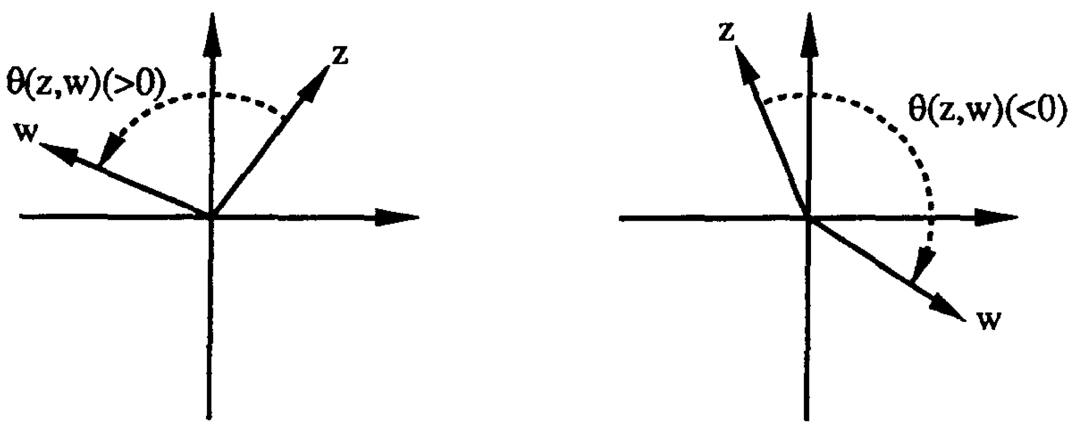


Figure 1.

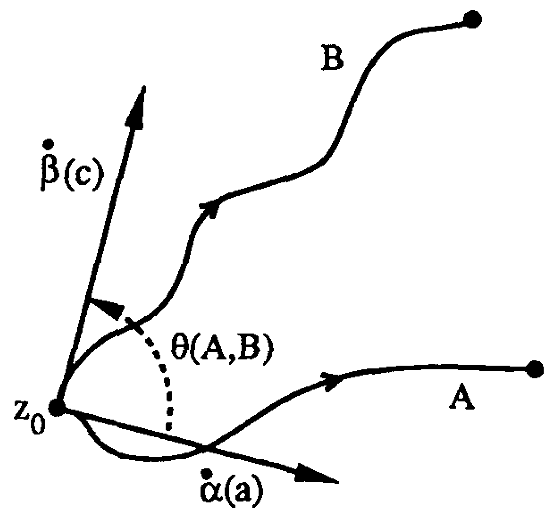


Figure 2.

segment passing through the origin, so no "smaller of the two angles" can be distinguished.) Observe that $\theta(w,z) = -\theta(z,w)$ and $\theta(\overline{z},\overline{w}) = -\theta(z,w)$ except when $\theta(z,w) = \pi$, in which case both $\theta(w,z)$ and $\theta(\overline{z},\overline{w})$ are π as well. Notice also that $\theta(cz,cw) = \theta(z,w)$ for any non-zero complex number c and that $\theta(rz,sw) = \theta(z,w)$ when r and s are positive real numbers.

As the name implies, the sides of a "curvilinear angle" are no longer required to be line segments, but instead are allowed to be more general smooth curves. To be precise, we introduce the expression ***regular arc*** to describe a set A that admits a representation of the sort $A = |\alpha|$, where $\alpha\colon [a,b] \to \mathbb{C}$ is a smooth, simple, non-closed path enjoying the added feature that $\dot{\alpha}(t) \neq 0$ for every t in $[a,b]$. Any such α is termed a ***regular parametrization*** of A. The points $\alpha(a)$ and $\alpha(b)$ are called the ***endpoints*** of A. Suppose that A and B are regular arcs which have one endpoint z_0 in common, but which are otherwise disjoint (Figure 2). In this event we say that A and B are the sides of a ***curvilinear angle with vertex at*** z_0, and we define $\theta(A,B)$, the ***oriented angle from A to B***, by

$$\theta(A,B) = \theta[\dot{\alpha}(a), \dot{\beta}(c)]\ , \tag{9.1}$$

where $\alpha\colon [a,b] \to \mathbb{C}$ and $\beta\colon [c,d] \to \mathbb{C}$ are arbitrary regular parametrizations of A and B, respectively, having initial point z_0. (N.B. Definition (9.1) does not really depend on the regular parametrizations α and β chosen for A and B, since

$$\theta[\dot{\alpha}(a), \dot{\beta}(c)] = \theta\left[\frac{\dot{\alpha}(a)}{|\dot{\alpha}(a)|}\ ,\ \frac{\dot{\beta}(c)}{|\dot{\beta}(c)|}\right]$$

and since the quantities $\dot{\alpha}(a)/|\dot{\alpha}(a)|$ and $\dot{\beta}(c)/|\dot{\beta}(c)|$ can be determined without reference to any parametrizations. For instance, it can be shown

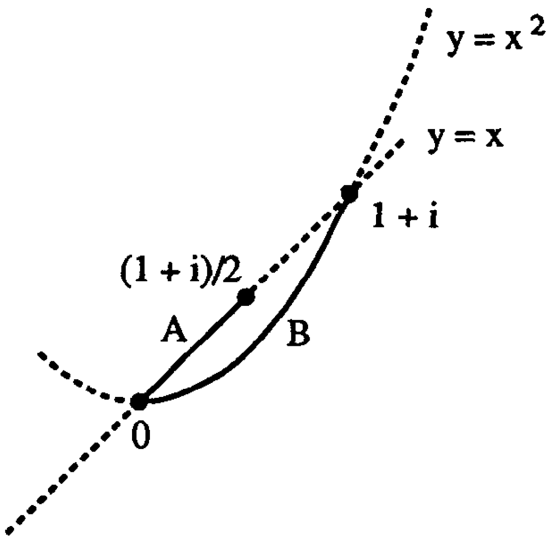


Figure 3.

that $\dot{\alpha}(a)/|\dot{\alpha}(a)|$ is just the limit of $(z - z_0)/|z - z_0|$ as z tends to z_0 along A. A similar statement holds for $\dot{\beta}(c)/|\dot{\beta}(c)|$.) As a simple illustration, consider the curvilinear angle pictured in Figure 3. Here the arc A lies on the line $y = x$, and B is a subset of the parabola $y = x^2$. We can use $\alpha(t) = t + it$ on the interval $[0, 1/2]$ and $\beta(t) = t + it^2$ on $[0, 1]$ as our regular parametrizations in computing

$$\theta(A, B) = \theta[\dot{\alpha}(0), \dot{\beta}(0)] = \theta(1 + i, 1) = \operatorname{Arg}[1/(1 + i)]$$
$$= \operatorname{Arg}[(1 - i)/2] = -\pi/4 .$$

1.2 Diffeomorphisms

An object of considerable importance both for the immediate discussion and for subsequent developments in this chapter is the *Jacobian* J_f of a function $f = u + iv$. Assuming that U is an open set in the complex plane and that $f\colon U \to \mathbb{C}$ is a member of the class $C^1(U)$, J_f is the continuous real-valued function defined in U by

$$J_f(z) = u_x(z)v_y(z) - u_y(z)v_x(z) , \tag{9.2}$$

or, expressed in terms of z- and $\bar{z}$-derivatives,

$$J_f(z) = |f_z(z)|^2 - |f_{\bar{z}}(z)|^2 . \tag{9.3}$$

In particular, if f happens to be differentiable at a point z_0 of U, then $f_z(z_0) = f'(z_0)$ and $f_{\bar{z}}(z_0) = 0$, with the result that

$$J_f(z_0) = |f'(z_0)|^2 . \tag{9.4}$$

In saying what is required of a function for it to qualify as a conformal mapping we shall presuppose a certain amount of regularity on the part of the function in question; specifically, we shall insist that it be continuously differentiable in the real sense and have non-vanishing Jacobian. Thus, given a domain D in the complex plane, we are going to consider univalent functions $f: D \to \mathbb{C}$ that belong to the class $C^1(D)$ and obey the condition $J_f(z) \neq 0$ for every z in D. A mapping of this type is termed a *diffeomorphism* — or, more accurately, a *C^1-diffeomorphism — of D onto its range.* (N.B. While it is possible to introduce conformal mappings without imposing such strong preconditions on their smoothness, the resulting discussion becomes far too technical for a book of this level.) We remark that, if $f: D \to \mathbb{C}$ is a diffeomorphism, then either $J_f > 0$ everywhere in D or $J_f < 0$ throughout the domain. This follows from the fact that D is connected and $J_f: D \to \mathbb{R}$ is a continuous, zero-free function: the set $J_f(D)$ is a connected set of real numbers that does not contain zero — hence, $J_f(D)$ is either a subset of $(-\infty, 0)$ or a subset of $(0, \infty)$. When J_f is positive in D, the diffeomorphism f is called *orientation-preserving* (or *sense-preserving*); a diffeomorphism with negative Jacobian is said to be *orientation-reversing* (or *sense-reversing*). Notice that the conjugate $\overline{f}$ of a diffeomorphism $f: D \to \mathbb{C}$ is also a diffeomorphism, one for which

$$J_{\overline{f}} = -J_f \ .$$

Therefore $\overline{f}$ is orientation-reversing when f is orientation-preserving, and vice versa.

Suppose now that $f: D \to \mathbb{C}$ is a diffeomorphism and that A is a regular arc located in D. Then $f(A)$ is also a regular arc. To see this, choose a regular parametrization for A, say $\alpha(t) = x(t) + iy(t)$ on $[a, b]$. Since f is univalent and is a member of the class $C^1(D)$, $\beta(t) = f[\alpha(t)]$ defines a smooth, simple, non-closed path with $|\beta| = f(A)$. Could it be the case that $\dot{\beta}(t_0) = 0$ for some t_0 in $[a, b]$? Assuming that this were so, we could write $f = u + iv$ and employ the chain rule to get

$$0 = \dot{\beta}(t_0) = u_x(z_0)\dot{x}(t_0) + u_y(z_0)\dot{y}(t_0) + i\left[v_x(z_0)\dot{x}(t_0) + v_y(z_0)\dot{y}(t_0)\right] \ ,$$

where $z_0 = \alpha(t_0)$. It would follow that

$$(9.5) \quad u_x(z_0)\dot{x}(t_0) + u_y(z_0)\dot{y}(t_0) = 0 \ , \ v_x(z_0)\dot{x}(t_0) + v_y(z_0)\dot{y}(t_0) = 0 \ .$$

The fact that $J_f(z_0) \neq 0$ would then permit us to solve (9.5) for $\dot{x}(t_0)$ and $\dot{y}(t_0)$. The result: $\dot{x}(t_0) = \dot{y}(t_0) = 0$. This would mean, however, that $\dot{\alpha}(t_0) = \dot{x}(t_0) + i\dot{y}(t_0) = 0$, contrary to the assumption that α is a regular parametrization of A. Accordingly, $\dot{\beta}(t) \neq 0$ must be true for every t in $[a, b]$, so β furnishes a regular parametrization of $f(A)$. A consequence of the preceding considerations is this: *a diffeomorphism $f: D \to \mathbb{C}$ transforms a curvilinear angle in D with sides A and B to a curvilinear angle in $D' = f(D)$ with sides $f(A)$ and $f(B)$* (Figure 4).

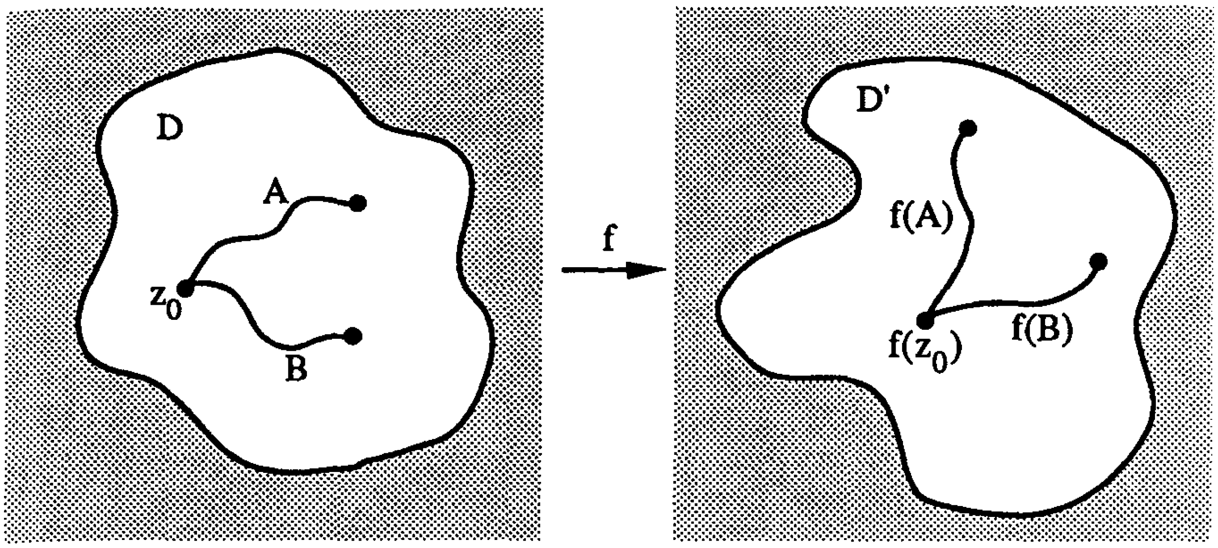


Figure 4.

1.3 Conformal Mappings

We are now ready to turn our earlier informal definition of an angle-preserving transformation into a proper one. As announced, we shall restrict attention to diffeomorphisms. A diffeomorphism $f: D \to \mathbb{C}$ is said to be ***angle-preserving*** (also called ***isogonal***) ***at a point*** z_0 of D under the condition that

$$(9.6) \qquad |\theta(A, B)| = |\theta[f(A), f(B)]|$$

whenever A and B are the sides of a curvilinear angle in D with vertex at z_0. It can be inferred from the soon-to-come Theorem 1.1 that condition (9.6) allows for two eventualities: if $J_f(z_0) > 0$, it reduces to

$$(9.7) \qquad \theta(A, B) = \theta[f(A), f(B)]$$

for all relevant A and B; if $J_f(z_0) < 0$, it becomes

$$(9.8) \qquad \theta(A, B) = \theta[f(B), f(A)] \; .$$

When (9.7) holds we say that f is ***conformal at*** z_0; in the alternative situation (9.8) we pronounce f ***anti-conformal at*** z_0. Thus, conformality demands that curvilinear angles be preserved not only in size, but also in sense. In the anti-conformal case, on the other hand, such angles are preserved in size, but their orientation gets reversed. It is not at all necessary, of course, for f to be angle-preserving at any point of D. Furthermore, since J_f does not undergo a change of sign in the domain D, it is not possible for f to be conformal at one point of D and anti-conformal at another. If the diffeomorphism f is conformal (anti-conformal) at every point of D, we call f a ***conformal*** (***anti-conformal***) ***mapping of*** D. For example, it is easily checked that a sense-preserving similarity transformation $f(z) = az + b$ —

remember that $a \neq 0$ here — is a conformal mapping of $\mathbb{C}$ onto itself. The canonical example of an anti-conformal mapping of the complex plane onto itself is found in the function $f(z) = \bar{z}$ (Exercise 6.5). A straightforward consequence of this last fact is that a diffeomorphism f is anti-conformal at a point z_0 if and only if its conjugate $\bar{f}$ is conformal there (Exercise 6.6).

The following theorem forges the link between the present discussion and the theory of analytic functions.

Theorem 1.1. *Let D be a domain in the complex plane, and let $f: D \to \mathbb{C}$ be a diffeomorphism. The following three statements concerning a point z_0 of D are equivalent: (i) f is differentiable at z_0; (ii) f is isogonal at z_0 and $J_f(z_0) > 0$; (iii) f is conformal at z_0.*

Proof. We first prove that (i) implies both (ii) and (iii). Assuming that f is differentiable at z_0, we infer from (9.4) that $J_f(z_0) = |f'(z_0)|^2 \geq 0$ and conclude, since $J_f(z_0) \neq 0$, that $J_f(z_0) > 0$ and $f'(z_0) \neq 0$. We establish the conformality of f at z_0, a property which encompasses isogonality at that point. Let A and B be the sides of a curvilinear angle in D with vertex at z_0. Choose regular parametrizations $\alpha: [a, b] \to \mathbb{C}$ for A and $\beta : [c, d] \to \mathbb{C}$ for B with $\alpha(a) = \beta(c) = z_0$. Then $\alpha_1(t) = f[\alpha(t)]$ and $\beta_1(t) = f[\beta(t)]$ provide regular parametrizations for $f(A)$ and $f(B)$, respectively. From the differentiability of f at z_0 it follows that

$$\dot{\alpha}_1(a) = f'[\alpha(a)]\dot{\alpha}(a) = f'(z_0)\dot{\alpha}(a)$$

and, similarly, $\dot{\beta}_1(c) = f'(z_0)\dot{\beta}(c)$. Because $f'(z_0) \neq 0$, we see that

$$\theta[f(A), f(B)] = \theta[\dot{\alpha}_1(a), \dot{\beta}_1(c)] = \theta[f'(z_0)\dot{\alpha}(a), f'(z_0)\dot{\beta}(c)]$$

$$= \theta[\dot{\alpha}(a), \dot{\beta}(c)] = \theta(A, B) ,$$

verifying the conformality of f at z_0.

We next derive (i) from (ii). Assume, therefore, that f is isogonal at z_0 and that $J_f(z_0) > 0$. As a diffeomorphism, the function f belongs to the class $C^1(D)$, which implies that it is differentiable in the real sense at each point of D (Theorem III.5.1). In particular, we are entitled to write

$$f(z) = f(z_0) + c\,(z - z_0) + d\,(\bar{z} - \bar{z}_0) + E(z) \tag{9.9}$$

for z in D, where $c = f_z(z_0)$, $d = f_{\bar{z}}(z_0)$, and E is a function satisfying $|E(z)|/|z - z_0| \to 0$ as $z \to z_0$. The problem rests in showing that $d = 0$, a condition which, being a restatement of the Cauchy-Riemann relations at z_0, makes plain that f is differentiable at z_0 with $f'(z_0) = c$.

We fix $r > 0$ with the property that the closed disk $\overline{\Delta}(z_0, r)$ is contained in D. When $0 \leq \psi \leq \pi$, we let A_ψ denote the regular arc parametrized by $\alpha_\psi(t) = z_0 + te^{i\psi}$ for $0 \leq t \leq r$. If $0 < \psi \leq \pi$, then the arcs A_0 and A_ψ are the sides of a curvilinear angle — in fact, a true angle — in D with vertex

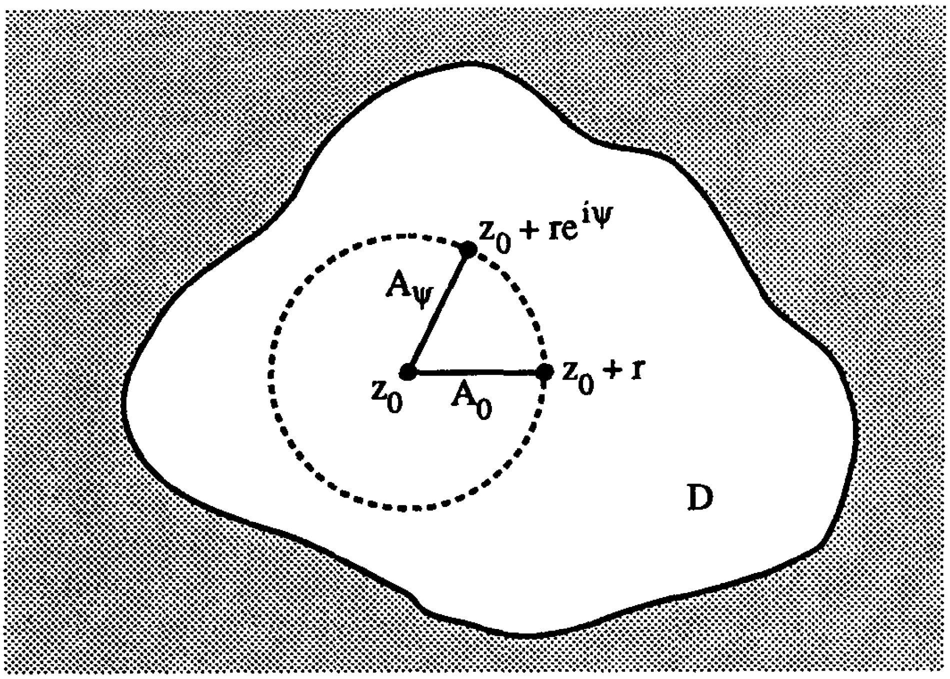


Figure 5.

at z_0 (Figure 5). Also, $\dot{\alpha}_\psi(0) = e^{i\psi}$, which implies that $\theta(A_0, A_\psi) = \psi$. The image arc $f(A_\psi)$ has the regular parametrization $\beta_\psi(t) = f[\alpha_\psi(t)] = f(z_0 + te^{i\psi})$ for $0 \le t \le r$. We use (9.9) to compute

$$\dot{\beta}_\psi(0) = \lim_{t \to 0+} \frac{f(z_0 + te^{i\psi}) - f(z_0)}{t}$$

$$= \lim_{t \to 0+} \left[ce^{i\psi} + de^{-i\psi} + \frac{e^{i\psi} E(z_0 + te^{i\psi})}{te^{i\psi}} \right] = ce^{i\psi} + de^{-i\psi} .$$

Note especially the implication that $ce^{i\psi} + de^{-i\psi} \neq 0$ for $0 \le \psi \le \pi$. We can thus safely write

$$\theta[f(A_0), f(A_\psi)] = \operatorname{Arg}\left(\frac{ce^{i\psi} + de^{-i\psi}}{c + d} \right)$$

when $0 < \psi \le \pi$. Since f is isogonal at z_0, we draw the conclusion that

$$\left| \operatorname{Arg}\left(\frac{ce^{i\psi} + de^{-i\psi}}{c + d} \right) \right| = \psi \tag{9.10}$$

for $0 \le \psi \le \pi$. Because $c\bar{d}e^{i\psi} + \bar{c}de^{-i\psi} = 2\operatorname{Re}(c\bar{d}e^{i\psi})$ is a real quantity and because — recall (9.3) — $|c|^2 - |d|^2 = J_f(z_0) > 0$, we learn that for such ψ

$$\operatorname{Im}\left(\frac{ce^{i\psi} + de^{-i\psi}}{c + d} \right) = \operatorname{Im}\left[\frac{(ce^{i\psi} + de^{-i\psi})(\bar{c} + \bar{d})}{|c + d|^2} \right]$$

$$= \frac{\operatorname{Im}\left(|c|^2 e^{i\psi} + c\bar{d}e^{i\psi} + \bar{c}de^{-i\psi} + |d|^2 e^{-i\psi} \right)}{|c + d|^2}$$

$$= \frac{\operatorname{Im}(|c|^2 e^{i\psi} + |d|^2 e^{-i\psi})}{|c+d|^2} = \frac{(|c|^2 - |d|^2)\sin\psi}{|c+d|^2} \geq 0 \,.$$

The fact that the imaginary part of $(ce^{i\psi} + de^{-i\psi})/(c+d)$ is non-negative implies that the principal argument of this number lies in the interval $[0, \pi]$, so (9.10) simplifies to

$$\text{(9.11)} \qquad \operatorname{Arg}\left(\frac{ce^{i\psi} + de^{-i\psi}}{c+d}\right) = \psi = \operatorname{Arg} e^{i\psi} \,.$$

This means that, for every ψ in $[0, \pi]$, $(ce^{i\psi} + de^{-i\psi})/(c+d)$ is a positive real multiple of $e^{i\psi}$ — or, what amounts to the same thing, the quantity $(c + de^{-2i\psi})/(c+d)$ is real and positive. Should $d \neq 0$, however, the set of points $\{(c + de^{-2i\psi})/(c+d) \colon 0 \leq \psi \leq \pi\}$ would be a circle with center $c/(c+d)$ and radius $|d|/|c+d|$, a set definitely not contained in the positive real axis. The alternative: $d = 0$ and $f'(z_0)$ exists, as claimed.

The proof that (iii) implies (i) is basically the same as the one presented for the implication (ii)⇒(i), only simpler. Namely, conformality on the part of f at z_0 allows one to by-pass the computation leading from (9.10) to (9.11); (9.11) is a direct consequence of conformality at z_0. ∎

There is an analogue of Theorem 1.1 treating the anti-conformal case. We formulate it in an exercise (Exercise 6.8). The global version of Theorem 1.1 marks plane conformal mapping theory as a special topic in the theory of analytic functions.

Theorem 1.2. *Let D be a domain in the complex plane. A function $f \colon D \to \mathbb{C}$ is a conformal mapping of D if and only if f is a univalent analytic function.*

Proof. If f is a conformal mapping of D, then f is by definition a diffeomorphism, while Theorem 1.1 affirms that f is differentiable at each point of D; i.e., f is analytic in D. Conversely, if f is a univalent analytic function, then certainly f is a member of the class $C^1(D)$, and, in view of Theorem VIII.3.9, $J_f(z) = |f'(z)|^2 > 0$ throughout D. Consequently, f is seen to be a diffeomorphism. From Theorem 1.1 we learn that f is conformal at every point of D, making f a conformal mapping of D. ∎

It follows from Theorem 1.2 without further ado that the composition of conformal mappings is conformal. Coupled with Theorem VIII.3.11, Theorem 1.2 tells us that the inverse of a conformal mapping is likewise conformal.

Let D be a plane domain. Some authors pin the label "conformal mapping" on any analytic function $f \colon D \to \mathbb{C}$ that obeys the condition $f'(z) \neq 0$ for every z in D, the requirement of univalence being waived. (The exponential function $f(z) = e^z$ in $D = \mathbb{C}$ would be an example.) We prefer to characterize a mapping f of this type as *locally conformal* in D. Indeed, Theorem VIII.3.10 asserts that there exists corresponding to any point z_0 of D a subdomain G of D containing z_0 in which f is both analytic and univalent; i.e., to which the restriction of f is a conformal mapping according to our understanding of the term.

1.4 Some Standard Conformal Mappings

We compile here for handy reference a short list of examples of elementary conformal mappings.

EXAMPLE 1.1. The function $f(z) = (1 - z)/(1 + z)$ maps the disk $D = \Delta(0, 1)$ conformally onto the half-plane $D' = \{w\colon \operatorname{Re} w > 0\}$ (Figure 6).

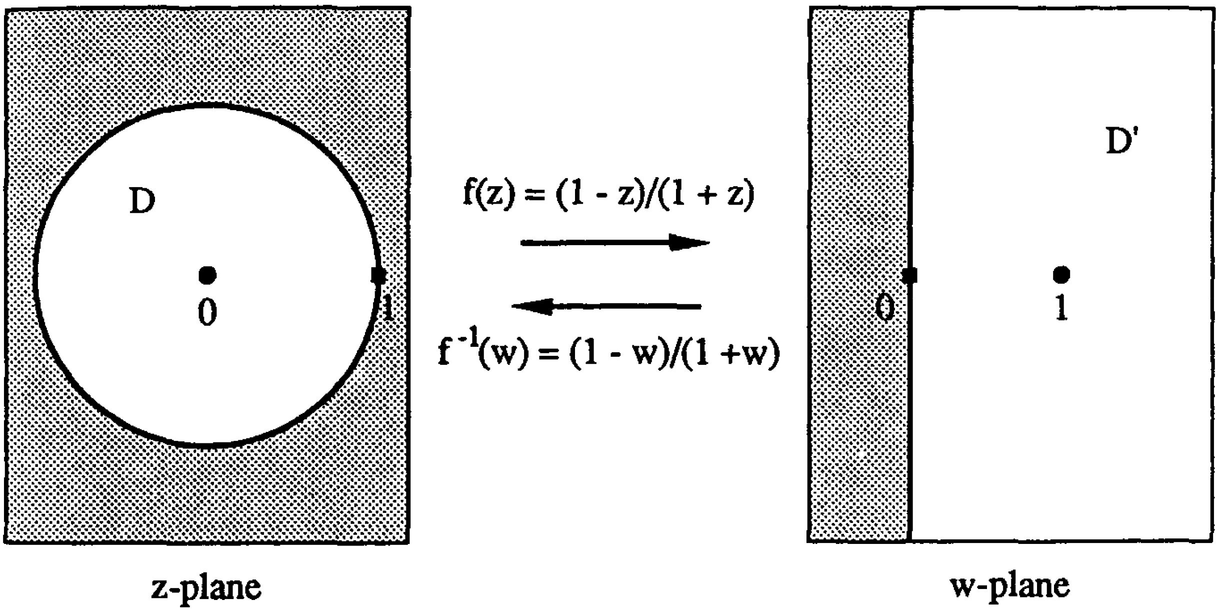


Figure 6.

The function f is analytic in D and maps D in a univalent fashion onto D'. (See Example I.3.4.) Theorem 1.2 certifies f as a conformal mapping of D onto D'. Note that $f^{-1} = f$ here, so f transforms D' conformally onto D as well.

EXAMPLE 1.2. Let $D = \mathbb{C} \sim (-\infty, 0]$ and let $0 < \lambda \leq 1$. The function $f\colon D \to \mathbb{C}$ defined by $f(z) = z^\lambda$ maps D conformally onto the angular region $D' = \{w\colon |\operatorname{Arg} w| < \lambda\pi\}$ (Figure 7).

For $0 < \lambda \leq 1$ and z in D we note that $|z^\lambda| = |z|^\lambda$ and $\operatorname{Arg}(z^\lambda) = \lambda \operatorname{Arg} z$. From these two facts it follows without difficulty that f is a univalent function with range D'. Since f is also plainly analytic, f is a conformal mapping of D onto D'. The inverse mapping $f^{-1}\colon D' \to D$ is given by $f^{-1}(w) = w^{1/\lambda}$. Observe that the restriction of f to $D \cap \Delta(0, r)$ gives a conformal mapping of this intersection onto $D' \cap \Delta(0, r^\lambda)$. (N.B. When $\lambda > 1$, $f(z) = z^\lambda$ is still analytic in D, but no longer univalent there (Ex-

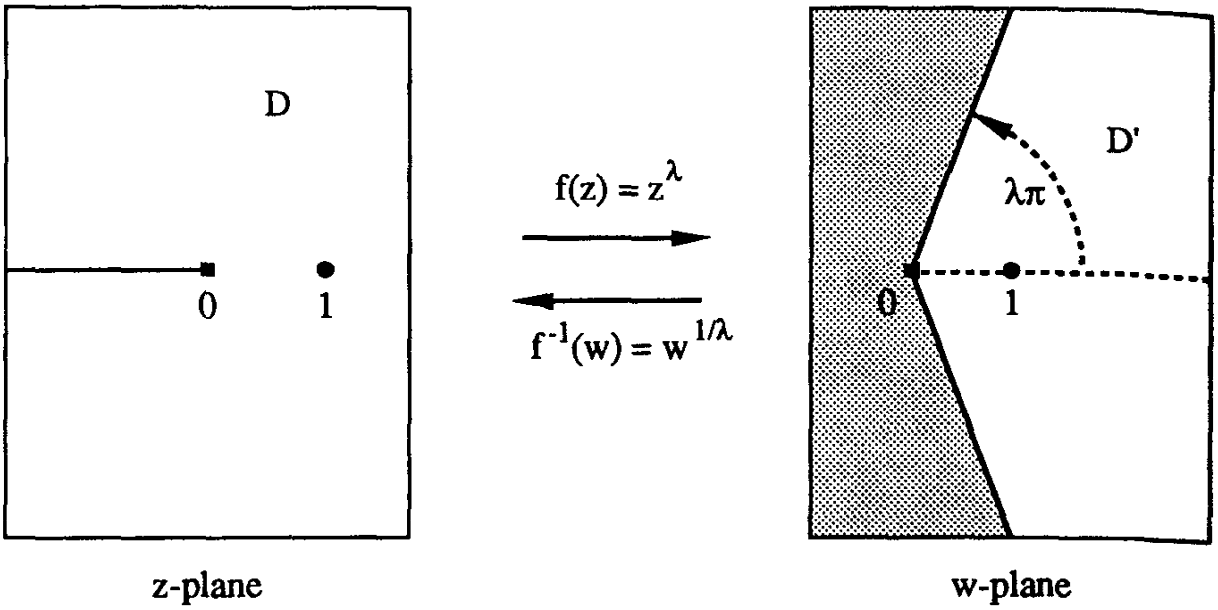


Figure 7.

ercise 6.10). Adhering to our definition of the concept, $f(z) = z^\lambda$ does not qualify as a conformal mapping of D for such λ. It does, however, meet the test to be a locally conformal mapping of D.)

EXAMPLE 1.3. If $-\infty \leq a < b \leq \infty$ and $-\pi \leq \alpha < \beta \leq \pi$, the exponential function $f(z) = e^z$ provides a conformal mapping of the domain $D = \{z : a < x < b, \alpha < y < \beta\}$ onto the region D' described by $D' = \{w : e^a < |w| < e^b,\ \ \alpha < \operatorname{Arg} w < \beta\}$ (Figure 8). The inverse $f^{-1}: D' \to D$ of this mapping is the restriction of the principal logarithm to D'; i.e., $f^{-1}(w) = \operatorname{Log} w$ for w in D'.

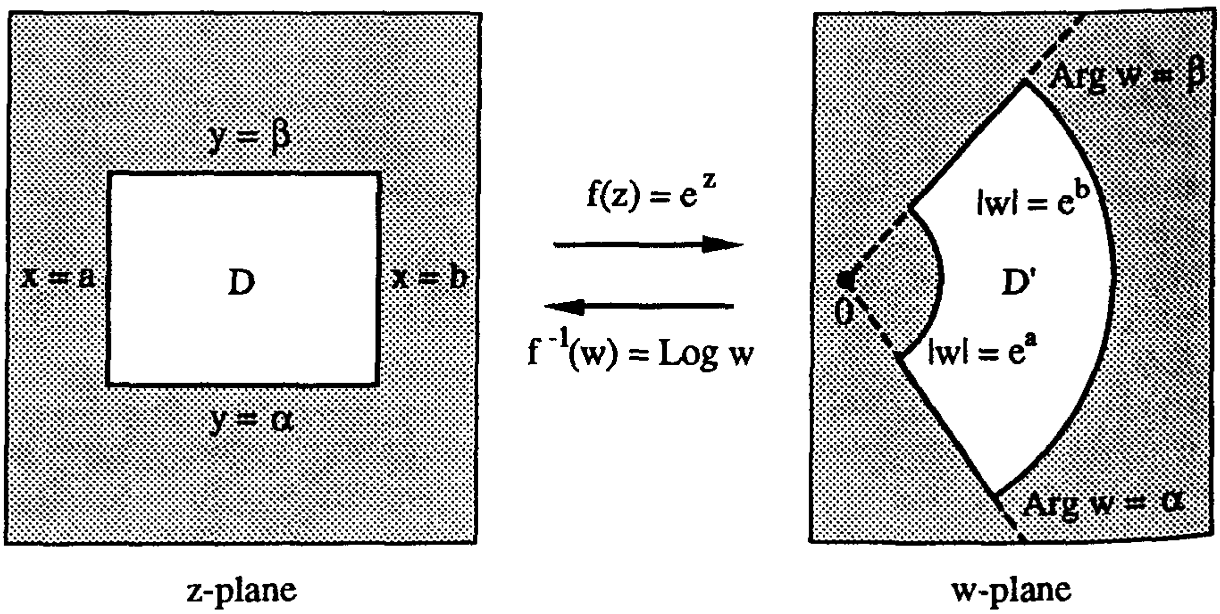


Figure 8.

EXAMPLE 1.4. The function $f(z) = \sin z$ maps the infinite strip $D = \{z : |\operatorname{Re} z| < \pi/2\}$ conformally onto the domain $D' = \mathbb{C} \sim \{w : |\operatorname{Re} w| \geq 1$ and $\operatorname{Im} w = 0\}$. (See Figure 9. We refer the reader to Example III.3.7 for details.) Here $f^{-1}: D' \to D$ is the restriction to D' of the principal arcsine function.

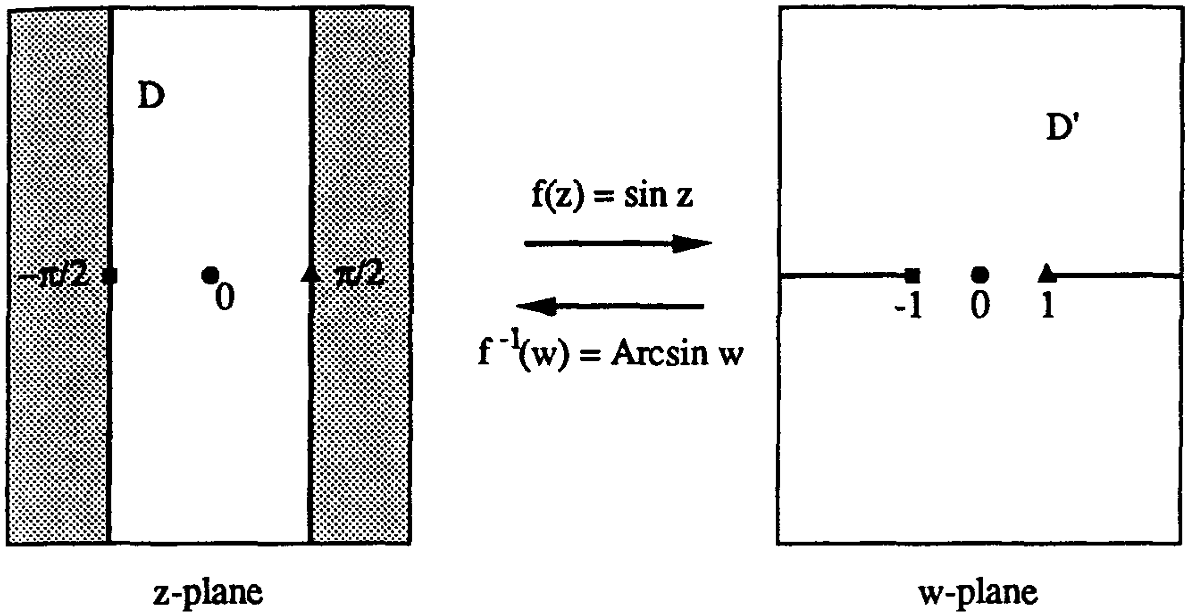


Figure 9.

EXAMPLE 1.5. The function $f(z) = \tan z$ is one that transforms the infinite strip $D = \{z : |\operatorname{Re} z| < \pi/2\}$ conformally to the domain $D' = \mathbb{C} \sim \{w : |\operatorname{Im} w| \geq 1$ and $\operatorname{Re} w = 0\}$. (See Figure 10. Recall, too, Example II.3.8.) The restriction of the principal arctangent function to D' furnishes the inverse mapping in this example.

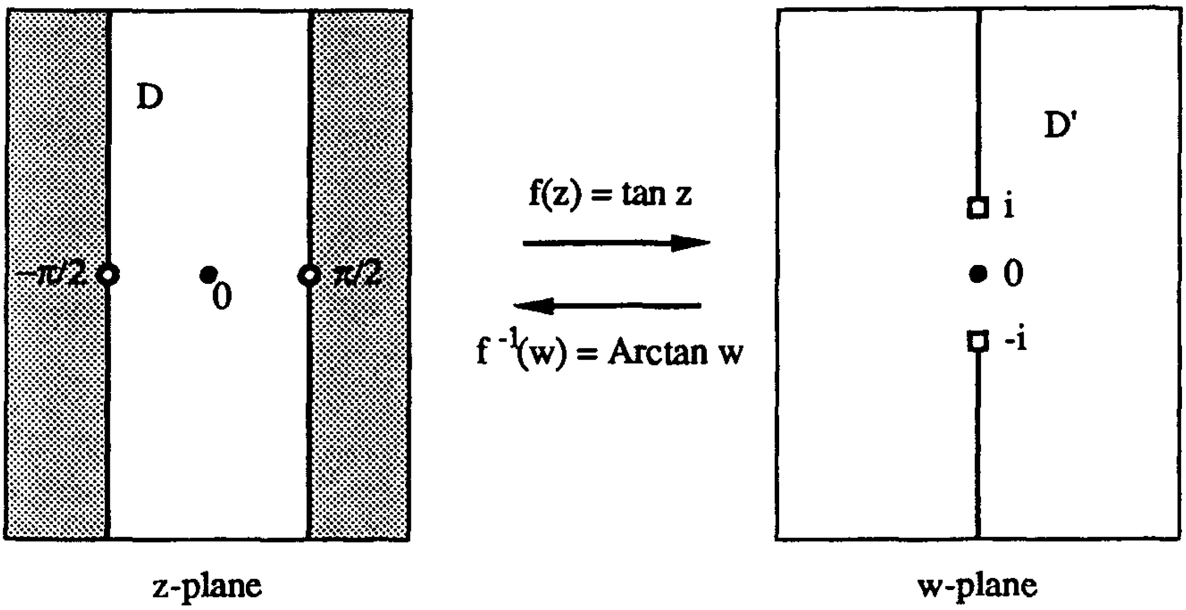


Figure 10.

Notice that $g(z) = i\sin z$ delivers a second function which maps the strip D conformally onto D', but maps it quite differently than f does. For example, f transforms the interval $(-\pi/2, \pi/2)$ to $(-\infty, \infty)$, whereas g maps $(-\pi/2, \pi/2)$ to the interval between $-i$ and i on the imaginary axis.

EXAMPLE 1.6. The restriction of the function $f(z) = (z + z^{-1})/2$ to the domain $D = \{z : |z| > 1\}$ gives a conformal mapping of D onto the domain $D' = \mathbb{C} \sim [-1, 1]$.

When $|z| = 1$, we observe that

$$f(z) = (z + z^{-1})/2 = (z + \bar{z})/2 = \operatorname{Re} z ,$$

so $f(z)$ is real and belongs to the interval $[-1, 1] = \mathbb{C} \sim D'$. For w other than 1 or -1 the equation $(z + z^{-1})/2 = w$ has exactly two solutions, $z_1 = w + \sqrt{w^2 - 1}$ and $z_2 = w - \sqrt{w^2 - 1}$. Furthermore, if w is not a point of $[-1, 1]$, then by the preceding remark neither z_1 nor z_2 can have unit modulus. Since $|z_1||z_2| = |z_1 z_2| = 1$, one of these points must lie in D, the other in the disk $\Delta(0, 1)$. In fact, it can be verified that $|z_1| > 1$ for any w in D' such that $\operatorname{Re} w > 0$ or such that $\operatorname{Re} w = 0$ and $\operatorname{Im} w > 0$, whereas $|z_2| > 1$ for all other elements w of D'. (We remind the reader of Exercise I.4.38.) On the basis of the preceding information we can assert that the restriction of f to D is a univalent function whose range is D'. As f is analytic in D, this restriction is a conformal mapping of D onto D'. The inverse mapping is a bit awkward to write down here, its values being provided by z_1 at some points w of D' and z_2 at others, as just pointed out. (N.B. The same function f maps the punctured disk $\Delta^*(0, 1)$ conformally onto D'.)

The remaining two examples indicate how the foregoing elementary mappings can be used to construct more complicated conformal mappings.

EXAMPLE 1.7. Find a conformal mapping of the infinite strip $D = \{z : |\operatorname{Im} z| < \pi/2\}$ onto the disk $D' = \Delta(0, 1)$.

As suggested by Figure 11, we can arrive at such a mapping f by first transforming D with a conformal mapping f_1 to the half-space D_1, then choosing a conformal mapping f_2 of D_1 onto D', and taking $f = f_2 \circ f_1$. We actually carry this out by using $f_1(z) = e^z$ and $f_2(z) = (1 - z)/(1 + z)$, although other choices would certainly be possible. The composite function $f(z) = (1 - e^z)/(1 + e^z)$ is thus a conformal mapping of D onto D'.

EXAMPLE 1.8. Construct a conformal mapping of the infinite strip $D = \{z : 0 < \operatorname{Re} z < \pi\}$ onto the domain $D' = \{w : |\operatorname{Arg} w| < \pi/4\} \sim [1, \infty)$.

Again we build f in stages: $f = f_3 \circ f_2 \circ f_1$, where the component mappings meet the requirements implicit in Figure 12. Specifically, we take $f_1(z) = z/2$, $f_2(z) = \sin z$, and $f_3(z) = \sqrt{z}$. They compose to produce $f(z) = \sqrt{\sin(z/2)}$, a function that maps D conformally onto D'.

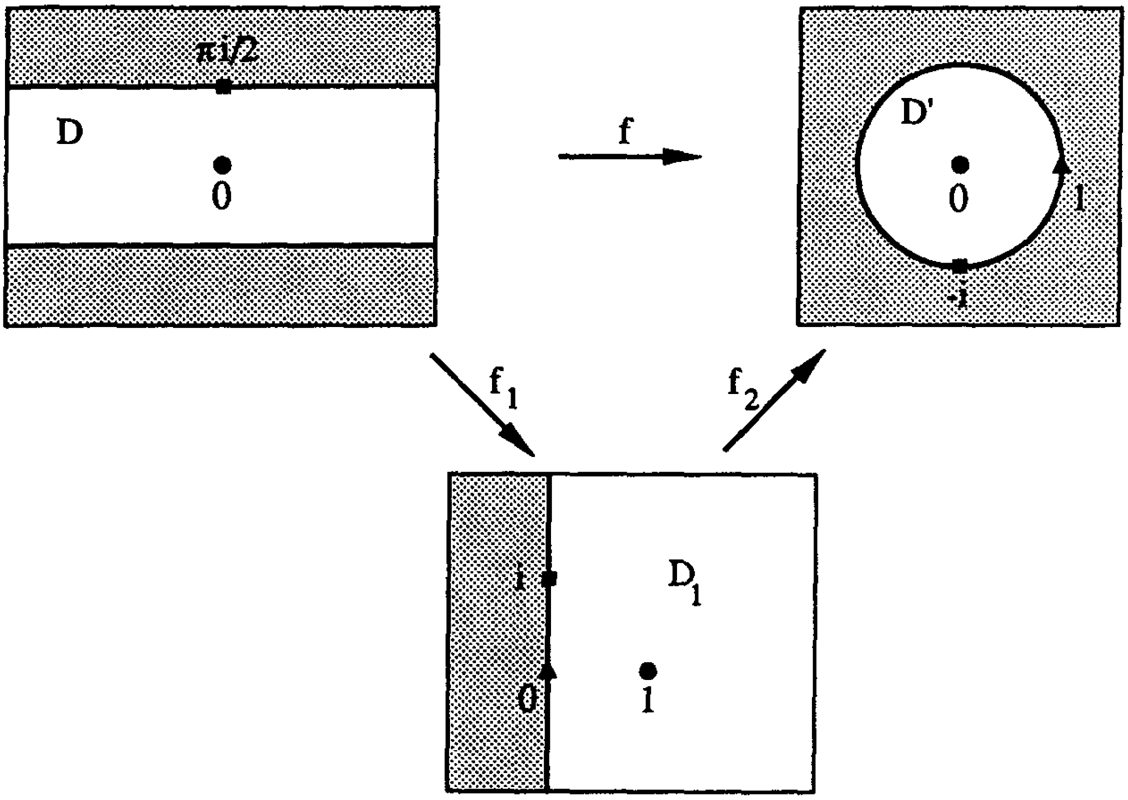


Figure 11.

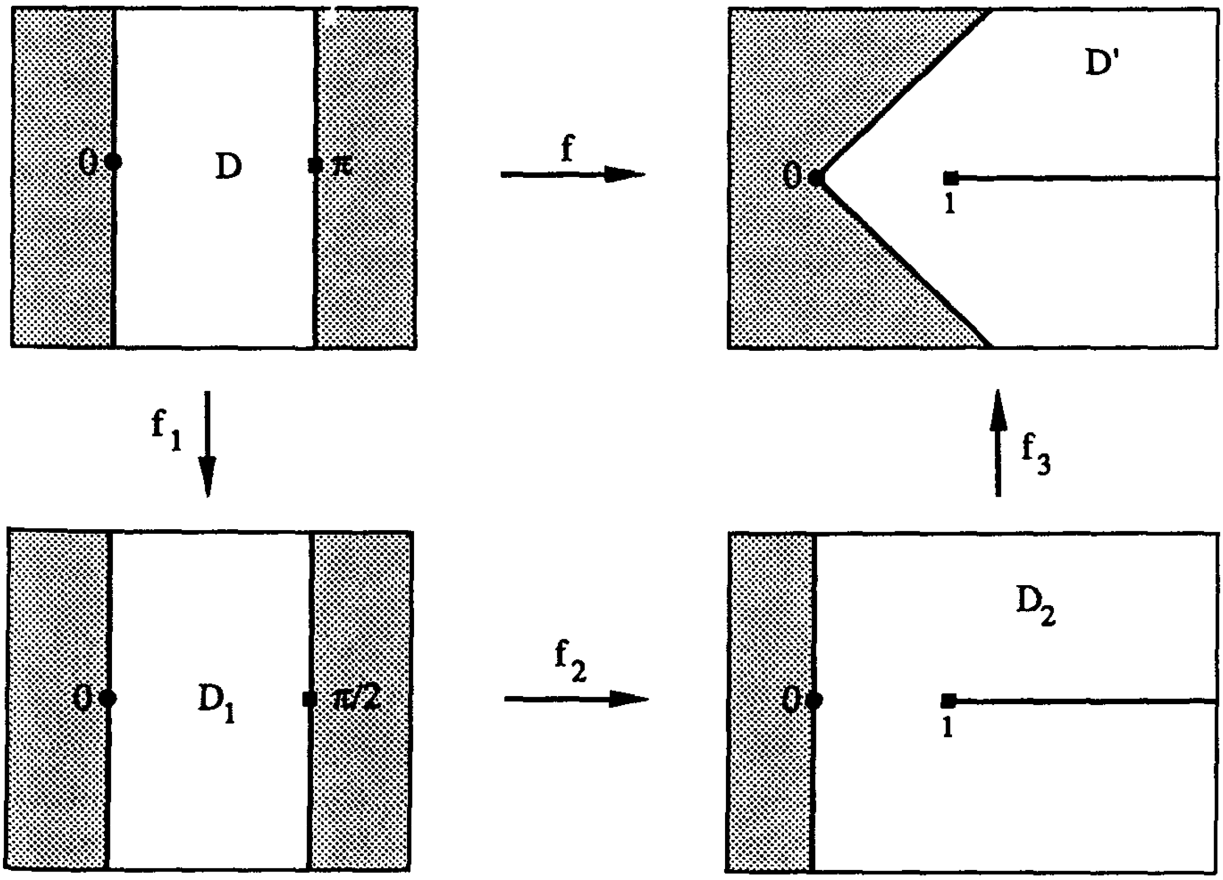


Figure 12.

1.5 Self-Mappings of the Plane and Unit Disk

Owing to Theorem 1.2 we can use the machinery of complex analysis to shed light on the structure of conformal mappings. The following two theorems illustrate this point. In the first we characterize the conformal mappings of the whole complex plane.

Theorem 1.3. *The conformal mappings $f: \mathbb{C} \to \mathbb{C}$ are the functions of the form $f(z) = az + b$, where a and b are complex numbers and $a \neq 0$. In particular, the range of any conformal mapping $f: \mathbb{C} \to \mathbb{C}$ is the entire complex plane.*

Proof. We already know that all functions of the stated description provide conformal mappings of the complex plane onto itself. We must show that there are no others. Theorem 1.2 tells us that a conformal mapping f of $\mathbb{C}$ is an entire function. As such, f admits a Taylor series expansion $f(z) = \sum_{n=0}^{\infty} a_n z^n$ valid throughout the complex plane. Since f is non-constant, $a_n \neq 0$ must be true for some $n > 0$, so the isolated singularity of f at ∞ is either an essential singularity or a pole. Could it be essential? Were this the case, the Casorati-Weierstrass theorem in concert with the open mapping theorem would imply that $f(U) \cap f(V) \neq \phi$, where $U = \Delta^*(\infty, 1)$ and $V = \Delta(0,1)$. Indeed, the set $f(U)$ would be dense in $\mathbb{C}$, a condition that forces it to intersect the open set $f(V)$. Any point w in $f(U) \cap f(V)$ would clearly possess at least two preimages in $\mathbb{C}$, one in U and one in V. The resulting conflict with the univalence of f rules out an essential singularity at ∞. Therefore, f has a pole of some order m at ∞; i.e., the Taylor series expansion of f reduces to $f(z) = a_0 + a_1 z + \cdots + a_m z^m$, where $m \geq 1$ and where $a_m \neq 0$. If $m > 1$, then the fundamental theorem of algebra would imply that f', a polynomial of positive degree, would have a root. In light of Theorem VIII.3.9, the assumption that f is univalent would again be violated. As a consequence, $m = 1$ and f takes the form $f(z) = az + b$ with $a = a_1 \neq 0$ and $b = a_0$. ■

The next result identifies the conformal self-mappings of the unit disk.

Theorem 1.4. *The functions that map the unit disk $D = \Delta(0,1)$ conformally onto itself are the functions $f: D \to \mathbb{C}$ of the form*

$$f(z) = e^{i\theta} \frac{z + c}{1 + \bar{c}z}, \qquad (9.12)$$

where θ is a real number and c is a complex number with $|c| < 1$.

Proof. We first check that each function of type (9.12) does represent a conformal self-mapping of D. Clearly $f = g \circ h$ in D, where $g(z) = e^{i\theta} z$ and $h(z) = (z + c)/(1 + \bar{c}z)$. The function g causes no problem: its effect is simply to rotate D about the origin. It plainly maps D conformally onto itself. The behavior of h is less transparent. Certainly the rational function h is analytic in D, for its denominator vanishes only at the point $-1/\bar{c}$, which lies outside D. A short computation — see Exercise I.4.21 — reveals that $|h(z)| < 1$ when $|z| < 1$. In other words, $h(D)$ is a subset of D. Next, we consider the function k defined by $k(z) = (z - c)/(1 - \bar{c}z)$. As k has the same structure that h does, with $-c$ replacing c, it is also true that $k(D)$ is contained in D. Furthermore, easy calculations show that $k[h(z)] = z$ and $h[k(z)] = z$ for every z in D. (The function k is just the inverse of h.) This implies in the first place that h is univalent in D and secondly that

$$D = [h \circ k](D) = h[k(D)] \subset h(D) \subset D ,$$

i.e., $h(D) = D$. Consequently, both g and h map D in a conformal manner onto itself. The composition $f = g \circ h$ does likewise.

It remains to be checked that an arbitrary conformal mapping f of D onto itself is of the kind described in (9.12). For this, we set $c = -f^{-1}(0)$ and look at the function $g: D \to D$ given by $g = f \circ k$, where as earlier $k: D \to D$ is the conformal mapping defined by $k(z) = (z - c)/(1 - \bar{c}z)$. Then g, too, is a conformal self-mapping of D, one that has

$$g(0) = f[k(0)] = f(-c) = f[f^{-1}(0)] = 0 .$$

Schwarz's lemma (Theorem V.3.14) informs us that $|g'(0)| \leq 1$. The function g^{-1} is another conformal self-mapping of D that fixes the origin, so a second appeal to the Schwarz lemma delivers the inequality $1/|g'(0)| = |[g^{-1}]'(0)| \leq 1$. As a result, equality holds — $|g'(0)| = 1$. According to Schwarz's lemma this only happens when g has the structure $g(z) = e^{i\theta}z$ for some real number θ. Finally, recalling from the first part of this proof that the inverse of k is the function $h: D \to D$ with rule of correspondence $h(z) = (z + c)/(1 + \bar{c}z)$, we conclude that $f = g \circ k^{-1} = g \circ h$ takes the form in (9.12). ∎

1.6 Conformal Mappings in the Extended Plane

For the sake of future discussions it will be convenient to have available the notion of a conformal mapping between domains in the extended complex plane. The most efficient way to introduce this concept (and the way best suited to our eventual needs) is to take a cue from Theorem 1.2 and to formulate the definition as follows: a function $f: D \to \widehat{\mathbb{C}}$, where D is a domain in $\widehat{\mathbb{C}}$, is a *conformal mapping of D* provided f is a univalent meromorphic function. (By Theorem 1.2 this reduces to the earlier definition of a conformal mapping if D and the range of f are subsets of $\mathbb{C}$. It follows from Theorem VIII.4.7, incidentally, that in the extended-plane setting the range of f is a domain in $\widehat{\mathbb{C}}$. The definition of a meromorphic function insures, of course, that f is continuous in the sense appropriate to mappings between sets in $\widehat{\mathbb{C}}$.) Since the requirement of univalence is imposed on f, there can be at most one point z_0 in D — there may be none — for which $f(z_0) = \infty$; i.e., f has at most one pole in D. If f does have a pole in D, it is necessarily a simple pole, for a meromorphic function is not univalent in the vicinity of a pole of order higher than one (Theorem VIII.4.10). The restriction of f to the finite domain $D_0 = \{z \in D: z \neq \infty, f(z) \neq \infty\}$ is both analytic and univalent — hence, is a conformal mapping of D_0 according to our original definition of the term. Theorem VIII.4.1 implies the following converse of this remark: if D is a domain in $\widehat{\mathbb{C}}$ and if $f: D \to \widehat{\mathbb{C}}$ is a univalent continuous function whose restriction to $D_0 = \{z \in D: z \neq \infty, f(z) \neq \infty\}$ is a conformal mapping of D_0, then f is a conformal mapping of D. As was the case in the finite plane, the composition of conformal mappings between domains in the extended plane is conformal, and the inverse of a conformal mapping is also conformal. By an *anti-conformal mapping* f of an extended-plane domain D we mean one whose conjugate $\overline{f}$ is a conformal mapping of D. Here we adopt the convention that $\overline{\infty} = \infty$.

A simple example of a conformal mapping $f: \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is the function $f(z) = z^{-1}$. (Remember: by our convention on extended domain-sets, $f(0) = \infty$ and $f(\infty) = 0$.) Being a rational function of z, f is meromorphic, and it is obviously univalent. This function also provides a useful conformal mapping of the disk $D = \Delta(0,1)$ onto its exterior $D' = \Delta(\infty, 1)$ and, since $f = f^{-1}$, of D' onto D.

The function $f(z) = z^{-1}$ is but one member of an extremely important class of conformal mappings, the class of *Möbius transformations*, named in honor of the geometer A.F. Möbius (1790-1868). These are the rational functions $f: \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ that have the form

$$f(z) = \frac{az + b}{cz + d}, \tag{9.13}$$

where a, b, c, and d are complex numbers with the property that $ad - bc \neq 0$. (Such functions are also known as *linear fractional transformations.*) As rational functions, Möbius transformations are meromorphic. The condition $ad - bc \neq 0$ makes certain that they are not constant functions in disguise. If $c = 0$, then $a \neq 0$ and (9.13) simplifies to an expression of the sort $f(z) = \alpha z + \beta$, with $\alpha = a/d \neq 0$; stated differently, when $c = 0$ the mapping f is a sense-preserving similarity transformation. In line with our extended domain-set convention we interpret (9.13) as saying that $f(\infty) = \infty$ in this case. If $c \neq 0$, then the Möbius transformation f has a pole at $-d/c$, so (9.13) is understood to include the assignment of value $f(-d/c) = \infty$. In this situation it is also implicit in (9.13) that $f(\infty) = \lim_{z \to \infty} f(z) = a/c$. To confirm that (9.13) does actually define a conformal self-mapping of $\widehat{\mathbb{C}}$ we still need to demonstrate that f is univalent. This is readily accomplished by writing down its inverse function:

$$f^{-1}(z) = \frac{dz - b}{-cz + a},$$

as a straightforward calculation verifies (Exercises I.4.36 and I.4.48). Notice that f^{-1} is itself a Möbius transformation. We conclude, in particular, that $f(\widehat{\mathbb{C}}) = \widehat{\mathbb{C}}$. One checks without difficulty (Exercise I.4.35) that the composition of Möbius transformations is another function from that class.

Theorems 1.3 and 1.4 reveal that the conformal self-mappings of $\mathbb{C}$ and those of the disk $\Delta(0,1)$ are merely Möbius transformations of special types. The following proposition is a companion to those two results.

Theorem 1.5. *The conformal mappings $f: \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ are the Möbius transformations. In particular, the range of any conformal mapping $f: \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is the entire extended plane.*

Proof. We have already observed that Möbius transformations map the extended plane conformally onto itself. Suppose that $f: \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is an arbitrary conformal mapping of $\widehat{\mathbb{C}}$. We define a function $h: \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ as follows: if $f(\infty) = \infty$, then we set $h = f$; if $f(\infty) = w_0 \neq \infty$, then we take $h = g \circ f$, where $g(z) = 1/(z - w_0)$. Since g is a Möbius transformation mapping w_0 to ∞, h is in either case a conformal mapping of $\widehat{\mathbb{C}}$ that leaves the point ∞ fixed. The restriction of h to the finite plane $\mathbb{C}$ is then a conformal mapping of $\mathbb{C}$, which means in view of Theorem 1.3 that h has the form $h(z) = az + b$ with $a \neq 0$. We conclude that h is a Möbius transformation. Because either $f = h$ or $f = g^{-1} \circ h$, f is a Möbius transformation as well. ∎

2 Möbius Transformations

2.1 Elementary Möbius Transformations

In this section we take a closer look at the class of Möbius transformation, of which the previous section afforded us a brief glimpse. We shall discover that these mappings have quite a number of interesting and useful properties. To launch the discussion we observe that every Möbius transformation can be built up as the composition of extremely simple ones, mappings we refer to as *elementary Möbius transformations*. These fall into four categories, as follows: (i) a *translation* is a Möbius transformation of the form $f(z) = z + b$; (ii) a mapping of the type $f(z) = az$, where a is both real and positive, is called a *dilation* (or *homothety*) *with respect to the origin*; (iii) if $f(z) = az$ with $|a| = 1$, f is a *rotation about the origin*; (iv) the final elementary transformation is the *inversion* $f(z) = z^{-1}$. (N.B. The *identity transformation*, the Möbius transformation f defined by $f(z) = z$ for every z in $\widehat{\mathbb{C}}$, fits each of the descriptions (i), (ii), and (iii). To avoid such an overlap of categories some authors prefer to place the identity in a separate class of its own.) Suppose now that $f(z) = (az + b)/(cz + d)$ is an arbitrary Möbius transformation. If $c = 0$, then f is a similarity transformation, in which event f is the composition of elementary transformations of types (i), (ii), and (iii) (cf., Example I.3.1); if $c \neq 0$, then we can write

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

in which case it is again clear that f can be obtained as the composition of (five or fewer) mappings of elementary type. The value of the foregoing remarks is this: to establish that a property is shared by all Möbius transformations it suffices, first, to verify that every elementary transformation enjoys the property in question and, secondly, to check that the property is preserved under composition. Often the verification of a property for elementary transformations involves little or no work.

2.2 Möbius Transformations and Matrices

With a Möbius transformation $f(z) = (az + b)/(cz + d)$ we can associate the non-singular 2×2 complex matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Conversely, each matrix A of the stated character furnishes the coefficients of a Möbius transformation, one we symbolize by f_A. The correspondence $A \to f_A$ has the nice feature that $AB \to f_A \circ f_B$ and $A^{-1} \to (f_A)^{-1}$, where AB indicates ordinary matrix multiplication and A^{-1} signifies the inverse matrix to A (Exercise 6.24). Thanks to this observation, many computations involving Möbius transformations can be reduced to simple matrix calculations, at great savings of time and avoidance of ponderous functional notation. As an illustration of how matrices can expedite such calculations, let us compute $f = g \circ h \circ k$, where $g(z) = (z + 1)/(z - 1)$, $h(z) = z/(z + i)$, and $k(z) = (z - i)/z$. By the above comments we can express the composition as $f = f_A$ for

$$A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & i \end{bmatrix} \begin{bmatrix} 1 & -i \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2+i & -2i \\ -i & 0 \end{bmatrix},$$

which yields

$$f(z) = \frac{(2+i)z - 2i}{(-i)z} = \frac{(-1+2i)z + 2}{z}.$$

Though hardly overwhelming, the computation of f without resorting to matrices would be considerably less efficient.

The correspondence which sends a non-singular matrix A to the Möbius transformation f_A is definitely not one-to-one. Indeed, the fact that for $\lambda \neq 0$ the formulas

$$f(z) = \frac{az + b}{cz + d} \tag{9.14}$$

and

$$f(z) = \frac{(\lambda a)z + \lambda b}{(\lambda c)z + \lambda d} \tag{9.15}$$

are alternate descriptions of one and the same function makes it clear that $f_A = f_{\lambda A}$ for every non-zero complex number λ. Presented with a Möbius transformation f in the form (9.14) we are free to choose λ for which $\lambda^2(ad - bc) = 1$ — namely, $\lambda = \pm(ad - bc)^{-1/2}$ — and to rewrite f as in (9.15). From this it becomes apparent that a Möbius transformation f always admits a representation of the type

$$f(z) = \frac{az + b}{cz + d} \quad , \qquad ad - bc = 1 \ . \tag{9.16}$$

When (9.16) holds we say that f is expressed in *normalized form*. The matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, which then has determinant one, is called a *normalized matrix* corresponding to f. For instance, the functions $f(z) = 4z + 1$ and $g(z) = (z - 1)/(z + i)$ can be put into normalized form as follows:

$$f(z) = \frac{2z + (1/2)}{0z + (1/2)} \quad , \quad g(z) = \frac{\left(2^{-1/4}e^{-\pi i/8}\right) z - 2^{-1/4}e^{-\pi i/8}}{\left(2^{-1/4}e^{-\pi i/8}\right) z + 2^{-1/4}e^{-\pi i/8}i} \ .$$

Normalized matrices corresponding to the elementary transformations are:

(i) $\begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix}$ for a translation $f(z) = z + b$;

(ii) $\begin{bmatrix} a^{1/2} & 0 \\ 0 & a^{-1/2} \end{bmatrix}$ for a transformation $f(z) = az$, which includes dilations ($a > 0$) and rotations ($|a| = 1$);

(iii) $\begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}$ for the inversion $f(z) = z^{-1}$.

Although at first glance it might seem that writing a Möbius transformation in normalized form only serves to complicate matters, for theoretical purposes it is really the opposite that is true: the use of normalized forms tends, in general, to simplify the treatment of these mappings. A perfect example of this is the discussion of fixed points found in Theorem 2.1. Even when restricted to matrices A with determinant one, the correspondence $A \to f_A$ fails to be one-to-one, for obviously $f_A = f_{-A}$. However, this is the only way in which it deviates from one-to-oneness: if $f_A = f_B$ and if $\det A = \det B = 1$, then either $A = B$ or $A = -B$ (Exercise 6.25).

2.3 Fixed Points

A giant step toward understanding the inner workings of a Möbius transformation f is taken with the realization that its set of fixed points, by which is meant the set of points z in $\widehat{\mathbb{C}}$ with the property that $f(z) = z$, has an extremely simple structure.

Theorem 2.1. *Let f be a Möbius transformation other than the identity. Then f has at most two fixed points. More precisely, if f is exhibited in normalized form as $f(z) = (az+b)/(cz+d)$, then f has exactly one fixed point when $a + d = \pm 2$ and exactly two fixed points otherwise.*

Proof. We may suppose that f is given to us in normalized form: $f(z) = (az+b)/(cz+d)$, with $ad - bc = 1$. Assume first that $c = 0$. Then $ad = 1$ and f is the similarity transformation $f(z) = \alpha z + \beta$, where $\alpha = a/d$ and $\beta = b/d$. Such a mapping always fixes ∞. In the event that $\alpha \neq 1$ it also fixes the point $-\beta/(\alpha - 1)$. Unless $\alpha = 1$ and $\beta = 0$ — this would make f the identity transformation — the equation $\alpha z + \beta = z$ has no other solutions, so f has no other fixed points. Notice that $\alpha = 1$ if and only if $a = d$. Since $(a+d)^2 = (a-d)^2 + 4ad = (a-d)^2 + 4$, $\alpha = 1$ holds if and only if $a + d = \pm 2$. In the case $c = 0$, therefore, f has exactly one fixed point when $a + d = \pm 2$ and exactly two such points otherwise.

We now consider the situation for non-zero c. In this case $f(-d/c) = \infty$ and $f(\infty) = a/c$, so neither $-d/c$ nor ∞ is fixed by f. Its only fixed points are the finite solutions z of the equation

$$\frac{az+b}{cz+d} = z ,$$

which, since we're concerned only with points where the denominator on the left is not zero, is equivalent to

$$cz^2 + (d-a)z - b = 0 .$$

Using the fact that $ad - bc = 1$, we find the roots of this quadratic equation to be

$$z = \frac{(a-d) \pm \sqrt{(a+d)^2 - 4}}{2c} .$$

Again when $c \neq 0$, the mapping f is seen to have a unique fixed point if $a + d = \pm 2$ and two fixed points if $a + d \neq \pm 2$. ∎

Möbius transformations f and g are said to be *conjugate* provided there exists a Möbius transformation h with the property that $g = h^{-1} \circ f \circ h$ or, equivalently, $h \circ g = f \circ h$. (N.B. The context will always make clear whether the term "conjugate" means the conjugate of a complex number, the conjugate of a harmonic function, or conjugate in the sense defined here for Möbius transformations.) Since $h[g(z)] = f[h(z)]$ is then true for every z in $\widehat{\mathbb{C}}$, it becomes clear that z is a fixed point of g precisely when $h(z)$ is fixed by f. In particular, conjugate Möbius transformations have the same number of fixed points. Many of the ideas and quantities naturally associated with Möbius transformations turn out to be invariant under conjugacy. Number of fixed points is the simplest example. We shall point out others as we proceed.

How much leeway exists for dictating the values of a Möbius transformation at specified points? The next two results provide the answer to this question.

Theorem 2.2. *Let (z_1, z_2, z_3) be an ordered triple of distinct points from $\widehat{\mathbb{C}}$. There exists a unique Möbius transformation f that has $f(z_1) = 1$, $f(z_2) = 0$, and $f(z_3) = \infty$.*

Proof. For the existence, assume initially that all three of the given points are finite. Then

$$f(z) = \frac{(z_1 - z_3)(z - z_2)}{(z_1 - z_2)(z - z_3)} \tag{9.17}$$

is a Möbius transformation with the features sought. (We have not bothered to express f in the form $f(z) = (az + b)/(cz + d)$, but it is obvious that we could do so if pressed.) If one of the points z_1, z_2, or z_3 is ∞, we can obtain f by letting the appropriate term in (9.17) tend to ∞. To be explicit, this process yields

$$\begin{cases} f(z) = \dfrac{z - z_2}{z - z_3} & \text{if } z_1 = \infty \ ; \\[2ex] f(z) = \dfrac{z_1 - z_3}{z - z_3} & \text{if } z_2 = \infty \ ; \\[2ex] f(z) = \dfrac{z - z_2}{z_1 - z_2} & \text{if } z_3 = \infty \ . \end{cases} \tag{9.18}$$

As for the uniqueness, suppose that both f and g are Möbius transformations which enjoy the stated property. Then $h = f^{-1} \circ g$ is a Möbius transformation that fixes z_1, z_2, and z_3. Theorem 2.1 implies that h is the identity transformation, which means that $g = (f^{-1})^{-1} = f$. This corroborates the uniqueness assertion. ∎

An immediate consequence of Theorem 2.2 is that, modulo the obvious restriction imposed by univalence, the values of a Möbius transformation can be prescribed arbitrarily at any three given points of $\widehat{\mathbb{C}}$.

Corollary 2.3. *Let (z_1, z_2, z_3) and (w_1, w_2, w_3) be ordered triples of distinct points from $\widehat{\mathbb{C}}$. There exists a unique Möbius transformation f that has $f(z_1) = w_1$, $f(z_2) = w_2$, and $f(z_3) = w_3$.*

Proof. We invoke Theorem 2.2 to choose Möbius transformations g and h for which $g(z_1) = 1$, $g(z_2) = 0$, and $g(z_3) = \infty$, while $h(w_1) = 1$, $h(w_2) = 0$, and $h(w_3) = \infty$. Then $f = h^{-1} \circ g$ is a Möbius transformation that maps z_1 to w_1, z_2 to w_2, and z_3 to w_3. Its uniqueness in this regard is confirmed by essentially the same argument used in the proof of Theorem 2.2 to demonstrate uniqueness there. ∎

2.4 Cross-ratios

In studying any class of mappings it is generally of value to identify quantities and structures that are preserved under application of those mappings. Later in this chapter we shall talk about a general "conformal invariant." For the time being we concentrate on three invariants — one numerical, two geometric — specific to the class of Möbius transformations. The first of these is the *cross-ratio* $[z_1, z_2, z_3, z_4]$ associated with an ordered quadruple (z_1, z_2, z_3, z_4) of distinct points from the extended complex plane. By definition, $[z_1, z_2, z_3, z_4]$ is the complex number $f(z_4)$, where f is the unique Möbius transformation that satisfies $f(z_1) = 1$, $f(z_2) = 0$, and $f(z_3) = \infty$. Referring to (9.17) and (9.18), we obtain the formula

$$[z_1, z_2, z_3, z_4] = \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_2)(z_3 - z_4)} \tag{9.19}$$

in case all four of these points are finite, whereas the situation in which one of the points is ∞ is covered by

$$\begin{cases} [\infty, z_2, z_3, z_4] = \dfrac{z_2 - z_4}{z_3 - z_4} \quad , \quad [z_1, \infty, z_3, z_4] = \dfrac{z_1 - z_3}{z_4 - z_3} \ , \\ [z_1, z_2, \infty, z_4] = \dfrac{z_2 - z_4}{z_2 - z_1} \quad , \quad [z_1, z_2, z_3, \infty] = \dfrac{z_1 - z_3}{z_1 - z_2} \ . \end{cases} \tag{9.20}$$

We stress that a cross-ratio is always a finite quantity different from 0 or 1. For example, we compute

$$[1, i, -1, -i] = \frac{2 \cdot (2i)}{(1 - i)(-1 + i)} = 2$$

and

$$[0, -1, i, \infty] = \frac{-i}{1} = -i \ .$$

The behavior of cross-ratios under Möbius transformations is the subject of the next theorem.

Theorem 2.4. *If f is a Möbius transformation, then $[z_1, z_2, z_3, z_4] = [f(z_1), f(z_2), f(z_3), f(z_4)]$ for every ordered quadruple (z_1, z_2, z_3, z_4) of distinct points from $\widehat{\mathbb{C}}$.*

Proof. Let g be the unique Möbius transformation that sends $f(z_1)$ to 1, $f(z_2)$ to 0, and $f(z_3)$ to ∞. Then $g \circ f$ transforms z_1 to 1, z_2 to 0, and z_3 to ∞. According to the definition of a cross-ratio

$$[z_1, z_2, z_3, z_4] = g \circ f(z_4) = g[f(z_4)] = [f(z_1), f(z_2), f(z_3), f(z_4)] \ ,$$

as asserted. ∎

A slight refinement of Theorem 2.4 tells us exactly when two cross-ratios are equal.

Corollary 2.5. *Cross-ratios* $[z_1, z_2, z_3, z_4]$ *and* $[w_1, w_2, w_3, w_4]$ *are equal if and only if there exists a Möbius transformation* f *satisfying* $f(z_j) = w_j$ *for* $j = 1, 2, 3, 4$.

Proof. If f exists as described, then Theorem 2.4 certifies that the two cross-ratios agree. Conversely, assume that these cross-ratios are equal. By definition $[w_1, w_2, w_3, w_4] = g(w_4)$, where g is the unique Möbius transformation that has $g(w_1) = 1$, $g(w_2) = 0$, and $g(w_3) = \infty$. Now let f be the unique Möbius transformation for which $f(z_1) = w_1$, $f(z_2) = w_2$, and $f(z_3) = w_3$ (Corollary 2.3). On the basis of Theorem 2.4 we conclude that

$$g[f(z_4)] = [w_1, w_2, w_3, f(z_4)] = [z_1, z_2, z_3, z_4] = [w_1, w_2, w_3, w_4] = g(w_4).$$

Since g is univalent, this implies that $f(z_4) = w_4$, so $f(z_j) = w_j$ for $j = 1, 2, 3, 4$. ∎

Theorem 2.4 introduces a convenient way to represent the Möbius transformation f that maps three given points z_1, z_2, and z_3 to designated image points w_1, w_2, and w_3; namely, for $z \neq z_1, z_2, z_3$ the value $f(z)$ is completely determined by the relation

$$(9.21) \qquad [z_1, z_2, z_3, z] = [w_1, w_2, w_3, f(z)]\ .$$

(One implication of this formula is that Möbius transformations are the only self-mappings of $\widehat{\mathbb{C}}$ which preserve cross-ratios.) For many purposes it is far preferable to work with (9.21) than with some cumbersome formula for f. Even if a formula for f is the objective, (9.21) often provides the most efficient route to it, as the following example illustrates.

EXAMPLE 2.1. Determine the Möbius transformation f that maps i to ∞, 0 to 1, and ∞ to $-i$.

For $z \neq i, 0, \infty$ we appeal to (9.21) and write

$$[i, 0, \infty, z] = [\infty, 1, -i, f(z)]\ ,$$

which reduces with the help of (9.20) to

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

Solving for $f(z)$ we arrive at the formula

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

2.5 Circles in the Extended Plane

From a purely geometric standpoint the most important class of objects preserved by Möbius transformations is the class of circles in the extended complex plane. We remind the reader that the expression "circle in $\widehat{\mathbb{C}}$" refers to a subset K of $\widehat{\mathbb{C}}$ that is either an honest-to-god circle in the complex plane or a "circle through ∞," meaning a set of the type $K = L \cup \{\infty\}$, where L is a straight line in $\mathbb{C}$. Thus, through any triad of points in $\widehat{\mathbb{C}}$ there passes one and only one circle in $\widehat{\mathbb{C}}$. It is easily checked that the locus of points z in $\mathbb{C}$ which satisfy an equation

$$(9.22) \quad A|z|^2 + Bz + \overline{B}\overline{z} + C = 0 \qquad (A \text{ and } C \text{ real}, |B|^2 - AC > 0)$$

is either a circle ($A \neq 0$) or a line ($A = 0$). Conversely, every circle or line in the complex plane can be described by an equation of this type. (Recall Exercise I.4.19.) We can therefore regard (9.22) as the general equation of a circle in $\widehat{\mathbb{C}}$, provided we make the convention that ∞ gets included in the locus of points satisfying (9.22) if and only if $A = 0$. After these preparatory remarks, we are set to prove:

Theorem 2.6. *Möbius transformations map circles in $\widehat{\mathbb{C}}$ to circles in $\widehat{\mathbb{C}}$.*

Proof. Since an arbitrary composition of mappings with the property of transforming circles in $\widehat{\mathbb{C}}$ to circles in $\widehat{\mathbb{C}}$ will again exhibit this property, it is enough to check that the image of K, an arbitrary circle in $\widehat{\mathbb{C}}$, under every elementary transformation is a circle in $\widehat{\mathbb{C}}$. If f is a translation or dilation or rotation, then it is perfectly obvious that $f(K)$ is another circle in $\widehat{\mathbb{C}}$. It is only for the inversion $f(z) = z^{-1}$ that the situation remains unclear and needs to be discussed. Choose an equation $A|z|^2 + Bz + \overline{B}\overline{z} + C = 0$ for K. Let complex numbers z and w be related by $w = z^{-1}$. (In particular, $z \neq 0$ and $w \neq 0$.) Then z lies on K if and only if — substitute w^{-1} for z in the equation of K — w obeys the condition

$$\frac{A}{|w|^2} + \frac{B}{w} + \frac{\overline{B}}{\overline{w}} + C = 0$$

or, equivalently, w is on the locus determined by

$$C|w|^2 + \overline{B}w + B\overline{w} + A = 0 \,.$$

The last line is the equation of another circle in $\widehat{\mathbb{C}}$, call it $\widetilde{K}$. The point $z = 0$ belongs to K precisely when $C = 0$, which occurs by the convention adopted above exactly when $w = \infty = 1/0$ belongs to $\widetilde{K}$. Similarly, $z = \infty$ is on K if and only if $w = 0 = 1/\infty$ is on $\widetilde{K}$. In summary, a point z of $\widehat{\mathbb{C}}$ is a point of K if and only if $w = z^{-1} = f(z)$ is a point of $\widetilde{K}$; i.e., $f(K) = \widetilde{K}$, a circle in $\widehat{\mathbb{C}}$. ∎

EXAMPLE 2.2. Identify the images of the circles $K = \mathbb{R} \cup \{\infty\}$ and $K' = K(0,1)$ under the Möbius transformation $f(z) = (z+i)/(z+1)$.

In view of Theorem 2.6 the set $f(K)$ is the circle in $\widehat{\mathbb{C}}$ that passes through the points $f(1) = (1+i)/2$, $f(0) = i$, and $f(-1) = \infty$ (Figure 13). Accordingly, $f(K)$ can be described by the (real) equation $y = -x+1$. The image of K' is likewise a circle in $\widehat{\mathbb{C}}$. It passes through $f(1) = (1+i)/2$ and $f(-1) = \infty$. By conformality $f(K')$ intersects $f(K)$ orthogonally at $(1+i)/2$, for K' is perpendicular to K at the point 1. This is enough information to identify $f(K')$ as the circle in $\widehat{\mathbb{C}}$ with equation $y = x$.

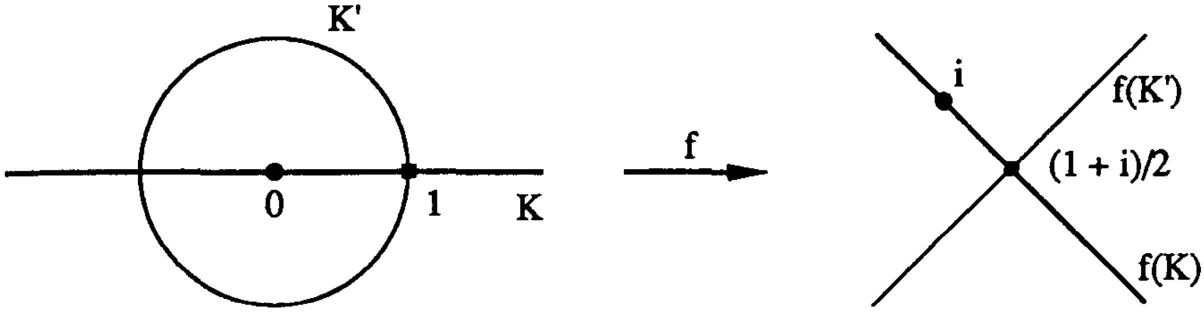


Figure 13.

2.6 Reflection and Symmetry

A third noteworthy invariant for Möbius transformations is the property of symmetry with respect to a circle in $\widehat{\mathbb{C}}$. In order to explain this notion we associate with each circle K in $\widehat{\mathbb{C}}$ a mapping $\rho_K : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ called the *reflection in* K. This we do as follows. Suppose first that K is the "true" circle in $\mathbb{C}$ with center z_0 and radius r. In this case we define ρ_K by

$$\rho_K(z) = z_0 + \frac{r^2}{\bar{z} - \bar{z}_0} = \frac{z_0\bar{z} + r^2 - |z_0|^2}{\bar{z} - \bar{z}_0} \tag{9.23}$$

for $z \neq z_0, \infty$, while $\rho_K(z_0) = \infty$ and $\rho_K(\infty) = z_0$. Put in geometric terms, ρ_K transforms a point z different from z_0 and ∞ to the unique point z^* that lies on the ray through z issuing from z_0 and that obeys the condition $|z - z_0||z^* - z_0| = r^2$ (Exercise 6.35). Figure 14 suggests how one can geometrically construct $\rho_K(z)$ from z when K is a genuine circle. Next, if $K = L \cup \{\infty\}$ for a line L with equation $Bz + \overline{B}\bar{z} + C = 0$ (here C is real and $B \neq 0$) then ρ_K is given by

$$\rho_K(z) = (-\overline{B}/B)\bar{z} - (C/B) \tag{9.24}$$

for $z \neq \infty$ and $\rho_K(\infty) = \infty$. In this situation ρ_K acts on the plane in the way one ordinarily thinks of a reflection acting: it sends a point $z(\neq \infty)$ to its mirror image with respect to K; i.e., to the point z^* such that K

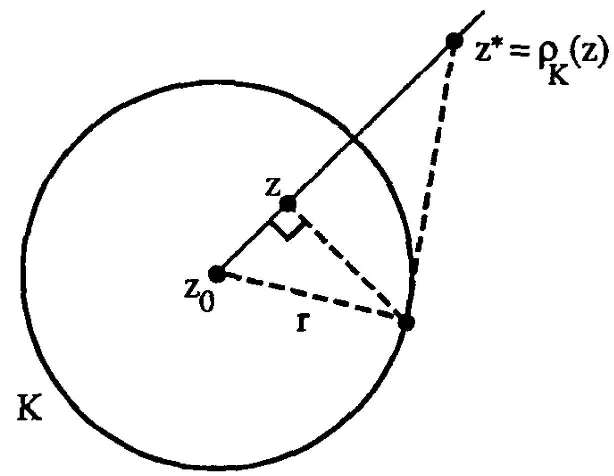


Figure 14.

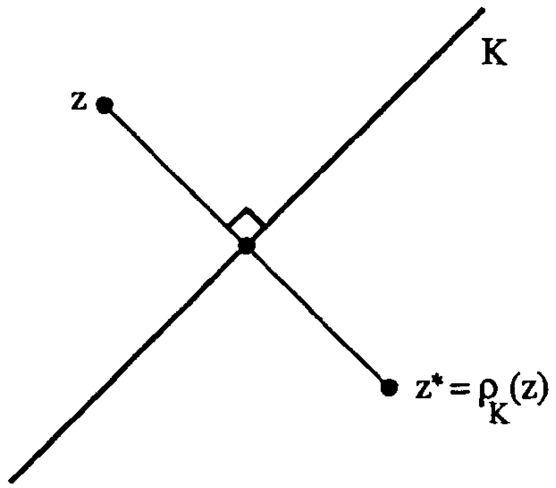


Figure 15.

is the perpendicular bisector of the line segment with endpoints z and z^*. (See Figure 15. See, too, Exercise 6.35.) In both (9.23) and (9.24) we observe that ρ_K fixes every point of K and that $\rho_K \circ \rho_K$ is the identity transformation of $\widehat{\mathbb{C}}$, from which it follows that $\rho_K^{-1} = \rho_K$.

It is evident from (9.23) and (9.24) that every reflection ρ_K belongs to the class of functions with the general structure

$$(9.25) \qquad f(z) = \frac{a\bar{z} + b}{c\bar{z} + d},$$

where a, b, c, and d are complex numbers for which $ad - bc \neq 0$. Stated differently, f is a function of the form $f = g \circ \rho$, where g is a Möbius transformation and ρ is the reflection in the real axis, $\rho(z) = \bar{z}$. (N.B. Many books retain the name Möbius transformation for a function of this kind, but in keeping with other usage in this book we prefer to describe such a function as an ***anti-Möbius transformation***. It should perhaps be emphasized that not every anti-Möbius transformation is a reflection. See Exercise 6.37.) Since ρ is the prototypical anti-conformal self-mapping of $\widehat{\mathbb{C}}$

and since Möbius transformations are the conformal mappings of $\widehat{\mathbb{C}}$, it follows that anti-Möbius transformations map $\widehat{\mathbb{C}}$ anti-conformally onto itself. Indeed, it follows from Theorem 1.5 and the definition of anti-conformality in $\widehat{\mathbb{C}}$ that an arbitrary anti-conformal mapping of $\widehat{\mathbb{C}}$ onto itself has the structure (9.25). As might be expected, functions f that fit the description in (9.25) have many properties in common with Möbius transformations. For instance, if K is a circle in $\widehat{\mathbb{C}}$, then $f(K)$ is again a circle in $\widehat{\mathbb{C}}$. It is readily verified that by composing a pair of anti-Möbius transformations one produces a Möbius transformation. More generally, any composition of Möbius transformations and anti-Möbius transformations that involves an even number of factors of the latter sort will be a Möbius transformation.

Suppose now that K is a circle in $\widehat{\mathbb{C}}$. Points z and z^* of $\widehat{\mathbb{C}} \sim K$ are said to be *symmetric with respect to* K if $z^* = \rho_K(z)$ — hence, also, $z = \rho_K(z^*)$. The next theorem establishes the symmetry principle for Möbius transformations that was hinted at earlier.

Theorem 2.7. *If f is a Möbius transformation and if K is a circle in $\widehat{\mathbb{C}}$, then*

$$f \circ \rho_K = \rho_{f(K)} \circ f \; . \tag{9.26}$$

In particular, if z and z^* are points that are symmetric with respect to K, then $f(z)$ and $f(z^*)$ are symmetric with respect to $f(K)$.

Proof. Statement (9.26) is equivalent to the assertion that the function $g = f^{-1} \circ \rho_{f(K)} \circ f \circ \rho_K$ is the identity transformation of $\widehat{\mathbb{C}}$. (Don't forget: $\rho_{f(K)}^{-1} = \rho_{f(K)}$.) As the composition of Möbius transformations and an even number of reflections, g is a Möbius transformation. Furthermore, it is evident that g fixes every point of K. Theorem 2.1 implies that g must fix every point of $\widehat{\mathbb{C}}$; i.e., g is the identity transformation, so (9.26) holds.

If z and z^* are symmetric with respect to K, then by (9.26) the points $w = f(z)$ and $w^* = f(z^*)$ satisfy

$$\rho_{f(K)}(w) = \rho_{f(K)}[f(z)] = f[\rho_K(z)] = f(z^*) = w^* \; ,$$

revealing that w and w^* are symmetric with respect to $f(K)$. ∎

The proof given for Theorem 2.7 would have worked just as well had f been a mapping of type (9.25) instead of a Möbius transformation. The conclusions of the theorem thus remain valid when f is an anti-Möbius transformation. Here is a simple application of the theorem.

EXAMPLE 2.3. Let K be the circle in $\widehat{\mathbb{C}}$ that passes through the three points z_1, z_2, and z_3. Show that the reflection ρ_K satisfies

$$\overline{[z_1, z_2, z_3, z]} = [z_1, z_2, z_3, \rho_K(z)] \tag{9.27}$$

for $z \neq z_1, z_2, z_3$. (Compare this with (9.21).)

Let f be the Möbius transformation that maps z_1 to 1, z_2 to 0, and z_3 to ∞. Then $f(K) = \mathbb{R} \cup \{\infty\}$. Moreover, for z other than z_1, z_2, or z_3 we infer from the very definition of a cross-ratio that $[z_1, z_2, z_3, z] = f(z)$ and $[z_1, z_2, z_3, \rho_K(z)] = f[\rho_K(z)]$. If z lies on K, then $\rho_K(z) = z$ and (9.27) reduces to the statement $\overline{f(z)} = f(z)$, which is true because $f(z)$ is real in this case. If z lies off K, then by Theorem 2.7 the points $f(z)$ and $f[\rho_K(z)]$ must be symmetric with respect to $\mathbb{R} \cup \{\infty\}$, which is another way to say that $\overline{f(z)} = f[\rho_K(z)]$. Again, relation (9.27) is seen to be in force.

2.7 Classification of Möbius Transformations

Theorem 2.1 is the starting point for a standard classification of the non-identity Möbius transformations into four types. To a large extent (but not entirely) the classification parallels the breakdown of the elementary transformations into their categories.

A Möbius transformation with exactly one fixed point is known as a *parabolic transformation*. Let f be such a mapping, and let z_0 be its fixed point. If $z_0 = \infty$ the proof of Theorem 2.1 reveals that f must be a translation:

$$f(z) = z + b \,, \tag{9.28}$$

where $b = f(0) \neq 0$. Next, under the assumption that $z_0 \neq \infty$, we conjugate f by the transformation $h(z) = z_0 + z^{-1} = (z_0 z + 1)/z$ — h is chosen because it is the simplest Möbius transformation sending ∞ to z_0 — to form $g = h^{-1} \circ f \circ h$. Since f fixes only $z_0 = h(\infty)$, g has ∞ for its sole fixed point, so it is also parabolic. By our previous comments g has the form $g(z) = z + b$ for some non-zero complex number b. (In fact, b can be computed directly from f —

$$b = \lim_{z \to z_0} \left[\frac{1}{f(z) - z_0} - \frac{1}{z - z_0} \right]$$

— as (9.29) makes plain.) Observing that $h^{-1}(z) = (z - z_0)^{-1}$ and exploiting the fact that $h^{-1}[f(z)] = g[h^{-1}(z)]$ for every z in $\widehat{\mathbb{C}}$, we find that

$$\frac{1}{f(z) - z_0} = b + \frac{1}{z - z_0} \,. \tag{9.29}$$

Equation (9.29) is quickly solved for $f(z)$ to yield

$$f(z) = z_0 + \frac{z - z_0}{1 + b(z - z_0)} \tag{9.30}$$

as the general form of a parabolic transformation that fixes the finite point z_0. (A reminder: it is required that $b \neq 0$ here.) We remark that a mapping f of type (9.30) takes ∞ to $z_0 + b^{-1}$ and $z_0 - b^{-1}$ to ∞, which means that

z_0, $f(\infty)$, and $f^{-1}(\infty)$ are collinear points in the finite plane. Let it be further noted that a Möbius transformation f is parabolic if and only if it is conjugate to a non-trivial translation g (in fact, one can take $g(z) = z+1$).

From (9.28) and (9.30) it is possible to read off immediately the *forward iterates* $f_2 = f \circ f$, $f_3 = f \circ f \circ f, \cdots$ and *backward iterates* $f_{-1} = f^{-1}$, $f_{-2} = f^{-1} \circ f^{-1}, f_{-3} = f^{-1} \circ f^{-1} \circ f^{-1}, \cdots$ of f: for any integer n it is true that

$$f_n(z) = z + nb \tag{9.31}$$

if $z_0 = \infty$, while

$$f_n(z) = z_0 + \frac{z - z_0}{1 + nb(z - z_0)} \tag{9.32}$$

if $z_0 \neq \infty$. (Implicit here is that $f_1 = f$. Also, f_0 is an alias for the identity transformation.) This assertion is essentially obvious in the former case and can be verified in the latter one by repeated application of (9.29). For instance, we have

$$\frac{1}{f_2(z) - z_0} = \frac{1}{f[f(z)] - z_0} = b + \frac{1}{f(z) - z_0} = 2b + \frac{1}{z - z_0},$$

and solving for $f_2(z)$ produces (9.32) with $n = 2$. Similarly, (9.29) gives

$$-b + \frac{1}{z - z_0} = -b + \frac{1}{f[f^{-1}(z)] - z_0} = \frac{1}{f^{-1}(z) - z_0} = \frac{1}{f_{-1}(z) - z_0},$$

which leads to the confirmation of (9.32) for $n = -1$. It should be pointed out that, except for $n = 0$, all the iterates f_n of f are likewise parabolic with the same fixed point z_0 and that, as $|n| \to \infty$, $f_n(z) \to z_0$ for every z in $\widehat{\mathbb{C}}$.

We now turn our attention to a Möbius transformation f that has exactly two fixed points. We label these points z_1 and z_2, establishing the notational convention for the present discussion and for similar situations later on that $z_1 \neq \infty$. If $z_2 = \infty$, then the proof of Theorem 2.1 shows that f has the appearance $f(z) = \kappa z + \beta$, in which $\kappa \neq 0, 1$. Since $f(z_1) = z_1$, it follows that $\beta = z_1 - \kappa z_1$. This allows us to rewrite f in the manner $f(z) = z_1 + \kappa(z - z_1)$. In particular, the relationship between z, $f(z)$, and z_1 is expressed by

$$f(z) - z_1 = \kappa(z - z_1) \tag{9.33}$$

for some complex number κ different from 0 and 1. If $z_2 \neq \infty$, we select a Möbius transformation h that sends 0 to z_1 and ∞ to z_2 — $h(z) = (z_2 z - z_1)/(z - 1)$ is our choice — and build the conjugate transformation $g = h^{-1} \circ f \circ h$. Then g has 0 and ∞ as its only fixed points, so $g(z) = \kappa z$ for some constant κ other than 0 or 1. Because $h^{-1}(z) = (z - z_1)/(z - z_2)$

and $h^{-1}[f(z)] = g[h^{-1}(z)]$ for every z in $\widehat{\mathbb{C}}$, we learn that f satisfies

$$\frac{f(z)-z_1}{f(z)-z_2} = \kappa\left(\frac{z-z_1}{z-z_2}\right). \tag{9.34}$$

The constant κ appearing in (9.33) and (9.34) is known as a *multiplier* for the transformation f; (9.33) and (9.34) represent f in so-called *multiplier-fixed point format.* In the case of (9.34) such a representation is not uniquely determined, for it clearly depends on which fixed point we elect to designate as z_1. Reversing the labels on the fixed points, however, has only the effect of replacing the multiplier κ in (9.34) by κ^{-1}. (It follows that in all cases the quantity $\kappa+\kappa^{-1}$ is independent of the specific designations of z_1 and z_2.) Notice that we can always use the formula $\kappa = f'(z_1)$ to determine a multiplier for f. This is clear from (9.33) when $z_2 = \infty$ and follows from (9.34) via the calculation

$$\begin{aligned}\kappa &= \lim_{z\to z_1}\left[\frac{f(z)-z_1}{z-z_1}\,\frac{z-z_2}{f(z)-z_2}\right]\\ &= f'(z_1)\,\frac{z_1-z_2}{f(z_1)-z_2} = f'(z_1)\,\frac{z_1-z_2}{z_1-z_2} = f'(z_1)\end{aligned}$$

when $z_2 \neq \infty$. In the latter instance a similar computation would produce an alternate description of the κ in (9.34) — namely, $\kappa = 1/f'(z_2)$.

It has already been remarked that in the case $z_2 = \infty$ the transformation f takes the form

$$f(z) = z_1 + \kappa(z-z_1)\ . \tag{9.35}$$

Under the assumption that both z_1 and z_2 are finite we can solve (9.34) for $f(z)$ and thereby extract from it the expression

$$f(z) = \frac{(z_1-\kappa z_2)z + (\kappa-1)z_1z_2}{(1-\kappa)z + \kappa z_1 - z_2} \tag{9.36}$$

for a general Möbius transformation that leaves fixed the two finite points z_1 and z_2. In practice, one should stress, it is frequently more convenient to work with (9.34) than with (9.36). It will prove useful for later considerations to register the fact that the function in (9.36) transforms ∞ to $(z_1-\kappa z_2)/(1-\kappa)$ and $(z_2-\kappa z_1)/(1-\kappa)$ to ∞.

There is no difficulty in passing from (9.33) and (9.34) to the corresponding representations for the iterate f_n of f, just as we did in the case of a parabolic transformation: for any integer n,

$$f_n(z) - z_1 = \kappa^n(z-z_1) \tag{9.37}$$

if $z_2 = \infty$, while

$$(9.38) \qquad \frac{f_n(z) - z_1}{f_n(z) - z_2} = \kappa^n \left(\frac{z - z_1}{z - z_2}\right)$$

if $z_2 \neq \infty$. For example, when $n = 2$ the verification of (9.38) reads

$$\frac{f_2(z) - z_1}{f_2(z) - z_2} = \frac{f[f(z)] - z_1}{f[f(z)] - z_2} = \kappa \left(\frac{f(z) - z_1}{f(z) - z_2}\right) = \kappa^2 \left(\frac{z - z_1}{z - z_2}\right) .$$

From (9.37) and (9.38) it is a short step to a closed form expression for f_n:

$$(9.39) \qquad f_n(z) = \begin{cases} z_1 + \kappa^n(z - z_1) & \text{if } z_2 = \infty ; \\ \dfrac{(z_1 - \kappa^n z_2)z + (\kappa^n - 1)z_1 z_2}{(1 - \kappa^n)z + \kappa^n z_1 - z_2} & \text{if } z_2 \neq \infty . \end{cases}$$

Several interesting phenomena are discernible in (9.39). First of all, in the situation where $|\kappa| \neq 1$ one of the fixed points of f has the property that $f_n(z)$ tends to this fixed point as $n \to \infty$ whenever z is an element of $\widehat{\mathbb{C}}$ different from the other fixed point of f. The fixed point of f that enjoys this property is named its *attracting fixed point*; the remaining fixed point is its *repelling fixed point*. Recalling that it is $\kappa = f'(z_1)$ which appears in (9.39), we can elicit from (9.39) the information that the fixed point z_1 is attracting when $|f'(z_1)| < 1$ and repelling when $|f'(z_1)| > 1$. A fixed point is plainly attracting for f if and only if it is repelling for f^{-1}. In the case of a Möbius transformation f for which $|\kappa| = 1$ no such distinction between its fixed points can be drawn. What does present itself as a possibility when $|\kappa| = 1$, however, is that $\kappa^n = 1$ for some integer n, in which event the iterate f_n of f reduces to the identity transformation. The smallest positive integer n with this property is then called *the order of* f. The multipliers κ for which this behavior is exhibited are those of the form $\kappa = e^{2\pi i\theta}$ with θ a rational number.

The classification of a Möbius transformation with exactly two fixed points hinges on the character of the multiplier κ in (9.33) or (9.34). (Recall: $\kappa \neq 1$.) If κ is real and positive, then f is termed a *hyperbolic transformation*; when $|\kappa| = 1$, we say that f is an *elliptic transformation*; in all remaining cases f is pronounced a *loxodromic transformation*. It is not difficult to verify — this was partially done in the derivation of (9.34) — that f is hyperbolic if and only if it is conjugate to a non-trivial dilation with respect to the origin, and elliptic if and only if it is conjugate to a non-trivial rotation about the origin (Exercise 6.47). A loxodromic transformation, on the other hand, can never be conjugated to one of the elementary transformations. When a Möbius transformation is presented in normalized form, its type can be detected instantly.

Theorem 2.8. *Let $f(z) = (az + b)/(cz + d)$ be a Möbius transformation that is not the identity and is displayed in normalized form. Then f is* *parabolic if and only if* $a+d=\pm 2$; f *is hyperbolic if and only if* $a+d$ *is real with* $|a+d|>2$; f *is elliptic if and only if* $a+d$ *is real with* $|a+d|<2$; f *is loxodromic if and only if* $a+d$ *is non-real.*

Proof. It is an immediate consequence of Theorem 2.1 that f is a parabolic transformation precisely in the circumstance that $a+d=\pm 2$. It remains to treat the case when f has two fixed points z_1 and z_2, where we continue to observe the convention that $z_1 \neq \infty$. Let κ be a multiplier for f. We claim first that

$$\kappa+\kappa^{-1}=(a+d)^2-2 . \tag{9.40}$$

If $z_2=\infty$, then $c=0$, $ad=1$, and $\kappa=f'(z_1)=a/d$. Therefore,

$$\kappa+\kappa^{-1}=\frac{a}{d}+\frac{d}{a}=\frac{a^2+2ad+d^2-2ad}{ad}=(a+d)^2-2 . \tag{9.41}$$

If $z_2 \neq \infty$, then $c \neq 0$ and we may assume that $\kappa=f'(z_1)=1/f'(z_2)$. In this case $\kappa+\kappa^{-1}=f'(z_1)+f'(z_2)$. Using the fact that $ad-bc=1$, one checks that

$$f'(z)=\frac{1}{(cz+d)^2} .$$

Moreover, as the proof of Theorem 2.1 indicates, we can identify the fixed points of f as

$$z_1=\frac{(a-d)+\sqrt{(a+d)^2-4}}{2c} , \quad z_2=\frac{(a-d)-\sqrt{(a+d)^2-4}}{2c} .$$

It follows that

$$\kappa+\kappa^{-1}=f'(z_1)+f'(z_2)=\frac{1}{(cz_1+d)^2}+\frac{1}{(cz_2+d)^2}$$

$$=\frac{4}{[(a+d)+\sqrt{(a+d)^2-4}]^2}+\frac{4}{[(a+d)-\sqrt{(a+d)^2-4}]^2}$$

$$=\frac{4[(a+d)-\sqrt{(a+d)^2-4}]^2+4[(a+d)+\sqrt{(a+d)^2-4}]^2}{\{[(a+d)+\sqrt{(a+d)^2-4}][(a+d)-\sqrt{(a+d)^2-4}]\}^2}$$

$$=\frac{16(a+d)^2-32}{16}=(a+d)^2-2 ,$$

which combined with (9.41) completes the verification of (9.40).

We proceed to write κ in polar form: $\kappa=re^{i\theta}$, with $\theta=\operatorname{Arg}\kappa$. For κ to be both real and positive, it is required that $\theta=0$. In this case $r=1$ is ruled out, as $\kappa \neq 1$. Therefore, a hyperbolic transformation f has

$$(a+d)^2=2+\kappa+\kappa^{-1}=2+r+r^{-1} ;$$

i.e., $(a+d)^2$ is real and positive. Furthermore, $(a+d)^2 > 4$, because when r varies over $(0,\infty)$ the quantity $r + r^{-1}$ attains its minimum value of 2 only for $r = 1$. Thus, for f to be hyperbolic it is necessary that $a+d$ be real and satisfy $|a+d| > 2$. Next, suppose that f is elliptic. Then $|\kappa| = 1$. Since $\kappa \neq 1$, this only happens when $r = 1$ and $\theta \neq 0$. The result is that

$$(a+d)^2 = 2 + \kappa + \kappa^{-1} = 2 + e^{i\theta} + e^{-i\theta} = 2 + 2\cos\theta < 4 ,$$

which implies that in the elliptic case $a+d$ is real and has $|a+d| < 2$. Looking at the remaining cases — $\theta \neq 0$ and $r \neq 1$ — we see that the quantity $\kappa + \kappa^{-1}$ is not real when $0 < |\theta| < \pi$ and belongs to the interval $(-\infty, -2)$ when $\theta = \pi$. Under these conditions $(a+d)^2$ is found to be a member of the set $\mathbb{C} \sim [0,\infty)$. As a consequence, $a+d$ must be non-real for loxodromic f. Since our list of categories of non-identity Möbius transformations is exhaustive and these categories are mutually exclusive, and since the same is true of the list of possibilities for $a+d$ in the statement of Theorem 2.8, the proof is complete. ∎

It is an easy consequence of Theorem 2.8 that conjugate Möbius transformations always fall into the same class (Exercise 6.48). We apply Theorem 2.8 and the discussion preceding it to a pair of examples.

EXAMPLE 2.4. Classify and then describe the important properties of the transformation $f(z) = (4z+6)/(2z+4)$.

We begin by rewriting f in normalized form: $f(z) = (2z+3)/(z+2)$. Here $a + d = 4$, so on the basis of Theorem 2.8 we can say that f is hyperbolic. Its fixed points are $z_1 = \sqrt{3}$ and $z_2 = -\sqrt{3}$. To find a multiplier for f we compute

$$\kappa = f'(z_1) = \frac{1}{(z_1+2)^2} = \frac{1}{7+4\sqrt{3}} .$$

Since $|f'(z_1)| < 1$, z_1 is the attracting fixed point of f and z_2 is its repelling fixed point.

EXAMPLE 2.5. Classify the transformation $f(z) = [(1+i)z - i]/z$, and indicate any special features that it might possess.

In normalized form this mapping becomes

$$f(z) = \frac{\sqrt{2}\,z - e^{\pi i/4}}{e^{-\pi i/4}\,z + 0} .$$

Thus $a+d = \sqrt{2}$, which stamps f as elliptic. It has fixed points $z_1 = 1$ and $z_2 = i$. For a multiplier we can use $\kappa = f'(1) = i$. Noting that $\kappa^4 = 1$, whereas $\kappa^n \neq 1$ for $1 \leq n \leq 3$, we infer that f is a transformation of order four.

2.8 Invariant Circles

Save for those loxodromic transformations whose multipliers are not real, every non-identity Möbius transformation f has a characteristic family of *invariant circles*. This name applies to any circle K in $\widehat{\mathbb{C}}$ having the property that $f(K) = K$. With the help of its invariant circles it is possible to draw a very instructive picture detailing how the transformation acts on $\widehat{\mathbb{C}}$. We shall carry this out for f, in turn, of parabolic, hyperbolic, and elliptic type. In each instance we deal first with an elementary transformation, where the geometry is quite obvious. The treatment of the general transformation in each category is reduced to the elementary case by conjugation, using the principle that K is an invariant circle for $h^{-1} \circ f \circ h$ if and only if $h(K)$ is an invariant circle for f (Exercise 6.52). Invariant circles, however, paint only half the picture. To complete it we track the image under f and its sundry (forward and backward) iterates f_n of a circle $\widetilde{K}_0$ selected from the family dual to the invariant circles of f, the latter being the family of circles $\widetilde{K}$ in $\widehat{\mathbb{C}}$ that are perpendicular to every f-invariant circle. We repeat for emphasis that a loxodromic transformation with a multiplier that is not a negative real number does not have any invariant circles (Exercise 6.56).

Parabolic Case. Suppose that f is a parabolic transformation with fixed point z_0. If $z_0 = \infty$, then f is an elementary transformation — a translation, to be exact — say $f(z) = z + b$ with $b \neq 0$. It is evident that $K_0 = L_0 \cup \{\infty\}$, where L_0 is the line through 0 and $b = f(0)$, is an invariant circle for f, as is $K = L \cup \{\infty\}$ for any line L parallel to L_0. These are readily seen to be the only circles in $\widehat{\mathbb{C}}$ invariant under f. If $\widetilde{L}_0$ denotes the line perpendicular to L_0 at the origin, then the images of $\widetilde{K}_0 = \widetilde{L}_0 \cup \{\infty\}$ under f and its iterates partition the plane into parallel strips. The geometry of this mapping is captured by Figure 16, where the arrows indicate the direction in which points move along invariant circles under the application of f. One message which this diagram (and later diagrams akin to it) seeks to convey is that f maps a shaded region onto an adjacent unshaded region, and vice versa.

If z_0 is a finite point, we consider $g = h^{-1} \circ f \circ h$, where h is a Möbius transformation taking ∞ to z_0. As g is parabolic and fixes ∞, we know the nature of its invariant circles. In particular, if we choose one invariant circle K_0^* for g, then we can characterize all remaining g-invariant circles as the circles in $\widehat{\mathbb{C}}$ that intersect K_0^* only at ∞, the fixed point of g. It follows from Exercise 6.52 that, once we pin down a single invariant circle K_0 for f, we shall have determined all invariant circles for this transformation — namely, K_0 and every circle K in $\widehat{\mathbb{C}}$ that intersects K_0 only at z_0. Recalling (9.30) we write,

$$f(z) = z_0 + \frac{z - z_0}{1 + b(z - z_0)} ,$$

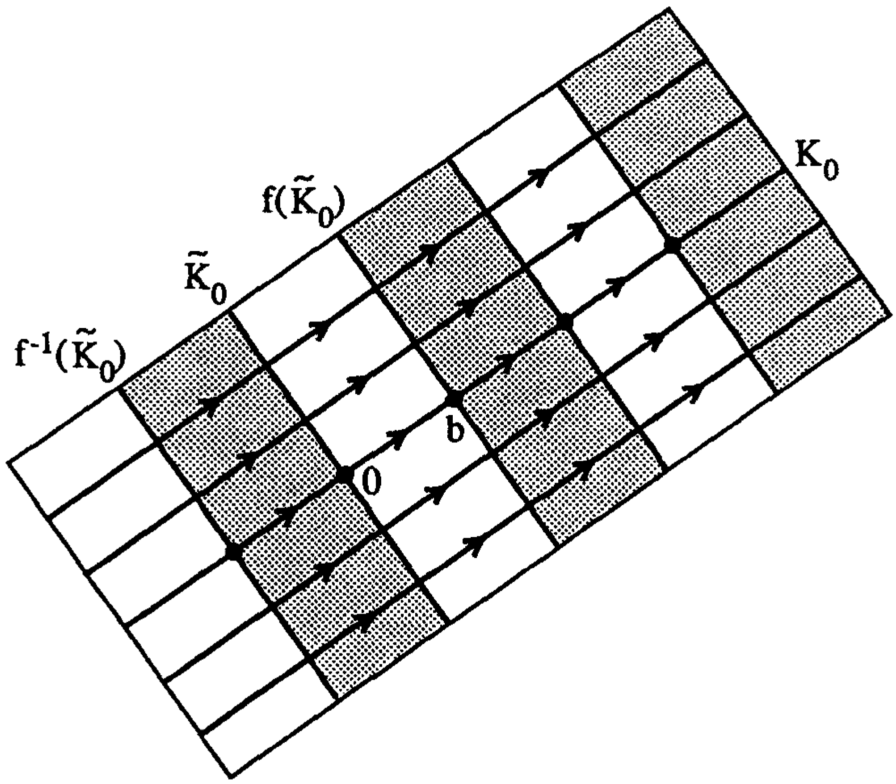


Figure 16.

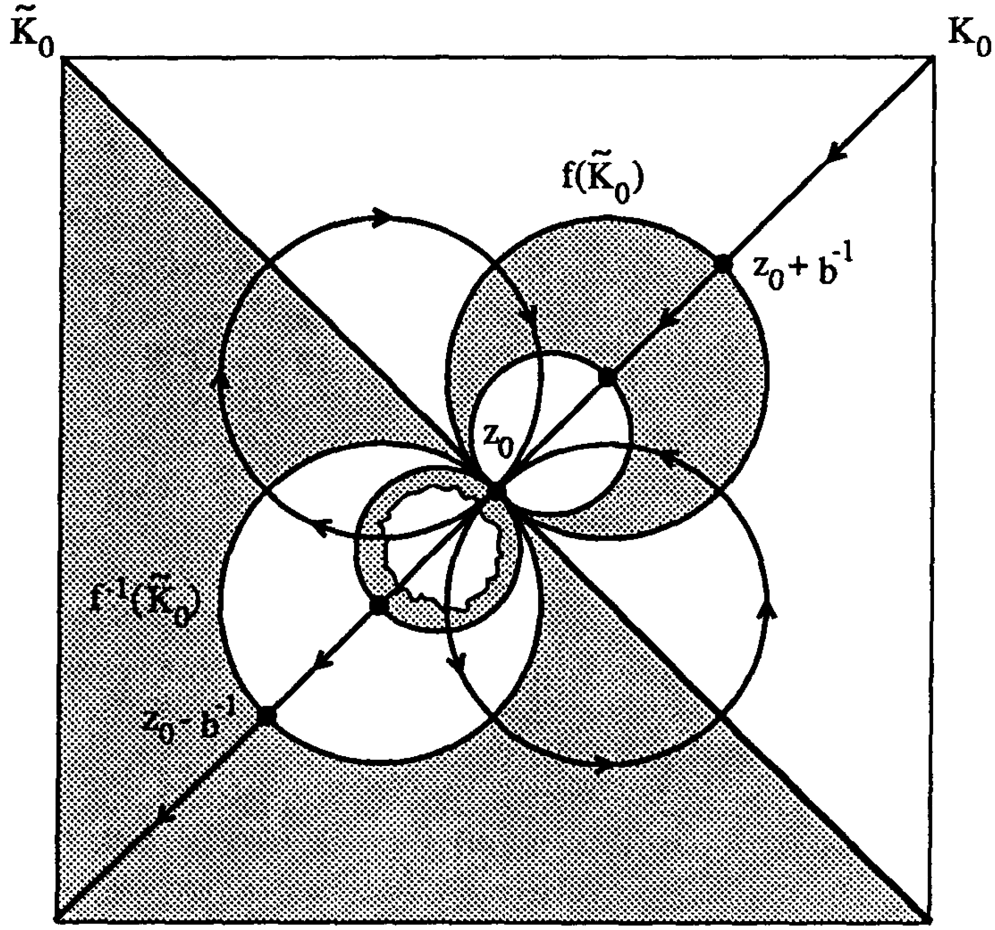


Figure 17.

where $b \neq 0$ is given by

$$b = \lim_{z \to z_0} \left[\frac{1}{f(z) - z_0} - \frac{1}{z - z_0} \right] .$$

As observed at the time, the points z_0, $f(\infty) = z_0 + b^{-1}$, and $f^{-1}(\infty) = z_0 - b^{-1}$ lie on a line L_0 in the finite plane, so $K_0 = L_0 \cup \{\infty\}$ must be an invariant circle for f. In light of the foregoing remarks its other invariant circles are just the authentic circles in $\mathbb{C}$ that are tangent to L_0 at z_0. We direct the reader to Figure 17, the analogue of Figure 16 for a parabolic transformation whose fixed point z_0 is finite. The direction of the arrows is dictated by the fact that f transforms $z_0 + b^{-1}$ to $z_0 + (2b)^{-1}$, a point situated between z_0 and $z_0 + b^{-1}$. Taking $\widetilde{K}_0 = \widetilde{L}_0 \cup \{\infty\}$, where $\widetilde{L}_0$ is the line orthogonal to L_0 at z_0, we observe that here the images of $\widetilde{K}_0$ under f and its iterates partition the set $\mathbb{C} \sim \{z_0\}$ into (with two unbounded exceptions) crescent-shaped pieces.

Hyperbolic Case. Let f be a hyperbolic transformation with fixed points z_1 and z_2 In the situation where $z_1 = 0$ and $z_2 = \infty$ the mapping is a simple dilation, $f(z) = \kappa z$ with κ positive and different from 1. The circles invariant under f are easily described: they are the circles in $\widehat{\mathbb{C}}$ that pass through both fixed points of f; i.e., the circles $K = L \cup \{\infty\}$, in which L is a line through the origin. A convenient member of the perpendicular family is the circle $\widetilde{K}_0 = K(0,1)$. Under iteration of f, $\widetilde{K}_0$ and its images subdivide the punctured plane $\mathbb{C} \sim \{0\}$ into concentric annular regions. (See Figure 18. There we have assumed that $\kappa < 1$, which makes the origin the attracting fixed point of f.)

In the general hyperbolic case, we conclude by means of the conjugation principle that the family of invariant circles of f still consists of all the circles in $\widehat{\mathbb{C}}$ passing through both z_1 and z_2. (Here we would choose the conjugating Möbius transformation h so as to send 0 to z_1 and ∞ to z_2.) Figure 19 illustrates the basic geometry of f when both z_1 and z_2 are finite and z_1 is the attracting fixed point. In this situation $K_0 = L_0 \cup \{\infty\}$, with L_0 the line through z_1 and z_2, is one of the invariant circles for f. As a circle from the perpendicular family we can take $\widetilde{K}_0 = \widetilde{L}_0 \cup \{\infty\}$, $\widetilde{L}_0$ being the line perpendicular to L_0 at $(z_1 + z_2)/2$. Incidentally, notice that the fixed points z_1 and z_2 of a hyperbolic transformation f are always symmetric with respect to any circle in $\widehat{\mathbb{C}}$ that is perpendicular to every f-invariant circle. This is clear when $z_1 = 0$ and $z_2 = \infty$; through conjugation it then follows from conformality and the symmetry principle (Theorem 2.7) for an arbitrary hyperbolic transformation.

Elliptic Case. Consider, finally, an elliptic transformation f with fixed points z_1 and z_2. In case $z_1 = 0$ and $z_2 = \infty$ we can write $f(z) = e^{i\theta} z$, where $\theta \neq 0$ and $-\pi < \theta \leq \pi$. One obvious class of circles preserved by such a rotation consists of the true circles centered at the origin or, described in terms of Möbius invariant concepts, the circles in $\widehat{\mathbb{C}}$ with respect to which the fixed points of f are symmetric. (Unless $\theta = \pi$, f has no other invariant

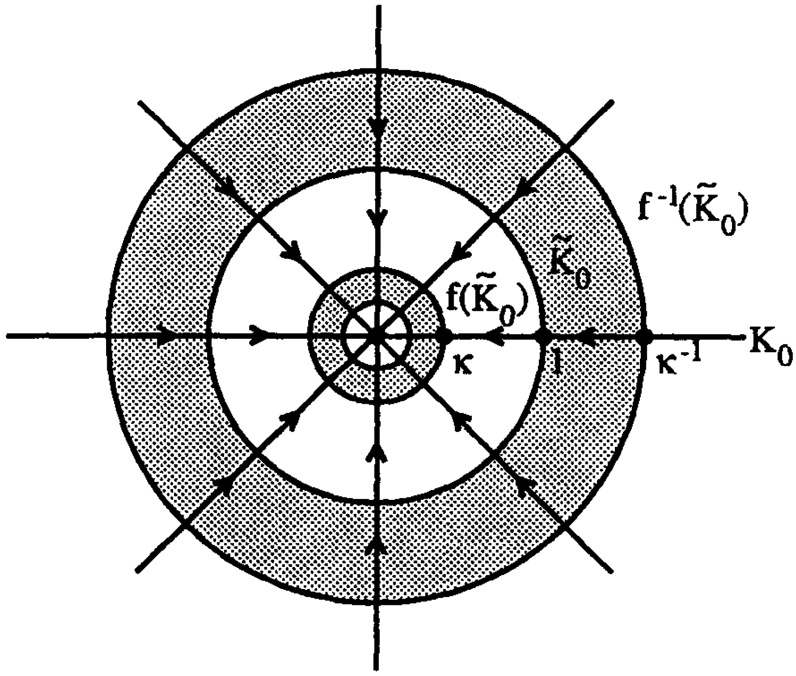


Figure 18.

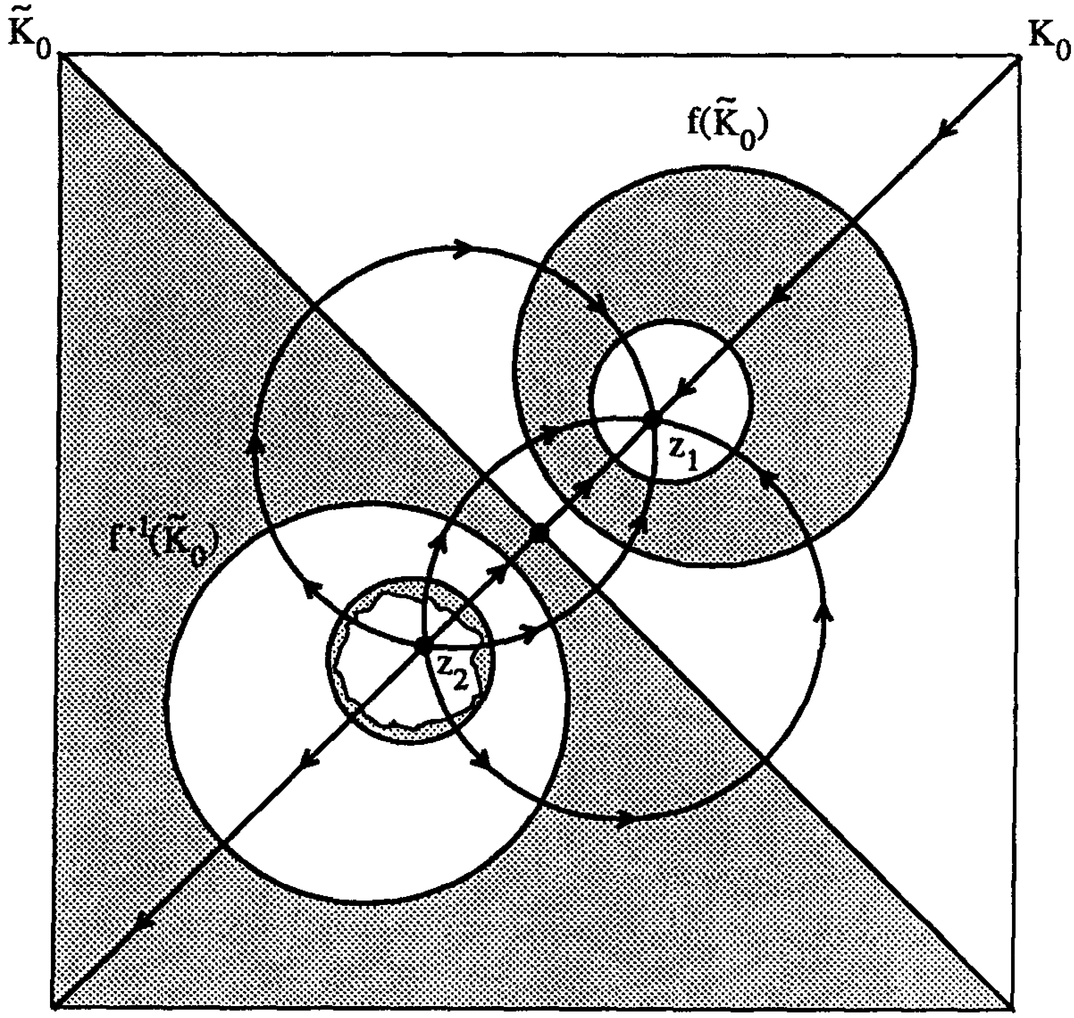


Figure 19.

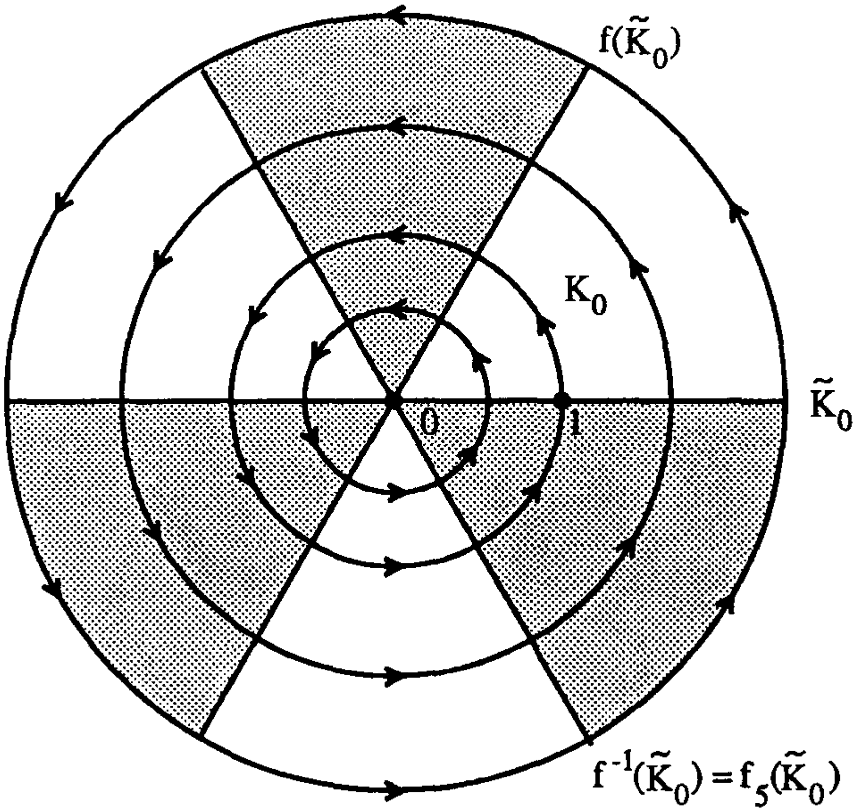


Figure 20.

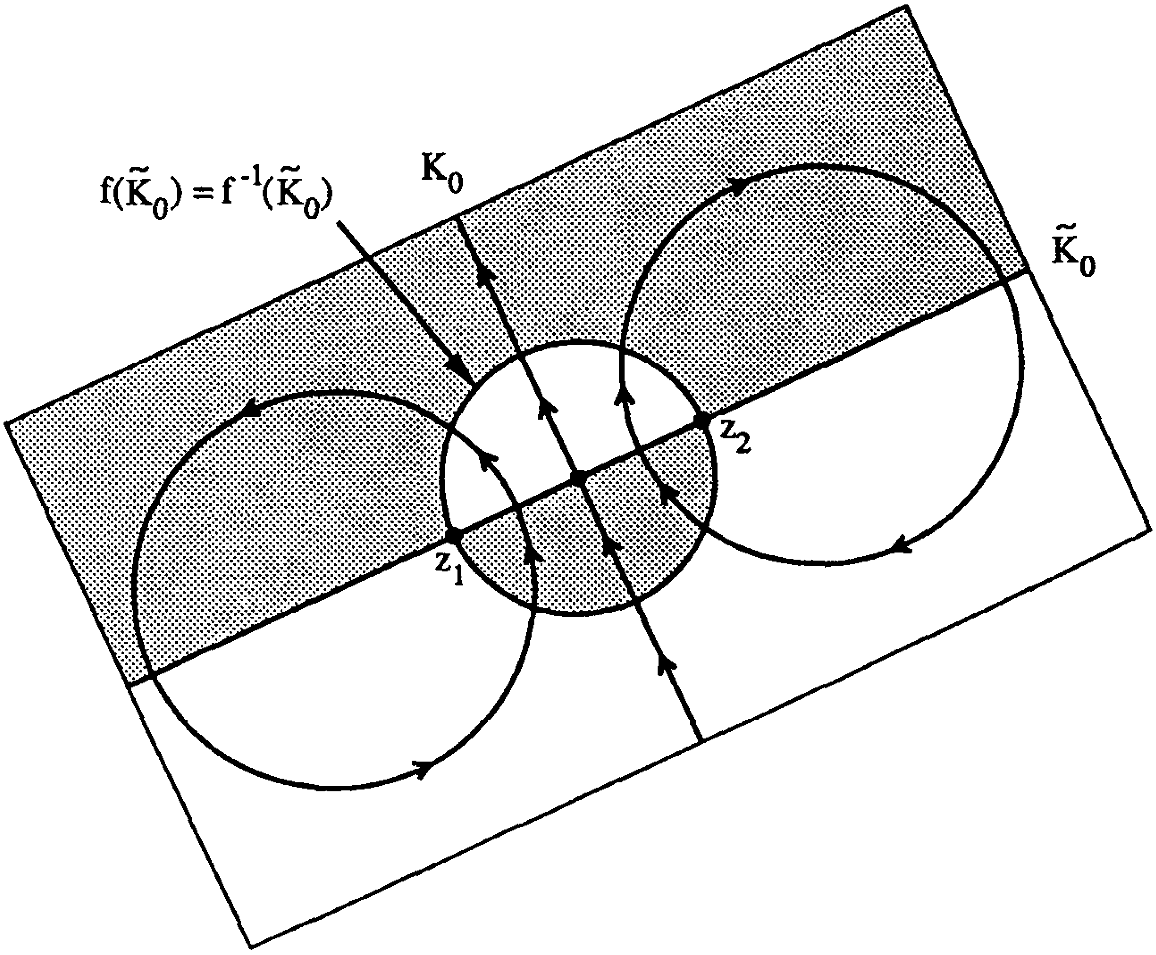


Figure 21.

circles. When $\theta = \pi$, which means that $f(z) = -z$, there is an additional family of f-invariant circles: in this special case any circle in $\widehat{\mathbb{C}}$ that passes through both 0 and ∞ is also invariant under f.) If $\widetilde{K} = \widetilde{L} \cup \{\infty\}$ is a circle in $\widehat{\mathbb{C}}$ that passes through both of these fixed points, then $f(\widetilde{K})$ is a second circle of the same kind and the oriented angle from $\widetilde{K}$ to $f(\widetilde{K})$ at the origin is θ. This statement applies, in particular, to $\widetilde{K}_0 = \mathbb{R} \cup \{\infty\}$. The images of $\widetilde{K}_0$ under f and its iterates do not typically partition the plane in any reasonable way; indeed, they normally seem to "fill up" the plane in a very chaotic fashion. An exception occurs when $\theta = 2\pi q$ for a rational number q; i.e., when f has finite order. In this case the images of $\widetilde{K}_0$ subdivide the plane into finitely many sectors, as pictured in Figure 20 for $\theta = \pi/3$.

Via conjugation we see that the general elliptic transformation f leaves invariant exactly those circles in $\widehat{\mathbb{C}}$ with respect to which its fixed points z_1 and z_2 are symmetric, supplemented when the multiplier of f is -1 by the circles in $\widehat{\mathbb{C}}$ that go through both z_1 and z_2. It maps an arbitrary circle passing through both fixed points to a new circle through those points. At z_1 — remember that by our notational convention $z_1 \neq \infty$ — this new circle intersects the original one at an angle $\theta = \operatorname{Arg} f'(z_1)$. Figure 21 depicts the situation for two finite fixed points when $\theta = \pi/2$. Of course, if both fixed points are finite, then the perpendicular bisector L_0 of the segment between z_1 and z_2 is part of an invariant circle for f, that being $K_0 = L_0 \cup \{\infty\}$. A natural choice of circle $\widetilde{K}_0$ from the orthogonal family is then found in $\widetilde{K}_0 = \widetilde{L}_0 \cup \{\infty\}$, where $\widetilde{L}_0$ is the line through z_1 and z_2.

By way of transition from the topic of Möbius transformations back to the subject of general conformal mappings, we close this section with two examples that demonstrate how Möbius transformations and the concepts associated with them enhance our ability to construct conformal mappings.

EXAMPLE 2.6. Determine a conformal mapping of the open half-disk $D = \{z : |z| < 1, \operatorname{Im} z > 0\}$ onto the full disk $D' = \Delta(0,1)$ (Figure 22).

We start by subjecting D to a Möbius transformation f_1 that sends 1 to 0 and -1 to ∞. (Reason: f_1 will transform D to a domain bounded by a pair of rays emanating from the origin. Such a domain is not hard to map to D'.) We choose $f_1(z) = (1-z)/(1+z)$, a mapping we have worked with before when we used it to map D' onto the half-plane $D'' = \{z : \operatorname{Re} z > 0\}$. Now $f_1(0) = 1$, which implies that f_1 transforms $\mathbb{R} \cup \{\infty\}$ to the circle in $\widehat{\mathbb{C}}$ through 0, 1, and ∞ — hence, to itself. It maps the interval $(-1,1)$ to the interval $(0,\infty)$. The symmetry property of Möbius transformations implies that the domains $f_1(D)$ and $f_1[\Delta(0,1) \sim \overline{D}]$ are symmetric with respect to the real axis. It follows that $f_1(D)$ is one of the two quarter planes into which $(0,\infty)$ separates D''. Lastly, since the curvilinear angle whose sides are the interval $A = [0,1]$ and the circular arc $B = \{z : |z| = 1, \operatorname{Im} z \geq 0\}$ has $\theta(A,B) = -\pi/2$, the conformality of f_1 tells us that $\theta[f_1(A), f_1(B)] = -\pi/2$. This added information is enough to nail

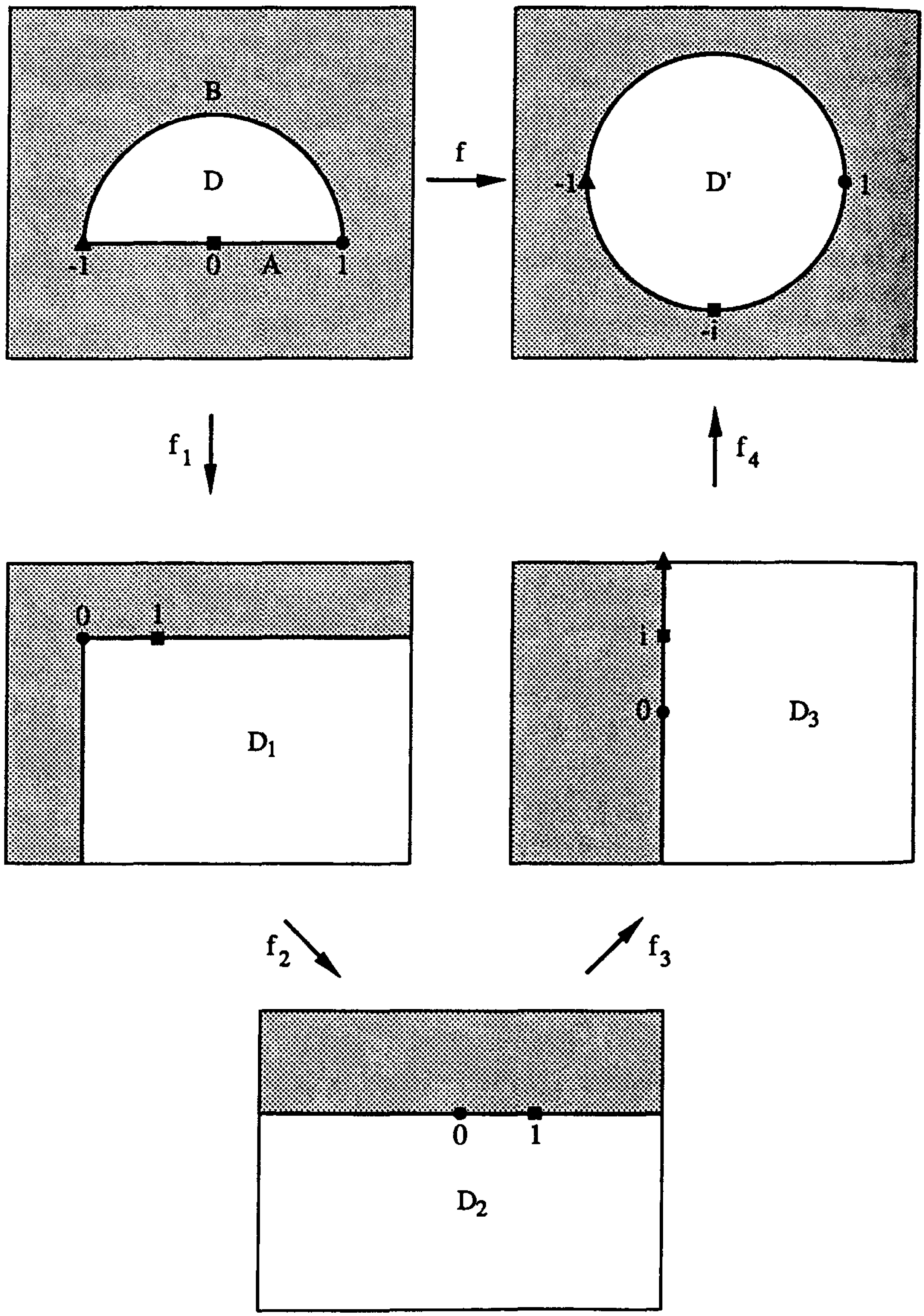


Figure 22.

down the image of D under f_1: $D_1 = f_1(D) = \{z : \operatorname{Re} z > 0, \operatorname{Im} z < 0\}$. We proceed with the construction of a mapping f in the manner proposed by Figure 22. We take $f = f_4 \circ f_3 \circ f_2 \circ f_1$ with f_1 as above, $f_2(z) = z^2$, $f_3(z) = iz$, and $f_4 = f_1$. Evaluation of this composition at a point z of D yields the formula

$$f(z) = \frac{(1+z)^2 - i(1-z)^2}{(1+z)^2 + i(1-z)^2}$$

for a conformal mapping of D onto D'.

EXAMPLE 2.7. Find a conformal mapping of the region D pictured in Figure 23 onto the strip $D' = \{z : |\operatorname{Im} z| < \pi/2\}$.

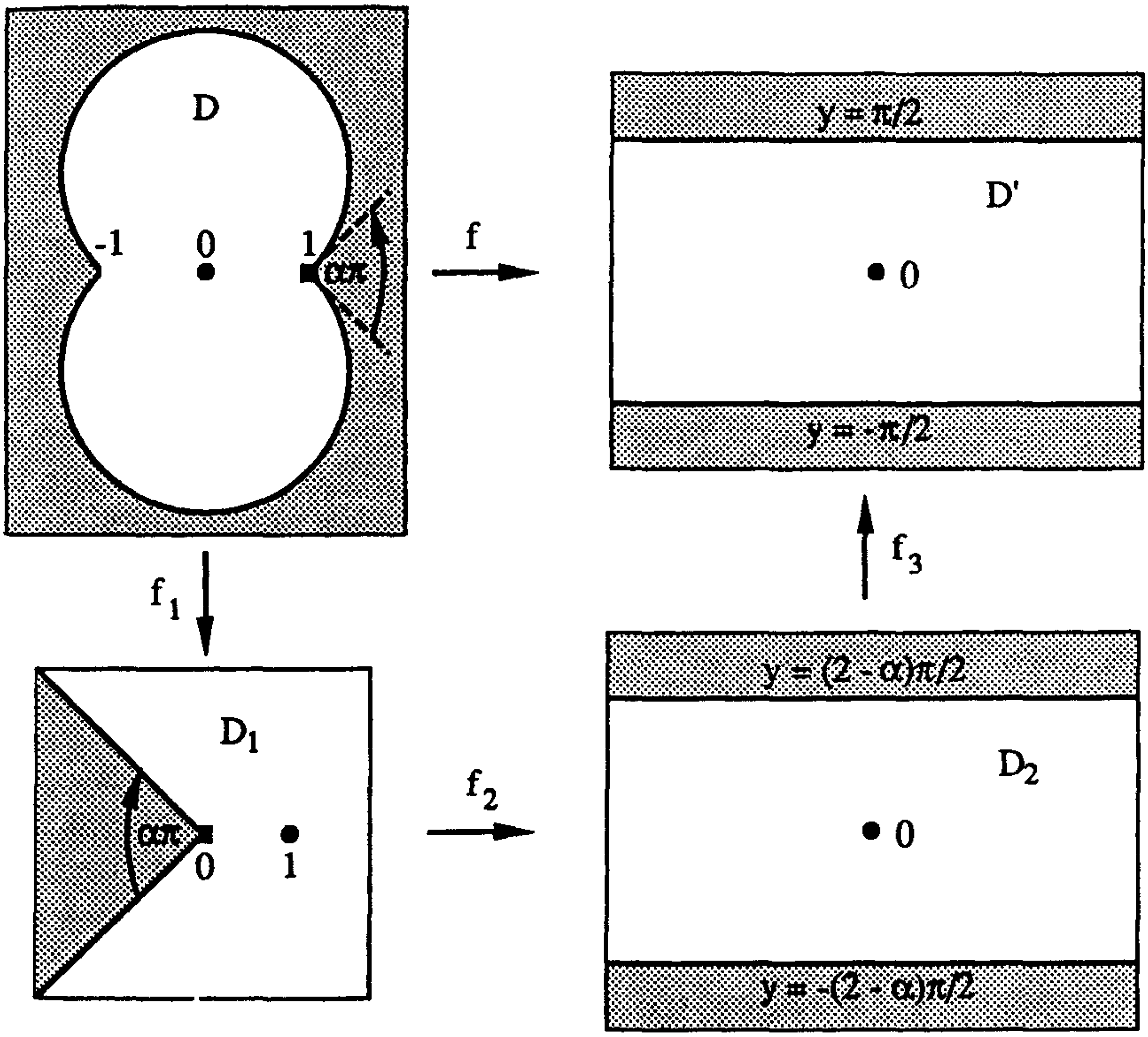


Figure 23.

Figure 23 is intended to indicate that the domain D is the union of two disks of the same radius whose bounding circles intersect in an angle $\alpha\pi, 0 < \alpha < 1$, at the points 1 and -1. In particular, D is symmetric with respect to the real axis. As in the previous example, the first step here is to "straighten out" the boundary of D. Once again this is accomplished by means of the mapping $f_1(z) = (1-z)/(1+z)$. The domain $D_1 = f_1(D)$ is symmetric with respect to the real axis, for D itself has this property and the Möbius transformation f_1 leaves the set $\mathbb{R} \cup \{\infty\}$ invariant. Also, D_1 is bounded by two rays issuing from the origin. Because the curvilinear angle exterior to D formed at the point 1 by the two circular arcs that compose ∂D has magnitude $\alpha\pi$, the same must be true of the angle exterior to D_1 formed by the aforementioned rays at the origin. Since f_1 maps the interval $(-1, 1)$ to $(0, \infty)$, the latter interval lies in D_1. The domain D_1 is now completely determined: $D_1 = \{z : |\operatorname{Arg} z| < (2-\alpha)\pi/2\}$. A mapping f of D onto D' is obtained in the form $f = f_3 \circ f_2 \circ f_1$ suggested by Figure 23. Here f_1 is as announced, $f_2(z) = \operatorname{Log} z$, and $f_3(z) = z/(2-\alpha)$. Therefore,

$$f(z) = \frac{1}{2-\alpha} \operatorname{Log}\left(\frac{1-z}{1+z}\right)$$

is a mapping that does the required job.

3 Riemann's Mapping Theorem

3.1 Preparations

It is a noteworthy item in the lore of complex function theory that the proof which Riemann offered for his mapping theorem was flawed, resting as it did on the faulty hypothesis that a certain minimization problem in the calculus of variations is guaranteed to have a solution. In all fairness to Riemann, few of his contemporaries would have batted an eye at this assumption: the existence of solutions to such problems was taken largely for granted by mathematicians of the day. Not until Weierstrass appeared on the scene and spearheaded a campaign to enforce stricter standards of rigor in analysis did existence questions of this type begin to attract closer attention. Furthermore, the basic features of Riemann's approach were eventually salvaged by David Hilbert (1862-1943) and Richard Courant (1888-1972), albeit some fifty years after Riemann's original work. In the interim other proofs for the mapping theorem were devised. The argument we present is based on Montel's normal family theorem and an idea of Paul Koebe (1882-1945), to whom this particular proof is usually credited. For convenience we include the more technical aspects of the proof in three preparatory lemmas. The first of these informs us that simple connectivity, as we have defined the term, is preserved under conformal mapping.

Lemma 3.1. *Let $f: D \to \mathbb{C}$ be a conformal mapping, where D is a simply connected domain in the complex plane. Then $D' = f(D)$ is also a simply connected domain.*

Proof. If $D' = \mathbb{C}$, the conclusion is obvious. We assume, therefore, that $D' \neq \mathbb{C}$. Since f is a non-constant analytic function, D' is certainly a domain. We must prove that $n(\beta, w) = 0$ whenever w is a point of $\mathbb{C} \sim D'$ and β is a closed, piecewise smooth path in D'. Fix w and β as described, say $\beta: [a, b] \to \mathbb{C}$. Define $\gamma: [a, b] \to \mathbb{C}$ by $\gamma = f^{-1} \circ \beta$, so that $\beta(t) = f[\gamma(t)]$. Then γ is a closed, piecewise smooth path in D. As D is simply connected, γ is homologous to zero in this domain. Because w does not belong to D', the function $f'/(f - w)$ is analytic in D. Cauchy's theorem thus leads to

$$0 = \int_\gamma \frac{f'(z)\,dz}{f(z) - w} = \int_a^b \frac{f'[\gamma(t)]\dot{\gamma}(t)\,dt}{f[\gamma(t)] - w} = \int_a^b \frac{\dot{\beta}(t)\,dt}{\beta(t) - w}$$

$$= \int_\beta \frac{d\zeta}{\zeta - w} = 2\pi i\, n(\beta, w) \ ,$$

which confirms that $n(\beta, w) = 0$. ∎

Our second preliminary lemma states that a simply connected subdomain of $\mathbb{C}$ which is not the whole plane can always be transformed conformally to a domain that is contained in the unit disk.

Lemma 3.2. *Let D be a simply connected domain in the complex plane, $D \neq \mathbb{C}$, and let z_0 be a point of D. There is a conformal mapping $f: D \to \mathbb{C}$ with the following properties: (i) the domain $f(D)$ is contained in $\Delta = \Delta(0, 1)$; (ii) $f(z_0) = 0$ and $f'(z_0) > 0$.*

Proof. We shall exhibit a mapping f with the stated properties as the composition $f = f_5 \circ f_4 \circ f_3 \circ f_2 \circ f_1$ of five relatively simple conformal mappings, which we now proceed to describe. To start, we choose a point b of $\mathbb{C} \sim D$ and define $f_1: D \to \mathbb{C}$ by $f_1(z) = z - b$. Since D is a proper subdomain of $\mathbb{C}$, the selection of b is possible. The mapping f_1 merely translates D to a simply connected domain $D_1 = f_1(D)$ that does not contain the origin. For the mapping f_2 we take any branch of $\log z$ in D_1. That such a branch f_2 exists is guaranteed by Theorem V.6.2. Moreover, we know that f_2 is a univalent analytic function. Fix a point w_0 in the domain $D_2 = f_2(D_1)$, together with a radius $r > 0$ for which the closed disk $\overline{\Delta}(w_0, r)$ lies in D_2. Setting $\widetilde{w}_0 = w_0 + 2\pi i$, we observe that $\overline{\Delta}(\widetilde{w}_0, r)$ and D_2 are necessarily disjoint. Indeed, should there be a point $\widetilde{w}$ in $\overline{\Delta}(\widetilde{w}_0, r) \cap D_2$, then on the one hand $\widetilde{w}$ would be of the form $\widetilde{w} = f_2(\widetilde{z})$ for some $\widetilde{z}$ in D_1, while on the other we could represent $\widetilde{w}$ as $\widetilde{w} = w + 2\pi i$ for some w in $\overline{\Delta}(w_0, r)$. Of course, it would also be true that $w = f_2(z)$ with z a point of D_1, implying that

$$\widetilde{z} = e^{f_2(\widetilde{z})} = e^{\widetilde{w}} = e^{w + 2\pi i} = e^w = e^{f_2(z)} = z \ .$$

It would follow that $w = f_2(z) = f_2(\widetilde{z}) = \widetilde{w} = w + 2\pi i$, an obvious contradiction. We conclude that $\overline{\Delta}(\widetilde{w}_0, r) \cap D_2 = \phi$, as claimed. Accordingly, $|z - \widetilde{w}_0| > r$ holds for every z in D_2. This fact implies that the image $D_3 = f_3(D_2)$ of D_2 under the Möbius transformation $f_3(z) = r/(z - \widetilde{w}_0)$ is a subset of Δ. The function $f_3 \circ f_2 \circ f_1$ thus furnishes a conformal mapping of D whose range is contained in Δ.

Next, set $c = f_3 \circ f_2 \circ f_1(z_0)$. According to Theorem 1.4 the function $f_4 : \Delta \to \Delta$ given by $f(z) = (z-c)/(1-\bar{c}z)$ maps Δ conformally onto itself. It carries the point c to the origin. The composition $f_4 \circ f_3 \circ f_1 \circ f_1$ then maps D conformally inside Δ and transforms z_0 to the origin.

Finally, $d = (f_4 \circ f_3 \circ f_2 \circ f_1)'(z_0) \neq 0$ (Theorem VIII.3.9). Write $u = e^{-i \operatorname{Arg} d}$ and define f_5 in Δ by $f_5(z) = uz$. Then $f = f_5 \circ f_4 \circ f_3 \circ f_2 \circ f_1$ provides a conformal mapping of D onto a subdomain of Δ. It has $f(z_0) = 0$ and $f'(z_0) = f_5'(0)(f_4 \circ f_3 \circ f_2 \circ f_1)'(z_0) = ud = |d| > 0$. In short, f is a conformal mapping of D that enjoys properties (i) and (ii). ∎

The final lemma in the present sequence articulates the idea of Koebe that was alluded to earlier.

Lemma 3.3. *Suppose that D is a simply connected domain in the complex plane, $D \neq \mathbb{C}$, that z_0 is a point of D, and that $f : D \to \mathbb{C}$ is a conformal mapping which exhibits properties (i) and (ii) in Lemma 3.2. Assume that $f(D) \neq \Delta$. Then there exists a conformal mapping $g : D \to \mathbb{C}$ that is also endowed with these two properties, but has $g'(z_0) > f'(z_0)$.*

Proof. Write $D_0 = f(D)$. We again obtain g in the form of a composition $g = g_3 \circ g_2 \circ g_1 \circ f$, where the mappings g_1, g_2, and g_3 are constructed as follows. First, select a point b in $\Delta \sim D_0$. Since $0 = f(z_0)$ is an element of D_0, $b \neq 0$. Theorem 1.4 tells us that the Möbius transformation defined by $g_1(z) = (z-b)/(1-\bar{b}z)$ maps Δ conformally onto itself — hence, maps the simply connected subdomain D_0 of Δ onto another such domain $D_1 = g_1(D_0)$. The domain D_1 does not contain the origin ($= g_1(b)$), but it does contain the point $-b = g_1(0)$. (See Figure 24.) Direct calculation shows that $g_1'(0) = 1 - |b|^2$.

Next, Theorem V.6.2 certifies the existence of a branch of $\log z$ in D_1. We choose one and call it L. More to the point, there is associated with L a branch g_2 of the square root function in D_1 — namely, the one given by $g_2(z) = \exp[L(z)/2]$. Then $|g_2(z)| = \sqrt{|z|} < 1$ for z in D_1, and g_2 is univalent in D_1: if $g_2(z) = g_2(\tilde{z})$, then $z = [g_2(z)]^2 = [g_2(\tilde{z})]^2 = \tilde{z}$. In other words, g_2 is a conformal mapping of D_1 onto some other simply connected domain $D_2 = g_2(D_1)$ lying in Δ (Figure 24). We write $c = g_2(-b)$, a point of D_2, and observe that $g_2'(-b) = 1/[2g_2(-b)] = 1/(2c)$.

Lastly, take g_3 to be the conformal self-mapping of Δ defined by $g_3(z) = u(z-c)/(1-\bar{c}z)$, with $u = \exp[i \operatorname{Arg} c]$. The domain $D_3 = g_3(D_2)$ is contained in Δ, the origin ($= g_3(c)$) belongs to D_3, and, as a simple computation reveals, $g_3'(c) = u/(1-|c|^2)$.

The composite mapping $g = g_3 \circ g_2 \circ g_1 \circ f$ transforms D conformally to D_3, sends z_0 to the origin, and has

$$g'(z_0) = g_3'(c)g_2'(-b)g_1'(0)f'(z_0) = \left(\frac{u}{1-|c|^2}\right)\left(\frac{1}{2c}\right)(1-|b|^2)f'(z_0)$$

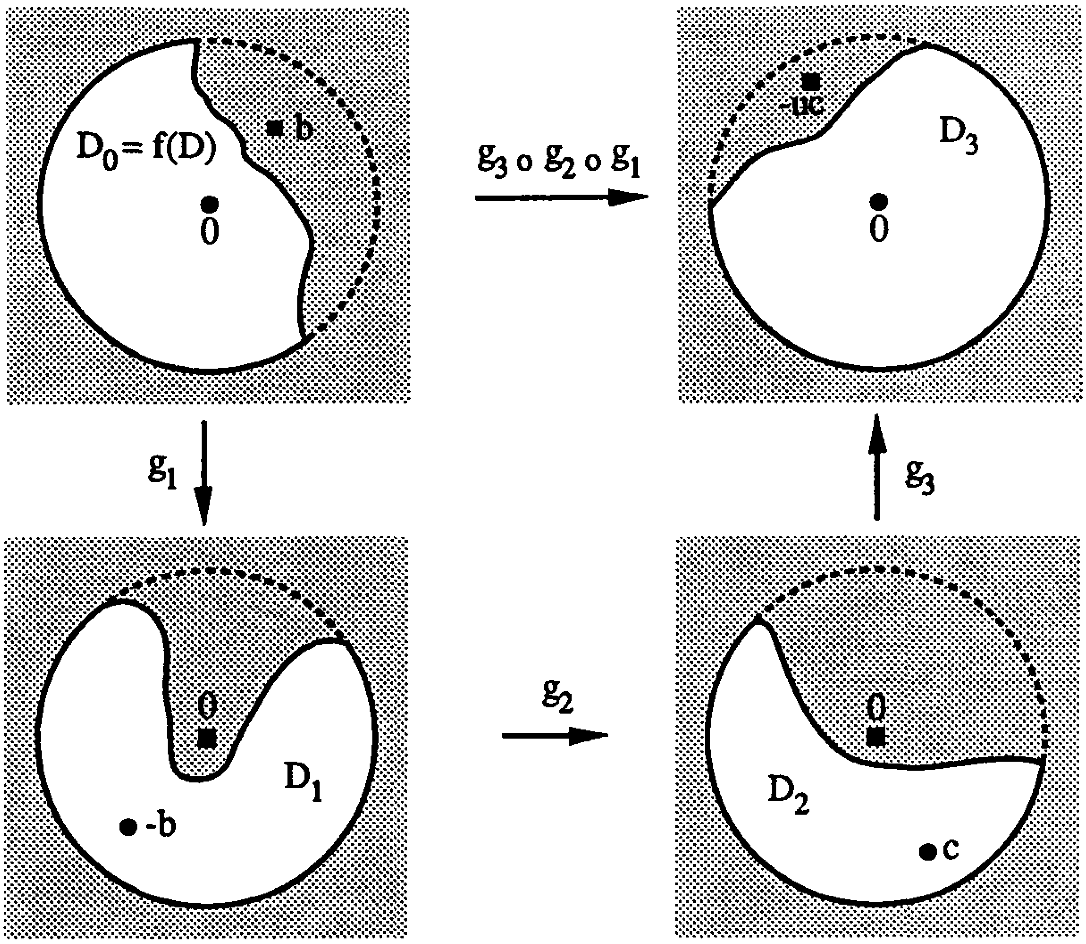


Figure 24.

$$= \left(\frac{1+|c|^2}{2|c|}\right) f'(z_0) > f'(z_0) ,$$

since $u/c = 1/|c|$, $|c|^2 = |g_2(-b)|^2 = |b|$, and $1 + |c|^2 > 2|c|$. ∎

3.2 The Mapping Theorem

We now move to the statement and proof of Riemann's mapping theorem. We preface the discussion with several remarks. First, we present the result in a normalized situation, requiring that the disk $\Delta = \Delta(0,1)$ be the image domain of the conformal mapping whose existence we establish. Nowadays this is quite standard, but it is not the way Riemann formulated his theorem. Secondly, just as Riemann sought to do, we arrive at a mapping by solving an extremal problem. The beauty of the Koebe-Montel approach, however, is that by avoiding the calculus of variations it skirts pitfalls like the one which was Riemann's misfortune. Thirdly, the conditions used in Theorem 3.4 to provide for the uniqueness of the mapping differ from those found in Riemann's own version of the theorem. All things considered, Theorem 4.11, which deals with domains bounded by Jordan curves, is more faithful to the original statement of the mapping theorem than are the results in the present section.

Theorem 3.4. **(Riemann's Mapping Theorem)** *Suppose that D is a simply connected domain in the complex plane, $D \neq \mathbb{C}$, and that z_0 is a point of D. There exists a unique conformal mapping f of D onto the disk $\Delta = \Delta(0,1)$ satisfying the conditions $f(z_0) = 0$ and $f'(z_0) > 0$.*

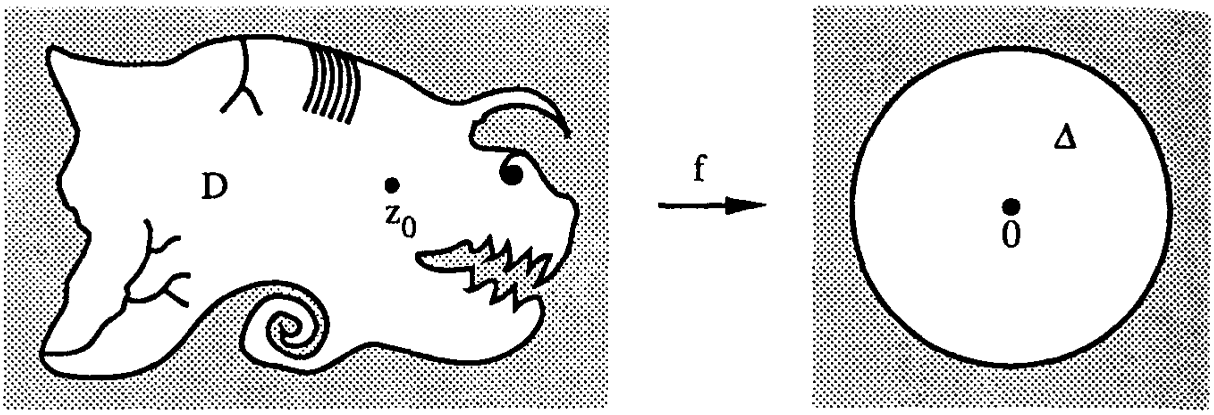


Figure 25.

Proof. Let $\mathcal{F}$ designate the family of all functions $f: D \to \mathbb{C}$ that meet the following specifications: f is a conformal mapping of D onto a subdomain of Δ, $f(z_0) = 0$, and $f'(z_0) > 0$. Lemma 3.3 insures that the family $\mathcal{F}$ is not empty. Suppose $r > 0$ has the property that $\Delta(z_0, r)$ lies in D. Cauchy's estimate (Theorem V.3.6) shows that $f'(z_0) = |f'(z_0)| \leq r^{-1}$ holds for every member f of $\mathcal{F}$. We infer that $\{f'(z_0) : f \in \mathcal{F}\}$ is a bounded set of positive real numbers. As such, this set has a least upper bound, which we denote by ℓ. For each positive integer n we can select a function f_n in $\mathcal{F}$ for which $\ell - n^{-1} \leq f_n'(z_0) \leq \ell$. Because the family $\mathcal{F}$ is manifestly locally bounded in D, Montel's theorem empowers us to extract from the sequence $\langle f_n \rangle$ a subsequence $\langle f_{n_k} \rangle$ that converges normally in D to a function f. This limit function is analytic in D. Also, $f(z_0) = \lim_{k\to\infty} f_{n_k}(z_0) = 0$ and $f'(z_0) = \lim_{k\to\infty} f'_{n_k}(z_0) = \ell > 0$. In particular, f is non-constant in D. Corollary VIII.3.13 asserts that f is univalent in this domain. Obviously $f(D)$ lies in $\overline{\Delta}$. By the open mapping theorem $f(D)$ is an open set and so must actually be a subset of Δ. The function f, therefore, is a member of the family $\mathcal{F}$. If it could be demonstrated that $f(D) = \Delta$, the existence portion of the proof would be complete. Were it true, however, that $f(D) \neq \Delta$, then Lemma 3.3 would enable us to produce a member of $\mathcal{F}$ whose derivative at z_0 exceeds $f'(z_0) = \ell$. Given the definition of ℓ, this is not possible. As a result, f must map D conformally onto Δ.

To address the question of uniqueness, assume that g is a second conformal mapping of D onto Δ which has $g(z_0) = 0$ and $g'(z_0) > 0$. Consider the function $\varphi: \Delta \to \Delta$, $\varphi = g \circ f^{-1}$. This function provides a conformal self-mapping of Δ that fixes the origin and satisfies $\varphi'(0) = g'(z_0)/f'(z_0) > 0$. Theorem 1.4 implies that the only conformal mappings φ of Δ onto itself with $\varphi(0) = 0$ are the rotations about the origin. The only such rotation with $\varphi'(0) > 0$ is the trivial rotation given by $\varphi(z) = z$. We conclude that $g(z) = \varphi[f(z)] = f(z)$ holds for every z in D; i.e., the alledged second mapping with the stated properties turns out to be nothing of the sort. ∎

Theorem 3.4 makes it possible to transform any simply connected proper subdomain D of $\mathbb{C}$ conformally to any other such domain. The mapping is uniquely determined once $f(z_0)$ and $\text{Arg}[f'(z_0)]$ are specified for some point z_0 of D.

Theorem 3.5. *Let D and D' be simply connected domains in $\mathbb{C}$, neither the whole complex plane. Corresponding to given z_0 in D, z_0' in D', and θ_0 in $(-\pi, \pi]$ there exists a unique conformal mapping f of D onto D' that obeys the conditions $f(z_0) = z_0'$ and $\text{Arg}[f'(z_0)] = \theta_0$.*

Proof. Let g be the conformal mapping of D onto $\Delta = \Delta(0,1)$ satisfying $g(z_0) = 0$ and $g'(z_0) > 0$, and let h be the corresponding mapping for the pair D' and z_0'. Then the function $f: D \to D'$ defined by $f(z) = h^{-1}[e^{i\theta_0} g(z)]$ gives a conformal mapping of D onto D' for which $f(z_0) = z_0'$ and $f'(z_0) = e^{i\theta_0} g'(z_0)/h'(z_0')$. Since $g'(z_0) > 0$ and $h'(z_0') > 0$, $\text{Arg}[f'(z_0)] = \theta_0$. To establish the uniqueness of f, consider an arbitrary conformal mapping f_0 of D onto D' with the prescribed features. Then $\varphi = g \circ f^{-1} \circ f_0 \circ g^{-1}$ is a conformal self-mapping of Δ which satisfies $\varphi(0) = 0$ and

$$\varphi'(0) = g'(0) \cdot \frac{1}{f'(z_0)} \cdot f_0'(z_0) \cdot \frac{1}{g'(0)} = \frac{f_0'(z_0)}{f'(z_0)} = \frac{|f_0'(z_0)|}{|f'(z_0)|} > 0 \ ,$$

since $\text{Arg}[f_0'(z_0)] = \text{Arg}[f'(z_0)]$. As in the proof of Theorem 3.4 we conclude that $\varphi(z) = z$ for all z in Δ, from which it follows easily that $f_0(z) = f(z)$ for every z in D. ∎

Theorems 3.4 and 3.5 fall under the heading of pure existence theorems, for they do not tell us how to construct in any explicit way the mappings whose existence they grant. In Section 5 we shall look at the problem of mapping a half-plane conformally to a polygon, a situation in which this state of affairs can be remedied to some extent: there one can represent the mapping by an elementary integral formula. The catch is that the formula contains a number of parameters which are not easily computed unless the target polygon displays a good deal of symmetry. Thus, even mappings of half-planes to polygons remain somewhat elusive creatures.

With the aid of the Riemann mapping theorem we can at last lay to rest any lingering confusion about what it means for a plane domain to be simply connected. We now show that the standard topological definition of the concept and the definition we have adopted are equivalent. We also take the occasion to introduce yet another characterization of such domains, one that involves neither homotopy nor winding numbers.

Theorem 3.6. *The following statements about a domain D in $\mathbb{C}$ are equivalent:*

(i) D is simply connected;

(ii) every closed path in D is contractible in D;

(iii) the complement of D in the extended plane $\widehat{\mathbb{C}}$ is connected.

Proof of (ii)⇒(i). This implication is a restatement of Theorem V.7.4.

Proof of (i)⇒(ii). Assume that D is simply connected, and let $\gamma: [a,b] \to \mathbb{C}$ be a closed path in D. If $D = \mathbb{C}$, then $H(t,s) = (1-s)\gamma(t)$ gives a free homotopy in D between γ and the constant path β defined on $[a,b]$ by $\beta(t) = 0$. If $D \neq \mathbb{C}$, then we can choose a conformal mapping f of D onto $\Delta = \Delta(0,1)$. In this instance $H(t,s) = f^{-1}\{(1-s)f[\gamma(t)]\}$ supplies a free homotopy in D from γ to the constant path β given by $\beta(t) = f^{-1}(0)$ for $a \le t \le b$. In either case γ is seen to be contractible in D.

Proof of (iii)⇒(i). We assume $D \neq \mathbb{C}$, for only in this case is the implication unclear. Let γ be an arbitrary closed and piecewise smooth path in D. Under the assumption that $\widehat{\mathbb{C}} \sim D$ is connected we are asked to prove that γ is homologous to zero in D; i.e., $n(\gamma, z) = 0$ for every z in $\mathbb{C} \sim D$. Let D_0 denote the component of $\widehat{\mathbb{C}} \sim |\gamma|$ that contains the point ∞. Since $|\gamma|$ lies in D, $\widehat{\mathbb{C}} \sim D$ is contained in $\widehat{\mathbb{C}} \sim |\gamma|$. Furthermore, because $\widehat{\mathbb{C}} \sim D$ is assumed to be a connected set, it must actually be contained in some component of $\widehat{\mathbb{C}} \sim |\gamma|$. As ∞ is one point that D_0 and $\widehat{\mathbb{C}} \sim D$ have in common, that unnamed component can be none other than D_0. In particular, $\mathbb{C} \sim D$ is a subset of $D_0 \sim \{\infty\}$, which is just the unbounded component of $\mathbb{C} \sim |\gamma|$. By Lemma V.2.1(ii), $n(\gamma, z) = 0$ for every z in $\mathbb{C} \sim D$.

Proof of (i)⇒(iii). We assume that D is simply connected and prove that $\widehat{\mathbb{C}} \sim D$ is connected. If $D = \mathbb{C}$, then $\widehat{\mathbb{C}} \sim D = \{\infty\}$ is certainly a connected set. We proceed under the assumption that $D \neq \mathbb{C}$, which allows us to choose a conformal mapping f of $\Delta = \Delta(0,1)$ onto D. For $0 < r < 1$ let D_r be the image of the disk $\Delta(0,r)$ under f. We show initially that the set $\widehat{\mathbb{C}} \sim D_r$ is connected. This conclusion is essentially trivial if one is prepared to invoke the Jordan curve theorem, but a direct proof is also possible. In the interest of keeping the present discussion self-contained — but at the risk of making it seem slightly long-winded — we present the latter argument.

Fix r in $(0,1)$ and write $K = f[\overline{\Delta}(0,r)]$. The set K is a compact subset of D. The first step in demonstrating that $\widehat{\mathbb{C}} \sim D_r$ is connected is to show that the open set $\mathbb{C} \sim K$ is connected. Setting this as our immediate objective, we remark that $f[\Delta \sim \overline{\Delta}(0,r)]$ is a connected subset of $\mathbb{C} \sim K$ and thus lies in some component — label it G — of $\mathbb{C} \sim K$. We claim that, in fact, $\mathbb{C} \sim K = G$. If so, $\mathbb{C} \sim K$ is definitely connected. Let w_0 be a point of $\mathbb{C} \sim K$. The compactness of K and the continuity of the function $d(w) = |w - w_0|$ insure the existence of a point w_0' of K with the property that $|w_0' - w_0| = \min\{|w - w_0| : w \in K\}$. Being a point of K, w_0' belongs to D. Let Δ_0 be an open disk centered at w_0' and contained in D. Then $\Delta_0 \sim K$ is contained in $f[\Delta \sim \overline{\Delta}(0,r)]$ and, for this reason, is contained in G. Consider $S = L \sim \{w_0'\}$, where L is the line segment with endpoints w_0 and w_0'. The set S is connected, it lies in $\mathbb{C} \sim K$, and it intersects Δ_0. It thus meets $\Delta_0 \sim K$ — hence, meets G. On the other hand, S is contained in some component of $\mathbb{C} \sim K$, a component that we can now assert to be G. In particular, the point w_0, an arbitrary point in $\mathbb{C} \sim K$, belongs to G; i.e., $\mathbb{C} \sim K = G$. It is then elementary to check that $\partial G = \partial D_r$ and to conclude that $\widehat{\mathbb{C}} \sim D_r = (\widehat{\mathbb{C}} \sim \overline{D}_r) \cup \partial D_r = G \cup \partial G \cup \{\infty\} = \overline{G} \cup \{\infty\} = \widehat{G}$, the closure of G in $\widehat{\mathbb{C}}$. Since the closure in $\widehat{\mathbb{C}}$ of a connected subset of $\widehat{\mathbb{C}}$ is again connected, $\widehat{\mathbb{C}} \sim D_r$ is connected.

To complete the proof of (i)⇒(iii), set $D_n = D_{r_n}$, where $r_n = 1 - 2^{-n}$ for $n = 1, 2, 3, \cdots$. Then $D_1 \subset D_2 \subset D_3 \subset \cdots$ and $D = \bigcup_{n=1}^{\infty} D_n$. As a result, $F = \widehat{\mathbb{C}} \sim D = \bigcap_{n=1}^{\infty} F_n$, in which $F_n = \widehat{\mathbb{C}} \sim D_n$. Each F_n is a compact, connected set in $\widehat{\mathbb{C}}$, and $F_1 \supset F_2 \supset F_3 \supset \cdots$. Suppose that F were disconnected. Then there would exist disjoint open sets U and V in $\widehat{\mathbb{C}}$ with $U \cap F \neq \phi$, $V \cap F \neq \phi$, and F contained in $U \cup V$. The sequence $F_1 \sim (U \cup V)$, $F_2 \sim (U \cup V)$, $F_3 \sim (U \cup V), \cdots$ would be a non-increasing sequence of compact sets in $\widehat{\mathbb{C}}$ with empty intersection. In view of Cantor's theorem (Theorem II.4.5) — or, to be more precise, its analogue in $\widehat{\mathbb{C}}$ — $F_n \sim (U \cup V) = \phi$ would have to be true for some n. For such an n we would find F_n contained in the union of two disjoint open sets, both of which intersect F_n. At this point we would run into a contradiction, for we already know that F_n is connected. The contradiction forces us to conclude that $F = \widehat{\mathbb{C}} \sim D$ is, after all, a connected set. ∎

We have confined our discussion in this section to the conformal mapping of simply connected domains. The existence of conformal mappings between domains of more complicated topological structure is a trickier business and will not be considered in this book.

4 The Carathéodory-Osgood Theorem

4.1 Topological Preliminaries

Suppose that D and D' are simply connected plane domains, neither of which is the entire complex plane. We are now in possession of the knowledge that conformal mappings f from D onto D' exist. Moreover, the function theory we have so far developed secures for us a reasonably good hold on the local geometric and analytical properties of such a mapping near any point of D. Left wide open to speculation is its behavior at the boundary of D: What can be said about $f(\zeta)$ as ζ approaches a point z of ∂D? Framed in such generality this question has no simple answer. Indeed, there are many facets of the problem that are not yet fully understood and are the subjects of continuing research. There is, however, one special set of circumstances in which, from a strictly topological viewpoint, the matter can be completely settled, thanks to a beautiful theorem discovered independently by Constantin Carathéodory (1873-1950) and William F. Osgood (1884-1943): if each of the domains D and D' is the inside of a plane Jordan curve, then f extends in a unique way to a continuous and univalent function $\tilde{f}$ that maps $\overline{D}$ onto $\overline{D}'$. It turns out that under these conditions $\tilde{f}^{-1}: \overline{D}' \to \overline{D}$ will automatically be continuous as well, making $\tilde{f}$ a homeomorphism of $\overline{D}$ onto $\overline{D}'$. (Recall: a function $h: A \to B$ between sets A and B in the complex plane — or, more generally, in the extended complex plane — is called a *homeomorphism of A onto B* if h is univalent, if its range is B, and if both h and h^{-1} are continuous.) This section is devoted to a proof, modulo one technical detail, of the Carathéodory-Osgood theorem and to a discussion of several of the theorem's many consequences. The proof that we give relies on estimates for an important conformal invariant, of interest in its own right, which is introduced in Subsection 4.4.

As we did in the case of the Riemann mapping theorem, we shall "normalize" our discussion of the extension problem, this time by considering initially only conformal mappings $f: \Delta \to \mathbb{C}$, where $\Delta = \Delta(0,1)$. Since we want to allow for the possibility of $f(\Delta)$ being unbounded, we shall work in the extended plane $\widehat{\mathbb{C}}$ when attempting to extend f. (We remind the reader that the notations $\widehat{A}$ and $\widehat{\partial} A$ indicate the closure in $\widehat{\mathbb{C}}$ and boundary in $\widehat{\mathbb{C}}$, respectively, of a subset A of the extended plane. In particular, for a domain D in the finite plane, $\widehat{D} = \overline{D}$ and $\widehat{\partial} D = \partial D$ if D is bounded, whereas $\widehat{D} = \overline{D} \cup \{\infty\}$ and $\widehat{\partial} D = \partial D \cup \{\infty\}$ in the unbounded case.) As a beginning step we summarize in the form of a lemma some essential topological background information.

Lemma 4.1. *Let $\Delta = \Delta(0,1)$, let $f: \Delta \to \mathbb{C}$ be a continuous function, and let $D = f(\Delta)$. Assume that $\lim_{\zeta \to z} f(\zeta)$ exists in $\widehat{\mathbb{C}}$ for every point z of $\partial \Delta$. Then the function $\tilde{f}: \overline{\Delta} \to \widehat{\mathbb{C}}$ defined by*

$$\tilde{f}(z) = \begin{cases} f(z) & \text{if } z \in \Delta\,, \\ \lim\limits_{\zeta \to z} f(\zeta) & \text{if } z \in \partial\Delta\,, \end{cases}$$

is the unique extension of f to a continuous function with domain-set $\overline{\Delta}$. Furthermore, $\tilde{f}(\overline{\Delta}) = \widehat{D}$. If, in addition, $\tilde{f}$ is univalent, then $\tilde{f}$ is a homeomorphism of $\overline{\Delta}$ onto $\widehat{D}$.

Proof. The relation $\tilde{f}(z) = \lim_{\zeta \to z} f(\zeta)$ is actually true for every point z

of $\overline{\Delta}$, by definition if z lies on $\partial\Delta$ and by the assumed continuity of f if z belongs to Δ. What needs to be checked first is that $\widetilde{f}$, which is obviously continuous at all points of Δ, is continuous at every point of $\partial\Delta$. Fix such a point, say z_0. We consider an arbitrary sequence $\langle z_n \rangle$ in $\overline{\Delta}$ with the property that $z_n \to z_0$ and verify that $\widetilde{f}(z_n) \to \widetilde{f}(z_0)$. Since $\widetilde{f}(z_n) = \lim_{\zeta \to z_n} f(\zeta)$, we are at liberty to choose a point ζ_n in Δ satisfying $|\zeta_n - z_n| < n^{-1}$ for which it is also true that $|f(\zeta_n) - \widetilde{f}(z_n)| < n^{-1}$ if $\widetilde{f}(z_n) \neq \infty$ and $|f(\zeta_n)| > n$ if $\widetilde{f}(z_n) = \infty$. Then

$$|\zeta_n - z_0| \leq |\zeta_n - z_n| + |z_n - z_0| \leq \frac{1}{n} + |z_n - z_0| \to 0$$

as $n \to \infty$; i.e., $\zeta_n \to z_0$. Since $\langle \zeta_n \rangle$ is a sequence in Δ, it follows from the definition of $\widetilde{f}(z_0)$ that $f(\zeta_n) \to \widetilde{f}(z_0)$. If $\widetilde{f}(z_0) \neq \infty$, it must be so that $|f(\zeta_n)| < n$ — hence, that $\widetilde{f}(z_n) \neq \infty$ — once n is suitably large. For large n, therefore, we have

$$|\widetilde{f}(z_n) - \widetilde{f}(z_0)| \leq |\widetilde{f}(z_n) - f(\zeta_n)| + |f(\zeta_n) - \widetilde{f}(z_0)| \leq \frac{1}{n} + |f(\zeta_n) - \widetilde{f}(z_0)| \to 0\,,$$

showing that $\widetilde{f}(z_n) \to \widetilde{f}(z_0)$ in the case $\widetilde{f}(z_0) \neq \infty$. If $\widetilde{f}(z_0) = \infty$, then by construction $|f(\zeta_n)| \to \infty$. Either $\widetilde{f}(z_n) = \infty$ or

$$|\widetilde{f}(z_n)| \geq |f(\zeta_n)| - \frac{1}{n}\,,$$

which facts make it evident that $\widetilde{f}(z_n) \to \infty = \widetilde{f}(z_0)$ in this case, too. Having checked the continuity of $\widetilde{f}$ at all points where it could conceivably be in doubt, we can assert that $\widetilde{f}$ is a continuous mapping of $\overline{\Delta}$ into $\widehat{\mathbb{C}}$. The very definition of continuity makes it clear that $\widetilde{f}$ is the only possible continuous function from $\overline{\Delta}$ into $\widehat{\mathbb{C}}$ which agrees with f in Δ.

We next prove that $\widetilde{f}(\overline{\Delta}) = \widehat{D}$. It follows almost immediately from the definition of $\widetilde{f}$ that $\widetilde{f}(\overline{\Delta})$ is a subset of $\widehat{D}$. As to the opposite containment, let w_0 belong to the set $\widehat{D}$. We must produce a point z_0 of $\overline{\Delta}$ for which $\widetilde{f}(z_0) = w_0$. To this end, we pick a sequence $\langle w_n \rangle$ in D such that $w_n \to w_0$. Because $D = f(\Delta)$, we are then entitled to choose a point z_n in Δ with $w_n = f(z_n) = \widetilde{f}(z_n)$. Because the set $\overline{\Delta}$ is compact, we can extract from the sequence $\langle z_n \rangle$ a subsequence $\langle z_{n_k} \rangle$ with the property that $z_{n_k} \to z_0$, some point of $\overline{\Delta}$. Finally, the continuity of $\widetilde{f}$ at z_0 gives

$$w_0 = \lim_{k\to\infty} w_{n_k} = \lim_{k\to\infty} \widetilde{f}(z_{n_k}) = \widetilde{f}(z_0)\,,$$

as desired. Thus, $\widetilde{f}(\overline{\Delta}) = \widehat{D}$.

Assuming that $\widetilde{f}$ is univalent, we can go further and state that its inverse $\widetilde{f}^{-1} : \widehat{D} \to \overline{\Delta}$ is another continuous function. If $\widetilde{f}^{-1}$ fails to be continuous at some point w_0 of $\widehat{D}$, then there is a sequence $\langle w_n \rangle$ in $\widehat{D}$ such that $w_n \to w_0$, whereas $z_n = \tilde{f}^{-1}(w_n) \not\to z_0 = \tilde{f}^{-1}(w_0)$. The last fact and the compactness of $\overline{\Delta}$ imply that $\langle z_n \rangle$ has some subsequence $\langle z_{n_k} \rangle$ with the property that $z_{n_k} \to z_0'$, where z_0' is a point of $\overline{\Delta}$ different from z_0. On the other hand, the continuity of $\tilde{f}$ tells us that

$$\tilde{f}(z_0') = \lim_{k\to\infty} \tilde{f}(z_{n_k}) = \lim_{k\to\infty} w_{n_k} = w_0 = \tilde{f}(z_0) ,$$

contradicting the univalence of $\tilde{f}$. Accordingly, if the extension $\tilde{f}$ is univalent, then it is a homeomorphism of $\overline{\Delta}$ onto $\hat{D}$. ∎

4.2 Double Integrals

In anticipation of the approaching material dealing with the conformal modulus of a path family, we review some basic facts about integrals of the type $\iint_S \rho(z)\,dxdy$, in which S is a subset of $\mathbb{C}$ and ρ is a real-valued function that is continuous on S. In multi-variable calculus courses it is typically shown that such an integral is meaningful when S has the form $S = \overline{D}$, where D is a bounded plane domain with the property that ∂D can be expressed as the union of a finite number of regular arcs. For want of a better term we call a set S of this kind a "standard integration region." For instance, a closed rectangle $S = \{z : a \le x \le b, c \le y \le d\}$ is a standard integration region and, in this instance, the double integral can be computed by means of iterated integrations:

$$\begin{aligned}\iint_S \rho(z)\,dxdy &= \int_a^b \left\{ \int_c^d \rho(x+iy)\,dy \right\} dx \\ &= \int_c^d \left\{ \int_a^b \rho(x+iy)\,dx \right\} dy .\end{aligned} \tag{9.42}$$

A second example is a set of the form $S = \{re^{i\theta} : a \le r \le b, \alpha \le \theta \le \beta\}$ with $0 \le a < b < \infty$ and $0 < \beta - \alpha \le 2\pi$. Here it is often convenient to make a switch to polar coordinates for the evaluation:

$$\begin{aligned}\iint_S \rho(z)\,dxdy &= \int_a^b \left\{ \int_\alpha^\beta \rho(re^{i\theta})\,d\theta \right\} rdr \\ &= \int_\alpha^\beta \left\{ \int_a^b \rho(re^{i\theta})\,rdr \right\} d\theta .\end{aligned} \tag{9.43}$$

This shift to polar coordinates is a special case of a more general "change of variables" formula: if $f: D \to \mathbb{C}$ is a diffeomorphism, if S is a standard integration region that is contained in D, and if $\rho: f(S) \to \mathbb{R}$ is a continuous function, then

$$\iint_{f(S)} \rho(w)\, dudv = \iint_{S} \rho[f(z)]|J_f(z)|\, dxdy\ . \tag{9.44}$$

We notice especially that

$$\iint_{f(S)} \rho(w)\, dudv = \iint_{S} \rho[f(z)]|f'(z)|^2\, dxdy \tag{9.45}$$

when f is a conformal mapping. Implicit in (9.44) and (9.45) is the fact that $f(S)$ is itself a standard integration region. To avoid confusion, we have indicated the variables of integration in $f(S)$ by u and v. Thus one is to think of making the change of variable $w = f(z)$ in the integral $\iint_{f(S)} \rho(w)\, dudv$, thereby transforming it into $\iint_S \rho[f(z)]|J_f(z)|\, dxdy$.

We shall also need to deal with "improper" double integrals of the sort $\iint_G \rho(z)\, dxdy$, where G is a "general integration region" in the complex plane — we use this expression to describe a set G that can be written in the form $G = \cup_{n=1}^{\infty} S_n$, in which $S_1 \subset S_2 \subset S_3 \subset \cdots$ is a sequence of standard integration regions — and where $\rho\colon G \to [0,\infty)$ is a continuous function. (A plane domain G is an example of a general integration region. Of course, any standard integration region also qualifies as a general integration region.) The definition of the integral in this situation reads

$$\iint_{G} \rho(z)\, dxdy = \sup\left\{ \iint_{S} \rho(z)\, dxdy\colon S \text{ stand. integ. reg.}, S \subset G \right\}. \tag{9.46}$$

We must allow here for the possibility that $\iint_G \rho(z)\, dxdy = \infty$, a state of affairs which arises when the set of numbers on the right-hand side of (9.46) is unbounded. The change of variables formula carries over to this setting and permits one to conclude, for instance, that

$$\iint_{f(G)} \rho(w)\, dudv = \iint_{G} \rho[f(z)]|f'(z)|^2\, dxdy \tag{9.47}$$

in the event $f\colon D \to \mathbb{C}$ is a conformal mapping of a domain D which contains G and $\rho\colon f(G) \to [0,\infty)$ is a continuous function.

4.3 Conformal Modulus

Let G be a general integration region. A continuous function $\rho\colon G \to [0,\infty)$ will be called a *density* in G. We can regard such a function as a vehicle for setting up in G a new system for measuring lengths and areas, as follows: the "ρ-length" $\ell_\rho(\gamma)$ of piecewise smooth path γ in G is given by

$$\ell_\rho(\gamma) = \int_\gamma \rho(z)|dz| ,$$

while the "ρ-area" $A_\rho(G')$ of any general integration region G' that is a subset of G is defined by

$$A_\rho(G') = \iint_{G'} [\rho(z)]^2 \, dx dy .$$

(N.B. When $\rho(z) = 1$ throughout G these reduce to the ordinary length $\ell(\gamma)$ of γ and area $A(G')$ of G'.) One is reminded of a situation in physics where G might serve as a model for a thin plate composed of some non-homogeneous substance and ρ would specify the "linear mass density" (= mass per unit length) at points of the plate. In this analogy, $\ell_\rho(\gamma)$ would represent the mass of a "wire" formed from the trajectory of γ and $A_\rho(G')$ would give the mass of the "subplate" G' of G.

We are going to consider configurations of sets E, F, and G, where E and F are non-empty, disjoint subsets of an integration region G (Figure 26). The notation $[E, F\colon G]$ is used to symbolize a configuration of this

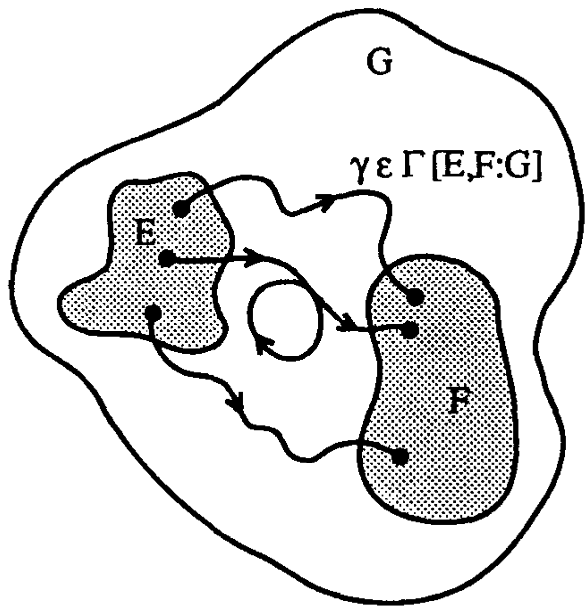


Figure 26.

kind. With each such configuration we shall associate an extended real number $M[E, F\colon G]$, the *conformal modulus* of $[E, F\colon G]$, in such a way that this quantity remains invariant under conformal mapping. Our method for doing this was introduced by Lars Ahlfors and Arne Beurling (1905-1986) in their landmark paper "Conformal invariants and function-theoretic null sets" in *Acta Mathematica*, Vol. 83, 1950. The invariant they defined there has subsequently developed into one of the major theoretical tools at the disposal of researchers in complex analysis. Let $\Gamma[E,F\colon G]$ designate the family of all piecewise smooth paths in G having initial point in E and terminal point in F. To define $M[E,F\colon G]$ we look at densities ρ in G with the extra feature that $\ell_\rho(\gamma) \geq 1$ for every path γ of $\Gamma[E,F\colon G]$. Such a function ρ is termed an ***admissible density*** for the configuration $[E,F\colon G]$. The collection of all densities in G that are admissible for $[E,F\colon G]$ is denoted by $\mathrm{Adm}[E,F\colon G]$. (*Warning*: there are situations in which no admissible densities exist; i.e., for which $\mathrm{Adm}[E,F\colon G] = \phi$.) We define

$$M[E,F\colon G] = \inf\{A_\rho(G) : \rho \in \mathrm{Adm}[E,F\colon G]\} \tag{9.48}$$

if $\mathrm{Adm}[E,F\colon G] \neq \phi$, the infimum being taken in the extended real interval $[0,\infty]$, and set $M[E,F\colon G] = \infty$ otherwise. In plain words, $M[E,F\colon G]$ is the smallest area a density ρ could assign to G, given that ρ is constrained to assign a length no less than one to every path in the family $\Gamma[E,F\colon G]$. (N.B. In the paper cited Ahlfors and Beurling found it more convenient to work with $\lambda[E,F\colon G] = 1/M[E,F\colon G]$, a quantity they dubbed the ***extremal length*** of $\Gamma[E,F\colon G]$, than with $M[E,F\colon G]$. The number $\lambda[E,F\colon G]$ is also known as the ***extremal distance between E and F in G***.) Since $-\gamma$ belongs to $\Gamma[F,E\colon G]$ if and only if γ belongs to $\Gamma[E,F\colon G]$ and since $\int_{-\gamma}\rho(z)\,|dz| = \int_\gamma \rho(z)\,|dz|$, it is clear that $M[E,F\colon G] = M[F,E\colon G]$.

It is not easy to gather from this admittedly abstruse definition what information the number $M[E,F\colon G]$ might encode, to say nothing of the relevance of that information for the theory of analytic functions. Nor shall we make any serious effort to explain this definition by delving into the ideas that underlie and motivate it. Instead, we hope that by seeing the concept in action in the proof of the Carathéodory-Osgood theorem the reader will gain an appreciation for the power and utility of this invariant, even though many aspects of it remain more than a little hazy. Perhaps the following example, which illustrates some common techniques for computing and estimating conformal moduli, may serve to shed a bit of light on the definition itself.

EXAMPLE 4.1. For $a > 0$ and $b > 0$ let G be the closed rectangle with vertices $0, a, a+ib$, and ib (Figure 27). If $E = \{x : 0 \leq x \leq a\}$ and $F = \{x + ib\colon 0 \leq x \leq a\}$, show that $M[E,F\colon G] = a/b$.

We first demonstrate that $A_\rho(G) \geq a/b$ for every density ρ in $\mathrm{Adm}[E,F\colon G]$. It will follow immediately that $M[E,F\colon G] \geq a/b$. Fix such a ρ. For each x satisfying $0 \leq x \leq a$ the path $\gamma_x\colon [0,b] \to \mathbb{C}$ defined by $\gamma_x(y) = x + iy$ is certainly a member of $\Gamma[E,F\colon G]$. By the definition of an admissible density, it must be the case that

$$1 \leq \int_{\gamma_x} \rho(z)|dz| = \int_0^b \rho(x+iy)\,dy\ .$$

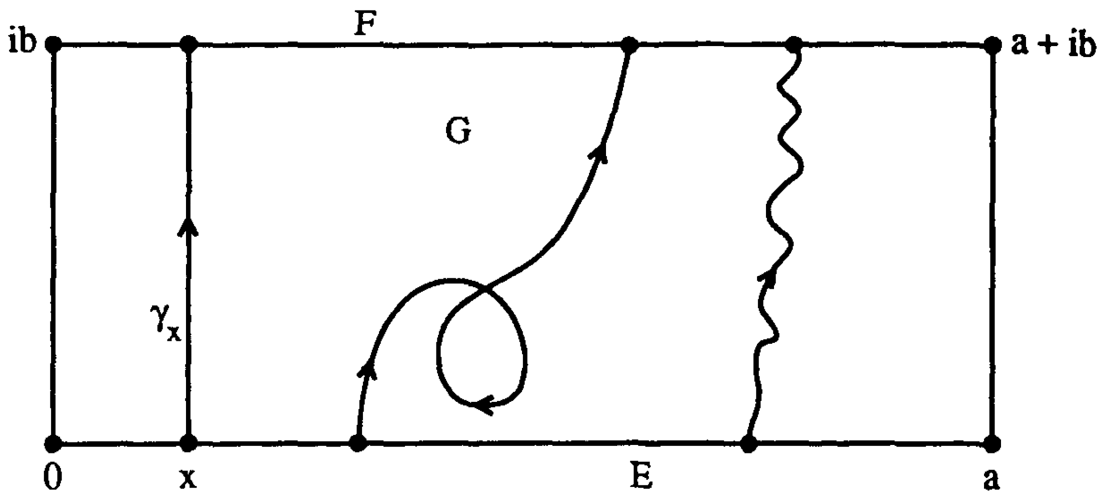


Figure 27.

Consequently,

$$1 \le \left[\int_0^b \rho(x+iy)\,dy\right]^2 \le \left\{\int_0^b [\rho(x+iy)]^2\,dy\right\}\left\{\int_0^b dy\right\}$$

$$= b\int_0^b [\rho(x+iy)]^2\,dy\ .$$

Here we have appealed to a classical inequality for Riemann integrals, the Cauchy-Schwarz inequality:

$$\left[\int_c^d f(y)g(y)\,dy\right]^2 \le \left\{\int_c^d [f(y)]^2\,dy\right\}\left\{\int_c^d [g(y)]^2\,dy\right\}\ , \tag{9.49}$$

if f and g are real-valued functions that are Riemann integrable over $[c,d]$. In our case $f(y) = \rho(x+iy)$ and $g(y) = 1$ on $[0,b]$. We infer that the inequality

$$\frac{1}{b} \le \int_0^b [\rho(x+iy)]^2\,dy$$

is valid for each x in $[0,a]$. As a result of (9.42),

$$\frac{a}{b} = \int_0^a \frac{dx}{b} \le \int_0^a \left\{\int_0^b [\rho(x+iy)]^2\,dy\right\}\,dx$$

$$= \iint_G [\rho(z)]^2\,dxdy = A_\rho(G)\ .$$

Therefore, the lower bound $A_\rho(G) \ge a/b$ is valid for every admissible density ρ — hence, $M[E,F\colon G] \ge a/b$.

To confirm that equality actually holds we consider a particular density $\rho_0\colon G \to [0,\infty)$, the one given by $\rho_0(z) = b^{-1}$. Since $\ell(\gamma) \geq b$ is without a doubt true of any path γ in $\Gamma[E,F\colon G]$, we see that

$$\ell_{\rho_0}(\gamma) = \int_\gamma \rho_0(z)\,|dz| = \int_\gamma \frac{|dz|}{b} = \frac{\ell(\gamma)}{b} \geq 1$$

for each such γ, which places ρ_0 in $\mathrm{Adm}[E,F\colon G]$. This means that

$$M[E,F\colon G] \leq A_{\rho_0}(G) = \iint_G [\rho_0(z)]^2\,dxdy = \iint_G \frac{dxdy}{b^2} = \frac{ab}{b^2} = \frac{a}{b}\,.$$

Coupled with the inequality $M[E,F\colon G] \geq a/b$, this gives $M[E,F\colon G] = a/b$.

The property that ultimately turns $M[E,F\colon G]$ into a useful commodity is its invariance under conformal mapping. The precise formulation of that invariance is the purpose of the next theorem.

Theorem 4.2. *Suppose that D is a domain in the complex plane and that $f\colon D \to \mathbb{C}$ is a conformal mapping. Then*

$$M[E,F\colon G] = M[f(E),f(F)\colon f(G)] \tag{9.50}$$

for any configuration $[E,F\colon G]$ with G contained in D.

Proof. Write $M = M[E,F\colon G]$ and $M' = M[f(E),f(F)\colon f(G)]$. It is enough to verify that $M \leq M'$, for the reverse inequality can be deduced from this one simply by applying it to f^{-1}. We may further assume that $M' < \infty$, since the inequality $M \leq M'$ holds trivially otherwise. Let $\widetilde{\rho}$ be an admissible density for the configuration $[f(E),\ f(F)\colon f(G)]$. We define a density ρ in G by $\rho(z) = \widetilde{\rho}[f(z)]|f'(z)|$. The claim is that ρ constitutes an admissible density for $[E,F\colon G]$. To see this, let $\gamma\colon [a,b] \to \mathbb{C}$ be a path in $\Gamma[E,F\colon G]$. Then $\beta = f\circ\gamma$ is a path belonging to the family $\Gamma[f(E),f(F)\colon f(G)]$. Since $\widetilde{\rho}$ is a member of $\mathrm{Adm}[f(E),f(F)\colon f(G)]$, we obtain

$$1 \leq \int_\beta \widetilde{\rho}(z)\,|dz| = \int_a^b \widetilde{\rho}[\beta(t)]|\dot{\beta}(t)|\,dt = \int_a^b \widetilde{\rho}\{f[\gamma(t)]\}|f'[\gamma(t)]||\dot{\gamma}(t)|\,dt$$

$$= \int_a^b \rho[\gamma(t)]|\dot{\gamma}(t)|\,dt = \int_\gamma \rho(z)\,|dz|\,,$$

which demonstrates that ρ belongs to $\mathrm{Adm}[E,F\colon G]$. Accordingly, M does not exceed $A_\rho(G)$. The change of variable formula (9.47) for double integrals then leads to

$$M \leq A_\rho(G) = \iint_G [\rho(z)]^2\,dxdy = \iint_G \{\widetilde{\rho}[f(z)]\}^2|f'(z)|^2\,dxdy$$

$$= \iint_{f(G)} [\tilde{\rho}(w)]^2 \, du dv = A_{\tilde{\rho}}[f(G)] \, .$$

Because $\tilde{\rho}$ was an arbitrary admissible density for $[f(E), f(F): f(G)]$, we are able to conclude that $M \leq M'$, as desired. ∎

In every application that we intend to make of Theorem 4.2 in this book we shall have $G = D$, but the reader must not get the impression that the utility of the theorem is limited to that case.

In general it is quite difficult to determine the quantity $M[E, F: G]$ exactly. One must ordinarily make do with upper or lower bounds for the modulus. Fortunately, such bounds are not exceedingly hard to come by and are all that is needed for many applications of the invariant. This point is dramatized by the proof we give of the Carathéodory-Osgood theorem, the crux of which resides in the modulus estimates afforded by the following sequence of lemmas. The first of these describes a situation in which we are assured that a certain modulus cannot be unduly small.

Lemma 4.3. *Let $\Delta = \Delta(0,1)$, let $E = \overline{\Delta}(0, 1/2)$, and let $z_1 = e^{i\theta_1}$ and $z_2 = e^{i\theta_2}$, where $0 \leq \theta_1 < \theta_2 < 2\pi$. Then*

$$(9.51) \qquad M[E, F: \Delta] \geq \min\{\theta_2 - \theta_1, 2\pi - (\theta_2 - \theta_1)\}$$

for every connected set F in $\Delta \sim E$ with the property that both z_1 and z_2 belong to $\overline{F}$ (Figure 28).

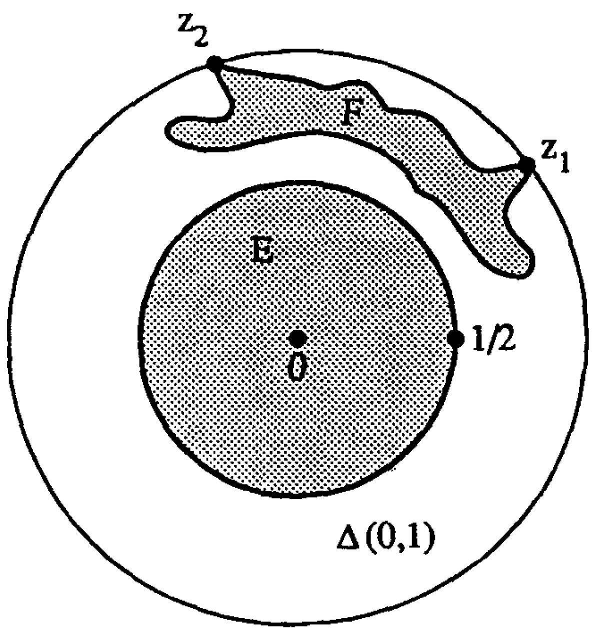


Figure 28.

Proof. By the action of performing a preliminary rotation and using the conformal invariance of the modulus, we reduce the proof to consideration of the case $z_1 = 1$ and $z_2 = e^{i\theta_0}$, where $0 < \theta_0 = \theta_2 - \theta_1 < 2\pi$. Fix a connected set F in $\Delta \sim E$ that has both 1 and $e^{i\theta_0}$ as points of its closure. For $0 \leq \theta < 2\pi$, let $S_\theta = \{re^{i\theta} : 1/2 \leq r < 1\}$. We start the argument with the observation that either F intersects S_θ for every θ in $(0, \theta_0)$ or it intersects S_θ for every θ in $(\theta_0, 2\pi)$. If not, there would exist angles ψ in $(0, \theta_0)$ and φ in $(\theta_0, 2\pi)$ such that both S_ψ and S_φ are disjoint from F. The open set $\Delta \sim (E \cup S_\psi \cup S_\varphi)$ would then consist of two components U and V, one of which has 1 as a boundary point and the other of which has $e^{i\theta_0}$ on its boundary. Due to the location of ψ and φ, U and V would be disjoint open sets which both meet F (remember that 1 and $e^{i\theta_0}$ belong to $\overline{F}$) and which include F in their union, a situation at odds with the hypothesis that F is a connected set. We shall proceed under the assumption that S_θ intersects F for every θ in $(0, \theta_0)$ and derive for $M[E, F : \Delta]$ the estimate

$$M[E, F : \Delta] \geq \theta_0 \ . \tag{9.52}$$

In the other case, similar reasoning would yield

$$M[E, F : \Delta] \geq 2\pi - \theta_0 \ . \tag{9.53}$$

Together (9.52) and (9.53) imply that

$$M[E, F : \Delta] \geq \min\{\theta_0, 2\pi - \theta_0\} \ ,$$

as announced in (9.51).

In verifying (9.52) we may assume that $M[E, F : \Delta] < \infty$, there being nothing to prove should this modulus be infinite. Let ρ be an arbitrary admissible density for $[E, F : \Delta]$. It is our task to demonstrate that $A_\rho(\Delta) \geq \theta_0$. To arrive at this lower bound for $A_\rho(\Delta)$ we first choose for each θ in $(0, \theta_0)$ a number b_θ in $(1/2, 1)$ with the feature that $b_\theta\, e^{i\theta}$ is a point of F. Such a choice is possible because by assumption S_θ does intersect F. The path $\gamma_\theta : [1/2, b_\theta] \to \mathbb{C}$ defined by $\gamma_\theta(r) = re^{i\theta}$ qualifies as a member of $\Gamma[E, F : \Delta]$, with the result that

$$1 \leq \int_{\gamma_\theta} \rho(z)\,|dz| = \int_{1/2}^{b_\theta} \rho(re^{i\theta})\,dr \tag{9.54}$$

by the definition of an admissible density. To certify that $A_\rho(\Delta) \geq \theta_0$ it is enough to check that

$$A_\rho(\Delta) \geq (1 - \epsilon)^2(\beta - \alpha) \tag{9.55}$$

whenever $0 < \alpha < \beta < \theta_0$ and $0 < \epsilon < 1$, for we then get $A_\rho(\Delta) \geq \theta_0$ by letting $\beta \to \theta_0$, $\alpha \to 0$, and $\epsilon \to 0$ on the right-hand side of (9.55). The argument that proves (9.55) is quite technical. We apologize for the technicality, but we find it necessary in order to avoid complications associated with iterated improper Riemann integrals. (A much shorter derivation of (9.52) — indeed, a much cleaner treatment of conformal moduli overall — is available in the context of Lebesgue integration.)

Fix α, β, and ϵ as indicated. We use (9.54) to establish the existence of a number b in $(1/2, 1)$ with the property that the inequality

$$\int_{1/2}^{b} \rho(re^{i\theta})\, dr \geq 1 - \epsilon \tag{9.56}$$

is true for every θ in the interval $[\alpha, \beta]$. Suppose that no such b existed. In particular, taking $b = b_n = 1 - 2^{-n-1}$ for a positive integer n would not meet the requirement, so we could find an angle θ_n in $[\alpha, \beta]$ for which

$$\int_{1/2}^{b_n} \rho(re^{i\theta_n})\, dr < 1 - \epsilon \ .$$

The sequence $\langle \theta_n \rangle$ in $[\alpha, \beta]$ thus generated would have an accumulation point in that interval. Pick one and call it θ. By passing to subsequences and relabeling, if necessary, we could assume that $\theta_n \to \theta$. Now $b_n \to 1$, implying that for large n we would have $b_n \geq b_\theta$, the number appearing in (9.54) for our special θ. Since ρ is non-negative, we see that

$$\int_{1/2}^{b_\theta} \rho(re^{i\theta_n})\, dr \leq \int_{1/2}^{b_n} \rho(re^{i\theta_n})\, dr < 1 - \epsilon$$

would hold for any such n. The continuity of ρ would allow us to infer that $\rho(re^{i\theta_n}) \to \rho(re^{i\theta})$ uniformly in r on the interval $[1/2, b_\theta]$, which would have the consequence that

$$\int_{1/2}^{b_\theta} \rho(re^{i\theta})\, dr = \lim_{n\to\infty} \int_{1/2}^{b_\theta} \rho(re^{i\theta_n})\, dr \leq 1 - \epsilon \ .$$

This is clearly in conflict with (9.54). The only way to avert this contradiction is to accept the existence of b with property (9.56).

We now select and fix b in $(1/2, 1)$ such that (9.56) is valid for every θ in $[\alpha, \beta]$. For each such θ we conclude with the help of the Cauchy-Schwarz inequality (9.49) that

$$(1-\epsilon)^2 \leq \left[\int_{1/2}^{b} \rho(re^{i\theta})\, dr\right]^2 = \left[\int_{1/2}^{b} \rho(re^{i\theta})\, r^{1/2} r^{-1/2}\, dr\right]^2$$

$$\leq \left\{\int_{1/2}^{b} [\rho(re^{i\theta})]^2\, r dr\right\} \left\{\int_{1/2}^{b} \frac{dr}{r}\right\} = \operatorname{Log}(2b) \int_{1/2}^{1} [\rho(re^{i\theta})]^2\, r dr \ .$$

Since $\operatorname{Log}(2b) < \operatorname{Log} 2 < 1$, we can thus be sure that

$$\int_{1/2}^{b} [\rho(re^{i\theta})]^2\, r dr \geq (1-\epsilon)^2 \tag{9.57}$$

when $\alpha \le \theta \le \beta$. Finally, let $S = \{re^{i\theta} : 1/2 \le r \le b, \alpha \le \theta \le \beta\}$, a standard integration region in Δ. After a shift to polar coordinates, we deduce from (9.43) and (9.57) that

$$A_\rho(\Delta) \ge A_\rho(S) = \iint_S [\rho(z)]^2 \, dx dy = \int_\alpha^\beta \left\{ \int_{1/2}^b [\rho(re^{i\theta})]^2 \, r dr \right\} d\theta$$

$$\ge \int_\alpha^\beta (1-\epsilon)^2 \, d\theta = (1-\epsilon)^2(\beta - \alpha) .$$

Inequality (9.55) is thereby confirmed, and with that the proof of the lemma is complete. ∎

The key feature of the lower bound in (9.51) is that it depends only on the points z_1 and z_2 (or, to be more precise, on $|z_1 - z_2|$), not on the connected set F. As stated earlier, there are circumstances under which a configuration $[E, F\colon G]$ can have $M[E, F\colon G] = \infty$. The next lemma presents one set of circumstances where this happens.

Lemma 4.4. *Let $H = \{z : \operatorname{Im} z > 0\}$. Then*

$$M[E, F\colon H] = \infty \tag{9.58}$$

for any disjoint pair of connected sets E and F in H with the property that the origin belongs to both $\overline{E}$ and $\overline{F}$.

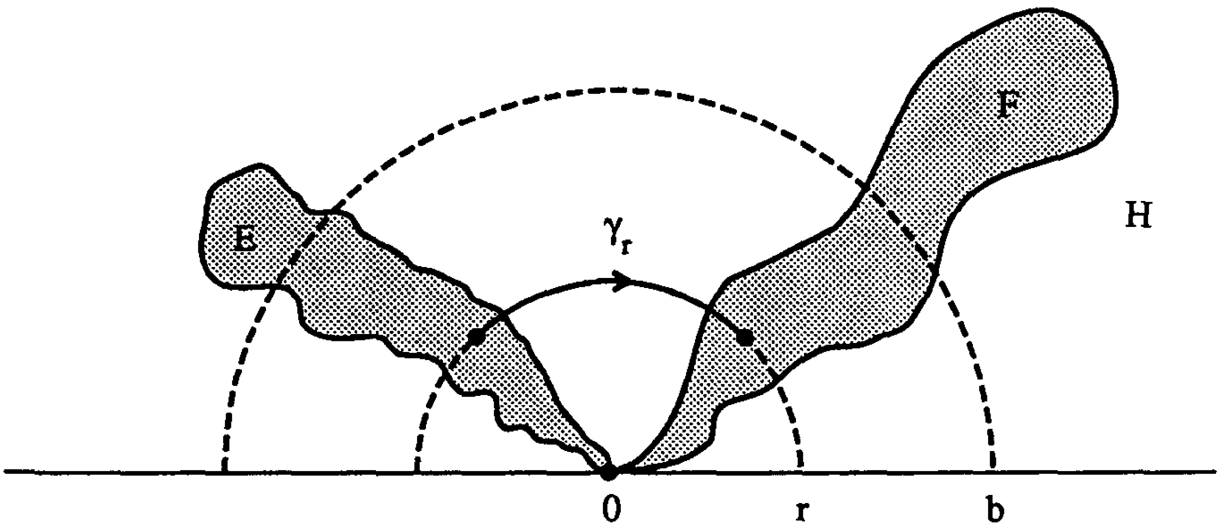


Figure 29.

Proof. Let E and F be sets as indicated (Figure 29), and let ρ be an arbitrary admissible density for the configuration $[E, F\colon H]$. (N.B. If no such densities ρ exist, (9.58) is true by definition.) We must show that $A_\rho(H) = \infty$. We fix a radius $b > 0$ so that both E and F contain points outside the disk $\overline{\Delta}(0, b)$ and prove that the inequality

$$A_\rho(H) \ge \frac{1}{\pi} \log \frac{b}{a} \tag{9.59}$$

holds whenever $0 < a < b$. This is enough to finish the proof of the lemma, for by letting $a \to 0$ in (9.59) we deduce that $A_\rho(H) = \infty$. The initial step in the verification of (9.59) is the remark that for each r in the interval $(0, b]$ the semi-circle $S_r = \{re^{i\theta} : 0 < \theta < \pi\}$ has to intersect both E and F. Suppose, for instance, that S_r did not meet E. Then $U = \Delta(0, r)$ and $V = H \sim \overline{\Delta}(0, r)$ would be disjoint open sets satisfying $E \cap U \neq \phi$ (by reason of the fact that 0 belongs to $\overline{E}$), $E \cap V \neq \phi$ (by the choice of b), and also $E \subset U \cup V$. This would violate the connectedness of E. Similarly, S_r must intersect F.

Given r in $(0, b]$ we are now free to pick angles φ_r and ψ_r from $(0, \pi)$ in such a way that $re^{i\varphi_r}$ lies in E and $re^{i\psi_r}$ in F. Define a path γ_r as follows: if $\varphi_r < \psi_r$ then $\gamma_r(\theta) = re^{i\theta}$ for $\varphi_r \le \theta \le \psi_r$; if $\psi_r < \varphi_r$, then $\gamma_r(\theta) = re^{i(\varphi_r + \psi_r - \theta)}$ for $\psi_r \le \theta \le \varphi_r$. (See Figure 29.) The path γ_r belongs to $\Gamma[E, F \colon H]$. Owing to the admissibility of ρ for $[E, F \colon H]$, we know that

$$1 \le \int_{\gamma_r} \rho(z)\,|dz| = \left| \int_{\varphi_r}^{\psi_r} \rho(re^{i\theta})\, r d\theta \right| .$$

In summary, for each r in $(0, b]$ we can fix a subinterval $[\alpha_r, \beta_r]$ of $(0, \pi)$ — take $[\varphi_r, \psi_r]$ if $\varphi_r < \psi_r$ and $[\psi_r, \varphi_r]$ if $\psi_r < \varphi_r$ — with the property that

$$\int_{\alpha_r}^{\beta_r} \rho(re^{i\theta})\, r d\theta \ge 1 . \tag{9.60}$$

We shall derive (9.59) from (9.60) in much the same way we obtained (9.55) from (9.54) in the proof of Lemma 4.3. Once again a rather technical step is necessary if we are to argue rigorously without at the same time being forced out of the confines of normal Riemann integration.

We fix a in $(0, b)$ and ϵ in $(0, 1)$. We contend that there exists an angle δ in $(0, \pi/2)$ for which

$$\int_{\delta}^{\pi - \delta} \rho(re^{i\theta})\, r d\theta \ge 1 - \epsilon \tag{9.61}$$

whenever $a \le r \le b$. If not, there would exist sequences $\langle \delta_n \rangle$ in $(0, \pi/2)$ and $\langle r_n \rangle$ in $[a, b]$ such that $\delta_n \to 0$, but such that

$$\int_{\delta_n}^{\pi - \delta_n} \rho(r_n e^{i\theta})\, r_n d\theta < 1 - \epsilon .$$

Replacing the original sequences by subsequences if necessary, we could assume that $r_n \to r$, a number in $[a, b]$. For large n the interval $[\delta_n, \pi - \delta_n]$ would contain the interval $[\alpha_r, \beta_r]$ that was chosen for this particular r in order to achieve (9.60). For large n, therefore, we would have

$$\int_{\alpha_r}^{\beta_r} \rho(r_n e^{i\theta})\, r_n d\theta \le \int_{\delta_n}^{\pi - \delta_n} \rho(r_n e^{i\theta})\, r_n d\theta < 1 - \epsilon .$$

Because the convergence of $\rho(r_n e^{i\theta})r_n$ to $\rho(re^{i\theta})r$ would be uniform on $[\alpha_r, \beta_r]$, we could infer that

$$\int_{\alpha_r}^{\beta_r} \rho(re^{i\theta})\, r d\theta = \lim_{n\to\infty} \int_{\alpha_r}^{\beta_r} \rho(r_n e^{i\theta})\, r_n d\theta \leq 1 - \epsilon\ ,$$

which would contradict (9.60). The only option left open is for $(0, \pi/2)$ to contain a δ with property (9.61). We choose such a δ and fix it for the rest of the proof.

Applying the Cauchy-Schwarz inequality to (9.61) we discover that

$$(1-\epsilon)^2 \leq r^2 \left[\int_{\delta}^{\pi-\delta} \rho(re^{i\theta})\, d\theta\right]^2 \leq r^2 \left\{\int_{\delta}^{\pi-\delta} [\rho(re^{i\theta})]^2\, d\theta\right\} \left\{\int_{\delta}^{\pi-\delta} 1\, d\theta\right\}$$

$$= (\pi - 2\delta)\, r^2 \int_{\delta}^{\pi-\delta} [\rho(re^{i\theta})]^2\, d\theta \leq \pi r^2 \int_{\delta}^{\pi-\delta} [\rho(re^{i\theta})]^2\, d\theta$$

for every r in $[a, b]$, so that

$$\text{(9.62)} \qquad \frac{(1-\epsilon)^2}{\pi r} \leq \int_{\delta}^{\pi-\delta} [\rho(re^{i\theta})]^2\, r d\theta$$

for such r. The set $S = \{re^{i\theta} : a \leq r \leq b, \delta \leq \theta \leq \pi - \delta\}$ is a standard integration region in H. By changing to polar coordinates and appealing to (9.43) and (9.62) we are put in a position to conclude that

$$A_\rho(H) \geq A_\rho(S) = \iint_S [\rho(z)]^2\, dxdy = \int_a^b \left\{\int_{\delta}^{\pi-\delta} [\rho(re^{i\theta})]^2\, r d\theta\right\} dr$$

$$\geq \int_a^b \frac{(1-\epsilon)^2\, dr}{\pi r} = \frac{(1-\epsilon)^2}{\pi} \operatorname{Log} \frac{b}{a}\ .$$

Because the estimate

$$A_\rho(H) \geq \frac{(1-\epsilon)^2}{\pi} \operatorname{Log} \frac{b}{a}$$

holds for every ϵ in $(0, 1)$, we can let $\epsilon \to 0$ to obtain (9.59). As observed, this finishes the proof. ∎

Our third lemma introduces one of the rare configurations whose conformal modulus can be calculated exactly and expressed in elementary terms.

Lemma 4.5. *Suppose that z_0 is a point in the complex plane and that $0 < r_0 < r_1 < \infty$. If $E = \overline{\Delta}(z_0, r_0)$ and $F = \mathbb{C} \sim \Delta(z_0, r_1)$, then*

$$\text{(9.63)} \qquad M[E, F\colon \mathbb{C}] = 2\pi \left(\operatorname{Log} \frac{r_1}{r_0}\right)^{-1}\ .$$

Proof. We treat the case $z_0 = 0$. The general case can be reduced to this special one by making a translation and invoking the conformal invariance of the modulus. Let ρ be an arbitrary admissible density for the configuration $[E, F\colon \mathbb{C}]$. We begin the proof by showing that

$$A_\rho(\mathbb{C}) \geq 2\pi \left(\operatorname{Log} \frac{r_1}{r_0}\right)^{-1} . \tag{9.64}$$

For each θ in $[0, 2\pi]$ the path γ_θ defined on $[r_0, r_1]$ by $\gamma_\theta(r) = re^{i\theta}$ is a member of $\Gamma[E, F\colon \mathbb{C}]$, from which it follows that

$$1 \leq \int_{\gamma_\theta} \rho(z)\,|dz| = \int_{r_0}^{r_1} \rho(re^{i\theta})\,dr .$$

An application of (9.49) produces the inequality

$$\begin{aligned} 1 \leq \left[\int_{r_0}^{r_1} \rho(re^{i\theta})\,dr\right]^2 &\leq \left\{\int_{r_0}^{r_1} [\rho(re^{i\theta})]^2\,rdr\right\}\left\{\int_{r_0}^{r_1} \frac{dr}{r}\right\} \\ &= \left(\operatorname{Log} \frac{r_1}{r_0}\right) \int_{r_0}^{r_1} [\rho(re^{i\theta})]^2\,rdr \end{aligned}$$

for every θ in $[0, 2\pi]$. If $S = \{z : r_0 \leq |z| \leq r_1\}$, then

$$\begin{aligned} A_\rho(\mathbb{C}) \geq A_\rho(S) &= \iint_S [\rho(z)]^2\,dxdy = \int_0^{2\pi} \left\{\int_{r_0}^{r_1} [\rho(re^{i\theta})]^2\,rdr\right\} d\theta \\ &\geq \int_0^{2\pi} \left(\operatorname{Log} \frac{r_1}{r_0}\right)^{-1} d\theta = 2\pi \left(\operatorname{Log} \frac{r_1}{r_0}\right)^{-1} , \end{aligned}$$

which establishes (9.64). Because the density ρ in $\operatorname{Adm}[E, F\colon \mathbb{C}]$ was arbitrary, (9.64) implies that

$$M[E, F\colon \mathbb{C}] \geq 2\pi \left(\operatorname{Log} \frac{r_1}{r_0}\right)^{-1} . \tag{9.65}$$

It remains to prove that the inequality sign in (9.65) can also be reversed. We accomplish this by first showing that

$$M[E, F\colon \mathbb{C}] \leq 2\pi \left(\operatorname{Log} \frac{s_1}{s_0}\right) \left(\operatorname{Log} \frac{r_1}{r_0}\right)^{-2} \tag{9.66}$$

whenever $0 < s_0 < r_0$ and $r_1 < s_1 < \infty$. Fix s_0 and s_1 obeying these conditions. Let $h\colon [0, \infty) \to [0, \infty)$ be the continuous function defined as follows: $h(r) = 0$ if $r < s_0$ or $r > s_1$; $h(r) = [\operatorname{Log}(r_1/r_0)]^{-1}$ if $r_0 \leq r \leq r_1$; h is linear on each of the intervals $[s_0, r_0]$ and $[r_1, s_1]$. We can then define a continuous function $\rho\colon \mathbb{C} \to [0,\infty)$ by $\rho(z) = |z|^{-1}h(|z|)$ if $z \neq 0$ and $\rho(0) = 0$. We maintain that ρ is an admissible density for $[E,F\colon\mathbb{C}]$; i.e., we claim that $\int_\gamma \rho(z)\,|dz| \geq 1$ for every path γ belonging to $\Gamma[E,F\colon\mathbb{C}]$. Let $\gamma\colon[a,b] \to \mathbb{C}$ be such a path. Determine numbers c and d by the following prescription:

$$c = \max\{t : t \in [a,b], |\gamma(t)| \leq r_0\} \quad , \quad d = \min\{t : t \in [c,b], |\gamma(t)| \geq r_1\}\ .$$

The fact that γ is a continuous function on $[a,b]$ having $|\gamma(a)| \leq r_0$ and $|\gamma(b)| \geq r_1$ makes certain that c and d are well-defined, that $|\gamma(c)| = r_0$ and $|\gamma(d)| = r_1$, and that $r_0 \leq |\gamma(t)| \leq r_1$ when $c \leq t \leq d$. In particular, $\rho[\gamma(t)] = [|\gamma(t)|\operatorname{Log}(r_1/r_0)]^{-1}$ for t in $[c,d]$. Therefore,

$$\int_\gamma \rho(z)\,|dz| = \int_a^b \rho[\gamma(t)]|\dot\gamma(t)|\,dt \geq \int_c^d \rho[\gamma(t)]|\dot\gamma(t)|\,dt$$

$$= \left(\operatorname{Log}\frac{r_1}{r_0}\right)^{-1} \int_c^d \frac{|\dot\gamma(t)|\,dt}{|\gamma(t)|} = \left(\operatorname{Log}\frac{r_1}{r_0}\right)^{-1} \int_c^d \frac{|\overline{\gamma(t)}\dot\gamma(t)|\,dt}{|\gamma(t)|^2}$$

$$\geq \left(\operatorname{Log}\frac{r_1}{r_0}\right)^{-1} \int_c^d \frac{\operatorname{Re}[\overline{\gamma(t)}\dot\gamma(t)]\,dt}{|\gamma(t)|^2} = \left(\operatorname{Log}\frac{r_1}{r_0}\right)^{-1} \int_c^d \frac{d}{dt}\{\operatorname{Log}|\gamma(t)|\}dt$$

$$= \left(\operatorname{Log}\frac{r_1}{r_0}\right)^{-1} \Big[\operatorname{Log}|\gamma(t)|\Big]_c^d = \left(\operatorname{Log}\frac{r_1}{r_0}\right)^{-1} \left(\operatorname{Log}\frac{r_1}{r_0}\right) = 1\ ,$$

putting ρ in $\operatorname{Adm}[E,F\colon\mathbb{C}]$. Remembering that $\rho(z) = 0$ when $|z| \leq s_0$ or $|z| \geq s_1$ and noticing that $\rho(z) \leq [|z|\operatorname{Log}(r_1/r_0)]^{-1}$ for all z, we deduce:

$$M[E,F\colon\mathbb{C}] \leq A_\rho(\mathbb{C}) = \iint\limits_{\mathbb{C}} [\rho(z)]^2\,dxdy$$

$$= \iint\limits_{s_0 \leq |z| \leq s_1} [\rho(z)]^2\,dxdy \leq \left(\operatorname{Log}\frac{r_1}{r_0}\right)^{-2} \iint\limits_{s_0\leq|z|\leq s_1} \frac{dxdy}{|z|^2}$$

$$= \left(\operatorname{Log}\frac{r_1}{r_0}\right)^{-2} \int_0^{2\pi} \left\{\int_{s_0}^{s_1} \frac{rdr}{r^2}\right\} d\theta = 2\pi \left(\operatorname{Log}\frac{s_1}{s_0}\right)\left(\operatorname{Log}\frac{r_1}{r_0}\right)^{-2}\ .$$

This gives (9.66). Letting $s_0 \to r_0$ and $s_1 \to r_1$ in (9.66) leads to

$$M[E,F\colon\mathbb{C}] \leq 2\pi\left(\operatorname{Log}\frac{r_1}{r_0}\right)^{-1}\ .$$

In tandem with (9.65) the last inequality yields (9.63). ∎

The final lemma in the present series is a monotonicity principle for conformal moduli.

Lemma 4.6. *Suppose that configurations* $[E_1, F_1 : G_1]$ *and* $[E_2, F_2 : G_2]$ *satisfy* $E_1 \subset E_2, F_1 \subset F_2$, *and* $G_1 \subset G_2$. *Then*

$$M[E_1, F_1 : G_1] \leq M[E_2, F_2 : G_2] . \tag{9.67}$$

Proof. We may assume that $M[E_2, F_2 : G_2] < \infty$, for (9.67) is a trivial statement otherwise. If ρ is an admissible density for $[E_2, F_2 : G_2]$, then its restriction to G_1 is obviously an admissible density for $[E_1, F_1 : G_1]$. Consequently,

$$M[E_1, F_1 : G_1] \leq A_\rho(G_1) \leq A_\rho(G_2) .$$

Since ρ in $\text{Adm}[E_2, F_2 : G_2]$ was arbitrary, (9.67) follows. ∎

4.4 Extending Conformal Mappings of the Unit Disk

It is definitely not the case that an arbitrary conformal mapping f of the unit disk $\Delta = \Delta(0,1)$ into $\mathbb{C}$ will extend to a continuous mapping of $\overline{\Delta}$ into $\widehat{\mathbb{C}}$. Such an extension will exist only when the boundary of the domain $D = f(\Delta)$ is suitably regular. In overly simplified terms the existence of an extension is predicated on a negative answer to the question: Does the boundary of D chop any small open disks centered at boundary points of D into infinitely many pieces? We can make this vague statement precise by introducing the following notion: a plane domain D is *finitely connected along its boundary* if corresponding to each point z of $\widehat{\partial} D$ and each $r > 0$ there exists an s in the interval $(0, r)$ such that $D \cap \Delta(z, s)$ intersects at most finitely many components of the open set $D \cap \Delta(z, r)$. As a rule, both the size of s and the number of components of $D \cap \Delta(z, r)$ that $D \cap \Delta(z, s)$ meets may vary wildly with z and r. In the nicest of situations it will be true that for every point z on $\widehat{\partial} D$ and every $r > 0$ a number s in $(0, r)$ corresponding to z and r can be found for which $D \cap \Delta(z, s)$ intersects — hence, is contained in — exactly one component of $D \cap \Delta(z, r)$. When the latter happens we say that D is *locally connected along its boundary*. In Figure 30, D_1 is locally connected along its boundary, D_2 and D_3 are finitely connected along their respective boundaries, but D_4 is not finitely connected along its boundary. For instance, the defining condition is violated by D_4 for small r at the point z_0 indicated. We emphasize that for an unbounded domain D to be finitely (or locally) connected along its boundary the pertinent condition must be satisfied at $z = \infty$, as well as at all points of ∂D. Just to illustrate this, in Figure 30 the quarter plane $D_5 = \{z : x > 0, y > 0\}$ is locally connected along its boundary, and the strip $D_6 = \{z : |y| < 1\}$ is finitely connected along its boundary. The fact that $D_6 \cap \Delta(\infty, r) = \{z \in D_6 : |z| > r^{-1}\}$ has two components whenever $0 < r < 1$ prevents D_6 from being locally connected along its boundary, the required behavior breaking down at ∞.

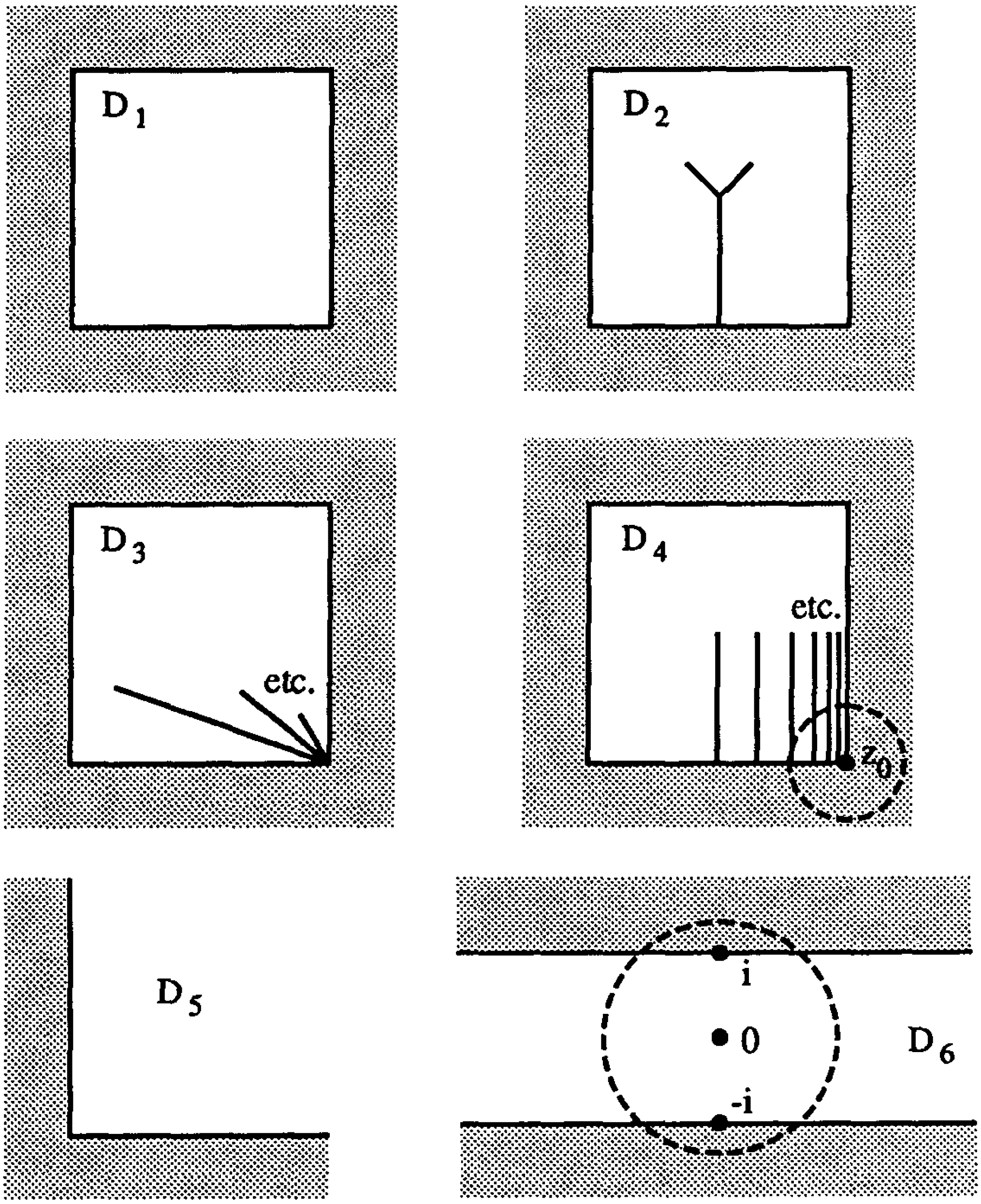


Figure 30.

The preceding definitions bring us to the point where we can establish our first extension result. To the best of the author's knowledge the theorem, as stated here, is due to Jussi Väisälä and Raimo Näkki.

Theorem 4.7. *Let f be a conformal mapping of the disk $\Delta = \Delta(0,1)$ onto a domain D in $\mathbb{C}$. Then f can be extended to a continuous mapping $\tilde{f}$ of $\overline{\Delta}$ onto $\widehat{D}$ if and only if D is finitely connected along its boundary.*

Proof. Assume first that D is finitely connected along its boundary. Owing to Lemma 4.1 the existence of $\tilde{f}$ will be established if it can be shown that $\lim_{\zeta \to z} f(\zeta)$ exists in $\widehat{\mathbb{C}}$ for each point z of $\partial\Delta$. We argue to this conclusion indirectly; i.e., we suppose that $\lim_{\zeta \to z_0} f(\zeta)$ fails to exist in $\widehat{\mathbb{C}}$ for some point z_0 of $\partial\Delta$ and derive a contradiction. Choose a sequence $\langle z_n \rangle$ in Δ such that $z_n \to z_0$. If $w_n = f(z_n)$, then due to the compactness of $\widehat{D}$ we may assume, possibly after passing to a subsequence and doing some relabeling, that $w_n \to w_0$, a point of $\widehat{D}$. In fact, w_0 must belong to $\widehat{\partial} D$. (Otherwise w_0 would be a point of D, $f^{-1}(w_0)$ would lie in Δ, and the continuity of f^{-1} at w_0 would lead to an immediate contradiction — namely, that

$$z_0 = \lim_{n\to\infty} z_n = \lim_{n\to\infty} f^{-1}(w_n) = f^{-1}(w_0) \in \Delta \text{ .)}$$

The assumption that $\lim_{\zeta\to z_0} f(\zeta)$ fails to exist means there has to be a second sequence $\langle z_n' \rangle$ in D such that $z_n' \to z_0$, yet $w_n' = f(z_n') \not\to w_0$. Once again passage to an appropriate subsequence allows us to assume that $w_n' \to w_0'$, where w_0' lies on $\widehat{\partial} D$ and $w_0' \neq w_0$. We select and fix $r > 0$ for which the disks $\overline{\Delta}(w_0, 2r)$ and $\overline{\Delta}(w_0', 2r)$ are disjoint.

Because D is finitely connected along its boundary, it is possible to pick an s in $(0, r)$ with the property that $D \cap \Delta(w_0, s)$ meets only a finite number of components of $D \cap \Delta(w_0, r)$. Since w_n lies in $D \cap \Delta(w_0, s)$ for all large n, we infer that at least one of those finitely many components must contain w_n for infinitely many values of n. Stated differently, it is possible to extract a subsequence $\langle w_{n_k} \rangle$ from $\langle w_n \rangle$ in such a way that w_{n_k} belongs to a fixed component E' of $D \cap \Delta(w_0, r)$ for every k. We can likewise single out a component F' of $D \cap \Delta(w_0', r)$ and a subsequence $\langle w_{m_k}' \rangle$ of $\langle w_n' \rangle$ such that w_{m_k}' is a point of F' for every k. Note that $E = f^{-1}(E')$ and $F = f^{-1}(F')$ are then disjoint connected sets in Δ.

We now examine the configurations $[E', F' \colon D]$ and $[E, F \colon \Delta]$. At least one of the points w_0 or w_0' must be a finite point — say $w_0 \neq \infty$. By construction E' is contained in $\widetilde{E} = \overline{\Delta}(w_0, r)$ and F' in $\widetilde{F} = \mathbb{C} \sim \Delta(w_0, 2r)$. According to Lemmas 4.6 and 4.5

$$M[E', F' \colon D] \leq M[\widetilde{E}, \widetilde{F} \colon \mathbb{C}] = \frac{2\pi}{\operatorname{Log} 2} < \infty \text{ .}$$

Next, $z_{n_k} = f^{-1}(w_{n_k})$ is an element of E for all k. Since $z_{n_k} \to z_0$ it follows that z_0 is a point of $\overline{E}$. For similar reasons z_0 lies in $\overline{F}$. Let h be a Möbius transformation that maps Δ to $H = \{z : \operatorname{Im} z > 0\}$ and sends z_0 to 0. The sets $h(E)$ and $h(F)$ are disjoint connected subsets of H having the origin as a common point of their closures. Theorem 4.2 and Lemma 4.4 tell us that

$$M[E, F \colon \Delta] = M[h(E), h(F) \colon H] = \infty \text{ .}$$

But then $M[E, F \colon \Delta] \neq M[E', F' \colon D]$, which contravenes the conformal invariance of the modulus. This contradiction arose from the assumption that $\lim_{\zeta\to z_0} f(\zeta)$ failed to exist in $\widehat{\mathbb{C}}$. Thus, when all is said and done, the limit in question must exist. The existence of the extension $\widetilde{f}$ is now promised by Lemma 4.1.

The proof of the converse is also by contradiction. We suppose that the extension $\widetilde{f}$ exists, but that D is not finitely connected along its boundary. There must then exist a point w_0 of $\widehat{\partial} D$ and a number $r > 0$ such that for each s in $(0, r)$ the set $D \cap \Delta(w_0, s)$ intersects infinitely many components of $D \cap \Delta(w_0, r)$. It follows easily that we can manufacture a sequence $\langle w_n \rangle$ in D with the properties that $w_n \to w_0$ and that for $n \neq m$ the points w_n and w_m lie in different components of $D \cap \Delta(w_0, r)$. Let $z_n = f^{-1}(w_n)$. Passing to a subsequence, if need be, we may assume that $z_n \to z_0$, necessarily a point of $\partial\Delta$. Since $\tilde{f}$ is continuous at z_0 and has

$$\tilde{f}(z_0) = \lim_{n \to \infty} f(z_n) = \lim_{n \to \infty} w_n = w_0 ,$$

there exists a $\delta > 0$ for which it is true that $\tilde{f}[\overline{\Delta} \cap \Delta(z_0, \delta)]$ is contained in $\Delta(w_0, r)$. The set $C = \Delta \cap \Delta(z_0, \delta)$ is connected, and $\tilde{f}(C) = f(C)$ is contained in $D \cap \Delta(w_0, r)$. Itself a connected set, $f(C)$ must therefore be a subset of some component of $D \cap \Delta(w_0, r)$. On the other hand, z_n is a point of C for all large n, which implies that $f(C)$ contains w_n for all large n, in disagreement with the fact that the points w_n all lie in different components of $D \cap \Delta(w_0, r)$. This contradiction shows that D has to be finitely connected along its boundary. ∎

Theorem 4.7 informs us, for example, that a conformal mapping of $\Delta = \Delta(0, 1)$ onto any one of the domains D_1, D_2, D_3, D_5, or D_6 in Figure 30 extends continuously to $\overline{\Delta}$, whereas a conformal mapping of Δ onto D_4 definitely does not admit a continuous extension to $\overline{\Delta}$.

Relying on slightly more sophisticated topological information, Carathéodory was able to prove the following variant of Theorem 4.7: *a conformal mapping f of Δ onto D can be extended to a continuous mapping $\tilde{f}$ of $\overline{\Delta}$ onto $\hat{D}$ if and only if $\hat{\partial} D$ is a locally connected subset of $\hat{\mathbb{C}}$.* (To say that a set A in $\hat{\mathbb{C}}$ is *locally connected* means this: corresponding to any point z of A and any $r > 0$ there exist an s in $(0, r)$ and a connected set C satisfying $A \cap \Delta(z, s) \subset C \subset A \cap \Delta(z, r)$.) A proof of this result of Carathéodory's can be found in the book of Pommerenke cited in the introduction to this chapter.

The analogue of Theorem 4.7 in which we insist that the extension $\tilde{f}$ be a homeomorphism merely replaces the requirement that D be finitely connected along its boundary with the more stringent demand that D be locally connected along its boundary. (*Warning*: Do not confuse "D is locally connected along its boundary" with "$\hat{\partial} D$ is locally connected.")

Theorem 4.8. *Let f be a conformal mapping of the disk $\Delta = \Delta(0, 1)$ onto a domain D in $\mathbb{C}$. Then f can be extended to a homeomorphism $\tilde{f}$ of $\overline{\Delta}$ onto $\hat{D}$ if and only if D is locally connected along its boundary.*

Proof. Attacking the sufficiency first, we assume that D is locally connected along its boundary. Theorem 4.7 certifies that f can be extended to a continuous function $\tilde{f}$ mapping $\overline{\Delta}$ onto $\hat{D}$. If we are able to demonstrate that $\tilde{f}$ is univalent, Lemma 4.1 will insure that $\tilde{f}$ is a homeomorphism. We show, in fact, that a breakdown in univalence on the part of $\tilde{f}$ leads to a contradiction. Since f is univalent and since $\tilde{f}(\partial\Delta)$ is contained in $\hat{\partial}D$, the only way that $\tilde{f}$ can fail to be univalent is for $\tilde{f}(z_0)$ to coincide with $\tilde{f}(z_0')$ for a pair of distinct points z_0 and z_0' of $\partial\Delta$. Suppose this occurs for $z_0 = e^{i\theta_1}$ and $z_0' = e^{i\theta_2}$, where $0 \le \theta_1 < \theta_2 < 2\pi$. Write $w_0 = \tilde{f}(z_0) = \tilde{f}(z_0')$, $E = \overline{\Delta}(0, 1/2)$, $E' = f(E)$, and $m = \min\{\theta_2 - \theta_1, 2\pi - (\theta_2 - \theta_1)\}$. The set E' is a compact set in D and w_0, a point of $\hat{\partial}D$, does not belong to E'. We can therefore choose $r_1 > 0$ with the property that E' and $\overline{\Delta}(w_0, r_1)$ are disjoint. Next, we pick r_0 in $(0, r_1)$ for which $2\pi[\mathrm{Log}(r_1/r_0)]^{-1} < m$. Finally, exploiting the fact that D is locally connected along its boundary, we select s in $(0, r_0)$ such that $D \cap \Delta(w_0, s)$ lies in a component F' of $D \cap \Delta(w_0, r_0)$. If $w_0 \ne \infty$, the set E' is contained in $\tilde{E} = \mathbb{C} \sim \Delta(w_0, r_1)$ and F' lies in $\tilde{F} = \overline{\Delta}(w_0, r_0)$; if $w_0 = \infty$, on the other hand, E' and F' are subsets of $\tilde{E} = \overline{\Delta}(0, r_1^{-1})$ and $\tilde{F} = \mathbb{C} \sim \Delta(0, r_0^{-1})$, respectively. In either case, Lemmas 4.6 and 4.5 imply that

$$(9.68) \qquad M[E', F' : D] \le M[\tilde{E}, \tilde{F} : \mathbb{C}] = 2\pi\left(\mathrm{Log}\,\frac{r_1}{r_0}\right)^{-1} < m\,.$$

The set $F = f^{-1}(F')$ is a connected set in $\Delta \sim E$. If $\langle z_n \rangle$ is a sequence in Δ such that $z_n \to z_0$, then from the continuity of $\tilde{f}$ we infer that $w_n = f(z_n) \to \tilde{f}(z_0) = w_0$. This dictates that w_n belong to $D \cap \Delta(w_0, s)$ — and so to F' — once n is suitably large. Accordingly, as soon as n gets sufficiently large, z_n is a point of F. Since $z_n \to z_0$, z_0 has to be a point of $\overline{F}$. A similar argument reveals that z_0' is also a point of $\overline{F}$. Lemma 4.3 and Theorem 4.2 then combine to give

$$m \le M[E, F : \Delta] = M[E', F' : D]\,,$$

contradicting (9.68). The upshot of the contradiction: $\tilde{f}$ is univalent — hence, in view of Lemma 4.1, is a homeomorphism of $\overline{\Delta}$ onto $\hat{D}$.

As for the converse, assume that f admits a homeomorphic extension $\tilde{f}$ to $\overline{\Delta}$. Let w_0 be a point of $\hat{\partial}D$ and let $r > 0$. Setting $z_0 = \tilde{f}^{-1}(w_0)$, we use the continuity of $\tilde{f}$ to choose $\delta > 0$ so that $\tilde{f}[\overline{\Delta} \cap \Delta(z_0, \delta)]$ is contained in $\Delta(w_0, r)$. The set $f[\Delta \cap \Delta(z_0, \delta)]$ is a connected set in $D \cap \Delta(w_0, r)$ and, for this reason, lies inside some component C of $D \cap \Delta(w_0, r)$. We now claim that, for sufficiently small s in $(0, r)$, $D \cap \Delta(w_0, s)$ is a subset of C. The alternative to this is the existence of a sequence $\langle w_n \rangle$ in $D \sim C$ such that $w_n \to w_0$. Write $z_n = f^{-1}(w_n)$. We can extract from $\langle z_n \rangle$ a convergent subsequence $\langle z_{n_k} \rangle$ whose limit z_0' is a point of $\overline{\Delta}$. Again by continuity

$$\tilde{f}(z_0') = \lim_{k \to \infty} f(z_{n_k}) = \lim_{k \to \infty} w_{n_k} = w_0\,.$$

From the univalence of $\tilde{f}$ we deduce that z_0' and z_0 are one and the same point. This implies, however, that z_{n_k} belongs to $\Delta \cap \Delta(z_0, \delta)$ once k is suitably large. For such k the point $w_{n_k} = f(z_{n_k})$ belongs to C, contrary to the selection of $\langle w_n \rangle$ as a sequence in $D \sim C$. In other words, given w_0 in $\widehat{\partial} D$ and $r > 0$, we have demonstrated that $D \cap \Delta(w_0, s)$ meets only one component of $D \cap \Delta(w_0, r)$, provided s in $(0, r)$ is taken appropriately small. Thus, D is locally connected along its boundary. ∎

4.5 Jordan Domains

Let D be a simply connected domain in $\mathbb{C}$, not the entire plane. Assume that D is locally connected along its boundary. Define a function $\gamma: [0, 2\pi] \to \widehat{\mathbb{C}}$ by $\gamma(t) = \tilde{f}(e^{it})$, where $\tilde{f}$ is the homeomorphic extension to $\overline{\Delta}$ of any conformal mapping f that transforms $\Delta = \Delta(0,1)$ to D. Then γ is a simple and closed path in $\widehat{\mathbb{C}}$, one whose trajectory is $\widehat{\partial} D$, so $\widehat{\partial} D$ is a Jordan curve in $\widehat{\mathbb{C}}$. A domain D in $\widehat{\mathbb{C}}$ with the property that $\widehat{\partial} D$ is a Jordan curve in $\widehat{\mathbb{C}}$ is called a *Jordan domain.* What we have just observed is this: if D is a simply connected domain in $\mathbb{C}$, $D \neq \mathbb{C}$, and if D is locally connected along its boundary, then D is a Jordan domain. The final link needed to complete the chain of proof for the Carathéodory-Osgood extension theorem is the converse of the preceding statement: if D is a Jordan domain in $\mathbb{C}$, then D is simply connected and is locally connected along its boundary. This converse is known to be true. Its standard proof is a by-product of the considerations that enter into the proof of the Jordan curve theorem. We shall not attempt to reproduce the proof, for to do so would divert us from our real mission in this chapter, which is to study conformal mappings. We shall merely accept the result as an established fact, just as we have done all along with the Jordan curve theorem itself, and go on. For the record we should point out that the Jordan curve theorem remains valid in $\widehat{\mathbb{C}}$: if J is a Jordan curve in $\widehat{\mathbb{C}}$, then $\widehat{\mathbb{C}} \sim J = D \cup D^*$, where D and D^* are disjoint domains in $\widehat{\mathbb{C}}$ having $\widehat{\partial} D = \widehat{\partial} D^* = J$. This implies, for instance, that a Jordan domain D in $\mathbb{C}$ is simply connected, because the set $\widehat{\mathbb{C}} \sim D$ is connected (Theorem 3.6). Indeed, if we apply the Jordan curve theorem to $J = \widehat{\partial} D$, we find that $\widehat{\mathbb{C}} \sim D = \widehat{D}^*$, a connected set. Making allowances for the preceding topological details, we have assembled all the components in the featured result of this section.

Theorem 4.9. (Carathéodory-Osgood Theorem) *A conformal mapping f of $\Delta = \Delta(0,1)$ onto a domain D in $\mathbb{C}$ can be extended to a homeomorphism of $\overline{\Delta}$ onto $\widehat{D}$ if and only if D is a Jordan domain.*

Having formally stated the Carathéodory-Osgood theorem, we wish to stress that in concrete situations Theorem 4.8, which amounts to the Carathéodory-Osgood theorem stripped of a few purely topological trappings, provides a perfectly worthy substitute for the higher powered result. In particular, it is not difficult to show that Theorem 4.8 applies to any bounded domain D in the complex plane whose boundary ∂D is the trajectory of a simple, closed, piecewise smooth path $\gamma:[a,b] \to \mathbb{C}$ with the property that $\dot{\gamma}(t) \neq 0$ for every t in $[a,b]$. (At possible points of non-differentiability this condition is interpreted to mean that both one-sided derivatives of γ are non-zero.) The domains falling under this description include, for instance, all bounded domains D for which ∂D is the trajectory of a simple and closed polygonal path.

For reference purposes we record the following useful corollary of Theorem 4.9.

Corollary 4.10. *Let f be a conformal mapping of a plane Jordan domain D onto another such domain D'. Then f extends to a homeomorphism $\tilde{f}$ of $\widehat{D}$ onto $\widehat{D}'$.*

Proof. We use the Riemann mapping theorem to choose a conformal mapping g of $\Delta = \Delta(0,1)$ onto D. Then $h = f \circ g$ maps Δ conformally to D'. By Theorem 4.9 both g and h have homeomorphic extensions to $\overline{\Delta}$, call them $\tilde{g}$ and $\tilde{h}$. We see that $\tilde{h} \circ \tilde{g}^{-1}$ is a homeomorphism of $\widehat{D}$ onto $\widehat{D}'$ that extends f. ∎

As we saw in Theorem 3.2, a conformal mapping f of one simply connected plane domain $D(\neq \mathbb{C})$ onto another such domain D' is uniquely determined once $f(z_0)$ in D' and $\operatorname{Arg}[f'(z_0)]$ are prescribed at some point z_0 of D. When D and D' are both Jordan domains, the Carathéodory-Osgood theorem opens the door to other means for uniquely specifying a conformal mapping between D and D'. The first of these occurs in the situation suggested by Figure 31. As a matter of fact, the associated theorem reproduces almost verbatim — concessions are made to present-day terminology — Riemann's own statement of his mapping theorem.

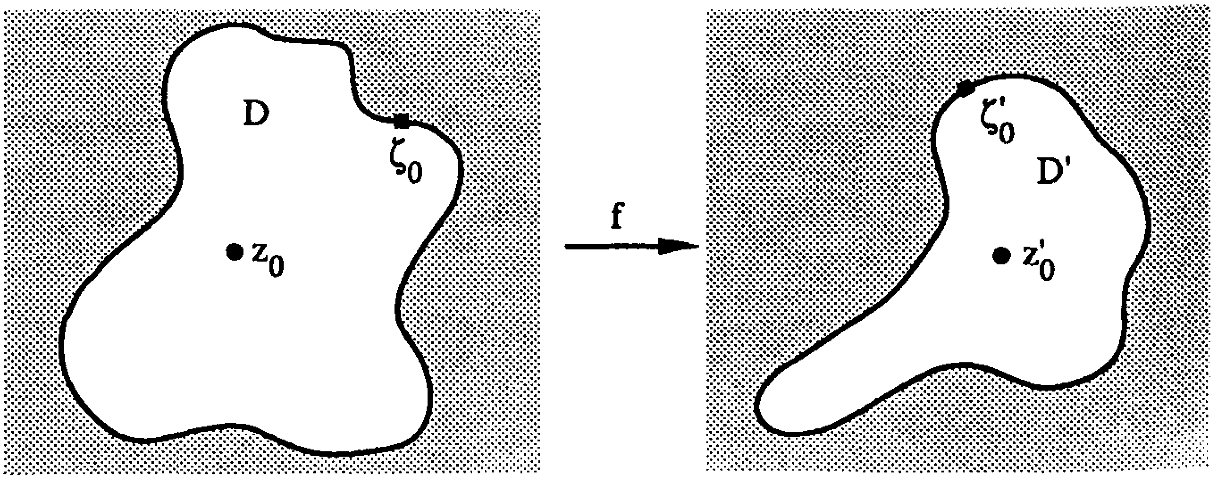


Figure 31.

Theorem 4.11. *Let D and D' be Jordan domains in the complex plane. Corresponding to any given points z_0 in D, ζ_0 on ∂D, z_0' in D', and ζ_0'*

on $\widehat{\partial}D'$ there is a unique homeomorphism f of $\widehat{D}$ onto $\widehat{D}'$ that maps D conformally onto D' and meets the conditions $f(z_0) = z_0'$ and $f(\zeta_0) = \zeta_0'$.

Proof. Let φ be any conformal mapping of $\Delta = \Delta(0,1)$ onto D taking 0 to z_0, and let $\widetilde{\varphi}$ be its homeomorphic extension to $\overline{\Delta}$. If $\widetilde{\varphi}^{-1}(\zeta_0) = e^{i\theta_0}$, then $g(z) = \widetilde{\varphi}(e^{i\theta_0}z)$ defines a homeomorphism g of $\overline{\Delta}$ onto $\widehat{D}$ that transforms Δ conformally to D with $g(0) = z_0$ and $g(1) = \zeta_0$. Similarly, one constructs a homeomorphism h of $\overline{\Delta}$ onto $\widehat{D}'$ that maps Δ to D' in a conformal fashion and satisfies $h(0) = z_0'$, $h(1) = \zeta_0'$. The function $f = h \circ g^{-1}$ furnishes a homeomorphism of $\widehat{D}$ onto $\widehat{D}'$ that has all the desired features. Furthermore, f is the unique such mapping. If, namely, f_0 is any homeomorphism of $\widehat{D}$ onto $\widehat{D}'$ that maps D conformally onto D' and that sends z_0 to z_0' and ζ_0 to ζ_0', then $h^{-1} \circ f_0 \circ g$ is a homeomorphism of $\overline{\Delta}$ onto itself that is conformal in Δ and fixes both 0 and 1. The only conformal self-mappings of Δ fixing the origin are rotations about the origin; the only such rotation also fixing 1 is the identity mapping. Thus $h^{-1} \circ f_0 \circ g(z) = z$ for every z in $\overline{\Delta}$. This implies that $f_0(z) = h \circ g^{-1}(z) = f(z)$ for all points z of $\widehat{D}$ — and so confirms the uniqueness of f. ∎

4.6 Oriented Boundaries

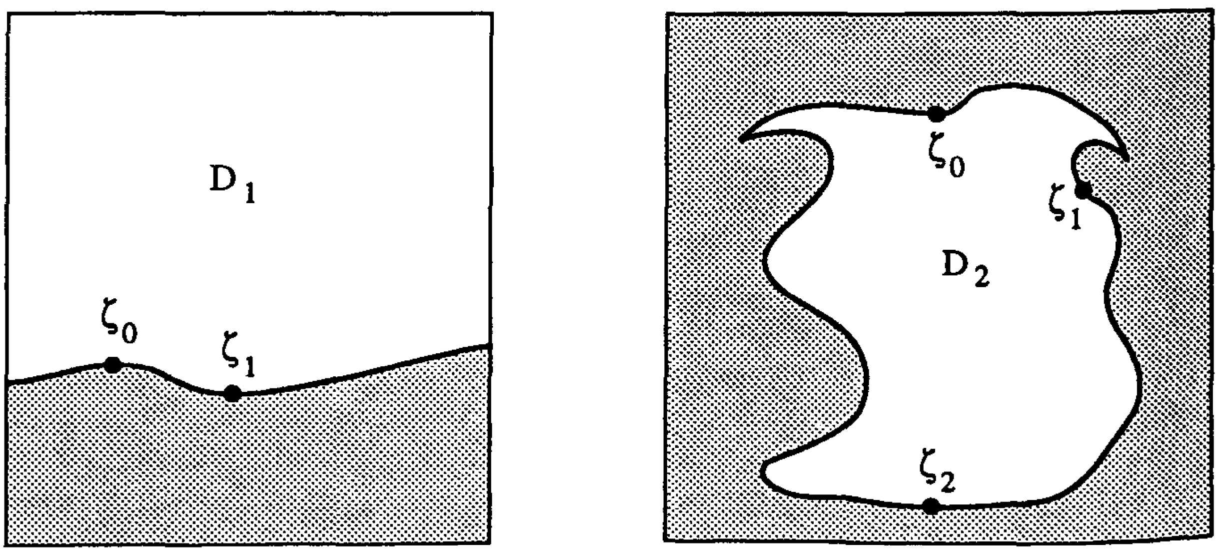
A second procedure for specifying in a unique way a conformal mapping of one plane Jordan domain D onto another is to dictate the values that the mapping should take — or, more accurately, the values that its homeomorphic extension to $\widehat{D}$ should take — at three given boundary points of D. Before this pronouncement can be incorporated into a theorem, a few words must be said concerning the notion of "orientation." Let D be a Jordan domain in $\mathbb{C}$, and let $\gamma: [a,b] \to \widehat{\mathbb{C}}$ be a simple, closed path that parametrizes $\widehat{\partial}D$. The intuitive meaning of the statement "γ is positively oriented with respect to D" is that, as t increases from a to b and $\gamma(t)$ traverses $\widehat{\partial}D$, D "stays to the left" of $\gamma(t)$. In the case where the domain D is bounded (we have heretofore referred to D as the "inside" of the Jordan curve $J = \partial D$ in this situation) and γ is piecewise smooth, we were able in Chapter V to turn this intuitive idea into a precise mathematical definition, always taking for granted the conclusions of the Jordan curve theorem: γ is positively oriented with respect to D — at the time we simply said that γ was positively oriented — if and only if $n(\gamma, z) = 1$ for every z in D. By revamping our treatment of winding numbers, a process that would entail some non-elementary topology, it would be possible to extend this definition to cover general D and γ. We shall again resist the temptation to go into detail on this topic, preferring to keep our attention centered on complex analysis rather than letting the focus shift to topology. The key fact we shall need — and shall use without proof — is this: *if f maps* *the disk* $\Delta = \Delta(0,1)$ *conformally onto a Jordan domain* D *in* $\mathbb{C}$, *if* $\tilde{f}$ *is the homeomorphic extension of* f *to* $\overline{\Delta}$, *and if* γ *is the parametrization of* $\widehat{\partial}D$ *defined on* $[0,2\pi]$ *by* $\gamma(t) = \tilde{f}(e^{it})$, *then* γ *is positively oriented with respect to* D. This statement is rendered at least plausible by the following argument. For $0 < r < 1$ we define a path γ_r on $[0,2\pi]$ by $\gamma_r(t) = f(re^{it})$. If z is a point of D and if $|f^{-1}(z)| < r < 1$, then the argument principle lets us know that

$$n(\gamma_r, z) = \frac{1}{2\pi i}\int_0^{2\pi} \frac{f'(re^{it})ire^{it}\,dt}{f(re^{it}) - z} = \frac{1}{2\pi i}\int_{|\zeta|=r} \frac{f'(\zeta)\,d\zeta}{f(\zeta)-z} = 1\,.$$

Whatever the exact definition of $n(\gamma, z)$ might be, it is not unreasonable to expect that $n(\gamma, z) = \lim_{r\to 1} n(\gamma_r, z)$. Thus, $n(\gamma, z) = 1$ ought to hold for every z in D.

An ordered triple $(\zeta_0, \zeta_1, \zeta_2)$ of distinct boundary points of a Jordan domain D in $\mathbb{C}$ is called *positively oriented relative to* D if, when starting at ζ_0 and traversing $\widehat{\partial}D$ according to any simple parametrization that is positively oriented with respect to D, one passes through the given points in the order $\zeta_0, \zeta_1, \zeta_2, \zeta_0$ (i.e., if a simple parametrization $\gamma\colon [a,b] \to \widehat{\partial}D$ is positively oriented with respect to D and has $\gamma(a) = \gamma(b) = \zeta_0$, it is required that the points t_1 and t_2 for which $\gamma(t_1) = \zeta_1$ and $\gamma(t_2) = \zeta_2$ satisfy $t_1 < t_2$.) If the points are encountered in the order $\zeta_0, \zeta_2, \zeta_1, \zeta_0$ for such a parametrization, the triple $(\zeta_0, \zeta_1, \zeta_2)$ is declared to be *negatively oriented relative to* D. (See Figure 32. The domain D_1 there is intended to be an unbounded Jordan domain.) We are now prepared to state:

Theorem 4.12. *Let* D *and* D' *be Jordan domains in the complex plane. Corresponding to any ordered triples of distinct points* $(\zeta_0, \zeta_1, \zeta_2)$ *from* $\widehat{\partial}D$



$(\zeta_0, \zeta_1, \infty)$ positively oriented relative to D_1

$(\zeta_0, \zeta_1, \zeta_2)$ negatively oriented relative to D_2

Figure 32.

and $(\zeta_0', \zeta_1', \zeta_2')$ *from* $\widehat{\partial} D'$ *that exhibit like orientations relative to* D *and* D'*, respectively, there exists a unique homeomorphism* f *of* $\widehat{D}$ *onto* $\widehat{D}'$ *that maps* D *conformally onto* D' *and has* $f(\zeta_j) = \zeta_j'$ *for* $j = 0, 1, 2$.

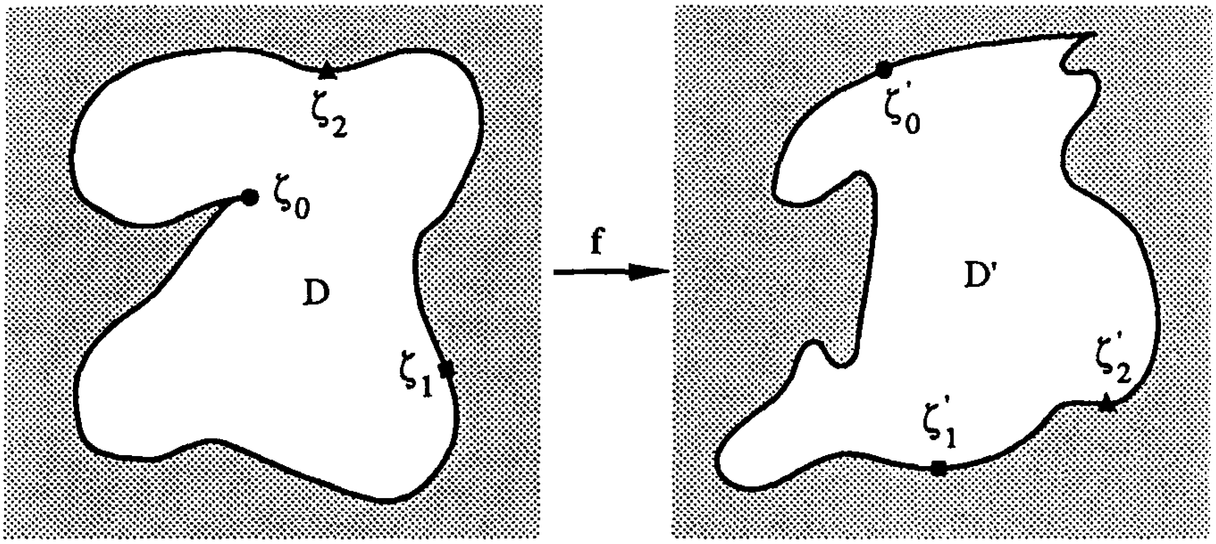


Figure 33.

Proof. We may assume that both triples are positively oriented. (If not, consider $(\zeta_0, \zeta_2, \zeta_1)$ and $(\zeta_0', \zeta_2', \zeta_1')$ instead of the original triples.) Let $\Delta = \Delta(0,1)$, let g be a homeomorphism of $\overline{\Delta}$ onto $\widehat{D}'$ that maps Δ conformally onto D and takes 1 to ζ_0, and let h be a homeomorphism of $\overline{\Delta}$ onto $\widehat{D}'$ that transforms Δ conformally to D' and has $h(1) = \zeta_0'$. Because the parametrization γ of $\widehat{\partial} D$ given by $\gamma(t) = g(e^{it})$ for $0 \le t \le 2\pi$ is positively oriented with respect to D and has initial point $\gamma(0) = g(1) = \zeta_0$ and because $(\zeta_0, \zeta_1, \zeta_2)$ is positively oriented relative to D, we can write $\zeta_1 = g(e^{i\theta_1})$ and $\zeta_2 = g(e^{i\theta_2})$, where $0 < \theta_1 < \theta_2 < 2\pi$. For similar reasons we can express $\zeta_1' = h(e^{i\psi_1})$ and $\zeta_2' = h(e^{i\psi_2})$, with $0 < \psi_1 < \psi_2 < 2\pi$. Next, let φ be the unique Möbius transformation that sends 1 to 1, $e^{i\theta_1}$ to $e^{i\psi_1}$, and $e^{i\theta_2}$ to $e^{i\psi_2}$. Then φ obviously maps the circle $K = K(0,1)$ to itself. It follows that either $\varphi(\Delta) = \Delta$ or $\varphi(\Delta) = \widehat{\mathbb{C}} \sim \overline{\Delta}$. If the latter were true, φ would have a single simple pole in Δ and be free of zeros there. The argument principle tells us that the parametrization of K given by $\beta(t) = \varphi(e^{it})$ for $0 \le t \le 2\pi$ would then have

$$n(\beta, 0) = \frac{1}{2\pi i} \int_0^{2\pi} \frac{\varphi'(e^{it}) i e^{it} dt}{\varphi(e^{it})} = \frac{1}{2\pi i} \int_{|z|=1} \frac{\varphi'(z)\, dz}{\varphi(z)} = -1 ,$$

implying that β would be negatively oriented with respect to Δ. This would force the triple $(\beta(0), \beta(\theta_1), \beta(\theta_2)) = (1, e^{i\psi_1}, e^{i\psi_2})$ to be negatively oriented relative to Δ, which it plainly isn't. We infer from these remarks that $\varphi(\overline{\Delta}) = \overline{\Delta}$ must be true and conclude that $f = h \circ \varphi \circ g^{-1}$ is a homeomorphism blessed with the requisite properties.

To establish the uniqueness of f, suppose that f_0 is an arbitrary homeomorphism of $\widehat{D}$ onto $\widehat{D}'$ meeting all the stated requirements. Then $\varphi_0 = g^{-1} \circ f_0^{-1} \circ f \circ g$ is a homeomorphism of $\overline{\Delta}$ onto itself. It maps Δ conformally to itself and fixes the points 1, $e^{i\theta_1}$, and $e^{i\theta_2}$. Theorem 1.4 implies that φ_0 is the restriction to $\overline{\Delta}$ of a Möbius transformation. Since φ_0 has three fixed points, that Möbius transformation is necessarily the identity transformation. In particular, $\varphi_0(z) = z$ for every z in $\overline{\Delta}$. From this it can be deduced that $f_0 = f$, which is the uniqueness assertion desired. ∎

The results in this section have many applications. We conclude the section with one sample — to wit, we use the Carathéodory-Osgood theorem to prove that any bounded Jordan domain is regular for the Dirichlet problem. This allows us to make good on a promise issued in Chapter VI.

Theorem 4.13. *Any bounded Jordan domain D in the complex plane is regular for the Dirichlet problem.*

Proof. Consider a continuous function $h: \partial D \to \mathbb{R}$. It is our job to produce a continuous function $u: \overline{D} \to \mathbb{R}$ that is harmonic in D and satisfies $u = h$ on ∂D. Let $\Delta = \Delta(0,1)$, and let f be a homeomorphism of $\overline{D}$ onto $\overline{\Delta}$ that transforms D conformally to Δ. The function $h_0 = h \circ f^{-1}$ is continuous on $\partial\Delta$. By Theorem VI.3.1 there exists a continuous function $u_0: \overline{\Delta} \to \mathbb{R}$ that is harmonic in Δ and agrees with h_0 on $\partial\Delta$. Finally, $u = u_0 \circ f$ is continuous on $\overline{D}$, $u = h$ on ∂D, and a direct calculation — see Exercise VI.4.5 — demonstrates that $u_{xx} + u_{yy} = 0$ in D; i.e., u is harmonic in D. ∎

The argument just presented also works for an unbounded Jordan domain D, provided the boundary data h is continuous not just on ∂D, but also at ∞. Since the Dirichlet problem in $\Delta(0,1)$ has a concrete solution given by the Poisson integral formula, the proof of Theorem 4.13 gives rise to an integral formula representing the solution of the Dirichlet problem in any Jordan domain for which the mapping f can be explicitly determined.

5 Conformal Mappings onto Polygons

5.1 Polygons

Let $\gamma = [z_1, z_2] + [z_2, z_3] + \cdots + [z_{n-1}, z_n] + [z_n, z_1]$ be a simple and closed polygonal path in the complex plane endowed with the extra feature that for $j = 1, 2, \ldots, n-1$ the points z_j, z_{j+1}, and z_{j+2} are not collinear. (To make this statement meaningful for $j = n-1$, set $z_{n+1} = z_1$.) Under these conditions we describe $P = \overline{D}$, where D is the inside of the Jordan curve $J = |\gamma|$, as the *closed polygon with vertices* $z_1, z_2, \ldots, z_n$. (See Figure 34.) We shall always tacitly assume that the vertices of P are listed in an order which causes the path γ to be positively oriented with respect to D. If λ_j is

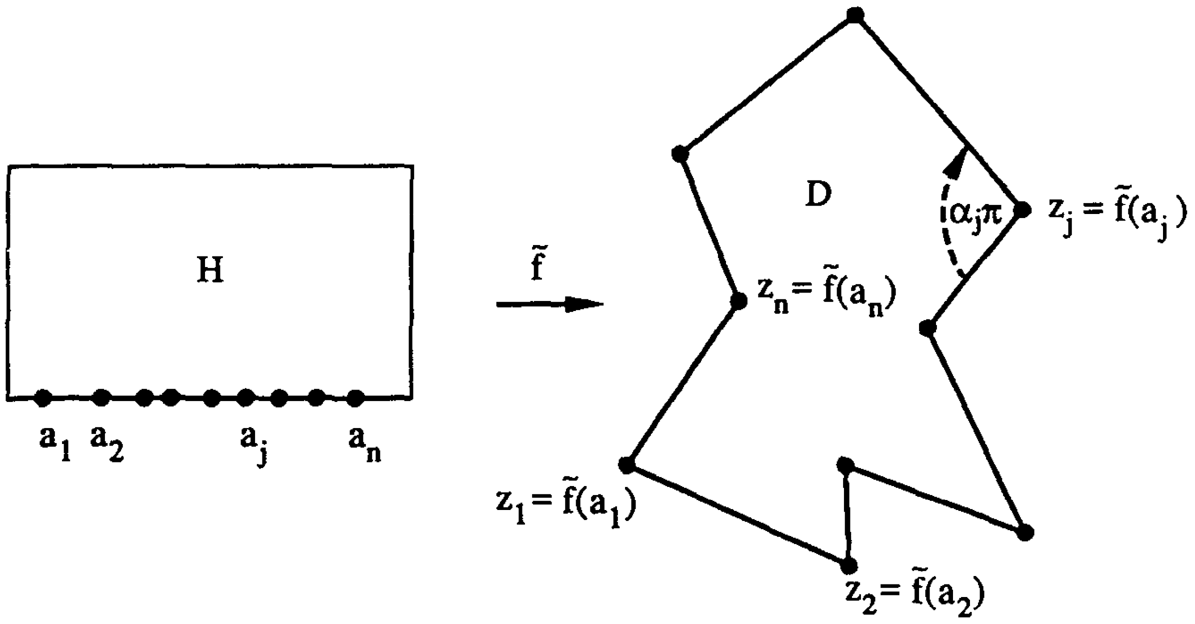


Figure 34.

the line segment joining z_j and z_{j+1}, then $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the *sides* of P. The (unoriented) interior angle of P at the vertex z_j will be written $\alpha_j \pi$, with $0 < \alpha_j < 2$. (N.B. The requirements we have imposed on γ insure that z_j is an "authentic" vertex of P, meaning that $\alpha_j \neq 1$.) It is a fact of elementary geometry that

$$\sum_{j=1}^{n} \alpha_j = n - 2 \; . \tag{9.69}$$

In this section we study conformal mappings of the half-plane $H = \{z : \operatorname{Im} z > 0\}$ onto the interior D of such a closed polygon P. It is our goal to obtain a concrete representation for such a mapping, a representation in the shape of a famous integral formula that bears the names of H.A. Schwarz and E.B. Christoffel (1829-1900). (As it is an easy matter to pass conformally from the disk $\Delta = \Delta(0,1)$ to H by means of a Möbius transformation, we shall also determine the structure of a conformal mapping from Δ onto D.) If f maps H conformally onto D, then f can be extended to a homeomorphism $\tilde{f}$ of $\widehat{H}$ onto P. We denote by $a_1, a_2, \ldots, a_n$ the preimages of the vertices $z_1, z_2, \ldots, z_n$ of P under $\tilde{f}$. We shall insist that the labeling of vertices be done in such a way that $-\infty < a_1 < a_2 < \cdots < a_n \leq \infty$, as in Figure 34.

5.2 The Reflection Principle

Underlying our derivation of the Schwarz-Christoffel formula is a method, due to Schwarz, of extending certain analytic functions from their given domain-sets to larger ones through a process of reflection. As a point of historical interest, we might add that Schwarz originally developed his "reflection principle" with the study of conformal mappings of H to polygons as the motivation. Our treatment of this result begins with the following refinement of Lemma V.1.1.

Lemma 5.1. *Let R be a closed rectangle in the complex plane, and let $f: R \to \mathbb{C}$ be a continuous function that is analytic in the interior of R. Then $\int_{\partial R} f(z)\,dz = 0$.*

Proof. The argument we make is a variation on the one used to prove Lemma V.1.2. Given $\epsilon > 0$, we verify that

$$\left|\int_{\partial R} f(z)\,dz\right| \le 4\epsilon L\ , \tag{9.70}$$

in which L is the perimeter of R. Letting $\epsilon \to 0$ in (9.70) we arrive at the stated conclusion. In view of the uniform continuity of f on R, we can choose an integer $n > 2$ such that $|f(z_1) - f(z_2)| < \epsilon$ for all points z_1 and z_2 of R that satisfy $|z_1 - z_2| < d/n$. Here d is the length of the diagonal of R. As in the proof of Lemma V.1.2, we subdivide R into n^2 congruent rectangles $R_{k\ell}$ and observe that

$$\left|\int_{\partial R} f(z)\,dz\right| = \left|\sum_{k,\ell}\int_{\partial R_{k\ell}} f(z)\,dz\right| \le \sum_{k,\ell}\left|\int_{\partial R_{k\ell}} f(z)\,dz\right| . \tag{9.71}$$

The rectangles $R_{k\ell}$ fall into two classes: $(n-2)^2$ of them lie in the interior of R; each of the remaining $4(n-1)$ rectangles has at least one side contained in ∂R. For any of the "interior" rectangles $R_{k\ell}$ we have $\int_{\partial R_{k\ell}} f(z)\,dz = 0$ by Lemma V.1.1. For a "border" rectangle $R_{k\ell}$ with center $z_{k\ell}$ we get

$$\begin{aligned}\left|\int_{\partial R_{k\ell}} f(z)\,dz\right| &= \left|\int_{\partial R_{k\ell}} f(z) - \int_{\partial R_{k\ell}} f(z_{k\ell})\,dz\right| \\ &\le \int_{\partial R_{k\ell}} |f(z) - f(z_{k\ell})|\,|dz| < \frac{\epsilon L}{n}\end{aligned}$$

by reason of the fact that $|z - z_{k\ell}| < d/n$ is true for every z in $\partial R_{k\ell}$ — this forces $|f(z) - f(z_{k\ell})| < \epsilon$ to hold for such z — and that $R_{k\ell}$ has perimeter L/n. Because the "border" rectangles are $4(n-1)$ in number, (9.71) leads to

$$\left|\int_{\partial R} f(z)\,dz\right| \le \frac{4(n-1)\epsilon L}{n} \le 4\epsilon L\ ,$$

confirming (9.70). ∎

The next lemma presents the reflection principle of Schwarz in its simplest setting.

Lemma 5.2. *Let D be a plane domain that is symmetric about the real axis, let G and G^* be the components of $D \sim \mathbb{R}$, and let $I = D \cap \mathbb{R}$. If $f: G \cup I \to \mathbb{C}$ is a continuous function that is analytic in G and real-valued on I, then the function $F: D \to \mathbb{C}$ defined by*

$$F(z) = \begin{cases} f(z) & \text{if } z \in G \cup I , \\ \overline{f(\overline{z})} & \text{if } z \in G^* , \end{cases}$$

is an analytic function.

Proof. (Recall Figure VI.2.) The function F is analytic in G by hypothesis. It is also analytic in G^*, for when z_0 lies in G^* we have

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

i.e., F is differentiable at z_0 with $F'(z_0) = \overline{f'(\overline{z_0})}$. The real-valuedness and continuity of f at the points of I insure that F, too, is continuous at these points. Therefore, F is at least a continuous function. Morera's theorem will testify to the analyticity of F if only it can be demonstrated that $\int_{\partial R} F(z)\, dz = 0$ for each closed rectangle R in D whose sides are parallel to the coordinate axes. Fix such a rectangle R. If R is contained in either $G \cup I$ or $G^* \cup I$, then the vanishing of the integral in question is a direct consequence of Lemma 5.1. In all other cases R is partitioned by I into two smaller rectangles R' and R'', where R' is located in $G \cup I$ and R'' in $G^* \cup I$. Thus, by the preceding remark,

$$\int_{\partial R} F(z)\, dz = \int_{\partial R'} F(z)\, dz + \int_{\partial R''} F(z)\, dz = 0 + 0 = 0 .$$

The invocation of Morera's theorem completes the proof. ∎

We generalize Lemma 5.2 as follows.

Theorem 5.3. (Schwarz Reflection Principle) *Let K and $\widetilde{K}$ be circles in $\widehat{\mathbb{C}}$ with associated reflections ρ and $\widetilde{\rho}$, let D be a domain in $\mathbb{C}$ that is symmetric about K (i.e., $\rho(D) = D$), let G and G^* be the components of $D \sim K$, and let $I = D \cap K$. If $f: G \cup I \to \mathbb{C}$ is a continuous function that is analytic in G and maps I to a subset of $\widetilde{K}$, then the function $F: D \to \widehat{\mathbb{C}}$ defined by*

$$F(z) = \begin{cases} f(z) & \text{if } z \in G \cup I , \\ \widetilde{\rho} \circ f \circ \rho(z) & \text{if } z \in G^* , \end{cases}$$

is a meromorphic function. In fact, F is an analytic function except when $\widetilde{K}$ is a true circle whose center belongs to the range of f, in which case the only poles of F are found at the points z of G^* with the property that $f[\rho(z)]$ is the center of $\widetilde{K}$.

Proof. Since the function $\widetilde{\rho} \circ f \circ \rho$ is continuous in $G^* \cup I$ and coincides with f on I, we see easily that the function F is continuous. It only assumes the value ∞ when $\widetilde{K}$ is a true circle with its center in the range of f. In this event $F(z) = \infty$ at precisely those points z of G^* for which $f[\rho(z)]$ is the center of $\widetilde{K}$. These are the only places where F could possibly have poles. We now fix an arbitrary open subarc I' of I such that $I' \neq I$ and $f(I') \neq \widetilde{K}$. We confirm that F is meromorphic in the domain $D' = G \cup I' \cup G^*$. The establishment of this fact for every such set I' clearly suffices to complete the proof of the theorem. Write $K_0 = \mathbb{R} \cup \{\infty\}$. We select Möbius transformations g and h with the following properties: $g(K) = K_0$; $h(K_0) = \widetilde{K}$; ∞ does not lie in $I_0 = g(I')$; $h(\infty)$ does not lie in $f(I')$. By Theorem 2.7 the plane domain $D_0 = g(D')$ is symmetric about the real axis. The components of $D_0 \sim \mathbb{R}$ are $G_0 = g(G)$ and $G_0^* = g(G^*)$, while $D_0 \cap \mathbb{R} = I_0$. Furthermore, the function $f_0 \colon G_0 \cup I_0 \to \mathbb{C}$ given by $f_0(z) = h^{-1} \circ f \circ g^{-1}(z)$ is continuous, it is analytic in G_0, and it is real-valued on I_0. We can appeal to Lemma 5.2 and assert that the function $F_0 \colon D_0 \to \mathbb{C}$ defined by

$$F_0(z) = \begin{cases} f_0(z) & \text{if } z \in G_0 \cup I_0 \,, \\ \overline{f_0(\overline{z})} & \text{if } z \in G_0^* \,, \end{cases}$$

is analytic. It follows that $h \circ F_0 \circ g$ is a meromorphic function with domain-set D'. We claim that in D' this function and F are one and the same. It is immediate from our definition of f_0 that $h \circ F_0 \circ g = f = F$ in the set $G \cup I'$. What happens in G^*? To answer this question we once more utilize information given to us by Theorem 2.7: $g[\rho(z)] = \overline{g(z)}$ and $\widetilde{\rho}[h(z)] = h(\overline{z})$ for every z in $\widehat{\mathbb{C}}$. We learn from it that

$$\begin{aligned} h \circ F_0 \circ g(z) &= h\{\overline{f_0[\overline{g(z)}]}\} = [(\widetilde{\rho} \circ h) \circ f_0 \circ (g \circ \rho)](z) \\ &= [\widetilde{\rho} \circ (h \circ f_0 \circ g) \circ \rho](z) = \widetilde{\rho} \circ f \circ \rho(z) = F(z) \end{aligned}$$

for every z in G^*. Agreeing with $h \circ F_0 \circ g$ in D', the function F is meromorphic in this set, as maintained. ∎

The function F constructed under the conditions detailed in Theorem 5.3 is called the *continuation of f across I by reflection.* The basic reflection principle that we have elected to state here admits further generalizations. A couple of these are indicated in the exercises.

5.3 The Schwarz-Christoffel Formula

Theorem 5.3 equips us with just the tool we need to expose the structure of a conformal mapping from the upper half-plane H to a polygonal region.

In an effort to streamline our presentation of the theorem that does this we collect some technical details in two preliminary lemmas.

Lemma 5.4. *Suppose that f is a conformal mapping of the half-plane $H = \{z: \operatorname{Im} z > 0\}$ onto the interior D of a closed polygon P in the complex plane and that $-\infty < a_1 < a_2 < \cdots < a_n \leq \infty$ is the list of points transformed to the vertices of P by $\tilde{f}$, the homeomorphic extension of f to $\widehat{H}$. Then the function f''/f' can be extended to a function g that is analytic in the domain $G = \mathbb{C} \sim \{a_1, a_2, \ldots, a_n\}$.*

Proof. Let $I_1 = (-\infty, a_1), \ldots, I_n = (a_{n-1}, a_n)$, and $I_{n+1} = (a_n, \infty)$ be the open intervals into which the real axis splits upon removal of the points $a_1, a_2, \ldots, a_n$. (In the event that $a_n = \infty$, disregard I_{n+1}.) The image of I_j under $\tilde{f}$ is contained in a side of P — hence, in a line L_j. Let ρ_j denote reflection in L_j, and let ρ be reflection in the real axis. Using the Schwarz reflection principle we can continue $\tilde{f}$ across I_j to define an analytic function F_j in the domain $D_j = H \cup I_j \cup H^*$. Here $H^* = \{z: \operatorname{Im} z < 0\}$, a set in which F_j is given by

$$F_j = \rho_j \circ f \circ \rho \,. \tag{9.72}$$

The function F_j is clearly univalent in both H and H^*. (It might fail to be univalent in D_j, however, for nothing in our hypotheses rules out the eventuality that the two sets $F_j(H)$ and $F_j(H^*)$ overlap.) In particular, $F_j'(z) \neq 0$ holds for every z in $H \cup H^*$ (Theorem VIII.3.9). It is also true that $F_j'(z) \neq 0$ if z belongs to I_j. To see this, fix a point z_0 of I_j. Then $w_0 = F_j(z_0) = \tilde{f}(z_0)$ is a point of $\partial D \cap L_j$, but w_0 is not a vertex of P. We can therefore choose $s > 0$ so that the set $D \cap \Delta(w_0, s)$ is a half-disk, one of the two components of $\Delta(w_0, s) \sim L_j$. Next, pick $r > 0$ small enough that $\mathbb{R} \cap \Delta(z_0, r)$ is a subset of I_j and that $\tilde{f}[H \cap \Delta(z_0, r)]$ is contained in $D \cap \Delta(w_0, s)$. The set $F_j[H \cap \Delta(z_0, r)]$ thus lies on one side of the line L_j, and its reflection in L_j — i.e., the set $F_j[H^* \cap \Delta(z_0, r)]$ — lies on the opposite side of L_j (Figure 35). This insures that F_j is univalent in $\Delta(z_0, r)$ and, as a result, that $F_j'(z_0) \neq 0$. An implication of the above considerations is that the function $g_j = F_j''/F_j'$ is analytic in D_j.

How are F_j and F_k related in H^* when $j \neq k$? The function F_j maps H^* conformally onto $\rho_j(D)$. We use (9.72) to compute $F_k \circ F_j^{-1}$ in $\rho_j(D)$:

$$F_k \circ F_j^{-1} = (\rho_k \circ f \circ \rho) \circ (\rho_j \circ f \circ \rho)^{-1} = \rho_k \circ f \circ \rho \circ \rho^{-1} \circ f^{-1} \circ \rho_j^{-1} = \rho_k \circ \rho_j \,.$$

(Remember: $\rho_j^{-1} = \rho_j$.) As the composition of two reflections in lines, $\rho_k \circ \rho_j$ has the form $\rho_k \circ \rho_j(z) = az + b$, where $a \neq 0$ and b are constants (they do depend on j and k, of course). Accordingly, we can write

$$F_k \circ F_j^{-1}(z) = az + b$$

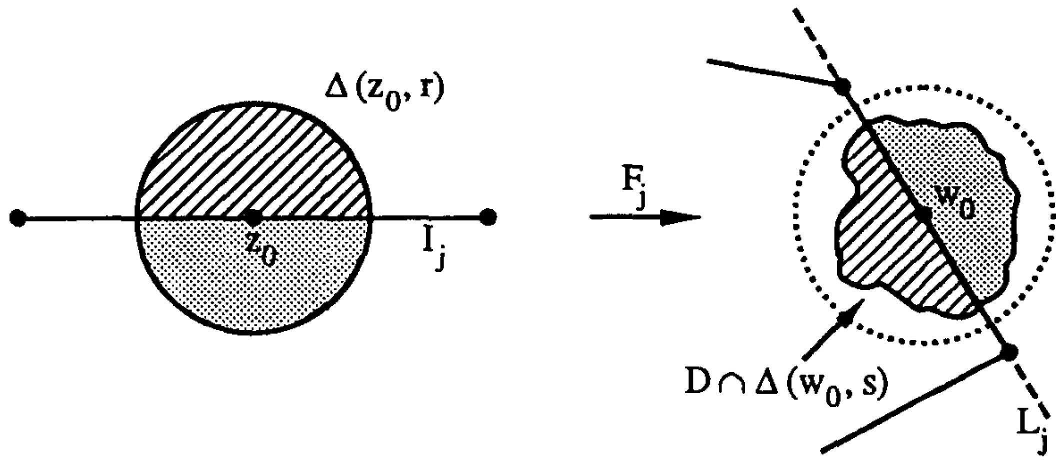


Figure 35.

for all z in $\rho_j(D)$ or, equivalently,

$$F_k(z) = aF_j(z) + b$$

for every z in H^*. The last relation leads to the conclusion that $g_k = F_k''/F_k' = F_j''/F_j' = g_j$ in H^* when $j \neq k$. If, as a consequence, we define a function g in the domain $G = \mathbb{C} \sim \{a_1, a_2, \ldots, a_n\} = \bigcup_j D_j$ by setting $g(z) = g_j(z)$ for z in D_j, then we succeed in creating an analytic function that coincides with f''/f' in H. ∎

The next lemma clarifies the nature of the the singularities of the function g just constructed.

Lemma 5.5. *Under the conditions of Lemma 5.4, let the interior angle of P at the vertex $\tilde{f}(a_j)$ be $\alpha_j \pi$. If $a_j \neq \infty$, the singularity of the function g at a_j is a simple pole with residue $\alpha_j - 1$. Furthermore, $g(z) \to 0$ as $z \to \infty$.*

Proof. To conserve on notation we write $a = a_j$ and $\alpha = \alpha_j$. In the first part of the proof we assume that $a \neq \infty$. Let $h(z) = g(z) - (\alpha - 1)(z - a)^{-1}$ for z in the domain G. We shall show that in some punctured disk $\Delta^*(a, \delta)$ this function h can be expressed in the form

$$h(z) = \frac{(\alpha + 1)f_0'(z) + (z - a)f_0''(z)}{\alpha f_0(z) + (z - a)f_0'(z)}, \tag{9.73}$$

where f_0 is a function that is both analytic and zero-free in the full disk $\Delta(a, \delta)$. Since the right-hand side of (9.73) plainly approaches the finite limit $(\alpha + 1)f_0'(a)/\alpha f_0(a)$ as $z \to a$, it becomes evident that h has a removable singularity at a. The singularity of g at a is then immediately recognized to be a simple pole at which g has residue $\alpha - 1$.

If $c \neq 0$ and d are complex constants, then $\varphi = cf + d$ also maps H conformally to a polygonal region, one similar to D. Its homeomorphic extension to $\hat{H}$ is $\tilde{\varphi} = c\tilde{f} + d$, so $\tilde{\varphi}$ also takes the points $a_1, a_2, \ldots, a_n$ to the vertices of its image polygon. It follows that Lemma 5.4 can be applied to

φ. The result is a function that is analytic in D and coincides with $\varphi''/\varphi' = f''/f'$ in H. By Corollary VIII.1.6 the only analytic extension of f''/f' to G is g, the function constructed in the proof of Lemma 5.4. In other words, it makes no difference whether we apply Lemma 5.4 to f or to φ — we obtain the same function g in either case. Consequently, in verifying (9.73) we are allowed to make the simplifying assumptions that the extension $\tilde{f}$ of f to $\widehat{H}$ satisfies $\tilde{f}(a) = 0$ and that the interior angle of P at the origin is bisected by the positive real axis, as in Figure 36. (If not, replace f by $cf + d$ for suitable c and d.) Under these assumptions we can choose and fix s for which $D \cap \Delta(0,s) = \{w \in \Delta(0,s): |\operatorname{Arg} w| < \alpha\pi/2\}$. We can then select r with the property that $\tilde{f}[\widehat{H} \cap \Delta(a,r)]$ is a subset of $\Delta(0,s)$. Finally, we can define a function $\psi: \widehat{H} \cap \Delta(a,r) \to \mathbb{C}$ as follows: $t = \psi(z) = [\tilde{f}(z)]^{1/\alpha}$. This function is a homeomorphism that maps $H \cap \Delta(a,r)$ conformally onto a domain contained in the half-plane $\{t : \operatorname{Re} t > 0\}$. It transforms the interval $I = (a - r, a + r)$ to an open interval on the imaginary axis in the t-plane (Figure 36).

We now continue the function ψ across I by reflection. Through this process we produce a conformal mapping $\Psi: \Delta(a,r) \to \mathbb{C}$ with the property that $f(z) = [\Psi(z)]^\alpha$ for every z in $H \cap \Delta(a,r)$. As a function analytic in $\Delta(a,r)$ whose only zero there is a simple one at a, Ψ can be rewritten in the

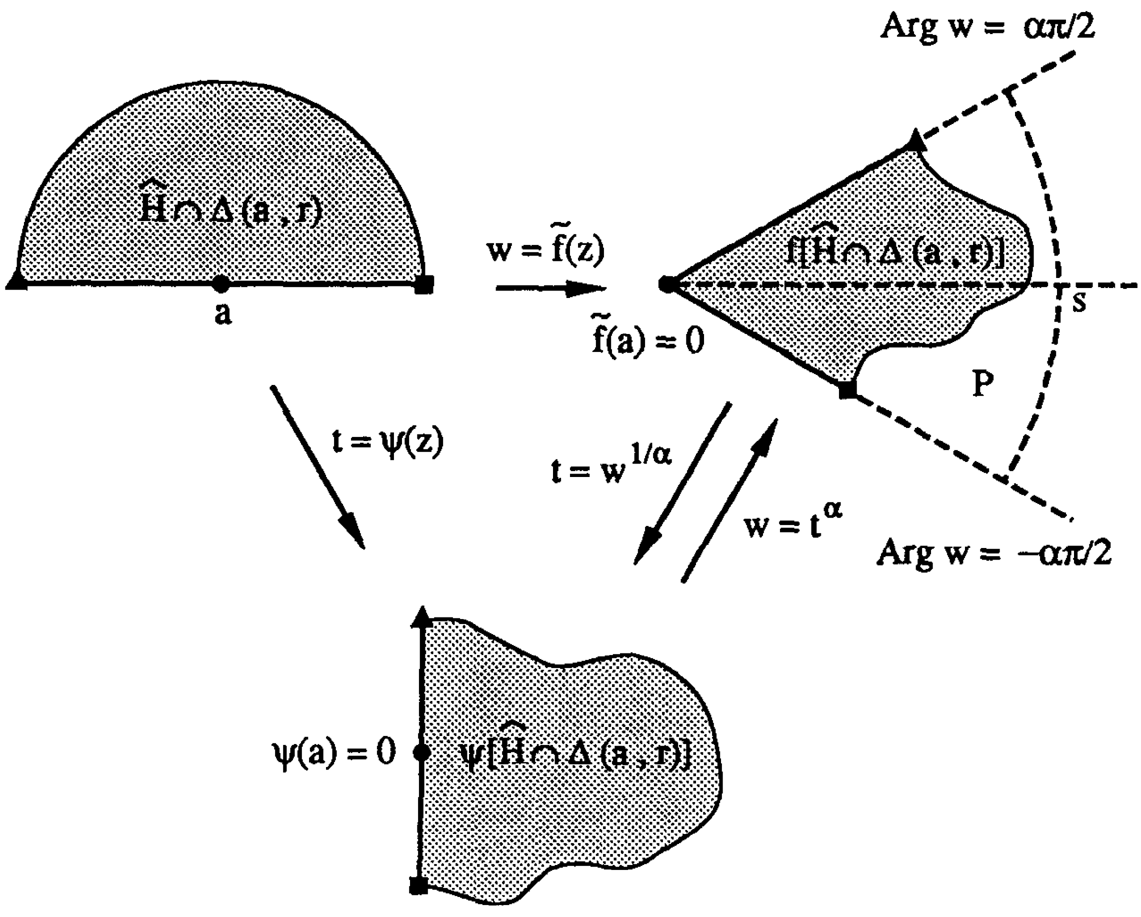


Figure 36.

form $\Psi(z) = (z-a)\Psi_0(z)$, where Ψ_0 is a function that is both analytic and free of zeros in $\Delta(a,r)$. Theorem V.6.2 insures the existence of branches of $\log \Psi_0(z)$ in $\Delta(a,r)$. We pick one and label it L. In fact, after possibly adjusting our original choice for L by adding to it an integral multiple of $2\pi i$, we may assume that $L(z) = \operatorname{Log}\Psi(z) - \operatorname{Log}(z-a)$ whenever z belongs to $H \cap \Delta(a,r)$. (Reason: the restriction of L to $H \cap \Delta(a,r)$ is a branch of $\log \Psi_0(z)$ in $H \cap \Delta(a,r)$, as is the function $\widetilde{L}$ given in that half-disk by $\widetilde{L}(z) = \operatorname{Log}\Psi(z) - \operatorname{Log}(z-a)$, conditions which imply that L and $\widetilde{L}$ differ in $H \cap \Delta(a,r)$ by an integral multiple of $2\pi i$.) For z in $H \cap \Delta(a,r)$ we can thus assert that

$$\begin{aligned} f(z) &= [\Psi(z)]^\alpha = e^{\alpha \operatorname{Log}\Psi(z)} = e^{\alpha[\operatorname{Log}\Psi(z) - \operatorname{Log}(z-a)] + \alpha \operatorname{Log}(z-a)} \\ &= (z-a)^\alpha e^{\alpha L(z)} . \end{aligned}$$

Note that the function f_0 defined by $f_0(z) = e^{\alpha L(z)}$ is actually analytic in the whole disk $\Delta(a,r)$, where it has no zeros. The purpose of these rather drawn out deliberations was to arrive at the representation $f(z) = (z-a)^\alpha f_0(z)$ for the function f in $H \cap \Delta(a,r)$. It leads by direct calculation to a representation for the function h in $H \cap \Delta(a,r)$ — namely,

$$h(z) = \frac{f''(z)}{f'(z)} - \frac{\alpha-1}{z-a} = \frac{(\alpha+1)f_0'(z) + (z-a)f_0''(z)}{\alpha f_0(z) + (z-a)f_0'(z)} .$$

Since the denominator in the last expression does not vanish at the point a, we can choose δ in $(0,r)$ so that the function defined by this formula is analytic in the disk $\Delta(a,\delta)$. Because that function and h coincide in $H \cap \Delta(a,r)$ and because both are analytic in $\Delta^*(a,\delta)$, they must, according to the principle of analytic continuation, agree everywhere in $\Delta^*(a,\delta)$. This completes the verification of (9.73) and, with it, the first half of the proof.

It remains to pin down the behavior of $g(z)$ as $z \to \infty$. For this, we consider the function $f_1 : H \to \mathbb{C}$ given by $f_1(z) = f(-z^{-1})$. (N.B. The relation $f(z) = f_1(-z^{-1})$ is equally valid.) Then f_1 is a second conformal mapping of H onto D, one satisfying $\widetilde{f}_1(0) = \widetilde{f}(\infty)$. Let g_1 be constructed from f_1 in the same way g was from f. Straightforward computation yields

$$g(z) = \frac{f''(z)}{f'(z)} = -\frac{2}{z} + \frac{f_1''(-z^{-1})}{z^2 f_1'(-z^{-1})}$$

for z in H; i.e., we conclude that in H

$$g(z) = -\frac{2}{z} + \frac{1}{z^2} g_1\left(-\frac{1}{z}\right) . \tag{9.74}$$

Each side of (9.74) describes a function that is analytic in the domain $U = \mathbb{C} \sim \{0, a_1, \ldots, a_n\}$. As these functions are the same in H, Corollary VIII.1.6 confirms that (9.74) remains valid throughout U. Two cases must now be distinguished. First, if $\widetilde{f}_1(0)$ is not a vertex of P, then g_1 is analytic in the vicinity of the origin, so $g_1(-z^{-1}) \to g_1(0) \neq \infty$ as $z \to \infty$. Coupled with (9.74) this implies that $g(z) \to 0$ as $z \to \infty$. If, on the other hand, $\widetilde{f}_1(0)$ happens to be a vertex of P, then by the first part of the lemma g_1 has a simple pole at the origin. As a result, $zg_1(z) \to \ell$, a finite limit, as $z \to 0$. Thus $z^{-1}g_1(-z^{-1}) \to -\ell$ as $z \to \infty$. In this instance, too, it follows from (9.74) that $g(z) \to 0$ as $z \to \infty$. ∎

We now come to the main result of this section.

Theorem 5.6. (Schwarz-Christoffel Formula) *Let f be a conformal mapping of the half-plane $H = \{z : \operatorname{Im} z > 0\}$ onto the interior D of a closed polygon P in the complex plane, let $-\infty < a_1 < a_2 < \cdots < a_n \leq \infty$ be the list of points transformed to the vertices of P by the homeomorphic extension $\widetilde{f}$ of f to $\widehat{H}$, and let $\alpha_j\pi$ be the interior angle of P at the vertex $\widetilde{f}(a_j)$. There exist constants A and B such that for every z in H*

$$(9.75) \qquad f(z) = A\int_i^z (\zeta - a_1)^{\alpha_1 - 1}(\zeta - a_2)^{\alpha_2 - 1}\cdots(\zeta - a_n)^{\alpha_n - 1}\, d\zeta + B$$

if $a_n \neq \infty$, while

$$(9.76) \qquad f(z) = A\int_i^z (\zeta - a_1)^{\alpha_1 - 1}(\zeta - a_2)^{\alpha_2 - 1}\cdots(\zeta - a_{n-1})^{\alpha_{n-1} - 1}\, d\zeta + B$$

if $a_n = \infty$.

Proof. Let g be the function constructed from f as in Lemma 5.4. We consider the function h defined in the domain-set G of g by

$$h(z) = g(z) - \sum_{j=1}^{n} \frac{\alpha_j - 1}{z - a_j}$$

if $a_n \neq \infty$, and

$$h(z) = g(z) - \sum_{j=1}^{n-1} \frac{\alpha_j - 1}{z - a_j}$$

if $a_n = \infty$. Then h is analytic in G. On the basis of Lemma 5.5 we can assert that the singularities of h at the points $a_1, a_2, \ldots, a_n$, and ∞ are all removable. We remove them and thereby turn h into a holomorphic function on $\widehat{\mathbb{C}}$. By Theorem VIII.4.2, h is constant on $\widehat{\mathbb{C}}$. Since $h(\infty) = 0$, we conclude that $h(z) = 0$ for every z in $\mathbb{C}$. We arrive in this way at an explicit formula for g.

In the event that $a_n \neq \infty$, we choose a branch ℓ of $\log f'(z)$ in H and observe that

$$\ell'(z) = \frac{f''(z)}{f'(z)} = g(z) = \sum_{j=1}^{n} \frac{\alpha_j - 1}{z - a_j} = \frac{d}{dz}\sum_{j=1}^{n}(\alpha_j - 1)\operatorname{Log}(z - a_j)$$

in this half-plane. (Such a function ℓ exists because f' is analytic and zero-free in H.) In H, therefore,

$$\ell(z) = \sum_{j=1}^{n} (\alpha_j - 1)\operatorname{Log}(z - a_j) + C$$

for some constant C. Exponentiation results in

$$f'(z) = e^{\ell(z)} = A(z-a_1)^{\alpha_1 - 1}(z-a_2)^{\alpha_2-1}\cdots(z-a_n)^{\alpha_n-1},$$

where $A = e^C$. Integration then produces (9.75) with $B = f(i)$. The argument in the case $a_n = \infty$ involves only minute changes and leads to (9.76) in place of (9.75). ∎

We accent the fact that all powers occurring in (9.75) and (9.76) are principal powers. Since the integrands in these formulas are continuous functions in $\overline{H} \sim \{a_1, a_2, \ldots, a_n\}$, it follows without any fuss that the integrals which appear there evaluate to $\tilde{f}(z)$ for every point z of the set $\mathbb{R} \sim \{a_1, a_2, \ldots, a_n\}$. As a matter of fact, even more is true: if $z = a_j$ for some j or if $z = \infty$, then the corresponding integral on the right-hand side of (9.75) or (9.76) still has the value $\tilde{f}(z)$, although the integral in question becomes improper when $z = a_j$ for a value of j such that the corresponding α_j is less than 1 or when $z = \infty$. (In the latter case use $\gamma(t) = ti$, $1 \le t \le \infty$, for the path of integration.) The choice of i as the initial point of integration in (9.75) and (9.76) was quite arbitrary. Any other point of $\overline{H}$ would serve just as well, the only thing affected by a change being the constant B. For instance, one frequently encounters these formulas rewritten as

$$(9.77)\quad f(z) = A\int_0^z (\zeta - a_1)^{\alpha_1-1}(\zeta-a_2)^{\alpha_2-1}\cdots(\zeta-a_n)^{\alpha_n-1}\,d\zeta + B$$

when $a_n \ne \infty$, and

$$(9.78)\quad f(z) = A\int_0^z (\zeta - a_1)^{\alpha_1-1}(\zeta-a_2)^{\alpha_2-1}\cdots(\zeta-a_{n-1})^{\alpha_{n-1}-1}\,d\zeta + B$$

when $a_n = \infty$. It should be pointed out that the integrals which turn up in conjunction with the Schwarz-Christoffel formula can seldom be evaluated explicitly in terms of elementary functions. They are, however, amenable to numerical techniques of integration, a fact that frees the way to numerical approximation of the conformal mapping f.

Theorem 5.6 actually covers a situation somewhat more general than we claimed. Apart from a bit of tinkering with details, the proof that we've given can be used to establish the validity of the Schwarz-Christoffel formula for a conformal mapping f of H onto any bounded domain D whose boundary ∂D is the trajectory of a closed — but not necessarily simple — polygonal path $\gamma = [z_1, z_2] + [z_2, z_3] + \cdots + [z_{n-1}, z_n] + [z_n, z_1]$ with the property that for $j = 1, 2, \ldots, n-1$ the points z_j, z_{j+1}, and z_{j+2} are not collinear. (As earlier, put $z_{n+1} = z_1$.) This description permits points of ∂D to be "repeated vertices." The domain pictured in Figure 37 offers an example. In this more general setting f still admits a continuous extension

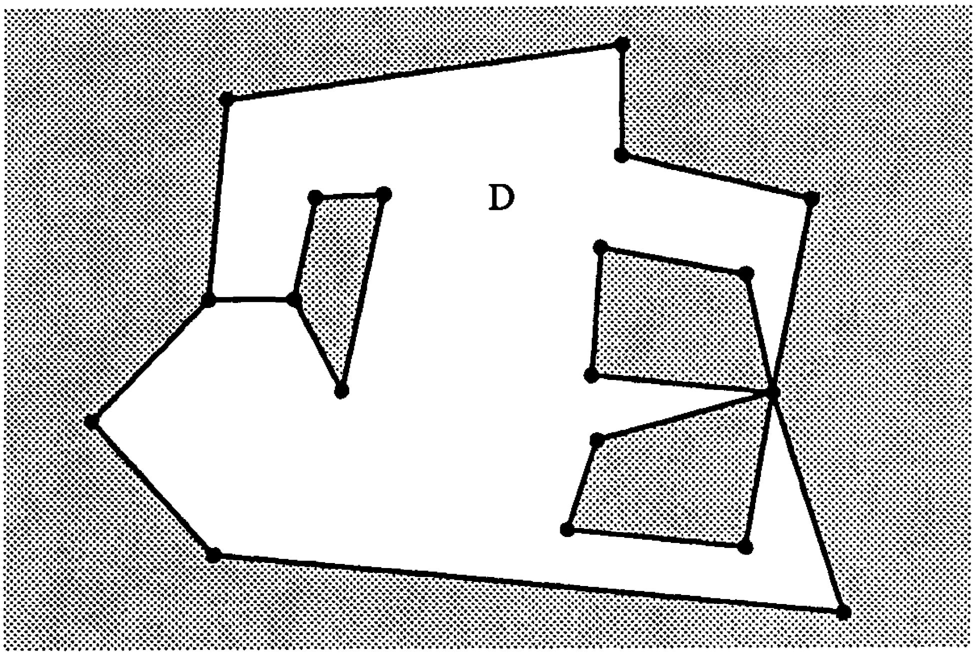


Figure 37.

$\tilde{f}$ to $\hat{H}$, though $\tilde{f}$ will not be a homeomorphism in the presence of repeated vertices, and we can once more list the points mapped by $\tilde{f}$ to the vertices in the manner $-\infty < a_1 < a_2 < \cdots < a_n \leq \infty$. We continue to express the angle at $\tilde{f}(a_j)$ interior to D as $\alpha_j \pi$. Should $\tilde{f}(a_j)$ happen to be a repeated vertex, we must take care that $\alpha_j \pi$ measures the particular interior angle corresponding to a_j, meaning the one whose sides are the images under $\tilde{f}$ of $[a_{j-1}, a_j]$ and $[a_j, a_{j+1}]$. (So that this will make sense for $j = 1$ and $j = n$, we set $[a_0, a_1] = [-\infty, a_1]$ and agree that $[a_n, a_{n+1}] = [a_n, \infty]$ if $a_n \neq \infty$, while $[a_n, a_{n+1}] = [-\infty, a_1]$ if $a_n = \infty$.) Subject to these provisions Theorem 5.6 remains true as it stands for mappings to such generalized polygons.

While Theorem 5.6 does accurately convey the structure of a conformal mapping f of H onto the interior of a given closed polygon P, it by no means hands one a foolproof method for describing f explicitly, even in integral form. A snag often develops in trying to get a hold on the points $a_1, a_2, \ldots, a_n$ that are mapped by $\tilde{f}$ to the vertices of P, something we must certainly do to make the Schwarz-Christoffel formula truly explicit. (After these points are identified, the computation of the constants A and B is usually quite manageable.) In light of Theorem 4.12 we are at liberty in constructing such a mapping f to specify the points that $\tilde{f}$ transforms to any three selected vertices of P, modulo constraints imposed by orientation. As an example of this kind of normalization, a triple $(\zeta_0, \zeta_1, \zeta_\infty)$ of consecutive vertices of P can be chosen, one that displays the positive orientation relative to the interior of P, and it can then be required that

$\tilde{f}$ take 0 to ζ_0, 1 to ζ_1, and ∞ to ζ_∞. For this mapping f we thus have $a_n = \infty$, $a_{n-1} = 1$, and $a_{n-2} = 0$. With that all flexibility in the matter comes to an end, for the values of $a_1, a_2, \ldots, a_{n-3}$ are dictated by the choices of a_{n-2}, a_{n-1}, and a_n. It can be a formidable problem to ascertain the values of these remaining a_j. Indeed, unless P is blessed with a considerable amount of symmetry, their exact determination is rarely possible. The role of symmetry is exemplified by the construction of a conformal mapping of H to a square in Example 5.2. Before this, we consider a more straightforward application of the Schwarz-Christoffel formula.

EXAMPLE 5.1. Find a conformal mapping of $H = \{z : \operatorname{Im} z > 0\}$ onto the interior D of the closed triangle T with vertices 0, 1, and ic, where $\operatorname{Im} c > 0$ (Figure 38).

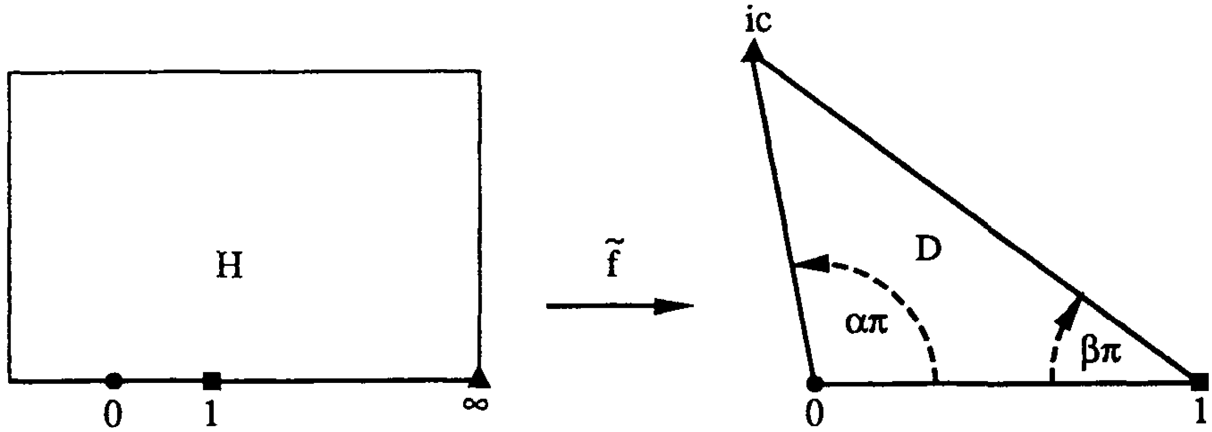


Figure 38.

Let $\alpha\pi$ and $\beta\pi$ be the interior angles of T at the vertices 0 and 1, respectively. We shall obtain the conformal mapping f of H onto D whose homeomorphic extension $\tilde{f}$ to $\widehat{H}$ sends 0 to 0, 1 to 1, and ∞ to ic. In view of (9.78) and other comments made in the aftermath of Theorem 5.6, we can express $\tilde{f}(z)$ for z belonging to $\widehat{H}$ in integral form,

$$\tilde{f}(z) = A \int_0^z \zeta^{\alpha-1}(\zeta - 1)^{\beta-1}\, d\zeta + B ,$$

for certain constants A and B. Since $\tilde{f}(0) = 0$ and since the integral here tends to zero when z does, we must have $B = 0$; since $\tilde{f}(1) = 1$, we then see that

$$A = \frac{1}{\int_0^1 \zeta^{\alpha-1}(\zeta - 1)^{\beta-1}\, d\zeta} .$$

Therefore, one conformal mapping of H onto D is given by

$$f(z) = \frac{\int_0^z \zeta^{\alpha-1}(\zeta - 1)^{\beta-1}\, d\zeta}{\int_0^1 \zeta^{\alpha-1}(\zeta - 1)^{\beta-1}\, d\zeta} .$$

Mapping H conformally to a square offers more of a challenge.

EXAMPLE 5.2. Determine a conformal mapping of $H = \{z : \operatorname{Im} z > 0\}$ onto the interior D of the closed square Q with vertices 0, 1, $1+i$, and i.

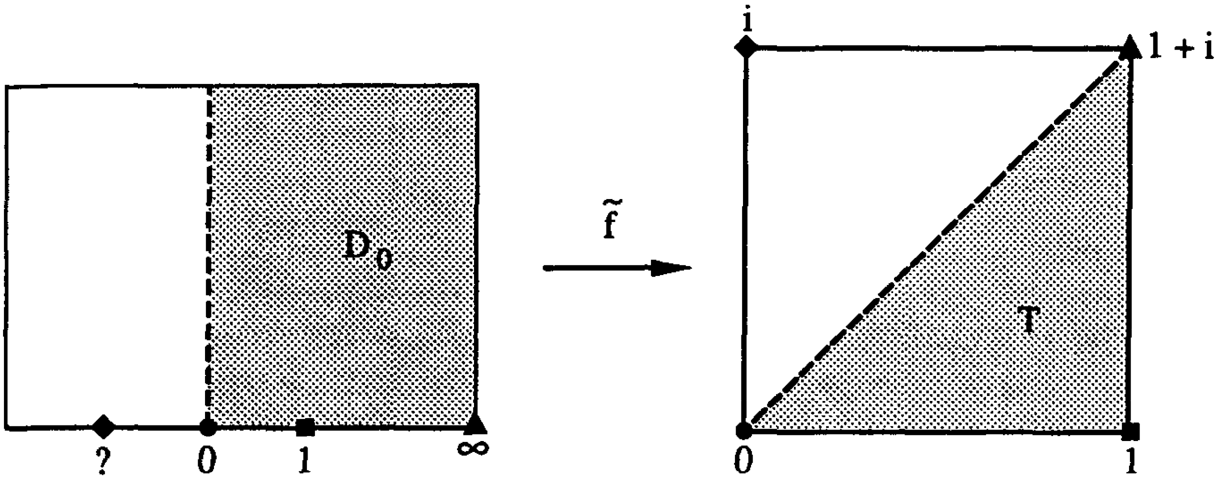


Figure 39.

We shall use Theorem 5.6 to identify the unique conformal mapping f of H onto D whose homeomorphic extension $\widetilde{f}$ to $\widehat{H}$ satisfies $\widetilde{f}(0) = 0$, $\widetilde{f}(1) = 1$, and $\widetilde{f}(\infty) = 1 + i$. Before we can apply the theorem, however, we must first settle the question: Which point of the interval $(-\infty, 0)$ gets mapped by $\widetilde{f}$ to i? We profit by the symmetry of Q to arrive at the answer. Consider the quarter plane $D_0 = \{z : \operatorname{Re} z > 0, \operatorname{Im} z > 0\}$ and the triangle T with vertices 0,1, and $1+i$ (Figure 39). Let g be the unique conformal mapping of D_0 onto the interior of T whose homeomorphic extension $\widetilde{g}$ to $\widehat{D}_0$ has $\widetilde{g}(0) = 0$, $\widetilde{g}(1) = 1$, and $\widetilde{g}(\infty) = 1+i$. The mapping $\widetilde{g}$ transforms the positive imaginary axis I to the diagonal of Q with endpoints 0 and $1+i$. We can continue $\widetilde{g}$ across I by reflection. Because a square is symmetric about its diagonals, the reflection process generates a homeomorphism G of $\widehat{H}$ onto Q that carries H conformally onto D and satisfies $G(0) = 0$, $G(1) = 1$, $G(\infty) = 1 + i$, and $G(-1) = i$. The uniqueness statement in Theorem 4.12 makes it clear that $G = \widetilde{f}$. In particular, we learn that $\widetilde{f}(-1) = i$. We are now able to invoke (9.78) and obtain an integral representation for $\widetilde{f}$:

$$\widetilde{f}(z) = A \int_0^z \zeta^{-1/2}(\zeta - 1)^{-1/2}(\zeta + 1)^{-1/2}\, d\zeta + B$$

for z in $\widehat{H}$. The conditions $\widetilde{f}(0) = 0$ and $\widetilde{f}(1) = 1$ allow us to evaluate the constants A and B. This results in the identification of

$$f(z) = \frac{\int_0^z \zeta^{-1/2}(\zeta - 1)^{-1/2}(\zeta + 1)^{-1/2}\, d\zeta}{\int_0^1 \zeta^{-1/2}(\zeta - 1)^{-1/2}(\zeta + 1)^{-1/2}\, d\zeta}$$

as a conformal mapping mapping of H onto D.

Theorem 5.6 admits a multitude of extensions and refinements. As most of these require for their development ideas that go well beyond the elementary arguments which led to Theorem 5.6, we say nothing about them here. (Some can be found in the book of Nehari mentioned at the opening of this chapter.) The one exception we make in this regard is the derivation of an integral representation for a conformal mapping of the disk $\Delta(0,1)$ onto the interior of a polygon — not barring one with repeated vertices — since this representation can be extracted painlessly enough from Theorem 5.6.

Theorem 5.7. *Let f be a conformal mapping of the disk $\Delta = \Delta(0,1)$ onto the interior D of a closed polygon P in the complex plane, let $b_1, b_2, \ldots, b_n$ be a list of the points of $\partial\Delta$ that are transformed to the vertices of P by $\tilde{f}$, the continuous extension of f to $\overline{\Delta}$, and let $\alpha_j\pi$ be the interior angle of P at the vertex $\tilde{f}(b_j)$. There exist constants A and B such that*

$$f(z) = A\int_0^z \left(\frac{b_1-\zeta}{1+\zeta}\right)^{\alpha_1-1}\left(\frac{b_2-\zeta}{1+\zeta}\right)^{\alpha_2-1}\cdots\left(\frac{b_n-\zeta}{1+\zeta}\right)^{\alpha_n-1}\frac{d\zeta}{(1+\zeta)^2}+B$$

for every z in Δ.

Proof. For convenience in the proof we assume the labeling of vertices has been arranged so that $-\pi < \operatorname{Arg} b_1 < \operatorname{Arg} b_2 < \cdots < \operatorname{Arg} b_n \le \pi$. Noting that the Möbius transformation $\varphi(z) = (1+iz)/(1-iz)$ maps the upper half-plane $H = \{z : \operatorname{Im} z > 0\}$ onto Δ, we see that $g = f\circ\varphi$ delivers a conformal mapping of H onto D. Since $\varphi^{-1}(z) = i[(1-z)/(1+z)]$, the points mapped to the vertices of P under the extension $\tilde{g} = \tilde{f}\circ\varphi$ of g to $\widehat{H}$ are the points $-\infty < a_1 < a_2 < \cdots < a_n \le \infty$ given by

$$a_j = \varphi^{-1}(b_j) = i\left(\frac{1-b_j}{1+b_j}\right).$$

Also, from the relation $f = g\circ\varphi^{-1}$ we infer that

$$f'(z) = g'[\varphi^{-1}(z)][\varphi^{-1}]'(z) = -\frac{2ig'[\varphi^{-1}(z)]}{(1+z)^2}. \tag{9.79}$$

Suppose initially that $\operatorname{Arg} b_n < \pi$; i.e., $b_n \ne -1$. Then $a_n \ne \varphi^{-1}(-1) = \infty$. On the strength of Theorem 5.6 we can assert that g' takes the form

$$g'(z) = A_0(z-a_1)^{\alpha_1-1}(z-a_2)^{\alpha_2-1}\cdots(z-a_n)^{\alpha_n-1} \tag{9.80}$$

for some constant A_0.

We now fix j. A simple calculation reveals that

$$\varphi^{-1}(z) - a_j = \frac{2i(b_j-z)}{(1+b_j)(1+z)}.$$

The Möbius transformation $\psi(z) = (b_j - z)/(1+z)$ maps the circle $K = K(0,1)$ to $\tilde{K} = L\cup\{\infty\}$, where L is a line through the origin. Since 0 and

∞ are symmetric with respect to K, $b_j = \psi(0)$ and $-1 = \psi(\infty)$ must be symmetric with respect to $\widetilde{K}$. This remark enables us to identify L as the perpendicular bisector of the segment joining b_j and -1. It follows easily that $\psi(\Delta)$ is one of the two open half-planes bounded by L — namely, the one that contains b_j. Since the complementary half-plane contains the real interval $(-\infty, 0]$, $\psi(\Delta)$ is disjoint from $(-\infty, 0]$. The consequence of this observation is that $\ell(z) = \text{Log}[(b_j - z)/(1+z)]$ defines a function that is analytic in Δ. The function ℓ_0 given by $\ell_0(z) = \text{Log}[\varphi^{-1}(z) - a_j]$ is likewise analytic in Δ. Furthermore, in this disk

$$e^{\ell_0(z)-\ell(z)} = \frac{[\varphi^{-1}(z) - a_j](1+z)}{(b_j - z)} = \frac{2i}{1+b_j} ,$$

a constant, from which one concludes without difficulty that ℓ_0 and ℓ also differ by a constant in Δ, say $\ell_0(z) = \ell(z) + \gamma_j$. As a result,

$$[\varphi^{-1}(z) - a_j]^{\alpha_j - 1} = e^{(\alpha_j - 1)\ell_0(z)} = e^{(\alpha_j - 1)\ell(z) + (\alpha_j - 1)\gamma_j}$$

$$= c_j \left(\frac{b_j - z}{1+z}\right)^{\alpha_j - 1}$$

for every z in Δ, where $c_j = e^{(\alpha_j - 1)\gamma_j}$. This is true for $j = 1, 2, \ldots, n$.

Referring to (9.79) and (9.80) we come to the conclusion that, at least when $b_n \neq -1$,

$$(9.81) \quad f'(z) = \frac{A}{(1+z)^2} \left(\frac{b_1 - z}{1+z}\right)^{\alpha_1 - 1} \left(\frac{b_2 - z}{1+z}\right)^{\alpha_2 - 1} \cdots \left(\frac{b_n - z}{1+z}\right)^{\alpha_n - 1}$$

for every z in Δ, where A is some constant. Should $b_n = -1$, Theorem 5.6 would give rise to an expression for $f'(z)$ similar to (9.81), but with the last factor in the product deleted. As a matter of fact, however, (9.81) still applies when $b_n = -1$, for in that event the last factor reduces to $(-1)^{\alpha_n - 1}$, a constant that can be absorbed into A. Integration of (9.81) results in the stated formula for f with $B = f(0)$. ∎

We again stress that the powers appearing in Theorem 5.7 are principal powers. In many books the conclusion of this result is reported as

$$(9.82) \quad f(z) = A \int_0^z (b_1 - \zeta)^{\alpha_1 - 1} (b_2 - \zeta)^{\alpha_2 - 1} \cdots (b_n - \zeta)^{\alpha_n - 1} \, d\zeta + B ,$$

a formula obtained when, with the aid of (9.69), one formally simplifies the integrand found in Theorem 5.7. Unfortunately, if the powers present in (9.82) are interpreted as principal powers, consistent with the notational conventions in force in this book, the integrand in this formula may have discontinuities in Δ. Formula (9.82) is valid, of course, but only with the understanding that the notation $(b_j - z)^{\alpha_j - 1}$ is merely being employed as shorthand to indicate some branch of the $(\alpha_j - 1)$-power of the analytic function $p_j(z) = b_j - z$ in Δ.

6 Exercises for Chapter IX

6.1 Exercises for Section IX.1

6.1. Assuming that $f: U \to V$ belongs to the class $C^1(U)$ and $g: V \to \mathbb{C}$ is a member of $C^1(V)$, check that $J_{g \circ f}(z) = J_g[f(z)]J_f(z)$ holds for every point z of U.

6.2. Let $f: \mathbb{C} \to \mathbb{C}$ be defined by $f(z) = x^3 + iy^3$. Is f a univalent C^1-function? Is f a diffeomorphism?

6.3. Let c and d be complex numbers such that $|c| \neq |d|$. Show that $f(z) = cz + d\bar{z}$ defines a diffeomorphism of $\mathbb{C}$ onto itself, and determine f^{-1}. Under what conditions on c and d is f orientation-preserving?

6.4. If $D = \mathbb{C} \sim \{0\}$, show that for any non-zero real number λ the function $f(z) = |z|^{\lambda - 1} z$ furnishes a diffeomorphism of D onto itself. For which λ is f orientation-preserving? What is f^{-1}? (*Hint.* Use (9.3) in computing J_f.)

6.5. Working from the definitions of conformality and anti-conformality, verify that a sense-preserving similarity transformation $f(z) = az + b$ gives a conformal mapping of the complex plane onto itself, but that a sense-reversing similarity $g(z) = a\bar{z} + b$ maps the plane anti-conformally onto itself.

6.6. Let $f: D \to \mathbb{C}$ be a diffeomorphism. Confirm that f is anti-conformal at a point z_0 of D if and only if its conjugate $\bar{f}$ is conformal at that point. Conclude that f is an anti-conformal mapping of D if and only if $\bar{f}$ is a conformal mapping of D.

6.7. For $j = 1, 2, \ldots, n$ let $f_j: D_j \to \mathbb{C}$ be either a conformal or an anti-conformal mapping of a domain D_j. Assuming that $f_j(D_j)$ is contained in D_{j+1} for $j = 1, 2, \ldots, n-1$, check that the composition $f = f_n \circ f_{n-1} \circ \cdots \circ f_1$ is likewise either conformal or anti-conformal. Show that f is conformal if and only if the number of anti-conformal factors in its make-up is even.

6.8. Let D be a domain in the complex plane, and let $f: D \to \mathbb{C}$ be a diffeomorphism. Prove that the following statements concerning a point z_0 of D are equivalent: (i) $\bar{f}$ is differentiable at z_0; (ii) f is isogonal at z_0 and $J_f(z_0) < 0$; (iii) f is anti-conformal at z_0.

6.9. Let $f: D \to \mathbb{C}$ be a diffeomorphism, and let z_0 be a point of D. Given the knowledge that $\lim_{z \to z_0} |f(z) - f(z_0)|/|z - z_0|$ exists, deduce that f is either conformal or anti-conformal at z_0. (*Hint.* Exercise III.6.65 is pertinent to this exercise.)

6.10. Let $D = \mathbb{C} \sim (-\infty, 0]$ and let $1 < \lambda < \infty$. Show that the function $f: D \to \mathbb{C}$ defined by $f(z) = z^\lambda$ is not a univalent function, but is a locally conformal mapping of D. What is its range?

6.11. Find a conformal mapping of the half-plane $D = \{z : x > 0\}$ onto the domain $D' = \{w : |\operatorname{Arg} w| < \lambda\pi\}$, where $0 < \lambda \leq 1$.

6.12. Construct a conformal mapping of the domain $D = \{z : x > |y|\}$ onto the disk $D' = \Delta(0,1)$.

6.13. Determine a conformal mapping of the semi-infinite strip $D = \{z : x < 0, 0 < y < 1\}$ onto the quarter-disk $D' = \{w : |w| < 1, v > |u|\}$.

6.14. Construct a conformal mapping of $D = \{z : |x| + |y| < \sqrt{2}/2\}$ onto the domain $D' = \{w : 1 < |w| < e^{\pi}, v > 0\}$.

6.15. Find a conformal mapping of the strip $D = \{z : |\operatorname{Re} z| < \pi/2\}$ onto itself that transforms the real interval $(-\pi/2, \pi/2)$ to the full imaginary axis.

6.16. Build a conformal mapping of the disk $D = \Delta(0,1)$ onto the domain $D' = \{w : u^2 - v^2 < 1\}$. (*Hint.* Remember Example III.3.7.)

6.17. Exhibit a conformal mapping of the disk $D = \Delta(0,1)$ onto the domain D' inside the cardioid whose polar equation is $r = 1 + \cos\theta$. (*Hint.* Recall Exercises I.4.55 and VIII.5.75.)

6.18. Find a conformal mapping of the half-plane $D = \{z : y > 0\}$ onto the domain $D' = \{w : u < v^2\}$. (*Hint.* Recall Exercise I.4.53.)

6.19. Display a conformal mapping of the strip $D = \{z : |y| < \pi/2\}$ onto the domain $D' = \{w : v < u^2\}$.

6.20. Construct a conformal mapping f of the disk $D = \Delta(0,1)$ onto the domain $D' = \mathbb{C} \sim (-\infty, -1/4]$ with $f(0) = 0$ and $f'(0) = 1$.

6.21. Let $D = \Delta(0,1)$, and let D' be the component of the open set $\{w : |w^2 - 1| < 1\}$ that contains the point -1. Find a conformal mapping of D onto D'. (*Hint.* Recall Exercise V.8.40.)

6.22. Let $a > 0$ and $b > 0$. Determine a conformal mapping of the domain $D = \Delta(\infty, 1)$ onto $D' = \widehat{\mathbb{C}} \sim \{w : (u/a)^2 + (v/b)^2 \leq 1\}$. (*Hint.* The function $f(z) = cz + dz^{-1}$ for properly chosen c and d will do the job.)

6.2 Exercises for Section IX.2

6.23. If $f(z) = z/(z+1)$, calculate $g = f \circ f \circ \cdots \circ f$ (n terms), the n^{th} iterate of f. (*Hint:* $f = f_A$ for $A = ?$)

6.24. Confirm that under the correspondence $A \to f_A$ of 2×2 non-singular matrices to Möbius transformations it is true that $AB \to f_A \circ f_B$ and $A^{-1} \to (f_A)^{-1}$.

6.25. Let A and B be 2×2 matrices such that $\det A = \det B = 1$ and $f_A = f_B$. Show that either $A = B$ or $A = -B$.

6.26. Express f in normalized form, and identify its fixed points: (i) $f(z) =$

$(z-3i)/(iz-1)$; (ii) $f(z)=(4z+i)/(4iz)$; (iii) $f(z)=(3z-4i)/(5iz+7)$.

6.27. Check that, when written in normalized form, a Möbius transformation f that maps the disk $\Delta(0,1)$ onto itself has the appearance $f(z)=(az+b)/(\bar{b}z+\bar{a})$, where $|a|^2-|b|^2=1$.

6.28. Calculate the cross-ratios: (i) $[1,i,-1,-i]$; (ii) $[1,-1,i,-i]$; (iii) $[0,\infty,1+i,1-i]$; (iv) $[0,a,b,\infty]$ for finite non-zero a and b, $a\neq b$; (v) $[c,c^{-1},1,0]$ for finite $c\neq 0,\pm1$; (vi) $[z,z^{-1},\bar{z},\infty]$ for finite, non-real z, $|z|\neq 1$; (vii) $[-1,1,w,w^{-1}]$ for $w\neq\pm1$; (viii) $[e^{i\theta},-e^{-i\theta},-e^{i\theta},e^{-i\theta}]$ for $\theta\in(0,\pi/2)$.

6.29. Let $\lambda=[z_1,z_2,z_3,z_4]$. Express in terms of λ the various cross-ratios that arise when the points z_1,z_2,z_3, and z_4 are permuted. (N.B. There are twenty-four possible permutations to deal with, but only six different values turn up among the associated cross-ratios. *Hint.* It suffices to consider $z_1=1,z_2=0$, and $z_3=\infty$. Why?)

6.30. Substantiate the following claim: four distinct points z_1,z_2,z_3, and z_4 in $\widehat{\mathbb{C}}$ lie on a circle in $\widehat{\mathbb{C}}$ if and only if the cross-ratio $[z_1,z_2,z_3,z_4]$ is real. (*Hint.* Reduce the problem to the case $z_1=1,z_2=0$, and $z_3=\infty$.)

6.31. Refine the preceding exercise as follows: four distinct points z_1,z_2,z_3, and z_4 in $\widehat{\mathbb{C}}$ lie on a circle in $\widehat{\mathbb{C}}$ and are so arranged on that circle that z_1 and z_3 separate z_2 and z_4 if and only if the condition

$$|[z_1,z_3,z_2,z_4]|+|[z_3,z_1,z_2,z_4]|=1$$

is fulfilled. (*Hint.* What does this condition boil down to when $z_1=0,z_3=1$, and $z_4=\infty$?)

6.32. Let z_1,z_2,z_3, and z_4 be distinct points of $\widehat{\mathbb{C}}$, finite except possibly for z_4, let K be the circle in $\widehat{\mathbb{C}}$ on which z_1,z_2, and z_4 lie, and let $\widetilde{K}$ be the circle in $\widehat{\mathbb{C}}$ that passes through z_1,z_3, and z_4. Confirm that $\operatorname{Arg}([z_1,z_2,z_3,z_4])=\theta(A,\widetilde{A})$, where A is the arc of K from z_1 to z_2 that does not contain z_4 and $\widetilde{A}$ is the arc of $\widetilde{K}$ from z_1 to z_3 that misses z_4.

6.33. Construct a Möbius transformation f with the specified effect: (i) f sends i to 1, 1 to ∞, and ∞ to i ; (ii) f transforms $K=K(i,1)$ to $\widetilde{K}=\{w:(1+i)w+(1-i)\bar{w}=0\}$; (iii) f takes $K=\mathbb{R}\cup\{\infty\}$ to $\widetilde{K}=K(0,1)$, leaving the point -1 fixed; (iv) f maps $D=\Delta(0,1)$ to $D'=\{w:\operatorname{Im}w>0\}$, with $f(0)=1+2i$; (v) f maps $K(0,1)$ to itself and $K(1/4,1/4)$ to $K(0,r)$ for some $r<1$.

6.34. Let $f(z)=(az+b)/(cz+d)$ be a normalized Möbius transformation for which $c\neq 0$. The circle $K(f)=\{z:|cz+d|=1\}$ is called the *isometric circle* of f for the reason that f preserves the distance between points of $K(f)$ — i.e., $|f(z_2)-f(z_1)|=|z_2-z_1|$ whenever z_1 and z_2 lie on $K(f)$. Certify that this is so. Show also that f transforms $K(f)$ to $K(f^{-1})$. To what set does f map the disk inside of $K(f)$?

6.35. Let K be a circle in $\widehat{\mathbb{C}}$. Certify that the reflection ρ_K defined by (9.23) (respectively, (9.24)) has the geometric effect described in the text.

6.36. How do cross-ratios behave under anti-Möbius transformations?

6.37. Show that the fixed point set of an anti-Möbius transformation f either contains at most two points or is a circle K in $\widehat{\mathbb{C}}$. Furthermore, prove that the latter occurs if and only if $f = \rho_K$, the reflection in K.

6.38. Let K be a circle in $\widehat{\mathbb{C}}$, and let z and w be different points of $\widehat{\mathbb{C}} \sim K$. Prove that z and w are symmetric with respect to K if and only if every circle in $\widehat{\mathbb{C}}$ that passes through both z and w is perpendicular to K. (*Hint.* What is the situation when $K = \mathbb{R} \cup \{\infty\}$?)

6.39. If z_1, z_2, and z_3 are distinct points of $\widehat{\mathbb{C}}$, verify that through z_1 passes one and only one circle in $\widehat{\mathbb{C}}$ with respect to which z_2 and z_3 are symmetric.

6.40. Let K and K' be circles in $\widehat{\mathbb{C}}$. If z_0 belongs to K, z_0' to K', w_0 to $\widehat{\mathbb{C}} \sim K$, and w_0' to $\widehat{\mathbb{C}} \sim K'$, then there is a unique Möbius transformation that maps K to K', z_0 to z_0', and w_0 to w_0'. Verify this.

6.41. Let $f(z) = (az+b)/(cz+d)$ be a normalized Möbius transformation, and let $H = \{z : \operatorname{Im} z > 0\}$. Demonstrate that $f(H) = H$ if and only if a, b, c, and d are real numbers. (*Hint.* For the sufficiency, look at $\operatorname{Im} f(z)$. For the necessity, consider $g(z) = \overline{f(\overline{z})}$.)

6.42. Show that the conformal self-mappings f of the half-plane $H = \{z : \operatorname{Im} z > 0\}$ are precisely the restrictions to H of the Möbius transformations that, when written in normalized form, have real coefficients. (*Hint.* Let $f: H \to H$ be a conformal mapping of H onto itself. Show first that f is the restriction to H of a Möbius transformation by considering $g = h^{-1} \circ f \circ h$ in $\Delta(0,1)$, where $h(z) = (z+i)/(iz+1)$.)

6.43. Demonstrate that any univalent function $f: \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ that preserves the class of circles in $\widehat{\mathbb{C}}$ — i.e., if K is a circle in $\widehat{\mathbb{C}}$, then so is $f(K)$ — is either a Möbius transformation or an anti-Möbius transformation. (*Hint.* Start by looking at a function f of the above kind that also fixes $0, 1$, and ∞. Verify that f is either the identity mapping of $\widehat{\mathbb{C}}$ or the reflection in the real axis. Try first to pin down the images under f of the real and imaginary axes, the line with equation $x = 1$, and the circle $K(1/2, 1/2)$. Assume, if need be, that f is continuous.)

6.44. Let S be the standard unit sphere in $\mathbb{R}^3$, and let $\pi: S \to \widehat{\mathbb{C}}$ be stereographic projection. If $F: S \to S$ is the mapping created by reflecting S in a plane P that passes through its center, then the mapping $f = \pi \circ F \circ \pi^{-1}$ of $\widehat{\mathbb{C}}$ is just the reflection ρ_K, where $K = \pi(S \cap P)$. Prove this. (*Hint.* Use Exercises 6.41 and 6.37 along with Exercise VIII.5.81. Notice that f could not be a Möbius transformation. Why not?)

6.45. As in the preceding exercise, let $\pi: S \to \widehat{\mathbb{C}}$ be the stereographic projection, but now let $F: S \to S$ be the antipodal map; i.e., $F(p) = -p$. Check that $f = \pi \circ F \circ \pi^{-1}$ is another anti-Möbius transformation — namely, $f(z) = -1/\bar{z}$.

6.46. Express f in normalized form and classify it; if f is non-parabolic, represent it in multiplier-fixed point format: (i) $f(z) = (z+i)/(iz+1)$; (ii) $f(z) = (2z+1)/(2z+3)$; (iii) $f(z) = (2z-1)/(2z-3)$; (iv) $f(z) = z/(iz+1)$.

6.47. Confirm that a Möbius transformation f is parabolic if and only if it is conjugate to the translation $g(z) = z + 1$, that f is hyperbolic if and only if it is conjugate to a non-trivial dilation with respect to the origin, and that f is elliptic if and only if it is conjugate to a non-trivial rotation about the origin.

6.48. Verify that conjugate Möbius transformations f and g always fall into the same category in the standard classification.

6.49. Assume that f is a Möbius transformation with exactly two fixed points z_1 and z_2. Let g be a Möbius transformation with the property that $g \circ f = f \circ g$. Show that either g fixes z_1 and z_2 or it interchanges these points, the latter case only being possible if g has order two.

6.50. Let f and g be Möbius transformations different from the identity. Assuming that f and g share the same set of fixed points, verify that $g \circ f = f \circ g$. Conversely, prove that the condition $g \circ f = f \circ g$, supplemented by the information that neither f nor g has order two, implies that f and g fix exactly the same points. Show by way of example that $g \circ f = f \circ g$ need not force a coincidence of fixed points when at least one of the transformations has order two. (*Hint.* First treat the case where f is parabolic. In the case of non-parabolic f Exercise 6.47 is relevant.)

6.51. Let S be the standard unit sphere in $\mathbb{R}^3$, let $F: S \to S$ be a rotation of S about an axis through the origin, and let $f = \pi \circ F \circ \pi^{-1}$, where $\pi: S \to \widehat{\mathbb{C}}$ is stereographic projection. Show that f is a Möbius transformation, one which appears in normalized form as $f(z) = (az - \bar{b})/(bz + \bar{a})$ with $|a|^2 + |b|^2 = 1$. Conversely, verify that any Möbius transformation which owns such a normalized form is generated in the above way by a rotation of S. (*Hint.* For topological reasons the mapping f must be sense-preserving. Use this information, without proof if necessary, in showing that f is a Möbius transformation. Then check that, unless f reduces to the identity transformation of $\widehat{\mathbb{C}}$, f is necessarily an elliptic transformation whose fixed points z_1 and z_2 are related by $z_2 = (\bar{z}_1)^{-1}$ and whose multiplier can be taken to be $\kappa = e^{i\theta}$, where θ is the rotation angle of the mapping F.)

6.52. Let f and g be conjugate Möbius transformations, say $g = h^{-1} \circ f \circ h$ for a Möbius transformation h. Verify that K is a g-invariant circle in $\widehat{\mathbb{C}}$ if and only if $h(K)$ is invariant under f.

6.53. Let $f(z) = z + b$, where $b \neq 0$, and let K be an f-invariant circle in $\widehat{\mathbb{C}}$. Give a careful proof that $K = L \cup \{\infty\}$, in which L is a line parallel to the one through 0 and b.

6.54. Given that $f(z) = \kappa z$, where $\kappa > 0$ and $\kappa \neq 1$, confirm that the invariant circles for f are the circles in $\widehat{\mathbb{C}}$ that pass through both 0 and ∞.

6.55. Let $f(z) = e^{i\theta}z$, where $\theta \neq 0$ and $-\pi < \theta \leq \pi$, and let K be an invariant circle for f. Assuming that $\theta \neq \pi$, prove that K must be a genuine circle centered at the origin. In case $\theta = \pi$, show that K is either a circle centered at the origin or a circle in $\widehat{\mathbb{C}}$ of the kind $L \cup \{\infty\}$, where L is a line through the origin.

6.56. It was claimed in the text that a loxodromic Möbius transformation f has no invariant circles, except when it has a negative real multiplier κ. Substantiate this claim. In addition, show that when $\kappa < 0$ the f-invariant circles are the circles in $\widehat{\mathbb{C}}$ that pass through both the fixed points of f.

6.57. Sketch a representative set of invariant circles for f: (i) $f(z) = (2z-1)/z$; (ii) $f(z) = (2z-1)/(-z+1)$; (iii) $f(z) = (2z-3)/(z-1)$.

6.58. Let $\widetilde{K}_0$ be a circle in $\widehat{\mathbb{C}}$, and let f be the reflection in $\widetilde{K}_0$. Prove that a circle K in $\widehat{\mathbb{C}}$ is an invariant circle for f if and only if K belongs to the family of circles in $\widehat{\mathbb{C}}$ that are orthogonal to $\widetilde{K}_0$. (*Hint.* Reduce the problem to the case $\widetilde{K}_0 = \mathbb{R} \cup \{\infty\}$.)

6.59. Construct a conformal mapping of the open quarter-disk $D = \{z : |z| < 1, x > 0, y > 0\}$ onto the open quadrant $D' = \{w : u > 0, v > 0\}$.

6.60. Let $D = \Delta_1 \cap \Delta_2$, where Δ_1 and Δ_2 are the open disks of equal radius whose bounding circles meet an angle (exterior to D) $\alpha\pi, 0 < \alpha < 1$, at the points i and $-i$. Construct a conformal mapping of D onto $D' = \Delta(0,1)$.

6.61. Let $D = \Delta(0,1) \sim [\overline{\Delta}(-1+i,1) \cup \overline{\Delta}(-1-i,1)]$. Map D conformally to the half-plane $D' = \{w : \operatorname{Im} w > 0\}$. (*Hint.* Start a construction by sending 1 to 0, 0 to 1, and -1 to ∞ with a Möbius transformation.)

6.62. Let $D = \Delta_1 \cap \Delta_2 \cap \Delta_3$, where Δ_1, Δ_2, and Δ_3 are open disks of the same radius whose boundary circles K_1, K_2, and K_3 obey the following conditions: K_1 intersects K_2 orthogonally at $i\sqrt{3}$, K_1 intersects K_3 orthogonally at -1, and K_2 intersects K_3 orthogonally at 1. Determine a conformal mapping of D onto the half-plane $D' = \{w : \operatorname{Im} w > 0\}$.

6.3 Exercises for Section IX.3

6.63. Suppose that f is a conformal mapping of a plane domain D onto $\Delta(0,1)$; assume it to be known that $f(z_0) = 0$ and $f'(z_0) = i$ at some specified point z_0 of D. Construct from f the conformal mapping g of D onto $\Delta(0,1)$ that obeys the conditions $g(z_0) = 1/2$ and $g'(z_0) > 0$. Find $g'(z_0)$.

6.64. If D is a simply connected plane domain, $D \neq \mathbb{C}$, and if z_0 is a point of D, then there is a unique radius $r > 0$ such that a conformal mapping f of D onto $\Delta(0,r)$ exists which satisfies $f(z_0) = 0$ and $f'(z_0) = 1$. Establish this fact.

6.65. Let $D(\neq \mathbb{C})$ be a simply connected plane domain that contains the origin and is symmetric with respect to it — the latter condition requires that $-z$ belong to D whenever z does — and let f be a conformal mapping of D onto $\Delta(0,1)$ with $f(0) = 0$. Demonstrate that f is necessarily an odd function. (*Hint.* Exploit the uniqueness assertion in the Riemann mapping theorem.)

6.66. Let $D(\neq \mathbb{C})$ be a simply connected plane domain that is symmetric with respect to a line L through the origin (i.e., $\rho(D) = D$, where ρ is the reflection in $K = L \cup \{\infty\}$), let z_0 be a point of $L \cap D$, and let f be the conformal mapping of D onto $\Delta(0,1)$ that has $f(z_0) = 0$ and $f'(z_0) > 0$. Show that f is symmetric about L, in the sense that $f[\rho(z)] = \rho[f(z)]$ for every point z of D. Using this information, conclude that $L \cap D$ is an open interval whose image under f is the set $L \cap \Delta(0,1)$. (*Hint.* Look at $g = \rho \circ f \circ \rho$ in D.)

6.67. The domain D inside the ellipse with equation $(x/a)^2 + (y/b)^2 = 1$ is mapped conformally onto $\Delta(0,1)$ by a function f. Given the information that $f(0) = 0$ and $f'(0) > 0$, ascertain the images under f of the (open) axes of the ellipse.

6.68. Let D be the interior of the square with vertices $1, i, -1$, and $-i$. Given that f maps D conformally onto $\Delta(0,1)$ with $f(0) = 0$ and $f'(0) < 0$, identify the images under f of the (open) diagonals of D and of the two (open) line segments that join the midpoints of its opposite sides.

6.69. If D is the interior of the triangle whose vertices are $1, e^{2\pi i/3}$, and $e^{4\pi i/3}$, determine the images of the (open) altitudes of D under the function f that maps D conformally onto $\Delta(0,1)$, fixes 0, and has $\operatorname{Arg} f'(0) = \pi/2$.

6.70. A domain D in $\widehat{\mathbb{C}}$ is called simply connected if either of the statements (ii) or (iii) in Theorem 3.6 is true of D. (The interpretation of "homologous in D" would require some serious reworking before our originally adopted definition of simple connectivity would make sense for domains that contain the point ∞.) Let D and D' be simply connected domains in $\widehat{\mathbb{C}}$ such that each of the sets $\widehat{\mathbb{C}} \sim D$ and $\widehat{\mathbb{C}} \sim D'$ contains at least two points, let z_0 be a point of D, and let z_0' belong to D'. Establish the existence of a conformal mapping of D onto D' with $f(z_0) = z_0'$. Is the same true if both $\widehat{\mathbb{C}} \sim D$ and $\widehat{\mathbb{C}} \sim D'$ are one-point sets? What if $\widehat{\mathbb{C}} \sim D$ contains at least two points but $\widehat{\mathbb{C}} \sim D'$ has fewer than two elements?

6.4 Exercises for Section IX.4

6.71. Under the assumption that $\ell(\gamma) \geq c > 0$ for every path γ in $\Gamma[E, F\colon G]$, obtain the estimate $M[E, F\colon G] \leq c^{-2}A(G)$.

6.72. Let $G = \{z : r_0 \leq |z - z_0| \leq r_1\}$ with $0 < r_0 < r_1 < \infty$, let $E = K(z_0, r_0)$, and let $F = K(z_0, r_1)$. Verify that $M[E, F\colon G] = 2\pi[\text{Log}(r_1/r_0)]^{-1}$.

6.73. If $G = \overline{\Delta}(z_0, r)$, $E = \{z_0\}$, and $F = K(z_0, r)$, then $M[E, F\colon G] = 0$. Prove this.

6.74. Suppose that f is the reflection in the real axis. Demonstrate that $M[E, F\colon G] = M[f(E), f(F)\colon f(G)]$ for any configuration $[E, F\colon G]$. Employ this information to check that the modulus $M[E, F\colon G]$ is also invariant under anti-conformal mappings.

6.75. Let L be a line in the complex plane, and let G and G^* be the complementary closed half-spaces determined by L. Show that

$$M[E_1, F\colon G] + M[E_2, F\colon G^*] \leq M[E_1 \cup E_2, F\colon \mathbb{C}]$$

for any configurations $[E_1, F\colon G]$ and $[E_2, F\colon G^*]$ with F contained in L.

6.76. Suppose that E lies in the interior of $G = \{z : \text{Im}\, z \geq 0\}$ and that E^* is the reflection of E in the real axis. Certify that $M[E \cup E^*, \mathbb{R}\colon \mathbb{C}] = 2M[E, \mathbb{R}\colon G]$. (*Hint.* Exercises 6.72 and 6.73 lead to an inequality in one direction. There is an easy way to construct an admissible density $\widetilde{\rho}$ for $[E \cup E^*, \mathbb{R}\colon \mathbb{C}]$ from an admissible density ρ for $[E, \mathbb{R}\colon G]$. This gives an equality in the other direction.)

6.77. Suppose that E lies in the interior of $G = \{z : \text{Im}\, z \geq 0\}$ and that E^* is the reflection of E in the real axis. Show that $M[E, \mathbb{R}\colon G] = 2M[E, E^*\colon \mathbb{C}]$. (*Hint.* To get an inequality in one direction consider $\rho \in \text{Adm}[E, \mathbb{R}\colon G]$, define $\widetilde{\rho}$ in $\mathbb{C}$ by $\widetilde{\rho}(z) = \rho(z)$ if $z \in G$ and $\widetilde{\rho}(z) = \rho(\bar{z})$ if $z \in G^*$, and show that $\widetilde{\rho}/2 \in \text{Adm}[E, E^*\colon \mathbb{C}]$. Conversely, given $\rho \in \text{Adm}[E, E^*\colon \mathbb{C}]$, define $\widetilde{\rho}\colon G \to [0, \infty)$ by $\widetilde{\rho}(z) = \rho(z) + \rho(\bar{z})$ and confirm that $\widetilde{\rho} \in \text{Adm}[E, \mathbb{R}\colon G]$. This will produce an inequality in the opposite direction.)

6.78. Assume that D is a domain in the complex plane and that Γ is an arbitrary family of piecewise smooth paths in D. Define an extended real number $M[\Gamma; D]$ as follows: $M[\Gamma\colon D] = \inf\{A_\rho(D) : \rho \in \text{Adm}[\Gamma\colon D]\}$ if $\text{Adm}[\Gamma\colon D] \neq \phi$ and $M[\Gamma\colon D] = \infty$ otherwise. Here, as one might anticipate, $\text{Adm}[\Gamma\colon D]$ indicates the class of densities ρ in D such that $\int_\gamma \rho(z)|dz| \geq 1$ for every path γ in Γ. (This is obviously a generalization of the conformal modulus of a configuration $[E, F\colon D]$, one in which the paths involved are no longer constrained to begin and end in preordained sets.) Show that $M[\Gamma\colon D]$ is conformally invariant: if $f\colon D \to \mathbb{C}$ is a conformal mapping, then $M[\Gamma\colon D] = M[f(\Gamma)\colon f(D)]$, where $f(\Gamma) = \{f \circ \gamma : \gamma \in \Gamma\}$.

6.79. A plane domain D is mapped conformally by f onto $\Delta(0,1)$. One is told that $|f'(z)| \geq c$ throughout D, where $c > 0$ is a constant. Derive the bound $M[\Gamma\colon D] \leq \pi(cL)^{-2}$ for any family Γ of piecewise smooth paths in D, in which $L = \inf\{\ell(\gamma) : \gamma \in \Gamma\}$.

6.80. Let $D = \{z : r_0 < |z| < r_1\}$, where $0 \leq r_0 < r_1 \leq \infty$, and let Γ be the collection of all closed, piecewise smooth paths γ in D with the property that $n(\gamma, 0) \neq 0$. Demonstrate that $M[\Gamma\colon D] = (2\pi)^{-1} \operatorname{Log}(r_1/r_0)$ if $0 < r_0 < r_1 < \infty$, while $M[\Gamma\colon D] = \infty$ if $r_0 = 0$ or $r_1 = \infty$.

6.81. If f is a conformal mapping of an annulus $D = \{z : r < |z| < 1\}$ onto an annulus $D' = \{w : s < |w| < 1\}$, where $0 \leq r < 1$ and also $0 \leq s < 1$, then necessarily $r = s$. Prove this. (*Hint.* Let Γ be the family of closed, piecewise smooth paths γ in D with the property that $n(\gamma, 0) \neq 0$. To start, show that any path $\beta = f \circ \gamma$, with γ a member of Γ, must have $n(\beta, 0) \neq 0$. If not, β would be homologous to zero in D'. Why? And why is this unacceptable?)

6.82. Show that any domain D in $\widehat{\mathbb{C}}$ such that $\widehat{\partial} D$ consists of two disjoint circles in $\widehat{\mathbb{C}}$ can be mapped by a Möbius transformation to $D' = \{z : r < |z| < 1\}$ for a unique r in the interval $(0,1)$.

6.83. Given that D is a Jordan domain in the complex plane and that E and F are disjoint connected subsets of D with the feature that $\widehat{E}$ and $\widehat{F}$ have a point of $\widehat{\partial} D$ in common, prove that $M[E, F\colon D] = \infty$. Show that, if $D = \{z : 0 < y < 1\}$, $E = \{z \in D : x \geq 1\}$, and $F = \{z \in D : x \leq -1\}$, then $M[E, F\colon D] < \infty$, even though $\infty \in \widehat{E} \cap \widehat{F}$. Are these two results in conflict?

6.84. Modify the proof of Lemma 4.3 to upgrade that result as follows: if $\Delta = \Delta(0,1)$ and $E = \overline{\Delta}(0, 1/2)$, if z_1 and z_2 are distinct points of $\partial\Delta$, and if $0 < r < |z_1 - z_2|/2$, then there is a constant $c > 0$ — it will depend on $|z_1 - z_2|$ and r — with the property that $M[E, F\colon \Delta] \geq c$ for every connected subset F of $\Delta \sim E$ which intersects both $\Delta(z_1, r)$ and $\Delta(z_2, r)$.

6.85. Let f be a conformal mapping of a bounded domain D onto $\Delta = \Delta(0,1)$. Establish the following extension result of Raimo Näkki: the function f extends to a continuous mapping $\tilde{f}$ of $\overline{D}$ onto $\overline{\Delta}$ if and only if D has the property (*) $\inf\{\ell(\gamma) : \gamma \in \Gamma[E, F\colon D]\} = 0$ whenever E and F are disjoint connected sets in D for which $\overline{E} \cap \overline{F} \cap \partial D \neq 0$. (N.B. The domain $D = \Delta \sim [0,1)$, for example, does not have the aforementioned property. *Hint.* For the necessity use Exercise 6.69 and Lemma 4.4. As to the sufficiency, it is enough to check that $\lim_{\zeta \to z} f(\zeta)$ exists for every point z of ∂D, for an analogue of Lemma 4.1 then applies. Suppose this not to be so at some z and argue to a contradiction. Use Exercise 6.81 and the following consequence of (*): if $E_0 = f^{-1}[\overline{\Delta}(0,1/2)]$, if E and F are disjoint connected sets in $D \sim f^{-1}[\overline{\Delta}(0,3/4)]$ with $\overline{E} \cap \overline{F} \cap \partial D \neq \phi$, and if $c > 0$, then there must be a path γ in $\Gamma[E, F{:}\, D]$ for which $M[E_0, F_0{:}\, D] < c$, where $F_0 = |\gamma|$.)

6.86. Let D be the domain inside the cardioid whose polar equation is $r = 2 + 2\cos\theta$. Find an integral representation for the solution of the Dirichlet problem in D with boundary data h.

6.5 Exercises for Section IX.5

6.87. A function f is analytic in a Jordan domain D in the complex plane. It is known that $\lim_{\zeta \to z} f(\zeta) = 0$ for all points z belonging to C, a connected subset of $\widehat{\partial} D$ which contains more than one point. Deduce that $f(z) = 0$ for every z in D. (*Hint.* First do the case where D is a half-plane.)

6.88. Let $D = \{z : 0 < x < 1, y > 0\}$, and let $f{:}\, \overline{D} \to \mathbb{C}$ be a bounded continuous function that is analytic in D and real-valued on the interval $(0,1)$. Assume that $|f(iy)| \leq m_0$ and $|f(1+iy)| \leq m_1$ for all $y \geq 0$, where m_0 and m_1 are constants. Prove that $|f(x+iy)| \leq m_0^{1-x} m_1^x$ holds whenever $0 < x < 1$ and $y \geq 0$.

6.89. A plane domain D is symmetric about the real axis, and G is one of the components of $D \sim \mathbb{R}$. Given that a function $f = u + iv$ is analytic in G and obeys the condition $\lim_{\zeta \to z} u(\zeta) = 0$ for every point z of $I = D \cap \mathbb{R}$, establish the existence of an analytic function $F{:}\, D \to \mathbb{R}$ that coincides with f in G. (*Hint.* First prove that $\lim_{\zeta \to z} f(\zeta)$ exists for every z in I. For this use Theorem VI.3.4 and the fact that harmonic functions always have harmonic conjugates in disks.)

6.90. Make the following changes in the hypotheses of Theorem 5.3: assume that D is now a domain in $\widehat{\mathbb{C}}$ and that $f{:}\, G \cup I \to \widehat{\mathbb{C}}$ is a continuous function which is meromorphic in G. Keeping the remaining hypotheses as they stand, prove that the function F defined there is still meromorphic.

6.91. Suppose that $D = \{z : 0 < x < a, 0 < y < b\}$ and that $D' = \{w : 0 < u < c, 0 < v < d\}$. Establish the fact that there is a conformal mapping f of D onto D' whose homeomorphic extension $\tilde{f}$ to $\overline{D}$ satisfies $\tilde{f}(0) = 0$, $\tilde{f}(a) = c$, $\tilde{f}(ib) = id$, and $\tilde{f}(a+ib) = c + id$ if and only if $a/b = c/d$. (*Hint.* Assuming that a mapping f of the given description exists, prove that it can be extended to a conformal self-mapping of the complex plane.)

6.92. Let $D = \{z : |z| > 1, |x| < 1, y > 0\}$, let $H = \{w : v > 0\}$, and let f be the homeomorphism of $\widehat{D}$ onto $\widehat{H}$ that maps D conformally onto H and satisfies $f(-1) = -1$, $f(1) = 1$, and $f(\infty) = \infty$. Show that f can be extended to an analytic function $F: H \to \mathbb{C}$ whose range is $\mathbb{C} \sim \{1, -1\}$.

6.93. Let f be a conformal mapping of $H = \{z : \operatorname{Im} z > 0\}$ onto the interior of a closed polygon P, and let z_0 be a point of ∂H. We can see from the proof of Lemma 5.4 that, if $\tilde{f}(z_0)$ is not a vertex of P, then $\lim_{z \to z_0} |f'(z)|$ exists and is not zero. Show that $\lim_{z \to z_0} |f'(z)| = \infty$ in case $\tilde{f}(z_0)$ is a vertex of P at which the interior angle is less than π, whereas $\lim_{z \to z_0} |f'(z)| = 0$ if $\tilde{f}(z_0)$ is a vertex of P at which the interior angle is more than π. Verify, also, that $|f'(z)| \to 0$ as $z \to \infty$ through H. (*Hint.* For the last part remember relation (9.69).)

6.94. Referring to Example 5.2, show that $\tilde{f}$ maps the arc $J = \widehat{H} \cap K(0,1)$ to the diagonal of Q which joins 1 and i. From this determine the point of H that gets mapped by f to the center of Q. (*Hint.* Consider $g = \rho_1 \circ f \circ \rho$ for suitable reflections ρ and ρ_1.)

6.95. Construct a conformal mapping of $H = \{z : \operatorname{Im} z > 0\}$ onto the interior of the closed rhombus P that has 1 and -1 as opposite vertices, the interior angle at each being $\alpha\pi$ for $0 < \alpha < 1$.

6.96. Let P be the regular polygon whose vertices are the n^{th}-roots of unity, where $n \geq 3$. Find an integral representation for the conformal mapping f of $\Delta(0,1)$ onto the interior of P such that $f(0) = 0$ and $\tilde{f}(1) = 1$. What are the points of $K(0,1)$ that get mapped by $\tilde{f}$ to the midpoints of the sides of P?

6.97. Let $f(z) = \int_0^z (\zeta - a)^{-1/2} (\zeta - b)^{-1/2} (\zeta - c)^{-1/2} d\zeta$, where $a < b < c$ are real numbers. Show directly that f maps $H = \{z : \operatorname{Im} z > 0\}$ conformally onto a rectangle with vertices $f(a), f(b), f(c)$, and $f(\infty)$. (*Hint.* Start by examining $\operatorname{Arg} \dot{\gamma}(t)$, where $\gamma(t) = f(t)$ for real t.)

6.98. Let $D = \{z : |x| < 1, |y| < c\}$, where $c > 0$. Demonstrate the existence and uniqueness of a number k in the interval $(0,1)$ — you will not be able to identify k unless $c = 1$ — with the property that there is a conformal mapping f of $H = \{z : \operatorname{Im} z > 0\}$ onto D for which $\tilde{f}(-1) = 1 + ic$, $\tilde{f}(-k) = -1 + ic$, $\tilde{f}(k) = -1 - ic$, and $\tilde{f}(1) = 1 - ic$. Thus f has the structure $f(z) = A \int_0^z (\zeta - 1)^{-1/2} (\zeta + 1)^{-1/2} (\zeta - k)^{-1/2} (\zeta + k)^{-1/2} d\zeta + B$, where A and B are constants. What point does f map to the center of D? What is k when $c = 1$? (*Hint.* Seek f in the form $f = g^{-1} \circ h$, where h is a Möbius transformation that maps H onto $\Delta(0,1)$ and g is the conformal mapping of D onto $\Delta(0,1)$ with $g(0) = 0$ and $g'(0) > 0$. Recall Exercises 6.66 and 6.28(viii).)

6.99. Let D be a Jordan domain in $\mathbb{C}$ and let (z_1, z_2, z_3, z_4) be a quadruple of distinct points from $\widehat{\partial} D$ that is positively oriented relative to D. Demonstrate that there is a unique number $M > 0$ with the following property: there exists a homeomorphism f of $\widehat{D}$ onto $\{z : 0 \leq x \leq M, 0 \leq y \leq 1\}$ that is conformal in D and maps z_1 to 0, z_2 to M, z_3 to $M + i$, and z_4 to i.

Chapter X

Constructing Analytic Functions

Introduction

It is our aim with the present chapter to impress upon the reader just how great the wealth of functions is to which the ideas related in this book apply. The theorem of Mittag-Leffler, Weierstrass's theorem, and the method of analytic continuation generate a supply of analytic and meromorphic functions rich beyond one's wildest expectations. At the same time the chapter is intended to serve as a gateway to more sophisticated topics in complex analysis. The subjects introduced here come to full fruition in such diverse areas as the theory of entire functions, the theory of value distribution for meromorphic functions, analytic number theory, and the theory of Riemann surfaces, to name but a few. Hopefully the reader will find in the contents of this chapter a stimulus for continuing the study of complex function theory at more advanced levels.

1 The Theorem of Mittag-Leffler

1.1 Series of Meromorphic Functions

The material in this section demands that we deal with function series $\sum_{n=1}^{\infty} f_n$ (and $\sum_{n=-\infty}^{\infty} f_n$) whose terms are meromorphic in a plane open set U. Since the value ∞ is allowed to appear in the range of any or all f_n, the notions of convergence we have heretofore considered do not strictly apply to such a series. The concept of normal convergence, in particular, requires some adjustment. That adjustment is conventionally made as follows: a series of functions $\sum_{n=1}^{\infty} f_n$ whose terms are meromorphic in an open set U is said to *converge normally in* U if corresponding to each compact subset K of U there exists an index N such that no term f_n with

$n > N$ has a pole in K and such that the truncated series $\sum_{n=N+1}^{\infty} f_n$ converges uniformly on K. (N.B. As was the case for the earlier definition of normal convergence, to insure this behavior it is sufficient that the stated condition be satisfied for every closed disk K in U.) A doubly infinite series $\sum_{n-\infty}^{\infty} f_n$ of functions meromorphic in U converges normally in U provided both of the series $\sum_{n=0}^{\infty} f_n$ and $\sum_{n=1}^{\infty} f_{-n}$ are normally convergent there. It goes without saying that this generalized understanding of normal convergence reduces to the original concept for series whose terms are actually analytic in U. To illustrate the new interpretation of normal convergence, recall Example VII.3.2. The argument presented there shows that $\sum_{n=-\infty}^{\infty}(z-n)^{-2}$, when viewed as a series of meromorphic functions, is normally convergent in the whole complex plane. By the old definition it was only regarded as normally convergent in $\mathbb{C} \sim \{0, \pm 1, \pm 2, \cdots\}$.

If a series $\sum_{n=1}^{\infty} f_n$ of functions that are meromorphic in an open set U converges normally there, then the set $E = \bigcup_{n=1}^{\infty} E_n$, where E_n is the set of poles of f_n in U, is a discrete subset of U. Indeed, the definition of normal convergence and the fact that each individual set E_n is a discrete subset of U imply that any closed disk $\overline{\Delta}(z_0, r)$ in U intersects E in at most finitely many points, eliminating z_0 as a potential limit point of E. All of the terms f_n are analytic in the open set $V = U \sim E$, and the series $\sum_{n=1}^{\infty} f_n$ converges normally in V. As a result, the function $f: V \to \mathbb{C}$ defined by $f(z) = \sum_{n=1}^{\infty} f_n(z)$ is analytic. It has an isolated singularity at every point of E. What is the nature of these singularities? Given z_0 in E, we can choose a closed disk $K = \overline{\Delta}(z_0, r)$ in U with the property that $K \cap E = \{z_0\}$. According to the definition of normal convergence we can then fix an index N such that f_n is free of poles in K — hence, is analytic in the open disk $\Delta = \Delta(z_0, r)$ — as soon as $n > N$ and such that the series $\sum_{n=N+1}^{\infty} f_n$ converges uniformly on K. It follows that $g(z) = \sum_{n=N+1}^{\infty} f_n(z)$ defines an analytic function in Δ. In the punctured disk $\Delta^* = \Delta^*(z_0, r)$ the function f admits the representation $f = f_1 + f_2 + \cdots + f_N + g$. Because each of the finitely many functions $f_1, f_2, \ldots, f_N$, and g has no worse than a pole at z_0, the same is true of their sum f. The conclusion: at each point of U the function f has at worst a pole, so f is meromorphic in U. By convention, f is automatically assigned values at the points of E so as to extend it to a continuous function from U into $\widehat{\mathbb{C}}$. We continue to speak of f as the *sum* of the series $\sum_{n=1}^{\infty} f_n$ in U and to write $f(z) = \sum_{n=1}^{\infty} f_n(z)$ for all z in U, despite the fact that $\sum_{n=1}^{\infty} f_n(z_0)$ is technically undefined when z_0 belongs to E: for such z_0 the value $f(z_0)$ is officially given by $f(z_0) = \lim_{z \to z_0} \sum_{n=1}^{\infty} f_n(z)$, the limit being taken in the extended plane $\widehat{\mathbb{C}}$. We summarize the preceding remarks in a theorem.

Theorem 1.1. *Suppose that each term in a function series $\sum_{n=1}^{\infty} f_n$ is meromorphic in an open set U and that the series converges normally in U, with the function f as its sum. Then f is meromorphic in U. Any pole of f is a pole of f_n for at least one value, but at most finitely many values, of* *n. Furthermore,* $f^{(k)} = \sum_{n=1}^{\infty} f_n^{(k)}$ *in* U *for every positive integer* k*. Each of these derived series converges normally in* U*.*

Proof. Only the statements concerning the series of derivatives are still in need of justification. Fix a positive integer k. We first establish that the series of meromorphic functions $\sum_{n=1}^{\infty} f_n^{(k)}$ converges normally in U. For this, let $K = \overline{\Delta}(z_0, r)$ be an arbitrary closed disk in U. We must prove that there is an index N for which $f_n^{(k)}$ is without poles in K whenever $n > N$ and for which the series $\sum_{n=N+1}^{\infty} f_n^{(k)}$ is uniformly convergent on K. Choose s satisfying $s > r$ such that $K_1 = \overline{\Delta}(z_0, s)$ is still contained in U. Since $\sum_{n=1}^{\infty} f_n$ is given to be normally convergent in U, we are guaranteed the existence of an index N such that no f_n with $n > N$ exhibits a pole in K_1 and such that the series $\sum_{n=N+1}^{\infty} f_n$ converges uniformly on K_1. It follows that $\sum_{n=N+1}^{\infty} f_n$ is a normally convergent series of analytic functions in the disk $\Delta = \Delta(z_0, s)$. Denote its sum there by g. When $n > N$ the function $f_n^{(k)}$, being itself analytic in Δ, has no poles in K. Also, in view of Theorem VII.3.2 the series $\sum_{n=N+1}^{\infty} f_n^{(k)}$ converges normally in Δ — hence, converges uniformly on K — where its sum is $g^{(k)}$. Notice that in Δ we have $f = f_1 + f_2 + \cdots + f_N + g$, from which we infer that in this disk

$$f^{(k)} = f_1^{(k)} + f_2^{(k)} + \cdots + f_N^{(k)} + g^{(k)}$$

$$= f_1^{(k)} + f_2^{(k)} + \cdots + f_N^{(k)} + \sum_{n=N+1}^{\infty} f_n^{(k)} = \sum_{n=1}^{\infty} f_n^{(k)} .$$

Therefore, we have shown that $\sum_{n=1}^{\infty} f_n^{(k)}$ is normally convergent in U and in the process demonstrated that each point of U is the center of an open disk in which $f^{(k)} = \sum_{n=1}^{\infty} f_n^{(k)}$, a fact which implies that $f^{(k)} = \sum_{n=1}^{\infty} f_n^{(k)}$ throughout U. ∎

An obvious analogue of Theorem 1.1 is valid for doubly infinite series.

1.2 Constructing Meromorphic Functions

The basic global existence theorem for meromorphic functions was discovered by the dynamic Swedish mathematician and entrepreneur Gösta Mittag-Leffler (1846-1927), whose mathematical legacy also includes the journal *Acta Mathematica*, which he founded and which remains to this day one of the world's most prestigious publications devoted to mathematical research. In stating Mittag-Leffler's result we use the terminology ***rational singular part at*** z_0 to describe a rational function S of the type

$$S(z) = \frac{a_{-m}}{(z - z_0)^m} + \frac{a_{-m+1}}{(z - z_0)^{m-1}} + \cdots + \frac{a_{-1}}{z - z_0} ,$$

in which $a_{-m} \neq 0$; i.e., S has exactly one pole in $\widehat{\mathbb{C}}$, it located at the point z_0, and $S(\infty) = 0$.

Let $E = \{z_1, z_2, z_3, \cdots\}$ be a discrete subset of an open set U in $\mathbb{C}$. (N.B. Implicit here is the fact that the elements of a discrete subset of U can be listed in a finite or infinite sequence, which for present purposes we always assume to be univalent — $z_n \neq z_m$ when $n \neq m$.) Assume that at each point z_n of E we are presented with a rational singular part S_n. Does there exist a meromorphic function in U whose set of poles in this set coincides with E and whose singular part at z_n is S_n for $n = 1, 2, \cdots$? Mittag-Leffler's theorem provides an affirmative answer to this question. Of course, when $E = \{z_1, z_2, z_3, \ldots, z_p\}$ is just a finite subset of U, it is no great feat to manufacture such a function: $f = S_1 + S_2 + \cdots + S_p$ does the job. In the more interesting case where E is infinite the natural temptation is to hope that $f = \sum_{n=1}^{\infty} S_n$ will meet the requirements, as indeed it will when this series happens to converge normally in U. Unfortunately, one cannot in general count on that convergence. Mittag-Leffler's insight was to realize that, upon subtracting from S_n an appropriate rational function R_n whose poles are outside U, one can produce a series $\sum_{n=1}^{\infty} (S_n - R_n)$ that does converge normally in U. The function $f = \sum_{n=1}^{\infty} (S_n - R_n)$ is then meromorphic in U, it is analytic in $U \sim E$, and it has a pole with the prescribed singular part at each point of E. (The last statement follows from the discussion preceding Theorem 1.1, which confirms that $f - S_n = -R_n + \sum_{k \neq n} (S_k - R_k)$ is analytic in a suitably small punctured disk centered at z_n. Remember: for $k \neq n$ the function $S_k - R_k$ is analytic in the neighborhood of z_n.) A collection of functions R_n for which Mittag-Leffler's procedure is successful is called a set of "convergence inducing summands" for the problem.

Theorem 1.2. (Mittag-Leffler's Theorem) *Let $E = \{z_1, z_2, z_3, \cdots\}$ be a discrete subset of an open set U in the complex plane, and for $n = 1, 2, 3, \cdots$ let S_n be a rational singular part at z_n. There exists a meromorphic function $f: U \to \widehat{\mathbb{C}}$ that has E as its set of poles and that for $n = 1, 2, 3, \cdots$ has singular part S_n at z_n. Any two functions fitting this description differ by a function that is analytic in U.*

Proof. Only when E is an infinite set does the existence of such a function remain in doubt, so it is an infinite set E that we consider in the proof. For technical reasons we shall assume that E does not include the origin. (This entails no loss of generality: if the origin does belong to E — say $z_1 = 0$ — then the function $f = S_1 + \widetilde{f}$, where $\widetilde{f}$ is a solution of the same problem for the set $\widetilde{E} = \{z_2, z_3, z_4, \cdots\}$ with the singular part S_n still specified at z_n for $n = 2, 3, 4, \cdots$, has all the features demanded.) Let δ_n designate the radius of the largest open disk centered at z_n that is contained in U. We distinguish two preliminary cases.

Case 1. $|z_n|\delta_n \geq 1$ for every n. (N.B. If $U = \mathbb{C}$, then $\delta_n = \infty$ for every n. By assumption $z_n \neq 0$, so $|z_n|\delta_n = \infty$ for every n. Hence, Case 1 is actually the general case when U is the whole complex plane.) We claim that in this situation $|z_n| \to \infty$ as $n \to \infty$. If not, we would be able to extract from the sequence $\langle z_n \rangle$ a subsequence $\langle z_{n_k} \rangle$ that converges to some finite accumulation point z_0 of $\langle z_n \rangle$. Should $U = \mathbb{C}$, this would instantly contradict the fact that E is a discrete subset of U. Assuming that $U \neq \mathbb{C}$, we know from the definition of δ_n and the assumption in Case 1 that

$$|z_n - z| \geq \delta_n \geq \frac{1}{|z_n|}$$

for every $n \geq 1$ and every z in $\mathbb{C} \sim U$. Accordingly, we would obtain

$$|z_0 - z| = \lim_{k\to\infty} |z_{n_k} - z| \geq \lim_{k\to\infty} \frac{1}{|z_{n_k}|} = \frac{1}{|z_0|}$$

for every z in $\mathbb{C} \sim U$, which both rule out $z_0 = 0$ and place z_0 in the set U. This would again mark z_0 as a limit point of E in U, another violation of discreteness. Therefore, $|z_n| \to \infty$ as declared. We shall see that the condition $|z_n| \to \infty$ leads to an even stronger conclusion than the one promised by the theorem.

The rational function S_n, whose unique pole is at z_n, is analytic in the disk $\Delta_n = \Delta(0, |z_n|)$. It follows that S_n can be represented in Δ_n as the sum of a normally convergent Taylor series centered at the origin, $S_n(z) = \sum_{k=0}^{\infty} a_k^{(n)} z^k$. As $d \to \infty$, $\sum_{k=0}^{d} a_k^{(n)} z^k \to S_n(z)$ uniformly on the closed disk $K_n = \bar{\Delta}(0, |z_n|/2)$. For this reason we are at liberty to select and fix an index $d(n)$ such that the polynomial $R_n(z) = \sum_{k=0}^{d(n)} a_k^{(n)} z^k$ satisfies $|S_n(z) - R_n(z)| < 2^{-n}$ for every z in K_n. We do this for $n = 1, 2, 3, \cdots$.

We now assert that the series $\sum_{n=1}^{\infty}(S_n - R_n)$ is normally convergent in $\mathbb{C}$. If so, $f = \sum_{n=1}^{\infty}(S_n - R_n)$ will define a function that is meromorphic in $\mathbb{C}$ — not just meromorphic in U, which is all we originally asked for — having the prescribed poles and singular parts. Let K be an arbitrary compact set in $\mathbb{C}$. Since $|z_n| \to \infty$, we can fix an index N with the property that K is contained in the disk K_n whenever $n > N$. If $n > N$, then the only pole in $\mathbb{C}$ of the function $S_n - R_n$, the one at z_n, is not a point of K. Also, for such n the estimate $|S_n(z) - R_n(z)| < 2^{-n}$ holds throughout K. On the basis of the Weierstrass M-test we can pronounce the series $\sum_{n=N+1}^{\infty}(S_n - R_n)$ uniformly convergent on K. We have thus established the normal convergence of $\sum_{n=1}^{\infty}(S_n - R_n)$ in $\mathbb{C}$. With this we have completed the proof of the theorem in Case 1, obtaining a slightly better result than the theorem stated.

Case 2: $|z_n|\delta_n < 1$ for every n. To handle Case 2 we first observe that its defining condition implies that $\delta_n \to 0$ as $n \to \infty$. Suppose this not to be so. Then there exists an $\epsilon > 0$ such that $\delta_n \geq \epsilon$ is true for infinitely many n. In other words, the sequence $\langle \delta_n \rangle$ has a subsequence $\langle \delta_{n_k} \rangle$ such that $\delta_{n_k} \geq \epsilon$ is satisfied for every k. Thus $|z_{n_k}| < \delta_{n_k}^{-1} < \epsilon^{-1}$ holds for all

k, making $\langle z_{n_k} \rangle$ a bounded sequence. By passing to a further subsequence and relabeling, if need be, we may assume that $\langle z_{n_k} \rangle$ is convergent. Let z_0 be its limit. For every $k \geq 1$ and every z in $\mathbb{C} \sim U$ it is then the case that

$$|z_{n_k} - z| \geq \delta_{n_k} \geq \epsilon \ .$$

For all such z, therefore,

$$|z_0 - z| = \lim_{k \to \infty} |z_{n_k} - z| \geq \epsilon > 0 \ ,$$

which means that z_0 is a point of U. As z_0 is also a limit point of E, a discrete subset of U, we have reached a contradiction. We are forced to conclude that $\delta_n \to 0$ in Case 2.

Since $\delta_n < \infty$ in Case 2, we can choose and fix for each n a point ζ_n of ∂U with the property that $|\zeta_n - z_n| = \delta_n$. (If no such ζ_n existed, $\Delta(z_n, \delta_n)$ would not be the largest open disk in U centered at z_n.) Because the function S_n is analytic in the annulus $G_n = \{z : \delta_n < |z - \zeta_n| < \infty\}$, it can be expanded there in a Laurent series centered at ζ_n. The structure of S_n permits us to extract this expansion from the identity

$$\frac{1}{z - z_n} = \sum_{k=1}^{\infty} \frac{(z_n - \zeta_n)^{k-1}}{(z - \zeta_n)^k} \ ,$$

valid for all z in G_n, by repeatedly differentiating it and by taking the appropriate linear combination of the resulting derived series, a process which reveals that the Laurent expansion of S_n in G_n has the form $S_n(z) = \sum_{k=1}^{\infty} a_k^{(n)} (z - \zeta_n)^{-k}$. Furthermore, this series converges uniformly on the set $A_n = \{z : 2\delta_n \leq |z - \zeta_n| < \infty\}$. This entitles us to select and fix an index $d(n)$ for which the rational function $R_n(z) = \sum_{k=1}^{d(n)} a_k^{(n)} (z - \zeta_n)^{-k}$ meets the following specification: $|S_n(z) - R_n(z)| < 2^{-n}$ is true for every z in A_n. Again, this is to be done for $n = 1, 2, 3, \cdots$. Notice that R_n is analytic in U, for its only pole ζ_n lies on ∂U.

To cap off the proof of the theorem in Case 2 we need only verify that the series $\sum_{n=1}^{\infty}(S_n - R_n)$ converges normally in U, for then $f = \sum_{n=1}^{\infty}(S_n - R_n)$ is a function blessed with all the properties insisted upon in the theorem. The poles of $S_n - R_n$ — namely, z_n and ζ_n — are not elements of A_n. If K is any compact subset of U, then there is a number $\delta > 0$ such that $|z - \zeta| \geq \delta$ holds for every z in K and every ζ in $\mathbb{C} \sim U$ (Lemma II.4.3). Since $2\delta_n \to 0$ as $n \to \infty$, it follows easily that there is an index N such that K is a subset of A_n once $n > N$. For such n the function $S_n - R_n$ is free of poles in K and obeys the estimate $|S_n(z) - R_n(z)| < 2^{-n}$ everywhere in K. By the Weierstrass M-test the series $\sum_{n=N+1}^{\infty}(S_n - R_n)$ is uniformly convergent on K. This certifies $\sum_{n=1}^{\infty}(S_n - R_n)$ as normally convergent in U and so finishes the proof of the theorem in Case 2.

Finally, we come to an infinite discrete subset $E = \{z_1, z_2, z_3, \cdots\}$ of U not covered by either Case 1 or Case 2. Set $J = \{n : |z_n|\delta_n \geq 1\}$,

$E_1 = \{z_n : n \in J\}$, and $E_2 = E \sim E_1$. Then E_1 and E_2 are non-empty, disjoint sets whose union is E. Moreover, E_1 is either a finite set or a set to which Case 1 applies, whereas E_2 either is finite or falls within the scope of Case 2. As a consequence, there exists for $j = 1, 2$ a meromorphic function $f_j : U \to \widehat{\mathbb{C}}$ that has pole-set E_j, at each point of which f_j has the singular part specified for that point in the data for E. The function $f = f_1 + f_2$ is meromorphic in U, its set of poles there is $E_1 \cup E_2 = E$, and it has at each point of E the prescribed singular part. If g is any other function with the same properties, then $f - g$ is analytic in $U \sim E$ and has a removable singularity at every point of E. Upon removal of these singularities, $f - g$ becomes analytic in U. The final assertion of the theorem is thus clear. ∎

Mittag-Leffler's theorem remains true essentially as stated — in the last sentence the work "analytic" must be replaced by "holomorphic" — for an open set U in the extended complex plane $\widehat{\mathbb{C}}$. By a *rational singular part at* ∞, in case ∞ is a point of E, is meant a polynomial function of z having positive degree and having a zero at the origin. The proof when U is a set containing ∞ reduces more or less to the proof of Case 2 in the argument presented. We leave the details as an exercise (Exercise 4.1).

There are many situations in which sets of convergence inducing summands for Mittag-Leffler's procedure can be written down explicitly. We indicate one instance of this. Suppose, namely, that $E = \{z_1, z_2, z_3, \cdots\}$ is an infinite discrete subset of $\mathbb{C}$, which for convenience is assumed not to have the origin as an element. A meromorphic function $f : \mathbb{C} \to \widehat{\mathbb{C}}$ is desired that has a simple pole with residue one at each of the points z_n and no other poles. If the series $\sum_{n=1}^{\infty} |z_n|^{-1}$ converges, then

$$f(z) = \sum_{n=1}^{\infty} \frac{1}{z - z_n} \tag{10.1}$$

already fills the bill; i.e., no convergence inducing modifications are needed. The reason: since $|z_n| \to \infty$ as $n \to \infty$ and since for any given compact subset K of $\mathbb{C}$ the inequality

$$\frac{1}{|z - z_n|} \le \frac{2}{|z_n|}$$

holds for every z in K as soon as n is large enough that $|z_n|$ exceeds the number $2\max\{|z| : z \in K\}$, it is evident that the series of meromorphic function in (10.1) is normally convergent in $\mathbb{C}$. If $\sum_{n=1}^{\infty} |z_n|^{-p}$ is convergent for an integer $p > 1$, then the function

$$f(z) = \sum_{n=1}^{\infty} \left(\frac{1}{z - z_n} + \frac{1}{z_n} + \frac{z}{z_n^2} + \cdots + \frac{z^{p-2}}{z_n^{p-1}} \right) \tag{10.2}$$

meets the demand. Indeed, for any compact set K in $\mathbb{C}$ and any n sufficiently large that $c = \max\{|z| : z \in K\} \le |z_n|/2$ we obtain the estimate

$$\left|\frac{1}{z-z_n} + \frac{1}{z_n} + \frac{z}{z_n^2} + \cdots + \frac{z^{p-2}}{z_n^{p-1}}\right| \le \frac{2c^{p-1}}{|z_n|^p} \tag{10.3}$$

for every z in K, as follows easily from the observation that

$$\frac{1}{z-z_n} + \frac{1}{z_n} + \frac{z}{z_n^2} + \cdots + \frac{z^{p-2}}{z_n^{p-1}} = \frac{1}{z-z_n} + \frac{1}{z_n}\,\frac{1-(z/z_n)^{p-1}}{1-(z/z_n)} = \frac{z^{p-1}}{z_n^{p-1}(z-z_n)}\,.$$

Coupled with the convergence of $\sum_{n=1}^{\infty} |z_n|^{-p}$, (10.3) implies that the series in (10.2) converges normally in the whole plane. It is worth noting that the functions in (10.1) and (10.2) do not depend on the particular way we elect to list the elements of E: since the defining series are absolutely convergent in $\mathbb{C} \sim E$, their terms can be rearranged at will (rearrangement is tantamount to relabeling the points of E) without affecting the sum.

A classic example of (10.2) is the function

$$f(z) = \frac{1}{z} + \sum_{|n|\ge 1}\left(\frac{1}{z-n} + \frac{1}{n}\right),$$

whose singularities in $\mathbb{C}$ are simple poles with residue one at all integers n. (Here we apply (10.2) to $E = \{\pm 1, \pm 2, \cdots\}$ with $p = 2$. The use of (10.2) is legitimate, because $\sum_{n=1}^{\infty} |z_n|^{-2} = 2\sum_{n=1}^{\infty} n^{-2} < \infty$. We have also thrown in a pole at the origin for good measure.) As luck would have it, we already know a function with these qualifications, the function $g(z) = \pi\cot(\pi z)$. What is the relationship between f and g? The next example furnishes the answer to this question.

EXAMPLE 1.1. Show that

$$\pi\cot(\pi z) = \frac{1}{z} + \sum_{|n|\ge 1}\left(\frac{1}{z-n} + \frac{1}{n}\right) \tag{10.4}$$

for every complex number z.

Since both sides of (10.4) describe meromorphic functions in $\mathbb{C}$ having poles — hence, taking the value ∞ — at all integers, we need only concern ourselves with verifying (10.4) for non-integral complex numbers z. Fix such a point z and consider the function h defined by

$$h(\zeta) = \frac{\pi\cot(\pi\zeta)}{\zeta(\zeta-z)} - \frac{1}{\zeta^2(\zeta-z)}\,.$$

This function is meromorphic in $\mathbb{C}$. It has a simple pole at the point z, with

$$\operatorname{Res}(z,h) = \frac{\pi\cot(\pi z)}{z} - \frac{1}{z^2}\,,$$

and a simple pole at each non-zero integer n, for which

$$\text{Res}(n,h) = \frac{1}{n(n-z)} = -\frac{1}{z}\left(\frac{1}{z-n} + \frac{1}{n}\right) .$$

The singularity of h at the origin is removable. Let N be any integer satisfying $N > |z|$, and let $Q_N = \{z : |x| \le N + (1/2), |y| \le N + (1/2)\}$. An application of the residue theorem yields

$$(10.5)\quad \frac{1}{2\pi i}\int_{\partial Q_N} h(\zeta)\,d\zeta = \frac{\pi\cot(\pi z)}{z} - \frac{1}{z^2} - \frac{1}{z}\sum_{1\le|n|\le N}\left(\frac{1}{z-n} + \frac{1}{n}\right) .$$

We claim that $\int_{\partial Q_N} h(\zeta)d\zeta \to 0$ as $N \to \infty$. Once this is demonstrated, we shall be in a position to let $N \to \infty$ in (10.5) and conclude that

$$\frac{\pi\cot(\pi z)}{z} - \frac{1}{z^2} - \frac{1}{z}\sum_{|n|\ge 1}\left(\frac{1}{z-n} + \frac{1}{n}\right) = 0 ,$$

from which (10.4) follows immediately.

We observe that

$$(10.6)\qquad |\cot(\pi\zeta)| \le m = \frac{1+e^{-3\pi}}{1-e^{-3\pi}}$$

for every ζ in ∂Q_N. To see this, observe first that for $\zeta = N + (1/2) + it$, where t is an arbitrary real number,

$$|\cot(\pi\zeta)| = \left|\cot(\pi N + 2^{-1}\pi + i\pi t)\right| = |\tan(i\pi t)|$$

$$= \left|\frac{e^{-\pi t} - e^{\pi t}}{e^{-\pi t} + e^{\pi t}}\right| \le \frac{e^{-\pi t} + e^{\pi t}}{e^{-\pi t} + e^{\pi t}} = 1 < m .$$

Similarly, if $\zeta = t + Ni + (i/2)$ for any real t, then

$$|\cot(\pi\zeta)| = \left|\frac{e^{i\pi\zeta} + e^{-i\pi\zeta}}{e^{i\pi\zeta} - e^{-i\pi\zeta}}\right| = \left|\frac{e^{i2\pi\zeta} + 1}{e^{i2\pi\zeta} - 1}\right| \le \frac{1 + |e^{i2\pi\zeta}|}{1 - |e^{i2\pi\zeta}|}$$

$$= \frac{1+e^{-(2N+1)\pi}}{1-e^{-(2N+1)\pi}} \le \frac{1+e^{-3\pi}}{1-e^{-3\pi}} = m .$$

Since $|\cot(\pi\zeta)| = |\cot(-\pi\zeta)|$, the same bound is valid on the other two sides of Q_N, which confirms (10.6). As $|\zeta| \ge N$ for ζ on ∂Q_N, we retrieve from (10.6) the estimate

$$|h(\zeta)| \le \frac{m\pi}{N(N-|z|)} + \frac{1}{N^2(N-|z|)} = \frac{m\pi N + 1}{N^2(N-|z|)}$$

for all such ζ. Because Q_N has perimeter $8N+4$,

$$\left|\int_{\partial Q_N} h(\zeta)\,d\zeta\right| \le \int_{\partial Q_N} |h(\zeta)||d\zeta| \le \frac{(8N+4)(m\pi N+1)}{N^2(N-|z|)} \to 0$$

as $N \to \infty$, just what we needed to know.

The beautiful identity (10.4) spawns a number of identities of the same general type. Here is one sample. Others can be found in the exercises.

EXAMPLE 1.2. Verify that

$$\frac{\pi^2}{\sin^2(\pi z)} = \sum_{n=-\infty}^{\infty} \frac{1}{(z-n)^2} \tag{10.7}$$

for every complex number z.

Owing to Theorem 1.1 a series representation for the derivative of $f(z) = \pi\cot(\pi z)$ can be arrived at by differentiating the right-hand side of (10.4) term by term, a step that results in

$$-\pi^2\csc^2(\pi z) = -\frac{1}{z^2} - \sum_{|n|\ge 1} \frac{1}{(z-n)^2} = -\sum_{n=-\infty}^{\infty} \frac{1}{(z-n)^2}$$

for every complex number z. This is plainly equivalent to (10.7).

1.3 The Weierstrass $\wp$-function

We could not abandon the present topic without saying at least a few words about one of the most significant functions to issue directly from the circle of ideas surrounding Mittag-Leffler's theorem, the so-called "$\wp$-function" of Weierstrass, which turns up in a host of advanced mathematical discussions. Its construction begins with a pair of non-zero complex numbers ω_1 and ω_2 whose ratio $\tau = \omega_2/\omega_1$ is not real. The set $\Omega = \Omega(\omega_1, \omega_2)$ consisting of all complex numbers ω of the form $\omega = n\,\omega_1 + m\,\omega_2$, where n and m are integers, is called the *lattice generated by* ω_1 *and* ω_2 (Figure 1). It is not difficult to see that Ω is a discrete subset of the complex plane. We use Ω^* to signify the set $\Omega \sim \{0\}$. The *Weierstrass function corresponding to the lattice* Ω is the function $\wp$ defined on $\mathbb{C}$ by

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega\in\Omega^*} \left[\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\right]. \tag{10.8}$$

The expanded notation $\wp_\Omega$ is sometimes used for this function, especially in situations where two or more different lattices are under consideration

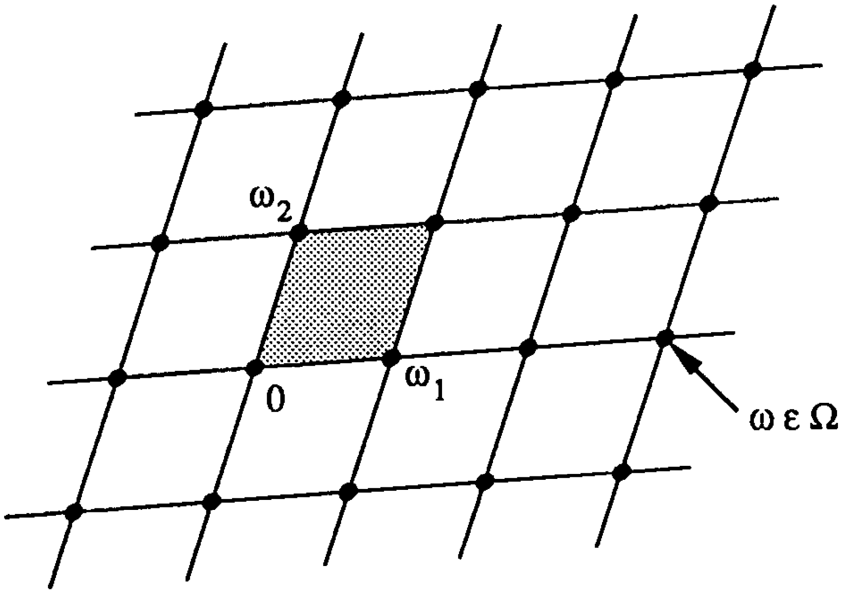


Figure 1.

simultaneously and confusion between their associated $\wp$-functions must be avoided.

We have not previously run into infinite series indexed by sets other than sets of integers, so (10.8) warrants some explanation. (It is not our intention, however, to get bogged down in a full-blown technical discussion of such series.) Since Ω^* is a discrete subset of $\mathbb{C}$, we are free to list its members in a sequence $z_1, z_2, z_3, \cdots$, where $z_n \neq z_m$ for $n \neq m$. We shall demonstrate that the series $\sum_{n=1}^{\infty} |z_n|^{-3}$ is convergent. Assume for a moment that this has already been done. An argument along the lines of those that established convergence in (10.1) and (10.2) can then be fashioned to show that the series of meromorphic functions $\sum_{n=1}^{\infty} \left[(z - z_n)^{-2} - z_n^{-2}\right]$ is normally convergent in $\mathbb{C}$ and absolutely convergent in $\mathbb{C} \sim \Omega^*$. Therefore, the formula

$$\wp(z) = \frac{1}{z^2} + \sum_{n=1}^{\infty} \left[\frac{1}{(z - z_n)^2} - \frac{1}{z_n^2}\right] \tag{10.9}$$

is seen to define a meromorphic function in $\mathbb{C}$ whose set of poles is the lattice Ω and whose singular part at any element ω of Ω is $S(z) = (z-\omega)^{-2}$. Furthermore, the absolute convergence of the series implies that $\wp$ depends only on the set Ω, not on our particular way of arranging the members of Ω^*: if $w_1, w_2, w_3, \cdots$ is a second enumeration of the points of Ω^*, then

$$\wp(z) = \frac{1}{z^2} + \sum_{n=1}^{\infty} \left[\frac{1}{(z - w_n)^2} - \frac{1}{w_n^2}\right]$$

gives an alternate description of the function in (10.9). The last fact justifies writing

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega \in \Omega^*} \left[\frac{1}{(z - \omega)^2} - \frac{1}{\omega^2}\right] ,$$

without any reference to the sequence $z_1, z_2, z_3, \cdots$. Notice, incidentally, that $-z_1, -z_2, -z_3, \cdots$ does represent another way of listing the elements of Ω^*, so we have

$$\wp(z) = \frac{1}{z^2} + \sum_{n=1}^{\infty} \left\{ \frac{1}{[z-(-z_n)]^2} - \frac{1}{(-z_n)^2} \right\} = \frac{1}{z^2} + \sum_{n=1}^{\infty} \left[\frac{1}{(z+z_n)^2} - \frac{1}{z_n^2} \right].$$

On the other hand, it is also true that

$$\wp(-z) = \frac{1}{(-z)^2} + \sum_{n=1}^{\infty} \left[\frac{1}{(-z-z_n)^2} - \frac{1}{z_n^2} \right] = \frac{1}{z^2} + \sum_{n=1}^{\infty} \left[\frac{1}{(z+z_n)^2} - \frac{1}{z_n^2} \right].$$

We deduce that $\wp(-z) = \wp(z)$ for every z in $\mathbb{C}$; i.e., $\wp$ is an even function. Of course, all of the above statements are contingent on the convergence of the series $\sum_{\omega\in\Omega^*} |\omega|^{-3} = \sum_{n=1}^{\infty} |z_n|^{-3}$, something we now confirm.

The function φ defined on the circle $K = K(0,1)$ by $\varphi(x+iy) = |x\,\omega_1 + y\,\omega_2|$ is non-negative and continuous. Were $\varphi(z) = 0$ to hold for some z on K, it would imply that ω_2/ω_1 is real, contrary to our assumption. We infer that $c = 2^{-1}\min\{\varphi(z) : z \in K\}$ is a positive number. We shall show that $\sum_{n=1}^{N} |z_n|^{-3} \le A = 4c^{-3}\sum_{k=1}^{\infty} k^{-2}$ holds for every N, from which the convergence of $\sum_{n=1}^{\infty} |z_n|^{-3}$ follows. Fix N and choose M so that Ω_M^*, the set of points $n\,\omega_1 + m\,\omega_1$ in Ω^* with $|n| + |m| \le M$, includes $z_1, z_2, \ldots, z_N$. Since the lower bound

$$|n\,\omega_1 + m\,\omega_2| = (n^2+m^2)^{1/2}|n(n^2+m^2)^{-1/2}\omega_1 + m(n^2+m^2)^{-1/2}\omega_2|$$

$$\ge 2c(n^2+m^2)^{1/2} \ge c(|n|+|m|)$$

is valid for every element $n\,\omega_1 + m\,\omega_2$ of Ω^* and since the number of integer pairs (n,m) for which $|n|+|m| = k$ is exactly $4k$ when $k \ge 1$, we obtain

$$\sum_{n=1}^{N} \frac{1}{|z_n|^3} \le \sum_{\omega\in\Omega_M^*} \frac{1}{|\omega|^3} = \sum_{k=1}^{M} \sum_{|n|+|m|=k} \frac{1}{|n\,\omega_1 + m\,\omega_1|^3}$$

$$\le \sum_{k=1}^{M} \sum_{|n|+|m|=k} \frac{1}{c^3(|n|+|m|)^3} = \frac{4}{c^3} \sum_{k=1}^{M} \frac{1}{k^2} \le A,$$

as desired.

We arrive at the derivative of the $\wp$-function by differentiating the series in (10.8) termwise,

$$\wp'(z) = -\sum_{\omega\in\Omega} \frac{2}{(z-\omega)^3}. \tag{10.10}$$

The latter series converges normally in $\mathbb{C}$ (Theorem 1.1), and, due to the convergence of $\sum_{\omega\in\Omega^*} |\omega|^{-3}$, it converges absolutely in $\mathbb{C} \sim \Omega$. If we pick a sequential listing $z_1, z_2, z_3, \cdots$ for the members of Ω, then we notice that $z_1 - \omega_1, z_2 - \omega_1, z_3 - \omega_1, \cdots$ provides another such listing. The absolute convergence in (10.10) dictates that

$$\wp'(z) = -\sum_{n=1}^{\infty} \frac{2}{(z - z_n)^3} = -\sum_{n=1}^{\infty} \frac{2}{(z + \omega_1 - z_n)^3}$$

for every z in $\mathbb{C}$. Using the first of these two representations, we remark that

$$\wp'(z + \omega_1) = -\sum_{n=1}^{\infty} \frac{2}{(z + \omega_1 - z_n)^3} .$$

We thus discover a noteworthy property of $\wp'$: $\wp'(z+\omega_1) = \wp'(z)$ for every z in $\mathbb{C}$. Similarly, $\wp'(z+\omega_2) = \wp'(z)$ in $\mathbb{C}$. (More generally, $\wp'(z+\omega) = \wp'(z)$ holds for every ω in Ω.) These identities bestow on $\wp'$ membership in an exclusive class of functions known as "elliptic functions," concerning which there is an extensive literature. (By definition a function f is an *elliptic function* provided it is meromorphic in $\mathbb{C}$ and *doubly periodic*. The last condition demands the existence of a lattice $\Omega = \Omega(\omega_1, \omega_2)$ generated by non-zero complex numbers ω_1 and ω_2 with non-real ratio $\tau = \omega_2/\omega_1$ such that $f(z) = f(z + \omega_1) = f(z + \omega_2)$ for every z in $\mathbb{C}$ or, equivalently, such that $f(z) = f(z + \omega)$ holds in $\mathbb{C}$ for every ω in Ω. It can be shown that, unless f is a constant function, the generators ω_1 and ω_2 here can always be chosen so that Ω includes every period of f, meaning every ω for which $f(z + \omega) = f(z)$ throughout $\mathbb{C}$. If so, we call Ω the *period lattice* of f and refer to ω_1 and ω_2 as a pair of *primitive periods* for f.) The $\wp$-function, too, is an elliptic function, although this fact is not quite as obvious as it was in the case of $\wp'$. To prove it, consider the function $g(z) = \wp(z) - \wp(z + \omega_1)$. Then $g'(z) = \wp'(z) - \wp'(z + \omega_1) = 0$ in $\mathbb{C} \sim \Omega$, making g constant in that domain. Because $\wp$ is an even function and $-\omega_1/2$ is not a point of Ω, $g(-\omega_1/2) = \wp(-\omega_1/2) - \wp(\omega_1/2) = 0$, so g vanishes identically in $\mathbb{C} \sim \Omega$; i.e., $\wp(z) = \wp(z + \omega_1)$ whenever z belongs to $\mathbb{C} \sim \Omega$. Since $\wp(z) = \wp(z + \omega_1) = \infty$ for every z in Ω, $\wp(z) = \wp(z + \omega_1)$ is true everywhere in the complex plane. The same argument shows that $\wp(z + \omega_2) = \wp(z)$ in $\mathbb{C}$. This makes $\wp$ an elliptic function, one which has ω_1 and ω_2 as a set of primitive periods. The function $\wp$ and its derivative play central roles in the general theory of elliptic functions, a subject we are unfortunately without space to probe more deeply here. (Some of the properties of elliptic functions, among them important properties of $\wp$, are explored in the exercises.)

2 The Theorem of Weierstrass

2.1 Infinite Products

Mittag-Leffler's theorem assures us that, given a discrete subset E of a plane open set U and given the assignment of a rational singular part to every point of E, we can construct a meromorphic function in U whose pole-set is E and whose singular part at each point of E is the one prescribed. Weierstrass's theorem answers in the affirmative a related question: Does there exist an analytic function $f: U \to \mathbb{C}$ that has the given discrete set E as its set of zeros, the order of each of these zeros being specified beforehand? As the usual method of creating such a function involves infinite products, it is to this topic that we first direct our attention.

Let $\langle z_n \rangle$ be a sequence of complex numbers. In giving meaning to the infinite product $\prod_{n=1}^{\infty} z_n$ we initially discuss the special case — as things turn out, it is not too special — where all of the factors z_n are non-zero. Mimicking the process that led from $\langle z_n \rangle$ via its corresponding sequence of partial sums s_n to the infinite series $\sum_{n=1}^{\infty} z_n$, we use $\langle z_n \rangle$ to generate an associated *sequence* $\langle p_n \rangle$ *of partial products:* $p_1 = z_1, p_2 = z_1 z_2, \ldots, p_n = z_1 z_2 \cdots z_n, \cdots$. It may happen that $p = \lim_{n\to\infty} p_n$ exists and that $p \neq 0$. Under these conditions we say that the infinite product $\prod_{n=1}^{\infty} z_n$ is *convergent* and that the value of the product is p, a state of affairs we symbolize by writing $p = \prod_{n=1}^{\infty} z_n$. If $\lim_{n\to\infty} p_n$ fails to exist or if $\lim_{n\to\infty} p_n = 0$, then $\prod_{n=1}^{\infty} z_n$ is said to be *divergent*. There are a number of reasons for insisting on the condition $p \neq 0$ as part of the definition of convergence here. Not the least of these is the simple desire to remain algebraically consistent with the situation for finite products: if $z_1, z_2, \ldots, z_n$ are all non-zero, then their product is non-zero as well. To illustrate the definition, let $z_n = 2^{2^{1-n}}$ for $n = 1, 2, 3, \cdots$. Then

$$p_n = 2 \cdot 2^{1/2} \cdots 2^{(1/2)^{n-1}} = 2^{1+(1/2)+\cdots+(1/2)^{n-1}} = 2^{2-(1/2)^{n-1}},$$

which yields

$$\prod_{n=1}^{\infty} 2^{2^{1-n}} = \lim_{n\to\infty} 2^{2-(1/2)^{n-1}} = 2^2 = 4.$$

As a second example, consider the sequence $z_n = n^{-1}$. In this instance $p_n = (n!)^{-1} \to 0$ as $n \to \infty$, so the product $\prod_{n=1}^{\infty} n^{-1}$ is by definition divergent. We remark that an infinite product $\prod_{n=1}^{\infty} z_n$ of non-zero complex numbers converges if and only if for each $N \geq 1$ the truncated product $\prod_{n=N+1}^{\infty} z_n = \prod_{n=1}^{\infty} z_{N+n}$ is convergent. Furthermore, it is then true that

$$\prod_{n=1}^{\infty} z_n = p_N \cdot \prod_{n=N+1}^{\infty} z_n \tag{10.11}$$

for every N.

The transition from the special case in which the factors are non-zero to a more general infinite product $\prod_{n=1}^{\infty} z_n$ is accomplished as follows: $\prod_{n=1}^{\infty} z_n$ is called *convergent* if there exists an index N such that $z_n \neq 0$ holds whenever $n > N$ and such that the truncated product $\prod_{n=N+1}^{\infty} z_n$, whose factors are non-zero, converges according to the earlier definition, in which event we define

$$\prod_{n=1}^{\infty} z_n = p_N \cdot \prod_{n=N+1}^{\infty} z_n .$$

Otherwise, $\prod_{n=1}^{\infty} z_n$ is pronounced *divergent.* Because of (10.11) it is evident that the value of $\prod_{n=1}^{\infty} z_n$ does not depend on which N is used here, as long as it has the two properties stated, and that the generalized understanding of an infinite product reduces to the original one when all the factors z_n are non-zero. Formula (10.11) remains in effect for an arbitrary convergent product. We emphasize: $\prod_{n=1}^{\infty} z_n$ is automatically divergent if $z_n = 0$ for infinitely many n; under the assumption that $\prod_{n=1}^{\infty} z_n$ is convergent, $\prod_{n=1}^{\infty} z_n = 0$ only when $z_n = 0$ for one or more values of n. It follows almost immediately from these definitions that, if $\prod_{n=1}^{\infty} z_n$ and $\prod_{n=1}^{\infty} w_n$ are convergent infinite products, then the product $\prod_{n=1}^{\infty}(z_n w_n)$ also converges, with

$$\prod_{n=1}^{\infty}(z_n w_n) = \left(\prod_{n=1}^{\infty} z_n\right)\left(\prod_{n=1}^{\infty} w_n\right) . \tag{10.12}$$

A simple condition which is necessary (but far from sufficient) for the convergence of $\prod_{n=1}^{\infty} z_n$ is that $\lim_{n\to\infty} z_n = 1$: assuming that $z_n \neq 0$ for every $n > N$ and writing $q_n = \prod_{k=1}^{n} z_{N+k}$, we compute

$$\lim_{n\to\infty} z_n = \lim_{n\to\infty} \frac{q_n}{q_{n-1}} = \frac{\prod_{n=N+1}^{\infty} z_n}{\prod_{n=N+1}^{\infty} z_n} = 1 .$$

It is sometimes convenient to express a product $\prod_{n=1}^{\infty} z_n$ in the form $\prod_{n=1}^{\infty}(1 + w_n)$ by writing w_n for $z_n - 1$. In this notation the necessary condition for convergence translates to $\lim_{n\to\infty} w_n = 0$.

The problem of testing an infinite product for convergence can be transformed to the task of checking whether a related infinite series converges or diverges.

Theorem 2.1. *Let $\langle z_n \rangle$ be a sequence of non-zero complex numbers. The infinite product $\prod_{n=1}^{\infty} z_n$ converges if and only if the associated infinite series $\sum_{n=1}^{\infty} \operatorname{Log} z_n$ converges, in which case*

$$\prod_{n=1}^{\infty} z_n = \exp\left(\sum_{n=1}^{\infty} \operatorname{Log} z_n\right) . \tag{10.13}$$

Proof. Write $p_n = z_1 z_2 \cdots z_n$ and $s_n = \operatorname{Log} z_1 + \operatorname{Log} z_2 + \cdots + \operatorname{Log} z_n$. Then $e^{s_n} = p_n$. If the series $\sum_{n=1}^{\infty} \operatorname{Log} z_n$ is convergent with sum s, then $p_n = e^{s_n} \to e^s$ as $n \to \infty$; i.e., $\prod_{n=1}^{\infty} z_n$ converges and (10.13) holds. The converse is slightly more delicate. Assume that the infinite product $\prod_{n=1}^{\infty} z_n$ converges, and call the product p. By definition, $p \neq 0$. We may further assume for the purposes of this proof that p is not a point of the negative real axis. (If p does happen to be in the interval $(-\infty, 0)$, we simply consider in place of $\langle z_n \rangle$ a new sequence $\langle w_n \rangle$; namely, $w_1 = -z_1$ and $w_n = z_n$ for $n \geq 2$. Then $\prod_{n=1}^{\infty} w_n = -p$ does not belong to $(-\infty, 0)$. The two-series $\sum_{n=1}^{\infty} \operatorname{Log} z_n$ and $\sum_{n=1}^{\infty} \operatorname{Log} w_n$, differing only in their first terms, either converge or diverge together.) Thus, we may assume that the principal logarithm function is continuous at p. Since $p_n \to p$, we conclude that $\operatorname{Log} p_n \to \operatorname{Log} p$. The relation $e^{s_n} = p_n$ stamps s_n as a logarithm of p_n. Unfortunately, s_n need not be the principal logarithm of p_n. We can, however, express this number in the manner $s_n = \operatorname{Log} p_n + 2k_n \pi i$ for some integer k_n. The convergence of $\prod_{n=1}^{\infty} z_n$ implies that $z_n \to 1$ as $n \to \infty$. Consequently,

$$\begin{aligned}\lim_{n\to\infty}(k_{n+1} - k_n) &= (2\pi i)^{-1} \lim_{n\to\infty}(s_{n+1} - s_n - \operatorname{Log} p_{n+1} + \operatorname{Log} p_n) \\ &= (2\pi i)^{-1} \lim_{n\to\infty}(\operatorname{Log} z_{n+1} - \operatorname{Log} p_{n+1} + \operatorname{Log} p_n) \\ &= (2\pi i)^{-1}(0 - \operatorname{Log} p + \operatorname{Log} p) = 0 .\end{aligned}$$

Because $k_{n+1} - k_n$ is an integer, there must exist an index N with the property that $k_{n+1} - k_n = 0$ once $n \geq N$; i.e., $k_n = k_N = k$ for every $n \geq N$. We infer that $s_n = \operatorname{Log} p_n + 2k_n \pi i \to \operatorname{Log} p + 2k\pi i$. In other words, the series $\sum_{n=1}^{\infty} \operatorname{Log} z_n$ converges, its sum being $\operatorname{Log} p + 2k\pi i$ for some integer k. We have already observed that the convergence of $\sum_{n=1}^{\infty} \operatorname{Log} z_n$ guarantees the validity of (10.13). ∎

If it is only known that $z_n \neq 0$ for $n > N$, then it is the convergence of $\sum_{n=N+1}^{\infty} \operatorname{Log} z_n$ that is equivalent to the convergence of $\prod_{n=1}^{\infty} z_n$.

There is a notion of absolute convergence for infinite products that is analogous to absolute convergence for infinite series. The definition is most easily stated for products of the type $\prod_{n=1}^{\infty}(1 + w_n)$: $\prod_{n=1}^{\infty}(1 + w_n)$ is ***absolutely convergent*** provided $\prod_{n=1}^{\infty}(1 + |w_n|)$ is convergent. Since

$$\lim_{w\to 0} \frac{|\operatorname{Log}(1+w)|}{|w|} = 1 ,$$

the comparison test implies that, for any complex sequence $\langle w_n \rangle$ tending to zero, the three series

$$(10.14) \quad \sum_{n=1}^{\infty} |\operatorname{Log}(1 + w_n)| \quad , \quad \sum_{n=1}^{\infty} \operatorname{Log}(1 + |w_n|) \quad , \quad \sum_{n=1}^{\infty} |w_n|$$

either converge or diverge in unison. (In the first series ignore those terms — there are at most a finite number of them — in which $w_n = -1$.) According to Theorem 2.1 the infinite product $\prod_{n=1}^{\infty}(1+|w_n|)$ converges if and only if the infinite series $\sum_{n=1}^{\infty} \operatorname{Log}(1+|w_n|)$ does. As a consequence, the absolute convergence of a product $\prod_{n=1}^{\infty}(1+w_n)$ is equivalent to the convergence of any (hence, all) of the series in (10.14). For example, the fact that the series $\sum_{n=1}^{\infty} n^{-2}$ converges implies that the infinite product $\prod_{n=1}^{\infty}(1+n^{-2}z)$ is absolutely convergent for every complex number z, whereas the divergence of $\sum_{n=1}^{\infty} n^{-1}$ means that $\prod_{n=1}^{\infty}[1+(-1)^n n^{-1}z]$ converges absolutely only if $z = 0$. For a product written in the straightforward fashion $\prod_{n=1}^{\infty} z_n$, with $z_n \to 1$, it is the first of the series in (10.14) that is the simplest to deal with when it comes to articulating a convenient criterion for absolute convergence: $\prod_{n=1}^{\infty} z_n$ converges absolutely if and only if $\sum_{n=N}^{\infty} |\operatorname{Log} z_n|$ is convergent, where N is taken large enough to insure that $z_n \neq 0$ is satisfied whenever $n \geq N$. Note especially that the absolute convergence of a product $\prod_{n=1}^{\infty} z_n$ imparts absolute convergence — and, with it, convergence — to the series $\sum_{n=N}^{\infty} \operatorname{Log} z_n$, as soon as N is suitably large. On the basis of Theorem 2.1 we can thus assert that an absolutely convergent infinite product is, in fact, convergent. Furthermore, the factors in an absolutely convergent infinite product can be rearranged at will without fear of upsetting the convergence or changing the value of the product: if $\prod_{n=1}^{\infty} z_n$ converges absolutely, then $\prod_{n=1}^{\infty} z_n = \prod_{n=1}^{\infty} z_{\sigma(n)}$ for every permutation σ of the positive integers. When all factors z_n are non-zero, this follows from (10.13) and the corresponding fact for absolutely convergent series,

$$\prod_{n=1}^{\infty} z_n = \exp\left(\sum_{n=1}^{\infty} \operatorname{Log} z_n\right) = \exp\left(\sum_{n=1}^{\infty} \operatorname{Log} z_{\sigma(n)}\right) = \prod_{n=1}^{\infty} z_{\sigma(n)} \ ;$$

when $z_n = 0$ for some n, it is not hard to see that $\prod_{n=1}^{\infty} z_n = \prod_{n=1}^{\infty} z_{\sigma(n)} = 0$.

2.2 Infinite Products of Functions

Assume that the domain-set of each function in a sequence $\langle f_n \rangle$ includes a subset A of the complex plane. We proclaim the infinite product $\prod_{n=1}^{\infty} f_n$ *pointwise convergent in* A if for each fixed z in A the infinite product $\prod_{n=1}^{\infty} f_n(z)$, whose n^{th} factor is the value of f_n at the given point, converges. This being the case, we can define a function $f: A \to \mathbb{C}$ through the rule of correspondence $f(z) = \prod_{n=1}^{\infty} f_n(z)$. Naturally, we write $f = \prod_{n=1}^{\infty} f_n$ in A. Owing to the way in which infinite products have been defined, the set of zeros of the product f is nothing but the union $\cup_{n=1}^{\infty} Z_n$ of the sets $Z_n = \{z \in A : f_n(z) = 0\}$.

As was true for infinite series of functions, pointwise convergence for infinite products of functions is just too weak a form of convergence to serve adequately our purposes. Once again the mode of convergence that best meets our needs is uniform convergence — or, to be more exact, normal convergence. An infinite product of functions $\prod_{n=1}^{\infty} f_n$ is said to *converge uniformly on a set* A provided (i) $\prod_{n=1}^{\infty} f_n$ is pointwise convergent in A and (ii) there exists an index N with the following two properties: f_n is free of zeros in A whenever $n > N$ and the partial product $f_{N+1}f_{N+2}\cdots f_{N+n}$ tends to the truncated infinite product $\prod_{n=N+1}^{\infty} f_n$ uniformly on A as $n \to \infty$. (N.B. Under this definition, which is a standard one, it is not automatically the case that the uniform convergence of $\prod_{n=1}^{\infty} f_n$ on a set A carries with it the uniform convergence on A of all the truncated products $\prod_{n=N+1}^{\infty} f_n, N = 1, 2, 3, \cdots$. Extra information is needed if such a conclusion is to be drawn. For instance, if each individual factor f_n happens to satisfy a condition $a_n \le |f_n(z)| \le b_n$ for every z in A, where a_n and b_n are positive constants — this is definitely true when the set A is compact and each factor f_n is both continuous and zero-free in A — then the uniform convergence of $\prod_{n=1}^{\infty} f_n$ on A does imply the same type of convergence for all truncated products.) The statement that an infinite product $\prod_{n=1}^{\infty} f_n$ *converges normally in an open set* U means that it converges uniformly on each compact subset of U. When the factors f_n are continuous functions in U, the normal convergence of $\prod_{n=1}^{\infty} f_n$ in this open set is implied by the uniform convergence of this product on every closed disk in U. The most frequently cited criterion for the uniform convergence of an infinite product is a carry-over of the Weierstrass M-test from series to products.

Theorem 2.2. *Suppose that each of the factors in an infinite product of functions $\prod_{n=1}^{\infty} f_n$ is defined on a set A. If there exists a sequence $\langle M_n \rangle$ of real numbers such that $|f_n(z) - 1| \le M_n$ is satisfied for every z in A and such that the series $\sum_{n=1}^{\infty} M_n$ converges, then $\prod_{n=1}^{\infty} f_n$ converges absolutely and uniformly on A.*

Proof. Set $M = \prod_{n=1}^{\infty}(1 + M_n)$. The fact that $\sum_{n=1}^{\infty} M_n$ converges and $M_n \ge 0$ accounts for the convergence of this product. (Recall the discussion surrounding (10.14).) Moreover, on the basis of the comparison test we can assert that for each fixed z in A the series $\sum_{n=1}^{\infty} |f_n(z) - 1|$ is convergent, a detail that insures the absolute convergence of the product $\prod_{n=1}^{\infty} f_n(z)$. It follows that $\prod_{n=1}^{\infty} f_n$ is at least pointwise convergent in A. We now select and fix an index N with the property that $\sum_{k=N+1}^{\infty} M_k < 1/2$. If $n > N$, then the function f_n can have no zeros in A, for the reason that the inequality $|f_n(z) - 1| \le M_n \le \sum_{k=N+1}^{\infty} M_k < 1/2$ is in force there. To finish the proof it is enough to demonstrate that the convergence of $g_n = f_{N+1}f_{N+2}\cdots f_{N+n}$ to $g = \prod_{n=N+1}^{\infty} f_n$ is uniform on A. We shall certainly accomplish this if we can verify that the estimate

$$|g(z) - g_n(z)| \le 2eM \sum_{k=N+n+1}^{\infty} M_k \tag{10.15}$$

is valid in A, for the right-hand side of (10.15) is independent of z and tends to 0 as $n \to \infty$. Fix $n \geq 1$. Since $|f_k(z) - 1| < 1/2$ holds everywhere in A once $k > N$ and since

$$|\operatorname{Log}(1+w)| = \left| \int_1^{1+w} \frac{dz}{z} \right| \leq \int_1^{1+w} \frac{|dz|}{|z|} \leq 2|w|$$

when $|w| < 1/2$, our choice of N determines that

$$\left| \sum_{k=N+n+1}^{\infty} \operatorname{Log} f_k(z) \right| \leq \sum_{k=N+n+1}^{\infty} |\operatorname{Log} f_k(z)|$$

$$\leq 2 \sum_{k=N+n+1}^{\infty} |f_k(z) - 1| \leq 2 \sum_{k=N+n+1}^{\infty} M_k < 1$$

for every point z of A. To repeat, the inequality

$$\left| \sum_{k=N+n+1}^{\infty} \operatorname{Log} f_k(z) \right| \leq 2 \sum_{k=N+n+1}^{\infty} M_k < 1 \tag{10.16}$$

holds throughout A. When $|w| \leq 1$ one obtains a simple bound for $|e^w - 1|$ via

$$|e^w - 1| = \left| \int_0^w e^z \, dz \right| \leq \int_0^w e^{\operatorname{Re} z} |dz| \leq e|w| ,$$

so referring first to (10.13) and then to (10.16) we estimate

$$|g(z) - g_n(z)| = |f_{N+1}(z)| \cdots |f_{N+n}(z)| \left| \exp\left[\sum_{k=N+n+1}^{\infty} \operatorname{Log} f_k(z) \right] - 1 \right|$$

$$\leq (1 + M_{N+1}) \cdots (1 + M_{N+n}) \cdot e \left| \sum_{k=N+n+1}^{\infty} \operatorname{Log} f_k(z) \right|$$

$$\leq 2e \prod_{n=1}^{\infty} (1 + M_n) \cdot \sum_{k=N+n+1}^{\infty} M_k = 2eM \sum_{k=N+n+1}^{\infty} M_k$$

for every z belonging to A. This confirms (10.15) and, in so doing, completes the proof of the theorem. ∎

2.3 Infinite Products and Analytic Functions

The analogue of Theorem VII.3.2 in the context of infinite products is the following proposition.

Theorem 2.3. *Suppose that each of the factors in an infinite product of functions $\prod_{n=1}^{\infty} f_n$ is analytic in an open set U and that $\prod_{n=1}^{\infty} f_n$ converges normally in U, with the function f as its product. Then f is analytic in U. Furthermore, under the assumption that f does not vanish identically in any component of U, the series of meromorphic functions $\sum_{n=1}^{\infty}(f_n'/f_n)$ is normally convergent in U, where its sum is the function f'/f.*

Proof. To prove that f is analytic in U we need only certify its analyticity in the interior of every closed disk that is contained in U. Let $K = \overline{\Delta}(z_0, r)$ be such a disk. The normal convergence of $\prod_{n=1}^{\infty} f_n$ in U demands the uniform convergence of this infinite product on K. We are thus able to fix an index N such that f_n is free of zeros in K as soon as $n > N$ and such that $g_n = f_{N+1}f_{N+2}\cdots f_{N+n}$ converges to the truncated product $g = \prod_{n=N+1}^{\infty} f_n$ uniformly on K as $n \to \infty$. It follows, in particular, that $g_n \to g$ normally in the open disk $\Delta = \Delta(z_0, r)$. Since each of the functions g_n is analytic in Δ, Theorem VII.3.1 attests to the fact that g is analytic in Δ. By (10.11) we are allowed to write $f = f_1 f_2 \cdots f_N g$ in Δ, which makes evident the analyticity of f there.

Assume now that f does not vanish identically in any component of U. Then plainly none of its factors f_n can be identically zero in a component of U. Accordingly, the functions f'/f and f_n'/f_n for $n \geq 1$ are all meromorphic in U. We show that, given a closed disk $K = \overline{\Delta}(z_0, r)$ in U, we can produce an index N such that no f_n with $n > N$ has a zero in K and such that the series $\sum_{n=N+1}^{\infty}(f_n'/f_n)$ converges uniformly on K. In the process we shall check that $f'/f = \sum_{n=1}^{\infty}(f_n'/f_n)$ in K. The obvious implication of these remarks is that, in the sense appropriate to series of meromorphic functions, $\sum_{n=1}^{\infty}(f_n'/f_n)$ converges normally in U, where its sum is f'/f. In order to carry out this program we are forced to work in a disk slightly larger than the given one. For this reason we fix $s > r$ with the property that the disk $K_0 = \overline{\Delta}(z_0, s)$ still lies in U. Due to the uniform convergence of $\prod_{n=1}^{\infty} f_n$ on K_0, we can fix an index N so that for every $n > N$ the function f_n is zero-free in K_0 and so that $g_n = f_{N+1}f_{N+2}\cdots f_{N+n} \to g = \prod_{n=N+1}^{\infty} f_n$ uniformly on K_0. Notice that the functions g_n and g are all free of zeros in K_0. As $g_n \to g$ normally in the disk $\Delta_0 = \Delta(z_0, s)$, Theorem VII.3.1 tells us that $g_n' \to g'$ normally in Δ_0. We make the further claim that in the present situation the convergence of g_n'/g_n to g'/g is also normal in Δ_0. Let us suppose for an instant that we can support this claim. An easy calculation reveals that

$$\frac{g_n'}{g_n} = \frac{f_{N+1}'}{f_{N+1}} + \frac{f_{N+2}'}{f_{N+2}} + \cdots + \frac{f_{N+n}'}{f_{N+n}},$$

the n^{th} partial sum of the series $\sum_{n=N+1}^{\infty}(f_n'/f_n)$. We infer that this series is normally convergent in Δ_0 — hence, uniformly convergent on K — its sum there being g'/g. Because $f = f_1 f_2 \cdots f_N g$ in Δ_0, we thus obtain

$$\frac{f'}{f} = \frac{f_1'}{f_1} + \frac{f_2'}{f_2} + \cdots + \frac{f_N'}{f_N} + \frac{g'}{g} = \sum_{n=1}^{\infty} \frac{f_n'}{f_n}$$

in this disk.

All that is left to prove is that $g_n'/g_n \to g'/g$ uniformly on each compact set in Δ_0. Fix such a set, say A. Because the functions involved are continuous and nowhere zero in A, the numbers $a_n = \min\{|g_n(z)| : z \in A\}$ and $a = \min\{|g(z)| : z \in A\}$ are all positive. Also, the fact that $g_n \to g$ uniformly on A implies that $a_n \to a$. As a result, it is clear that $b = \min\{a, a_1, a_2, \cdots\} > 0$. Let $c = b^{-2} \max\{|g(z)| + |g'(z)| : z \in A\}$. Then for every z in A and every $n \geq 1$ we have

$$\begin{aligned}\left|\frac{g'(z)}{g(z)} - \frac{g_n'(z)}{g_n(z)}\right| &= \frac{|g_n(z)g'(z) - g(z)g_n'(z)|}{|g(z)||g_n(z)|} \\ &\leq \frac{|g'(z)||g_n(z) - g(z)| + |g(z)||g'(z) - g_n'(z)|}{|g(z)||g_n(z)|} \\ &\leq c|g(z) - g_n(z)| + c|g'(z) - g_n'(z)| \,.\end{aligned}$$

The last expression tends to zero uniformly on A, thereby forcing the uniform convergence of g_n'/g_n to g'/g on this set. ∎

The efforts we have invested in the last two theorems have an immediate payoff in the form of a lovely identity.

EXAMPLE 2.1. Show that

$$\sin(\pi z) = \pi z \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{n^2}\right) \tag{10.17}$$

for every complex number z.

For $n = 1, 2, 3, \cdots$ set $f_n(z) = 1 - (z/n)^2$. If K is a compact set in $\mathbb{C}$ and if $c = \max\{|z|^2 : z \in K\}$, then

$$|f_n(z) - 1| = \frac{|z|^2}{n^2} \leq \frac{c}{n^2}$$

for every z in K. Since the series $\sum_{n=1}^{\infty} n^{-2}$ converges, Theorem 2.2 allows us to declare the product $\prod_{n=1}^{\infty} f_n$ uniformly convergent on K. It follows that $\prod_{n=1}^{\infty} f_n$ converges normally in the whole complex plane. We conclude by way of Theorem 2.3 that the formula $f(z) = z \prod_{n=1}^{\infty}[1 - (z/n)^2]$ defines an entire function, one with a simple zero at every integer and no other zeros. Theorem 2.3 also provides a little extra information about f — namely, that the meromorphic function f'/f admits the representation

$$\frac{f'(z)}{f(z)} = \frac{1}{z} + \sum_{n=1}^{\infty} \frac{2z}{z^2 - n^2}$$

in $\mathbb{C}$. In Example 1.1 we learned that for all complex numbers z

$$\pi\cot(\pi z) = \frac{1}{z} + \sum_{|n|\geq 1}\left(\frac{1}{z-n}+\frac{1}{n}\right)$$

$$= \frac{1}{z} + \sum_{n=1}^{\infty}\left(\frac{1}{z-n}+\frac{1}{n}+\frac{1}{z+n}-\frac{1}{n}\right)$$

$$= \frac{1}{z} + \sum_{n=1}^{\infty}\frac{2z}{z^2-n^2}\,.$$

The identity of f'/f is thus revealed to us.

The function $g(z) = \sin(\pi z)$ is a second entire function whose only zeros are simple zeros at the integers. Therefore, the singularities of $h = f/g$ in $\mathbb{C}$ are removable. Upon their removal h becomes an entire function — in fact, an entire function without zeros. What is more,

$$h'(z) = \frac{h'(z)h(z)}{h(z)} = h(z)\left[\frac{f'(z)}{f(z)} - \frac{g'(z)}{g(z)}\right] = h(z)[\pi\cot(\pi z) - \pi\cot(\pi z)] = 0$$

throughout the complex plane, making h a constant function there. Since $h(0) = \lim_{z\to 0}[f(z)/g(z)] = \pi^{-1}$, the sole value that h assumes is π^{-1}. This leads to

$$\sin(\pi z) = g(z) = \frac{f(z)}{h(z)} = \pi z\prod_{n=1}^{\infty}\left(1-\frac{z^2}{n^2}\right)$$

for every z.

Consider again a plane open set U and a discrete subset $E = \{z_1, z_2, \cdots\}$ of U. Given a sequence $\langle m_n\rangle$ of positive integers, we would like to construct a function f that is analytic in U and that has E as its set of zero there, the zero at z_n having order m_n. If $E = \{z_1, z_2, \ldots, z_p\}$ is a finite set, this creates no problem: we just take $f(z) = (z-z_1)^{m_1}(z-z_2)^{m_2}\cdots(z-z_p)^{m_p}$. One might expect that $f(z) = \prod_{n=1}^{\infty}(z-z_n)^{m_n}$ would do the trick in the case of infinite E, but such hopes are dashed by the fact that this infinite product does not, in general, converge normally in U. Something akin to the convergence inducing summands of Mittag-Leffler's construction is called for here. Fortunately, precisely the right thing is available, as the proof of the main result in this section makes apparent.

Theorem 2.4. (Weierstrass's Theorem) *Let $E = \{z_1, z_2, z_3, \cdots\}$ be a discrete subset of an open set U in the complex plane, and for $n = 1, 2, 3, \cdots$ let m_n be a positive integer. There exists an analytic function $f: U \to \mathbb{C}$ that has E as its set of zeros, the zero at z_n being one of order m_n for $n = 1, 2, 3, \cdots$. The quotient of any two functions fitting this description is both analytic and zero-free in U.*

Proof. We assume that the set E is infinite, having already dealt with the situation for a finite set E in essentially trivial fashion. We also assume for the purposes of the construction in this proof that the origin is not a point of E. (If the origin does belong to E, we simply carry out the ensuing construction for the set $E \sim \{0\}$ and then multiply the result by z^m, where m is the order stipulated for the zero at the origin.) Let δ_n denote the radius of the largest open disk centered at z_n that is contained in U. Mirroring the proof of Mittag-Leffler's theorem, we initially consider two special cases.

Case 1: $|z_n|\delta_n \geq 1$ for every n. (N.B. If $U = \mathbb{C}$, then Case 1 is the general case.) We recall that $|z_n| \to \infty$ as $n \to \infty$ in this case. The function $L_n(z) = \text{Log}(1 - z_n^{-1}z)$ is analytic in the open disk $\Delta(0, |z_n|)$, where its Taylor series expansion reads

$$\text{Log}\left(1 - \frac{z}{z_n}\right) = -\sum_{k=1}^{\infty} \frac{1}{k}\left(\frac{z}{z_n}\right)^k .$$

For $n = 1, 2, 3, \cdots$ we choose and fix an index $d(n)$ with the property that

$$m_n \sum_{k=d(n)+1}^{\infty} \frac{1}{k}\left(\frac{1}{2}\right)^k \leq \left(\frac{1}{2}\right)^n . \tag{10.18}$$

Our reason for doing this is that the bound

$$\left| m_n \,\text{Log}\left(1 - \frac{z}{z_n}\right) + m_n \sum_{k=1}^{d(n)} \frac{1}{k}\left(\frac{z}{z_n}\right)^k \right| \leq \left(\frac{1}{2}\right)^n \tag{10.19}$$

then holds for every z in the closed disk $K_n = \overline{\Delta}(0, |z_n|/2)$, as follows from (10.18) via the computation

$$\left| m_n \,\text{Log}\left(1 - \frac{z}{z_n}\right) + m_n \sum_{k=1}^{d(n)} \frac{1}{k}\left(\frac{z}{z_n}\right)^k \right| = \left| -m_n \sum_{k=d(n)+1}^{\infty} \frac{1}{k}\left(\frac{z}{z_n}\right)^k \right|$$

$$\leq m_n \sum_{k=d(n)+1}^{\infty} \frac{1}{k}\left(\frac{|z|}{|z_n|}\right)^k \leq m_n \sum_{k=d(n)+1}^{\infty} \frac{1}{k}\left(\frac{1}{2}\right)^k \leq \left(\frac{1}{2}\right)^n .$$

We now define a function f_n by

$$f_n(z) = \left(1 - \frac{z}{z_n}\right)^{m_n} \exp\left[m_n \sum_{k=1}^{d(n)} \frac{1}{k}\left(\frac{z}{z_n}\right)^k\right] .$$

This function is clearly an entire function whose only zero is one of order m_n at z_n. Referring to (10.19) and remembering that $|e^w - 1| \leq e|w|$ when

$|w| \le 1$, we compute for any z in the set K_n

$$|f_n(z) - 1| = \left| \exp\left[m_n \operatorname{Log}\left(1 - \frac{z}{z_n}\right) + m_n \sum_{k=1}^{d(n)} \frac{1}{k}\left(\frac{z}{z_n}\right)^k \right] - 1 \right|$$

$$\le e \left| m_n \operatorname{Log}\left(1 - \frac{z}{z_n}\right) + m_n \sum_{k=1}^{d(n)} \frac{1}{k}\left(\frac{z}{z_n}\right)^k \right| \le 2^{-n} e \ ;$$

i.e., the estimate

$$|f_n(z) - 1| \le 2^{-n} e \tag{10.20}$$

is valid throughout K_n. Let K be an arbitrary compact set in the complex plane. The fact that $|z_n| \to \infty$ as $n \to \infty$ means that K is a subset of K_n — hence, that (10.20) holds for every z in K — once n is sufficiently large. This implies that the series $\sum_{n=1}^{\infty} M_n$ is convergent, where we take $M_n = \max\{|f_n(z) - 1| : z \in K\}$. On the strength of Theorem 2.2 we conclude that the product $\prod_{n=1}^{\infty} f_n$ converges uniformly on K. Consequently, this product is normally convergent in $\mathbb{C}$. Theorem 2.3 informs us that $f = \prod_{n=1}^{\infty} f_n$ is an entire function, one whose zero-set is E and whose zero at z_n has order m_n for $n = 1, 2, 3, \cdots$. In Case 1, therefore, we have exhibited a function with all the desired features (and even more, for f is an entire function, not just a function that is analytic in U).

Case 2: $|z_n|\delta_n < 1$ for every n. Here, as we recall from the proof of Mittag-Leffler's theorem, $\delta_n \to 0$ as $n \to \infty$. Copying what we did in that argument, we select and fix for $n = 1, 2, 3, \cdots$ a point of ζ_n of ∂U for which $|z_n - \zeta_n| = \delta_n$. Then $|(z_n - \zeta_n)/(z - \zeta_n)| < 1$ holds for every point z of the annulus $G_n = \{z : \delta_n < |z - \zeta_n| < \infty\}$, which means that the function L_n defined in G_n by

$$L_n(z) = \operatorname{Log}\left[1 - \left(\frac{z_n - \zeta_n}{z - \zeta_n}\right)\right] = \operatorname{Log}\left(\frac{z - z_n}{z - \zeta_n}\right)$$

is analytic. If we represent L_n in G_n by its Laurent expansion centered at ζ_n, we learn that

$$\operatorname{Log}\left(\frac{z - z_n}{z - \zeta_n}\right) = -\sum_{k=1}^{\infty} \frac{1}{k}\left(\frac{z_n - \zeta_n}{z - \zeta_n}\right)^k$$

for every z in G_n. Also, if $d(n)$ is again selected to insure that (10.18) is true, we discover in the present situation that

$$\left| m_n \operatorname{Log}\left(\frac{z - z_n}{z - \zeta_n}\right) + m_n \sum_{k=1}^{d(n)} \frac{1}{k}\left(\frac{z_n - \zeta_n}{z - \zeta_n}\right)^k \right| \le \left(\frac{1}{2}\right)^n \tag{10.21}$$

whenever z belongs to the set $A_n = \{z : 2\delta_n \le |z - \zeta_n| < \infty\}$.

In Case 2 we introduce the function g_n:

$$g_n(z) = \left(\frac{z - z_n}{z - \zeta_n}\right)^{m_n} \exp\left[m_n \sum_{k=1}^{d(n)} \frac{1}{k}\left(\frac{z_n - \zeta_n}{z - \zeta_n}\right)^k\right].$$

Since ζ_n lies on ∂U, g_n is analytic in U, where its sole zero is a zero of multiplicity m_n at z_n. From (10.21) we conclude by means of calculations similar to those which produced (10.20) in Case 1 that the bound

$$|g_n(z) - 1| \le 2^{-n} e \tag{10.22}$$

holds for every element z of A_n. The fact that $\delta_n \to 0$ implies that any given compact subset K of D is contained in A_n as soon as n is suitably large. In conjunction with (10.22) and Theorem 2.2 this observation lets us know that the product $\prod_{n=1}^{\infty} g_n$ converges uniformly on each such K — hence, normally in U. Once more invoking Theorem 2.3 we find in Case 2 that the function $f = \prod_{n=1}^{\infty} g_n$ meets all the requirements of the theorem.

An infinite discrete subset E of U that does not already come under the umbrella of Case 1 or Case 2 can be written as a disjoint union $E = E_1 \cup E_2$, where E_1 either is finite or fits into Case 1 and where E_2 is either finite or covered by Case 2. We can thus construct analytic functions g and h in U whose respective zero-sets are E_1 and E_2, each zero being of the order stipulated in the specifications for E. Then $f = gh$ has all the properties demanded by the theorem. If $\tilde{f}$ is a second function with the same properties, then the quotient $\tilde{f}/f$ has only removable singularities in U, at each of which it has a non-zero limit. Upon removal of these singularities, $\tilde{f}/f$ becomes analytic and zero-free in U, which explains the final assertion of the theorem. ∎

Except for changing the word "analytic" to "holomorphic," Theorem 2.4 remains valid for an open set U in the extended complex plane $\widehat{\mathbb{C}}$, provided $U \ne \widehat{\mathbb{C}}$. (Since the only functions holomorphic in all of $\widehat{\mathbb{C}}$ are constant functions, Theorem 2.4 cannot be true for $U = \widehat{\mathbb{C}}$.) The proof is left as an exercise for the reader, with the following reminder: for a function f to have a zero of order m at ∞ means that $g(z) = f(z^{-1})$ has a zero of order m at the origin.

An immediate consequence of Theorem 2.4 is that an entire function f (to avoid trivial cases, assume that f is not identically zero but has an infinite number of zeros) admits representations of the type

$$f(z) = g(z)\, z^m \prod_{n=1}^{\infty} \left(1 - \frac{z}{z_n}\right)^{m_n} E_n(z)\,,$$

where g is a zero-free entire function, m is a non-negative integer, the numbers $z_1, z_2, z_3, \cdots$ are the non-zero roots of f, m_n is the order of the zero that f has at z_n, and E_n is a function of the form

$$E_n(z) = \exp\left[\sum_{k=1}^{d(n)} \frac{1}{k}\left(\frac{z}{z_n}\right)^k\right]$$

for a suitably large non-negative integer $d(n)$. (We do not exclude the possibility of being able to choose $d(n) = 0$, in which case we would interpret this formula to mean $E_n = 1$.) Any such representation is called a *Weierstrass product expansion* of f.

Because we approached Theorem 2.4 primarily as an existence theorem, no serious effort was made in its proof to be efficient about the choice of "convergence inducing factors." In concrete situations these factors can often be selected much more economically than would be suggested by that proof, if not dispensed with entirely. As a case in point, just consider a discrete subset $E = \{z_1, z_2, z_3, \cdots\}$ of the complex plane with the property that the series $\sum_{n=1}^{\infty} |z_n|^{-p-1}$ is convergent, where p is a non-negative integer. If $p = 0$, then the function $f_n(z) = 1 - (z/z_n)$ satisfies

$$|f_n(z) - 1| = \frac{|z|}{|z_n|}$$

for every z in $\mathbb{C}$; if $p \geq 1$, then for all z belonging to the disk $K_n = \overline{\Delta}(0, |z_n|/2)$ we have

$$\left|\operatorname{Log}\left(1 - \frac{z}{z_n}\right) + \sum_{k=1}^{p} \frac{1}{k}\left(\frac{z}{z_n}\right)^k\right| = \left|-\sum_{k=p+1}^{\infty} \frac{1}{k}\left(\frac{|z|}{|z_n|}\right)^k\right|$$

$$\leq \frac{1}{2}\sum_{k=p+1}^{\infty}\left(\frac{|z|}{|z_n|}\right)^k = \frac{1}{2}\left(\frac{|z|}{|z_n|}\right)^{p+1}\left(1 - \frac{|z|}{|z_n|}\right)^{-1} \leq \left(\frac{|z|}{|z_n|}\right)^{p+1},$$

so for such z the function

$$f_n(z) = \left(1 - \frac{z}{z_n}\right)\exp\left[\sum_{k=1}^{p} \frac{1}{k}\left(\frac{z}{z_n}\right)^k\right]$$

is readily seen to obey the estimate

$$|f_n(z) - 1| \leq e\left(\frac{|z|}{|z_n|}\right)^{p+1}.$$

(Don't forget that $|e^w - 1| \leq e|w|$ when $|w| \leq 1$.) In either case the convergence of $\sum_{n=1}^{\infty} |z_n|^{-p-1}$ forces the uniform convergence of the product $\prod_{n=1}^{\infty} f_n$ on each compact set in $\mathbb{C}$. As a consequence,

$$f(z) = \prod_{n=1}^{\infty}\left(1 - \frac{z}{z_n}\right) \tag{10.23}$$

when $p = 0$ or

$$f(z) = \prod_{n=1}^{\infty} \left(1 - \frac{z}{z_n}\right) \exp\left[\sum_{k=1}^{p} \frac{1}{k} \left(\frac{z}{z_n}\right)^k\right] \tag{10.24}$$

when $p \geq 1$ defines an entire function with a simple zero at each point of E and with no other zeros. Significant here is that in the case of (10.23) no convergence inducing factors are needed, while in (10.24) the limits on the sums appearing in those factors are independent of n.

An important consequence of the Weierstrass theorem is a clarification of the structure of meromorphic functions.

Theorem 2.5. *Suppose that a function f is meromorphic in an open subset U of the complex plane. Then f can be represented in U as $f = g/h$, where $g: U \to \mathbb{C}$ and $h: U \to \mathbb{C}$ are analytic functions that have no common zeros.*

Proof. If f is already analytic in U, we simply use $g = f$ and $h = 1$. If not, the set E of poles of f in U is a non-empty, discrete subset of U. Appealing to Weierstrass's theorem, we construct an analytic function h in U whose zero-set there is E and whose zero at any point of E has the same order as the pole of f at that point. The function $g = fh$ is then meromorphic in U, where its set of singularities is E. A moment's thought reveals that g possesses a finite, non-zero limit at each such singularity. These singularities are thus seen to be removable. Upon their removal g becomes an analytic function in U, one for which $f = g/h$ in this open set. By construction, the set of zeros of g is disjoint from E, the zero-set of h. ∎

In the representation $f = g/h$ of Theorem 2.5 the zeros of f in U are the zeros of g, while its poles are the zeros of h. Incidentally, this representation is not unique, since we can obviously write $f = (gk)/(hk)$ for any function k that is both analytic and free of zeros in U. On the other hand, it can be shown that this is the only departure from uniqueness in such a representation. Theorem 2.5 carries over to any open set U in $\widehat{\mathbb{C}}$, apart from $U = \widehat{\mathbb{C}}$, provided one transcribes "analytic" to "holomorphic."

Mittag-Leffler's theorem and the Weierstrass theorem can also be effective when used in concert, to which the proof of the next theorem bears witness. The theorem states that, given a discrete subset of a plane open set U, we can construct a function analytic in U whose Taylor series expansions about the points of E have predetermined initial segments.

Theorem 2.6. *Let $E = \{z_1, z_2, z_3, \cdots\}$ be a discrete subset of an open set U in the complex plane, and for $n = 1, 2, 3, \cdots$ let $p_n(z) = \sum_{k=0}^{d_n} a_k^{(n)} z^k$ be a polynomial of degree d_n. There exists an analytic function $f: U \to \mathbb{C}$ that for $n = 1, 2, 3, \cdots$ has a Taylor series expansion about z_n of the form*

$$f(z) = p_n(z - z_n) + O[(z - z_n)^{1+d_n}] \ .$$

Proof. (We refer the reader to Example VIII.2.3. for a review of "big O notation.") Through application of the Weierstrass theorem we can create an analytic function $g: U \to \mathbb{C}$ that has E as its set of zeros, the zero at z_n being one of order $1 + d_n$ for $n = 1, 2, 3, \cdots$. The Taylor expansion of g about z_n thus has the appearance

$$g(z) = c_n(z - z_n)^{1+d_n} + O[(z - z_n)^{2+d_n}] ,$$

in which $c_n \neq 0$. The function h_n defined in U by $h_n(z) = p_n(z - z_n)/g(z)$ has a pole at z_n. Let S_n be the singular part of h_n at z_n. Using Mittag-Leffler's theorem we construct a meromorphic function $h: U \to \widehat{\mathbb{C}}$ whose pole-set is E and whose singular part at z_n is S_n for $n = 1, 2, 3, \cdots$. For z in any suitably small punctured disk centered at z_n we have

$$h(z) = h_n(z) + O(1) = \frac{p_n(z - z_n)}{g(z)} + O(1) .$$

It follows that the function $f = gh$, which is clearly meromorphic in U and which has E as its singular set in U, shows a Laurent expansion of the form

$$f(z) = p_n(z - z_n) + g(z)O(1) = p_n(z - z_n) + O[(z - z_n)^{1+d_n}]$$

in such a punctured disk. We conclude that the singularities of f in U are all removable. Removing them turns f into a function that is analytic in U and has the specified structure near every point of E. ∎

2.4 The Gamma Function

One of the most famous special functions in mathematics comes up quite naturally in the context of infinite products of analytic functions. It is the "gamma function" of Leonhard Euler (1707-1783), a function denoted by Γ and defined by the formula

$$\Gamma(z) = \frac{e^{-\gamma z}}{z} \prod_{n=1}^{\infty} \left(1 + \frac{z}{n}\right)^{-1} e^{z/n} . \tag{10.25}$$

Here γ is "Euler's constant,"

$$\gamma = \lim_{n \to \infty} \left(1 + \frac{1}{2} + \cdots + \frac{1}{n} - \operatorname{Log} n\right) . \tag{10.26}$$

(N.B. Since the sequence $\gamma_n = \sum_{k=1}^{n} k^{-1} - \operatorname{Log} n$ is easily seen to be decreasing and non-negative, $\gamma = \lim_{n\to\infty} \gamma_n$ exists. In rough approximation, $\gamma \approx 0.577$.) The fact that $\sum_{n=1}^{\infty} n^{-2}$ converges implies — recall (10.24) — that the function f given by

$$f(z) = ze^{\gamma z} \prod_{n=1}^{\infty} \left(1 + \frac{z}{n}\right) e^{-z/n} \tag{10.27}$$

is an entire function, one with a simple zero at each of the non-positive integers and with no other zeros. Therefore, the function $\Gamma = 1/f$ is meromorphic in the whole complex plane, it has simple poles at $0, -1, -2, \cdots$, and it is zero-free. The infinite product in (10.27) is normally convergent in the complex plane. The function defined by this product is bounded away from zero on each compact subset of $U = \mathbb{C} \sim \{-1, -2, -3, \cdots\}$. From these two pieces of information it follows without difficulty that the product in (10.25) converges normally in U. With justification provided by Theorem 2.3 we can state that in $\mathbb{C} \sim \{0, -1, -2, \cdots\}$

$$\frac{\Gamma'(z)}{\Gamma(z)} = -\gamma - \frac{1}{z} - \sum_{n=1}^{\infty} \left[\frac{1}{z+n} - \frac{1}{n}\right] . \tag{10.28}$$

As a matter of fact, in the sense fitting to series of meromorphic functions this series converges normally in $\mathbb{C}$, so (10.28) actually holds throughout $\mathbb{C}$, both sides taking the value ∞ at $0, -1, -2, \cdots$.

If $z \neq 0, -1, -2, \cdots$, then we obtain from (10.25) and (10.26)

$$\Gamma(z) = \frac{1}{z} \exp\left[z \lim_{n\to\infty} \left(\operatorname{Log} n - \sum_{k=1}^{n} \frac{1}{k}\right)\right] \lim_{n\to\infty} \prod_{k=1}^{n} \left(1 + \frac{z}{k}\right)^{-1} e^{z/k}$$

$$= \lim_{n\to\infty} \left\{\frac{1}{z} \exp\left[z \left(\operatorname{Log} n - \sum_{k=1}^{n} \frac{1}{k}\right)\right] \prod_{k=1}^{n} \left(\frac{k}{z+k}\right) e^{z/k}\right\} ,$$

which simplifies to "Gauss's formula":

$$\Gamma(z) = \lim_{n\to\infty} \frac{n!\, n^z}{z(z+1)\cdots(z+n)} . \tag{10.29}$$

Therefore, for $z \neq 0, -1, -2, \cdots$ we find that

$$\Gamma(z+1) = \lim_{n\to\infty} \frac{n!\, n^{z+1}}{(z+1)(z+2)\cdots(z+n+1)}$$

$$= \lim_{n\to\infty} \left[\frac{nz}{z+n+1} \cdot \frac{n!\, n^z}{z(z+1)\cdots(z+n)}\right] = z\,\Gamma(z) .$$

In this way we arrive at the "functional equation" of the gamma function:

$$\Gamma(z+1) = z\,\Gamma(z) . \tag{10.30}$$

From (10.29) we also learn that

$$\Gamma(1) = \lim_{n\to\infty} \frac{n!\, n}{(n+1)!} = \lim_{n\to\infty} \frac{n}{n+1} = 1 .$$

We conclude using (10.30) that $\Gamma(n) = n!$ for $n = 1, 2, \cdots$. Referring to (10.12) and Example 2.1, we notice that the function f in (10.27) satisfies

$$f(z)f(-z) = \left[ze^{\gamma z} \prod_{n=1}^{\infty} \left(1 + \frac{z}{n}\right) e^{-z/n}\right] \left[(-z)e^{-\gamma z} \prod_{n=1}^{\infty} \left(1 - \frac{z}{n}\right) e^{z/n}\right]$$

$$= -z^2 \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{n^2}\right) = -\frac{z \sin(\pi z)}{\pi} .$$

In conjunction with the functional equation for Γ, this observation leads to another oft-cited identity satisfied by the gamma function,

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)} . \tag{10.31}$$

Indeed,

$$\Gamma(z)\Gamma(1-z) = -z\,\Gamma(z)\Gamma(-z) = -\frac{z}{f(z)f(-z)} = \frac{\pi}{\sin(\pi z)} .$$

Thus we see, for instance, that $[\Gamma(1/2)]^2 = \pi$. Because Γ is obviously a positive function on the interval $(0, \infty)$, we deduce that $\Gamma(1/2) = \sqrt{\pi}$.

There are an assortment of other ways to represent $\Gamma(z)$, most of them applicable for limited ranges of z. As one example, we point to

$$\Gamma(z) = \lim_{n\to\infty} \int_0^n t^{z-1} \left(1 - \frac{t}{n}\right)^n dt , \tag{10.32}$$

valid when $\operatorname{Re} z > 0$. (N.B. When $0 < \operatorname{Re} z < 1$ the integral involved here is improper at its lower limit, but it is convergent.) We evaluate the integral in (10.32) by making the change of variable $t = ns$, $dt = n\,ds$ and then doing integration by parts repeatedly:

$$\int_0^n t^{z-1} \left(1 - \frac{t}{n}\right)^n dt = n^z \int_0^1 s^{z-1}(1-s)^n\, ds$$

$$= n^z \cdot \frac{n}{z} \int_0^1 s^z (1-s)^{n-1}\, ds = n^z \cdot \frac{n(n-1)}{z(z+1)} \int_0^1 s^{z+1}(1-s)^{n-2}\, ds$$

$$\cdots = \frac{n^z\, n!}{z(z+1)\cdots(z+n-1)} \int_0^1 s^{z+n-1}\, ds = \frac{n!\, n^z}{z(z+1)\cdots(z+n)} .$$

Statement (10.32) now follows from (10.29). A second well-known integral formula for the gamma function — in fact, a formula often used as the definition of $\Gamma(z)$ in a real-variable setting — is

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt , \tag{10.33}$$

valid under the assumption that $\operatorname{Re} z > 0$, which causes this improper integral to be convergent. In deriving (10.33) we make use of the elementary inequalities

$$1 - \frac{t^2}{n} \le e^t \left(1 - \frac{t}{n}\right)^n \le 1 , \tag{10.34}$$

which hold when $0 \le t \le n$. For $z = x + iy$ with $x > 0$, we determine that

$$\left| \int_0^n t^{z-1} e^{-t}\, dt - \int_0^n t^{z-1} \left(1 - \frac{t}{n}\right)^n dt \right| \le \int_0^n \left| t^{z-1} \left[e^{-t} - \left(1 - \frac{t}{n}\right)^n \right] \right| dt$$

$$= \int_0^n t^{x-1} e^{-t} \left[1 - e^t \left(1 - \frac{t}{n}\right)^n \right] dt \le \int_0^n t^{x-1} e^{-t} \left[1 - \left(1 - \frac{t^2}{n}\right) \right] dt$$

$$= \frac{1}{n} \int_0^n t^{x+1} e^{-t}\, dt \to 0$$

as $n \to \infty$, since $\lim_{n\to\infty} \int_0^n t^{x+1} e^{-t}\, dt = \int_0^\infty t^{x+1} e^{-t}\, dt$ is finite. Recalling (10.32), we infer that

$$\int_0^\infty t^{z-1} e^{-t}\, dt = \lim_{n\to\infty} \int_0^n t^{z-1} e^{-t}\, dt$$

$$= \lim_{n\to\infty} \int_0^n t^{z-1} \left(1 - \frac{t}{n}\right)^n dt = \Gamma(z)$$

when $\operatorname{Re} z > 0$ and in this manner obtain (10.33).

The gamma function has a multitude of fascinating properties which a shortage of space prevents us from going into here. Interested readers are urged to look at Emil Artin's beautiful little monograph *The Gamma Function* (Holt, Reinhart, and Winston, New York, 1964) to find out more about this function.

3 Analytic Continuation

3.1 Extending Functions by Means of Taylor Series

The final method we shall discuss for constructing analytic functions has a definite "organic" flavor to it. Much of the special vocabulary associated with the method would look perfectly at home in a biology text! Analytic continuation can be likened to growing a tree from a seedling. Here the "seedling" takes the form of a Taylor series $\sum_{n=0}^\infty a_n(z-z_0)^n$ with a positive radius of convergence ρ_0 or, to be more precise, the analytic function g_0 that this series defines in the disk $\Delta_0 = \Delta(z_0, \rho_0)$ through the rule of

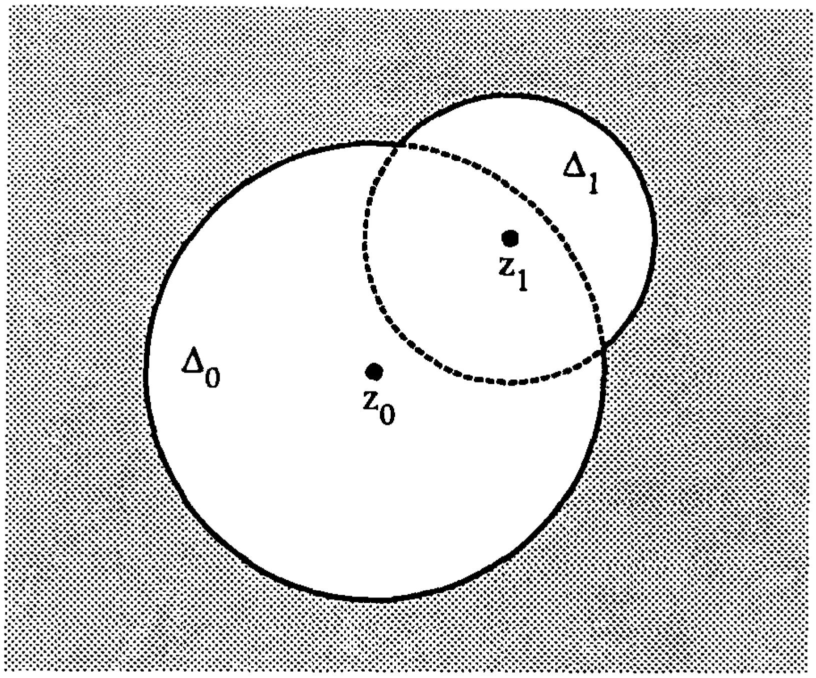


Figure 2.

correspondence $g_0(z) = \sum_{n=0}^{\infty} a_n(z - z_0)^n$. (In fact, one of the technical expressions frequently encountered in the context of analytic continuation describes g_0 not as a "seedling," but as an "analytic germ at z_0.") We can envision g_0 "growing" by a process we now sketch. Given a point z_1 of Δ_0, we can expand g_0 in a Taylor series centered at z_1 — say $g_0(z) = \sum_{n=0}^{\infty} b_n(z - z_1)^n$. We are assured by Theorem VII.3.4 that the radius of convergence ρ_1 of this series is no smaller than $\rho_0 - |z_1 - z_0|$. In particular, $\rho_1 > 0$. The function g_1 defined in $\Delta_1 = \Delta(z_1, \rho_1)$ by the formula $g_1(z) = \sum_{n=0}^{\infty} b_n(z - z_1)^n$ is analytic, and it agrees with g_0 in $\Delta_0 \cap \Delta_1$. If it happens, as well it may, that $\rho_1 = \rho_0 - |z_1 - z_0|$, then g_0 does not experience any "growth" at z_1. In this case g_1 is nothing but the restriction of g_0 to the disk Δ_1. (N.B. There exist situations in which $\rho_1 = \rho_0 - |z_1 - z_0|$ for every choice of z_1 in the disk Δ_0. Such severely "stunted growth" is exemplified by $g_0(z) = \sum_{n=1}^{\infty} z^{n!}$ in $\Delta_0 = \Delta(0,1)$. We refer the reader to Exercise VII.5.64.) If, on the other hand, it is true that ρ_1 is larger than $\rho_0 - |z_1 - z_0|$, then the domain $D_1 = \Delta_0 \cup \Delta_1$ properly contains Δ_0 and the function f_1 given in D_1 by

$$f_1(z) = \begin{cases} g_0(z) & \text{if } z \in \Delta_0 \ , \\ g_1(z) & \text{if } z \in \Delta_1 \ , \end{cases}$$

is analytic (Figure 2). At the risk of working our analogy to death we might say that g_0 has sprouted a "branch" g_1 at z_1. In precise technical language we speak of g_0 being "analytically continued" to g_1 in Δ_1 (and also to f_1 in D_1).

Nothing prevents us from repeating the above construction; i.e., we can choose a point z_2 of D_1, expand f_1 in a Taylor series about z_2, denote by g_2 the analytic function defined by this series in its full disk of convergence

Δ_2, and attempt to define a function f_2 in the domain $D_2 = \Delta_0 \cup \Delta_1 \cup \Delta_2$ by insisting that $f_2 = g_j$ in the disk Δ_j for $j = 0, 1, 2$. In other words,

$$f_2(z) = \begin{cases} f_1(z) & \text{if } z \in D_1 , \\ g_2(z) & \text{if } z \in \Delta_2 . \end{cases}$$

It is not difficult to see that, under these conditions, the set $D_1 \cap \Delta_2$ is still a domain. By construction, $f_1 = g_2$ in some small open disk centered at z_2. On the authority of Corollary VIII.1.6 we can say that $f_1 = g_2$ throughout $D_1 \cap \Delta_2$. As a consequence, f_2 is unambiguously defined in D_2 and is plainly analytic there. The possibility exists, of course, that Δ_2 is contained in D_1, in which event $D_2 = D_1$, $f_2 = f_1$, and this attempt to extend f_1 will have failed. If, however, Δ_2 does not lie completely in D_1, then we shall have succeeded in continuing g_0 analytically to a domain D_2 that properly includes D_1.

On the face of it we ought to be able to iterate the foregoing construction indefinitely, hoping thereby to continue g_0 analytically to larger and larger domains. Unfortunately, the process can develop a hitch. To see what that might be, let us suppose that we have successfully managed to carry out the extension procedure through $n(\geq 2)$ stages. Thus, we have produced functions $g_0, g_1, \ldots, g_n$, each the sum of a Taylor series in its disk of convergence. We label those disks $\Delta_0, \Delta_1, \ldots, \Delta_n$. Our assumption is that for $j = 1, 2, \ldots, n$ the center z_j of Δ_j is an element of the domain $D_{j-1} = \Delta_0 \cup \Delta_1 \cup \cdots \cup \Delta_{j-1}$, but that Δ_j is not contained in D_{j-1}. Moreover, the functions $g_0, g_1, \ldots, g_n$ are assumed to be "compatible," in the sense that $g_j = g_k$ throughout $\Delta_j \cap \Delta_k$ when Δ_j and Δ_k have non-empty intersection. Subject to these hypotheses, we obtain a well-defined analytic function f_n in $D_n = \Delta_0 \cup \Delta_1 \cup \cdots \cup \Delta_n$ by setting $f_n(z) = g_j(z)$ for z belonging to Δ_j, $0 \leq j \leq n$. We seek to push this construction one step further by picking a point z_{n+1} of D_n, expanding f_n in a Taylor series about z_{n+1}, and using the series obtained in this way to define an analytic function g_{n+1} in Δ_{n+1}, the disk of convergence of the series. So far, so good. We would like then to define a function f_{n+1} in the domain $D_{n+1} = \Delta_0 \cup \Delta_1 \cup \cdots \cup \Delta_{n+1}$ by requiring that $f_{n+1} = g_j$ in Δ_j for $0 \leq j \leq n+1$. It is here that things can go awry, for a real possibility exists that there is an index $j(0 \leq j \leq n)$ such that the *intersection* $\Delta_{n+1} \cap \Delta_j$ is not empty, but such that g_{n+1} does not coincide with g_j in this intersection. If that happens, then f_{n+1} is not a well-defined function in $\Delta_j \cap \Delta_{n+1}$ — and the continuation process grinds temporarily to a halt. We could, to be sure, try our luck with a different choice for z_{n+1}, but there is no guarantee that the same scenario wouldn't be repeated. We emphasize that this obstruction to the further analytic continuation of g_0 is likely to make its presence felt at any stage of the continuation process beyond the first two.(N.B. There is another circumstance in which the attempt to continue g_0 beyond D_n through the method described would end in failure — namely, it is conceivable that for every choice of z_{n+1} in D_n the disk Δ_{n+1} is contained in D_n! As previously noted, this is the fate that the function g_0 defined in $\Delta_0 = \Delta(0,1)$ by $g_0(z) = \sum_{n=1}^{\infty} z^{n!}$ suffers at the initial step: g_0 admits no analytic continuation beyond Δ_0. The latter kind of breakdown in the extension procedure is not nearly as traumatic as the former. It simply indicates that we have continued g_0 to what might be styled a "maximal domain of analyticity." We say no more about this case, even though it is not without interest.)

There are several options open to us for dealing with the above impediment to analytic continuation. One of these is to isolate conditions which prevent its occurrence and under which we are at least occasionally in a position to state with certainty that our original function g_0 is analytically continuable to a given domain D which contains Δ_0. (See Theorem 3.4.) This is to some extent a head-in-the-sand approach, for it does not confront the general issue of what the interruption of the continuation process is trying to tell us about the nature of analyticity. Nevertheless, it is the approach we intend to take. Our avowed purpose in the present chapter is, after all, to present various methods of constructing analytic functions — in this instance, returning to our earlier analogy, the "cultivation" of functions from g_0. A more sophisticated treatment of the phenomenon would incorporate it into a full-scale theory of "multi-valued functions" and would lead ultimately to the study of functions that are analytic not in plane domains, but on so-called "Riemann surfaces." Space limitations make it impossible for us to give more than an inkling of that theory in this book. We refer the reader to more advanced texts — e.g., *An Introduction to Riemann Surfaces* by George Springer (Addison-Wesley, Reading, Mass., 1957) — for a thorough discussion of such ideas.

In this section we implemented the process of analytic continuation by making use of Taylor series. When formalizing the concept it actually pays to adopt a slightly more general (and more flexible) point of view. That is what we do in the next section.

3.2 Analytic Continuation

By an *analytic function element* — we frequently abbreviate this expression to "function element," omitting the adjective "analytic" — is meant a pair (f, D), where D is a domain in the complex plane and f is an analytic function whose domain-set is D. The prototypical analytic function elements are those of the form (g, Δ), in which $\Delta = \Delta(z_0, \rho)$ is the disk of convergence of a Taylor series $\sum_{n=0}^{\infty} a_n(z-z_0)^n$ and g is the function given in Δ by $g(z) = \sum_{n=0}^{\infty} a_n(z-z_0)^n$. We reserve a special name for function elements of this type, calling them *analytic germs at* z_0. (We normally refer to such a germ (g, Δ) simply as the "germ g," it being implicit that the domain-set involved is the disk of convergence of the Taylor series which defines g.) An arbitrary function element (f, D) obviously determines an analytic germ at each point z_0 of D — to wit, the germ generated by expanding f in a Taylor series about z_0.

We say that one analytic function element (f_1, D_1) is a *direct analytic continuation* of another such element (f_0, D_0) provided $D_1 \cap D_0$ is nonempty and $f_1(z) = f_0(z)$ for every z in the intersection. (Since this definition is plainly symmetric in f_0 and f_1, it is correct to speak of (f_0, D_0) and (f_1, D_1) as direct analytic continuations of each other.) For example, if (g_0, Δ_0) is an analytic germ at z_0 and if z_1 is an arbitrary point of Δ_0, then the germ g_1 determined by g_0 at z_1 is a direct analytic continuation of g_0. Notice that, when (f_0, D_0) and (f_1, D_1) are direct analytic continuations of one another, we can house both of these function elements in a "larger" function element (f, D) by setting $D = D_1 \cup D_2$ and defining f in D through the requirement that f coincide with f_0 in D_0 and with f_1 in D_1.

Observe that the relation between function elements of being mutual direct analytic continuations respects the process of differentiation: if (f_0, D_0) and (f_1, D_1) are direct analytic continuations of each other, then this property persists for (f_0', D_0) and (f_1', D_1), (f_0'', D_0) and (f_1'', D_1), etc. There are other relationships involving functions or their derivatives that are also preserved under direct analytic continuation. To illustrate this point, suppose that $P(z, w) = a_0(z) + a_1(z)w + \cdots + a_n(z)w^n$ is a polynomial in the complex variable w with coefficients $a_0(z), a_1(z), \ldots, a_n(z)$ that are entire functions of z. Assuming it to be true of an analytic function element (f_0, D_0) that $P[z, f_0(z)] = 0$ for every z in D_0, then any direct analytic continuation (f_1, D_1) of (f_0, D_0) will automatically satisfy $P[z, f_1(z)] = 0$ for all z in D_1 (Exercise 4.36). Similarly, if $w = f_0(z)$ provides a solution in D_0 to an ordinary differential equation

$$c_n(z)\frac{d^n w}{dz^n} + c_{n-1}(z)\,\frac{d^{n-1}w}{dz^{n-1}} + \cdots + c_1(z)\,\frac{dw}{dz} + c_0(z)\,w = 0$$

whose coefficients $c_0(z), c_1(z), \ldots, c_n(z)$ are entire functions, then $w = f_1(z)$ will furnish a solution of the same equation in D_1 (Exercise 4.37).

To say that an analytic function element (f, D) is merely an *analytic continuation* of another such element (f_0, D_0) — again this relationship is a reciprocal one — means that it is possible to pass from (f_0, D_0) to (f, D) via a finite chain of direct analytic continuations; i.e., there exists a finite sequence (f_0, D_0), $(f_1, D_1), \ldots, (f_{n-1}, D_{n-1})$, $(f_n, D_n) = (f, D)$ of analytic function elements such that (f_{j+1}, D_{j+1}) is a direct analytic continuation of (f_j, D_j) for $j = 0, 1, \ldots, n-1$. (N.B. These conditions neither assume nor imply that $f_k = f_j$ in $D_k \cap D_j$ when $k \neq j + 1$. Thus, unlike what happened in the case of a direct analytic continuation, it is not generally possible to consolidate the members of such a chain into a "super" function element $(\widetilde{f}, \widetilde{D})$ by taking $\widetilde{D} = \cup_{j=0}^{n} D_j$ and demanding that $\widetilde{f} = f_j$ in D_j for $j = 0, 1, \ldots, n$. The reason: $\widetilde{f}$ may fail to be a well-defined function.) Consider, by way of example, the analytic function elements (f_0, D_0) and (f, D), in which $D_0 = D = \{z : \operatorname{Re} z > 0\}$ and the two functions are defined in that set by $f_0(z) = \operatorname{Log} z$, $f(z) = \operatorname{Log} z + 2\pi i$. These function elements are manifestly not direct analytic continuations of each other. On the other hand, they definitely are analytic continuations of one another. They can, for instance, be connected by a string of direct analytic continuations (f_0, D_0), (f_1, D_1), (f_2, D_2), (f_3, D_3), (f, D) as follows: $D_1 = \{z : \operatorname{Im} z > 0\}$, $f_1(z) = \operatorname{Log} z$; $D_2 = \{z : \operatorname{Re} z < 0\}$, $f_2(z) = \operatorname{Log}|z| + i\theta(z)$, where θ is the branch of $\arg z$ in D_2 that takes its values in the interval $(\pi/2, 3\pi/2)$; $D_3 = \{z : \operatorname{Im} z < 0\}$, $f_3(z) = \operatorname{Log} z + 2\pi i$. We remark that the functional relationships preserved under direct analytic continuation are also preserved under its weaker cousin, "indirect" analytic continuation.

3.3 Analytic Continuation Along Paths

The important link between analytic continuation and plane topology is most conveniently expressed through the medium of "analytic continuation along a path." We shall operate with this concept exclusively on the level of germs, where the uniqueness of such a continuation accompanies its existence. Suppose then that g is an analytic germ at a point z_0 and that $\gamma : [a, b] \to \mathbb{C}$ is a path with z_0 as its initial point. We say that g can be *analytically continued along* γ if for every t in $[a, b]$ there is an analytic germ g_t at the point $\gamma(t)$ such that $g_a = g$ and such that the following compatibility condition is met: there exists a constant $\delta > 0$ with the property that g_s and g_t are direct analytic continuations of each other whenever the elements s and t of $[a, b]$ satisfy $|s - t| < \delta$. In these circumstances we describe the one-parameter family of germs $\{g_t\}$ as an *analytic continuation of* $g = g_a$ *along* γ *to* $\tilde{g} = g_b$. The germ $\tilde{g}$ really is an analytic continuation of g according to the strict meaning of that term, for we can move from g to $\tilde{g}$ through the chain of direct analytic continuations $g = g_a, g_{t_1}, g_{t_2}, \ldots, g_{t_{n-1}}, g_b = \tilde{g}$ simply by taking an arbitrary partition $a = t_0 < t_1 < \cdots < t_n = b$ of $[a, b]$ with the feature that $t_k - t_{k-1} < \delta$ for $1 \le k \le n$. What the idea of continuing g to $\tilde{g}$ along a path lends to this picture is the sense of $\tilde{g}$ developing from g as the end-product of a continuous evolution, rather than being the last stop in a sequence of jumps. (In actual fact analytic continuation and analytic continuation along a path are equivalent notions.) If g can be continued analytically along γ — this is by no means a foregone conclusion — then the continuation is unique, as the following result proclaims.

Lemma 3.1. *Suppose that* $\{g_t\}$ *and* $\{g_t^*\}$ *are analytic continuations of an analytic germ* g *along a path* $\gamma : [a, b] \to \mathbb{C}$. *Then* $g_t = g_t^*$ *for every* t *in* $[a, b]$.

Proof. Define E to be the set of t in $[a, b]$ such that $g_s = g_s^*$ for every s satisfying $a \leq s \leq t$. Since $g_a = g_a^* = g$, the point a belongs to E. Let t_0 designate the least upper bound of this non-empty, bounded set. Then t_0 belongs to $[a, b]$. We shall be finished with the proof if we can only show that (i) t_0 is a member of E and (ii) $t_0 = b$. To verify (i) we first select a point t_1 from the set E such that $|t_1 - t_0| < \min\{\delta, \delta^*\}$, in which δ and δ^* are compatibility condition constants for the continuations $\{g_t\}$ and $\{g_t^*\}$, respectively. Because t_1 is an element of E, we know that $g_{t_1} = g_{t_1}^*$. By the definitions of δ and δ^* we are told that both g_{t_0} and $g_{t_0}^*$ are direct analytic continuations of g_{t_1}. If ρ and ρ^* denote the radii of convergence of the Taylor series which define g_{t_0} and $g_{t_0}^*$, then we know that $g_{t_0}(z) = g_{t_0}^*(z) = g_{t_1}(z)$ for every z in the non-empty open set $\Delta_0 \cap \Delta_1$, where Δ_0 is the open disk of radius $\min\{\rho, \rho^*\}$ centered at $\gamma(t_0)$ and Δ_1 is the domain-set of g_{t_1}. From the principle of analytic continuation (Corollary VIII.1.6) we infer that $g_{t_0}(z) = g_{t_0}^*(z)$ throughout the disk Δ_0. As a result, these functions determine the same analytic germ at $\gamma(t_0)$; i.e., $g_{t_0} = g_{t_0}^*$. The definition of E implies that $g_s = g_s^*$ when $a \leq s < t_0$. We have just added to $s = t_0$ to the list of points for which these germs coincide and so identified t_0 as a point of E.

We turn next to (ii). Assume, to the contrary, that $t_0 < b$. We can then choose t_2 satisfying $t_0 < t_2 < b$ and $|t_2 - t_0| < \min\{\delta, \delta^*\}$. For any s in the interval $[t_0, t_2]$ both of the germs g_s and g_s^* at $\gamma(s)$ are direct analytic continuations of g_{t_0}, which implies that g_s and g_s^* coincide with g_{t_0} in some non-empty open set. Appealing once again to the principle of analytic continuation, we conclude just as we did above that $g_s = g_s^*$ whenever $t_0 \leq s \leq t_2$. But this clearly puts t_2, a number larger than t_0, in the set E, an impossibility when one considers the definition of t_0. The only way to avoid such a contradiction is to have $t_0 = b$, which is the second thing we needed to prove. ∎

Lemma 3.1 entitles us to speak of "the" analytic continuation of a germ along a path, assuming that a continuation does exist in the first place. Let it be stated emphatically, however, that an analytic germ at a point z_0 need not be analytically continuable along every path originating at z_0. Consider, as an example, the germ g generated at the point $z_0 = 1$ by the function $f(z) = \sqrt{z}$. The function g is nothing more than the restriction of f to the disk $\Delta = \Delta(1, 1)$. Its Taylor series representation in Δ reads

$$g(z) = \sum_{n=0}^{\infty} \binom{1/2}{n} (z-1)^n \ ,$$

in which we make use of the symbol $\binom{\lambda}{n}$ to indicate a generalized binomial coefficient: if λ is any complex number, then

$$\binom{\lambda}{n} = \frac{\lambda(\lambda-1)\cdots(\lambda-n+1)}{n!}$$

for $n = 1, 2, 3, \cdots$, while $\binom{\lambda}{0} = 1$. Let γ be the path defined by $\gamma(t) = 1-t$ for $0 \leq t \leq 1$. We claim that g does not admit an analytic continuation along γ. Were $\{g_t\}$ to be such a continuation, then the relation $[g(z)]^2 = z$, which is satisfied by $g = g_0$ in Δ, would be preserved by g_t for all t in the interval $[0,1]$, meaning that $[g_t(z)]^2 = z$ would hold for every z in Δ_t, the domain-set of g_t. This would, of course, make the germ g_1 a branch of the square root function in a disk containing the origin, an entity we know to be non-existent. (Recall Chapter III.4.2.) The consequence: no such continuation is possible. On the other hand, the given germ g can be continued analytically along the path β that is defined on $[0, 2\pi]$ by $\beta(t) = e^{it}$. (More generally, g can be continued along any path α starting at 1 whose trajectory stays away from the origin.) To present this continuation explicitly, write $\Delta_t = \Delta(e^{it}, 1)$ for $0 \leq t \leq 2\pi$ and define $g_t \colon \Delta_t \to \mathbb{C}$ as follows: when $0 \leq t \leq \pi/2$, $g_t(z) = \sqrt{z}$ for all z in Δ_t; when $\pi/2 < t < 3\pi/2$,

$$g_t(z) = \begin{cases} \sqrt{z} & \text{if } z \in \Delta_t \text{ and } \operatorname{Im} z \geq 0 , \\ -\sqrt{z} & \text{if } z \in \Delta_t \text{ and } \operatorname{Im} z < 0 ; \end{cases}$$

when $3\pi/2 \leq t \leq 2\pi$, $g_t(z) = -\sqrt{z}$ for all z in Δ_t. Each of the functions g_t is an analytic germ. To see this, notice that g_t is continuous in Δ_t and satisfies $[g_t(z)]^2 = z$ throughout Δ_t. It follows from discussions in Chapter III that g_t provides a branch of the square root function in Δ_t. In particular, g_t is an analytic function. A short computation provides the Taylor expansion of g_t about e^{it}:

$$g_t(z) = e^{it/2} \sum_{n=0}^{\infty} e^{-int} \binom{1/2}{n} (z - e^{it})^n .$$

The radius of convergence of this series is easily found to be 1, which implies that the series represents g_t throughout its disk of convergence, that disk being none other than Δ_t. By definition, this makes g_t an analytic germ at e^{it}. It is not difficult to check here that g_s and g_t are direct analytic continuations of one another as soon as $|s-t| < \pi$, so $\{g_t\}$ does supply an analytic continuation of g along β. As a matter of fact, it continues g to $\tilde{g} = g_{2\pi} = -g$. In a similar way we can analytically continue $-g$ along β — the family $\{-g_t\}$ furnishes the continuation — and arrive back at g! (In general, when g gets continued along a closed path α in $\mathbb{C} \sim \{0\}$ beginning and ending at 1, the continuation terminates in the germ $(-1)^n g$, where $n = n(\alpha, 0)$ is the winding number of α about the origin.)

We make the observation that if g is an analytic germ at a point z_0, if g has domain-set $\Delta(z_0, \rho)$, and if $\gamma: [a, b] \to \mathbb{C}$ is any path with initial point z_0 whose trajectory lies in the disk $\Delta(z_0, \rho/2)$, then the analytic continuation $\{g_t\}$ of g along γ definitely exists: g_t is just the germ determined by g at $\gamma(t)$. That the necessary compatibility condition holds is essentially trivial here. Each germ g_t is a direct analytic continuation of g and the domain-set of g_t has radius at least $\rho - |\gamma(t) - z_0|$, which is larger than $\rho/2$. This means that g_t and g_s agree with g in some open set containing z_0 — hence, are direct analytic continuations of each other — for all t and s in $[a, b]$.

We record a few general remarks about the analytic continuation $\{g_t\}$ of an analytic germ g along a path $\gamma: [a, b] \to \mathbb{C}$. Fixing a compatibility condition constant δ for this continuation, we let $\Delta_t = \Delta[\gamma(t), \rho(t)]$ be the domain-set of the germ g_t. Our first comment is that the estimate

$$(10.35) \qquad \rho(s) \leq \rho(t) + |\gamma(s) - \gamma(t)|$$

is in force whenever $|s - t| < \delta$. If, namely, it were the case that $\rho(s)$ exceeded $\rho(t) + |\gamma(s) - \gamma(t)|$, then the disk Δ_t would be contained in Δ_s. As g_s is a direct analytic continuation of g_t, the former germ would have to agree with g_t throughout Δ_t. This would imply that g_t and g_s give rise to the same Taylor series at $\gamma(t)$. Since g_s is analytic in Δ_s, the Taylor series it generates at $\gamma(t)$ would have a radius of convergence no smaller than $\rho(s) - |\gamma(s) - \gamma(t)|$, a number that is by assumption larger than $\rho(t)$. This state of affairs would be inconsistent with the definition of $\rho(t)$. Accordingly, (10.35) must hold once $|s - t| < \delta$. By symmetry,

$$(10.36) \qquad \rho(t) \leq \rho(s) + |\gamma(s) - \gamma(t)|$$

is also true for such s and t. From (10.35) and (10.36) we extract the information that either $\rho(t) = \infty$ for every t in $[a, b]$ or the inequality

$$(10.37) \qquad |\rho(s) - \rho(t)| \leq |\gamma(s) - \gamma(t)|$$

is valid whenever s and t satisfy $|s - t| < \delta$. An immediate consequence of (10.37) and the continuity of γ is that $\rho(t)$ is a continuous function of t on $[a, b]$, unless $\rho(t) = \infty$ for all t in this interval. Because $\rho(t) > 0$ for every t in $[a, b]$, we are thus in a position to conclude that

$$(10.38) \qquad \rho = \min\{\rho(t) : a \leq t \leq b\} > 0 .$$

Incidentally, we enjoy a certain measure of flexibility in our choice of the constant δ, always having the freedom to replace our initial selection by any smaller positive number. Once aware of (10.38) we can reap the benefits of hindsight and, by taking advantage of the uniform continuity of γ on

$[a, b]$, shrink our original δ to a size that compels $|\gamma(s) - \gamma(t)| < \rho$ to hold whenever $|s - t| < \delta$. Under this added constraint on δ the statement

$$(10.39) \qquad |\gamma(s) - \gamma(t)| < \min\{\rho(s), \rho(t)\}$$

is true for all s and t in $[a, b]$ with $|s - t| < \delta$. The consequence of (10.39) for such s and t is this: not only are the germs g_t and g_s direct analytic continuations of one another, but g_t is actually obtained by expanding g_s in a Taylor series about $\gamma(s)$, and vice versa. We have no plans to exploit this fact, but we make note of it in passing for the simple reason that some treatments of analytic continuation include (10.39) as part of the compatibility condition in the definition of a continuation of g along γ. While we are on the subject, we should also point out that some authors prefer to state the aforementioned compatibility condition in a local version, phrasing it so: corresponding to each t in $[a, b]$ there exists a constant $\delta_t > 0$ with the property that g_s is a direct analytic continuation of g_t whenever s in $[a, b]$ satisfies $|s - t| < \delta_t$. We leave as an exercise the verification that this formulation is equivalent to the one we have given.

Let $\{g_t\}$ be the analytic continuation of a germ g to a germ $\widetilde{g}$ along a path $\gamma: [a, b] \to \mathbb{C}$. Then $\widetilde{g}$ can be analytically continued back to g along the path $-\gamma$, the germ corresponding to t in the reverse continuation being g_{b+a-t}. If $\widetilde{g}$ admits a further analytic continuation — denote it by $\{\widetilde{g}_s\}$ — along a path $\widetilde{\gamma}: [c, d] \to \mathbb{C}$, then the germ g can be continued along the path $\gamma^* = \gamma + \widetilde{\gamma}$. Recalling that γ^* has for its domain-set the interval $[a, b + d - c]$, we patch $\{g_t\}$ and $\{\widetilde{g}_s\}$ together to form the continuation $\{g_r^*\}$ of g along the composite path as follows: $g_r^* = g_r$ if $a \le r \le b$ and $g_r^* = \widetilde{g}_{r+c-b}$ if $b \le r \le b+d-c$. That $\{g_r^*\}$ obeys the necessary compatibility condition is not immediately obvious (by contrast, the validity of the local compatibility condition referred to above is instantly clear), but this fact is not hard to verify, especially if compatibility condition constants for $\{g_t\}$ and $\{\widetilde{g}_s\}$ are selected in such a way that (10.39) and its counterpart for $\{\widetilde{g}_s\}$ are satisfied. Finally, if a path $\gamma_1: [a_1, b_1] \to \mathbb{C}$ is derived from γ by a change of parameter $h: [a_1, b_1] \to [a, b]$, then the assignment of the germ $g_{h(u)}$ to u defines an analytic continuation of g along γ_1. In this instance the compatibility condition can be checked by appealing to the uniform continuity of h on the interval $[a_1, b_1]$.

The last observation in this section can be summarized as follows: if a germ g can be analytically continued along a path γ and if $\widetilde{\gamma}$ is a path that lies sufficiently "close" to γ, then g is continuable along $\widetilde{\gamma}$ as well.

Lemma 3.2. *Suppose that $\{g_t\}$ is the analytic continuation of a germ g along a path $\gamma: [a, b] \to \mathbb{C}$ and that $0 < \rho < \min\{\rho(t) : a \le t \le b\}$, where $\Delta_t = \Delta[\gamma(t), \rho(t)]$ is the domain-set of g_t. If $\widetilde{\gamma}: [a, b] \to \mathbb{C}$ is any path with the properties that $\gamma(a) = \widetilde{\gamma}(a)$ and that $|\gamma(t) - \widetilde{\gamma}(t)| < \rho/4$ for every t in $[a, b]$, then g has an analytic continuation $\{\widetilde{g}_t\}$ along $\widetilde{\gamma}$. Furthermore, when $\gamma(b) = \widetilde{\gamma}(b)$ the two continuations lead to the same terminal germ; i.e., $g_b = \widetilde{g}_b$.*

Proof. The hypotheses imply that $\widetilde{\gamma}(t)$ is an element of Δ_t for each t in $[a,b]$. Because g_t is analytic in Δ_t, it determines a germ at $\widetilde{\gamma}(t)$. We label that germ $\widetilde{g}_t$ and write its domain-set $\Delta[\widetilde{\gamma}(t),\widetilde{\rho}(t)]$ in abbreviated form as $\widetilde{\Delta}_t$. We can be certain that

$$\widetilde{\rho}(t) \geq \rho(t) - |\widetilde{\gamma}(t) - \gamma(t)| \geq \rho - \frac{\rho}{4} = \frac{3\rho}{4} \,.$$

Moreover, the functions g_t and $\widetilde{g}_t$ coincide in the intersection $\Delta_t \cap \widetilde{\Delta}_t$. The obvious candidate for the analytic continuation of g along $\widetilde{\gamma}$ is $\{\widetilde{g}_t\}$. Plainly $\widetilde{g}_a = g_a = g$, so only the compatibility condition for $\{\widetilde{g}_t\}$ needs to be verified. We choose $\delta > 0$ small enough to guarantee the validity of the following statements for all s and t in $[a,b]$ satisfying $|s-t| < \delta$: (i) $|\gamma(s) - \gamma(t)| < \rho/2$; (ii) $|\widetilde{\gamma}(s) - \widetilde{\gamma}(t)| < \rho/2$; (iii) g_s and g_t are direct analytic continuations of each other. For (i) and (ii) the selection of δ is made possible by the uniform continuity of γ and $\widetilde{\gamma}$ on $[a,b]$; to achieve (iii) we invoke the compatibility condition for $\{g_t\}$. We now claim that $\widetilde{g}_s$ and $\widetilde{g}_t$ are related by direct analytic continuation whenever $|s-t| < \delta$. Fix s and t subject to this constraint. The inequalities

$$|\gamma(t) - \widetilde{\gamma}(t)| < \frac{\rho}{4} \leq \widetilde{\rho}(t) \,,$$

$$|\gamma(t) - \gamma(s)| < \frac{\rho}{2} \leq \rho(s) \,,$$

$$|\gamma(t) - \widetilde{\gamma}(s)| \leq |\gamma(t) - \widetilde{\gamma}(t)| + |\widetilde{\gamma}(t) - \widetilde{\gamma}(s)| < \frac{\rho}{4} + \frac{\rho}{2} = \frac{3\rho}{4} \leq \widetilde{\rho}(s) \,,$$

demonstrate that $\gamma(t)$ is a point of the set $U = \Delta_t \cap \widetilde{\Delta}_t \cap \Delta_s \cap \widetilde{\Delta}_s$, making U a non-empty open set. Moreover, by design $\widetilde{g}_t = g_t = g_s = \widetilde{g}_s$ in U. The principle of analytic continuation then forces $\widetilde{g}_t$ and $\widetilde{g}_s$ to agree throughout $\widetilde{\Delta}_t \cap \widetilde{\Delta}_s$; i.e., $\widetilde{g}_t$ and $\widetilde{g}_s$ are direct analytic continuations of each other. We have thus succeeded in certifying $\{\widetilde{g}_t\}$ as the analytic continuation of g along $\widetilde{\gamma}$. Finally, if $\gamma(b) = \widetilde{\gamma}(b)$, then the above construction and the uniqueness of analytic continuation along a path combine to yield $\widetilde{g}_b = g_b$ for any path $\widetilde{\gamma}$ with all the other stated properties. ∎

3.4 Analytic Continuation and Homotopy

In Chapter V.7.1 we introduced the notion of homotopy for paths in the complex plane. At the time our primary concern was the relationship between winding numbers and topology, so the focus was on free homotopy of closed paths. We did, however, pay lip-service to the idea of "homotopy with fixed endpoints," which is the brand of homotopy that comes into play in conjunction with analytic continuation. To recall the definition of the concept assume that $\alpha\colon [a,b] \to \mathbb{C}$ and $\beta\colon [a,b] \to \mathbb{C}$ are paths in a plane domain D and that these two paths have both the same initial point and the same terminal point. For α and β to be ***homotopic with fixed endpoints in D*** means that there exists a continuous function $H: R = \{(t,s): a \le t \le b, 0 \le s \le 1\} \to D$ which is blessed with the following two properties:

$$\begin{cases} \text{(i) } H(t,0) = \alpha(t) \,,\ H(t,1) = \beta(t) & \text{for } a \le t \le b \,; \\ \text{(ii) } H(a,s) = \alpha(a) \,,\ H(b,s) = \alpha(b) & \text{for } 0 \le s \le 1 \,. \end{cases}$$

(Any such H is called a ***fixed-endpoint homotopy from α to β in D.***) With H we associate a one-parameter family of paths $\gamma_s: [a,b] \to D$ $(0 \le s \le 1)$, each of which has the same endpoints as α and β: γ_s is defined by $\gamma_s(t) = H(t,s)$. One thinks of the homotopy H as a mechanism whereby the path $\alpha = \gamma_0$ can be continuously deformed to the path $\beta = \gamma_1$ in the domain D, the intermediate path γ_s representing the state of the deformation process at the instant s.

A fundamental connection between analytic continuation and plane topology is signalled by a result known as the "Monodromy Theorem."

Theorem 3.3. (Monodromy Theorem) *Let z_0 and z_1 be points of a plane domain D, and let g be an analytic germ at z_0 that can be analytically continued along every path in D with initial point z_0 and terminal point z_1. If two paths α and β of this description are homotopic with fixed endpoints in D, then the continuations of g along α and β produce the same terminal germ at z_1.*

Proof. Suppose that α and β are parametrized on the interval $[a,b]$. We choose a fixed-endpoint homotopy H from α to β in D, and let γ_s be the intermediate path associated with H for $0 \le s \le 1$. We denote by $\widetilde{g}$ the germ at z_1 obtained as the end-product of the analytic continuation of g along α. Consider the set E consisting of all s in $[0,1]$ for which continuation of g along the path γ_s leads to $\widetilde{g}$ as a final germ. Since $\gamma_0 = \alpha$, E is non-empty. This set has a least upper bound s_0, which clearly satisfies $0 \le s_0 \le 1$. To finish the proof we shall argue that (i) s_0 is a member of E and (ii) $s_0 = 1$. For this, designate by $\widetilde{g}_0$ the germ at z_1 that results from the continuation of g along γ_{s_0}. The message of Lemma 3.2 is that we can produce a number $\epsilon > 0$ with the following property: if $\gamma: [a,b] \to \mathbb{C}$ is any path from z_0 to z_1 fulfilling the requirement that $|\gamma(t) - \gamma_{s_0}(t)| < \epsilon$ for every t in $[a,b]$, then not only is the analytic continuation of g along γ possible, but this continuation necessarily ends with the germ $\widetilde{g}_0$. Next, because the continuous function H is uniformly continuous on the compact set R (Theorem II.4.8), we can fix $\eta > 0$ so as to make certain that

$$|\gamma_s(t) - \gamma_{s_0}(t)| = |H(t,s) - H(t,s_0)| < \epsilon$$

holds for every t in $[a,b]$ whenever $|s - s_0| < \eta$. By the definition of a least upper bound we can assert the existence of a number s_1 in E for which

$s_0 - \eta < s_1 \le s_0$. The fact that s_1 is in E means that the continuation of g along γ_{s_1} leads to $\widetilde{g}$; on the other hand, the preceding remarks pin down the outcome of that continuation as $\widetilde{g}_0$. We conclude that $\widetilde{g}_0 = \widetilde{g}$, which places s_0 in E. If s_0 were less than 1, then we could select a number s_2 in the interval $(s_0, 1)$ satisfying $s_2 < s_0 + \eta$. In continuing g along γ_{s_2} we would again arrive at $\widetilde{g}_0 = \widetilde{g}$; i.e., s_2 would be an element of E larger than s_0, an unacceptable situation. Consequently, $s_0 = 1$ and the continuation of g along $\beta = \gamma_1$ does terminate in the germ $\widetilde{g}$, as claimed. ■

The monodromy theorem is most readily applied in a simply connected domain D, where any pair of paths $\alpha, \beta: [a, b] \to D$ that share a common initial point and a common terminal point are homotopic with fixed endpoints: if $D = \mathbb{C}$, then $H(t, s) = (1 - s)\alpha(t) + s\beta(t)$ gives a fixed-endpoint homotopy from α to β in D; if $D \neq \mathbb{C}$, such a homotopy is delivered by

$$H(t,s) = f^{-1}\{(1-s)\, f[\alpha(t)] + s\, f[\beta(t)]\} \ ,$$

in which f is a conformal mapping of D onto $\Delta(0, 1)$. A simply connected domain D is the setting for the next theorem, which formulates the third and final method of constructing analytic functions that we discuss in this book.

Theorem 3.4. *Let z_0 be a point of a simply connected domain D in the complex plane, and let g be an analytic germ at z_0 that can be analytically continued along every path in D which has z_0 for its initial point. Then there exists a unique analytic function element (f, D) whose germ at z_0 is g.*

Proof. Define a function $f: D \to \mathbb{C}$ as follows: given z in D, choose an arbitrary path γ in D with initial point z_0 and terminal point z, continue g analytically along γ to a terminal germ $\widetilde{g}$ at the point z, and set $f(z) = \widetilde{g}(z)$. The number $f(z)$ so obtained does not depend on the choice of the path γ. Indeed, after a preliminary change of parameter — such a change does not affect the terminal germ of the continuation — we may always assume that γ has for its domain-set the interval $[0, 1]$. Because D is simply connected, any two paths $\gamma: [0, 1] \to D$ and $\widetilde{\gamma}: [0, 1] \to D$ with initial point z_0 and terminal point z are homotopic with fixed endpoints in D. By the monodromy theorem the continuations of g along γ and $\widetilde{\gamma}$ produce the same terminal germ. As a result, the independence of $f(z)$ from the selection of γ is confirmed. In other words, f is a well-defined function in D.

To show that f is analytic in D, it suffices to verify that for each point z_1 of D this function is analytic in some open disk centered at z_1. Fix such a point z_1 and fix along with it a path $\gamma: [0, 1] \to D$ that starts at z_0 and ends at z_1. Let $\widetilde{g}$ be the germ at z_1 arising from the continuation $\{g_t\}$ of g along γ, let $\Delta(z_1, \rho)$ be the domain-set of $\widetilde{g}$, and let $\Delta = \Delta(z_1, r)$ be an open disk with $0 < r < \rho/2$ such that Δ is contained in D. If z belongs to Δ, the analytic continuation $\{g_s^*\}$ of g along the path $\gamma^* = \gamma + [z_1, z]$ is not hard to describe: $g_s^* = g_s$ when $0 \le s \le 1$ and g_s^* is the germ determined at $\gamma^*(s)$ by $\widetilde{g}$ when $1 \le s \le 2$. (Since for $1 \le s \le 2$ the germ g_s^* is obviously a direct analytic continuation of $\widetilde{g} = g_1^*$ and since the domain-set of g_s^* includes z_1 — its radius is at least $\rho/2$ — the compatibility condition for $\{g_s^*\}$ is easily checked.) By definition $f(z) = g_2^*(z) = \widetilde{g}(z)$. Such being the case for every z in Δ, the functions f and $\widetilde{g}$ agree in Δ, which fact makes evident the analyticity of f in this disk. We have therefore managed to construct an analytic function element (f, D). The argument just presented applies to the special case where we take $z_1 = z_0$ and $\gamma(t) = z_0$ for $0 \le t \le 1$ (then $\widetilde{g} = g$) and demonstrates that f and g coincide in some open disk centered at z_0, so g is the germ determined by f at z_0. The principle of analytic continuation vouches for the uniqueness of any function element (f, D) exhibiting this property. ∎

We stress that in using Theorem 3.4 to construct an analytic function it is necessary to prove that the original germ g is analytically continuable along all relevant paths. This is not going to be the case automatically.

3.5 Algebraic Function Elements

In this section we put Theorem 3.4 to work in showing how analytic functions arise through the solution of algebraic equations. The discussion starts with a polynomial function P of two complex variables z and w. We write P in the form

$$P(z, w) = p_0(z) + p_1(z)w + \cdots + p_n(z)w^n \ , \qquad (10.40)$$

where $p_0(z), p_1(z), \ldots, p_n(z)$ are polynomial functions of z. To avoid certain degenerate or trivial situations we shall always tacitly assume that $n \ge 1$ and that the leading coefficient p_n is not the zero polynomial. A simple example of what we have in mind here is the function $P(z, w) = w^n - z$. Another example that turns out to be significant in various contexts is the quadratic (in w) polynomial $P(z, w) = w^2 - 4(z-a)(z-b)(z-c)$, where a, b, and c are distinct complex constants. By holding the value of z in (10.40) fixed — say $z = z_0$ — we obtain a polynomial function φ of the variable w alone, $\varphi(w) = P(z_0, w)$. Without giving the matter much thought we might hope that φ would have degree n and possess n different roots. Inevitably, however, there is a set of points z_0 for which φ fails to have n distinct zeros. We name any point z_0 of this type an *exceptional point for P*. In case P is "irreducible" — this means that P cannot be expressed as the product of two non-constant polynomial functions of z and w — the set of exceptional points is necessarily a finite set, as we shall later see; in the "reducible" case the exceptional set may become infinite. For example, every complex number z_0 is exceptional for the polynomial $P(z, w) = z^2 + 2zw + w^2 = (z+w)^2$; if $P(z, w) = w^n - z$, then the exceptional set consists of the origin; for the polynomial $P(z, w) = w^2 - 4(z-a)(z-b)(z-c)$ the exceptional set is just $\{a,b,c\}$. In general, a point z_0 can qualify as exceptional for one of two reasons: (i) z_0 is a zero of the coefficient p_n in (10.40), in which event the degree of φ drops below n; (ii) $p_n(z_0) \neq 0$ — so φ is of degree n — but φ has a multiple root. In the latter instance there exists a point w_0 (any multiple zero w_0 of φ will have the property) such that both $P(z_0,w_0) = \varphi(w_0) = 0$ and $P_w(z_0,w_0) = \varphi'(w_0) = 0$; i.e., P and P_w, its partial derivative with respect to the variable w, have a common zero in $\mathbb{C}^2$ of the form (z_0,w_0).

The first result of this section is a version of the "Implicit Function Theorem." It informs us that at any non-exceptional point z_0 for P the equation $P(z,w) = 0$ implicitly defines a collection of n analytic germs $g_1, g_2, \ldots, g_n$ which obey the relationship $P[z, g_j(z)] = 0$ throughout their respective domain-sets.

Theorem 3.5. *Assume that z_0 is not an exceptional point for the polynomial $P(z,w) = p_0(z) + p_1(z)w + \cdots + p_n(z)w^n$. Let $w_1, w_2, \ldots, w_n$ be the solutions of the equation $P(z_0,w) = 0$. Corresponding to each w_j there exists a unique analytic germ g_j at z_0 such that $g_j(z_0) = w_j$ and such that $P[z, g_j(z)] = 0$ for every z in the domain-set of g_j. In fact, if D is any plane domain containing z_0 and $f: D \to \mathbb{C}$ is any continuous function satisfying $P[z, f(z)] = 0$ for every z in D, then there is an open disk centered at z_0 in which f coincides with the unique germ g_j among $g_1, g_2, \ldots, g_n$ that takes the value $f(z_0)$ at z_0.*

Proof. Since the n^{th}-degree polynomial $\varphi(w) = P(z_0, w)$ finds in the points $w_1, w_2, \ldots, w_n$ a full complement of distinct roots, each of these is necessarily a simple root. Therefore, $\varphi'(w_j) \neq 0$ for every j. We use the continuity of φ' to select a number $s > 0$ with the following properties: the open disks $\Delta_1 = \Delta(w_1, s)$, $\Delta_2 = \Delta(w_2, s), \ldots, \Delta_n = \Delta(w_n, s)$ are pairwise disjoint and $\varphi'(w) \neq 0$ holds for every w in the set $K = \cup_{j=1}^n \partial\Delta_j$. Naturally $\varphi(w) \neq 0$ for all w in K, so $|\varphi|$ has a positive minimum value on this compact set. Call that minimum value m. Appealing to the uniform continuity of the polynomial P on the compact set $E = \{(z,w) : z \in \overline{\Delta}(z_0, 1), w \in K\}$ in $\mathbb{C}^2$, we can choose and fix a radius r in $(0,1)$ with the property that

$$|P(z,w) - P(z_0,w)| < m$$

whenever z belongs to $\Delta = \Delta(z_0, r)$ and w to K.

Consider a point z of Δ. Owing to the selection of r the polynomial ψ defined by $\psi(w) = P(z,w)$ satisfies

$$|\varphi(w) - \psi(w)| < m \leq |\varphi(w)|$$

for every w on $\partial\Delta_j$. According to Rouché's theorem the functions φ and ψ must have the same number of zeros in Δ_j. By construction φ has a simple zero at w_j as its one and only zero in this disk. As a result, ψ has exactly one root in Δ_j. Since this is the case for $j = 1, 2, \ldots, n$ and since

Δ_j and Δ_k are disjoint when $j \neq k$, we conclude that for each fixed z in Δ the equation $P(z,w) = 0$ has exactly n solutions, one in each of the disks Δ_j. This fact enables us to define functions $f_j : \Delta \to \Delta_j$, $1 \leq j \leq n$, by the following prescription: if z is in Δ, then $f_j(z)$ is the unique point of Δ_j for which $P[z, f_j(z)] = 0$. Notice especially that no point of Δ can be exceptional for P. It follows that, when $P(z,w) = 0$ and z lies in Δ, $P_w(z,w) \neq 0$.

For fixed z in Δ the formula $Q(w) = wP_w(z,w)/P(z,w)$ defines a rational function of w whose only singularities in $\mathbb{C}$ are located at the points $\zeta_1 = f_1(z)$, $\zeta_2 = f_2(z), \ldots, \zeta_n = f_n(z)$, each being no worse than a simple pole. (N.B. If $\zeta_j = 0$ for some j, then the singularity of Q at ζ_j is actually a removable singularity, rather than a pole.) Through an application of the residue theorem we discover that

$$\int_{|w-w_j|=s} Q(w)\,dw = 2\pi i \operatorname{Res}(\zeta_j, Q) = 2\pi i \lim_{w \to \zeta_j} (w - \zeta_j) Q(w)$$

$$= 2\pi i \lim_{w \to \zeta_j} \left[w\, P_w(z,w)\, \frac{w - \zeta_j}{P(z,w)} \right] = 2\pi i\, \zeta_j P_w(z, \zeta_j) \lim_{w \to \zeta_j} \frac{w - \zeta_j}{P(z,w)}$$

$$= 2\pi i\, \zeta_j\, P_w(z,\zeta_j)\, \frac{1}{P_w(z,\zeta_j)} = 2\pi i\, \zeta_j = 2\pi i\, f_j(z)\ .$$

In this way we see that f_j admits an integral representation in Δ; namely,

$$f_j(z) = \frac{1}{2\pi i} \int_{|w-w_j|=s} \frac{w\, P_w(z,w)\, dw}{P(z,w)}\ . \tag{10.41}$$

If z is a point of Δ and if $\langle z_k \rangle$ is any sequence in Δ such that $z_k \to z$, then we infer from the uniform continuity of the function P_w/P on the set $\{(\zeta, w) : |\zeta - z| \leq (r - |z - z_0|)/2, w \in K\}$ that, as $k \to \infty$,

$$\frac{w\, P_w(z_k, w)}{P(z_k, w)} \longrightarrow \frac{w\, P_w(z,w)}{P(z,w)}$$

uniformly on $\partial\Delta_j$. This implies that

$$\lim_{k\to\infty} f_j(z_k) = \lim_{k\to\infty} \frac{1}{2\pi i} \int_{|w-w_j|=s} \frac{w\, P_w(z_k,w)\, dw}{P(z_k,w)}$$

$$= \frac{1}{2\pi i} \int_{|w-w_j|=s} \frac{w\, P_w(z,w)\, dw}{P(z,w)} = f_j(z)\ ,$$

which reveals that f_j is a continuous function.

Suppose next that R is an arbitrary closed rectangle in Δ. Formula (10.41), combined with identity (V.5.20) and the observation that for fixed

w in $\partial\Delta_j$ the rational function $\widetilde{Q}$ of z given by $\widetilde{Q}(z) = P_w(z,w)/P(z,w)$ is analytic in Δ, lets us conclude that

$$\int_{\partial R} f_j(z)\,dz = \frac{1}{2\pi i}\int_{\partial R}\left\{\int_{|w-w_j|=s}\frac{w\,P_w(z,w)\,dw}{Pz,w)}\right\}dz$$

$$= \frac{1}{2\pi i}\int_{|w-w_j|=s} w\left\{\int_{\partial R}\frac{P_w(z,w)\,dz}{P(z,w)}\right\}dw = \frac{1}{2\pi i}\int_{|w-w_j|=s} 0\,dw = 0\ .$$

On the basis of Morera's theorem we can now assert that f_j is analytic in Δ. Let g_j be the analytic germ determined by f_j at z_0. Then $g_j = f_j$ in Δ, though the domain-set of g_j may actually be a disk of radius larger than r. Clearly $g_j(z_0) = f_j(z_0) = w_j$ and $P[z, g_j(z)] = P[z, f_j(z)] = 0$ for every z in Δ. The principle of analytic continuation sees to it that $P[z, g_j(z)] = 0$ throughout the domain-set of g_j.

Finally, assume that D is a plane domain which contains z_0 and that $f\colon D \to \mathbb{C}$ is a continuous function which satisfies $P[z, f(z)] = 0$ for every z in D. Because $P[z_0, f(z_0)] = 0$, it can only be the case that $f(z_0) = w_j$ for some j. The continuity of f then allows us to choose a t in $(0, r)$ for which the disk $\Delta(z_0, t)$ is contained in D and $f[\Delta(z_0, t)]$ is contained in Δ_j. The fact that $P[z, f(z)] = 0$ for all z in $\Delta(z_0, t)$, coupled with the definition of f_j, makes it evident that f coincides with f_j — hence, with the germ g_j — in $\Delta(z_0, t)$. ∎

We earlier made the statement that for an irreducible polynomial P of the type (10.40) there are at most a finite number of exceptional points. This important fact is obvious when $n = 1$. The key to confirming it for larger n lies in the realization that an irreducible polynomial P and its partial derivative P_w are related by identities of the form

$$A(z,w)P(z,w) + B(z,w)P_w(z,w) = C(z)\ , \tag{10.42}$$

in which A and B are polynomial functions of z and w, while C is a non-zero polynomial function of the single variable z. (N.B. Among all non-zero polynomial functions $C(z)$ expressible in the manner (10.42) there is a unique monic polynomial — i.e., its highest coefficient is 1 — of smallest degree. It is referred to in the literature as the *discriminant* of P.) Given the identity (10.42) we observe that for a point (z_0, w_0) in $\mathbb{C}^2$ to obey $P(z_0, w_0) = P_w(z_0, w_0) = 0$ it is necessary that $C(z_0) = 0$. As a consequence, the exceptional points for P must appear among the finitely many zeros of C and the finitely many zeros of the coefficient p_n in (10.40) — hence, must themselves be finite in number.

We shall only sketch the derivation of (10.42), for it hinges on a non-trivial piece of algebraic information concerning the irreducible polynomial P: *if F is a non-constant polynomial function of z and w whose degree in w is less than n and if F is a factor of a product GP, where G is also a polynomial function of z and w, then F must be a factor of G.* A proof of this fact can be found in many elementary abstract algebra texts, e.g., *Topics in Algebra* by I.N. Herstein (Xerox College Publishing, Lexington, Mass., 1964). Assuming that $n \geq 2$, we arrive at (10.42) through an explicit construction. We start by dividing P_w into P (we are thinking here of old-fashioned long division with respect to the variable w) and express the result in the manner

$$P(z,w) = Q(z,w)P_w(z,w) + R(z,w) \ . \tag{10.43}$$

The functions Q and R, the quotient and remainder that come out of the division process, have the general structure

$$r_0(z) + r_1(z)w + \cdots + r_\ell(z)w^\ell \ ,$$

in which the coefficients $r_0, r_1, \ldots, r_\ell$ are rational functions of z. We can multiply both sides of (10.43) by an appropriate non-zero polynomial function of z — label it A_0 — to clear the denominators from all the rational coefficients appearing on its right-hand side. We thereby transform (10.43) into an identity

$$A_0(z)P(z,w) = B_1(z,w)P_w(z,w) + C_1(z,w) \ , \tag{10.44}$$

where B_1 and C_1 are polynomial functions of z and w. Note that the degree of C_1 with respect to w is smaller than $n-1$, the degree of P_w in w. If the degree of C_1 in w is positive, we repeat the above procedure, this time dividing C_1 into P_w. By so doing we obtain an identity analogous to (10.44) relating P_w and C_1,

$$A_1(z)P_w(z,w) = B_2(z,w)C_1(z,w) + C_2(z,w) \ ,$$

in which the degree of C_2 in w is at most $n-3$. If that degree is again positive, we continue by dividing C_2 into C_1 to get a relation

$$A_2(z)C_1(z,w) = B_3(z,w)C_2(z,w) + C_3(z,w) \ .$$

Here the degree of C_3 with respect to w is at most $n-4$. This process must come to a halt after a finite number $m (1 \leq m \leq n-1)$ division steps. It terminates in an identity

$$A_{m-1}(z)C_{m-2}(z,w) = B_m(z,w)C_{m-1}(z,w) + C(z) \ ,$$

where C is a polynomial function of the variable z alone. The function C is not the zero polynomial. If it were, we would have $A_{m-1}\, C_{m-2} = B_m C_{m-1}$, making C_{m-1} a factor of $A_{m-1}C_{m-2}$. Thus, C_{m-1} would divide the right-hand side of the equation — to simplify the notation we henceforth suppress any mention of the variables z and w —

$$A_{m-1}A_{m-2}C_{m-3} = B_{m-1}A_{m-1}C_{m-2} + A_{m-1}C_{m-1} \ ,$$

and so would also be a factor of $A_{m-1}A_{m-2}C_{m-3}$. Backtracking in this way we would find C_{m-1} among the factors of GP, with $G = A_{m-1}A_{m-2}\cdots A_0$. The irreducibility of P would force C_{m-1} to be a factor of G, which it clearly is not, for C_{m-1} has positive degree with respect to w and G doesn't. The conclusion: C is not the zero polynomial. Lastly, we can express C as

$$C = A_{m-1}C_{m-2} + (-B_m)C_{m-1} .$$

Since $C_{m-1} = A_{m-2}C_{m-3} - B_{m-1}C_{m-2}$, this leads to a second representation for C,

$$C = (-B_mA_{m-2})C_{m-3} + (A_{m-1} - B_mB_{m-1})C_{m-2} .$$

Through the repetition of this back-substitution process we eventually arrive at an identity of the form (10.42).

Modulo our ability to locate the roots of p_n and C, the preceding discussion furnishes us with an algorithm to help pin down the exceptional points for an irreducible polynomial P. (*Warning.* Unless C is actually the discriminant of P, there may be zeros of C that are not exceptional for P.) Another spin-off from this discussion and from Theorem 3.5 is the identification of a new situation in which analyticity can be inferred from continuity.

Theorem 3.6. *Let $P(z,w) = p_0(z) + p_1(z)w + \cdots + p_n(z)w^n$ be an irreducible polynomial in z and w. If D is a domain in $\mathbb{C}$ and if $f: D \to \mathbb{C}$ is a continuous function that satisfies $P[z, f(z)] = 0$ for every z in D, then f is an analytic function.*

Proof. Since P is irreducible, the set E of exceptional points for P that lie in D is a finite set. Theorem 3.5 implies that f is analytic in $D \sim E$. The continuity of f insures that it has no worse than removable singularities at the points of E — hence, is actually analytic in D. ∎

An analytic function element (f, D) with the feature that $P[z, f(z)] = 0$ for every z in D, where P is a non-zero polynomial function of two complex variables, is known as an *algebraic function element.* The crowning result of this section is an existence theorem for algebraic function elements.

Theorem 3.7. *Suppose that a simply connected domain D in the complex plane contains none of the exceptional points for a polynomial $P(z,w) = p_0(z) + p_1(z)w + \cdots + p_n(z)w^n$. If (z_0, w_0) is a point of $\mathbb{C}^2$ such that z_0 belongs to D and $P(z_0, w_0) = 0$, then there is a unique analytic function element (f, D) such that $f(z_0) = w_0$ and $P[z, f(z)] = 0$ for every z in D.*

Proof. By Theorem 3.5 there is a unique analytic germ g at z_0 about which is true that $g(z_0) = w_0$ and $P[z, g(z)] = 0$ throughout the domain-set $\Delta_0 = \Delta(z_0, \rho_0)$ of g. We shall prove that g can be analytically continued along every path in D originating at z_0. Once this has been shown, we can quote Theorem 3.4 and assert the existence of a unique analytic function element (f, D) whose germ at z_0 is g. Then $f(z_0) = g(z_0) = w_0$ and, since $P[z, f(z)] = P[z, g(z)] = 0$ for all z in some open disk centered at z_0, the principle of analytic continuation dictates that $P[z, f(z)] = 0$ everywhere in D. Furthermore, any analytic function element (f, D) with these two properties must, according to the last statement in Theorem 3.5, have g as its germ at z_0. The uniqueness aspect of the present result is then seen to be a consequence of uniqueness in Theorem 3.4. The proof thus comes down to demonstrating the continuability of g along the necessary paths in the domain D.

Let $\gamma: [a, b] \to D$ be a path with initial point z_0. For each c satisfying $a < c \le b$ denote by γ_c the restriction of γ to the interval $[a, c]$. We consider the set E of all c in $(a, b]$ such that g admits an analytic continuation along the path γ_c. If c is close enough to a that $\gamma([a, c])$ is contained in the disk $\Delta(z_0, \rho_0/2)$, then, as previously noted, the continuation $\{g_t\}$ of g along γ_c exists: for $a \le t \le c$, g_t is just the germ determined by g at $\gamma(t)$. Therefore, the set E is not empty. Let d be its least upper bound — then $a < d \le b$ — and set $z_0^* = \gamma(d)$. We verify (i) that d is a member of E and (ii) that $d = b$. Because z_0^* is not an exceptional point for P, the equation $P(z_0^*, w) = 0$ has n different roots $w_1, w_2, \ldots, w_n$. In view of Theorem 3.5 there exists for each w_j a unique analytic germ g_j at z_0^* satisfying the requirements $g_j(z_0^*) = w_j$ and $P[z, g_j(z)] = 0$ for every z in $\Delta_j = \Delta(z_0^*, \rho_j)$, the domain-set of g_j. We choose and fix a number r with the feature that $0 < r < \min\{\rho_j/2 : 1 \le j \le n\}$. Next we select a number c in E, $c < d$, enjoying the property that $\gamma([c, d])$ lies in $\Delta(z_0^*, r)$. If β is the restriction of γ to the interval $[c, d]$, then each of the germs g_j can be analytically continued along the path $-\beta$, for the trajectory of this path is a subset of $\Delta(z_0^*, \rho_j/2)$. We use $\widetilde{g}_j$ to indicate the germ at the point $[-\beta](d) = \gamma(c)$ obtained when g_j is so continued. Being in fact a direct analytic continuation of g_j, the germ $\widetilde{g}_j$ obeys the condition $P[z, \widetilde{g}_j(z)] = 0$ throughout its domain-set, a set that includes the point z_0^*. Since their respective values at the point z_0^* — namely, $w_1, w_2, \ldots, w_n$ — are different, the germs $\widetilde{g}_1, \widetilde{g}_2, \ldots, \widetilde{g}_n$ are distinct. Of course, $\widetilde{g}_j$ can be continued along $\beta = -(-\beta)$ to g_j. By the definition of E the germ g is analytically continuable along the path γ_c to a germ $\widetilde{g}$ at $\gamma(c)$. As an analytic continuation of g, the germ $\widetilde{g}$ satisfies $P[z, \widetilde{g}(z)] = 0$ for every z in its domain-set. A glance at the final sentence in Theorem 3.5 reveals that $\widetilde{g}$ must be one of the germs $\widetilde{g}_1, \widetilde{g}_2, \ldots, \widetilde{g}_n$. As such, $\widetilde{g}$ is analytically continuable along β. We infer that g can be continued along the path $\gamma_c + \beta = \gamma_d$. This makes d an element of E. Finally, if d were less than b, we could pick a number c_1 such that $d < c_1 \le b$ and such that $\gamma([d, c_1])$ is contained in $\Delta(z_0^*, r)$. Then g^*, the end-product of the continuation of g along γ_d, could be further continued along α, the restriction of γ to the interval $[d, c_1]$. Therefore, g would admit an analytic continuation along the path $\gamma_d + \alpha = \gamma_{c_1}$, in conflict with the definition of d. It follows that $d = b$ and that g can be analytically continued along $\gamma = \gamma_b$, as insisted. ∎

3.6 Global Analytic Functions

If one starts with an analytic function element (f_0, D_0) and builds the collection of all function elements (f, D) that are analytic continuations of (f_0, D_0), one is faced with what appears to be a colossal hodgepodge of functions and domains. By no means is it evident that these function elements, these "patches" of function, fit together in any coherent and harmonious way to form a "whole" function. Nor should it be evident, for the very good reason that the construction which manages to consolidate all these elements into a single function generally forces one out of the familiar territory of the complex plane and requires one to do complex analysis on a "Riemann surface." In this section we afford the reader a glimpse of that construction and its implications. As indicated earlier, however, limitations of space prevent us from giving anything beyond an overview of the subject. Those who find their interest piqued by the ideas that follow should at this stage be adequately prepared to pursue them in a more advanced text.

The fundamental concept involved here is that of a "global analytic function." A *global analytic function* $\mathcal{F}$ (the expression *complete analytic function* is sometimes used to describe the same object) is a non-empty family of analytic function elements that enjoys the following two special properties: (i) any pair of function elements belonging to $\mathcal{F}$ are analytic continuations of each other and (ii) any analytic function element that is related through analytic continuation to any function element in $\mathcal{F}$ is itself a member of $\mathcal{F}$. In short, $\mathcal{F}$ is the totality of analytic function elements that can be obtained through the arbitrary analytic continuation of any one of its members. Each of the individual function elements that make up $\mathcal{F}$ is known as a *branch of* $\mathcal{F}$. The reader may be excused for looking with suspicion at the title "global analytic function" we have bestowed on $\mathcal{F}$. It would seem to be a serious misnomer: $\mathcal{F}$ is not a function at all, but a whole collection of functions! A justification for the terminology will be offered in due course.

Everyone's favorite example of a global analytic function is the succeeding one.

EXAMPLE 3.1. Let $\mathcal{L}$ denote the collection of all analytic function elements (f, D) with the property that $e^{f(z)} = z$ for every z in D. Show that $\mathcal{L}$ is a global analytic function.

If (f, D) belongs to $\mathcal{L}$ and if (f_1, D_1) is a direct analytic continuation of (f, D), then $e^{f_1(z)} = e^{f(z)} = z$ for every z in $D \cap D_1$. By the principle of analytic continuation $e^{f_1(z)} = z$ holds everywhere in D_1, so (f_1, D_1) is a member of $\mathcal{L}$. Since an arbitrary analytic continuation of (f, D) is linked to it by a finite chain of direct analytic continuations, any such continuation stays in $\mathcal{L}$. All that remains for us to check is that any two members of $\mathcal{L}$ are analytic continuations of each other. Because the relation here is a transitive one and because a function element is obviously an analytic continuation of every germ that it generates, it suffices to demonstrate that every germ in $\mathcal{L}$ is an analytic continuation of some fixed member of $\mathcal{L}$. For this member we pick the function element (f_0, D_0), in which $D_0 = \mathbb{C} \sim (-\infty, 0]$ and $f_0(z) = \operatorname{Log} z$.

Suppose now that $\widetilde{g}$, an analytic germ at a point z_0, belongs to $\mathcal{L}$. Then $\widetilde{g}$ constitutes a branch of $\log z$ in its domain-set $\widetilde{\Delta} = \Delta(z_0, \rho)$, so $z_0 \neq 0$. A straightforward calculation reveals that

$$\widetilde{g}(z) = \widetilde{g}(z_0) + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}(z - z_0)^n}{n\, z_0^n}$$

in $\widetilde{\Delta}$. In particular, we see that $\rho = |z_0|$. Next, we write $z_0 = |z_0|e^{i\theta_0}$ with $0 \le \theta_0 < 2\pi$. Then

$$\widetilde{g}(z_0) = \operatorname{Log}|z_0| + i(\theta_0 + 2k_0\pi)$$

for some integer k_0. We proceed assuming $k_0 \ge 0$, the case $k_0 < 0$ being handled similarly. Define a path γ by $\gamma(t) = |z_0|e^{it}$ for $0 \le t \le \theta_0 + 2k_0\pi$. (N.B. Should $\theta_0 = k_0 = 0$, then $\widetilde{g}$ would be the germ determined by f_0 at z_0, and none of what ensues would be necessary. For this reason we shall suppose that either $\theta_0 > 0$ or $k_0 > 0$. Incidentally, in the case where $k_0 < 0$ the path γ would be replaced in the following discussion by $\gamma(t) = |z_0|e^{-it}$, $0 \le t \le -\theta_0 - 2k_0\pi$.) We claim that $\widetilde{g}$ is obtained via the analytic continuation of g, the germ generated by f_0 at the point $|z_0|$ of the positive real axis, along the path γ. If this is true, it certainly marks $\widetilde{g}$ as an analytic continuation of g — hence, of (f_0, D_0). Observe that g is given by

$$g(z) = \operatorname{Log}|z_0| + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}(z - |z_0|)^n}{n|z_0|^n}$$

in the disk $\Delta = \Delta(|z_0|, |z_0|)$. For $0 \le t < \infty$ let g_t designate the analytic germ defined by

$$g_t(z) = \operatorname{Log}|z_0| + it + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}(z - |z_0|e^{it})^n}{n|z_0|^n e^{int}}$$

in the domain-set $\Delta_t = \Delta(|z_0|e^{it}, |z_0|)$; i.e., g_t is the unique branch of $\log z$ in Δ_t whose value at the point $|z_0|e^{it}$ is $\operatorname{Log}|z_0| + it$. Plainly $g_0 = g$ and $g_{\theta_0 + 2k_0\pi} = \widetilde{g}$. Furthermore, when $|s - t| < \pi$ the germs g_t and g_s are direct analytic continuations of each other. Perhaps the easiest way to make this clear is to express g_t in closed form: when $0 \le t \le \pi/2$, $g_t(z) = \operatorname{Log} z$ for every z in Δ_t; when $\pi/2 < t < 3\pi/2$,

$$g_t(z) = \begin{cases} \operatorname{Log} z & \text{if } z \in \Delta_t \text{ and } \operatorname{Im} z \geq 0 , \\ \operatorname{Log} z + 2\pi i & \text{if } z \in \Delta_t \text{ and } \operatorname{Im} z < 0 ; \end{cases}$$

when $3\pi/2 \leq t \leq 2\pi$, $g_t(z) = \operatorname{Log} z + 2\pi i$ for every z in Δ_t; for t satisfying $2k\pi \leq t < 2(k+1)\pi$ with $k \geq 1$, $g_t = 2k\pi i + g_{t-2k\pi}$. This description leaves no doubt that $g_t = g_s$ in $\Delta_t \cap \Delta_s$ when $|s-t| < \pi$. It follows that $\{g_t\}$, where we now confine t to the interval $[0, \theta_0 + 2k_0\pi]$, is an analytic continuation of g to $\widetilde{g}$ along γ. We have thus verified that $\mathcal{L}$ is a global analytic function.

The global analytic function $\mathcal{L}$ in Example 3.1 is called the *global logarithm function*. A similar discussion would reveal that $\mathcal{F}$, the family of all analytic function elements (f, D) that satisfy $[f(z)]^2 = z$ for every z in D, is a global analytic function, the *global square root function* (Exercise 4.43). More generally, if $P(z,w) = p_0(z) + p_1(z)w + \cdots + p_n(z)w^n$ is an irreducible polynomial function of z and w, then the collection $\mathcal{A}_P$ consisting of all analytic function elements (f, D) such that $P[z, f(z)] = 0$ throughout D is a global analytic function, the *global algebraic function corresponding to* P. The irreducibility of P is crucial to this pronouncement. The argument which shows that all members of $\mathcal{A}_P$ are related by analytic continuation is non-trivial. Because it involves a number of ideas unavailable in this book, we omit the proof.

Given a global analytic function $\mathcal{F}$ we carry out the following construction. For each fixed z in $\mathbb{C}$ we build what can be envisioned as a stack of points above z, the elements of which are in one-to-one correspondence with the collection of analytic germs at z that belong to $\mathcal{F}$. We write $p = (z, g)$ to signify that p is the point of the stack associated with the germ g at z. Of course, it is entirely possible that $\mathcal{F}$ contains no germs at z, in which event the stack above z is empty. This happens, for instance, in the case of the global logarithm function $\mathcal{L}$ at $z = 0$. Over any point z_0 other than the origin the stack determined by $\mathcal{L}$ comprises infinitely many points $p_0, p_1, p_{-1}, p_2, p_{-2}, \cdots$: for each integer k we set $p_k = (z_0, g_k)$, where g_k is the unique branch of $\log z$ in the disk $\Delta(z_0, |z_0|)$ whose value at z_0 is $\operatorname{Log} z_0 + 2k\pi i$. If $P(z,w) = p_0(z) + p_1(z)w + \cdots + p_n(z)w^n$ is an irreducible polynomial in z and w, then the global analytic function $\mathcal{A}_P$ generates a stack of n points above every point z of $\mathbb{C}$ that is not exceptional for P, one stack-point corresponding to each solution w of the equation $P(z,w) = 0$ (Theorem 3.5). Over the exceptional points for P the stacks contain fewer than n points and may well be empty, as the one above the origin is when $P(z,w) = w^2 - z$. If we now assemble all the stacks to which a global analytic function $\mathcal{F}$ gives rise in this way, we obtain an object called the "Riemann surface" of $\mathcal{F}$. To formalize the definition: the *Riemann surface of the global analytic function* $\mathcal{F}$ is the set $S(\mathcal{F})$ whose elements are the ordered pairs $p = (z, g)$, where z is a point of the complex plane and g is an analytic germ at z that belongs to $\mathcal{F}$. (N.B. To say that $p_1 = p_2$ for two such pairs $p_1 = (z_1, g_1)$ and $p_2 = (z_2, g_2)$ means that $z_1 = z_2$ and that g_1 and g_2 are the same germ at z_1. The latter condition demands more than just $g_1(z_1) = g_2(z_1)$. It insists that g_1 and g_2 be represented by the same Taylor series at z_1; i.e., it insists that $g_1^{(n)}(z_1) = g_2^{(n)}(z_1)$ for $n = 0, 1, 2, \cdots$.) The image which $S(\mathcal{F})$ is supposed to conjure up is that of a multi-tiered surface suspended over a portion of the complex plane. The technical name for what we have heretofore described as the "stack of points" generated by $\mathcal{F}$ above z is the *fiber of* $S(\mathcal{F})$ *over* z.

(*Remark.* As defined, the Riemann surface $S(\mathcal{F})$ of a global analytic function $\mathcal{F}$ is, properly speaking, a subset of the cartesian product $\mathbb{C} \times \mathcal{F}$. In many instances, however, $S(\mathcal{F})$ is realizable as an actual surface in the four-dimensional space $\mathbb{C}^2$. Take, for example, the case of the global logarithm function $\mathcal{L}$. The Riemann surface $S(\mathcal{L})$ can be legitimately identified with $\{(z, w) \in \mathbb{C}^2 : e^w = z\}$, a concrete two-dimensional surface in $\mathbb{C}^2$. If we regard the complex plane as sitting inside $\mathbb{C}^2$ disguised as the plane $\{(z, w) \in \mathbb{C}^2 : w = 0\}$, then the fiber of $S(\mathcal{L})$ over any non-zero point $(z, 0)$ of the complex plane — the preposition "over" is slightly misleading here — reduces to the set $\{(z, \operatorname{Log} z + 2k\pi i) : k = 0, \pm 1, \pm 2, \cdots\}$ in $\mathbb{C}^2$. Similarly, when $\mathcal{F}$ is the global square root function, the Riemann surface $S(\mathcal{F})$ can be thought of as $\{(z, w) \in \mathbb{C}^2 : z \neq 0, w^2 = z\}$. The fiber of $S(\mathcal{F})$ over $(z, 0)$ in this situation consists of the two points $(z, \sqrt{z})$ and $(z, -\sqrt{z})$.)

There are two functions of particular importance defined on the Riemann surface $S(\mathcal{F})$. The first of these is the function $\mathcal{P}: S(\mathcal{F}) \to \mathbb{C}$ with rule of correspondence $\mathcal{P}(z, g) = z$. It is termed the *projection of* $S(\mathcal{F})$ *to the complex plane.* We now force the symbol $\mathcal{F}$ to do double duty, in that we retain it to denote the second function in question here, for it is the function on $S(\mathcal{F})$ that incorporates into a single function the diverse analytic function elements which make up the global analytic function $\mathcal{F}$. The function $\mathcal{F}: S(\mathcal{F}) \to \mathbb{C}$ is defined by $\mathcal{F}(z, g) = g(z)$. (N.B. As g is a germ at z, the expression $g(z)$ makes perfectly good sense.) Let (f, D) be any of the analytic function elements that constitute $\mathcal{F}$. We can associate with (f, D) a replica of the domain D on the surface $S(\mathcal{F})$, the set $\widetilde{D}$ consisting of all points (z, g) on $S(\mathcal{F})$ such that z lies in D and g is the germ determined at z by f. (The function $\mathcal{P}$ projects $\widetilde{D}$ in a one-to-one fashion onto D, and this projection has every nice property imaginable.) Let (z, g) be an element of $\widetilde{D}$. Since g is the germ of f at z, we see that $\mathcal{F}(z, g) = g(z) = f(z)$. Thus, $\mathcal{F}$ assigns to a point of $\widetilde{D}$ exactly the same value that f assigns to the corresponding point of D. As things turn out, the mimicry of f that $\mathcal{F}$ displays in $\widetilde{D}$ is complete. For all intents and purposes we might just as well regard D as a subset of $S(\mathcal{F})$ and f as the restriction of $\mathcal{F}$ to that subset. It is in this sense that $\mathcal{F}$ combines all the function elements of its namesake global analytic function into one function. A price is exacted for this consolidation: we must abandon the complex plane and work instead on $S(\mathcal{F})$. Fortunately, at least as far as the function $\mathcal{F}$ is concerned, the transition to complex analysis on $S(\mathcal{F})$ is quite straightforward. For example, differentiation can be done as follows: $\mathcal{F}'(z,g) = g'(z)$, $\mathcal{F}''(z,g) = g''(z)$, etc.

Let $p_0 = (z_0, g_0)$ be a point of $S(\mathcal{F})$, where g_0 has domain-set $\Delta(z_0, \rho_0)$. For r satisfying $0 < r \leq \rho_0$ we designate as the *open disk on $S(\mathcal{F})$ with center p_0 and radius r* the set $\Delta(p_0, r)$ composed of all points (z, g) on $S(\mathcal{F})$ with the property that z is an element of $\Delta(z_0, r)$ and g is the germ generated at z by g_0. The disk $\Delta(p_0, r)$ is a carbon copy of the plane disk $\Delta(z_0, r)$ and lies above it on the surface $S(\mathcal{F})$. We emphasize that $\Delta(p_0, r)$ is defined not for arbitrary $r > 0$, but only for r in the interval $(0, \rho_0]$. If Δ_1 and Δ_2 are open disks on $S(\mathcal{F})$ and if p is a point of $\Delta_1 \cap \Delta_2$, it is not difficult to check that this intersection contains some open disk centered at p. Also, if p_1 and p_2 are distinct points of $S(\mathcal{F})$, then there exists an $r > 0$ for which $\Delta(p_1, r)$ and $\Delta(p_2, r)$ are disjoint.

With the notion of an open disk on $S(\mathcal{F})$ thus established, the way is cleared for the installation in $S(\mathcal{F})$ of the full apparatus of topology, embracing all of the concepts discussed previously in the contexts of the complex plane $\mathbb{C}$ and extended complex plane $\widehat{\mathbb{C}}$: open set, closed set, convergent sequence, connected set, domain, compact set, etc. Indeed, since the earlier definitions were couched in terms of open disks, they transfer to the present setting with little or no formal change . One important topological fact about $S(\mathcal{F})$ is that it is connected. Another is that open disks on $S(\mathcal{F})$ are domains; i.e., they are open, connected sets. We are also entitled to speak of continuity for functions of the type $\Phi: A \to S(\mathcal{F})$, where A is a subset of $\mathbb{C}$ (or $\widehat{\mathbb{C}}$), and those of the kind $\Phi: A \to \mathbb{C}$ (or $\Phi: A \to \widehat{\mathbb{C}}$), in which A is a subset of $S(\mathcal{F})$. Once again, the previous definitions, which were formulated with the aid of disks, carry over essentially verbatim. To cite two examples, both the projection $\mathcal{P}$ and the function $\mathcal{F}$ are readily seen to be continuous functions from $S(\mathcal{F})$ into the complex plane. As a matter of fact, the restriction of $\mathcal{P}$ to any open disk $\Delta(p, r)$ on $S(\mathcal{F})$ furnishes a homeomorphism between $\Delta(p, r)$ and the plane disk $\Delta[\mathcal{P}(p), r]$.

Paths on $S(\mathcal{F})$ enjoy a special significance. A path Γ on $S(\mathcal{F})$ is just a continuous function of the type $\Gamma: [a, b] \to S(\mathcal{F})$. For reasons that will be clear in a second, we express $\Gamma(t)$ in the manner $\Gamma(t) = (\gamma(t), g_t)$, where $\gamma: [a, b] \to \mathbb{C}$ is the path in the complex plane defined by $\gamma(t) = \mathcal{P}[\Gamma(t)]$. (We refer to γ as the *projection of Γ into the complex plane.*) Let $p_1 = (z_1, g)$ be the initial point of Γ, and let $p_2 = (z_2, \widetilde{g})$ be its terminal point. For each t in $[a, b]$ we are presented by Γ with a germ g_t at $\gamma(t)$. By definition, $g_a = g$ and $g_b = \widetilde{g}$. Furthermore, a moment's reflection should convince one that the definition of continuity for Γ at a point t_0 of $[a, b]$ is a restatement of the fact that $\{g_t\}$ satisfies at t_0 the local compatibility condition for an analytic continuation along γ. The result: $\{g_t\}$ gives the analytic continuation of g to $\widetilde{g}$ along γ. Conversely, if $\{g_t\}$ is known to be the analytic continuation of a germ g to a germ $\tilde{g}$ along a path $\gamma: [a,b] \to \mathbb{C}$, then $\Gamma(t) = (\gamma(t), g_t)$ defines a path on $S(\mathcal{F})$ starting at the point $p_1 = (\gamma(a), g)$ and terminating at $p_2 = (\gamma(b), \tilde{g})$. (The path Γ is called a *lift* of γ to $S(\mathcal{F})$.) One consequence of the definition of a global analytic function is that any two points p_1 and p_2 of $S(\mathcal{F})$ can be joined by a path on that surface.

It is possible to do contour integration on $S(\mathcal{F})$. If $\Gamma: [a,b] \to S(\mathcal{F})$ is a piecewise smooth path — this condition simply asks that the projection γ of Γ into $\mathbb{C}$ be piecewise smooth — and if Φ is a continuous, complex-valued function on the trajectory of Γ, then we make the definitions

$$\int_\Gamma \Phi(p)\, dp = \int_a^b \Phi[\Gamma(t)]\, \dot{\gamma}(t)\, dt$$

and

$$\int_\Gamma \Phi(p)\, |dp| = \int_a^b \Phi[\Gamma(t)] |\dot{\gamma}(t)|\, dt\ .$$

The first equation defines the *complex line integral of* Φ *along* Γ, the second gives meaning to the *integral of* f *along* Γ *with respect to arclength.* As an illustration, we remark that

$$\int_\Gamma \mathcal{F}'(p)\, dp = \mathcal{F}[\Gamma(b)] - \mathcal{F}[\Gamma(a)]\ ,$$

a formula whose verification is left to the reader as an instructive exercise. (See Exercise 4.49.)

Finally, we introduce the counterpart of analyticity for functions whose domain-sets are subsets of $S(\mathcal{F})$. Let $\Delta = \Delta(p_0, r)$ be an open disk on $S(\mathcal{F})$, and let $\mathcal{P}_\Delta$ designate the restriction of the function $\mathcal{P}$ to Δ. Then $\mathcal{P}_\Delta$ projects Δ homeomorphically onto the plane disk $\mathcal{P}(\Delta) = \Delta[\mathcal{P}(p_0), r]$. In particular, $\mathcal{P}_\Delta$ has an inverse function $\mathcal{P}_\Delta^{-1}: \mathcal{P}(\Delta) \to \Delta$. A complex-valued function Φ whose domain-set includes Δ is pronounced *holomorphic in* Δ provided the function $\varphi = \Phi \circ \mathcal{P}_\Delta^{-1}$, which is just an ordinary complex-valued function of a complex variable, is analytic. (The literature talks of $\mathcal{P}_\Delta$ as a "local coordinate chart" on $S(\mathcal{F})$ at p_0. The statement that $\Phi \circ \mathcal{P}_\Delta^{-1}$ is analytic is often rendered by the phrase "Φ is analytic in Δ with respect to the local coordinate system given by $\mathcal{P}_\Delta$.") To differentiate Φ at a point p of Δ we take derivatives of φ at the corresponding point of $\mathcal{P}(\Delta)$: $\Phi'(p) = \varphi'[\mathcal{P}(p)]$, $\Phi''(p) = \varphi''[\mathcal{P}(p)]$, etc. For example, both $\mathcal{P}$ and $\mathcal{F}$ are holomorphic functions in Δ: if $p_0 = (z_0, g_0)$, then we obtain

$$\mathcal{P} \circ \mathcal{P}_\Delta^{-1}(z) = z \quad , \quad \mathcal{F} \circ \mathcal{P}_\Delta^{-1}(z) = g_0(z)$$

for every z in $\mathcal{P}(\Delta)$, which confirms that $\mathcal{P} \circ \mathcal{P}_\Delta^{-1}$ and $\mathcal{F} \circ \mathcal{P}_\Delta^{-1}$ are analytic there. A function Φ is said to be *holomorphic in* U, an open subset of $S(\mathcal{F})$, if Φ is holomorphic in every open disk on $S(\mathcal{F})$ that is contained in U. For such Φ it makes sense to take derivatives $\Phi'(p)$, $\Phi''(p)$, $\Phi''(p), \cdots$

at a point p of U: we merely choose an open disk Δ in U that contains p and compute the derivatives of $\Phi \circ \mathcal{P}_\Delta^{-1}$ at the point $\mathcal{P}(p)$. The results do not depend on which disk Δ is selected. Prime examples of functions that are holomorphic on the whole surface $S(\mathcal{F})$ are, as might be expected, $\mathcal{P}$ and $\mathcal{F}$. Other simple examples include $\mathcal{P}^2$, $\mathcal{P}+\mathcal{F}$, $\mathcal{P}\mathcal{F}^2+i\mathcal{F}'$, and $\mathcal{F}'e^{\mathcal{F}}$. The concept of meromorphicity for a function defined in an open subset of $S(\mathcal{F})$, but taking values in $\widehat{\mathbb{C}}$, is treated in a similar fashion.

We have tried in this section to indicate how many of the concepts met throughout the book can be transported from the complex plane to the Riemann surface $S(\mathcal{F})$. Naturally, this whole procedure raises many questions. What changes does the theory undergo in the new setting? What form do fundamental theorems, such as Cauchy's theorem, take on $S(\mathcal{F})$? What is the significance of points of $\mathbb{C}$ whose fibers in $S(\mathcal{F})$ are empty or "deficient" in some way? And most importantly, is there any real purpose to generalizing complex analysis along these lines? The answers to these and other questions will have to be sought elsewhere. We end here by taking a parting look at two of the Riemann surfaces that have received occasional mention in the preceding pages.

A point p on the Riemann surface $S(\mathcal{L})$ of the global logarithm function $\mathcal{L}$ has the form $p=(z,g)$, where g is a germ of $\mathcal{L}$ at z. The value of g at z is $\operatorname{Log} z + 2k\pi i$ for some integer k. We can use this integer as a handy means of identifying p, writing $p=(z,k)$ in place of $p=(z,g)$. For a fixed integer k we refer to the set $S_k = \{(z,k) \in S(\mathcal{L}) : z \neq 0\}$ as the k^{th} *sheet of* $S(\mathcal{L})$. Of course, $S(\mathcal{L}) = \cup_{k=-\infty}^{\infty} S_k$. What do these sheets actually look like and just how do they fit together to form $S(\mathcal{L})$; i.e., how can we utilize them to picture this surface? To answer these questions we assume that we have at our disposal an infinite number of facsimiles of the complex plane, one copy for each integer k. We distinguish between the various copies by writing $\mathbb{C}_k$ to specify the copy corresponding to k. A model of the sheet S_k is produced by slitting $\mathbb{C}_k$ along the entire length of its negative real axis — i.e., forming $\mathbb{C}_k \sim (-\infty,0]$ — and then restoring the points of the interval $(-\infty,0)$ to the upper edge of the slit (Figure 3). (Think of physically cutting a piece of paper along a ray, but imagine doing it so deftly that all the paper molecules on the cut-line itself wind up on the upper edge of the cut!) To build $S(\mathcal{L})$ from these sheets we perform a massive gluing job: for each integer k we paste S_k and S_{k+1} together along half of their respective boundaries, each point on the upper edge (bold line) of the slit that borders S_k being matched with the corresponding point on the lower edge (dotted line) of the slit bordering S_{k+1}. In this way we form from the sundry sheets S_k a continuous surface $S(\mathcal{L})$, S_k blending smoothly with S_{k-1} below it and S_{k+1} above it for every integer k. In moving on $S(\mathcal{L})$ across a former slit line one makes a transition from one sheet to an adjoining sheet. Perhaps the most evocative image of $S(\mathcal{L})$ that emerges from this construction is that of an immense spiral ramp, as one sometimes finds in parking garages, winding endlessly upward and

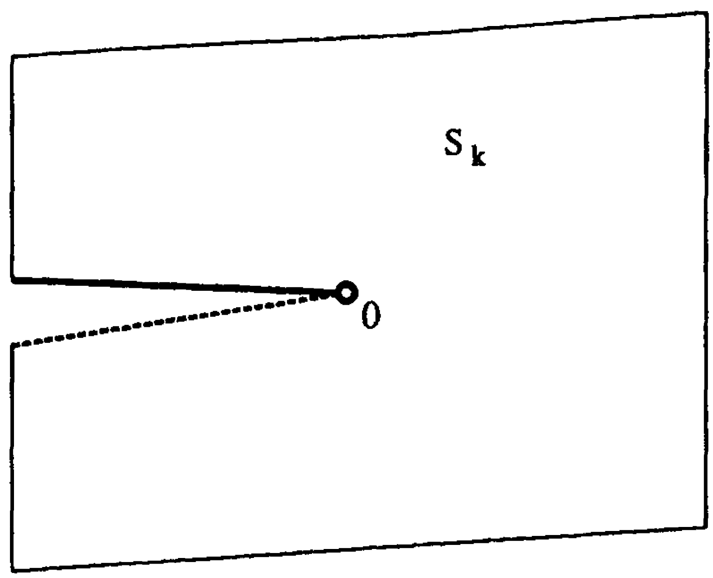


Figure 3.

downward around a central axis. In the above notation the holomorphic function $\mathcal{L}$ is given on $S(\mathcal{L})$ by $\mathcal{L}(z,k) = \operatorname{Log} z + 2k\pi i$. Its derivative has the formula $\mathcal{L}'(z,k) = z^{-1}$. Furthermore, $\mathcal{L}$ provides a conformal mapping (i.e., a univalent holomorphic function) of $S(\mathcal{L})$ onto the complex plane. From a conformal point of view, $S(\mathcal{L})$ is thus the same as the complex plane!

If $\mathcal{F}$ is the global square root function, then a point p on the Riemann surface $S(\mathcal{F})$ has the structure $p = (z, g)$, where the germ g is a branch of the square root function in the disk $\Delta(z, |z|)$. At the point z itself the value of g is either $\sqrt{z}$ or $-\sqrt{z}$. If we relabel the point p for which the former is true $p = (z, +)$ and the point for which the latter holds $p = (z, -)$, then we achieve a decomposition of $S(\mathcal{F})$ into two sheets, $S_+ = \{(z,+) : z \neq 0\}$ and $S_- = \{(z,-) : z \neq 0\}$. To construct the Riemann surface $S(\mathcal{F})$ we start with two copies of the complex plane — call them $\mathbb{C}_+$ and $\mathbb{C}_-$ — we slit each along its negative real axis, just as we did in constructing $S(\mathcal{L})$, in order to obtain concrete realizations of S_+ and S_-, and we then fasten S_+ and S_- together according to the following prescription: the upper edge of the slit that borders S_+ is joined to the lower edge of the slit on the boundary of S_- — and vice versa. As things stand the result of this cut-and-paste operation is admittedly hard to visualize. If a correct topological picture of $S(\mathcal{F})$ is all that is desired, however, a simple one is available. We merely have to substitute for the complex plane slit along the negative real axis its topological equivalent, the Riemann sphere slit from north pole to south pole along the great semi-circle that corresponds to the interval $(-\infty, 0)$ under stereographic projection. In this model S_+ and S_- amount topologically to hemispheres which, when joined together in accordance with the above instructions, form the whole Riemann sphere $\widehat{\mathbb{C}}$, except for punctures at the points 0 and ∞ (Figure 4). Even more is true. The holomorphic function $\mathcal{F}$, whose rule of correspondence can be

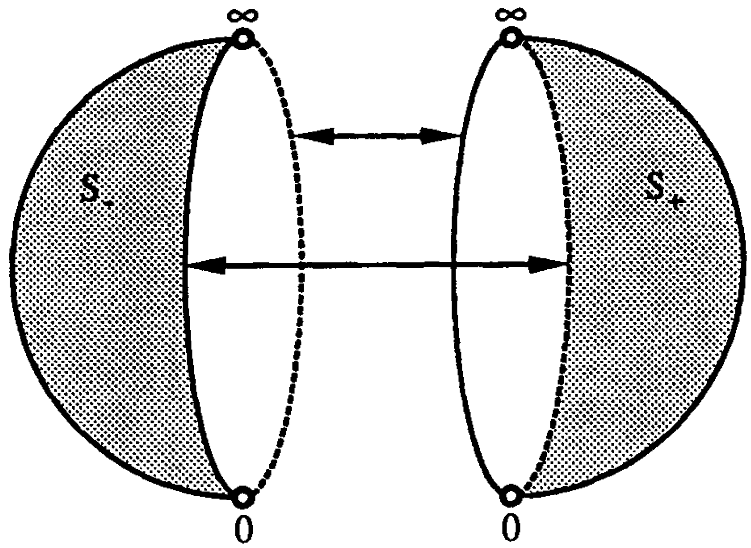


Figure 4.

expressed by $\mathcal{F}(z,+) = \sqrt{z}$ and $\mathcal{F}(z,-) = -\sqrt{z}$, provides a conformal mapping of $S(\mathcal{F})$ onto $\mathbb{C} \sim \{0\}$.

4 Exercises for Chapter X

4.1 Exercises for Section X.1

4.1. Prove Mittag-Leffler's theorem for an open subset U of $\widehat{\mathbb{C}}$ that contains the point ∞. (*Hint.* Let f_0 be the meromorphic function in $U_0 = U \sim \{\infty\}$ constructed in the proof of Theorem 1.2 for the set $E_0 = E \sim \{\infty\}$ and the given assignment of singular parts. Check that f_0 has a removable singularity at ∞. Remember: should the set E_1 that appears in the aforementioned proof be a finite set, then the contribution to f_0 arising from E_1 would require no convergence inducing summands.)

4.2. From (10.4) derive the identity

$$\pi \tan(\pi z) = -2 \sum_{n=-\infty}^{\infty} \left(\frac{1}{2z - 2n + 1} + \frac{1}{2n - 1} \right)$$

for z in $\mathbb{C}$. (*Hint.* At one point the fact that $\tan 0 = 0$ comes into play.)

4.3. Verify that

$$\pi \csc(\pi z) = \frac{1}{z} + \sum_{|n| \geq 1} (-1)^n \left(\frac{1}{z - n} + \frac{1}{n} \right)$$

for z in $\mathbb{C}$. (*Hint.* $2\csc(\pi z) = \tan(\pi z/2) + \cot(\pi z/2)$.)

4.4. Identify the meromorphic function f defined in $\mathbb{C}$ by $f(z) =$

$\sum_{n=-\infty}^{\infty}(-1)^n(z-n)^{-2}$.

4.5. Prove that $f(z)=\sum_{n=-\infty}^{\infty}(z^3-n^3)^{-1}$ defines a meromorphic function in $\mathbb{C}$. Determine this function explicitly.

4.6. If $\wp$ is the Weierstrass function associated with a lattice Ω, confirm that $\wp'(z)=0$ for every "half-lattice" point z; i.e., for every z such that $2z$ belongs to Ω, but z itself does not.

4.7. Let Ω be the lattice generated by ω_1 and ω_2, and let $\wp$ be its associated Weierstrass function. Demonstrate that

$$\zeta(z)=\frac{1}{z}+\sum_{\omega\in\Omega^*}\left(\frac{1}{z-\omega}+\frac{1}{\omega}+\frac{z}{\omega^2}\right)$$

defines a meromorphic function in $\mathbb{C}$ and that $\zeta'(z)=-\wp(z)$. Conclude that $\zeta(z+\omega_1)=\zeta(z)+\eta_1$ and $\zeta(z+\omega_2)=\zeta(z)+\eta_2$ for every z in $\mathbb{C}$, where η_1 and η_2 are constants. Prove that these constants obey "Legendre's relation": $\eta_1\omega_2-\eta_2\omega_1=2\pi i$. (*Hint.* For the last part integrate $\zeta(z)$ along the boundary of a certain parallelogram centered at the origin.)

4.8. Show that the only entire elliptic functions are the complex-valued constant functions on $\mathbb{C}$.

4.9. Let $\wp$ be the Weierstrass function associated with a lattice Ω. Since $\wp$ is an even function with singular part $S(z)=z^{-2}$ at the origin and since $\wp(z)-S(z)\to 0$ as $z\to 0$ (Why is this?), the Laurent expansion of $\wp$ in a small punctured disk $\Delta^*(0,r)$ has the form $\wp(z)=z^{-2}+a_2z^2+a_4z^4+O(z^6)$. Check that $a_2=3\sum_{\omega\in\Omega^*}\omega^{-4}$ and $a_4=5\sum_{\omega\in\Omega^*}\omega^{-6}$. Then show that the elliptic function $f=(\wp')^2-4\wp^3+20a_2\wp+28a_4$ satisfies $f(z)=O(z^2)$ in $\Delta^*(0,r)$, whence $f(z)\to 0$ as $z\to 0$. Deduce from this information that $f(z)=0$ for every z in $\mathbb{C}$. Conclude that $\wp$ and $\wp'$ are related by the equation $(\wp')^2=4\wp^3-20a_2\wp-28a_4$.

4.10. If f is a non-constant elliptic function whose period lattice Ω is generated by ω_1 and ω_2, if P is the closed parallelogram with vertices $0,\omega_1,\omega_2$, and $\omega_1+\omega_2$, and if E is the set (modulo Ω) of poles that f has in P, then E is non-empty and $\sum_{z\in E}\operatorname{Res}(z,f)=0$. Justify this statement. Conclude that it is not possible for f to have only a single (modulo Ω) simple pole in P. (*N.B.* In dealing with the zeros and poles of an elliptic function f — or, more generally, with solutions of equations of the sort $f(z)=w$ — in such a "period parallelogram" P, it is customary to work "modulo Ω," meaning that points z and $z+\omega$ with ω in Ω are considered to be one and the same point. For example, we would say that the Weierstrass function $\wp=\wp_\Omega$ has two poles in P, these accounted for by the double pole at the origin, for the actual poles which $\wp$ has at $0,\omega_1,\omega_2$, and $\omega_1+\omega_2$ coalesce modulo Ω. We abide by this convention in forming the set E in the present problem — we take all the poles of f, we eliminate duplicates modulo Ω, and only then do we sum the residues — and in similar situations later. *Hint.* By replacing f with $g(z) = f(z+c)$ for a suitable constant c, one can reduce the proof to the case in which f is free of poles on ∂P. Notice that no two points in the interior of P coincide modulo Ω.)

4.11. Let f and P be as in the preceding exercise. Show that, modulo Ω, the number of zeros that f has in P is equal to its number of poles in this set, provided multiplicity is taken into account. (N.B. If points z_0 and $\widetilde{z}_0$ coincide modulo Ω, then the multiplicity of f at z_0 is the same as its multiplicity at $\widetilde{z}_0$. Thus the multiplicity of a zero or pole is independent of which representative we pick for it modulo Ω.)

4.12. Again taking f and P as in Exercise 4.10, prove that $f(P) = \widehat{\mathbb{C}}$. More precisely, show that for any w in $\widehat{\mathbb{C}}$ the number of solutions (modulo Ω) in P of the equation $f(z) = w$ is independent of w, assuming that solutions are counted with proper regard for multiplicity.

4.13. Let $\wp$ be the Weierstrass function associated with the plane lattice Ω that is generated by ω_1 and ω_2, and let P be the closed parallelogram with vertices $0, \omega_1, \omega_2$, and $\omega_1+\omega_2$. Certify that $\wp'$ has simple zeros at the points $\omega_1/2$, $\omega_2/2$, and $(\omega_1+\omega_2)2$, but has (modulo Ω) no additional zeros in P. Use this information to verify that for given w in $\widehat{\mathbb{C}}$ the equation $\wp(z) = w$ has a unique solution (modulo Ω) in P when $w = \infty$, $\wp(\omega_1/2)$, $\wp(\omega_2/2)$, or $\wp[(\omega_1+\omega_2)/2]$, and exactly two solutions (modulo Ω) in P otherwise. Conclude, in particular, that the complex numbers $\wp(\omega_1/2)$, $\wp(\omega_2/2)$, and $\wp[(\omega_1+\omega_2)/2]$ are distinct. (*Hint.* Recall Exercises 4.6 and 4.12.)

4.14. Let $\Omega = \Omega(\omega_1, \omega_2)$ be a plane lattice, let $\wp$ be the corresponding Weierstrass function, let $a = 60\sum_{\omega\in\Omega^*}\omega^{-4}$ and $b = 140\sum_{\omega\in\Omega^*}\omega^{-6}$, and let C denote the locus of points (z,w) in $\mathbb{C}^2$ such that $w^2 = 4z^3 - az - b$. Show that for each element (z,w) of C there is a unique (modulo Ω) point ζ of the closed parallelogram P with vertices $0, \omega_1, \omega_2$, and $\omega_1+\omega_2$ for which $\wp(\zeta) = z$ and $\wp'(\zeta) = w$. (*Hint.* Utilize Exercises 4.9 and 4.13.)

4.15. If $\wp$ is the Weierstrass function corresponding to the plane lattice $\Omega = \Omega(\omega_1, \omega_2)$ and if $e_1 = \wp(\omega_1/2)$, $e_2 = \wp(\omega_2/2)$ and $e_3 = \wp[(\omega_1+\omega_2)/2]$, verify that $(\wp')^2 = 4(\wp - e_1)(\wp - e_2)(\wp - e_3)$. Conclude with the help of Exercise 4.9 that $e_1 + e_2 + e_3 = 0$. (*Hint.* Consider the function $f = (\wp - e_1)(\wp - e_2)(\wp - e_3)/(\wp')^2$. Use Exercise 4.8.)

4.16. Let $\omega_1 = 1$ and $\omega_2 = i$. If $\wp$ is the Weierstrass function associated with the lattice that ω_1 and ω_2 generate, check that $\wp(iz) = -\wp(z)$ and $\wp(\bar{z}) = \overline{\wp(z)}$. Deduce that, in the notation of the preceeding exercise, $e_2 = -e_1$, e_1 is real, and $e_3 = 0$.

4.17. Suppose that D is the interior of a closed rectangle R in the complex plane with one of its vertices at the origin and that f is a conformal mapping of D onto a half-plane D'. Confirm that f is the restriction to D of an elliptic function F having R as one quarter of a period parallelogram.

Would the same be true if, rather than a rectangle, R were some other parallelogram?

4.2 Exercises for Section X.2

4.18. Compute the infinite product $\prod_{n=1}^{\infty}[(n+1)^2(n^2+2n)^{-1}]$.

4.19. Test the following infinite products for convergence; in case of convergence, decide whether it is absolute: (i) $\prod_{n=1}^{\infty}[(-1)^n n(2n+1)^{-1}]$; (ii) $\prod_{n=1}^{\infty}[1+(-1)^n n^{-3/2}]$; (iii) $\prod_{n=1}^{\infty}\cos[n^{-1}]$; (iv) $\prod_{n=1}^{\infty}\sqrt[n]{n}$; (v) $\prod_{n=2}^{\infty}[1+(-1)^n n^{-1}\operatorname{Log}^{-2} n]$.

4.20. For which values of z does $\prod_{n=1}^{\infty}(1+z^n)$ converge. What about $\prod_{n=1}^{\infty}\sqrt[n]{z}$?

4.21. Let $\langle x_n\rangle$ be a sequence such that $0 \le x_n < 1$ for every n. Given that the infinite product $\prod_{n=1}^{\infty}(1-x_n)$ is convergent; prove that it is absolutely convergent, i.e., $\prod_{n=1}^{\infty}(1+x_n)$ converges. (*Hint.* Show that the sequence of partial products $\langle p_n\rangle$ corresponding to $\prod_{n=1}^{\infty}(1+x_n)$ is monotone and bounded. How do $1+x$ and $1-x$ compare when $0 < x \le 1$?)

4.22. Let $\lambda > 0$. Show that the product $\prod_{n=2}^{\infty}[1+(-1)^n n^{-\lambda}]$ converges absolutely when $\lambda > 1$, merely converges when $1/2 < \lambda \le 1$, and diverges when $0 < \lambda \le 1/2$. (*Hint.* In the case $0 < \lambda \le 1/2$ verify that $\prod_{n=2}^{2N}[1+(-1)^n n^{-\lambda}] \to 0$ as $N \to \infty$.)

4.23. Use Example 2.1 to determine $\prod_{n=2}^{\infty}(1-n^{-2})$.

4.24. Establish the identity $(1-z^2)^{-1} = \prod_{n=1}^{\infty}(1+z^{2^n})$ for $|z| < 1$.

4.25. Let $p_1 < p_2 < p_3 < \cdots$ be the sequence of prime natural numbers; i.e., $p_1 = 2, p_2 = 3, p_3 = 5, \cdots$. Demonstrate that the infinite product $\prod_{k=1}^{\infty}(1-p_k^{-z})^{-1}$ converges absolutely and normally in the half-plane $U = \{z : \operatorname{Re} z > 1\}$. Go on to prove that the function defined in U by this product is none other than Riemann's zeta-function! Through this process obtain the "Euler product formula" for $\zeta(z)$ in U: $\zeta(z) = \prod_{k=1}^{\infty}(1-p_k^{-z})^{-1}$. Infer from it that ζ is free of zeros in U. (*Hint.* To obtain the product formula make use of the fact that any integer $n \ge 2$ has a unique factorization of the type $n = p_{k_1}^{\alpha_1}p_{k_2}^{\alpha_2}\cdots p_{k_r}^{\alpha_r}$ where $k_1 < k_2 < \cdots < k_r$ and where $\alpha_1, \alpha_2, \ldots, \alpha_r$ are positive integers. Observing that $(1-p_k^{-z})^{-1} = \sum_{\ell=0}^{\infty} p_k^{-\ell z}$, try to get a feel for what's going on by checking what happens when one forms the products $(1-2^{-z})^{-1}(1-3^{-z})^{-1}$ and $(1-2^{-z})^{-1}(1-3^{-z})^{-1}(1-5^{-z})^{-1}$.)

4.26. Show that the infinite product $\prod_{n=1}^{\infty}\exp(n^{-z}\operatorname{Log} z)$ converges normally in the half-plane $U = \{z : \operatorname{Re} z > 1\}$, and identify the function that this product defines in U.

4.27. Let $\langle z_n\rangle$ be a sequence of non-zero complex numbers in the disk $\Delta = \Delta(0,1)$ with the property that the series $\sum_{n=1}^{\infty}(1-|z_n|)$ converges.

Verify that the formula

$$f(z) = cz^p \prod_{n=1}^{\infty} \frac{\bar{z}_n}{|z_n|} \left(\frac{z_n - z}{1 - \bar{z}_n z} \right) ,$$

where p is a non-negative integer and c is a constant of unit modulus, defines an analytic function in Δ, one that satisfies $|f(z)| \leq 1$ for every z in Δ. Where are the zeros of f and what are their multiplicities?

4.28. Establish the Weierstrass product theorem for an open set U in $\widehat{\mathbb{C}}$ — $U \neq \widehat{\mathbb{C}}$ — that contains the point ∞.

4.29. Let $E = \{z_1, z_2, z_3, \cdots\}$ be a discrete subset of an open set U in the complex plane, let $\langle w_n \rangle$ be an arbitrary sequence of complex numbers, and let $\langle m_n \rangle$ be an arbitrary sequence of positive integers. Confirm the existence of an analytic function $f: U \to \mathbb{C}$ that for $n = 1, 2, 3, \cdots$ takes the value w_n with multiplicity m_n at z_n.

4.30. Let D be a domain in the complex plane. Construct a discrete subset E of D with the property that every point of ∂D is a limit point of E. Use E to prove: there exists an analytic function $f: D \to \mathbb{C}$ that admits no extension to an analytic function $F: \widetilde{D} \to \mathbb{C}$ with $\widetilde{D}$ a domain that properly contains D.

4.31. Find an explicit Weierstrass product expansion for the entire function f: (i) $f(z) = \cos z$; (ii) $f(z) = e^z - 1$; (iii) $f(z) = e^{a\pi z} - e^{b\pi z}$; (iv) $f(z) = \sqrt{z}\sin(\pi\sqrt{z})$.

4.32. With the aid of the functional equation (10.30) confirm that $\operatorname{Res}[-n, \Gamma(z)] = (-1)^n (n!)^{-1}$ for $n = 0, 1, 2, \cdots$.

4.33. Obtain the representation

$$\Gamma(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!(z+n)} + \int_1^{\infty} t^{z-1} e^{-t}\, dt$$

for every z in $\mathbb{C} \sim \{0, -1, -2, \cdots\}$.

4.34. The logarithmic derivative $\Psi = \Gamma'/\Gamma$ of the gamma function is known as the *digamma function* or, alternatively, as the *Gauss psi-function.* Derive the identity $\Psi'(z) = \sum_{n=0}^{\infty} (z+n)^{-2}$. Also, show that Ψ satisfies the functional equations

$$\Psi(1+z) - \Psi(z) = z^{-1} \quad , \quad \Psi(1-z) - \Psi(z) = \pi \cot(\pi z) .$$

4.35. Define f by $f(z) = \Psi(z) + \Psi[z + (1/2)] - 2\Psi(2z)$ for z in the domain $D = \mathbb{C} \sim \{0, -1/2, -1, -3/2, \cdots\}$. By calculating f', show that f is constant in D. Remembering that Ψ is the logarithmic derivative of Γ, argue from the constancy of f to the fact that $\Gamma(z)\Gamma[z+(1/2)] = e^{az+b}\Gamma(2z)$ for every z in D, where a and b are constants. By comparing the values of both sides of this identity when $z = 1/2$ and $z = 1$, determine a and b. In this way arrive at the so-called "duplication formula" of Legendre:

$$\Gamma(2z) = \pi^{-1/2} 2^{2z-1} \Gamma(z)\Gamma[z + (1/2)] \ .$$

4.3 Exercises for Section X.3

4.36. Let $P(z, w) = a_0(z) + a_1(z)w + \cdots + a_n(z)w^n$ be a polynomial in w whose coefficients $a_0(z), a_1(z), \ldots, a_n(z)$ are entire functions of z. Given that an analytic function element (f_0, D_0) has $P[z, f_0(z)] = 0$ for every z in D_0 and that a second function element (f, D) is an analytic continuation of (f_0, D_0), confirm that $P[z, f(z)] = 0$ for every z in D.

4.37. If an analytic function element (f_0, D_0) has the property that $w = f_0(z)$ furnishes a solution in D_0 to the differential equation

$$c_n(z)\frac{d^n w}{dz^n} + c_{n-1}(z)\frac{d^{n-1} w}{dz^{n-1}} + \cdots + c_1(z)\frac{dw}{dz} + c_0(z)w = 0 \ ,$$

where $c_0(z), c_1(z), \ldots, c_n(z)$ are entire functions of z, and if (f, D) is an analytic continuation of (f_0, D_0), then $w = f(z)$ gives a solution to the same equation in D. Prove this.

4.38. Exhibit the analytic continuation $\{g_t\}$ of g, the analytic germ generated at the point 1 by the function $f(z) = \operatorname{Log} z$, along the path γ given by $\gamma(t) = e^{it}$ for $0 \le t \le 2\pi$. How is the terminal germ $\tilde{g}$ of this continuation related to g? What terminal germ is obtained when g is continued along the path $\gamma + \gamma$? What about $-\gamma$? Or a general path $\alpha: [a, b] \to \mathbb{C} \sim \{0\}$ for which $\alpha(a) = \alpha(b) = 1$?

4.39. Let g be the analytic germ determined at the point 1 by the function $f(z) = z^{2/3}$. Find the analytic continuation $\{g_t\}$ of g along the path γ defined on $[0, 2\pi]$ by $\gamma(t) = e^{it}$. In what relation do g and the terminal germ $\tilde{g}$ of this continuation stand?

4.40. It is known that an analytic germ g at a point z_0 of a domain D can be continued analytically along all closed paths γ in D starting — hence, terminating — at z_0. Assuming that such a path γ is homotopic in D, with fixed endpoints, to a constant path at z_0, what is the end-product of the continuation of g along γ?

4.41. Determine all exceptional points z for the polynomial $P(z, w)$:
(i) $P(z, w) = w^3 - z^2 + z$; (ii) $P(z, w) = z^2 w^2 + (2z + 2)w + 1$;
(iii) $P(z, w) = 2zw^3 + 3zw^2 + 1$.

4.42. Let $P(z, w) = w^3 - z^2$, a polynomial whose only exceptional point is at the origin, and let $D = \mathbb{C} \sim [0, \infty)$. According to Theorem 3.7 there is an analytic function $f: D \to \mathbb{C}$ such that $f(-1) = 1$ and $P[z, f(z)] = 0$

for every z in D. Find a formula for f.

4.43. Show that the only exceptional points for the polynomial $P(z,w) = w^3 - z^2w^2 + (4/27)$ are 1 and -1. By Theorem 3.7 there is a unique analytic function $f: \Delta(0,1) \to \mathbb{C}$ such that $f(0) = -\sqrt[3]{4}/3$ and $P[z, f(z)] = 0$ for every z in $\Delta(0,1)$. Verify that f is a zero-free, even function which is real-valued on the interval $(-1,1)$, and check that the only point z of $\Delta(0,1)$ for which $f'(z) = 0$ is $z = 0$. What is the image of the interval $(-1,1)$ under f?

4.44. Consider a linear differential equation

$$c_n(z)\frac{d^n w}{dz^n} + c_{n-1}(z)\frac{d^{n-1}w}{dz^{n-1}} + \cdots + c_0(z)w = 0 \ , \tag{10.45}$$

say with entire functions for coefficients, in a simply connected domain D where the leading coefficient c_n is zero-free. Prove that for any given point z_0 of D and any preassigned complex numbers $w_0, w_1, \ldots, w_{n-1}$ there exists a unique analytic function $f: D \to \mathbb{C}$ such that $w = f(z)$ solves (10.45) in D and such that $f(z_0) = w_0$, $f'(z_0) = w_1, \ldots, f^{(n-1)}(z_0) = w_{n-1}$. (*Hint.* Apply Theorem 3.4. Feel free to use the following basic existence and uniqueness theorem from the theory of ordinary differential equations: if ζ_0 is a point of the complex plane for which $c_n(\zeta_0) \neq 0$ and if $\omega_0, \omega_1, \ldots, \omega_{n-1}$ are arbitrary complex numbers, then there is a unique analytic germ g at ζ_0 such that $w = g(z)$ satisfies (10.45) throughout the domain-set of g and such that $g(\zeta_0) = \omega_0$, $g'(\zeta_0) = \omega_1, \ldots, g^{(n-1)}(\zeta_0) = \omega_{n-1}$.)

4.45. Demonstrate that $\mathcal{F}$, the family of all analytic function elements (f, D) with the property that $[f(z)]^2 = z$ for every z in D, is a global analytic function.

4.46. Let $P(z,w) = z^2w^2 - (z^2+1)w + 1$. Show that the collection of analytic function elements (f, D) with the property that $P[z, f(z)] = 0$ for every z in D is not a global analytic function by producing two such elements that are not analytic continuations of each other. What does this tell you about P?

4.47. Let $\Delta_1 = \Delta(p_1, r_1)$ and $\Delta_2 = \Delta(p_2, r_2)$ be open disks on the Riemann surface $S(\mathcal{F})$ of a global analytic function $\mathcal{F}$, and let p be a point of $\Delta_1 \cap \Delta_2$. Demonstrate that $\Delta(p, r)$ is contained in $\Delta_1 \cap \Delta_2$ when $r > 0$ is suitably small.

4.48. Suppose that p_1 and p_2 are different points of the Riemann surface $S(\mathcal{F})$. Confirm that there is a radius $r > 0$ for which the disks $\Delta(p_1, r)$ and $\Delta(p_2, r)$ are disjoint.

4.49. Let U be an open set on $S(\mathcal{F})$, let $\Phi: U \to \mathbb{C}$ be a continuous function, and let $\Psi: U \to \mathbb{C}$ be a primitive for Φ in U, meaning that Ψ is holomorphic in U and has $\Psi'(p) = \Phi(p)$ for every p in U. Show that $\int_\Gamma \Phi(p)dp = \Psi[\Gamma(b)] - \Psi[\Gamma(a)]$ for any piecewise smooth path $\Gamma: [a,b] \to U$.

(*Hint.* First treat the case in which the trajectory of Γ lies in an open disk on $S(\mathcal{F})$ that is contained in U.)

Appendix A

Background on Fields

1 Fields

1.1 The Field Axioms

The definition of a field starts with a set F that is equipped with two binary operations. These operations are conventionally called *addition* and *multiplication* and are denoted by $+$ and $\cdot$. Thus, to each pair of elements a and b from F they assign a *sum* $a+b$ and a *product* $a \cdot b$ (usually abbreviated to ab), also belonging to the set F. The definition goes on to require that the following axioms be satisfied:

F.1 (Associative Laws): $(a+b)+c = a+(b+c)$ and $(ab)c = a(bc)$ for all a, b, and c in F;

F.2 (Commutative Laws): $a+b = b+a$ and $ab = ba$ for all a and b in F;

F.3 (Distributive Law): $a(b+c) = ab+ac$ for all a, b, and c in F;

F.4 (Existence of Identities): there exist a pair of distinct elements of F, by tradition designated 0 and 1, that are neutral elements for the operations of addition and multiplication, respectively — i.e., $a+0 = a$ and $a \cdot 1 = a$ for every a in F;

F.5 (Existence of Inverses): corresponding to each a in F there exists an element of F, ordinarily represented by $-a$, with the property that $a+(-a) = 0$; corresponding to each a in $F^* = F \sim \{0\}$ there exists an element of F, normally symbolized by a^{-1} or $1/a$, such that $aa^{-1} = 1$.

When the above requirements are met the set F endowed with the given operations is called a *field*. The axiom system is known as the set of *field axioms*. These axioms imply that the *additive identity* 0 and the *multiplicative identity* 1 are uniquely determined. The requirement $0 \neq 1$ prevents the set $\{0\}$ from trivially satisfying the field axioms, and so forces any field to contain at least two elements. (There is a field with exactly two elements!)

Also implicit in the axioms is the uniqueness of the *additive inverse* $-a$ of an arbitrary element a of F and the *multiplicative inverse* a^{-1} of a non-zero element a of F. *Subtraction* and *division* in F are defined in the expected manner: $a-b=a+(-b)$ for a and b in F; $a/b=a\cdot b^{-1}$ for a in F and b in F^*. The meaning of a^n for a in F and n an integer is also the anticipated one: $a^n=a\cdot a\cdots a$ (n factors) if n is positive, $a^{-n}=(a^{-1})^n$ if n is positive and $a\neq 0$, and $a^0=1$ if $a\neq 0$.

The three fields that have a direct bearing on this book are the fields $\mathbb{Q}$ of rational numbers, $\mathbb{R}$ of real numbers,and $\mathbb{C}$ of complex numbers. The general study of fields is an important component of present-day algebra.

1.2 Subfields

Let F be a field. If a subset E of F is closed with respect to the operations of F — meaning that $a+b$ and ab belong to E whenever a and b do — if the elements 0 and 1 from F are members of E, and if E is itself a field with respect to the operations it thus inherits from F, then E is termed a *subfield* of F. This definition insures that, if a is an element of E, then its additive inverse $-a$ in the field F also belongs to E and, naturally, furnishes the additive inverse of a with respect to the field structure of E. Likewise, the multiplicative inverse in the field E of any non-zero member a of E coincides with its inverse a^{-1} from F. The field $\mathbb{Q}$, for instance, is a subfield of $\mathbb{R}$. In Hamilton's construction of the complex number field $\mathbb{C}$ the set $E=\{(x,0): x\in\mathbb{R}\}$ forms a subfield of $\mathbb{C}$.

1.3 Isomorphic Fields

Consider a pair of fields F and F'. It may happen that there exists a function $\varphi: F\to F'$ which is one-one, which has range F', and which satisfies

$$\varphi(a+b)=\varphi(a)+\varphi(b) \quad , \quad \varphi(ab)=\varphi(a)\varphi(b)$$

for all a and b in F. (N.B. The addition and multiplication on the left-hand sides of these equations are performed in F; the operations on the right-hand sides are carried out in F'.) If so, we say that F and F' are *isomorphic fields* and that φ is an *isomorphism of F onto F'*. The function φ supplies a dictionary for passing from algebraic statements made in the context of F to their exact translations in the language of F'. The existence of such a function implies that there is nothing intrinsic to the field structure of F to distinguish it from F'. For all practical purposes F and F' are the same field! Referring again to Hamilton's construction of the complex numbers, we note that $\varphi(x)=(x,0)$ defines an isomorphism of $\mathbb{R}$ onto $E=\{(x,0): x\in\mathbb{R}\}$, a subfield of $\mathbb{C}$. It is this fact that we had in mind when in Chapter I we spoke of $\mathbb{R}$ and E as "structurally indistinguishable" fields.

2 Order in Fields

2.1 Ordered Fields

Roughly speaking, an "ordered field" is one in which there is established a sense of "positive" and "negative" that is compatible with the field structure. To make this precise, suppose that a field F has a subset P (for "positive elements") which exhibits the following two properties:

OF.1: $a + b$ and ab belong to P whenever both a and b do;

OF.2: P does not contain the element 0; if a is any non-zero element of F, then P contains either a or $-a$, but not both.

We then say that F is *ordered by* P. Indeed, we can define an order relation $>$ on F by declaring that $b > a$ if and only if $b - a$ lies in P. (N.B. The statement that $a < b$ is synonymous with $b > a$. The notation $b \geq a$ indicates that either $b > a$ or $b = a$.) Thus, asserting that $a > 0$ is equivalent to stating that a is a member of P. Using (OF.1) and (OF.2) one easily checks that the ordinary rules for operating with inequalities become valid in F. It is conceivable, of course, that there is more than one subset P of F satisfying (OF.1) and (OF.2). As a result, there may be many ways of turning a given field F into an *ordered field,* by which is meant a field in which an order has been specified through a choice of P. It is also possible that F has no subsets P enjoying these properties. When the latter situation occurs, F is called an *unorderable field.* In such a field one must live without an algebraically consistent notion of order.

The field $\mathbb{Q}$, when thought of as an ordered field, is typically assigned its *standard ordering*: a rational number $x = m/n$, where m and n are integers and $n \neq 0$, is declared positive if and only if mn belongs to $\mathbb{N}$, the set of natural numbers. The usual constructions of the real field $\mathbb{R}$ from $\mathbb{Q}$ show how to extend this standard ordering of $\mathbb{Q}$ to the *standard ordering* of $\mathbb{R}$. The field $\mathbb{C}$, on the other hand, is unorderable. To see this, suppose that P were a subset of $\mathbb{C}$ endowed with properties (OF.1) and (OF.2). Since $i \neq 0$, either i would belong to P or $-i$ would. In both cases (OF.1) would imply that $i^2 = (-i)^2 = -1$ is an element of P. Again by (OF.1), $1 = (-1)^2$ would have to be in P, placing both 1 and -1 there. This would violate condition (OF.2). The conclusion: no such P exists, so $\mathbb{C}$ cannot be ordered.

2.2 Complete Ordered Fields

Let F be an ordered field, say with order relation $>$, and let S be a non-empty subset of F. To say that S is *bounded above* means there exists an element b of F with the property that the statement $b \geq x$ is true for every x in S. Such an element b is called an *upper bound for S* in F. If S is bounded above, it may be the case — it does not happen automatically — that among all upper bounds for S in F there is a smallest one, call it u; i.e., u is an upper bound for S and $b > u$ holds for every other upper bound b. In these circumstances we refer to u as the *least upper bound* (or *supremum*) *of S in F*. We express this relationship between S and u by writing $u = \sup S$. In a similar fashion one can say what it means for S to be *bounded below* and for an element of F to be a *lower bound* for S. If S has a *greatest lower bound* (also called an *infimum*) ℓ in F, then we indicate this fact symbolically by means of $\ell = \inf S$. Take, for example, the set $S = \{x \in \mathbb{Q} : x^2 < 2\}$. If $F = \mathbb{Q}$ with its standard ordering, then S is bounded above but has no supremum in F. By contrast, if $F = \mathbb{R}$ with its standard ordering, then the same set S has $\sup S = \sqrt{2}$.

A field F ordered by $>$ is pronounced *complete* relative to the given ordering if every non-empty subset S of F that is bounded above has a supremum in F. This implies the dual property, the existence of an infimum in F for every non-empty subset of F that is bounded below. A fundamental statement about the real number system is:

Theorem 2.1. *The field $\mathbb{R}$ of real numbers with its standard ordering is a complete ordered field.*

Even more is true. The standardly ordered field $\mathbb{R}$ is up to an order-preserving isomorphism the only complete ordered field: if F is any complete ordered field, then there exists an isomorphism φ of $\mathbb{R}$ onto F with the added property that $\varphi(b) > \varphi(a)$ in the order in F whenever $b > a$ with respect to the standard ordering of $\mathbb{R}$.

2.3 Implications for Real Sequences

A sequence $\langle x_n \rangle$ of real numbers is said to be *monotone* if it is either non-decreasing ($x_1 \leq x_2 \leq x_3 \leq \cdots$) or non-increasing ($x_1 \geq x_2 \geq x_3 \geq \cdots$). Theorem 2.1 has the following implication for such sequences: *a monotone and bounded sequence $\langle x_n \rangle$ of real numbers is convergent.* In fact, it is a simple exercise to check that

$$\lim_{n\to\infty} x_n = \sup\{x_n : n = 1, 2, 3, \cdots\}$$

in the non-decreasing case, whereas

$$\lim_{n\to\infty} x_n = \inf\{x_n : n = 1, 2, 3, \cdots\}$$

in the non-increasing case.

Consider now an arbitrary bounded sequence $\langle x_n \rangle$ of real numbers. Let us assume that $|x_n| \leq c$ for all n. We define a new sequence $\langle y_n \rangle$ by setting

$$y_n = \sup\{x_k : k = n, n+1, n+2, \cdots\}$$

for $n \geq 1$. Then $-c \leq y_n \leq c$ and $y_1 \geq y_2 \geq y_3 \geq \cdots$. According to the above remarks, $M = \lim_{n\to\infty} y_n$ exists. This number M is known as the *limit superior* of the given sequence and is denoted by $\limsup_{n\to\infty} x_n$. To repeat: for a bounded sequence $\langle x_n \rangle$ of real numbers,

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup\{x_k : k = n, n+1, n+2, \cdots\} .$$

Similarly, the *limit inferior* of such a sequence, denoted $\liminf_{n\to\infty} x_n$, is defined by

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \inf\{x_k : k = n, n+1, n+2, \cdots\} .$$

The important observation to be made about $M = \limsup_{n\to\infty} x_n$ and $m = \liminf_{n\to\infty} x_n$ is that these numbers are accumulation points of $\langle x_n \rangle$. As a matter of fact, M is the largest accumulation point of $\langle x_n \rangle$ and m the smallest. Furthermore, $\langle x_n \rangle$ is convergent with limit ℓ if and only if $m = M = \ell$. The proofs of the preceding observations are straightforward exercises. As one important spin-off from this discussion we record:

Theorem 2.2. (Bolzano-Weierstrass Theorem) *Any bounded sequence of real numbers has at least one real accumulation point.*

Appendix B

Winding Numbers Revisited

1 Technical Facts About Winding Numbers

1.1 The Geometric Interpretation

We furnish here the rigorous argument that justifies the geometric interpretation of winding numbers presented in Chapter V. Let us first recall the situation there. Given a closed, piecewise smooth path $\gamma: [a, b] \to \mathbb{C}$ and a point z of $\mathbb{C} \sim |\gamma|$, we gave geometric meaning to the winding number $n(\gamma, z)$ as follows: we fixed a circle $K = K(z, r)$ centered at z; we defined an auxiliary path $\beta: [a, b] \to \mathbb{C}$, the radial projection of γ on K, by $\beta(t) = z + r\{[\gamma(t) - z]/|\gamma(t) - z|\}$ and denoted by $v(t)$ the radial vector from z to $\beta(t)$; we decided what it should mean for $v(t)$ to perform one complete revolution over a subinterval $[c, d]$ of $[a, b]$; we characterized each such revolution as either positive or negative; we then made an assertion that we now turn into a theorem.

Theorem 1.1. *The winding number $n(\gamma, z)$ gives the net number of complete revolutions — meaning the number of positive revolutions minus the number of negative revolutions — executed by the vector $v(t)$ as t increases from a to b.*

Proof. For convenience we take $r = 1$. (The choice of r affects only the length of $v(t)$, not the number of revolutions it performs.) With a view toward minimizing notational difficulties, we also assume that $z = 0$ and that $\beta(a) = \beta(b) = 1$. (N.B. The general case can be reduced to this special one by considering in place of γ the path $\gamma_1: [a, b] \to \mathbb{C}$ defined via $\gamma_1(t) = e^{i\varphi}[\gamma(t) - z]$, where $\varphi = -\operatorname{Arg}[\beta(a) - z]$. Then $n(\gamma, z) = n(\gamma_1, 0)$, as a simple calculation confirms, and the radial projection β_1 of γ_1 onto the circle $K_1 = K(0, 1)$ is given by $\beta_1(t) = e^{i\varphi}[\beta(t) - z]$. Thus, since we have chosen $r = 1$, $\beta_1(a) = \beta_1(b) = 1$. Finally, the revolutions that $v(t)$ performs exactly mirror those of its counterpart $v_1(t)$.) We may further suppose that β is a non-constant path, for the contents of the theorem are essentially trivial otherwise.

Because β is uniformly continuous on $[a,b]$, we can choose and fix $\delta > 0$ so as to insure that $|\beta(t) - \beta(s)| < 1$ holds whenever t and s are points of $[a,b]$ for which $|t-s| < \delta$. We remark next that the set $A = \{t \in [a,b] : \beta(t) \neq 1\}$ is a non-empty open set in the topology of the real line $\mathbb{R}$. As such, A is expressible as a disjoint union of open intervals, possibly an infinite number of them, and this breakdown of A into open intervals is unique. Consider an arbitrary component interval of A — call it (c,d). The definitions of A and (c,d) dictate that $\beta(t) \neq 1$ when $c < t < d$, whereas $\beta(c) = \beta(d) = 1$. In other words, $[c,d]$ is an interval over which $v(t)$ potentially executes a complete revolution. (Conversely, if $v(t)$ does perform a complete revolution over a subinterval $[c,d]$ of $[a,b]$, then (c,d) is easily seen to be one of the component intervals of A.) If the point -1 happens to belong to $\beta([c,d])$, then necessarily $d-c \geq \delta$. (If $d-c < \delta$, then the fact that $\beta(c) = 1$ would, by the choice of δ, force $\beta([c,d])$ to lie in the disk $\Delta(1,1)$, which does not contain -1.) It follows that the number of intervals (c,d) in the decomposition of A with the feature that -1 belongs to $\beta([c,d])$ cannot exceed $(b-a)/\delta$ — hence, is finite. Suppose that $(c_1,d_1),(c_2,d_2),\ldots,(c_p,d_p)$ is a complete list of the component intervals of A with this property, labeled so that $c_1 < d_1 \leq c_2 < d_2 \leq \cdots \leq c_p < d_p$. These intervals create a subdivision $a = d_0 \leq c_1 < d_1 \leq c_2 \cdots \leq c_p < d_p \leq c_{p+1} = b$ of the interval $[a,b]$. (Of course, if -1 is not in the trajectory of β to begin with, this becomes the trivial partition $a = d_0 < c_1 = b$.) For $0 \leq k \leq p$ we let β_k^* denote the restriction of β to the interval $[d_k, c_{k+1}]$ and for $1 \leq k \leq p$ let β_k designate the restriction of β to $[c_k, d_k]$. Then $\beta = \beta_0^* + \beta_1 + \beta_1^* + \cdots + \beta_p + \beta_p^*$. (In the not unlikely event that $d_k = c_{k+1}$ one can actually ignore the term β_k^* in this expression, for its deletion has absolutely no effect on the path-sum.) We use γ_k^* and γ_k to indicate the restrictions of γ to the corresponding intervals, so that $\gamma = \gamma_0^* + \gamma_1 + \gamma_1^* + \cdots + \gamma_p + \gamma_p^*$.

By construction the point -1 is not a member of $\beta([d_k, c_{k+1}])$, which fact implies that $\gamma([d_k, c_{k+1}])$ lies in $\mathbb{C} \sim (-\infty, 0]$, a domain where the function $f(\zeta) = \operatorname{Log}\zeta$ is analytic. As $\operatorname{Arg}[\gamma(d_k)] = \operatorname{Arg}[\gamma(c_{k+1})] = \operatorname{Arg} 1 = 0$, we conclude that for $0 \leq k \leq p$

$$\int_{\gamma_k^*} \frac{d\zeta}{\zeta} = \operatorname{Log}[\gamma(c_{k+1})] - \operatorname{Log}[\gamma(d_k)] = \operatorname{Log}|\gamma(c_{k+1})| - \operatorname{Log}|\gamma(d_k)| .$$

In particular,

$$\text{(B.1)} \qquad \operatorname{Im}\left(\int_{\gamma_k^*} \frac{d\zeta}{\zeta}\right) = 0$$

for $k = 0,1,\ldots,p$. Next, set $D = \mathbb{C} \sim [0,\infty)$ and let Ψ be the branch of the argument in D whose values lie in the interval $(0, 2\pi)$. Then $g(\zeta) = \operatorname{Log}|\zeta| + i\Psi(\zeta)$ defines a branch of the logarithm function in D. In Chapter V we defined revolutions of $v(t)$ in terms of the behavior of a certain reference angle $\theta(t)$. That angle is represented here by $\theta(t) = \Psi[\beta(t)] = \Psi[\gamma(t)]$, provided t belongs to any of the intervals (c_k, d_k). Writing $\ell(t) = g[\gamma(t)] = \operatorname{Log}|\gamma(t)| + i\,\theta(t)$ and exploiting the fact that $\gamma[(c_k, d_k)]$ is contained in D — hence, that the derivative of ℓ is given by $\dot{\ell}(t) = g'[\gamma(t)]\dot{\gamma}(t) = \dot{\gamma}(t)/\gamma(t)$ at every point of differentiability t which γ has in (c_k, d_k) — we compute for $k = 1, 2, \ldots, p$:

$$\int_{\gamma_k} \frac{d\zeta}{\zeta} = \int_{c_k}^{d_k} \frac{\dot{\gamma}(t)\,dt}{\gamma(t)} = \lim_{\substack{s\to d_k^- \\ u\to c_k^+}} \int_u^s \frac{\dot{\gamma}(t)\,dt}{\gamma(t)}$$

$$= \lim_{\substack{s\to d_k^- \\ u\to c_k^+}} \int_u^s \dot{\ell}(t)\,dt = \lim_{s\to d_k^-} \ell(s) - \lim_{u\to c_k^+} \ell(u)$$

$$= \operatorname{Log}|\gamma(d_k)| - \operatorname{Log}|\gamma(c_k)| + i\left[\lim_{s\to d_k^-} \theta(s) - \lim_{u\to c_k^+} \theta(u)\right] .$$

Recalling what it means for the vector $v(t)$ to perform a single revolution over $[c, d]$, we note that the imaginary part of the final expression is equal to 2π if $v(t)$ performs a positive revolution over $[c_k, d_k]$, -2π if $v(t)$ makes a negative revolution over $[c_k, d_k]$, and 0 if $\beta([c_k, d_k])$ is not the full circle $K(0, 1)$. These are the only three possibilities. Because $n(\gamma, 0)$ is a real number, we can combine this information with (B.1) to obtain

$$n(\gamma, 0) = \frac{1}{2\pi i}\int_\gamma \frac{d\zeta}{\zeta} = \operatorname{Re}\left(\frac{1}{2\pi i}\int_\gamma \frac{d\zeta}{\zeta}\right) = \frac{1}{2\pi}\operatorname{Im}\left(\int_\gamma \frac{d\zeta}{\zeta}\right)$$

$$= \frac{1}{2\pi}\operatorname{Im}\left(\int_{\gamma_0^*} \frac{d\zeta}{\zeta} + \int_{\gamma_1} \frac{d\zeta}{\zeta} + \cdots + \int_{\gamma_p} \frac{d\zeta}{\zeta} + \int_{\gamma_p^*} \frac{d\zeta}{\zeta}\right) = P - N ,$$

where P is the number of positive revolutions performed by $v(t)$ and N is the number of negative revolutions. ∎

Notice that Theorem 1.1 expresses $n(\gamma, z)$ in terms of quantities that make sense for an arbitrary closed path $\gamma: [a, b] \to \mathbb{C} \sim \{z\}$, even one that is not piecewise smooth. Thus, Theorem 1.1 can serve as a definition of $n(\gamma, z)$ in this more general setting. Most of the properties of winding numbers observed for piecewise smooth paths (e.g., those listed in Lemma V.2.1) remain valid in the generalized situation, although they require new confirmation.

1.2 Winding Numbers and Jordan Curves

We turn next to the proof of Lemma V.2.1(iii). We remind the reader that, in stating and proving this result, we accept the Jordan curve theorem as a given. Needless to say, a much shorter argument would be possible if we had

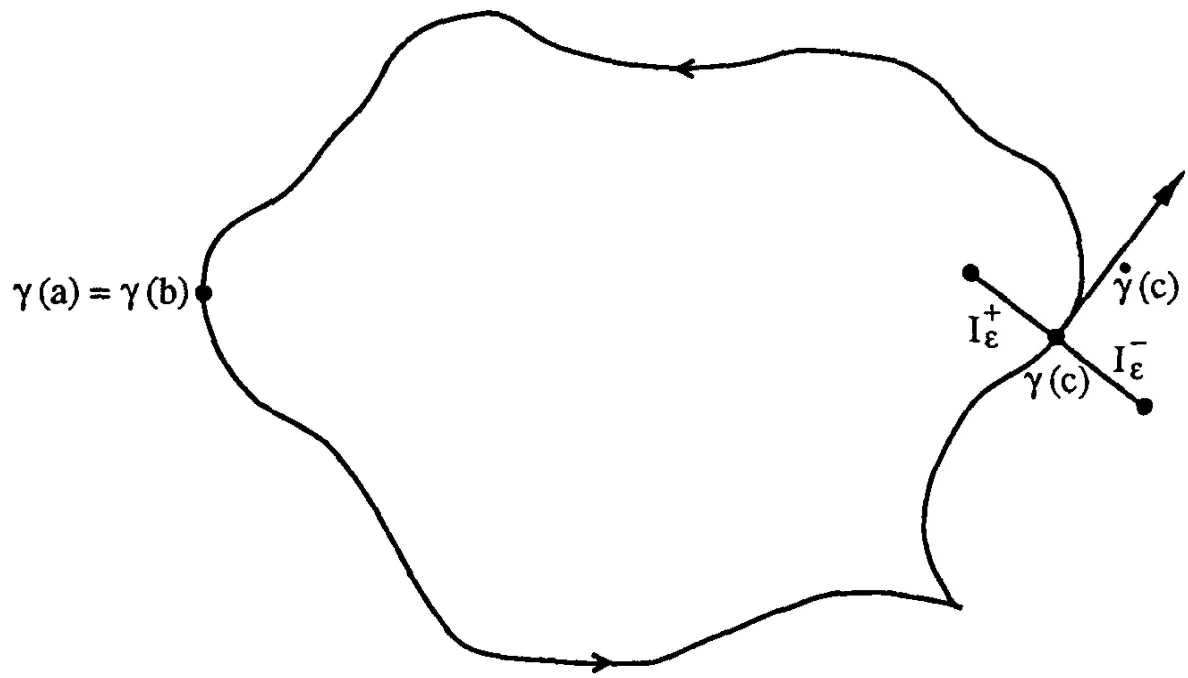


Figure 1.

access to more sophisticated topological machinery. The chief ingredient in the proof that we furnish is the ensuing lemma. (See Figure 1.)

Lemma 1.2. *Let $\gamma\colon [a,b] \to \mathbb{C}$ be a simple, closed path, and let c be a point of (a,b) at which γ is differentiable with $\dot\gamma(c) \neq 0$. There exists an $\epsilon > 0$ for which it is true that the sets $I_\epsilon^+ = \{\gamma(c) + si\dot\gamma(c) : 0 < s \le \epsilon\}$ and $I_\epsilon^- = \{\gamma(c) + si\dot\gamma(c) : -\epsilon \le s < 0\}$ lie in different components of $\mathbb{C} \sim |\gamma|$.*

Proof. Without loss of generality we may suppose that $\gamma(c) = 0$ and that $\dot\gamma(c)$ is both real and positive. (The proof can always be reduced to this case by subjecting γ to a preliminary translation and rotation.) With this normalization the sets under consideration are $I_\epsilon^+ = \{si\,\dot\gamma(c) : 0 < s \le \epsilon\}$ and $I_\epsilon^- = \{si\,\dot\gamma(c) : -\epsilon \le s < 0\}$, which are intervals on the positive and negative imaginary axis, respectively. We first note that there exists an $\epsilon > 0$ with the property that the set $I_\epsilon = \{si\,\dot\gamma(c) : -\epsilon \le s \le \epsilon\}$ meets the Jordan curve $J = |\gamma|$ only at the origin. If not, there would be a sequence of non-zero real numbers $\langle s_n \rangle$ such that $s_n \to 0$, but such that $s_n i\,\dot\gamma(c)$ lies on J. Thus, $s_n i\,\dot\gamma(c) = \gamma(t_n)$ for some t_n in $[a,b]$. After passing to a subsequence and relabeling, if necessary, one could assume that $t_n \to t_0$, a point of $[a,b]$. Now $\gamma(t_n) = s_n i\,\dot\gamma(c) \to 0$ by construction and $\gamma(t_n) \to \gamma(t_0)$ by the continuity of γ. Therefore $\gamma(t_0) = 0 = \gamma(c)$. From the simplicity of γ we would infer that $t_0 = c$. Because $\gamma(t_n) \neq 0$ for $n = 1, 2, 3, \cdots$, the simplicity of γ would also make sure that $t_n \neq c$ for such n. These facts would allow us to write

$$\dot\gamma(c) = \lim_{n\to\infty} \frac{\gamma(t_n) - \gamma(c)}{t_n - c} = \lim_{n\to\infty} \frac{\gamma(t_n)}{t_n - c} = \lim_{n\to\infty} \left[\frac{s_n i\,\dot\gamma(c)}{t_n - c}\right].$$

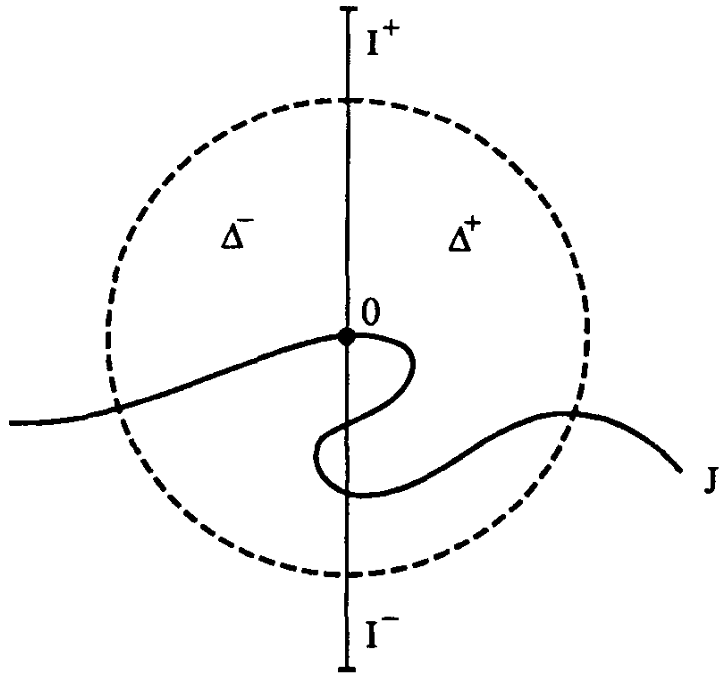


Figure 2.

By assumption $\dot{\gamma}(c) \neq 0$, so we would have $s_n/(t_n - c) \to -i$. But $s_n/(t_n - c)$ is real for all n, producing a contradiction. Accordingly, we are at liberty to choose and fix $\epsilon > 0$ for which $I_\epsilon \cap J = \{0\}$. We abbreviate $I = I_\epsilon$, $I^+ = I_\epsilon^+$, and $I^- = I_\epsilon^-$ for the rest of the proof. We claim that I^+ and I^- are contained in different components of $\mathbb{C} \sim J$. Since I^+ and I^- are connected sets and neither of them meets J, each of these sets does lie in a component of $\mathbb{C} \sim J$. We shall suppose that they are contained in the same component of $\mathbb{C} \sim J$ and derive another contradiction.

The critical step in obtaining a contradiction is to make a small observation. Let $\Delta = \Delta(0, r)$, where $0 < r < \epsilon\dot{\gamma}(c)$. (Don't forget: $\dot{\gamma}(c) > 0$.) If $\Delta^+ = \{z \in \Delta : \operatorname{Re} z > 0\}$ and $\Delta^- = \{z \in \Delta : \operatorname{Re} z < 0\}$, then we observe that J must have points in both Δ^+ and Δ^-. (See Figure 2.) Indeed, because γ is differentiable at c and because $\gamma(c) = 0$, we can express $\gamma(t)$ in the form

$$\gamma(t) = \dot{\gamma}(c)(t - c) + E(t) \ ,$$

where the function $E: [a, b] \to \mathbb{C}$ satisfies $\lim_{t \to c} |E(t)|/|t - c| = 0$. If $t \neq c$ is near enough to c that $|E(t)| < \dot{\gamma}(c)|t - c|$, then $\dot{\gamma}(c)(t - c)$ and the real part of $\gamma(t)$ must have the same sign. Since $\dot{\gamma}(c) > 0$, we conclude that $\operatorname{Re}[\gamma(t)] > 0$ must hold for all t near c with $t > c$, whereas $\operatorname{Re}[\gamma(t)] < 0$ for all t near c with $t < c$. This fact, together with the continuity of γ, implies that J meets both Δ^+ and Δ^-.

Assume now that I^+ and I^- lie in the same component of the set $\mathbb{C} \sim J$ — call that component D. An elementary argument shows that there is a simple polygonal path β in D with initial point $\epsilon i\, \dot{\gamma}(c)$, with terminal point $-\epsilon i\, \dot{\gamma}(c)$, and with $|\beta|$ otherwise disjoint from I. It follows that the path $\alpha = \beta + [-\epsilon i\, \dot{\gamma}(c), \epsilon i\, \dot{\gamma}(c)]$ is simple and closed, and that the

Jordan curve $J_1 = |\alpha| = |\beta| \cup I$ meets J only at the origin. We next choose r with $0 < r < \epsilon\dot{\gamma}(c)$ and with the property that $\Delta = \Delta(0, r)$ does not intersect $|\beta|$. Employing the notation introduced in the last paragraph, we conclude that $\Delta \sim J_1 = \Delta \sim I = \Delta^+ \cup \Delta^-$. The Jordan curve theorem applied to J_1 demands that Δ^+ and Δ^- be subsets of different components of $\mathbb{C} \sim J_1$. (Otherwise the disk Δ would intersect only one component of $\mathbb{C} \sim J_1$, which would prevent the origin from being a boundary point of its second component, as required by the Jordan curve theorem.) The set $J \sim \{0\}$ is connected and does not meet J_1. This means that $J \sim \{0\}$ has to be contained in a single component of $\mathbb{C} \sim J_1$. As a consequence, J must fail to intersect one of the sets Δ^- or Δ^+. We have thereby arrived at the anticipated contradiction. The alternative is for I^- and I^+ to be contained in different components of $\mathbb{C} \sim J$. ∎

After this preparation, we are ready for the proof of:

Theorem 1.3. *Let $\gamma: [a, b] \to \mathbb{C}$ be a simple, closed, piecewise smooth path and let D be the bounded component of $\mathbb{C} \sim |\gamma|$. Then either $n(\gamma, z) = 1$ for every z in D or $n(\gamma, z) = -1$ for all such z.*

Proof. In light of Lemma V.2.1(i) it suffices to exhibit just one point z_0 of D for which $n(\gamma, z_0) = 1$ or $n(\gamma, z_0) = -1$. To accomplish this we begin by choosing c in (a, b) with the property that $\dot{\gamma}(c)$ exists and is not zero. As γ is piecewise smooth and non-constant, the existence of such a number c is evident. The previous lemma permits us to select $\epsilon > 0$ so that the sets I_ϵ^+ and I_ϵ^- described there lie in different components of $\mathbb{C} \sim |\gamma|$. In order to simplify notation in this proof we shall assume that $\gamma(c) = 0$, that $\text{Arg}[\dot{\gamma}(c)] = \pi/2$, and that $\epsilon = 1$. One can always reduce the proof to this situation through preliminary translation, rotation, and dilation of γ. With these normalizations $I_\epsilon^+ = [-1, 0)$ and $I_\epsilon^- = (0, 1]$. We proceed under the assumption that $[-1, 0)$ is contained in D and $(0, 1]$ is contained in D^*, the unbounded component of $\mathbb{C} \sim |\gamma|$. We shall locate a point z_0 of D for which $n(\gamma, z_0) = 1$. In the opposite case (i.e., when $(0, 1]$ is contained in D and $[-1, 0)$ in D^*) a similar argument would identify a point z_0 of D with $n(\gamma, z_0) = -1$.

To aid in the proof we select an open disk $\Delta = \Delta(0, R)$ of radius $R > 1$ that contains the bounded set $|\gamma|$ — hence, that encompasses the domain D — and apply Theorem II.3.6 to choose a polygonal arc A in D^* whose endpoints are 1 and R. As a final preparatory step we fix r in $(0, 1)$ for which the closed disk $\overline{\Delta}(-r, r)$ is contained in Δ and is disjoint from A. (See Figure 3.)

By construction, the point $z_0 = -r$ belongs to D. We claim that $n(\gamma, z_0) = 1$. For the proof of this let γ_1 and γ_3 designate the restrictions of γ to the intervals $[a, c]$ and $[c, b]$, respectively, and let $\gamma_2(t) = -r + re^{it}$ for $0 \le t \le 2\pi$. Then $\beta = \gamma_1 - \gamma_2 + \gamma_3$ is a closed and piecewise smooth

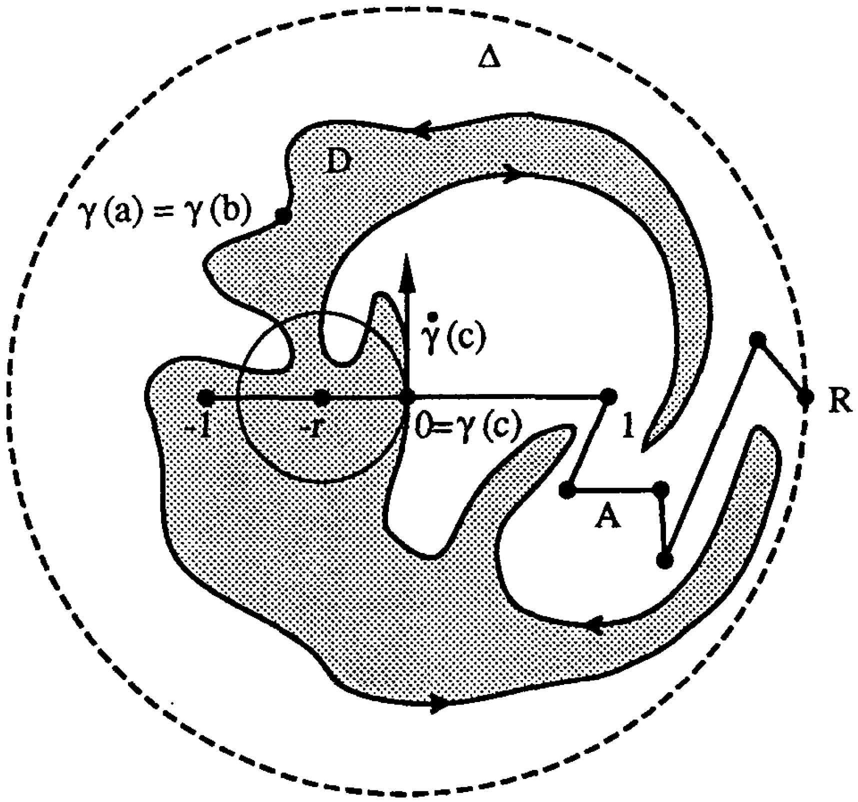


Figure 3.

path. We notice that

$$n(\beta, z_0) = \frac{1}{2\pi i}\int_{\gamma_1}\frac{d\zeta}{\zeta - z_0} - \frac{1}{2\pi i}\int_{\gamma_2}\frac{d\zeta}{\zeta - z_0} + \frac{1}{2\pi i}\int_{\gamma_3}\frac{d\zeta}{\zeta - z_0}$$

$$= \frac{1}{2\pi i}\int_{\gamma}\frac{d\zeta}{\zeta - z_0} - \frac{1}{2\pi i}\int_{\gamma_2}\frac{d\zeta}{\zeta - z_0} = n(\gamma, z_0) - n(\gamma_2, z_0)\ ,$$

while direct calculation yields $n(\gamma_2, z_0) = 1$. The path β is parametrized on the interval $[a, b + 2\pi]$, with $\beta(t) = 0$ only for $t = c$ and $t = c + 2\pi$. Since $\gamma(c) = 0$ and $\mathrm{Arg}[\dot\gamma(c)] = \pi/2$, it follows that $\mathrm{Im}[\gamma(t)] < 0$ for all t sufficiently close to c with $t < c$, whereas $\mathrm{Im}[\gamma(t)] > 0$ for all t near to c with $t > c$. (Recall the verification of a similar fact in the proof of Lemma 1.2.) Also, $\mathrm{Im}[\gamma_2(t)] > 0$ when $0 < t < \pi$ and $\mathrm{Im}[\gamma_2(t)] < 0$ when $\pi < t < 2\pi$. These facts permit us to select intervals $[t_1, s_1]$ and $[t_2, s_2]$ with $t_1 < c < s_1 < t_2 < c + 2\pi < s_2$ concerning which the following statements are valid: $\mathrm{Im}[\beta(t_1)] < 0$ and $\mathrm{Im}[\beta(s_1)] < 0$; $\mathrm{Im}[\beta(t_2)] > 0$ and $\mathrm{Im}[\beta(s_2)] > 0$; $\beta([t_1, s_1])$ and $\beta([t_2, s_2])$ are subsets of a small disk $\Delta_0 = \Delta(0, r_0)$, so chosen that Δ_0 does not contain the point z_0, does not intersect A, and is itself a subset of Δ.

To complete the proof, we consider a closed, piecewise smooth path α that is constructed by modifying β on each of the intervals $[t_1, s_1]$ and $[t_2, s_2]$. Namely, the restriction of α to $[t_k, s_k]$ $(k = 1, 2)$ is just some smooth parameterization on this interval of the directed line segment from $\beta(t_k)$ to

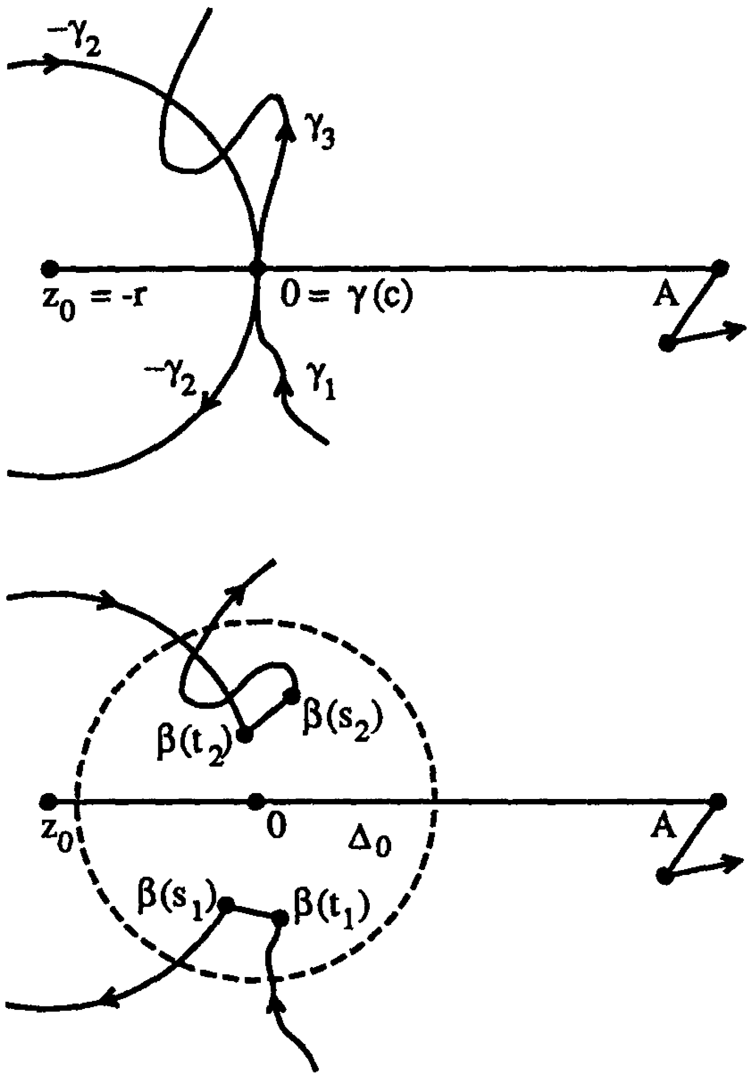


Figure 4.

$\beta(s_k)$. Otherwise, α agrees with β. (See Figure 4.) If β_k and α_k designate the restrictions of β and α to $[t_k, s_k]$ we see that

$$\int_\beta \frac{d\zeta}{\zeta - z_0} - \int_\alpha \frac{d\zeta}{\zeta - z_0} = \int_{\beta_1 - \alpha_1} \frac{d\zeta}{\zeta - z_0} + \int_{\beta_2 - \alpha_2} \frac{d\zeta}{\zeta - z_0} = 0 \ .$$

The last assertion is a consequence of the definition of α and of the local Cauchy theorem applied to the function $f(\zeta) = (\zeta - z_0)^{-1}$ and the paths $\beta_k - \alpha_k (k = 1, 2)$ in the disk Δ_0.

We infer that $n(\beta, z_0) = n(\alpha, z_0)$. On the other hand, since the whole construction of α was engineered to put z_0 in the unbounded component of $\mathbb{C} \sim |\alpha|$, Lemma V.2.1(ii) allows us to conclude that $n(\alpha, z_0) = 0$. Consequently,

$$n(\gamma, z_0) = n(\gamma_2, z_0) + n(\beta, z_0) = 1 + n(\alpha, z_0) = 1 \ ,$$

as asserted. ∎

Index

(continued from page ii)

Malitz: Introduction to Mathematical Logic.
Marsden/Weinstein: Calculus I, II, III. Second edition.
Martin: The Foundations of Geometry and the Non-Euclidean Plane.
Martin: Transformation Geometry: An Introduction to Symmetry.
Millman/Parker: Geometry: A Metric Approach with Models. Second edition.
Moschovakis: Notes on Set Theory.
Owen: A First Course in the Mathematical Foundations of Thermodynamics.
Palka: An Introduction to Complex Function Theory.
Pedrick: A First Course in Analysis.
Peressini/Sullivan/Uhl: The Mathematics of Nonlinear Programming.
Prenowitz/Jantosciak: Join Geometries.
Priestley: Calculus: An Historical Approach.
Protter/Morrey: A First Course in Real Analysis. Second edition.
Protter/Morrey: Intermediate Calculus. Second edition.
Ross: Elementary Analysis: The Theory of Calculus.
Samuel: Projective Geometry. *Readings in Mathematics.*
Scharlau/Opolka: From Fermat to Minkowski.
Sigler: Algebra.
Silverman/Tate: Rational Points on Elliptic Curves.
Simmonds: A Brief on Tensor Analysis. Second edition.
Singer/Thorpe: Lecture Notes on Elementary Topology and Geometry.
Smith: Linear Algebra. Second edition.
Smith: Primer of Modern Analysis. Second edition.
Stanton/White: Constructive Combinatorics.
Stillwell: Elements of Algebra: Geometry, Numbers, Equations.
Stillwell: Mathematics and Its History.
Strayer: Linear Programming and Its Applications.
Thorpe: Elementary Topics in Differential Geometry.
Troutman: Variational Calculus and Optimal Control with Elementary Convexity. Second edition.
Valenza: Linear Algebra: An Introduction to Abstract Mathematics.
Whyburn/Duda: Dynamic Topology.
Wilson: Much Ado About Calculus.

Printed in the United Kingdom
by Lightning Source UK Ltd.
121307UK00006B/28/A

9 780387 974279

This book provides a rigorous yet elementary introduction to the theory of analytic functions of a single complex variable. The book is aimed primarily at advanced undergraduate and beginning graduate students in mathematics. While presupposing in its readership a degree of mathematical maturity, it insists on no formal prerequisites beyond a sound knowledge of calculus. Thus, starting with basic definitions, the text slowly and carefully develops the ideas of complex analysis to the point where such landmarks of the subject as Cauchy's theorem, the Riemann mapping theorem, and the theorem of Mittag-Leffler can be treated without sidestepping any issues of rigor. The emphasis throughout is a geometric one, most pronounced in an extensive chapter on conformal mapping which amounts essentially to a "short course" in that important area of complex function theory. Each chapter concludes with a wide selection of exercises, ranging from straightforward computations to problems of a more conceptual and thought-provoking nature.

ISBN 0-387-97427-X
90000
9 780387 974279