# Class Notes
# Math 618: Complex Variables II
# Spring 2016

Harold P. Boas

updated April 29, 2016

## Contents

# 1 Introduction

There are three general themes for this semester:

- convergence and approximation of holomorphic (and harmonic) functions,
- conformal mapping, and
- the range of holomorphic functions.

The first item includes infinite products, the Weierstrass factorization theorem, Mittag-Leffler's theorem, normal families, and Runge's approximation theorem. The second item includes the Riemann mapping theorem. The third item includes Picard's theorems. The emphasis is on techniques that are constructive, at least in principle.

# 2 Normal families and the Riemann mapping theorem

The Riemann mapping theorem says that every simply connected planar region (other than the whole plane $\mathbb{C}$ and the empty set) is conformally equivalent to the open unit disk. In other words, there exists a bijective analytic function mapping the region onto the unit disk. [A satisfactory working definition of "simply connected" in the setting of open subsets of $\mathbb{C}$ is a connected open set whose complement with respect to the extended plane $\mathbb{C}_\infty$ is connected.]

The inverse function is automatically analytic by Corollary 7.6 to the open mapping theorem in §IV.7 of the textbook. Thus the Riemann mapping function could be called "bianalytic," but a more common terminology is "biholomorphic" (since "holomorphic function" is a standard synonym for "analytic function"). The Riemann mapping function can also be called a "conformal mapping," since injective (or even locally injective) analytic functions are conformal (by Exercise 4 in §IV.7 and Theorem 3.4 in §III.3).

The word "conformal" is used ambiguously in the literature. Some authors assume that conformal mappings are injective, but other authors allow conformal mappings to be locally injective but not globally injective. The exponential function $e^z$, for example, preserves angles between intersecting curves but is not globally injective. The term "biholomorphic mapping" is unambiguous: such a mapping is both bijective and holomorphic. Another reason to prefer the term "biholomorphic mapping" is that this concept generalizes usefully to higher dimension, but "conformal mapping" does not. You can change angles in $\mathbb{C}^2$ even with a complex-linear transformation, so biholomorphic mappings in $\mathbb{C}^2$ typically are not conformal. In the early days of multidimensional complex analysis, biholomorphic mappings were called "pseudoconformal," but that terminology was subsequently discarded.

The Riemann mapping theorem can be viewed as saying that the nonvoid simply connected planar regions consist of precisely two biholomorphic equivalence classes: one class is a singleton consisting of the whole plane, and the other equivalence class contains all other simply connected planar regions. Accordingly, function theory on simply connected regions bifurcates into two subtheories: (1) the theory of entire functions and (2) the theory of analytic functions on the unit disk. Both of these theories will be addressed this semester.

For multiply connected planar regions, on the other hand, there are infinitely many different biholomorphic equivalence classes. Two annuli $\{ z \in \mathbb{C} : r_1 < |z| < R_1 \}$ and $\{ z \in \mathbb{C} : r_2 < |z| < R_2 \}$ are biholomorphically equivalent if and only if the ratios $R_1/r_1$ and $R_2/r_2$ are equal to each other (in which case the first annulus can be mapped onto the second by a dilation; see Exercise 8 in §VII.4). In higher dimension, the situation is vastly more complicated than in $\mathbb{C}^1$: even for simply connected domains in $\mathbb{C}^2$, there are infinitely many distinct biholomorphic equivalence classes.

## 2.1 Outline of the proof

The standard modern proof of the Riemann mapping theorem consists of three steps.

1. Formulate a suitable extremal problem in the space of analytic functions mapping the given simply connected region into (but not necessarily onto) the unit disk.

2. Show that an extremal function exists.

3. Show that the extremal function is surjective (because if not, a new function could be constructed that contradicts the extremality).

The second step has the flavor of finding a limit point. A natural tool for implementing this step is a metric on the space of analytic functions on an open set. Developing that tool is the next topic.

## 2.2 A metric on analytic functions

A convenient starting point is a metric on the space $C(K)$ of continuous complex-valued functions defined on a compact subset $K$ of $\mathbb{C}$. There is a standard norm on the space $C(K)$, the supremum norm:

$$\|f\|_K = \max\{\, |f(z)| \,:\, z \in K \,\}.$$

The norm induces a metric (the distance between $f$ and $g$ is $\|f - g\|_K$), and convergence with respect to this metric is *uniform convergence* (so the metric is called the *uniform metric*). To say that $f_n \to f$ uniformly on $K$ is precisely the statement that $\|f_n - f\|_K \to 0$.

You know from real analysis that in the metric space $\mathbb{C}$ (and more generally in Euclidean space of arbitrary dimension), the compact sets are precisely the sets that are simultaneously closed and bounded (the Heine–Borel theorem, named after the German mathematician Eduard Heine [1821–1881] and the French mathematician Émile Borel [1871–1956]).[1]

The analogous equivalence is not valid in the space $C(K)$. Indeed, if $K$ is the closed unit disk, then the sequence $\{z^n\}_{n=1}^{\infty}$ of monomials has no subsequence converging to a continuous function on $K$, for the sequence converges pointwise to 0 on the open unit disk but is constantly equal to 1 at the point 1. Thus the sequence is not compact as a subset of $C(K)$, although the sequence is bounded (being a subset of the $C(K)$ closed unit ball) and is trivially closed (since the sequence has no limit point in the space).

You may know a generalization of the Heine–Borel theorem to general metric spaces: namely, a subset of a metric space is compact if and only if the set is simultaneously complete (Cauchy

[1] Apparently, this theorem was in the air in the second half of the nineteenth century, Heine being only one of several mathematicians who used the idea; Borel seems to have made the first explicit statement.

sequences converge) and totally bounded (the set can be covered by a finite number of arbitrarily small balls).

The characterization of compact subsets of the metric space $C(K)$ in function-theoretic terms is a famous proposition from the late nineteenth century. The theorem uses the following two notions. A set $S$ of functions on $K$ is called pointwise bounded if for each point $z$ in $K$ there exists a constant $M$ such that $|f(z)| \leq M$ for every function $f$ in $S$ (the value of $M$ is allowed to depend on the point $z$ but not on the function $f$). A set $S$ of functions on $K$ is called equicontinuous if for every point $z$ in $K$ and for every positive $\varepsilon$ there is a positive $\delta$ such that $|f(z) - f(w)| < \varepsilon$ whenever $f \in S$ and $|z - w| < \delta$ (the value of $\delta$ possibly depending on the point $z$ but not depending on the function $f$).

**Theorem** (Arzelà–Ascoli theorem)**.** *A subset of $C(K)$ is compact if and only if the subset is simultaneously closed, pointwise bounded, and equicontinuous.*

*Exercise.* On a compact set, equicontinuity at every point is equivalent to uniform equicontinuity: the value of $\delta$ actually can be taken to be independent both of the point $z$ and of the function $f$. (The proof is analogous to the proof that a continuous function on a compact set is automatically uniformly continuous.)

Although pointwise boundedness on a compact set is not equivalent to uniform boundedness (think of a sequence of triangle functions with increasingly steep peaks condensing at the origin), the proof of the theorem yields that in the presence of equicontinuity, pointwise boundedness does imply uniform boundedness on compact sets.

The theorem is due to the Italian mathematician Giulio Ascoli (1843–1896) in a paper of 1884 in which he introduced the notion of equicontinuity. It seems that Cesare Arzelà (1847–1912) actually published the idea of equicontinuity a year or so earlier than Ascoli did. Subsequently, in 1889 and in 1896, Arzelà (notice the grave accent) clarified, extended, and applied the theorem. So technically it may be Ascoli's theorem, but Arzelà's work popularized the theorem, and Arzelà even had the key concept earlier.

*Proof of the Arzelà–Ascoli theorem.* In a metric space, compactness is the same as sequential compactness. Accordingly, what needs to be shown for the sufficiency of the conditions is that if $\{f_n\}$ is a pointwise bounded, equicontinuous sequence in $C(K)$, then there is a subsequence that converges uniformly on $K$ (necessarily to an element of $C(K)$, since the uniform limit of continuous functions is continuous). An equivalent statement is that there is a subsequence for which Cauchy's criterion for convergence holds uniformly.

Take a dense sequence $\{z_n\}$ in $K$. (To construct the sequence, cover the set $K$ with a mesh of closed squares with sides of length $1/k$, pick a point of $K$ in each cell that intersects $K$, increase $k$, and iterate to produce the dense sequence. For a nice set $K$, say the closure of an open set, you could take the points of $K$ having both coordinates rational.)

The sequence of complex numbers $\{f_n(z_1)\}$ is bounded (by one of the hypotheses), so the Bolzano–Weierstrass theorem provides an initial increasing sequence $\{j_1(n)\}$ of natural numbers such that the sequence $\{f_{j_1(n)}(z_1)\}$ converges. There is a subsequence $\{j_2(n)\}$ such that $\{f_{j_2(n)}(z_2)\}$

converges (and $\{f_{j_2(n)}(z_1)\}$, being a subsequence of $\{f_{j_1(n)}(z_1)\}$, converges too). Iterate this process. The diagonal sequence $\{f_{j_n(n)}\}$ converges at the point $z_k$ for every $k$. Call this diagonal sequence of functions $\{g_n\}$ for short.

The uniform equicontinuity forces this diagonal sequence $\{g_n\}$ to converge everywhere on $K$ (and uniformly) by the following argument. If a positive $\varepsilon$ is specified, then there is a positive $\delta$ such that if $|z-w|<\delta$, then $|f(z)-f(w)|<\varepsilon$ for every function $f$ in the original sequence of functions. The triangle inequality implies that

$$|g_n(z)-g_m(z)| \le |g_n(z)-g_n(z_k)| + |g_n(z_k)-g_m(z_k)| + |g_m(z_k)-g_m(z)|$$

for an arbitrary value of $k$. For each fixed $z$, there is some point $z_k$ in the specified dense set such that $|z-z_k|<\delta$. Hence the first and third terms on the right-hand side of the preceding inequality each can be made less than $\varepsilon$ by a suitable choice of $z_k$. For a fixed $z_k$, the middle term will be less than $\varepsilon$ when $n$ and $m$ are sufficiently large, in view of the convergence of the sequence $\{g_n\}$ at the points of the dense set. Consequently, the diagonal sequence satisfies Cauchy's criterion uniformly on $K$.

To prove the converse direction of the theorem, first observe that a compact subset of a metric space always is closed. If a compact set of functions fails to be uniformly bounded, then there is a sequence $\{f_n\}$ of functions in the set and a sequence $\{z_n\}$ of points in $K$ such that $|f_n(z_n)|>n$ for each natural number $n$. There is an increasing sequence $\{n(k)\}$ such that the sequence $\{z_{n(k)}\}$ converges to some point $z$ in the compact set $K$ and the sequence $\{f_{n(k)}\}$ converges uniformly to some function $f$ in $C(K)$. Then $f_{n(k)}(z_{n(k)}) \to f(z)$, but also $f_{n(k)}(z_{n(k)}) \to \infty$. The contradiction shows that a compact set of functions is necessarily uniformly bounded on $K$.

If a compact set of functions fails to be uniformly equicontinuous, then there exists some positive $\varepsilon$ such that for every natural number $n$ there are points $z_n$ and $w_n$ and a function $f_n$ in the set such that $|z_n-w_n|<1/n$ but $|f_n(z_n)-f_n(w_n)|\ge\varepsilon$. Compactness implies that there is an increasing sequence $\{n(k)\}$ such that the sequence $\{z_{n(k)}\}$ converges to a point $z$ in $K$, the sequence $\{w_{n(k)}\}$ converges to a point $w$ in $K$, and the sequence $\{f_{n(k)}\}$ converges uniformly to a continuous function $f$ such that $|f(z)-f(w)|\ge\varepsilon$, but $|z-w|\le 0$. The contradiction shows that a compact set of functions must be uniformly equicontinuous after all. □

There is a standard method for bootstrapping the metric on $C(K)$ to a metric on $C(G)$, the space of continuous functions on an open set $G$ in $\mathbb{C}$. First notice that $\|f-g\|_K/(1+\|f-g\|_K)$ defines a bounded metric that determines the same topology (the same convergent sequences) on $C(K)$ as does the metric $\|f-g\|_K$. (To verify the triangle inequality, observe that the real-valued function $x/(1+x)$ on the positive real numbers is both increasing and subadditive.)

A standard construction produces an increasing sequence $\{K_n\}$ of nonempty compact sets that exhaust $G$ in the sense that the union $\bigcup_{n=1}^{\infty} K_n = G$, and every compact subset $K$ of $G$ is a subset of $K_n$ when $n$ is sufficiently large. If $G=\mathbb{C}$, then take $K_n$ to be the closed ball of radius $n$ centered at the origin. If $G$ is a proper subset of $\mathbb{C}$, then fix a particular point $z_0$ in $G$, and define $K_n$ to be the set of points of $G$ whose distance from $z_0$ is less than or equal to $n$ and whose distance from the complement of $G$ is greater than or equal to $1/n$ times the distance of $z_0$ from the complement of $G$.

This exhaustion has two additional useful properties. The construction implies that for each $n$, the set $K_n$ is contained in the interior of $K_{n+1}$. Moreover, the set $K_n$ has no unnecessary holes, in the sense that each component of $\mathbb{C}_\infty \setminus K_n$ contains a component of $\mathbb{C}_\infty \setminus G$. To see why, observe that $\mathbb{C}_\infty \setminus G$ is a subset of $\mathbb{C}_\infty \setminus K_n$, so each component of $\mathbb{C}_\infty \setminus G$ is contained in some component of $\mathbb{C}_\infty \setminus K_n$. The only question, then, is whether every component of $\mathbb{C}_\infty \setminus K_n$ intersects $\mathbb{C}_\infty \setminus G$. If $z$ is an arbitrary point in $\mathbb{C}_\infty \setminus K_n$, then either $z = \infty$, or $|z| > n$, or there is a point $w$ in the boundary of $G$ such that $|w - z| < 1/n$. In the first two cases, the component of $\mathbb{C}_\infty \setminus K_n$ containing $z$ intersects $\mathbb{C}_\infty \setminus G$ at $\infty$. In the third case, the open ball $B(w; 1/n)$ is a connected set that lies in $\mathbb{C}_\infty \setminus K_n$ and hence lies in the component of $\mathbb{C}_\infty \setminus K_n$ containing $z$, which thus intersects $\mathbb{C}_\infty \setminus G$ at $w$.

Having fixed an exhaustion $\{K_n\}$ of $G$ once and for all, define $d(f, g)$ as follows:

$$d(f,g) = \sum_{n=1}^{\infty} \frac{\|f - g\|_{K_n}}{1 + \|f - g\|_{K_n}} \cdot \frac{1}{2^n}.$$

Evidently this function $d$ defines a metric on $C(G)$ that is bounded by 1.

The metric depends on the choice of the exhaustion, but the topology is independent of the choice. Indeed, convergence with respect to this metric is the same as uniform convergence on every compact subset of $G$, for the following reason. Convergence in this metric implies, in particular, uniform convergence on each individual compact set $K_n$ and hence on an arbitrary compact set. Conversely, if uniform convergence happens on each compact set, and a positive $\varepsilon$ is fixed, then chopping off the tail at a point where $\sum_{n \geq N} 1/2^n < \varepsilon/2$ and invoking uniform convergence on $K_N$ shows that convergence happens in the metric $d$.

**Theorem** (Arzelà–Ascoli revisited)**.** *A subset of $C(G)$ is relatively compact (that is, has compact closure) if and only if this set of functions is pointwise bounded and pointwise equicontinuous.*

*Proof.* If $\{f_n\}$ is a pointwise bounded and pointwise equicontinuous sequence in $C(G)$, then the first version of the theorem produces a subsequence that converges uniformly on $K_1$, the first set in the exhaustion of $G$. There is a further subsequence that converges uniformly on $K_2$, and so on. The diagonal subsequence converges uniformly on every compact subset of $G$, that is, converges in $C(G)$. This conclusion proves one direction of the theorem.

For the converse, suppose a sequence of functions is not bounded at some point. Then there is a subsequence that blows up at the point. Evidently no further subsequence can converge in $C(G)$. Therefore the original sequence cannot be a relatively compact subset of $C(G)$. Similarly, a relatively compact sequence of functions cannot fail equicontinuity at a point, for passing to a locally uniformly convergent subsequence gives a contradiction by a $3\varepsilon$ argument. □

## 2.3 Compactness in $H(G)$

The next goal is to characterize compactness in the space $H(G)$ of holomorphic functions on the open set $G$. The first observation is that $H(G)$ is a *closed* subspace of $C(G)$. In other words, if a sequence of holomorphic functions converges uniformly on compact sets to a (necessarily) continuous limit function, then the limit function is holomorphic. Since holomorphicity can be tested by integration over closed curves, which are compact sets, the observation follows directly from Morera's theorem or from Cauchy's integral formula.

**Theorem** (One of Montel's theorems)**.** *A set of functions in $H(G)$ is relatively compact if and only if the set of functions is locally bounded.*

The theorem comes from Paul Montel's 1907 thesis, written under the direction of Borel and Lebesgue. A set of functions satisfying the condition of the theorem is called a *normal family* in Montel's terminology. Montel published in 1927 a book titled *Leçons sur les familles normales de fonctions analytiques et leurs applications*. Some authors (including Montel himself) allow the term "normal family" to be a bit more general, allowing the possibility that the limit of a subsequence is identically equal to the constant $\infty$. This generalization amounts to working inside the space $C(G, \mathbb{C}_\infty)$ of continuous functions taking values in the extended complex numbers (equipped with the spherical metric) instead of inside the space $C(G, \mathbb{C})$ of continuous functions taking values in the finite complex numbers.

Incidentally, Texas A&M Distinguished Professor Emeritus Ciprian Foias is a mathematical grandson of Montel (1876–1975) through Miron Nicolescu (1903–1975), and two mathematical grandsons of Foias himself are on the faculty (Michael Anshelevich and Kenneth Dykema, both students of Dan-Virgil Voiculescu). Visiting Assistant Professor Paul Skoufranis is a mathematical great-grandson of Foias through Voiculescu and his student Dimitri Shlyakhtenko.

*Proof.* For the sufficiency, what needs to be shown is that a locally bounded family of holomorphic functions is equicontinuous at each point. By Cauchy's estimate for the first derivative, the family of derivatives of a locally bounded family is again a locally bounded family (one has to shrink disks, but the property is local, so shrinking is allowable). Hence the functions in the original family are Lipschitz with a Lipschitz constant that is locally bounded independently of the function. Equicontinuity evidently follows.

The converse, that a normal family of holomorphic functions must be locally bounded, follows from a previous observation that pointwise boundedness in the presence of equicontinuity implies local boundedness. $\square$

*Example.* The set of holomorphic functions mapping an open set $G$ into the unit disk is a normal family. Indeed, the family is not only locally bounded but even bounded. This example will be used in the proof of the Riemann mapping theorem.

*Example.* The set of holomorphic functions mapping an open set $G$ into the upper half-plane is a normal family. To see why, observe that $H(G)$ is a metric space, so compactness is the same as sequential compactness. What needs to be shown, then, is that if $\{f_n\}_{n=1}^{\infty}$ is a sequence of holomorphic functions taking values in the upper half-plane, then there is a subsequence converging uniformly on each compact subset of $G$.

Let $\varphi$ denote the linear fractional function sending a point $z$ in the upper half-plane to the point $\dfrac{z-i}{z+i}$. Points in the upper half-plane are closer to $i$ than to $-i$, that is, $|z-i| < |z+i|$, so $\varphi$ maps the upper half-plane to the unit disk. The sequence $\{\varphi \circ f_n\}$ of composite functions fits the previous example, so there is a subsequence converging uniformly on compact sets. Composing with the biholomorphic mapping $\varphi^{-1}$ preserves the convergence, so the corresponding subsequence of the original sequence $\{f_n\}$ converges uniformly on compact subsets of $G$.

The preceding examples can be pushed further. The philosophy is that if the common range of a family of holomorphic functions is not too big, then the family is normal. How big is too big?

Evidently the whole of $\mathbb{C}$ is too big, for the family of functions $\exp(nz)$ (where $n$ runs through the natural numbers) is not locally bounded for $z$ in the unit disk. The same example shows that the punctured plane $\mathbb{C} \setminus \{0\}$ is still too big for the range of the functions. Remarkably, the twice-punctured plane is small enough.

**Theorem** (Montel's fundamental normality criterion)**.** *The family of holomorphic functions mapping an open set $G$ into $\mathbb{C} \setminus \{0, 1\}$ (the twice-punctured plane) is a normal family in the extended sense that every sequence of such functions either admits a subsequence that converges uniformly on compact sets to a holomorphic function or admits a subsequence that converges uniformly to $\infty$.*

Of course, the values 0 and 1 could be replaced by two arbitrary distinct complex numbers $a$ and $b$ (the same numbers for all members of the family of functions). Simply make a linear fractional transformation that fixes $\infty$ and moves the points $a$ and $b$ to 0 and 1.

Notice that the theorem provides a sufficient condition for normality, but the condition is not necessary. A family of functions might be normal even though there is no omitted value. A simple example in the setting of entire functions is the family consisting of a single nonconstant polynomial: the range is all of $\mathbb{C}$, and a singleton set is always compact.

The name "fundamental criterion" (*critère fondamental*) is due to Montel himself. Montel's fundamental criterion is quite deep. Indeed, an easy proof of Picard's theorem is a consequence. The proofs of Montel's criterion and Picard's theorem are deferred until later. (Both theorems appear in Chapter XII of the textbook.)

## 2.4 Aside on non-normal families

Concerning Montel's necessary and sufficient condition for the restricted notion of normality, the question arises of exhibiting a family of functions in $H(G)$ that is pointwise bounded but not locally bounded. The difficulty of writing down a concrete example is revealed by the following observation.

**Theorem.** *If $G$ is an open subset of $\mathbb{C}$, and $\mathcal{F}$ is a pointwise bounded family of continuous functions on $G$, then there is a dense open subset of $G$ on which the family $\mathcal{F}$ is locally bounded.*

*Proof.* Consider an arbitrary closed disk $D$ contained in the open set $G$. Equipped with the standard topology inherited from $\mathbb{C}$, the disk $D$ is a complete metric space, so the Baire category theorem is applicable. [Baire's theorem says that a complete metric space cannot be expressed as the union of countably many nowhere dense sets; equivalently, the intersection of countably many dense open sets is still dense.] For each natural number $n$, let $E_n$ denote the set of points of $D$ at which all functions belonging to the family $\mathcal{F}$ are bounded (in absolute value) by the value $n$. Continuity of the functions implies that each set $E_n$ is a closed set. By hypothesis, every point of $D$ lies in some $E_n$. By the Baire category theorem, there is some value of $n$ for which the set $E_n$ has nonvoid interior.

Let $S_D$ denote the interior of the indicated subset $E_n$ of $D$. Now let $D$ vary over all closed disks contained in $G$, and let $S$ denote the union of the corresponding $S_D$ sets. Then $S$ is an open subset of $G$; and $S$ is dense in $G$ because $S$ intersects every neighborhood; and every point of $S$ has a neighborhood on which the family $\mathcal{F}$ is bounded. Thus $S$ is the required dense open subset of $G$ on which the family $\mathcal{F}$ is locally bounded. □

The preceding theorem reveals that examples of pointwise bounded but not locally bounded families of holomorphic functions must be somewhat tricky, since local boundedness has to hold on a dense subset. One way to construct examples is to apply a significant upcoming theorem, Runge's approximation theorem (the proof of which is deferred until later).

**Theorem** (First version of Runge's approximation theorem)**.** *If $K$ is a compact subset of $\mathbb{C}$, possibly disconnected but without holes (that is, $\mathbb{C} \setminus K$ is connected), then every function that is holomorphic on an open neighborhood of $K$ can be uniformly approximated on $K$ by holomorphic polynomials.*

The wording "holomorphic polynomial" sounds redundant but is not. The function $\operatorname{Re}(z^2)$ is a nonholomorphic polynomial (a polynomial in the underlying real variables). What is wanted in the theorem is a polynomial in the complex variable. A more precise statement of the conclusion of the theorem is that if $f(z)$ is holomorphic in a neighborhood of $K$, then there is a sequence $\{p_n(z)\}_{n=1}^{\infty}$ of polynomials such that $\max_{z\in K} |f(z) - p_n(z)| \to 0$ when $n \to \infty$.

*Example.* There exists a sequence $\{p_n(z)\}$ of polynomials converging *pointwise* everywhere in $\mathbb{C}$, the limit being identically equal to 0 in the open upper half-plane and identically equal to 1 in the closed lower half-plane. The convergence is certainly not uniform on compact sets, since the limit function is not continuous. Since the sequence of functions is pointwise convergent, the sequence is pointwise bounded. The sequence is not locally bounded, since the family is not normal.

To construct the example, apply Runge's theorem on an increasing sequence $\{K_n\}$ of disconnected compact sets. Let $K_n$ be the union of two closed rectangles, one in the closed lower half-plane with vertices at the points $-n$, $n$, $n - in$, and $-n - in$, and the other in the open upper half-plane with vertices at $-n + i/n$, $n + i/n$, $n + in$, and $-n + in$. Evidently these rectangles form an increasing sequence whose union is the whole plane.

The piecewise-constant function that equals 0 when $\operatorname{Im} z > 1/(2n)$ and 1 when $\operatorname{Im} z < 1/(2n)$ is holomorphic on an open set containing $K_n$, a compact set having no holes, so by Runge's theorem there exists a holomorphic polynomial $p_n$ that approximates this piecewise-constant function uniformly on $K_n$ with error less than $1/n$. In other words, $|p_n(z)| < 1/n$ when $z$ is in the top rectangle, and $|p_n(z) - 1| < 1/n$ when $z$ is in the bottom rectangle.

Consequently, $p_n(z) \to 0$ locally uniformly in the open upper half-plane, and $p_n(z) \to 1$ uniformly on each compact subset of the closed lower half-plane. The convergence is not uniform on any neighborhood of a point on the real axis.

## 2.5 The Julia set

Normal families appear in the theory of iteration of holomorphic functions. Suppose $p(z)$ is a polynomial, and consider the sequence $p \circ p$, $p \circ p \circ p$, ... of iterates. The largest open subset of $\mathbb{C}$ on which the sequence of iterates is a normal family is called the *Fatou set* of $p$. The complement of the Fatou set is the *Julia set*. These notions are interesting already for quadratic polynomials. The names honor the French mathematicians Pierre Fatou (1878–1929), who is known not only for his work in complex analysis but also for Fatou's lemma in the theory of integration; and Gaston Julia (1893–1978), whose work on iteration of rational functions made him famous at age 25.

## 2.6 Normal families on the qualifying examination

Here are some problems involving normal families that appeared on past qualifying examinations.

- Problem 4 on the January 2010 qualifying exam asks you to prove normal convergence given convergence at one point and convergence of the real parts.
- Problem 9 on the August 2010 qualifying exam asks you to prove for a bounded family of harmonic functions that convergence on a subdomain implies normal convergence throughout.
- Problem 8 on the August 2011 qualifying exam asks about local boundedness of the iterates of the sine function.
- Problem 5 on the January 2013 qualifying exam asks about normality of functions in the unit disk satisfying the bound in the Bieberbach conjecture.
- Problem 9 on the August 2014 qualifying exam asks about self-mappings of the unit disk converging normally to a constant.
- Problem 9 on the January 2015 qualifying exam asks about normality of iterates of the logarithm on the upper half-plane.

## 2.7 Applications of convergence in $H(G)$

**Theorem.** *If a sequence of holomorphic functions converges normally (uniformly on compact sets), then so does the sequence of derivatives.*

**Theorem** (Hurwitz)**.** *If $G$ is a connected open set, and $\{f_n\}$ is a sequence of zero-free holomorphic functions converging uniformly on compact sets to a limit function $f$, then either $f$ is zero-free or $f$ is identically equal to zero.*

**Corollary.** *If $G$ is a connected open set, and $\{f_n\}$ is a sequence of injective holomorphic functions converging uniformly on compact sets to a limit function $f$, then either $f$ is injective or $f$ is constant.*

This corollary is Exercise 10 in §2 of Chapter VII.

*Proof that derivatives inherit normality.* The derivative of a holomorphic function is represented by an integral, and uniform convergence of the integrands implies convergence of the integrals. Since the Cauchy integral kernel is uniformly bounded when the free variable is bounded away from the integration curve, the convergence is uniform on compact sets. □

*Proof of Hurwitz's theorem.* Notice that the second case can occur: consider, for example, the sequence $\{z^n\}$ on the open unit disk with a puncture at the origin.

If $f(z_0) = 0$, but $f$ is not identically equal to 0, then $f$ has no zeroes in some punctured neighborhood of $z_0$ (since the zeroes of $f$ are isolated). Therefore if $D$ is a sufficiently small disk centered at 0 whose closure is contained in $G$, the function $f$ has no zero on the boundary of $D$. Then

$$\frac{1}{2\pi i}\int_{\partial D}\frac{f_n'(z)}{f_n(z)}\,dz \to \frac{1}{2\pi i}\int_{\partial D}\frac{f'(z)}{f(z)}\,dz.$$

The integral counts the number of zeroes of the function inside $D$. Since the approximating integrals all are equal to 0, and the limiting integral is equal to 1, a contradiction arises. □

*Proof of corollary.* Fix a point $z_0$ in $G$. The function that sends $z$ to $f_n(z) - f_n(z_0)$ is zero-free on the region $G \setminus \{z_0\}$ by hypothesis. Hurwitz's theorem implies that the limit function $f(z) - f(z_0)$ is either zero-free or identically zero on $G \setminus \{z_0\}$. Since $z_0$ is arbitrary, the function $f$ on $G$ takes each value in its range only once, unless the function is constant. □

**Theorem** (Vitali's theorem)**.** *If $\{f_n\}$ is a normal family of holomorphic functions on a connected open set $G$, and if the sequence converges pointwise on a subset of $G$ that has an accumulation point in the interior of $G$, then the sequence of functions converges normally on all of $G$.*

The theorem is named for the Italian mathematician Giuseppe Vitali (1875–1932), who is known also for an example of a nonmeasurable set of real numbers. The result is sometimes called the Vitali–Porter theorem, since M. B. Porter discovered the theorem independently at about the same time. Vitali's theorem is Exercise 4 in §2 of Chapter VII.

*Proof.* Every subsequence of $\{f_n\}$ has a further subsequence that converges normally to a holomorphic limit function. By hypothesis, all of these limit functions agree on a set that has a limit point, so by the identity theorem, all of the limit functions agree identically on $G$. Call this unique common limit $g$. If there were a compact set $K$ on which the sequence $\{f_n\}$ fails to converge uniformly to $g$, then there would be a positive $\varepsilon$ and a subsequence $\{f_{n_k}\}$ such that $\|f_{n_k} - g\|_K \ge \varepsilon$ for every $k$. But the subsequence $\{f_{n_k}\}$ has a further subsequence that does converge uniformly on $K$ to $g$. This contradiction completes the proof. □

## 2.8 Proof of the Riemann mapping theorem

The first complete proof apparently is due to Carathéodory in 1912; he produced the map as the limit of a sequence of maps. The modern proof via an extremal problem is due to Fejér and Riesz, published with permission by Radó in 1922. The square-root trick to cook up an improved mapping apparently is due to Carathéodory and Koebe. Fejér and Riesz make the explicit computation; the method for avoiding the computation seems to be due to Ostrowski and Carathéodory. Carathéodory's method (given below) avoids taking derivatives.

*Remark.* The Riemann map certainly is not unique, for one can post-compose with an arbitrary automorphism of the disk. The map can be made unique in various ways. For instance, if a point $z_0$ in $G$ is chosen that maps to 0, and if the derivative of the mapping at $z_0$ is specified to be a positive real number, then the mapping is unique. Indeed, if $f$ and $g$ are two such maps, then $f \circ g^{-1}$ is an automorphism of the unit disk fixing the origin and having positive derivative at the origin. The Schwarz lemma (discussed below) implies that such a map is a rotation, and the positivity of the derivative forces this composite map to be the identity rotation.

Another way to ensure uniqueness is to choose two distinct points $z_0$ and $z_1$ in $G$ and demand that $z_0$ maps to 0 and $z_1$ maps to a positive real value. Again, if $f$ and $g$ are two such maps, then $f \circ g^{-1}$ is an automorphism of the disk that fixes 0 and maps some positive real number to a positive real number. Hence $f \circ g^{-1}$ is a trivial rotation.

*Proof of existence.* The outline of the proof is the following. Consider the family of all injective holomorphic functions that map the given simply connected region $G$ into (not necessarily onto) the unit disk, taking a specified point $z_0$ to 0. The goal is to find a mapping in this normal family that makes the image fill out as much of the disk as possible. Namely, there is an extremal function that maximizes the modulus of the value of the map at a second specified point $z_1$. This extremal function must be the required holomorphic bijection, else a new function could be constructed that increases the value at $z_1$.

An alternative extremal problem—the one used in the textbook—is to maximize the absolute value of the derivative of the mapping function at $z_0$. "There is more than one way to do it."

Numerous details need to be filled in.

First of all, are there *any* injective holomorphic functions mapping $G$ into the unit disk? If $G$ were the whole plane, then there would be no nonconstant maps (by Liouville's theorem), hence the exclusion of the plane is necessary in the statement of the theorem.

If $G$ is a *bounded* region, then there are lots of injective maps into the unit disk: translate $G$ to move $z_0$ to the origin, then dilate by a suitable factor less than 1.

What if $G$ is unbounded? Since $G$ is not the whole plane, there is at least one point $b$ in the complement of $G$. If $b$ is an interior point of $\mathbb{C} \setminus G$, then an inversion with respect to $b$ maps $G$ into a bounded region, and the previous case can be invoked. So the hard case is the case in which the complement of $G$ has empty interior: the region $G$ could be the plane with a slit, for instance.

If $b$ is a point in the complement of $G$, then $z - b$ is a zero-free holomorphic function on $G$ (a simply connected region), so there is a holomorphic branch of $\sqrt{z-b}$ on $G$. Evidently this square root is injective (if not, then $z - b$ would not be), so what needs to be shown is that the image of $G$ under $\sqrt{z-b}$ omits some disk, thus reducing to the previous case.

Now if $c$ is a point in the image of $\sqrt{z-b}$ (and $c$ is necessarily different from zero, since $z-b$ is zero-free on $G$), then the point $-c$ is not in the image: for if both $\sqrt{z_2 - b} = c$ and $\sqrt{z_3 - b} = -c$, then squaring shows that $z_2 - b = z_3 - b$, a contradiction. Since $\sqrt{z-b}$ is an open map, a whole neighborhood of $c$ is in the image, so a neighborhood of $-c$ is not in the image. Thus the previous case produces an injective map of $G$ into the unit disk. Composing with a suitable automorphism of the disk will arrange that the specified point $z_0$ goes to the origin.

Fix a point $z_1$ in $G$ different from $z_0$. Take a sequence in the family for which the absolute value of the function at $z_1$ approaches the least upper bound of all such values. There is a subsequence converging normally to a holomorphic limit function $f$. Evidently $f(z_0) = 0$, and $|f(z_1)|$ achieves the extreme value in the family. In particular, $f(z_1)$ is different from $f(z_0)$. Therefore the limit function is not constant, so $f$ is injective, being the limit of injective holomorphic functions. The function $f$ a priori maps into the closed unit disk, but by the maximum principle the image lies in the open unit disk. Thus $f$ is indeed an extremal function within the family.

What remains to show is that the extremal function is surjective. The argument is by contradiction. Suppose a nonzero point $c$ in the disk is not in the image of $f$. The goal is to produce a contradiction by finding a new function in the family whose value at $z_1$ has absolute value larger than $|f(z_1)|$.

## 2.9 Aside on self-mappings of the disk

The plan is to compose $f$ with holomorphic mappings of the disk into itself, and such mappings are interesting for their own sake. Here is the first observation.

**Lemma** (Schwarz lemma)**.** *Suppose $f$ is a holomorphic function (not necessarily injective, not necessarily surjective) that maps the unit disk into itself, fixing the origin. Then either $f$ is a rotation, or $|f(z)| < |z|$ when $z \neq 0$. In the latter case, $|f'(0)| < 1$.*

*Proof.* Since $f(0) = 0$, the quotient $f(z)/z$ has a removable singularity at the origin. On a circle of radius $r$ less than 1, this quotient has absolute value bounded above by $1/r$. The maximum principle implies that $|f(z)/z| \leq 1/r$ when $|z| \leq r$. Keeping $z$ fixed, let $r$ tend to 1 to deduce that $|f(z)/z| \leq 1$ for every value of $z$ in the unit disk. In other words, $|f(z)| \leq |z|$ for every $z$ in the unit disk.

If equality holds for some nonzero value of $z$, then the function $|f(z)/z|$ attains a maximum at an interior point, hence is constant by the maximum principle. The constant value has modulus equal to 1, so $f$ is a rotation. The removable singularity of $f(z)/z$ is removed by setting the value at the origin equal to $f'(0)$, so $|f'(0)| \leq 1$, and the maximum principle again implies that equality holds only if $f$ is a rotation. $\square$

The Schwarz lemma implies that the solution of the extremal problem under consideration in the proof of the Riemann mapping theorem is necessarily surjective if $G$ is already the unit disk.

Indeed, suppose $z_0$ is 0, and consider a holomorphic function $f$ (whether injective or not) that maps the unit disk into itself, taking 0 to 0. The Schwarz lemma implies that if $z_1$ is another point in the unit disk, then $|f(z_1)| \le |z_1|$, and equality obtains if and only if the function $f$ is a rotation. Thus the extremal function—the function that realizes equality—is bijective.

What if $z_0$ is some point other than 0? This situation can be reduced to the preceding special case by using the knowledge that all points in the unit disk are equivalent to each other from the point of view of complex analysis. More precisely, there exists a holomorphic bijection of the disk that takes any prescribed point of the disk to any other prescribed point. In other words, the holomorphic automorphism group of the disk is transitive.

To see why all points of the disk are equivalent, let $c$ be a point of the open unit disk, and define a function $\varphi_c$ as follows:

$$\varphi_c(z) = \frac{c - z}{1 - \overline{c}z}.$$

Notice that $\varphi_c(c) = 0$, and $\varphi_c(0) = c$, so this function interchanges the points 0 and $c$. A routine computation shows that

$$\left|\frac{c - z}{1 - \overline{c}z}\right|^2 = 1 - \frac{(1 - |c|^2)(1 - |z|^2)}{|1 - \overline{c}z|^2}.$$

Since $1 - |c|^2 > 0$, the preceding identity shows that $|\varphi_c(z)| < 1$ when $|z| < 1$, and $|\varphi_c(z)| = 1$ when $|z| = 1$. The composite function $\varphi_c \circ \varphi_c$ maps the unit disk into itself and fixes the points 0 and $c$. By the remark following the proof of the Schwarz lemma, this composite function must be the identity rotation. Accordingly, the function $\varphi_c$ is a self-inverse, holomorphic bijection of the unit disk. (You could alternatively compute $\varphi_c \circ \varphi_c$ algebraically to see that this composite function reduces to the identity function.)

The most general holomorphic bijection of the unit disk is the composition of a rotation with some $\varphi_c$. Indeed, if $h$ is a holomorphic bijection of the disk fixing the origin, then the Schwarz lemma implies that $|h(z)| \le |z|$ and $|h^{-1}(z)| \le |z|$ for every point $z$ in the disk. Therefore $|h(z)| = |z|$ for every $z$, and the equality case of the Schwarz lemma implies that $h$ is a rotation. Next, if $h$ is a holomorphic bijection of the disk that moves some point $c$ to 0, then $h \circ \varphi_c$ is a holomorphic bijection of the disk that fixes 0, hence is a rotation.

Here is a formula for a holomorphic automorphism of the unit disk that interchanges two prescribed points $z_1$ and $z_2$:

$$\varphi_{z_1} \circ \varphi_{\varphi_{z_1}(z_2)} \circ \varphi_{z_1}.$$

Being a composition of disk automorphisms, this function is a disk automorphism. Routine checking shows that indeed this composite function takes $z_1$ to $z_2$ and $z_2$ to $z_1$.

## 2.10 Conclusion of the proof of the Riemann mapping theorem

Returning to the proof of the Riemann mapping theorem, notice that the hypothesized missing point $c$ is not 0, since $f(z_0) = 0$. Under the hypothesis that $c$ is not in the image of $f$, the function $\varphi_c \circ f$ is zero-free in the simply connected region $G$, hence has a holomorphic square root, say $g$. This function $g$ is injective, for otherwise the square would not be injective. Now $g$ maps the region $G$ into the unit disk, but $g$ does not belong to the specified family of functions, for $g$ is not normalized at $z_0$. Indeed, $\varphi_c \circ f(z_0) = c$, so $g(z_0) = \sqrt{c}$ for one of the two possible values of the square root.

Set $h$ equal to $\varphi_{\sqrt{c}} \circ g$. Then $h$ again maps $G$ into the unit disk, and now $h(z_0) = 0$. What remains to show (to reach the desired contradiction) is that $|f(z_1)| < |h(z_1)|$. The plan now is to unwind the definitions to relate $f$ to $h$.

On the one hand, $g^2 = \varphi_c \circ f$, so $f = \varphi_c \circ g^2$. (Notice that the notation $g^2$ represents an algebraic square $g \times g$, not a composition.) On the other hand, $g = \varphi_{\sqrt{c}} \circ h$, so $g^2 = (\varphi_{\sqrt{c}} \circ h)^2 = (\varphi_{\sqrt{c}})^2 \circ h$. Therefore $f = \varphi_c \circ (\varphi_{\sqrt{c}})^2 \circ h$. Now the function $\varphi_c \circ (\varphi_{\sqrt{c}})^2$ maps the unit disk to itself, fixing the origin. By the Schwarz lemma, $|\varphi_c \circ (\varphi_{\sqrt{c}})^2(z)| \leq |z|$ for every point $z$ in the unit disk. Moreover, if equality holds in the Schwarz lemma for even one nonzero point, then the function has to be a rotation. But the map $\varphi_c \circ (\varphi_{\sqrt{c}})^2$ evidently is not a rotation, since this map is two-to-one (because of the square). Therefore $|\varphi_c \circ (\varphi_{\sqrt{c}})^2(z)| < |z|$ for every nonzero point $z$ in the disk, with strict inequality.

Replacing $z$ with $h(z_1)$ in this inequality shows that

$$|f(z_1)| = |\varphi_c \circ (\varphi_{\sqrt{c}})^2(h(z_1))| < |h(z_1)|,$$

so the function $h$ violates the extremality of $f$. The contradiction shows that the map $f$ must be surjective after all. Thus $f$ is the required holomorphic bijection from $G$ to the unit disk. □

## 2.11 Remarks on normality of families of meromorphic functions

Section 3 of Chapter VII of the textbook proves a necessary and sufficient condition for normality of a family of meromorphic functions on an open subset of $\mathbb{C}$, where the notion of convergence is with respect to the spherical metric. A family of meromorphic functions is normal in this sense if every sequence in the family has a subsequence that converges uniformly on compact sets with respect to the spherical distance either to a meromorphic function or to the constant $\infty$. The condition is local boundedness of the quantity

$$\frac{|f'(z)|}{1 + |f(z)|^2}.$$

This quantity is known as the *spherical derivative*, and the theorem is due to Frédéric Marty (1911–1940).

Marty introduced the notion of spherical derivative in his 1931 doctoral dissertation[2] written under the direction of Paul Montel. The characterization of normality of families of meromorphic functions is Theorem 5 in §2 of Chapter I of the dissertation.

Marty was a casualty of the second world war. A lieutenant in the French air force, Marty had a mission on June 14, 1940 (the day that Paris fell to German forces, and nine days before Marty's

[2] F. Marty, Recherches sur la répartition des valeurs d'une fonction méromorphe, *Annales de la faculté des sciences de Toulouse*, third series, **23** (1931), 183–261.

29th birthday) to carry French diplomatic mail from Tallinn, Estonia across the Gulf of Finland to Helsinki. At the time, both Estonia and Finland were technically neutral countries. But the Soviet Union began a military blockade of Estonia on June 14, and the aircraft *Kaleva* on which Marty was a passenger was intercepted by Soviet bombers and shot down—killing all nine on board. The Soviet Union invaded Estonia two days later. War between Finland and the Soviet Union broke out a year later, in June 1941.

The spherical derivative of $f$ arises by dividing the spherical distance between $f(z)$ and $f(w)$ by the Euclidean distance $|z - w|$ and taking the limit as $w$ approaches $z$. This calculation apparently makes sense only away from the poles of the meromorphic function; but since the spherical distance between $f(z)$ and $f(w)$ equals the spherical distance between $1/f(z)$ and $1/f(w)$, poles can be handled by the same method! Thus the elaborate computation on page 157 of the textbook to show that the spherical derivative $\mu(f)$ is well defined at poles is obviated by the observation that $\mu(f) = \mu(1/f)$.

## 2.12 Remark on Exercise 9 in §4 of Chapter VII

The exercise asks for a proof of the existence of an analytic function without critical points (in other words, an analytic function having zero-free derivative), the domain being the unit disk punctured at the origin and the image being the whole unit disk. Although the problem can be solved by available tools, the solution is far from obvious. The following remarks indicate why the problem is hard (some natural attempts to solve the problem fail) and how you might possibly be led to a solution.

The required function is locally injective (since the derivative is never zero) but cannot possibly be globally injective. Indeed, the punctured disk is not homeomorphic to the disk, so there certainly is no analytic bijection between these two open sets. (You could also apply the argument of Exercise 2 in the same section to see that global injectivity is impossible.) Moreover, the required function cannot even be a covering map (in the sense of topology), since a doubly connected space cannot cover a simply connected space. Accordingly, the required function is a local homeomorphism that does not evenly cover the image. (In topology, "even covering" is a technical term meaning that each point in the image has a neighborhood whose inverse image consists of disjoint open sets each of which is mapped homeomorphically onto the neighborhood.)

On the other hand, the required function is somewhat nice in the sense that the function extends to be analytic on the whole disk (by Riemann's theorem on removable singularities). Therefore a natural idea is to try to build the required function by taking the restriction to the punctured disk of a suitable analytic function from the whole disk to the disk. The squaring function ($z$ maps to $z^2$) does not work: this function maps the disk onto the disk, and the derivative has no zeroes in the punctured disk, but the restriction of the function to the punctured disk fails to be surjective, since 0 is not in the image of the restriction. If $c$ is some nonzero point in the disk, and $\varphi_c$ is the standard disk automorphism that interchanges 0 and $c$, then the square $\varphi_c^2$ fails for a different reason. The restriction of $\varphi_c^2$ to the punctured disk now maps surjectively to the disk (for $\varphi_c$ moves the puncture to $c$, and the squaring function maps $c$ and $-c$ to the same value, so the hole fills in), but $\varphi_c^2$ has a critical point at $c$. Similar considerations show that the map sending $z$ to $z\varphi_c(z)$ and the map sending $z$ to $\varphi_b(z)\varphi_c(z)$ fail to solve the problem: both functions are surjective but have a critical point.

Surprisingly, the related problem of mapping a *twice*-punctured disk onto the disk by an analytic function without critical points can be solved by modifying the preceding failed attempts. This claim sounds strange at first, because making the domain smaller would seem to increase the difficulty of filling out the desired image. The advantage of making the domain smaller is that there is then more flexibility for avoiding critical points.

Indeed, fix some nonzero point $c$ in the unit disk, and consider the function $g$ that maps $z$ to $z^2\varphi_c(z)$. This function has exactly two critical points in the unit disk: one critical point at 0 and another critical point $b$ (different from both 0 and $c$) that you can compute by solving a quadratic equation (whence the critical point is nondegenerate: $g'(b) = 0$ but $g''(b) \neq 0$). Viewed as a mapping from the whole disk to the disk, the mapping $g$ is three-to-one. (Evidently the value 0 is taken three times, at 0, 0, and $c$; since $|g(z)| = 1$ when $|z| = 1$, the argument principle can be applied to show that every value in the disk is taken three times, counting multiplicity.) The three points that map to the nonzero value $g(b)$ are $b$, $b$, and something else that is neither 0 nor $b$. The upshot is that if $g$ is restricted to the disk with the points 0 and $b$ removed, then $g$ has no critical points on this twice-punctured disk, and $g$ maps the twice-punctured disk surjectively onto the whole disk. (The points 0 and $g(b)$ in the image are covered once each; all other points are covered three times.)

In view of the preceding example, the original problem reduces to finding an analytic function without critical points that maps the once-punctured disk onto a twice-punctured disk. Composing this function with the one just constructed produces the required mapping. (The chain rule implies that composing two functions with zero-free derivatives yields another function with the same property.)

Solving the new problem requires yet another example. The claim is that there exists an analytic function $h$ without critical points mapping the unit disk onto the disk punctured at the origin, values in the open upper half of the punctured disk being taken twice and values in the open lower half of the punctured disk being taken once. Suppose for the moment that such an $h$ is known. Let $a$ be the unique point in the unit disk such that such that $h(a) = -i|b|$. Then the function $i\frac{b}{|b|} \cdot h \circ \varphi_a$ has no critical point and maps the disk punctured at 0 onto the disk punctured at 0 and $b$.

All that remains, then, is to construct the special function $h$ mapping the unit disk onto the punctured disk. By the Riemann mapping theorem, there is a biholomorphic mapping from the unit disk to the half-strip $\{ z \in \mathbb{C} : \operatorname{Re}(z) < 0 \text{ and } 0 < \operatorname{Im}(z) < 3\pi \}$. [This map can even be computed explicitly. Start with the linear fractional transformation $\dfrac{1-z}{1+z}$ mapping the unit disk to the right-hand half-plane. Compose with the principal branch of the square root. Multiply by $e^{i\pi/4}$ to rotate the image to the first quadrant. Compose with the linear fractional transformation $\dfrac{z-1}{z+1}$ to map to the upper half of the unit disk. Compose with the principal branch of the logarithm to map to the half-strip $\{ z \in \mathbb{C} : \operatorname{Re}(z) < 0 \text{ and } 0 < \operatorname{Im}(z) < \pi \}$. Dilate by a factor of 3.] Now compose with the exponential function to map onto the punctured disk, points in the upper half being covered twice and points in the lower half being covered once. Notice that each step uses a function with zero-free derivative. The composition of all these functions is the indicated function $h$. Thus Exercise 9 is solved.

# 3 Infinite products and applications

If you know the zeroes of a polynomial (including multiplicity), then you know the polynomial up to multiplication by a nonzero constant. Indeed, a polynomial whose zeroes are the points $z_1$, $z_2$, …, $z_n$ can be written as $c\prod_{k=1}^{n}(z - z_k)$ for some complex number $c$.

If you think of an entire function as a polynomial of infinite degree, then you might expect that knowing the zeroes of the function would essentially determine the function. The situation is more complicated than for true polynomials, however, because the indeterminate nonzero scale factor in general is a function, not a constant.

If $f$ and $g$ are two entire functions having the same zeroes with the same multiplicity, then the quotient $f/g$ has only removable singularities and so can be viewed as a zero-free entire function. Since the complex plane is simply connected, a zero-free entire function has a holomorphic logarithm. Thus $f/g$ can be written in the form $e^h$ for some entire function, and $f = ge^h$.

Accordingly, an entire function in principle is determined by its zeroes up to multiplication by a zero-free factor $e^h$. But how can the entire function be explicitly constructed from the zeroes? If there are infinitely many zeroes (as is the case for the sine function, for example), then mimicking the process for polynomials leads to infinite products. The immediate goal is to develop enough theory about infinite products to make sense of formulas like the following one proved in §6 of Chapter VII the textbook:

$$\sin(\pi z) = \pi z \prod_{n=1}^{\infty}\left(1 - \frac{z^2}{n^2}\right)$$

(Euler's product formula for the sine function).

## 3.1 Convergence of infinite products

What should it mean to say that an infinite product $\prod_{n=1}^{\infty} b_n$ converges? Apparently, the natural definition would be existence of the limit $\lim_{N\to\infty}\prod_{n=1}^{N} b_n$. That definition will not do, however, because if $b_1 = 0$, then the limit of partial products exists (and equals 0) for completely arbitrary values of the other factors. But the *existence* of a limit ought not to depend on the value of the first term (or on the values of finitely many terms).

One could insist on considering products having no factors equal to 0, but the application to entire functions needs precisely the case in which some factors are equal to 0. On the other hand, if there were infinitely many factors equal to 0, then the limit of partial products could only be 0, and the limit would be independent of the values of the subsequence of nonzero terms.

Accordingly, the standard definition of convergence of infinite products requires that there be only finitely many factors equal to 0 and that the limit of the partial products of the nonzero factors exists and that this limit is not equal to 0. If the limit of nonzero factors exists and equals 0, then the product is said to diverge to 0.

One reason for excluding 0 as a limit is that one would like to pass back and forth between infinite series and infinite products by using the exponential and logarithm functions. Another reason is to preserve the property that a product is equal to zero if and only if some factor is equal to zero.

*Example.* The infinite product $\prod_{n=1}^{\infty} 1/n$ diverges to 0. The corresponding series $\sum_{n=1}^{\infty} \log(1/n)$ (with the principal branch of the logarithm) diverges to $-\infty$.

*Example.* The infinite product $\prod_{n=1}^{\infty}\left(1+\frac{1}{n}\right)$ diverges. Indeed, the partial product $\prod_{n=1}^{k}\left(1+\frac{1}{n}\right)$ telescopes to the value $k+1$, which does not approach a finite value.

*Example.* The product $\prod_{n=1}^{\infty}\left(1-\frac{1}{n^2}\right)$ converges to 0 for the following reason.

Notice that 0 is an allowed value for the limit, but only if the nonzero factors converge to a nonzero limit. In this example, the first term equals 0. Moreover,

$$\prod_{n=2}^{k}\left(1-\frac{1}{n^2}\right)=\prod_{n=2}^{k}\frac{(n-1)(n+1)}{n^2}.$$

The product telescopes: each natural number appears twice in the numerator and twice in the denominator, except for special terms at the beginning and the end. The product equals

$$\frac{1}{2}\cdot\frac{k+1}{k},$$

which has limit $1/2$ when $k\to\infty$. Therefore the original infinite product converges to 0.

For a product of nonzero terms $b_n$ to converge to a nonzero limit $L$, a necessary condition is that $b_n \to 1$. Indeed, if a positive $\varepsilon$ less than $|L|$ is specified, then convergence of the product implies the existence of a natural number $N$ such that $\left|L-\prod_{n=1}^{k} b_n\right|<\varepsilon$ when $k\geq N$. Write $p_k$ for $\prod_{n=1}^{k} b_n$. Then

$$1-b_k=1-p_k/p_{k-1}=\frac{(p_{k-1}-L)-(p_k-L)}{(p_{k-1}-L)+L},$$

so $|1-b_k|<2\varepsilon/(|L|-\varepsilon)$ when $k>N$.

Accordingly, the general term of an infinite product usually is written in the form $1+a_n$. A necessary condition for convergence of an infinite product is then that $a_n \to 0$. What about sufficient conditions?

**Proposition.** *If no term* $(1+a_n)$ *equals* $0$*, then the infinite product* $\prod_{n=1}^{\infty}(1+a_n)$ *and the infinite series* $\sum_{n=1}^{\infty}\log(1+a_n)$ *(with the principal branch of the logarithm) either both converge or both diverge.*

*Example.* Consider the product $\prod_{n=1}^{\infty}\left(1-\frac{z^2}{n^2}\right)$.

If $|z|\leq R$, say, then $|z^2/n^2|\leq R^2/n^2$, and $\sum_{n=1}^{\infty} R^2/n^2$ converges. Therefore the series $\sum_{n=1}^{\infty}|z^2/n^2|$ converges uniformly for $z$ in a compact set by the Weierstrass $M$-test. Use the observation that $\lim_{w\to 0} \frac{\log(1+w)}{w} = 1$ to deduce that the series $\sum_{n=1}^{\infty} \log\left(1 - \frac{z^2}{n^2}\right)$ converges uniformly on compact sets. Therefore the original infinite product converges uniformly on compact sets. (Exponentiation is a continuous operation, so if the partial sums of the series are close to a limiting value, then the corresponding partial products obtained by exponentiating are closed to the exponential of the limiting value.)

Notice that the discussion so far does not prove Euler's product for the sine function. The infinite product does converge, and multiplying by $z$ gives an entire function having the same zeroes as the sine function. Consequently, the two functions have a ratio that is a zero-free entire function, hence of the form $e^{g(z)}$ for some entire function $g$. More work is needed to show that $g$ is the 0 function.

*Proof of the Proposition.* If the partial sums of the infinite series converge, then the exponentials of the partial sums converge; hence the partial products converge (by continuity of the exponential function). The converse argument is more delicate. If the partial products have a limit, then so does the sequence of logarithms of partial products, but the logarithm of a product is not necessarily equal to the sum of the logarithms for a fixed branch of the logarithm.

Suppose that the partial products do converge (to a nonzero limit). Since

$$\log(1 + a_n) = \log|1 + a_n| + i\arg(1 + a_n),$$

what needs to be checked is that both $\sum_{n=1}^{\infty} \log|1 + a_n|$ and $\sum_{n=1}^{\infty} \arg(1 + a_n)$ converge. If the partial products converge, then continuity of the modulus implies that the partial products of the moduli converge, and to a positive real number. Continuity of the real logarithm function implies that $\sum_{n=1}^{\infty} \log|1 + a_n|$ converges.

Now write $1+a_n = |1+a_n|\,e^{i\theta_n}$, where $\theta_n = \arg(1+a_n)$. Since the partial products $\prod_{n=1}^{N}(1+a_n)$ and $\prod_{n=1}^{N} |1 + a_n|$ both converge to nonzero limits, it follows that the partial products $\prod_{n=1}^{N} e^{i\theta_n}$ converge, say to some $e^{i\varphi}$. Consequently, there is a sequence of integers $m_N$ such that $\varphi + 2\pi m_N - \sum_{n=1}^{N} \theta_n \to 0$ as $N \to \infty$. But $a_n \to 0$ as $n \to \infty$, so $\theta_n \to 0$, and the integer $m_N$ must eventually stabilize at a constant value (since eventually $m_N$ and $m_{N+1}$ differ by less than 1). Consequently, the series $\sum_{n=1}^{\infty} \theta_n$ converges to some value $\varphi + 2\pi m$. □

A simple sufficient condition for convergence of an infinite product $\prod_{n=1}^{\infty}(1 + a_n)$ is that the series $\sum_{n=1}^{\infty} |a_n|$ converges. The intuitive idea is that $\log(1 + a_n) \approx a_n$ when $a_n$ is close to 0, so the hypothesis implies absolute convergence of the series of logarithms, hence convergence of the infinite product. Indeed, since the condition implies that $a_n \to 0$, there is no loss of generality in supposing that $|a_n| < 1/2$, say. Now integrating the geometric series gives a series for the logarithm:

$$\log(1 + z) = z - \tfrac{1}{2}z^2 + \tfrac{1}{3}z^3 - \cdots \qquad \text{when } |z| < 1,$$

so $|\log(1 + z)| \le |z| + |z|^2 + |z|^3 + \cdots = |z|/(1 - |z|)$. Hence $|\log(1 + a_n)| \le 2|a_n|$.

This simple sufficient condition for convergence of an infinite product is not necessary.

*Example.* Consider $\prod_{n=1}^{\infty}\left(1+\frac{(-1)^n}{n^{2/3}}\right)$.

Now $\log(1+a_n) = a_n - \frac{1}{2}a_n^2 + \cdots$, so

$$\log\left(1+\frac{(-1)^n}{n^{2/3}}\right) = \frac{(-1)^n}{n^{2/3}} - \frac{1}{2}\cdot\frac{1}{n^{4/3}} + \cdots = \frac{(-1)^n}{n^{2/3}} + O(1/n^{4/3}).$$

The alternating series $\sum_{n=1}^{\infty}(-1)^n/n^{2/3}$ converges, and the remainder series converges absolutely. Hence the infinite product converges (conditionally).

## 3.2 The Weierstrass factorization theorem for entire functions

Suppose $\{z_n\}$ is a discrete set of points in $\mathbb{C}$ (no accumulation point), and $\{m_n\}$ is a sequence of natural numbers. The goal is to construct an entire function having a zero of order $m_n$ at $z_n$ for every $n$.

The first try is an infinite product of the form $\prod_{n=1}^{\infty}(z-z_n)^{m_n}$, but this attempt fails. Since the factors do not tend to 1, the product diverges.

The second try is an infinite product of the form $\prod_{n=1}^{\infty}\left(1-\frac{z}{z_n}\right)^{m_n}$. This attempt succeeds if $z_n$ tends to infinity fast enough to make the product converge. This method handles, for instance, the construction of a function with simple zeroes at the squares of the natural numbers, since $\sum_{n=1}^{\infty} z/n^2$ converges for every $z$ (and converges uniformly on compact sets, so the limit function is holomorphic). But the method fails to construct a function with a simple zero at each natural number, since $\sum_{n=1}^{\infty}\left(1-\frac{z}{n}\right)$ diverges when $z \neq 0$.

The third try, which succeeds, is to introduce nonvanishing convergence factors. For instance,

$$\prod_{n=1}^{\infty}\left(1-\frac{z}{n}\right)\exp\left(\frac{z}{n}\right)$$

does converge, uniformly on compact sets, since

$$\log\left(1-\frac{z}{n}\right)+\frac{z}{n} = -\frac{1}{2}\cdot\frac{z^2}{n^2}+\cdots = z^2\cdot O(1/n^2).$$

Convergence factors were introduced by Weierstrass and caused a sensation at the time.

**Theorem** (Existence of an entire function with prescribed zeroes)**.** *Let $\{z_n\}$ be a sequence of nonzero complex numbers, possibly with repetitions, but with no limit point. There exists an entire function with zeroes precisely at the points of the sequence, the order of each zero being equal to the multiplicity of the point in the sequence.*

The following lemma is useful in establishing the general result.

**Lemma.** *If* $|z| \le 1/2$*, then the principal branch of the logarithm satisfies the following estimates:*

$$|z + \log(1-z)| \le |z|^2,$$
$$\left|\left(z + \frac{z^2}{2}\right) + \log(1-z)\right| \le |z|^3,$$
$$\left|\left(z + \frac{z^2}{2} + \frac{z^3}{3}\right) + \log(1-z)\right| \le |z|^4,$$

*and so on. Also,* $\log|1-z| \le |z|$.

*Proof.* By Taylor's theorem,

$$\log(1-z) = -z - \frac{z^2}{2} - \frac{z^3}{3} - \cdots \qquad \text{when } |z| < 1.$$

Hence

$$\left|\sum_{n=1}^{k} \frac{z^n}{n} + \log(1-z)\right| = \left|\sum_{n=k+1}^{\infty} \frac{z^n}{n}\right| \le \frac{1}{k+1} \sum_{n=k+1}^{\infty} |z|^n = \frac{1}{k+1} \cdot \frac{|z|^{k+1}}{1-|z|}.$$

Now

$$\frac{1}{k+1} \cdot \frac{1}{1-|z|} \le \frac{2}{k+1} \le 1 \qquad \text{when } |z| \le 1/2 \text{ and } k \ge 1.$$

For the final statement in the lemma, observe that $\log|1-z| \le \log(1+|z|) \le |z|$ by concavity of the real logarithm function. □

A corollary (not actually needed below) is that

$$\left|1 - (1-z)e^{z+\frac{1}{2}z^2+\frac{1}{3}z^3+\cdots+\frac{1}{n}z^n}\right| \le |z|^{n+1} \qquad \text{when } |z| \le 1 \text{ and } n \ge 1,$$

which is Lemma 5.11 in Chapter VII of the textbook (page 168). Indeed, the preceding considerations show that the entire function inside the absolute value on the left-hand side is divisible by $z^{n+1}$. The derivative of the function is easily seen by explicit computation to have nonnegative Maclaurin coefficients, so the function itself has nonnegative Maclaurin coefficients. Hence the function divided by $z^{n+1}$ has modulus that is maximized in the closed unit disk when $z = 1$, where the value is equal to 1.

The expression

$$(1-z)\exp\left(z + \frac{z^2}{2} + \frac{z^3}{3} + \cdots + \frac{z^n}{n}\right)$$

is known as a Weierstrass elementary factor, denoted $E_n(z)$.

*Construction of the convergent Weierstrass product.* Behold:

$$\prod_{n=1}^{\infty} \left(1 - \frac{z}{z_n}\right) \exp\left(\frac{z}{z_n} + \frac{1}{2}\left(\frac{z}{z_n}\right)^2 + \cdots + \frac{1}{n}\left(\frac{z}{z_n}\right)^n\right).$$

The claim is that this product converges uniformly on compact sets, in which case the limit is an entire function that evidently has the required zeroes.

Indeed, $|z_n| \to \infty$ by hypothesis, so if $z$ is confined to a compact set, then $|z/z_n|$ is bounded uniformly by $1/2$ when $n$ is sufficiently large. The lemma then implies that the logarithm of the general term of the product has modulus bounded by $1/2^{n+1}$. Since these bounds are the terms of a convergent infinite series, the infinite product converges uniformly on compact sets by the Weierstrass $M$-test. □

*Remark.* The proof reveals that the sum in the exponent could be stopped at the term with power $n-1$ or $n-17$ instead of the term with power $n$. This refinement is not especially interesting. The interesting question is whether the sum in the exponent can be stopped at a power that is independent of $n$. That question is answered by the Hadamard factorization theorem (covered in Chapter XI).

To construct an entire function whose zero set includes the origin, simply multiply the infinite product by a suitable power of $z$.

**Theorem** (Weierstrass factorization theorem for entire functions, 1876)**.** *Every entire function $f(z)$ can be expressed in the following form:*

$$z^k e^{g(z)} \prod_{n=1}^{\infty} \left(1 - \frac{z}{z_n}\right) \exp\left(\frac{z}{z_n} + \frac{1}{2}\left(\frac{z}{z_n}\right)^2 + \cdots + \frac{1}{m_n}\left(\frac{z}{z_n}\right)^{m_n}\right),$$

*where $k$ (possibly $0$) is the order of the zero of $f$ at the origin, $g$ is some other entire function, the sequence $\{z_n\}$ is the list of nonzero zeroes of $f$ (repeated according to multiplicity), and $\{m_n\}$ is a suitable sequence of natural numbers (it will do to take $m_n = n$).*

*Proof.* Dividing $f(z)$ by a convergent infinite product with the same zeroes as $f$ (of the same orders) produces a zero-free entire function (after removing the removable singularities). Such a function has a holomorphic logarithm, that is, can be written in the form $\exp g$. □

The representation in the Weierstrass factorization theorem is not unique, for the sequence $\{m_n\}$ can be replaced by an arbitrary sequence of larger numbers. Changing this sequence of natural numbers will change the function $g$ in the factorization. Moreover, the factorization changes if the zeroes are reordered. A convenient normalization is to order the zeroes by increasing absolute value, but the preceding proof does not depend on the order of the zeroes.

**Corollary.** *Every function that is meromorphic in the whole plane can be written as the quotient of two entire functions.*

The corollary appeared as problem 9 on the August 2012 qualifying exam!

## 3.3 Prescribed zeroes on general regions

The extension of Weierstrass's theorem to proper subsets of the complex plane was accomplished not by Weierstrass himself but by other researchers (Picard and Mittag-Leffler). Here is the statement and a proof.

**Theorem** (Weierstrass theorem for general regions)**.** *Suppose $G$ is an open subset of $\mathbb{C}$, and $\{z_n\}$ is a sequence of points (possibly with repetitions) in $G$ having no limit point inside $G$. Then there is a holomorphic function on $G$ having zeroes precisely at the points $\{z_n\}$ (with order corresponding to the multiplicity of the point in the sequence).*

*Example.* On an arbitrary open set $G$, there is a holomorphic function that cannot be analytically continued across any boundary point whatsoever.

Indeed, take a sequence in $G$ that has every boundary point as an accumulation point, and use the theorem to construct a holomorphic function with zeroes at the points of the sequence. This function cannot extend to a neighborhood of any boundary point, for the zeroes of the function accumulate inside that neighborhood, which would contradict the identity principle.

To build the indicated sequence, take a dense sequence $\{a_n\}$ in the boundary of $G$. Create a new sequence $\{b_n\}$ that contains each $a_k$ infinitely often. For instance, the sequence $a_1$, $a_1$, $a_2$, $a_1$, $a_2$, $a_3$, $a_1$, $a_2$, $a_3$, $a_4$, $\ldots$ will do. Then take $z_n$ to be a point in $G$ at distance less than $1/n$ from the point $b_n$.

*Proof of the general Weierstrass theorem.* The proof in the textbook is the standard modern proof that throws one point to infinity. Here instead is a perhaps more intuitive proof that works directly on the original open set.

The first idea is to split the sequence of points into two parts, depending on whether the points are close to the boundary of $G$ or close to $\infty$. Namely, view $G$ as the union of the following two disjoint sets:

$$S := \{\, z \in G : |z| \operatorname{dist}(z, \partial G) \geq 1 \,\} \qquad \text{and} \qquad T := \{\, z \in G : |z| \operatorname{dist}(z, \partial G) < 1 \,\}.$$

Observe that $S \cap \{z_n\}$ must be either a finite set or a sequence tending to $\infty$. For in the contrary case, there would be infinitely many of these points confined to a bounded set. The definition of $S$ implies that these points would be bounded away from $\partial G$. The Bolzano–Weierstrass theorem then implies that these points would have a limit point inside $G$, contrary to hypothesis.

Consequently, the first version of the Weierstrass theorem implies that there is an *entire* function with zeroes precisely at the points of the sequence that lie in $S$. That entire function certainly is holomorphic on $G$.

All that remains is to construct a holomorphic function on $G$ that has zeroes at the points of the sequence lying in $T$. The product of this function with the entire function from the preceding paragraph solves the problem.

Accordingly, let $\{a_n\}$ denote the subsequence of points in the original sequence that happen to lie in $T$. If there are infinitely many such terms of the sequence, then $\operatorname{dist}(a_n, \partial G)$ must approach 0. If not, there would be a (further) subsequence bounded away from $\partial G$. The definition of $T$ implies that the subsequence would be bounded, hence would have a limit point inside $G$, contrary to hypothesis.

Now let $b_n$ be a point of $\partial G$ such that $|a_n - b_n| = \operatorname{dist}(a_n, \partial G)$. Let $E_n(z)$ denote the Weierstrass elementary factor

$$(1 - z)\exp\left(z + \tfrac{1}{2}z^2 + \cdots + \tfrac{1}{n}z^n\right).$$

The claim is that the following product provides the required holomorphic function:

$$\prod_{n=1}^{\infty} E_n\left(\frac{a_n - b_n}{z - b_n}\right).$$

The argument of $E_n$ takes the value 1 precisely when $z = a_n$, so the $n$th factor in the infinite product vanishes precisely when $z = a_n$. The singularity of the argument is on the boundary of $G$, so each factor is holomorphic inside $G$.

What remains to show is that the infinite product converges uniformly on compact subsets of $G$. When $z$ is confined to a compact set, then $z$ necessarily is bounded away from $\partial G$, so the denominator $z - b_n$ is bounded away from 0. On the other hand, $a_n - b_n \to 0$ by construction. Consequently, $|a_n - b_n|/|z - b_n| < 1/2$ for sufficiently large $n$, so the infinite product converges uniformly on compact sets. □

# 4 Approximation

The basic version of Runge's approximation theorem has already been stated and applied in § 2.4. The theorem is due to the German mathematician Carl Runge (1856–1927), a student of Weierstrass, in a paper in *Acta Mathematica* in 1885 (the same year that Weierstrass published his approximation theorem for continuous functions). Runge is known also for the Runge–Kutta method, a numerical method for finding approximate solutions of ordinary differential equations. [The second name is another German mathematician, Martin Wilhelm Kutta (1867–1944).]

**Theorem** (Runge's theorem for general compact sets)**.** *If $K$ is a compact subset of $\mathbb{C}$, and the function $f$ is holomorphic on a neighborhood of $K$, then $f$ is the uniform limit on $K$ of a sequence of rational functions with poles in the holes of $K$. Moreover, within each hole, the position of the pole can be prescribed arbitrarily.*

A *hole* in $K$ means a bounded component of $\mathbb{C} \setminus K$. If $K$ has no holes, then the approximation is by holomorphic polynomials.

## 4.1 Sketch of the proof of Runge's theorem

There are two main ideas in the proof: (i) approximate Cauchy's integral formula using Riemann sums and (ii) push the poles to new locations. Both ideas go back to Runge.

*Proof of step (i).* Suppose that $f$ is holomorphic in a neighborhood of a compact set $K$. The first step is to produce a cycle $\gamma$ (a union of finitely many oriented simple closed curves) contained in the open set where $f$ is holomorphic and having winding number 1 around each point of $K$. The precise meaning of the winding number of $\gamma$ around $z$ is

$$\frac{1}{2\pi i}\int_\gamma \frac{1}{w-z}\,dw.$$

If the compact set $K$ has a simple structure (a closed Jordan region, for example), then the existence of $\gamma$ is evident: draw a curve just outside the boundary of $K$. But if $K$ is a complicated set—perhaps a fractal set or a set having infinitely many components—then some work is needed to demonstrate the existence of $\gamma$ convincingly.

Suppose for the moment that $\gamma$ has been constructed. By Cauchy's integral formula,

$$f(z) = \frac{1}{2\pi i}\int_\gamma \frac{f(w)}{w-z}\,dw \qquad \text{when } z \in K.$$

Since $f$ is uniformly continuous on $\gamma$ (which is a compact set), and $w - z$ is bounded away from 0 when $z \in K$ and $w \in \gamma$, the integral can be approximated uniformly for such $z$ and $w$ by Riemann sums. These sums are linear combinations of rational functions of $z$ with first-order poles at certain points of $\gamma$ (the partition points). Thus $f$ is approximated uniformly on $K$ by rational functions with poles off $K$.

(The error in approximation by a Riemann sum depends on the modulus of continuity of the function. Evidently the dependence on $z$ is uniform when $z$ has distance from $\gamma$ bounded away from zero.)

The construction of $\gamma$ can be carried out in the following way. Cover the plane by a honeycomb mesh of regular hexagons having diameter less than the distance from $K$ to the boundary of the open set where the function $f$ is holomorphic. Collect all the hexagonal cells whose closures intersect $K$, and orient the boundaries counterclockwise. The claim is that $\gamma$ can be taken to consist of a subset of the oriented edges of these hexagons: namely, those edges that do not intersect $K$.

Observe that if $z$ is a point of $K$ not on any of the gridlines, then the sum of the Cauchy integrals of $f$ over the boundaries of all the closed hexagons that intersect $K$ equals $f(z)$, for $z$ is inside exactly one of these hexagons. If an edge of a hexagon intersects $K$, then there is an adjacent hexagon containing the same edge with opposite orientation. Cancelling the integrals over such edges shows that $f(z)$ equals the Cauchy integral of $f$ over $\gamma$ when $z$ is not on a gridline. By continuity, the integral still equals $f(z)$ even when $z$ is on a gridline.

Why is $\gamma$ a union of simple closed curves? Observe that at the terminal point of an arbitrary oriented edge of $\gamma$, there is a unique adjacent edge of $\gamma$. (This property is the reason that hexagons are more convenient than squares.) Only finitely many hexagons intersect the compact set $K$, so starting with an arbitrary edge of $\gamma$ and following successor edges eventually leads back to the starting point, generating a closed curve. If the edges of $\gamma$ have not been used up, then start over with a new edge and generate another closed curve. The process terminates after a finite number of steps. (Actually, the information that $\gamma$ is a union of closed curves is not really needed in the proof. Viewed merely as a union of edges, the arc $\gamma$ still gives an integral representing the function, and that integral can be approximated by Riemann sums.) □

*Proof of pole pushing.* The idea behind pole pushing is shown by the following calculation, in which $z_0$ and $z_1$ are two arbitrary distinct complex numbers:

$$\frac{1}{z-z_0} = \frac{1}{(z-z_1)-(z_0-z_1)} = \frac{1}{z-z_1}\cdot\frac{1}{1-\dfrac{z_0-z_1}{z-z_1}} = \frac{1}{z-z_1}\sum_{n=0}^{\infty}\left(\frac{z_0-z_1}{z-z_1}\right)^n,$$

with convergence when $|z_0 - z_1| < |z - z_1|$. Taking partial sums of the infinite series shows that the rational function $1/(z - z_0)$ can be approximated by rational functions with pole at $z_1$ when $z$ is farther away from $z_1$ than $z_0$ is. The convergence is even uniform on sets whose distance from $z_1$ is strictly greater than the distance from $z_1$ to $z_0$.

In particular, if $z_0$ is in a hole of the compact set $K$, then a rational function with pole at $z_0$ can be approximated uniformly on $K$ by rational functions with pole at an arbitrary point of the hole at slightly less than half the distance of $z_0$ to $K$. Iterating this observation shows that the pole can be pushed to an arbitrary location inside the hole.

What if $z_0$ is in the unbounded component of the complement of $K$? By the preceding reasoning, the pole can be pushed to an arbitrary location in the unbounded component. Suppose the pole has been pushed to a point $z_1$ outside a disk so large that the disk contains the compact set $K$. The Taylor series of $1/(z - z_1)$ about the center of the disk converges uniformly on $K$, and the partial sums of this series are polynomials. In other words, the pole in the unbounded component of the complement of $K$ can be pushed to infinity. □

## 4.2 Mergelyan's theorems

The hypothesis in Runge's theorem is somewhat unnatural: athough the approximation takes place only *on* the compact set, the function being approximated is assumed to be holomorphic in a *neighborhood* of the set. The following significant improvement of Runge's theorem is due to the Armenian mathematician Sergey Nikitovich Mergelyan (1928–2008). The main idea in the proof (not presented here) is to extend the function in a smooth way to a neighborhood of the compact set and to correct the extended function by solving a $\overline{\partial}$-problem (that is, solving the inhomogeneous Cauchy–Riemann equations). Then use Runge's theorem to approximate the extended function. The difficulty—which Mergelyan overcame—is to control the size of the correction term.

**Theorem** (Mergelyan's theorems, 1951–1952)**.** *1. If $K$ is a compact set with no holes, and $f$ is a continuous function on $K$ that is holomorphic on the interior of $K$, then there is a sequence $\{p_n\}$ of polynomials such that* $\max_{z\in K} |f(z) - p_n(z)| \to 0$.

2. *If $K$ is a compact set with finitely many holes, and $f$ is a continuous function on $K$ that is holomorphic on the interior of $K$, then $f$ is the uniform limit on $K$ of a sequence of rational functions with poles in the holes.*

3. *More generally, the conclusion holds in the case of infinitely many holes if the diameters of the holes are bounded away from zero.*

Here is an example of a compact set to which the third case applies:

$$\{ iy : |y| \leq 1 \} \cup \bigcup_{n=1}^{\infty} \{ n^{-1} + iy : |y| \leq 1 \} \cup \{ x + i : |x| \leq 1 \} \cup \{ x - i : |x| \leq 1 \}.$$

Each of the infinitely many holes has diameter greater than 2.

A counterexample with infinitely many shrinking holes was found by the Swiss mathematician Alice Roth (1905–1977) in her 1938 dissertation (prior to Mergelyan's work). The example is built by removing a suitable sequence of open disks from the closed unit disk, leaving a compact set with empty interior. Such a set is now called a "Swiss cheese" (which is the generic name in English for a pale-yellow cheese having many holes).

The case of infinitely many holes was definitively settled in 1966 by the famous blind Russian mathematician A. G. Vitushkin (1931–2004). He found a necessary and sufficient condition on the compact set (in terms of a notion called capacity) for rational approximation to hold.

## 4.3 Mittag-Leffler's theorem

Magnus Gustaf (Gösta) Mittag-Leffler (1846–1927), founder of *Acta Mathematica* (1882) and namesake of the Mittag-Leffler Institute in the suburbs of Stockholm, was a Swedish mathematician who attended some of Weierstrass's lectures and subsequently generalized the theorem of Weierstrass about constructing functions with prescribed zeroes. (Leffler was the father's name, and Mittag was the mother's name; Mittag-Leffler joined the names himself as an adult, apparently because of his interest in women's rights. His influence made it possible for Sonya Kovalevsky [1850–1891] to be appointed professor of mathematics in Stockholm.)

The theorem of Weierstrass says that there exists an entire function with prescribed zeroes (subject to the zeroes not accumulating). A consequence is the existence of meromorphic functions with prescribed poles (just take the reciprocal of a function with prescribed zeroes). Mittag-Leffler's main contribution was to construct functions not just with prescribed poles but with prescribed singular parts (also known as principal parts). Here is one version of the theorem.

**Theorem** (Mittag-Leffler's theorem, 1876–1884)**.** *Suppose $G$ is an open subset of $\mathbb{C}$ and $E$ is a discrete subset of $G$. Suppose given, for each point $b$ in $E$, a holomorphic function $p_b$ on a punctured neighborhood of $b$. Then there exists a holomorphic function $f$ on $G \setminus E$ such that for each point $b$ in $E$, the function $f - p_b$ has a removable singularity at $b$.*

In the statement of the theorem, a discrete set means a set that has no accumulation point in $G$.

A typical case is that $p_b$ is a finite linear combination of powers of $1/(z-b)$. The theorem guarantees that there exists a function with prescribed poles and prescribed principal parts at the singularities. In other words, there exists a meromorphic function with prescribed principal parts.

The theorem is more general, allowing essential singularities to be prescribed. For instance, the theorem produces a meromorphic function in the plane sharing the whole singular part of $\sin\left(\dfrac{1}{z-k}\right)$ at each integer $k$.

*Proof.* The standard modern proof uses Runge's theorem, although Runge's work actually was motivated by Mittag-Leffler's.

The first step is to exhaust $G$ by a sequence $\{K_n\}$ of compact sets such that each set is contained in the interior of the next, and no $K_n$ has unnecessary holes. In other words, each hole in the compact set contains a hole in $G$. This construction was done in § 2.2.

Each function $p_b$ has a Laurent series that converges in a punctured neighborhood of $b$. The part of the series involving powers of $(z-b)$ with nonnegative exponents is a Taylor series that converges in some disk centered at $b$. The part of the series involving powers of $(z-b)$ with negative exponents converges in $\mathbb{C}\setminus\{b\}$. Since the conclusion of the theorem addresses only the second part of the Laurent series, there is no loss of generality in replacing each $p_b$ by the singular part. In other words, there is no loss of generality in supposing that each $p_b$ is holomorphic on $\mathbb{C}\setminus\{b\}$.

The main idea in the proof of the theorem is to build the function as an infinite series. The first try is simply to add together all the functions $p_b$ as $b$ runs over the countable set $E$. This method certainly works when the set $E$ is finite. In general, there will be an infinite series, and this series need not converge. The idea is to add convergence factors that are holomorphic on all of $G$ and hence do not affect the singular parts.

To implement this idea, observe first that there are only finitely many singular points inside the compact set $K_1$. Add the corresponding functions $p_b$ together and call that sum $f_1$. In general, let $f_n$ denote the sum of the functions $p_b$ for $b$ in the set $K_n\setminus K_{n-1}$. Notice that when $n>1$, the function $f_n$ is holomorphic on $K_{n-1}$. Use Runge's theorem to find a rational function $g_n$ with poles in $\mathbb{C}\setminus G$ such that $\left|f_n(z)-g_n(z)\right|<1/2^n$ when $z\in K_{n-1}$.

The required function is $f_1+\sum_{n=2}^{\infty}(f_n-g_n)$. Indeed, if a compact set $K$ is fixed, then the terms in the tail of the series eventually are holomorphic on $K$, and the tail of the series converges uniformly on $K$ by comparison with a convergent geometric series. Hence the series represents a holomorphic function on $G\setminus E$. Moreover, at a particular point $b$ in $E$, all the terms of the sum are holomorphic in a neighborhood of $b$ except for one term, which by construction has the required singular part. $\square$

# 5 Hadamard's factorization theorem

The structure of the Weierstrass factorization of an entire function can be made more precise if some information is known about the growth of the function. But the question of rate of growth is more complicated for entire functions than for polynomials.

On a circle of large radius $r$, the absolute value of a polynomial of degree $n$ is comparable to $r^n$, uniformly with respect to the angle. An entire function, however, can have different growth rates in different directions. For example, the exponential function $e^z$ grows fast on the positive part of the real axis but decays to 0 on the negative part of the real axis. The simplest way to describe the growth of an entire function is to quantify the maximal rate of growth (independent of the angle).

## 5.1 Order

The *order* of an entire function $f$ is the greatest lower bound of the positive values of $t$ for which $f(z)\exp(-|z|^t)$ is a bounded function of $z$ in the plane. For example, the order of $ze^z$ is 1: the expression $ze^z e^{-|z|}$ is unbounded on the positive part of the real axis, but the expression $ze^z e^{-|z|^{1+\varepsilon}}$ is bounded for every positive $\varepsilon$. (This example shows that the infimum in the definition of order need not be attained.)

For an arbitrary positive $\varepsilon$, every constant times $|z|^t$ grows slower than $|z|^{t+\varepsilon}$ when $|z| \to \infty$, so an equivalent way to describe the order is the infimum of positive values of $t$ for which there exist constants $c_1$ and $c_2$ such that $|f(z)| \le c_1 \exp(c_2|z|^t)$ for every $z$. Only the growth rate is significant, so another equivalent description of the order is the infimum of positive values of $t$ for which $|f(z)| < \exp(|z|^t)$ when $|z|$ is sufficiently large.

*Exercise.* (A version of this exercise is posted, with hints; due March 24, 2016.) Write $M(r)$ for the maximum of $|f(z)|$ when $|z| \le r$. The following are equivalent definitions of the order of the entire function $f$. (Exclude the identically zero function, an uninteresting special case.)

- $\inf\left\{ t \in \mathbb{R} : |f(z)|\exp(-|z|^t) \text{ is a bounded function of } z \text{ in } \mathbb{C} \right\}$.
- $\inf\{ t \in \mathbb{R} : \lim_{r\to\infty} r^{-t}\log M(r) = 0 \}$.
- $\limsup_{r\to\infty} \dfrac{\log\log M(r)}{\log r}$.
- $\limsup_{n\to\infty} \dfrac{n\log n}{\log \frac{1}{|c_n|}}$, where $\sum_{n=0}^{\infty} c_n z^n$ is the Maclaurin series of $f(z)$.

*Example.* If $f$ is a constant function, then the order of $f$ equals 0. More generally, the order of every polynomial is 0. Indeed, a polynomial of degree $n$ grows like a constant times $|z|^n$, which is a slower growth rate than $\exp(|z|^\varepsilon)$ for an arbitrary positive $\varepsilon$.

*Example.* If $g$ is a polynomial of degree $k$, then the order of $e^g$ is $k$. Indeed, $\log M(r)$ equals the maximum of $\operatorname{Re} g(z)$ when $|z| = r$. On a circle of large radius $r$, the function $g$ is essentially a power function wrapping the circle $k$ times around an image curve that is nearly a circle of radius comparable to $r^k$, so the maximum of $\operatorname{Re} g(z)$ is comparable to $r^k$.

*Example.* The function $e^{e^z}$ is an entire function of infinite order. Indeed, $\log M(r) = e^r$, so there is no power $t$ such that $r^{-t}\log M(r)$ tends to 0 when $r$ tends to infinity.

**Lemma** (Generalization of Liouville's theorem)**.** *If an entire function grows no faster than a polynomial, then the entire function* is *a polynomial. More precisely, if $f$ is entire, and there* *exist nonnegative constants $A$, $B$, and $C$ such that $|f(z)| \le A + B|z|^C$ for every $z$, then $f$ is a polynomial, and the degree is no larger than $C$.*

*Proof.* (The statement is essentially Exercise 1 in §3 of Chapter IV in the textbook.) Let $k$ be the unique integer such that $k \le C < k+1$, let $R$ be a large radius, and let $z$ be an arbitrary point in the unit disk. Cauchy's formula for derivatives implies that

$$\begin{aligned}
|f^{(k+1)}(z)| &= \left|\frac{(k+1)!}{2\pi i}\int_{|w|=R}\frac{f(w)}{(w-z)^{k+2}}\,dw\right| \\
&\le \frac{(k+1)!}{2\pi}\int_0^{2\pi}\frac{A+BR^C}{(R-1)^{k+2}}R\,d\theta \\
&= \frac{(k+1)!\,R(A+BR^C)}{(R-1)^{k+2}}.
\end{aligned}$$

Letting $R$ go to infinity shows that $f^{(k+1)}$ is identically equal to 0 in the unit disk, hence in the whole plane. Therefore $f$ is a polynomial of degree at most $k$. □

*Remark.* The hypothesis can be weakened. The proof needs the growth bound on $f$ not for every radius $R$ but merely for a sequence of values of $R$ tending to infinity. This improvement of the lemma will be used in § 5.7.

*Example.* There exist nonpolynomial entire functions of order 0. One example is the infinite product $\prod_{n=1}^{\infty}\left(1-\frac{z}{n!}\right)$ from the midterm examination.

To see why this function has order 0, fix a positive $\varepsilon$, and let $C_\varepsilon$ denote the sum of the convergent infinite series $\sum_{n=1}^{\infty} n^{1/\varepsilon}/n!$ (either the root test or the ratio test establishes the convergence). When $|z| = r$, an upper bound for the absolute value of the infinite product is

$$\prod_{1\le n\le r^\varepsilon}\left(1+\frac{r}{n!}\right)\prod_{n>r^\varepsilon}\left(1+\frac{r}{n!}\right).$$

The first factor is bounded above by $(1+r)^{r^\varepsilon}$. The following chain of inequalities shows that the second factor is bounded:

$$\prod_{n>r^\varepsilon}\left(1+\frac{r}{n!}\right) < \prod_{n>r^\varepsilon}\exp\frac{r}{n!} < \prod_{n>r^\varepsilon}\exp\frac{n^{1/\varepsilon}}{n!} < \exp\sum_{n=1}^{\infty}\frac{n^{1/\varepsilon}}{n!} = \exp C_\varepsilon.$$

Therefore

$$r^{-2\varepsilon}\log M(r) \le C_\varepsilon r^{-2\varepsilon} + r^{-\varepsilon}\log(1+r),$$

and the right-hand side tends to 0 when $r$ goes to infinity. Since $\varepsilon$ is an arbitrary positive number, the order of the entire function is 0.

*Example.* The expression $\cos\sqrt{z}$ apparently does not define an entire function, for $\sqrt{z}$ is not analytic in a neighborhood of the origin. But the cosine is an even function, so the Maclaurin series involves only even powers, and halving each power produces an entire function that can reasonably be called $\cos\sqrt{z}$. This function has order $1/2$. Thus the order need not be a natural number. (The exercise above, which relates the order to the Maclaurin coefficients, reveals that there exist entire functions whose order is any prescribed nonnegative real number.)

*Remark.* The function $\cos\sqrt{z}$ has zeroes roughly at the squares of integers and grows roughly like $e^{|z|^{1/2}}$. The function $\cos z$ has zeroes roughly at the integers and grows roughly like $e^{|z|}$. Thus a greater concentration of zeroes corresponds to a larger rate of growth. This observation is a general property. Although paradoxical at first sight, the property corresponds with your knowledge that a polynomial having a lot of zeroes has high degree and hence grows fast.

## 5.2 Statement of the theorem

**Theorem** (Hadamard's factorization theorem)**.** *If $f$ is an entire function (not identically zero) of order $\rho$, then $f(z)$ can be expressed in the form*

$$z^m e^{g(z)} \prod_{n=1}^{\infty} \left(1 - \frac{z}{z_n}\right) \exp\left(\frac{z}{z_n} + \frac{1}{2}\left(\frac{z}{z_n}\right)^2 + \cdots + \frac{1}{k}\left(\frac{z}{z_n}\right)^k\right),$$

*where $m$ is the order of the zero of $f$ at the origin (possibly $m = 0$), the function $g$ is a polynomial of degree no larger than $\rho$, the natural number $k$ is no larger than $\rho$ (if $k = 0$, then the exponential factor is absent), and the numbers $z_n$ are the zeroes of $f$ at points other than the origin, repeated according to multiplicity. If $f$ has only finitely many zeroes, then the factorization reduces to a polynomial times $e^{g(z)}$.*

*Example.* Show that the function $z + e^z$ has infinitely many zeroes.

*Solution.* If not, then there is a polynomial $p(z)$ and a constant $b$ such that $z + e^z = p(z)e^{bz}$ (since $z + e^z$ has order 1). Then

$$ze^{-z} + 1 = p(z)e^{(b-1)z}.$$

When $z \to \infty$ along the positive real axis, the left-hand side approaches the limit 1. The only way the right-hand side can have limit equal to 1 is for $b$ to equal 1 and $p$ to be the constant polynomial 1. But then $ze^{-z}$ on the left-hand side must be identically 0, which is absurd. Accordingly, the expression $z + e^z$ must have infinitely many zeroes, so that $p(z)$ is replaced by an infinite product. □

*Example.* An entire function of finite, non-integer order takes every complex value infinitely often. (In other words, there is no Picard exceptional value.) This statement is Corollary 3.8 in Chapter XI of the textbook.

*Proof.* Seeking a contradiction, suppose that such a function $f$ takes the value $c$ only a finite number of times. Hadamard's theorem implies that $f(z) - c$ can be written as a polynomial $p(z)$ times $e^{g(z)}$ for some polynomial $g$. But the order of $c + p(z)e^{g(z)}$ evidently is an integer. □

The preceding examples illustrate the power of Hadamard's factorization theorem. The proof requires some preparation. Before diving into the proof, seeing the classical statement of the theorem is useful.

## 5.3 Genus

Suppose $\{z_n\}_{n=1}^{\infty}$ is the sequence of nonzero zeroes of an entire function $f$ (not identically zero), where multiple zeroes are repeated in the sequence according to multiplicity. Estimates on the Maclaurin series of $\log(1-z)$ from § 3.2 imply that if $\sum_{n=1}^{\infty} 1/|z_n|^{k+1} < \infty$, then $\prod_{n=1}^{\infty} E_k(z/z_n)$ converges, where $E_k(w) = (1-w)\exp(w + \frac{1}{2}w^2 + \cdots + \frac{1}{k}w^k)$ if $k$ is a positive integer, and $E_0(w) = 1 - w$. If there exists some such nonnegative integer $k$, then there is a smallest one, called the *rank* of $f$ (or the rank of the sequence $\{z_n\}$ of zeroes).

If $f$ has finite rank $k$, then dividing $f(z)$ by the convergent infinite product $\prod_{n=1}^{\infty} E_k(z/z_n)$ yields an entire function whose only possible zero is at the origin. Further dividing by a suitable power $z^m$ gives a zero-free entire function, and such a function can be written in the form $e^g$ for some other entire function $g$. If $g$ is a polynomial, then the *genus* of $f$ is the maximum of the rank of $f$ and the degree of $g$. Notice that the genus is necessarily an integer, in contrast to the order.

**Theorem** (Classical statement of Hadamard's theorem)**.** *The order $\rho$ of an entire function is finite if and only if the genus $\mu$ is finite, and*

$$\mu \le \rho \le \mu + 1.$$

*Proof of the right-hand inequality.* The order of a product of two entire functions does not exceed the maximum of the orders of the two factors. Indeed, if $|f_1(z)| < \exp(|z|^{\rho_1+\varepsilon})$ when $|z|$ is sufficiently large, and $|f_2(z)| < \exp(|z|^{\rho_2+\varepsilon})$ when $|z|$ is sufficiently large, then $|f_1(z)f_2(z)| < \exp(2|z|^{\max(\rho_1,\rho_2)+\varepsilon}) < \exp(|z|^{\max(\rho_1,\rho_2)+2\varepsilon})$ when $|z|$ is sufficiently large.

Therefore the factor $z^m e^{g(z)}$ and the infinite product $\prod_{n=1}^{\infty} E_k(z/z_n)$ can be discussed separately. If $g$ is a polynomial of degree $k$, then the order of $e^g$ is $k$. But $z^m$ has order 0, so the product $z^m e^{g(z)}$ has order not exceeding $k$ (in fact, equal to $k$). What remains is to show that an infinite product $\prod_{n=1}^{\infty} E_k(z/z_n)$ of rank $k$ has order no larger than $k+1$. That conclusion is part of the following more precise lemma. □

Define the cutoff convergence exponent $\sigma$ of a sequence $\{z_n\}_{n=1}^{\infty}$ of nonzero complex numbers to be the infimum of real values of $t$ for which $\sum_{n=1}^{\infty} 1/|z_n|^t$ converges. If $\sigma$ is not an integer, then the rank $k$ is the unique integer such that $k < \sigma < k+1$. If $\sigma$ is an integer, then the rank $k$ has the property that $k \le \sigma \le k+1$. Equality holds in the left-hand inequality when the cutoff is not attained, and equality holds in the right-hand inequality when the cutoff is attained.

**Lemma.** *The cutoff convergence exponent of a sequence* $\{z_n\}_{n=1}^{\infty}$ *of rank* $k$ *is equal to the order of the infinite product* $\prod_{n=1}^{\infty} E_k(z/z_n)$.

The lemma implies that the double inequality in Hadamard's theorem cannot be improved: equality can occur on either side. Indeed, the function $\exp(z^k)$ has genus $k$ and order $k$; the convergent infinite product

$$\prod_{n=2}^{\infty}\left(1-\frac{z}{n(\log n)^2}\right)$$

has rank 0 (also genus 0, since there is no factor $e^g$) and order 1, since the cutoff convergence exponent of 1 is attained.

The immediately relevant part of the lemma—that the order of the infinite product is less than or equal to the cutoff convergence exponent—can be proved now. On the other hand, establishing the lower bound for the order requires a tool to be developed in the next section.

*Proof of half the lemma.* If the cutoff convergence exponent $\sigma$ is strictly less than $k+1$, then let $\varepsilon$ be an arbitrary positive number such that $\sigma+\varepsilon \le k+1$. If $\sigma = k+1$, then let $\varepsilon$ be 0. In either case, the series $\sum_{n=1}^{\infty} 1/|z_n|^{\sigma+\varepsilon}$ converges by hypothesis. The goal is to bound the absolute value of the infinite product $\prod_{n=1}^{\infty} E_k(z/z_n)$ from above by the exponential of a constant times $|z|^{\sigma+\varepsilon}$.

First suppose that $k \ge 1$. When $|w| \le 1/2$, the lemma in § 3.2 implies that $|E_k(w)| \le \exp(|w|^{k+1}) \le \exp(|w|^{\sigma+\varepsilon})$. If $|w| > 1/2$, then

$$\left|w+\frac{1}{2}w^2+\cdots+\frac{1}{k}w^k\right| \le \left(2^{k-1}+2^{k-2}+\cdots+1\right)|w|^k < 2^k|w|^k.$$

Therefore

$$\exp(-2^k|w|^k) < \left|\exp\left(w+\frac{1}{2}w^2+\cdots+\frac{1}{k}w^k\right)\right| < \exp(2^k|w|^k) \quad \text{when } |w|>1/2.$$

The right-hand side of this double inequality is of interest now, and the left-hand side will be invoked in § 5.7. The assumptions that $|w| > 1/2$ and $k \ge 1$ additionally imply that

$$|1-w| \le 1+|w| < \exp(|w|) \le \exp(2^{k-1}|w|^k);$$

but $2^k + 2^{k-1} < 2^{k+1}$, so $|E_k(w)| < \exp(2^{k+1}|w|^k) < \exp(2^{\sigma+\varepsilon+1}|w|^{\sigma+\varepsilon})$ when $|w| > 1/2$. In summary, if $k \ge 1$, then the absolute value of $E_k(w)$ is bounded above by the exponential of a constant times $|w|^{\sigma+\varepsilon}$ for every $w$, whether $|w| \le 1/2$ or $|w| > 1/2$.

If $k = 0$, then $E_k(w)$ reduces to $1-w$. If $|w| \le 1$, then

$$|1-w| \le 1+|w| \le \exp(|w|) \le \exp(|w|^{\sigma+\varepsilon}).$$

Let $N$ be the least integer greater than or equal to $1/(\sigma+\varepsilon)$. If $|w| > 1$, then

$$\exp(N|w|^{\sigma+\varepsilon}) = 1+\sum_{n=1}^{\infty}\frac{(N|w|^{\sigma+\varepsilon})^n}{n!} > 1+\frac{(N|w|^{\sigma+\varepsilon})^N}{N!} > 1+|w| \ge |E_0(w)|.$$

Therefore $|E_0(w)| \le \exp(N|w|^{\sigma+\varepsilon})$ for every $w$, whether $|w| \le 1$ or $|w| > 1$.

The two preceding paragraphs show that whatever the value of the rank, there is a positive number $c$ depending only on $\sigma + \varepsilon$ (the maximum of $2^{\sigma+\varepsilon+1}$ and $\lceil 1/(\sigma+\varepsilon) \rceil$ will do) such that $|E_k(w)| \le \exp(c|w|^{\sigma+\varepsilon})$ for every $w$. Accordingly,

$$\left| \prod_{n=1}^{\infty} E_k(z/z_n) \right| \le \exp\left( c|z|^{\sigma+\varepsilon} \sum_{n=1}^{\infty} \frac{1}{|z_n|^{\sigma+\varepsilon}} \right).$$

By hypothesis, the series $\sum_{n=1}^{\infty} 1/|z_n|^{\sigma+\varepsilon}$ converges. Thus the absolute value of the infinite product is bounded above by the exponential of a constant times $|z|^{\sigma+\varepsilon}$, so the order $\rho$ does not exceed $\sigma + \varepsilon$. Letting $\varepsilon$ go to 0 shows that the order of the infinite product is no larger than the cutoff convergence exponent $\sigma$. $\square$

The other half of the lemma, that the order of the infinite product is no smaller than the cutoff convergence exponent $\sigma$, follows from Jensen's formula, the next topic.

## 5.4 Jensen's formula

The harder half of Hadamard's theorem is the statement that the genus is no larger than the order. The key tool needed in the proof is a quantitative estimate showing that the growth rate of an entire function controls the number of zeroes of the function in large disks.

**Theorem** (Jensen's formula)**.** *Suppose $f$ is holomorphic in a neighborhood of the closed disk $\overline{B}(0;r)$, and let $z_1, \ldots, z_n$ be the zeroes of $f$ in the open disk $B(0;r)$, each zero repeated according to its multiplicity. If $f(0) \ne 0$, then*

$$\log|f(0)| + \sum_{k=1}^{n} \log \frac{r}{|z_k|} = \frac{1}{2\pi} \int_0^{2\pi} \log|f(re^{i\theta})|\, d\theta.$$

The theorem is due to the Danish mathematician Johan Jensen (1859–1925). Actually his profession was telephone engineer, a job that he took to support himself while he pursued his love of mathematics. Thus he was an amateur mathematician. He found his formula while trying unsuccessfully to prove the Riemann hypothesis. He is known in real analysis and probability for Jensen's inequality (about convex functions).

You can easily adjust the formula in case $f(0) = 0$. If $f$ has a zero of order $m$ at the origin, then apply the formula to $f(z)/z^m$. See Exercise 1 in §1 of Chapter XI. The following statement is essentially Exercise 2 in the same section.

**Corollary.** *If $f$ is entire, normalized such that $f(0) = 1$, and $n(r)$ denotes the number of zeroes of $f$ in $\overline{B}(0;r)$, counted according to multiplicity, then $n(r)\log(1+c) \le \log M((1+c)r)$ for each positive $c$ and each positive radius $r$. In particular, taking $c$ equal to 2 shows that $n(r) < M(3r)$.*

Again, the condition that $f(0) = 1$ is merely a convenient normalization. If $f(0)$ is an arbitrary nonzero value, then apply the result to $f(z)/f(0)$. If $f$ has a zero of order $m$ at the origin, then apply the result to $f(z)/z^m$. The details of the formula change, but the essence remains: namely, control on the modulus of the entire function on a disk gives control on the number of zeroes in a smaller disk.

*Example.* If $f(z) = \cos(z)$, then $M(3r) = \frac{1}{2}(e^{3r} + e^{-3r})$, so $\log M(3r) \approx 3r$. On the other hand, $n(r) \approx 2r/\pi$. This example shows that the corollary is sensible, and the inequality cannot be improved by much.

*Proof of Corollary.* Jensen's formula [with $r$ replaced by $(1 + c)r$] implies that

$$\begin{aligned}\log M((1+c)r) &\ge \frac{1}{2\pi}\int_0^{2\pi} \log|f((1+c)re^{i\theta})|\,d\theta = \sum_{|z_k|<(1+c)r} \log\frac{(1+c)r}{|z_k|}\\ &\ge \sum_{|z_k|\le r} \log\frac{(1+c)r}{|z_k|} \ge n(r)\log(1+c). \qquad \square\end{aligned}$$

*Proof of Jensen's formula.* The proof will be built up through a sequence of steps. If $f$ has no zeroes on the closed disk, then there is a holomorphic branch of $\log f$, and the formula is simply the real part of the mean-value property of $\log f$.

Next suppose that $f$ has some zeroes inside the disk $B(0; r)$ but none on the boundary. The zeroes inside the disk can be canceled by multiplying by a suitable linear fractional transformation (a Blaschke product) that has modulus equal to 1 on the boundary of $B(0; r)$. Namely, if $|z_k| < r$, then $|z_k/r| < 1$, so the Möbius transformation

$$\frac{\frac{z_k}{r} - \frac{z}{r}}{1 - \frac{\overline{z_k}}{r}\frac{z}{r}} \qquad \text{or} \qquad \frac{r(z_k - z)}{r^2 - \overline{z_k}z}$$

has a zero when $z = z_k$ and has modulus 1 when $z$ is on the boundary of $B(0; r)$. Set $F(z)$ equal to

$$f(z)\Big/\prod_{k=1}^{n} \frac{r(z_k - z)}{r^2 - \overline{z_k}z}.$$

Then $|F(z)| = |f(z)|$ when $|z| = r$, and the singularities of $F$ are removable, so applying the preceding case shows that

$$\begin{aligned}\frac{1}{2\pi}\int_0^{2\pi} \log|f(re^{i\theta})|\,d\theta &= \frac{1}{2\pi}\int_0^{2\pi} \log|F(re^{i\theta})|\,d\theta = \log|F(0)|\\ &= \log|f(0)| - \sum_{k=1}^{n} \log\frac{|z_k|}{r},\end{aligned}$$

as required.

Finally, suppose that some of the zeroes of $f$ lie on the boundary of $B(0;r)$. This case is not needed in applications, since making a small perturbation of the radius reduces to the preceding case, but pinning down the general case is interesting for completeness. Notice that if $a$ is a zero on the boundary, then $\log(|a|/r) = 0$. Seemingly, the result should follow from a suitable convergence theorem for integrals as the radius approaches $r$. But the situation does not fit any of the standard convergence theorems!

Here is a different argument. To get started, suppose there is just one zero on the boundary, say at $a$. The preceding proof applies to $f(z)/(z-a)$. Consequently, what needs to be shown to verify the validity of Jensen's formula when a zero lies on the boundary is that

$$\frac{1}{2\pi}\int_0^{2\pi} \log|re^{i\theta} - a|\,d\theta = \log|a|,$$

or, equivalently,

$$\frac{1}{2\pi}\int_0^{2\pi} \log|1 - (r/a)e^{i\theta}|\,d\theta = 0.$$

Now $r/a$ is a complex number of modulus equal to 1, so adding a constant to the integration variable and using periodicity of the exponential function reduces the problem to showing that

$$\int_0^{2\pi} \log|1 - e^{i\theta}|\,d\theta = 0.$$

An equivalent formulation of this statement is that

$$\operatorname{Re}\int_{|z|=1} \log(1-z)\,\frac{dz}{iz} = 0.$$

Replacing the integration path by one that avoids the singularity at 1 by bumping inward with a small semi-circle of radius $\delta$ makes an error of order $\delta\log\delta$, which tends to 0 with $\delta$. But Cauchy's integral formula applies on the bumped contour and evaluates the integral as $2\pi\log 1$, or 0. Multiple zeroes on the boundary can be handled the same way. □

## 5.5 Application of Jensen's formula to the rank

*Proof that the rank does not exceed the order.* Suppose the entire function $f$ has finite order $\rho$. There is no loss of generality in supposing that $f(0) = 1$, for dividing $f$ by a constant or even by a constant times a power of $z$ changes neither the order of the entire function nor the rank. Let $\{z_n\}_{n=1}^{\infty}$ be the sequence of zeroes of $f$ (repeated according to multiplicity) arranged in order of increasing (or nondecreasing) modulus.

The plan is to show that the cutoff convergence exponent for the sequence $\{z_n\}$ is no larger than the order $\rho$. That conclusion completes the proof of the lemma from § 5.3 and simultaneously shows that the rank does not exceed the order.

Suppose that $\lambda$ is an arbitrary real number strictly larger than $\rho$, and fix a positive number $\beta$ such that $\rho + 2\beta < \lambda$. The corollary to Jensen's formula implies that

$$j \le n(|z_j|) < \log M(3|z_j|).$$

Since $|z_j| \to \infty$ when $j \to \infty$, the definition of order implies that if $j$ is sufficiently large, then $\log M(3|z_j|) < (3|z_j|)^{\rho+\beta}$. If additionally $j$ is so large that $3^{\rho+\beta} < |z_j|^{\beta}$, then

$$\log M(3|z_j|) < |z_j|^{\rho+2\beta}.$$

The two preceding inequalities imply that $j < |z_j|^{\rho+2\beta}$ when $j$ is sufficiently large, so

$$\frac{1}{|z_j|} < \left(\frac{1}{j}\right)^{1/(\rho+2\beta)}, \qquad \text{and} \qquad \frac{1}{|z_j|^{\lambda}} < \left(\frac{1}{j}\right)^{\lambda/(\rho+2\beta)}.$$

Since $\rho + 2\beta < \lambda$, the series $\displaystyle\sum_j \frac{1}{|z_j|^{\lambda}}$ converges by the comparison test.

Thus the cutoff convergence exponent $\sigma$ is less than or equal to $\lambda$. But $\lambda$ is an arbitrary number larger than $\rho$, so $\sigma \le \rho$. If $k$ is the rank, then the series $\displaystyle\sum_j \frac{1}{|z_j|^{k}}$ diverges, so $k \le \sigma$. □

The remaining part of Hadamard's factorization theorem is that if an exponential factor $e^g$ is present, then $g$ is a polynomial, and the degree of $g$ does not exceed the order of the entire function. One way to get this conclusion is to apply the so-called Poisson–Jensen formula (see §1.3 in Chapter XI of the textbook), which is obtained by composing Jensen's formula with a linear fractional transformation. The following proof, different from the one in the textbook, is based on another famous and useful inequality.

## 5.6 Carathéodory's inequality

If you control the size of the real part of a holomorphic function, do you control the size of the function? Evidently not, for adding $100i$ to the function does not affect the real part. A way to avoid this problem is to ask additionally for control on the whole function at one point. Another difficulty is illustrated by the holomorphic function $-i\log(1-z)$ on the open unit disk: the real part is $\arg(1-z)$, which is bounded, but the imaginary part is $\ln(1/|1-z|)$, which is unbounded near the boundary point 1. A way to avoid this problem is to stay away from the boundary of the domain.

Carathéodory's inequality (which is Exercise 4 in §2 of Chapter VI in the textbook) can be interpreted as saying that the two indicated problems are the only ones that can arise. One version of the inequality (a little stronger than the statement in the textbook) says that if $f$ is holomorphic on a disk $B(0; R)$, and the real part of $f$ is bounded above by $A$, and $0 < r < R$, then

$$M(r) \le \frac{2r}{R-r}A + \frac{R+r}{R-r}\,|f(0)|,$$

where $M(r)$ is the maximum of $|f(z)|$ for $z$ in the closed disk $\overline{B}(0, r)$. Notice that the number $A$ is allowed to be negative, being an upper bound for the real part (not for the absolute value of the real part).

*Proof.* First consider the case that $f(0) = 0$. By hypothesis, the function $f$ maps into the half-plane $\{ z \in \mathbb{C} : \operatorname{Re} z \leq A \}$, and $A \geq 0$. Suppose, without loss of generality, that $f$ is not constant (for the inequality is trivial for constant functions). By the open mapping theorem, the image of $f$ must lie in the open half-plane, and the number $A$ must be strictly positive.

Let $\varphi(w)$ denote $w/(2A - w)$. This linear fractional transformation takes $A$ to 1 and $\infty$ to $-1$ and 0 to 0, so $\varphi$ maps the indicated half-plane to the unit disk. The composite function $\varphi \circ f$ then maps the disk of radius $R$ to the unit disk, fixing the origin. The Schwarz lemma, adjusted to a disk of radius $R$, implies that $|\varphi \circ f(z)| \leq |z|/R$ for every $z$ in the disk of radius $R$. The required inequality for $f$ follows by observing that $f = \varphi^{-1} \circ (\varphi \circ f)$, and $\varphi^{-1}(w) = 2Aw/(1 + w)$, so

$$f(z) = 2A \frac{\varphi \circ f(z)}{1 + \varphi \circ f(z)}, \qquad \text{whence} \qquad M(r) \leq 2A \frac{r/R}{1 - (r/R)} = \frac{2r}{R - r} A.$$

To obtain the general case, when $f(0)$ is not necessarily equal to 0, apply the special case to the function $f(z) - f(0)$. Then the number $A$ changes to $A - \operatorname{Re} f(0)$, and

$$|f(z) - f(0)| \leq \frac{2r}{R - r} (A - \operatorname{Re} f(0)) \qquad \text{when } |z| \leq r.$$

Hence

$$M(r) \leq |f(0)| + \frac{2r}{R - r} A + \frac{2r}{R - r} |f(0)| = \frac{2r}{R - r} A + \frac{R + r}{R - r} |f(0)|,$$

as claimed. □

A consequence of Carathéodory's inequality is that if the real part of an entire function grows no faster than a polynomial, then the entire function is a polynomial. This statement generalizes the lemma from § 5.1, itself a generalization of Liouville's theorem. Here is a precise formulation.

**Lemma.** *If $f$ is entire, and there exist nonnegative constants $A$, $B$, and $C$ such that* $\operatorname{Re} f(z) \leq A + B|z|^C$ *for every $z$, then $f$ is a polynomial, and the degree is no larger than $C$.*

*Proof.* Applying Carathéodory's inequality with $R$ equal to $2r$ shows that

$$M(r) \leq 2(A + B(2r)^C) + 3|f(0)|.$$

Thus $|f(z)|$ inherits a growth estimate of the same form as the estimate satisfied by $\operatorname{Re} f(z)$, but with different constants. The conclusion now follows from the lemma in § 5.1. □

## 5.7 Conclusion of the proof of Hadamard's factorization theorem

Suppose $f$ has order $\rho$ and rank $k$, and

$$f(z) = z^m e^{g(z)} \prod_{n=1}^{\infty} \left(1 - \frac{z}{z_n}\right) \exp\left(\frac{z}{z_n} + \dots + \frac{1}{k}\left(\frac{z}{z_n}\right)^k\right).$$

What remains to show is that $g$ must be a polynomial, moreover a polynomial of degree at most $\rho$.

The problem is how to leverage the growth estimate on $f$ to obtain information about $g$. The proof in the textbook takes the logarithmic derivative to extract $g$ from the exponent. The following proof takes the more intuitive approach of direct size estimates.

To get started, suppose that $f$ has no zeroes at all, that is, $f(z) = e^{g(z)}$. By hypothesis, for every positive $\varepsilon$ there is a constant $C_\varepsilon$ such that $|f(z)| \le C_\varepsilon \exp(|z|^{\rho+\varepsilon})$ for every $z$. But $|\exp(g(z))| = \exp(\operatorname{Re} g(z))$, so the hypothesis implies that $\operatorname{Re}(g(z)) \le \log(C_\varepsilon) + |z|^{\rho+\varepsilon}$. The lemma in § 5.6 implies that $g$ is a polynomial of degree at most $\rho + \varepsilon$. Since $\varepsilon$ is arbitrary, the degree of $g$ is at most $\rho$.

If $f$ has some zeroes, but only finitely many, then $f(z) = p(z)e^{g(z)}$ for some polynomial $p$. Since $|p(z)| \to \infty$ when $|z| \to \infty$, the order of the entire function $f(z)/p(z)$ is no greater than the order of $f$ (actually equal to the order of $f$), so the preceding argument still applies to show that $g$ is a polynomial of degree at most $\rho$.

The difficult case occurs when $f$ has infinitely many zeroes. To obtain an upper bound on $|e^g|$ from an upper bound on $|f|$ apparently requires a lower bound on the modulus of the convergent infinite product $\prod_{n=1}^{\infty} E_k(z/z_n)$, at least for values of $z$ of large modulus. But the infinite product has zeroes of arbitrarily large modulus, so this approach appears doubtful. Nonetheless, the growth of $e^g$ can indeed be controlled by the growth of $f$ even in the presence of infinitely many zeroes, as the following argument shows.

Fix a positive $\varepsilon$. The plan is to demonstrate the existence of a constant $C_\varepsilon$ such that if $|z| > 1$ and also $|z - z_n| > 1/|z_n|^{k+1}$ for every $n$, then the infinite product $\prod_{n=1}^{\infty} E_k(z/z_n)$ has modulus no smaller than $C_\varepsilon \exp(-|z|^{\rho+2\varepsilon})$. By the definition of rank, the sum of the radii of the excluded disks is a convergent series, so there are arbitrarily large circles on which this lower bound for the magnitude of the infinite product is valid. But $|f(z)/z^m|$ is bounded above by $\exp(|z|^{\rho+\varepsilon})$ on large circles, so combining the upper bound for $|f|$ with the lower bound for the infinite product shows that $|e^{g(z)}|$ is bounded above by a constant times $\exp(|z|^{\rho+3\varepsilon})$ on a sequence of circles with radii tending to infinity. The resulting polynomial growth bound on $\operatorname{Re} g$ combines with Carathéodory's inequality and the remark in § 5.1 to show that the entire function $g$ is a polynomial of degree at most $\rho + 3\varepsilon$. Since $\varepsilon$ can be arbitrarily close to zero, the degree of $g$ is at most $\rho$.

To obtain the required lower bound on the infinite product when $1 < |z| = r$, split the factors into two subsets, depending on whether $|z_n| < 2r$ or $2r \le |z_n|$. For the second case, apply the lemma in § 3.2 to deduce that

$$|E_k(z/z_n)| \ge \exp(-|z/z_n|^{k+1}) \qquad \text{when } |z/z_n| \le 1/2 \text{ and } k \ge 1.$$

When $k = 0$, observe that $\operatorname{Re}\log(1 - w) = \operatorname{Re}[w + \log(1 - w)] - \operatorname{Re} w \geq -|w|^2 - |w| \geq -2|w|$ when $|w| \leq 1/2$, so $|E_0(z/z_n)| \geq \exp(-2|z/z_n|)$ when $|z/z_n| \leq 1/2$. Thus for every $k$, whether 0 or a positive integer, $|E_k(z/z_n)| \geq \exp(-2|z/z_n|^{k+1})$ when $|z/z_n| \leq 1/2$. If $\rho + \varepsilon \leq k + 1$, then $|z/z_n|^{k+1} \leq |z/z_n|^{\rho+\varepsilon}$ when $|z/z_n| \leq 1/2$, so

$$\left|\prod_{2r\leq|z_n|} E_k(z/z_n)\right| \geq \exp\left(-2r^{\rho+\varepsilon} \sum_{n=1}^{\infty} 1/|z_n|^{\rho+\varepsilon}\right).$$

The infinite series in the exponent converges (since $\rho$ equals the cutoff convergence exponent), so the indicated part of the infinite product is bounded below by the exponential of a negative constant times $r^{\rho+\varepsilon}$. On the other hand, if $\rho + \varepsilon > k + 1$, then

$$\left|\prod_{2r\leq|z_n|} E_k(z/z_n)\right| \geq \exp\left(-2r^{k+1} \sum_{n=1}^{\infty} 1/|z_n|^{k+1}\right) \geq \exp\left(-2r^{\rho+\varepsilon} \sum_{n=1}^{\infty} 1/|z_n|^{k+1}\right)$$

when $r > 1$, so again this part of the infinite product is bounded below by the exponential of a negative constant times $r^{\rho+\varepsilon}$. What about the other part of the infinite product?

If $|z_n| < 2|z| = 2r$, and $|z - z_n| > 1/|z_n|^{k+1}$, then

$$\left|1 - \frac{z}{z_n}\right| > \frac{1}{|z_n|^{k+2}} \geq \frac{1}{(2r)^{k+2}}.$$

Since $2r > 1$, each such lower bound is less than 1, and

$$\prod_{|z_n|<2r} \left|1 - \frac{z}{z_n}\right| \geq \frac{1}{(2r)^{(k+2)n(2r)}}.$$

Now $n(2r)$ is bounded above by $\log M(6r)$ by the corollary to Jensen's formula from § 5.4, and $f$ has order $\rho$, so the preceding expression is bounded below by the exponential of a negative constant times $r^{\rho+\varepsilon}$ when $r$ is sufficiently large. The lower bound from the proof of the lemma in § 5.3 shows that

$$\left|\exp\left[\frac{z}{z_n} + \frac{1}{2}\left(\frac{z}{z_n}\right)^2 + \cdots + \frac{1}{k}\left(\frac{z}{z_n}\right)^k\right]\right| \geq \exp\left(-2^k\left|\frac{z}{z_n}\right|^k\right) \geq \exp\left(-2^{\rho+\varepsilon}\left|\frac{z}{z_n}\right|^{\rho+\varepsilon}\right)$$

when $|z_n| < 2|z|$. Therefore the product of these terms over all zeroes for which $|z_n| < 2|z|$ is bounded below by

$$\exp\left(-2^{\rho+\varepsilon}|z|^{\rho+\varepsilon} \sum_{n=1}^{\infty} \frac{1}{|z_n|^{\rho+\varepsilon}}\right).$$

In summary, when $z$ lies outside the excluded disks, and $r$ is sufficiently large, each of the three pieces of the infinite product is bounded below by the exponential of a negative constant times $r^{\rho+\varepsilon}$. Therefore the whole infinite product is bounded below by a positive constant times $\exp(-r^{\rho+2\varepsilon})$ when $r$ is sufficiently large and $z$ lies outside the excluded disks. This deduction is the last step needed to complete the proof of Hadamard's factorization theorem.

# 6 Harmonic functions

## 6.1 Definition

The following theorem says that several different properties are equivalent. A real-valued function satisfying any one of the properties is called a harmonic function. (Some authors call a complex-valued function "harmonic" when both the real part and the imaginary part are harmonic.)

**Theorem.** *The following properties of a real-valued function $u$ on an open set $G$ are equivalent.*

1. *On each disk contained in $G$, there exists a holomorphic function whose real part equals $u$.*
2. *Every point in $G$ is the center of some disk on which there exists a holomorphic function whose real part equals $u$.*
3. *The function $u$ is twice continuously differentiable (that is, the second-order real partial derivatives $u_{xx}$, $u_{xy}$, $u_{yx}$, and $u_{yy}$ exist and are continuous), and $u_{xx} + u_{yy} = 0$ (that is, the function $u$ satisfies Laplace's equation).*
4. *The function $u$ is continuous and satisfies the small-circle mean-value property: namely, for each point $z_0$ in $G$, there is a positive radius $r_0$ (allowed to depend on $z_0$) such that if $0 < r < r_0$, then*
$$u(z_0) = \frac{1}{2\pi} \int_0^{2\pi} u(z_0 + re^{i\theta})\, d\theta.$$
5. *The function $u$ is continuous and satisfies the mean-value property on every closed disk contained in $G$.*

*Proof.* Evidently item 1 implies item 2. The plan of the rest of the proof is to show that item 2 implies item 3 which in turn implies item 1. Hence the first three properties are equivalent to each other. Evidently item 5 implies item 4. What remains is to show that item 4 implies one of the first three properties, and one of the first three properties implies item 5.

If item 2 holds, then each point of $G$ has a neighborhood in which there is a function $v$ such that $u + iv$ is holomorphic. Then $u_x = v_y$ and $u_y = -v_x$ by the Cauchy–Riemann equations. Differentiating the first equation with respect to the first variable and the second equation with respect to the second variable and adding shows that $u_{xx} + u_{yy} = v_{yx} - v_{xy} = 0$. (Since the holomorphic function $u + iv$ has continuous derivatives of all orders, so do the functions $u$ and $v$, whence the mixed second-order partial derivatives of $v$ match.) Thus item 2 implies item 3.

If item 3 holds, then fix a disk in $G$ with center point $(x_0, y_0)$, and define a "harmonic conjugate" function $v$ in the disk by the following line integral:

$$v(x, y) = \int_{(x_0,y_0)}^{(x,y)} u_s(s,t)\, dt - u_t(s,t)\, ds.$$

The integral is well defined—independent of the path joining the two points—because Laplace's equation implies that the integral over a closed loop is zero by Green's theorem. (In fancier language, the integral is path independent because the integrand is a closed differential form.) The fundamental theorem of calculus implies that if $v$ is defined by this formula, then $v$ is a continuously differentiable function with the property that $v_x = -u_y$ and $v_y = u_x$. In other words, the Cauchy–Riemann equations hold, so the function $u + iv$ is holomorphic in the disk. Thus item 3 implies item 1.

If item 1 holds, and $D$ is a closed disk contained in $G$ with center $z_0$ and radius $r$, then there is a holomorphic function $f$ in $D$ of which $u$ is the real part. The mean-value property of holomorphic functions implies that

$$f(z_0) = \frac{1}{2\pi} \int_0^{2\pi} f(z_0 + re^{i\theta})\, d\theta.$$

Taking the real part shows that item 5 holds.

The remaining step in the proof is to show that item 4 implies one of the first three properties. This step requires the development of a new tool (the Poisson integral) and is therefore postponed.

□

Observe that the construction of a harmonic conjugate function in the proof did not really require working on a disk: the same argument applies on an arbitrary simply connected region. On the other hand, if the region is not simply connected, then a harmonic function need not be equal to the real part of a holomorphic function globally.

*Exercise.* The function $\log|z|^2$ is well defined and harmonic on $\mathbb{C} \setminus \{0\}$, the punctured plane, but there is no holomorphic function $f$ on the punctured plane such that $\operatorname{Re} f(z)$ equals $\log|z|^2$.

*Exercise.* If $u$ is harmonic and $g$ is holomorphic, then the composite function $u \circ g$ is harmonic.

## 6.2 Poisson integral on the disk

Experience with the Cauchy integral suggests the possibility of recovering a holomorphic function from its real part by a suitable integral over the boundary of a region. A preliminary step is to see how to recover a harmonic function from a boundary integral. In the case of the unit disk, the formula is named for Siméon Denis Poisson (1781–1840), a contemporary of Cauchy.

Here is a magic trick for deriving the Poisson integral from the mean-value property. If $u$ is continuous on the closed unit disk and harmonic on the open disk, then the mean-value property yields that

$$u(0) = \frac{1}{2\pi} \int_0^{2\pi} u(e^{i\theta})\, d\theta = \frac{1}{2\pi i} \int_{|w|=1} u(w)\, \frac{dw}{w}.$$

Let $a$ be an arbitary point in the open unit disk, and compose with the disk automorphism $\varphi_a$ [recall that $\varphi_a(z) = (a - z)/(1 - \overline{a}z)$] to see that

$$u(a) = u \circ \varphi_a(0) = \frac{1}{2\pi i} \int_{|w|=1} u(\varphi_a(w))\, \frac{dw}{w}.$$

Make a change of variable, replacing $w$ by $\varphi_a(w)$ and remembering that $\varphi_a$ is self-inverse. Since

$$\frac{dw}{w} = d(\log w) \qquad \text{locally,}$$

and $w = 1/\overline{w}$ on the boundary of the disk,

$$\frac{d\varphi(w)}{\varphi(w)} = \left(\frac{1}{w-a} + \frac{\bar{a}}{1-\bar{a}w}\right) dw = \left(\frac{w}{w-a} + \frac{\bar{a}}{\overline{w}-\bar{a}}\right) \frac{dw}{w} \qquad \text{when } |w| = 1.$$

The expression in parentheses simplifies to $\dfrac{1-|a|^2}{|w-a|^2}$, so

$$u(a) = \frac{1}{2\pi} \int_0^{2\pi} u(e^{i\theta}) \frac{1-|a|^2}{|e^{i\theta}-a|^2}\, d\theta. \tag{1}$$

This formula is the Poisson integral representation for a harmonic function $u$ on the unit disk. Moreover,

$$\frac{1-|a|^2}{|e^{i\theta}-a|^2} = \operatorname{Re} \frac{w+a}{w-a} \qquad \text{when } w = e^{i\theta},$$

so the real-valued harmonic function $u(a)$ is the real part of the holomorphic function

$$\frac{1}{2\pi i} \int_{|w|=1} u(w) \frac{w+a}{w-a} \cdot \frac{dw}{w} \qquad \text{when } |a| < 1.$$

This second formula, named for Hermann Amandus Schwarz (1843–1921), explicitly determines a holomorphic function (up to an additive purely imaginary constant) from the real part.

Notice that the Poisson kernel

$$\frac{1}{2\pi} \cdot \frac{1-|a|^2}{|e^{i\theta}-a|^2}$$

is a positive function whose integral from 0 to $2\pi$ is equal to 1 (as follows from (1) when $u$ is identically equal to 1). Accordingly, the Poisson integral (1) exhibits the value $u(a)$ as a weighted average of the boundary values of $u$.

This interpretation of the Poisson integral immediately yields a local maximum principle for real-valued harmonic functions: if the restriction of a harmonic function to a closed disk attains a (weak) maximum at an interior point $a$, then the function reduces to a constant. Indeed, if the weighted average $u(a)$ is a maximum, then the values of $u$ on the boundary must all be equal to $u(a)$. Invoking (1) again at a different interior point shows that $u$ is constantly equal to $u(a)$ everywhere in the disk. Considering the negative of $u$ shows that harmonic functions satisfy a minimum principle too.

## 6.3 The Dirichlet problem for the disk

The preceding discussion shows that the Poisson integral reproduces harmonic functions on the unit disk. A little more work shows that the Poisson integral additionally solves the "Dirichlet problem" of finding a harmonic function on the disk with prescribed boundary values.

Suppose that a real-valued function $u$ is given on the boundary circle. Define the Poisson integral of $u$ at a point $a$ in the disk to be

$$\frac{1}{2\pi}\int_0^{2\pi} u(e^{i\theta})\frac{1-|a|^2}{|e^{i\theta}-a|^2}\,d\theta \qquad \text{or, equivalently,} \qquad \frac{1}{2\pi}\int_{-\pi}^{\pi} u(e^{i\theta})\frac{1-|a|^2}{|e^{i\theta}-a|^2}\,d\theta.$$

The integral makes sense if the function $u$ is continuous, or (more generally) Riemann integrable, or (still more generally) Lebesgue integrable. When the point $a$ lies inside the disk, the integral defines a function of $a$ that is harmonic, because the Poisson kernel is the real part of a holomorphic function. Is the limit of this function when $a$ tends to a boundary point equal to the value of the original function $u$ at that point?

The following argument shows that the answer is affirmative if $u$ is continuous at the specified boundary point. In view of the rotational invariance, verifying the claim at the specific boundary point 1 suffices. The Poisson integral certainly reproduces constant functions, so the difference between the value of the Poisson integral of $u$ at $a$ and the constant value $u(1)$ is

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1-|a|^2}{|e^{i\theta}-a|^2}\bigl(u(e^{i\theta})-u(1)\bigr)\,d\theta.$$

Fix an arbitrary positive $\varepsilon$, and invoke the continuity of the function $u$ at the point 1 to choose a positive $\delta$ such that $|u(e^{i\theta})-u(1)|<\varepsilon$ when $|\theta|<\delta$. Split the integral into the part for which $|\theta|<\delta$ and the part for which $|\theta|\ge\delta$. The integral over the first part is at most

$$\frac{\varepsilon}{2\pi}\int_{|\theta|<\delta} \frac{1-|a|^2}{|e^{i\theta}-a|^2}\,d\theta,$$

which by the positivity of the Poisson kernel does not exceed

$$\frac{\varepsilon}{2\pi}\int_{-\pi}^{\pi} \frac{1-|a|^2}{|e^{i\theta}-a|^2}\,d\theta, \qquad \text{or} \qquad \varepsilon,$$

the inequality being independent of the value of $a$ inside the disk. The integral over the second part, where $|\theta|\ge\delta$, tends to 0 when $a$ tends to 1 since the Poisson kernel converges to 0 uniformly on that piece. Since $\varepsilon$ is arbitrary, the limit of the value of the Poisson integral of $u$ at $a$ tends to $u(1)$ when $a$ tends to 1.

When the boundary function is continuous on the whole boundary, uniqueness of the solution of the Dirichlet problem in the disk is easy. Indeed, the difference of two solutions is a harmonic function with boundary value identically equal to zero; by the maximum and minimum principles, such a function is identically equal to zero inside the disk.

## 6.4 Badly behaved harmonic conjugates

The question arose in class of whether there can exist a holomorphic function on the open unit disk whose real part extends to be continuous on the closed disk but whose imaginary part does not extend to be continuous on the closed disk. When the real part has continuous boundary values, the imaginary part can be obtained in the interior of the disk from the formula of Schwarz worked out in § 6.2. But the proof in § 6.3 about good boundary behavior of the Poisson integral does not carry over to the imaginary part of the Schwarz integral. The question therefore is an interesting one. One way to see that examples do exist is to apply the following theorem.

**Theorem** (Carathéodory's theorem on boundary behavior of the Riemann map)**.** *Every bijective holomorphic map from the open unit disk onto a Jordan region necessarily extends to be a bijective continuous map between the closed unit disk and the closure of the Jordan region.*

According to the famous Jordan curve theorem, a simple closed curve (that is, the image of the unit circle under an injective continuous map) divides the plane into precisely two components, one bounded and the other unbounded. The bounded component is a Jordan region.

Notice that if the Riemann map extends to the boundary as a continuous injection of the closed disk, then the image region is necessarily a Jordan region. Accordingly, the hypothesis in Carathéodory's theorem is the natural assumption. An elementary proposition of point-set topology says that a continuous bijection between compact Hausdorff spaces is a homeomorphism (that is, the inverse map is continuous too), so the conclusion of Carathéodory's theorem can be rephrased as saying that the Riemann map extends to be a homeomorphism of the closed regions.

The theorem was proved by Carathéodory[3] and is usually referred to by his name, but William F. Osgood and Edson H. Taylor published the result the same year as Carathéodory as a special case of a more general theory.[4] Incidentally, finding the correct analogue of Carathéodory's theorem for holomorphic mappings in higher dimension is an unsolved problem.

In view of Carathéodory's theorem, the construction of a good harmonic function with a bad harmonic conjugate function can equivalently be carried out on some Jordan region other than the unit disk. Consider the Jordan region in the first quadrant bounded by the segment of real axis from 0 to 1, a vertical line segment from the point 1 to the point $1+i$, and the arc of the parabola defined in standard real coordinates by the property that $y = x^2$.

The function that sends the complex variable $z$ to $-i\log(z)$ is holomorphic on the open region. The real part is the harmonic function $\arg(z)$, or $\theta$ in polar coordinates. Evidently $\theta$ is continuous on the open right-hand half-plane. Moreover, the ratio $y/x$, which equals $\tan^{-1}\theta$ in the open Jordan region, tends to 0 when $(x, y)$ is a point of the region and $x \to 0$. Thus $\theta$ extends to be

[3] Über die gegenseitige Beziehung der Ränder bei der konformen Abbildung des Inneren einer Jordanschen Kurve auf einen Kreis, *Mathematische Annalen* **73**, no. 2 (1913) 305–320.

[4] See Corollary 1 on page 294 of their paper, Conformal transformations on the boundaries of their regions of definition, *Transactions of the American Mathematical Society* **14**, no. 2 (1913) 277–298. The main results of this paper were announced by Osgood a decade earlier: On the transformation of the boundary in the case of conformal mapping, *Bulletin of the American Mathematical Society* **9** (1903) 233–235.

a continuous function on the closed Jordan region, the value at the origin being 0. On the other hand, the imaginary part of the specified holomorphic function is $\log(1/|z|)$, which is unbounded at the origin and so does not extend to be continuous on the closed Jordan region.

## 6.5 The small-circle mean-value property

Just as Morera's theorem gives a way to characterize holomorphic functions by integration instead of by differentiation, the small-circle mean-value property gives a way to characterize harmonic functions via integration. The tools are in hand now to complete the missing step in the proof of the theorem stated in § 6.1.

The claim is that if $u$ is a continuous real-valued function on some open set, and if for every point $z$ in the set there is a positive radius $r$ (depending on $z$) such that the average of $u$ on every circle centered at $z$ of radius less than $r$ equals $u(z)$, then $u$ is necessarily harmonic. (In particular, the function $u$ turns out to be not merely continuous but actually infinitely differentiable.)

Harmonicity is a local property, so there is no loss of generality in supposing that the domain of $u$ is a disk and that $u$ is continuous on the closure of this disk. Scaling and translation do not affect the problem, so there is no loss of generality in taking the disk to be the unit disk centered at 0.

A key observation is that the small-circle mean-value property implies a maximum principle: the continuous function $u$ must attain its maximum on the boundary of the disk. Why? Since $u$ is continuous on a compact set, a maximum is attained somewhere. If the maximum is attained at an interior point, then the mean value of $u$ on every small circle centered at that point equals the maximum, so $u$ must be constantly equal to the central value on small circles. Hence $u$ is locally equal to the maximum. A standard connectedness argument now shows that $u$ is constantly equal to the maximum. So the maximum is taken on the boundary in any case.

Consider $P[u]$, the Poisson integral of the boundary value of $u$. The solution of the Dirichlet problem on the disk shows that $P[u]$ matches $u$ on the boundary.

Being harmonic inside the disk, the function $P[u]$ has the small-circle mean-value property. The difference $u - P[u]$ satisfies the small-circle mean-value property since both $u$ and $P[u]$ do. The observation in the preceding paragraph shows that $u - P[u]$ attains its maximum on the boundary. This boundary value equals 0, so $u - P[u] \leq 0$ inside the disk. Precisely the same argument applies to the difference $P[u] - u$, so $P[u] - u \leq 0$. The two inequalities combine to show that $u - P[u]$ is identically equal to 0.

Accordingly, a function $u$ satisfying the small-circle mean-value property is harmonic because locally $u$ matches a known harmonic function. This conclusion completes the proof of the theorem in § 6.1.

## 6.6 Harnack's principle

**Proposition.** *An increasing sequence of harmonic functions on a connected open set converges uniformly on compact subsets either to $+\infty$ or to a harmonic function.*

The proposition is named for Axel Harnack (1851–1888), a Baltic German mathematician. A corresponding statement holds for a decreasing sequence of harmonic functions, since the negative of a harmonic function is again a harmonic function.

The proof depends on Harnack's inequality for positive harmonic functions: namely, if $u$ is harmonic and nonnegative in the unit disk, and $0 < r < 1$, then

$$u(0)\frac{1-r}{1+r} \le u(re^{i\theta}) \le u(0)\frac{1+r}{1-r}.$$

Indeed, since $u$ is nonnegative, the Poisson integral representation and the mean-value property imply that

$$u(re^{i\theta}) = \frac{1}{2\pi}\int_0^{2\pi} u(e^{i\varphi})\frac{1-r^2}{|re^{i\theta}-e^{i\varphi}|^2}\,d\varphi \le \frac{1}{2\pi}\int_0^{2\pi} u(e^{i\varphi})\frac{1-r^2}{(1-r)^2}\,d\varphi = u(0)\frac{1+r}{1-r}.$$

(Strictly speaking, one should integrate over a slightly smaller circle and take the limit.) The other inequality follows in the same way, using that $|re^{i\theta}-e^{i\varphi}|^2 \le (1+r)^2$.

*Proof of Harnack's principle.* Replacing the increasing sequence $\{u_n\}$ by $\{u_n - u_1\}$ reduces to the case of nonnegative functions, so one might as well assume from the beginning that the functions are nonnegative. Harnack's inequality is then in force. Suppose the domain contains the unit disk. The increasing sequence $\{u_n(0)\}$ of real numbers either tends to $+\infty$ or is a Cauchy sequence. In the former case, Harnack's inequality implies that the sequence $\{u_n\}$ converges uniformly on compact sets to $+\infty$. In the latter case, the same reason implies that the sequence is uniformly Cauchy on compact subsets of the disk. The continuous limit function is represented by the Poisson integral and so is harmonic. The generalization from convergence on disks to convergence on general connected open sets is a routine compactness argument. □

# 7 Dirichlet problem on general domains

## 7.1 Dirichlet problem

The Poisson integral solves the Dirichlet problem on a disk. The corresponding problem in a general region is not always solvable.

*Example.* In the punctured disk $\{\, z \in \mathbb{C} : 0 < |z| < 1 \,\}$, there is no harmonic function $u$ such that $u$ has boundary value 0 on the outer boundary and boundary value 1 on the inner boundary.

Why? Seeking a contradiction, suppose that such a harmonic $u$ exists. The maximum principle implies that $u$ is bounded between 0 and 1, but more is true. Fix a positive $\varepsilon$, and apply the maximum principle to the harmonic function $u(z) + \varepsilon \log|z|$ on the annulus with outer radius 1 and inner radius $\exp(-1/\varepsilon)$. This function is negative on the inner boundary and approaches 0 at the outer boundary, so $u(z) + \varepsilon \log|z| < 0$ for $z$ in the annulus. Holding $z$ fixed, let $\varepsilon$ go to 0 to deduce that $u(z) \le 0$ when $0 < |z| < 1$. This conclusion contradicts the assumption that $u$ has boundary value 1 at the origin.

This example reveals the basic obstruction to solvability of the Dirichlet problem: thinness of the boundary. An upcoming theorem says that the Dirichlet problem is solvable when every component of the boundary contains more than one point (and even more generally).

The method to be considered is due[5] to the German mathematician Oskar Perron (1880–1975), who is noted for beautiful expository books, especially one on continued fractions. Additionally, he is remembered for the Perron integral, for a formula in analytic number theory, and for the Perron–Frobenius theorem in linear algebra about eigenvalues of matrices with positive entries (a result that has applications to internet search engines).

## 7.2 Subharmonic functions

A key tool in Perron's method for solving the Dirichlet problem is a class of functions known nowadays as *subharmonic functions*. The philosophy is that holomorphic functions and harmonic functions can be inconveniently rigid: the values of the function on an open set determine the values of the function everywhere. In particular, there are no holomorphic or harmonic partitions of unity. By contrast, subharmonic functions are flexible, enabling cut-and-paste operations. Yet there is a way to pass from flexible subharmonic functions to rigid harmonic functions through taking upper envelopes.

Roughly speaking, subharmonic functions sit underneath harmonic functions in the same way that convex functions sit underneath affine linear functions. Like convex functions, subharmonic functions need not be everywhere differentiable. In fact, subharmonic functions need not be continuous (although continuous ones will do for a basic solution to the Dirichlet problem).

The natural context for subharmonic functions is the class of real-valued upper semicontinuous functions. A function $u$ (defined on an open subset of a topological space) taking values in $[-\infty, \infty)$ is called upper semicontinuous if any of the following equivalent conditions holds:

- $\limsup_{z\to z_0} u(z) \le u(z_0)$ for every point $z_0$ in the domain of $u$.
- Reinterpretation of the preceding statement: For every number $M$ larger than $u(z_0)$, there is a neighborhood of $z_0$ such that $u(z) < M$ when $z$ is in that neighborhood. If $u(z_0) \ne -\infty$, then the number $M$ can be written conveniently in the form $u(z_0) + \varepsilon$.
- The set $\{\, z : u(z) < c \,\}$, the inverse image of $[-\infty, c)$ under $u$, is open for every real number $c$.

The word "upper" in the definition corresponds to the upper half of the inequality that characterizes continuity. What the condition says about the graph of the function is that the dot at a discontinuity fills in at (or above) the high point.

A reason for allowing the value $-\infty$ but excluding the value $+\infty$ is that upper semicontinuous functions arise naturally as pointwise limits of decreasing sequences of continuous finite-valued functions. Such limits can attain the value $-\infty$ but not the value $+\infty$.

**Proposition.** *An upper semicontinuous function is bounded above on every compact set and attains the least upper bound.*

[5] Oskar Perron, Eine neue Behandlung der ersten Randwertaufgabe für $\Delta u = 0$, *Math. Z.* **18** (1923), no. 1, 42–54.

*Proof.* The hypothesis implies that every point $z$ in the compact set $K$ has a neighborhood on which the function $u$ is bounded above by $u(z) + 1$. Finitely many such neighborhoods cover $K$. Hence $u$ is bounded above on $K$.

If the least upper bound $M$ is not attained, then the compact set $K$ is covered by the sequence of open sets of the form $\{ z : u(z) < M - \frac{1}{n} \}$ (where $n$ runs through the natural numbers), but there is no finite subcover. The contradiction shows that the bound $M$ must be attained after all. □

If $G$ is an open set in $\mathbb{C}$, then an upper semicontinuous function $u$ is called subharmonic[6] if for every disk in $G$ and for every harmonic function $v$ on the disk, the difference $u - v$ satisfies a local maximum principle: namely, the function $u - v$ cannot have a strict local maximum and can attain a weak local maximum at a point only if $u - v$ is constant in a neighborhood of the point. Thus if $u \le v$ on the boundary of the disk, then $u \le v$ in the interior of the disk.

This property evidently is local. The property needs to hold merely on all sufficiently small disks. In other words, for every point $z_0$ there should be a radius $r_0$ such that the property holds on each disk $B(z_0; r)$ when $0 < r < r_0$.

*Example.* If $f$ is holomorphic, then $\log |f|$ is subharmonic. (The function is defined to be equal to $-\infty$ at zeroes of $f$.)

Indeed, if $v$ is harmonic, then $\log |f| - v$ evidently cannot attain a local maximum at a zero of $f$ (except in the trivial case that $f$ is identically equal to 0). And away from the zeroes of $f$, there is a locally defined branch of $\log f$, so $\log |f|$ is harmonic, whence the difference $\log |f| - v$ is harmonic too. Hence there cannot be a local maximum unless the function is constant.

*Example.* If $u(x, y) = \min(0, x^2 - y^2)$ in $\mathbb{C}$, then $u$ is *not* subharmonic.

Indeed, if $v(x, y)$ is the harmonic function $x^2 - y^2$, then $u(x, y) - v(x, y)$ is equal to 0 when $x^2 - y^2 \le 0$ and is equal to the negative quantity $-(x^2 - y^2)$ when $x^2 - y^2 > 0$. Hence $u - v$ attains a maximal value of 0 but is not constant in a neighborhood of any point at which $x = y$, violating the maximum principle.

The initial definition of subharmonicity appears hard to verify in concrete examples. For functions having some regularity, there are equivalent properties that are easier to check than the original definition.

If $u$ is continuous, then an equivalent property is the local sub-mean-value property. In other words, for each point $z_0$ there is a radius $r_0$ such that

$$u(z_0) \le \frac{1}{2\pi} \int_0^{2\pi} u(z_0 + re^{i\theta})\, d\theta \qquad \text{when } 0 < r < r_0.$$

Indeed, if $u$ satisfies the sub-mean-value property, then so does $u - v$ when $v$ is harmonic. Hence $u - v$ satisfies the local maximum principle (for if the average value at the center of some disk is maximal, then the integrand must be constant on the disk). Conversely, if $u - v$ satisfies the local maximum principle for every harmonic $v$, then in a small disk let $v$ be the Poisson integral of $u$. The maximum principle implies that the value of $u$ at the center is at most the value of

[6] Some authors exclude the function that is identically equal to $-\infty$.

the Poisson integral of $u$ at the center, which equals the average of the values of $u$ around the boundary circle. Hence $u$ has the sub-mean-value property. The same argument shows that if $u$ has the local sub-mean-value property, then $u$ has the global sub-mean-value property on every disk whose closure lies inside the domain of $u$.

(The local sub-mean-value property can be used to characterize subharmonicity in general, when $u$ is merely upper semicontinuous instead of continuous, if you are willing to accept the Lebesgue integral. You need to go back to the discussion of the Poisson integral on the unit disk and check that the Poisson integral of a merely upper semicontinuous function produces a harmonic function whose lim sup at the boundary sits below the boundary value.)

If $u$ is twice continuously differentiable, then an equivalent condition to subharmonicity is that $\Delta u \geq 0$, where $\Delta$ is the Laplace operator. For the proof, suppose first that $\Delta u > 0$ with strict inequality. If $v$ is harmonic, then $\Delta(u - v) = \Delta u > 0$. Hence $u - v$ cannot have a local maximum, because at a local maximum, the second derivatives $\partial^2/\partial x^2$ and $\partial^2/\partial y^2$ of a function must be negative or zero. So $u - v$ does indeed satisfy the local maximum principle.

Next suppose only that $\Delta u \geq 0$. The goal is to show that if $v$ is a harmonic function in a small disk, say in $B(0; r)$, and if $u \leq v$ on the boundary of the disk, then $u \leq v$ inside the disk. If $\varepsilon$ is an arbitrary positive number, then $u(z) + \varepsilon|z|^2$ has strictly positive Laplacian, and $u(z) + \varepsilon|z|^2 \leq v(z) + \varepsilon r^2$ on the boundary of the disk, so the previous case implies that $u(z) + \varepsilon|z|^2 \leq v(z) + \varepsilon r^2$ inside the disk. Now let $\varepsilon$ go to zero.

Conversely, suppose that a twice continuously differentiable function $u$ is subharmonic. Why is $\Delta u \geq 0$? In the contrary case, $\Delta u$ would be negative on some open set. By what was just proved, the function $-u$ would be subharmonic on that set. Then both $u - v$ and $-u - (-v)$ would satisfy the maximum principle for every harmonic function $v$. Setting $v$ equal to the Poisson integral of $u$ on a small disk implies that $u$ is equal to its local Poisson integral, that is, $u$ is harmonic. Hence $\Delta u$ cannot be negative after all.

*Example.* Here are some standard ways to produce subharmonic functions.

- $|f|$ when $f$ is holomorphic. (The subharmonicity is easy to check from the mean-value property.)
- $|f|^p$ when $p$ is a positive number and $f$ is holomorphic. (At zeroes of $f$, the sub-mean-value property is immediate. Away from zeroes of $f$, there is a local holomorphic branch of $f^p$, so the subharmonicity follows from the preceding example.)
- $u \circ f$, where $u$ is subharmonic and $f$ is holomorphic. (When $u$ is twice continuously differentiable, compute that $\Delta(u \circ f) = |f'|^2 (\Delta u) \circ f$. In general, approximate $u$ by smooth subharmonic functions, which can be done by convolving with a mollifier.)
- $\alpha u_1 + \beta u_2$, where $u_1$ and $u_2$ are subharmonic, and $\alpha$ and $\beta$ are *nonnegative* real numbers. (This case is clear from the sub-mean-value property.)
- $\max(u_1, u_2)$ (pointwise maximum), where $u_1$ and $u_2$ are subharmonic. (This case is clear from the sub-mean-value property.)
- More generally, suppose $\{u_t\}_t$ is a family of subharmonic functions, and consider the pointwise supremum $\sup_t u_t(z)$. In general, this envelope need not be upper semicontinuous, but if the envelope is upper semicontinuous, then the envelope is subharmonic.

[Aside: Here is an example of failure of upper semicontinuity of the envelope. The function $(1/n)\log|z|$ is subharmonic and negative in the unit disk for each natural number $n$. The pointwise supremum of this sequence of functions equals 0 on the punctured disk but $-\infty$ at the center, hence is not upper semicontinuous. On the other hand, for a family that is locally bounded above, the upper semicontinuous regularization of the envelope is subharmonic.] To see why the envelope is subharmonic, apply the sub-mean-value property. If a positive $\varepsilon$ is specified, and a point $z_0$ is specified, then there is some parameter value $t_0$ such that

$$\begin{aligned}\sup_t u_t(z_0) \le u_{t_0}(z_0) + \varepsilon &\le \frac{1}{2\pi}\int_0^{2\pi} u_{t_0}(z_0 + re^{i\theta})\,d\theta + \varepsilon \\ &\le \frac{1}{2\pi}\int_0^{2\pi} \sup_t u_t(z_0 + re^{i\theta})\,d\theta + \varepsilon.\end{aligned}$$

Letting $\varepsilon$ go to 0 shows that the upper envelope has the sub-mean-value property.

- $\log(1+|z|)$ is subharmonic. In principle, the subharmonicity can be verified by computing second derivatives, but the calculation is nasty. Here is a trick. Observe that

  $$\log(1+|z|) = \sup_\theta \log|1+e^{i\theta}z|,$$

  by the triangle inequality. For each fixed $\theta$, the function $\log|1+e^{i\theta}z|$ is subharmonic, being the logarithm of the modulus of a holomorphic function. The envelope is not only upper semicontinuous but even continuous. Hence the preceding example shows that $\log(1+|z|)$ is subharmonic.
- If $u$ is subharmonic in a region, and $D$ is a closed disk in the region, build a new function by replacing $u$ inside $D$ by the Poisson integral of the value of $u$ on $\partial D$. Then $u$ satisfies the mean-value property at points inside $D$ and the sub-mean-value property at points outside $D$. What about points on $\partial D$? The original function satisfies the sub-mean-value property at these points, and the Poisson integral is at least as large as $u$ inside $D$, so the average value of the new function increases. Hence the sub-mean-value property can only improve. Thus local "Poissonization" of a subharmonic function produces a new subharmonic function.

## 7.3 Hadamard's theorems on three lines and three circles

Since the modulus of a holomorphic function is subharmonic, some versions of the maximum principle are most naturally stated in the context of subharmonic functions. Here is a family of examples that appear in applications.

**Theorem.** *Suppose that $u$ is a subharmonic function in a strip $\{\,(x,y)\in\mathbb{R}^2 : a<x<b\,\}$, and $u$ is bounded above. Let $M(x)$ denote* $\sup\{\,u(x,y) : y\in\mathbb{R}\,\}$. *Then $M(x)$ is a convex function of $x$ on the interval $(a,b)$.*

The word “convex” is to be understood in the usual sense of real analysis: namely, if $x_1$ and $x_2$ are two arbitrary points in the interval $(a, b)$, and $t$ is a real number between 0 and 1, then

$$M(tx_1 + (1-t)x_2) \le tM(x_1) + (1-t)M(x_2).$$

The geometric content of the inequality is that the graph of the function $M$ lies below each chord: convex functions are “sublinear.” The reason for the name “three lines” is that bounds on the function on two lines control the size of the function on any third line in between.

*Proof.* Since subharmonicity is a property that is preserved by translations and by dilations, there is no loss of generality in supposing that $a = -\pi/2$ and $b = \pi/2$. Suppose $x_1$ and $x_2$ are two arbitrary numbers such that $-\pi/2 < x_1 < x_2 < \pi/2$. What needs to be shown is that if $p$ is a first-degree polynomial of one real variable such that $M(x_1) \le p(x_1)$ and $M(x_2) \le p(x_2)$, then $M(x) \le p(x)$ whenever $x_1 < x < x_2$.

The second derivative of $p$ is identically equal to zero, so $p(x)$ can be viewed as a harmonic function that happens to be independent of $y$. Consider for an arbitrary positive $\varepsilon$ the function

$$u(x, y) - p(x) - \varepsilon \operatorname{Re} \cos(x + iy). \tag{2}$$

Since the real part of $\cos(x + iy)$ equals $\cos(x)\cosh(y)$, which is strictly positive in the strip where $|x| < \pi/2$, the indicated function (2) is negative on the vertical lines where $x = x_1$ and $x = x_2$. Moreover, the real part of $\cos(x+iy)$ tends to $+\infty$ when $|y| \to \infty$, and the convergence is uniform with respect to $x$ between $x_1$ and $x_2$. By hypothesis, the function $u$ is bounded above, so for sufficiently large $R$, the function (2) is negative on the horizontal line segments where $y = \pm R$ and $x_1 \le x \le x_2$.

The function (2) is the difference between a subharmonic function and a harmonic function, so the maximum principle for bounded regions implies that for every sufficiently large $R$, the expression (2) is negative on the rectangular region where $x_1 \le x \le x_2$ and $|y| \le R$. Letting $R$ tend to infinity shows that the expression (2) is negative on the whole strip where $x_1 \le x \le x_2$. Letting $\varepsilon$ tend to zero shows that $u(x, y) \le p(x)$ when $x_1 \le x \le x_2$. Taking the supremum over $y$ shows that $M(x) \le p(x)$ when $x_1 \le x \le x_2$, as claimed. □

*Remark.* The three-lines theorem can be viewed as a maximum principle that applies to a special unbounded region. If $u$ is bounded above by some unknown constant in a strip, and $u$ is bounded above on the sides of the strip by a specific constant $C$, then the theorem implies that $u$ is bounded above by the same constant $C$ inside the strip.

The proof reveals that the hypothesis of boundedness of $u$ can be relaxed. What is needed is that $u$ does not grow too fast at infinity. For instance, if there are positive constants $A$ and $B$, with $B$ strictly less than 1, such that $u(x, y) \le Ae^{B|y|}$ when $-\pi/2 < x < \pi/2$, then an upper bound for $u$ on two lines implies the same upper bound on the region between the two lines. For a general interval $(a, b)$, the requirement is that $B < \pi/(b - a)$. Generalizations along these lines are part of so-called Phragmén–Lindelöf theory.

**Corollary.** *Suppose that $f$ is holomorphic, not identically zero, and bounded in a vertical strip. Let $M(x)$ denote* $\sup\{\,|f(x+iy)| : y \in \mathbb{R}\,\}$*. Then* $\log M(x)$ *is a convex function; equivalently, if $x_1$ and $x_2$ are real numbers in the strip, and $0 < t < 1$, then*

$$M(tx_1 + (1-t)x_2) \le M(x_1)^t M(x_2)^{1-t}.$$

*Proof.* Apply the preceding theorem to the subharmonic function $\log|f|$ and exponentiate the convexity inequality. □

*Remark.* Since the geometric mean is no larger than the arithmetic mean, the corollary implies (but is stronger than) the statement that $M(x)$ is a convex function.

The exponential function wraps a vertical line around a circle. Accordingly, the preceding results produce analogous theorems for circular geometry.

**Theorem.** *Suppose that $u$ is a subharmonic function in an annulus $\{\, z \in \mathbb{C} : a < |z| < b \,\}$ with inner radius $a$ and outer radius $b$. Let $m(r)$ denote the maximum of $u(z)$ when $|z| = r$. If $a < r_1 < r < r_2 < b$, then*

$$m(r) \le \frac{\log r_2 - \log r}{\log r_2 - \log r_1} m(r_1) + \frac{\log r - \log r_1}{\log r_2 - \log r_1} m(r_2).$$

Sometimes this inequality is described by saying that $m(r)$ is "a convex function of $\log r$." In other words, the composite function $m(e^r)$ is convex. An equivalent formulation of the inequality is that

$$m(r) \le \frac{\log \frac{r_2}{r}}{\log \frac{r_2}{r_1}} m(r_1) + \frac{\log \frac{r}{r_1}}{\log \frac{r_2}{r_1}} m(r_2).$$

*Proof.* Applying the three-lines theorem to the function $u(e^z)$, which is subharmonic in the strip where $\log a < \operatorname{Re} z < \log b$, shows that

$$m(e^x) \le \frac{x_2 - x}{x_2 - x_1} m(e^{x_1}) + \frac{x - x_1}{x_2 - x_1} m(e^{x_2})$$

when $\log a < x_1 < x < x_2 < \log b$. The required inequality follows by setting $x$ equal to $\log r$ and $x_1$ equal to $\log r_1$ and $x_2$ equal to $\log r_2$.

Alternatively, let $v(z)$ denote the harmonic function

$$\frac{\log r_2 - \log|z|}{\log r_2 - \log r_1} m(r_1) + \frac{\log|z| - \log r_1}{\log r_2 - \log r_1} m(r_2)$$

in the annulus where $a < |z| < b$. Then $v(z)$ takes the constant value $m(r_1)$ when $|z| = r_1$ and the constant value $m(r_2)$ when $|z| = r_2$, so the definition of subharmonic function implies that $u(z) \le v(z)$ when $r_1 < |z| < r_2$. Taking the supremum when $|z| = r$ gives the required inequality. □

**Corollary.** *Suppose that $f$ is a holomorphic function in an annulus $\{ z \in \mathbb{C} : a < |z| < b \}$. Let $M(r)$ denote the maximum of $|f(z)|$ when $|z| = r$. If $a < r_1 < r < r_2 < b$, then*

$$M(r) \leq M(r_1)^{\lambda} M(r_2)^{1-\lambda}, \quad \textit{where } \lambda = \frac{\log \frac{r_2}{r}}{\log \frac{r_2}{r_1}}.$$

In other words, the function $\log M(r)$ is a convex function of $\log r$. The proof follows directly from the preceding theorem by applying the statement to the subharmonic function $\log |f|$ and exponentiating the inequality.

## 7.4 Perron's method

Suppose $\varphi$ is a given function on the boundary of a bounded region. Consider the class of all subharmonic functions in the region whose boundary values do not exceed those of $\varphi$. Take the pointwise supremum of all such subharmonic functions. If there is a solution of the Dirichlet problem, then this construction must yield the solution, for the putative solution is in the class. Moreover, the putative solution is an upper bound for all subharmonic functions with the given boundary values.

The question, then, is whether the upper envelope actually does solve the Dirichlet problem. The counterexample mentioned earlier (the punctured disk) shows that some information about the boundary has to come into play. The essential element turns out to be the existence (or the non-existence) of subharmonic peak functions. A *peak function* corresponding to a boundary point $z_0$ of a region $G$ means a negative function $u$ on $G$ such that $\lim_{z \to z_0} u(z) = 0$ and $\limsup_{z \to w} u(z) < 0$ when $w \in \partial G \setminus \{z_0\}$.

**Theorem** (Solvability of the Dirichlet problem)**.** *If $G$ is a bounded region in $\mathbb{C}$ such that $G$ admits a subharmonic peak function corresponding to each boundary point, and if $\varphi$ is a continuous function on the boundary of $G$, then there exists a harmonic function $u$ on $G$ such that $\lim_{z \to w} u(z) = \varphi(w)$ for every point $w$ in the boundary of $G$.*

*Moreover, if $\mathcal{F}_\varphi$ is the Perron family consisting of every subharmonic function $v$ on $G$ such that $\limsup_{z \to w} v(z) \leq \varphi(w)$ for each $w$ in the boundary of $G$, then $u(z) = \sup_{v \in \mathcal{F}_\varphi} v(z)$ for every $z$ in $G$.*

The proof has two parts. The first part is to show that the envelope of the Perron family is harmonic. That conclusion holds even without the hypothesis of the existence of peak functions. The second part is to show that the peak functions force the envelope of the Perron family to have the right boundary values.

In Perron's method, a needed fact is that if $\varphi$ is the boundary value of a function $u$ that is subharmonic in a neighborhood of the closed disk, then the Poisson integral of $\varphi$ is at least as large as $u$ inside the disk. Since the previous discussion about the Poisson integral used continuity of the boundary values, some further argument is needed to handle subharmonic boundary values.

The necessary proposition is that every upper semicontinuous function on a compact set (or on any set where the function is bounded above) is the limit of a decreasing sequence of continuous functions. Namely, let $\varphi_n(w)$ be $\sup_t\{\varphi(t)-n|t-w|\}$. (To see the point of this definition, consider the case of a function that is constant except for a jump at one point.) When $t = w$, the expression in brackets equals $\varphi(w)$, so $\varphi_n(w) \geq \varphi(w)$. Moreover, for each fixed $t$ the expression in brackets decreases as $n$ increases, so the sequence $\{\varphi_n\}$ is decreasing. If $M$ is an arbitrary number larger than $\varphi(w)$, then by upper semicontinuity there is a neighborhood of $w$ such that $\varphi(t) < M$ for $t$ in the neighborhood. On the other hand, the quantity $|t - w|$ is bounded away from 0 outside the neighborhood, and $\varphi$ is bounded above, so $\varphi(t) - n|t - w| \to -\infty$ uniformly outside the neighborhood when $n \to \infty$. It follows that $\varphi_n(w) < M$ for large $n$. Since $M$ is arbitrary, the limit to which the decreasing sequence $\{\varphi_n(w)\}$ converges is $\varphi(w)$. What remains to see is that $\varphi_n$ is continuous. For arbitrary points $w_1$ and $w_2$, the triangle inequality implies that

$$\varphi(t) - n|t - w_1| \geq \varphi(t) - n|t - w_2| - n|w_1 - w_2| \qquad \text{for each } t,$$

so $\varphi_n(w_1) \geq \varphi_n(w_2) - n|w_1 - w_2|$. Interchanging $w_1$ and $w_2$ then shows that $\varphi_n$ is a Lipschitz function with Lipschitz constant equal to $n$. In particular, $\varphi_n$ is continuous.

Returning to the Poisson integral, suppose that $v$ is the Poisson integral of the boundary value of a subharmonic function $u$. Approximate the boundary value by a decreasing sequence $\{u_n\}$ of continuous functions. Let $v_n$ be the Poisson integral of $u_n$. Then $v_n$ has the boundary values of $u_n$, so $v_n$ is a harmonic function than exceeds $u$ on the boundary, whence $v_n$ exceeds $u$ inside the disk. The functions $v_n$ decrease inside the disk by the maximum principle. By the monotone convergence theorem for integrals, the functions $v_n$ converge to $v$, which therefore dominates $u$ inside the disk.

This argument has a further implication. By Harnack's principle, the limiting function $v$ is harmonic and not identically $-\infty$ (unless $u$ is identically $-\infty$). Consequently, a subharmonic function (not identically $-\infty$) is integrable on each circle (that is, the integral is not $-\infty$). For similar reasons, subharmonic functions are area-integrable.

## 7.5 Return to the solution of the Dirichlet problem

*Proof of the harmonicity of the Perron envelope.* Suppose that $G$ is a bounded domain, and $\varphi$ is a bounded function on the boundary. (For this part of the proof, the continuity of $\varphi$ is not needed.) The goal is to show that the upper envelope $u$ of the Perron family $\mathcal{F}_\varphi$ is harmonic.

Recall that a function $v$ belongs to the Perron family if and only if $v$ is subharmonic, and $\limsup_{z\to w} v(z) \leq \varphi(w)$ for every point $w$ in the boundary of $G$. If $M$ is a constant larger than the supremum of $\varphi$, then every function $v$ in the Perron family has the property that $v - M$ is negative near the boundary of $G$ and hence is negative everywhere inside $G$ (by the maximum principle; the boundedness of the domain $G$ is used here). Therefore every function in the Perron family is bounded above by $M$. Hence $u$, the upper envelope, is bounded above by $M$.

It suffices to verify harmonicity—a local property—on an arbitrary disk $B(z_0; r)$ whose closure is contained in $G$. Let $\{v_n\}$ be a sequence of subharmonic functions in the Perron family such that the sequence $\{v_n(z_0)\}$ increases up to $u(z_0)$. Replacing each $v_k$ by $\max\{v_1, \dots, v_k\}$ ensures that the sequence $\{v_n\}$ is increasing at each point of $G$.

Next replace each $v_k$ with its "Poissonization" inside $B(z_0; r)$ to ensure that $v_k$ is harmonic inside the disk. The modified sequence $\{v_n\}$ now is an increasing sequence in the Perron family, and inside $B(z_0; r)$ this sequence is an increasing sequence of harmonic functions that converges at $z_0$ to $u(z_0)$. By Harnack's principle, the limit of the sequence $\{v_n\}$ inside $B(z_0; r)$ is a harmonic function, say $v^*$.

The proof is not finished, for what is known so far is that $u$, the Perron envelope, matches $v^*$, a harmonic function, at one point. Does $u$ match $v^*$ at other points of $B(z_0; r)$ besides $z_0$?

Suppose $z_1$ is an arbitrary point of $B(z_0; r)$. Repeat the preceding construction to obtain an increasing sequence $\{u_n\}$ in the Perron family such that the sequence $\{u_n(z_1)\}$ converges to $u(z_1)$. Replacing each $u_k$ by $\max(u_k, v_k)$ gives a new increasing sequence of subharmonic functions in the Perron family that converges to the upper envelope $u$ at both $z_0$ and $z_1$. Poissonizing as before produces a harmonic limit function $u^*$ in $B(z_0; r)$ that matches $u$ at both $z_0$ and $z_1$.

By construction, $v^* - u^* \leq 0$ in $B(z_0; r)$, and $v^*(z_0) = u(z_0) = u^*(z_0)$. By the maximum principle, the harmonic function $v^* - u^*$ is identically equal to 0 in $B(z_0; r)$. Consequently, $v^*(z_1) = u^*(z_1) = u(z_1)$. Since $z_1$ is arbitrary, the function $v^*$ is a harmonic function in $B(z_0; r)$ that equals the envelope $u$ in all of $B(z_0; r)$. Thus the envelope $u$ is indeed harmonic (in all of $G$, since $z_0$ is arbitrary). $\square$

*Proof that peak functions imply the right boundary values.* Suppose now that the bounded function $\varphi$ is continuous at $z_0$ and that there is a subharmonic peak function for $z_0$. The claim is that the Perron envelope function $u$ has limit at $z_0$ equal to $\varphi(z_0)$.

There is no loss of generality in supposing that $\varphi(z_0) = 0$. (Simply subtract the constant value $\varphi(z_0)$ from all functions.) Two arguments are needed, one to show that the envelope is not too big and another to show that the envelope is not too small.

Fix a positive $\varepsilon$. The goal is to find a neighborhood of $z_0$ such that $-\varepsilon < u(z) < \varepsilon$ when $z$ is a point of $G$ lying in the neighborhood. Since $\varphi$ is continuous at $z_0$, there is a radius $r$ such that $-\varepsilon/2 < \varphi(z) < \varepsilon/2$ whenever $z$ is a point of $\partial G$ for which $|z - z_0| < r$.

Let $\psi$ be a subharmonic peak function corresponding to $z_0$. Since $G$ is bounded, the intersection of the boundary of $G$ with the closed set $\{\, z \in \mathbb{C} : |z - z_0| \geq r \,\}$ is compact, and each point $z$ of this compact set has a neighborhood $N_z$ such that the upper semicontinuous function $\psi$ is negative on $N_z \cap G$. Taking a finite subcover produces an open neighborhood $U$ of the compact set $\{\, z \in \mathbb{C} : |z - z_0| \geq r \,\} \cap \partial G$ such that the function $\psi$ has a negative upper bound on $U \cap G$, say $-\delta$.

Showing that the upper envelope $u$ is not too small near $z_0$ requires constructing a particular member of the Perron family that is not too small near $z_0$. Let $M$ be a positive constant so large that $M\delta$ exceeds the supremum of $|\varphi|$ on $\partial G$. If $w$ is a point of $\partial G$ at distance at least $r$ from $z_0$, then $\limsup_{z\to w} M\psi(z) \leq -M\delta < \varphi(w)$. On the other hand, if $w$ is a point of $\partial G$ within distance $r$ from $z_0$, then $\limsup_{z\to w} M\psi(z) \leq 0 < \varphi(w) + \varepsilon/2$. Therefore the function $M\psi - \varepsilon/2$ belongs to the Perron family associated to the boundary function $\varphi$. Accordingly, $M\psi(z) - \varepsilon/2 \leq u(z)$ for every point $z$ in $G$. By the definition of peak function, $\lim_{z\to z_0} M\psi(z) = 0$, so there is a neighborhood of $z_0$ in which $-\varepsilon/2 < M\psi(z)$. In this neighborhood, $-\varepsilon < u(z)$.

Showing that the upper envelope $u$ is not too big near $z_0$ requires finding an upper bound on every member of the Perron family. The construction in the preceding paragraph implies that $\limsup_{z\to w} M\psi(z) < -\varphi(w)$ when $w$ is a point of $\partial G$ at distance at least $r$ from $z_0$, and $\limsup_{z\to w} M\psi(z) \le 0 < -\varphi(w) + \varepsilon/2$ when $w$ is a point of $\partial G$ within distance $r$ from $z_0$. Consequently, if $v$ is an arbitrary member of the Perron family, then $\limsup_{z\to w}(v + M\psi - \varepsilon/2) < 0$ for every point $w$ in $\partial G$. Since $v + M\psi - \varepsilon/2$ is subharmonic, the maximum principle implies that $v + M\psi - \varepsilon/2$ is negative everywhere inside $G$. Thus $v < -M\psi + \varepsilon/2$ inside $G$. Taking the pointwise supremum over functions $v$ in the Perron family shows that $u \le -M\psi + \varepsilon/2$. Since $\lim_{z\to z_0} -M\psi(z) = 0$, there is a neighborhood of $z_0$ in which $-M\psi(z) < \varepsilon/2$. In this neighborhood, $u(z) < \varepsilon$.

In conclusion, there is a neighborhood of $z_0$ such that $-\varepsilon < u(z) < \varepsilon$ when $z$ is a point of $G$ in the neighborhood. But $\varepsilon$ is arbitrary, so $\lim_{z\to z_0} u(z) = 0$, as claimed. □

## 7.6 Remarks on barriers

### 7.6.1 Lebesgue

The term "barrier" was introduced by Henri Lebesgue[7] as a name for a harmonic peak function (more precisely, a family of functions). The details of the definition of barrier vary in modern sources. The proof in § 7.5 uses subharmonic functions that are strict peak functions. As indicated below, the hypothesis can be weakened somewhat, making the theorem more general but the proof more difficult.

Lebesgue even provided an algorithm for solving the Dirichlet problem when there is a solution. Suppose given a continuous function on the boundary of a bounded open set in the plane. Extend the function arbitrarily to a continuous function on the closed region, say by the Tietze extension theorem. Execute the following recursive construction.

Replace the value of the function at each point by the average value over the largest disk centered at the point and contained in the region (two-dimensional average over the disk, not one-dimensional average over the boundary circle). Repeat the averaging process for the new function that arises, and iterate.

Lebesgue showed that this sequence of averages converges uniformly to the solution of the Dirichlet problem, assuming the existence of a harmonic barrier at each boundary point. The same argument works assuming the existence of a subharmonic peak function at each boundary point.

### 7.6.2 Necessity and sufficiency

The proof in § 7.5 shows that if there exists a subharmonic peak function on $G$ corresponding to the boundary point $z_0$, and $\varphi$ is a bounded function on $\partial G$ that is continuous at $z_0$, then the upper envelope of the Perron family corresponding to $\varphi$ has limit at $z_0$ equal to $\varphi(z_0)$. Some authors say in this situation that the point $z_0$ is a "regular point" for the Dirichlet problem.

[7] Sur le problème de Dirichlet, *Comptes rendus hebdomadaires des séances de l'Académie des sciences* **154** (1912) 335–337.

What about the converse? If $z_0$ is a regular point, that is, if the upper envelope of the Perron family $\mathcal{F}_\varphi$ has limit $\varphi(z_0)$ at $z_0$ whenever $\varphi$ is continuous at $z_0$, must there exist a subharmonic peak function corresponding to $z_0$?

The answer is affirmative, for the following reason. The function that sends $z$ to $-|z - z_0|$ certainly is continuous on $\partial G$. The upper envelope of the corresponding Perron family is then a harmonic function $h$ on $G$ having limit at $z_0$ equal to zero. If $v$ is an arbitrary member of the Perron family, then $v(z)+|z-z_0|$ is subharmonic (being the sum of two subharmonic functions), and the defining property of the Perron family implies that $\limsup_{z\to w\in\partial G}(v(z)+|z-z_0|) \le 0$ for every point $w$ in the boundary of $G$. The maximum principle for subharmonic functions implies that $v(z) + |z - z_0| \le 0$ for every point $z$ inside $G$. Passing to the upper envelope shows that $h(z) \le -|z - z_0|$ when $z \in G$. Therefore the upper limit of $h$ at every boundary point other than $z_0$ is strictly negative. Thus $h$ is a (sub)harmonic peak function in the sense defined in § 7.4.

### 7.6.3 Locality

The existence of a subharmonic peak function corresponding to a boundary point $z_0$ turns out to be a local property of the boundary of the region $G$ near $z_0$. In other words, if for some disk $B(z_0; r)$ there is a negative subharmonic function $\chi$ on $G \cap B(z_0; r)$ such that $\lim_{z\to z_0} \chi(z) = 0$ and $\limsup_{z\to w} \chi(z) < 0$ when $w \in B(z_0; r) \cap \partial G$, then there is a subharmonic peak function $\psi$ on $G$ corresponding to $z_0$.

To see how to construct the global peak function $\psi$ from the local peak function $\chi$, first shrink to the disk $B(z_0; r/2)$. The hypotheses imply that the function $\chi$ has a strictly negative upper bound on $G \cap \partial B(z_0; r/2)$. Choose a positive number $C$ so large that $C \cdot \chi < -1$ on $G \cap \partial B(z_0; r/2)$. Define $\psi$ as follows.

$$\psi(z) = \begin{cases} -1, & \text{if } z \in G \setminus B(z_0; r/2), \\ \max\{-1, C\chi(z)\}, & \text{if } z \in G \cap B(z_0, r/2). \end{cases}$$

Evidently $\psi$ is negative in $G$ and has limit equal to 0 at $z_0$ (because $\chi$ has limit 0). Is the function $\psi$ subharmonic? Certainly $\psi$ is subharmonic on the open set $G\setminus\overline{B}(z_0; r/2)$, where $\psi$ has the constant value $-1$. And $\psi$ is subharmonic on the open set $G \cap B(z_0; r/2)$, since the pointwise maximum of two subharmonic functions is subharmonic. At every point of $G \cap \partial B(z_0; r/2)$, the function $\psi$ attains a weak local minimum, hence automatically satisfies the small-circle sub-mean-value property. So the constructed function $\psi$ is indeed subharmonic on $G$. Moreover, $\limsup_{z\to w} \psi(z) < 0$ when $w \in B(z_0; r)\cap\partial G$ because $\chi$ has this property, and $\limsup_{z\to w} \psi(z) = -1$ if $w \in (\partial G) \setminus B(z_0; r/2)$. Thus $\psi$ is the required global subharmonic peak function.

### 7.6.4 Bouligand

The definition in § 7.4 says that a subharmonic peak function corresponding to a boundary point $z_0$ should have three properties: (i) the values of the function are strictly negative inside the region; (ii) the limit of the function at the boundary point $z_0$ should be zero; and (iii) the upper limit of the function should be strictly negative at all other boundary points. Such a function might be called more precisely a "strict" peak function.

Since the function is supposed to be negative inside the region, the upper limit at the boundary is certainly less than or equal to zero. Dropping property (iii) allows the function to have limit zero at some boundary points other than $z_0$. A function satisfying merely properties (i) and (ii) might be called a "weak" peak function. A technical refinement in the theory of the Dirichlet problem is that the existence of a weak subharmonic peak function (or even a local weak subharmonic peak function) actually implies the existence of a strict subharmonic peak function. This nontrivial statement (not proved here) can be established through an argument due to G. Bouligand.[8]

## 7.7 Construction of peak functions

When do subharmonic peak functions exist? Examples in the homework assignment reveal that there cannot be a subharmonic peak function at the center of a punctured disk.

But subharmonic peak functions do exist at reasonable boundary points. The construction is easy at points where there is a supporting line, straightforward at points that are accessible from the exterior by a line segment, and difficult for boundary points about which all that is known is that the point is not a singleton boundary component.

*Example.* If $G$ is a convex domain, in the sense that at each boundary point there is a supporting line that intersects the (open) domain at no other point, then there is a harmonic peak function. Indeed, a translation puts the boundary point at the origin, and a rotation makes the imaginary axis the supporting line, with the domain lying in the right-hand half-plane. If the domain is strongly convex (no boundary point besides the origin lies on the imaginary axis), then $-\operatorname{Re} z$ is a harmonic peak function (hence a subharmonic peak function). If the domain is only weakly convex, then $-\operatorname{Re}\sqrt{z}$ is a harmonic peak function.

*Example.* Suppose $z_0$ is a boundary point of a domain with the property that there is a line segment lying in the complement of the domain with one endpoint at $z_0$. (Part or all of the line segment is allowed to lie in the boundary of the domain.) Then there is a peak function at $z_0$.

In particular, a domain bounded by a finite number of smooth curves admits peak functions at all boundary points. The boundary curves can even have cusps. Moreover, the region can have some straight slits.

To construct the peak function, let $z_1$ be a second point on the indicated line segment. Use the linear fractional transformation $(z - z_0)/(z - z_1)$ to send $z_0$ to 0 and $z_1$ to $\infty$, and make a rotation to ensure that the line segment maps to the negative part of the real axis. If $z_0$ is the only point of the original line segment that lies on the boundary of the region, then use $\sqrt{z}$ to map into the right-hand half-plane, and take the negative of the real part of the variable as the peak function (as in the preceding example). If the original line segment touches the boundary of the region at more than one point, then use a fourth root instead of a square root in this construction.

[8] Sur le problème de Dirichlet, *Annales de la Société Polonaise de Mathématique* **4** (1926) 59–112.

The goal now is to establish the much more general statement that if $z_0$ is a boundary point of $G$, and the connected component $K$ of the complement of $G$ containing $z_0$ contains at least one other point, then there is a subharmonic peak function at $z_0$.

(Notice that singleton boundary components can arise as punctures in the domain but also in other ways. For instance, consider the unit disk with a slit along the interval $[2^{-n}, 2^{-n} + 4^{-n}]$ of the real axis for each natural number $n$ and with the origin removed too. The origin is a nonisolated boundary point that also is a singleton connected component of the boundary.)

To construct the peak function, first make a linear fractional transformation to put the point $z_0$ at 0 and a second point of $K$ at $\infty$. The transformation amounts to a holomorphic change of coordinates, so there is no harm in continuing to use the letters $G$ and $K$ to represent the image of the domain and the image of the boundary component.

In the transformed picture, the complement $\mathbb{C} \setminus K$ is a simply connected region containing $G$, so there is a holomorphic branch of $\log(z)$ on $\mathbb{C} \setminus K$ that can be restricted to $G$. On the intersection of $G$ with the unit disk, the real part of $\log(z)$ is a negative harmonic function. On the same open set $G \cap B(0; 1)$, the real part of $1/\log(z)$ is a negative harmonic function, and $\operatorname{Re}\{1/\log(z)\} \to 0$ when $z \to 0$.

Accordingly, the real part of $1/\log(z)$ is a weak local harmonic peak function. The discussion in § 7.6.3 shows that the existence of a local peak function implies the existence of a global peak function. A technical complication arises here, because the imaginary part of $\log(z)$ could be unbounded if $G$ has a spiral structure. Therefore the real part of $1/\log(z)$ is not necessarily a strict local peak function. But Bouligand's lemma mentioned in § 7.6.4 implies that the existence of a weak local peak function suffices.

## 7.8 Wiener's criterion for regularity

A necessary and sufficient condition for a boundary point of a region to be regular for the Dirichlet problem was found in 1924 by Norbert Wiener. The details are beyond the scope of this course, but here is the statement.

A relevant new concept is the *logarithmic capacity* of a compact subset of $\mathbb{C}$. The notion arises from physics. If a unit charge is put on a piece of metal, then the charge settles into some equilibrium position of minimum energy. In mathematical terms, the conductor is modeled by a compact set $K$, the charge distribution is modeled by a probability measure $\mu$ on $K$, and the associated energy is

$$\iint \log \frac{1}{|z-w|}\, d\mu(z)\, d\mu(w).$$

Rather than worry about whether there actually is a measure on the set $K$ that achieves the minimum, consider the infimum $I$ of the integrals over all probability measures on $K$. The integral might be divergent, and the standard procedure is to consider the related quantity $e^{-I}$, which is necessarily positive (if $I$ is finite) or zero (if $I$ is $+\infty$). This quantity $e^{-I}$ is called the logarithmic capacity of $K$.

If $K$ is a single point, then the only available measure is a point mass, so $I = +\infty$, and the capacity is 0. The logarithmic capacity of a line segment turns out to be one-quarter the length of the segment. The logarithmic capacity of a disk is equal to the radius of the disk.

Wiener's idea is to fix a positive parameter $\lambda$ less than 1 and to chop up a neighborhood of $z_0$ into annuli by considering the set

$$\{ z \in \partial G : \lambda^{n+1} \le |z - z_0| \le \lambda^n \}.$$

Let $c_n$ denote the logarithmic capacity of this compact set. The theorem is that $z_0$ is a regular point for the Dirichlet problem if and only if the infinite series

$$\sum_n \frac{n}{\log(1/c_n)} \qquad \text{diverges.}$$

(If some $c_n$ equals 0, then interpret the whole fraction as 0.)

Sanity check: If $z_0$ is a puncture in $G$, then all of the indicated sets are empty, so the series is a convergent series $0 + 0 + \cdots$, and $z_0$ is irregular. If the boundary of $G$ near $z_0$ is a smooth curve, then $c_n$ is comparable to $\lambda^n$, so the series diverges by comparison with $\sum_n 1$, and $z_0$ is a regular point.

An interesting example is the unit disk with a Cantor set removed from the real axis. The Cantor set is totally disconnected (that is, the only connected subsets are singletons), but no point of the Cantor set is isolated. Wiener's criterion can be used to show that every point of the Cantor set is a regular point for the Dirichlet problem.

# 8 The range of holomorphic functions

## 8.1 Bloch's theorem

If $f$ is a nonconstant holomorphic function defined on the unit disk, what can be said about the size of the range of $f$? Not much, for by the Riemann mapping theorem, the range can be an arbitrary simply connected proper subdomain of $\mathbb{C}$. When $f$ is not injective, the range can even be all of $\mathbb{C}$ (for instance, use a linear fractional transformation to map to the half-plane where $\operatorname{Re} z > -1$, and then compose with the squaring function). And the scaling mapping that sends $z$ to $z/n$ for a large natural number $n$ shows that the range can be a disk of tiny radius.

To rule out such rescaling, suppose that the derivative of the function at the origin has absolute value equal to 1. This normalization prevents shrinking the range in the obvious way, but could you perhaps make the range small in a subtle way? The surprise is that the normalization of the derivative at a single point constrains the range to be somewhere fat.

**Theorem** (André Bloch, 1924)**.** *There is a positive constant $\beta$ such that if $f$ is a holomorphic function defined on the unit disk, and $|f'(0)| = 1$, then the range of $f$ contains a (schlicht) disk of radius at least $\beta$.*

The point is that $\beta$ is independent of $f$. A *schlicht* disk in the range of $f$ means a disk that is the biholomorphic image under $f$ of some open set. In other words, a schlicht disk is a disk on which a branch of $f^{-1}$ is well defined. The range might contain non-schlicht disks of radius greater than $\beta$ if the image wraps around and covers some points more than once. The theorem says that if the derivative at the origin is normalized to magnitude 1, then there is some open subset of the unit disk that $f$ maps one-to-one onto some disk of radius at least $\beta$.

For applications, a precise value for $\beta$ is not needed: what is important is that some such constant exists. The proof below shows that radius $1/25$ works (but is not optimal).

Considerable effort has been devoted to seeking the supremum of the values of $\beta$ for which the conclusion of the theorem holds. This unknown cutoff value is called *Bloch's constant*. Lars V. Ahlfors (1907–1996), recipient of the first Fields Medal, published an influential paper in 1938 that generalizes the Schwarz lemma;[9] an application at the end of the article yields the value $\frac{1}{4}\sqrt{3}$ (approximately 0.433) as a lower bound for Bloch's constant. There was no quantitative improvement for half a century. In his 1988 thesis, Mario Bonk was able to increase the lower bound to $10^{-14} + \frac{1}{4}\sqrt{3}$, the point being that this value is a concrete number strictly larger than the bound of Ahlfors.[10] In 1937, Ahlfors and H. Grunsky had given an example[11] of a holomorphic function that constrains Bloch's constant to be no greater than

$$\sqrt{\frac{\sqrt{3}-1}{2}} \cdot \frac{\Gamma(1/3)\Gamma(11/12)}{\Gamma(1/4)}$$

(approximately 0.472), where $\Gamma$ is the standard Gamma function that interpolates the values of the factorial function. This upper bound is conjectured to be the precise value of Bloch's constant, but the problem remains open.

A curious bit of history[12] is that Bloch (1893–1948) was confined to a psychiatric hospital after murdering three family members in 1917. (The circumstances are murky; perhaps Bloch suffered from post-traumatic stress syndrome resulting from his war service.) He proved his theorem while institutionalized.

*Proof of Bloch's theorem.* To get started, suppose that $f$ is holomorphic in a neighborhood of the closed disk. (This special assumption will be removed at the end of the proof.) Under this assumption, both $f$ and $f'$ are bounded functions in the disk.

If $\varphi$ is an automorphism of the unit disk, then $f$ and $f\circ\varphi$ have the same range, but the composite function $f\circ\varphi$ is not normalized at the origin. Explicit computation shows that if $c = \varphi(0)$, then $|(f\circ\varphi)'(0)| = |f'(c)|(1-|c|^2)$. The right-hand side tends to 0 when $|c| \to 1$ (under the assumption from the first paragraph that $f'$ is bounded), so there is some point $c$ inside the disk for which $|f'(c)|(1-|c|^2)$ is maximized. That maximal value of $|(f\circ\varphi)'(0)|$ is no smaller than 1 (which is the particular value corresponding to $c$ equal to 0).

[9] An extension of Schwarz's lemma, *Transactions of the American Mathematical Society* **43** (1938), no. 3, 359–364.
[10] Mario Bonk, On Bloch's constant, *Proceedings of the American Mathematical Society* **110** (1990), no. 4, 889–894.
[11] Über die Blochsche Konstante, *Mathematische Zeitschrift* **42** (1937), no. 1, 671–673.
[12] Douglas M. Campbell, Beauty and the beast: The strange case of Andre Bloch, *The Mathematical Intelligencer* **7** (1985), no. 4, 36–38.

Let $g$ denote $f \circ \varphi$ for an automorphism $\varphi$ (not necessarily unique) that realizes the maximum. If $b$ is an arbitary nonzero point in the unit disk, and $\varphi_b$ is the disk automorphism that interchanges 0 and $b$, then $|(g \circ \varphi_b)'(0)| = |g'(b)|(1 - |b|^2)$. Since $g$ is obtained by composing $f$ with a disk automorphism, so is $g \circ \varphi_b$, and the maximality of $g$ implies that $|g'(b)|(1 - |b|^2) \le |g'(0)|$. Equality holds in this inequality when $b = 0$.

Now let $h(z)$ denote $\{g(z) - g(0)\}/g'(0)$. The reason for defining $h$ this way is that $h(0) = 0$ and $h'(0) = 1$. If the range of $h$ can be shown to contain a schlicht disk of radius $1/25$, then the range of $g$ will contain a schlicht disk of radius $|g'(0)|/25$. Since $|g'(0)| \ge 1$, the range of $f$ will contain a schlicht disk of radius at least $1/25$.

The lower bound on $|g'(0)|$ implies that if $|z|^2 \le 1/2$, then

$$|h'(z)| = \frac{|g'(z)|}{|g'(0)|} \le \frac{1}{1 - |z|^2} \le \frac{1}{1 - \frac{1}{2}} = 2.$$

Evidently $h(z) = h(z) - h(0) = \displaystyle\int_0^1 \frac{d}{dt} h(tz)\, dt$. If $|z| \le 1/2$, then $|z|^2 \le 1/4$, and

$$|h(z)| \le \left| \int_0^1 z h'(tz)\, dt \right| \le |z| \int_0^1 |h'(tz)|\, dt \le \tfrac{1}{2} \cdot 2 = 1.$$

Cauchy's estimate for derivatives, applied on the disk of radius $1/2$, implies that the $n$th Maclaurin coefficient of $h$ has modulus no greater than $2^n$. So if $|z| \le 1/8$, then

$$|h(z) - z| \le \sum_{n=2}^{\infty} 2^n |z|^n = \frac{(2|z|)^2}{1 - 2|z|} \le \frac{\left(\frac{1}{4}\right)^2}{1 - \frac{1}{4}} = \frac{1}{12}.$$

If $w$ is a point whose modulus is less than $1/24$, and $|z| = 1/8$, then the preceding inequality implies that $|(h(z) - w) - (z - w)| \le 1/12 < |z - w|$. By Rouché's theorem, there is exactly one point inside the disk of radius $1/8$ that $h$ maps to $w$. In other words, the range of $h$ contains a schlicht disk of radius $1/24$ centered at the origin. Consequently, the range of $f$ contains a schlicht disk of radius at least $1/24$ centered at $g(0)$.

The preceding analysis assumes that $f$ is holomorphic in a neighborhood of the closed disk. If instead $f$ is holomorphic only on the open disk, then consider the function sending $z$ to $r^{-1} f(rz)$ when $r$ is a real number less than 1 and greater than $24/25$. This function is normalized at the origin and is holomorphic in a neighborhood of the closed disk, so the range contains a schlicht disk of radius $1/24$. Therefore the range of the function sending $z$ to $f(rz)$ contains a schlicht disk of radius $1/25$. Equivalently, the restriction of $f$ to the disk of radius $r$ maps some open set biholomorphically onto a disk of radius $1/25$. Hence the original function $f$ does so too. □

## 8.2 Schottky's theorem

The theorem is named for the German mathematician Friedrich Hermann Schottky (1851–1935). The theorem dates from the first decade of the twentieth century. Subsequently, many authors found new proofs and refinements. The 1938 paper of Ahlfors cited above proves a strong version of the theorem. The proof given below deduces Schottky's theorem from Bloch's theorem.

**Theorem** (Schottky's theorem)**.** *There is a function $\varphi$ of two real variables, increasing with respect to each variable, such that if $f$ is a holomorphic function in the unit disk that takes neither of the values* 0 *and* 1*, then* $|f(z)| \leq \varphi(|z|, |f(0)|)$ *for every point $z$ in the disk.*

The main point is that $\varphi$ is a universal function, independent of $f$. In particular, if $|f(0)|$ is bounded above by some value $s$, and $z$ lies in the disk of radius $r$ (where $r$ is less than 1), then $\max_{|z|\leq r} |f(z)| \leq \varphi(r, s)$. The proof will actually produce an explicit formula for $\varphi$.

**Corollary** (Montel's theorem)**.** *The family of holomorphic functions in the unit disk whose range omits the values* 0 *and* 1 *is a normal family in the generalized sense that every sequence of functions in the family either has a subsequence converging uniformly on compact sets to a holomorphic function or has a subsequence converging uniformly on compact sets to the constant* $\infty$*.*

*Proof of Montel's theorem.* Suppose $\{f_n\}$ is a sequence of holomorphic functions mapping the unit disk into $\mathbb{C} \setminus \{0, 1\}$, the twice-punctured plane. First suppose there are infinitely many values of $n$ for which $|f_n(0)| \leq 1$. Schottky's theorem implies that the corresponding family of functions is locally bounded: on the compact set where $|z| \leq r$, the functions have absolute value bounded above by $\varphi(r, 1)$, a bound that is independent of the holomorphic function. The easy theorem of Montel from § 2.3 implies that there is a normally convergent subsequence.

On the other hand, if there are only finitely many such values of $n$, then there are infinitely many values of $n$ for which $|1/f_n(0)| < 1$. Again by Schottky's theorem, there is a subsequence $\{1/f_{n_k}\}$ that converges uniformly on compact sets to a holomorphic function. By Hurwitz's theorem from § 2.7, that limit function is either nowhere zero or identically zero. If the limit is identically 0, then the sequence $\{f_{n_k}\}$ converges normally to $\infty$. Otherwise, the sequence $\{f_{n_k}\}$ converges normally to a holomorphic function. □

Normality is a local property, and there is nothing special about the particular disk $B(0; 1)$. Thus Montel's theorem immediately generalizes to say that if $G$ is an arbitrary open set in $\mathbb{C}$, then the family of holomorphic functions mapping $G$ into $\mathbb{C} \setminus \{0, 1\}$ is a normal family in the generalized sense.

*Proof of Schottky's theorem.* Suppose that $f$ is holomorphic in the unit disk, and the range of $f$ omits the values 0 and 1. The plan is to cook up a new function whose value at 0 is controlled by $|f(0)|$ and whose range omits a lattice. By Bloch's theorem, that information will give control on the derivative, hence control on the function by integrating. Bloch's theorem from the preceding section implies that there is a constant $\beta$ (the value $1/25$ will do) such that if $f$ is holomorphic in a disk $B(c; R)$, then the range of $f$ contains a (schlicht) disk of radius at least $\beta R|f'(c)|$.

Observe that if $g$ is a holomorphic function defined on a simply connected region and omitting the values 1 and $-1$, then $g^2 - 1$ is never equal to 0, so a holomorphic branch of $\sqrt{g^2 - 1}$ can be defined on the region. And $g + \sqrt{g^2 - 1}$ is never equal to 0 (since $g^2 \neq g^2 - 1$), so a holomorphic branch of $\log(g + \sqrt{g^2 - 1}\,)$ can be defined. A routine calculation shows that

$\cos i \log(g + \sqrt{g^2 - 1}) = g$. Thus a holomorphic function omitting the values 1 and $-1$ on a simply connected region can be written as the cosine of another holomorphic function on the region.

If $f$ omits the values 0 and 1, then $2f - 1$ omits the values 1 and $-1$, so $2f - 1 = \cos(\pi f_1)$ for some holomorphic function $f_1$. The function $f_1$ is not unique, for any integer multiple of 2 can be added to $f_1$. Hence $f_1$ can be chosen to have the property that $-1 \le \operatorname{Re} f_1(0) \le 1$. Moreover, the imaginary part of $f_1(0)$ can be controlled by $|f(0)|$. Indeed, $|y| \le |\sinh(y)| \le |\cos(x + iy)|$, so

$$\pi |\operatorname{Im} f_1(0)| \le |\cos(\pi f_1(0))| \le 2|f(0)| + 1.$$

Accordingly,

$$|f_1(0)| \le |\operatorname{Re} f_1(0)| + |\operatorname{Im} f_1(0)| \le 1 + \frac{2}{\pi}|f(0)| + \frac{1}{\pi} < 2 + |f(0)|.$$

(The goal here is not to get an optimal bound but rather to get a simple and explicit bound.)

Since $\cos(\pi f_1)$ omits the values 1 and $-1$, the function $f_1$ omits all integer values. In particular, the function $f_1$ omits the values 1 and $-1$, so there is a function $f_2$ such that $f_1 = \cos(i\pi f_2)$. Since $f_2$ is determined only up to addition of an integer multiple of $2i$, the function $f_2$ can be chosen to have the property that $-1 \le \operatorname{Im} f_2(0) \le 1$. Reasoning as before shows that

$$\pi |\operatorname{Re} f_2(0)| \le |\cos(i\pi f_2(0))| = |f_1(0)| < 2 + |f(0)|.$$

Therefore

$$|f_2(0)| \le 1 + \frac{2}{\pi} + \frac{1}{\pi}|f(0)| < 2 + |f(0)|.$$

The function $f_2$ omits a whole lattice of values. Indeed, for an arbitary branch of the logarithm, the cosine of $\pm i \log(k + \sqrt{k^2 - 1})$ is equal to $k$, so the function $f_2$ omits all possible values of $\pm\pi^{-1} \log(k + \sqrt{k^2 - 1})$ for every natural number $k$. The spacing between omitted values in the imaginary direction is 2. The spacing between omitted values in the real direction is variable and decreases as $k$ increases. The maximal spacing in the real direction is therefore no more than $\pi^{-1}(\log(2 + \sqrt{3}) - \log 1)$, which is less than $\pi^{-1} \log 4$. The radius of the largest disk that misses all lattice points is less than half the diagonal spacing, hence less than half the sum $2 + \pi^{-1} \log 4$, hence less than 2.

If $|z| \le r < 1$, then $f_2$ is holomorphic on a disk centered at $z$ of radius at least $1 - r$. By Bloch's theorem, there is a disk in the range of $f_2$ of radius at least $\beta(1 - r)|f_2'(z)|$. Hence $|f_2'(z)| < 2\beta^{-1}(1 - r)^{-1}$. Integrating along a line segment from 0 to $z$ shows that $|f_2(z)| \le |f_2(0)| + 2\beta^{-1}r(1 - r)^{-1}$ when $|z| \le r$.

This inequality induces an estimate on $|f(z)|$. Since $|\cos z| \le \cosh |z|$, taking $\beta$ equal to $1/25$ shows that if $|f(0)| \le s$ and $|z| \le r$, then

$$\begin{aligned}
|f(z)| &\le \tfrac{1}{2} + \tfrac{1}{2}|\cos(\pi \cos(\pi i f_2(z)))| \\
&\le \tfrac{1}{2} + \tfrac{1}{2}\cosh(\pi \cosh(\pi(2 + |f(0)| + 2\beta^{-1}r(1 - r)^{-1}))) \\
&< \cosh(\pi \cosh(\pi(2 + s + 50r(1 - r)^{-1}))). \qquad \square
\end{aligned}$$

Modest improvements in the upper bound are available through essentially the same method. The main point is that there *is* an upper bound depending only on the upper bound for $|f(0)|$ and the distance of $|z|$ from the boundary of the unit disk.

## 8.3 Proofs of Picard's theorems

Picard's great theorem says that in every neighborhood of an essential singularity, a holomorphic function takes every complex value with one possible exception. Repeatedly shrinking the neighborhood shows that every value except possibly one is actually taken infinitely many times.

To prove the great theorem, suppose without loss of generality that $f$ has an essential singularity at the origin. Seeking a contradiction, suppose there are two distinct complex numbers $a$ and $b$ that $f$ takes only finitely many times. Shrinking the neighborhood reduces to the case that these two values are not taken at all. And considering the function $(f - a)/(b - a)$ reduces to the case that the omitted values are 0 and 1. Dilating the independent variable shows that the punctured neighborhood can be taken to be the punctured unit disk.

For each natural number $n$, define $f_n$ to be the function sending $z$ to $f(z/n)$, and consider the family $\{f_n\}$ in the punctured disk. By Montel's fundamental normality criterion, this family is normal in the extended sense. There are two cases.

First suppose there is a subsequence $\{f_{n_k}\}$ converging normally to a holomorphic function. The circle of radius $1/2$ is a compact set, so there is some bound $M$ such that $|f_{n_k}(z)| \leq M$ for every $k$ when $|z| = 1/2$. Unwinding the definition of $f_n$ shows that there is a sequence of annuli with outer radius $1/2$ and inner radius approaching 0 such that $|f|$ is bounded by $M$ on the boundary of each annulus, hence on the whole annulus (by the maximum principle). Therefore $|f|$ is bounded by $M$ on the union of the annuli, which is the whole punctured disk of radius $1/2$. By Riemann's theorem on removable singularities, the singularity at the origin is removable, contrary to the hypothesis.

Next suppose there is a subsequence converging normally to $\infty$. The corresponding subsequence of reciprocals converges normally to 0. By the preceding argument, the reciprocal of the original function has a removable singularity, and the singularity is removed by setting the value at the origin to be 0. Hence the original function has a pole, again contrary to the hypothesis.

Thus the assumption that $f$ omits two values contradicts the hypothesis that the singularity is essential. The proof of Picard's great theorem is complete.

Picard's little theorem says that a nonpolynomial entire function takes every complex value, with one possible exception, infinitely many times. This result is a corollary of the great theorem because a nonpolynomial entire function has an essential singularity at infinity.