Lecture Notes in Complex Analysis

Eric T. Sawyer

McMaster University, Hamilton, Ontario
E-mail address: `sawyer@mcmaster.ca`

Contents

Preface

These notes grew out of lectures given three times a week in a fourth year undergraduate course in complex analysis at McMaster University January to April 2009. I'd like to thank all the students taking the course, in particular Martin Munoz and Preston Wake, for a number of interesting observations and for correcting many mistakes.

Analysis in the complex field $\mathbb{C}$ has many seemingly magical advantages over analysis in the real field $\mathbb{R}$. For example, if a function f from an open subset $\Omega \subset \mathbb{C}$ to $\mathbb{C}$ has at every point $z \in \Omega$, a *derivative* in the complex sense:

$$f'(z) = \lim_{w(\in\mathbb{C})\to z} \frac{f(w) - f(z)}{w - z},$$

then f actually has a power series expansion

$$f(z) = \sum_{n=0}^{\infty} a_n (z - w)^n ,$$

valid for all $z \in B(w, R)$ whenver $B(w, R) \subset \Omega$. In particular f is *infinitely* differentiable in Ω. This is in striking contrast to the situation for real-valued functions f that are differentiable on an interval in the usual real sense - the function $f(x) = x^2 \sin \frac{1}{x}$ is differentiable on the real line, but its derivative f' fails to be even continuous at the origin. A famous example of Weierstrass exhibits a continuously differentiable function f that nowhere has a second derivative!

There is a plethora of special properties enjoyed by functions having a complex derivative at every point in an open subset of the complex plane. These include in particular Cauchy's theorem and representation formula, the maximum modulus principle, the open mapping theorem, uniform convergence and Montel's theorem, the residue theorem and many others as given in Rudin [**5**] and Boas [**1**]. An application to the Prime Number Theorem is given following Stein and Shakarchi [**6**] where a very important complex differentiable function is studied - the Riemann zeta function and Euler's formula,

$$\zeta(s) = \sum_{n=1} \frac{1}{n^s} = \prod_p \frac{1}{1 - \frac{1}{p^s}},$$

where the infinite product is taken over all prime numbers.

On the other hand, the famous Riemann Mapping Theorem shows that there are lots of these special functions, enough to realize any homeomorphism of the unit disk to a proper subset of the plane by a biholomorphic function! This latter theorem requires antidifferentiation, hence integration over paths, in simply connected domains. We treat this matter using taxicab paths as introduced in [**1**], together with an elementary approach to the Jordan Curve Theorem for such paths initiated in [**6**].

In the second part of these notes we turn to Carathéodory's characterization of when a Riemann map extends to a homeomorphism up to the boundary. This introduces the study of continuous closed curves in the complex plane, including simple boundary points and the Jordan Curve Theorem for general curves.

Moreover, we require for the first time Lebesgue's theory of the integral in order to obtain Fatou's Theorem on the existence of radial limits for bounded holomorphic functions in the unit disk. Lebesgue's theory renders the space of square integrable functions on the circle *complete*, and thus allows both the Cauchy integral and the Poisson integral to integrate boundary values on the unit circle. The Poisson integral has two important advantages over the Cauchy integral: its much smaller size permits the study of radial limits, and its real (positive) values permit the development of a theory of harmonic functions. Together with the maximal theorem from real analysis, this circle of ideas provides one of the early triumphs of Lebesgue's integral over that of Riemann, which until then had sufficed for the purposes of complex analysis.

Part 1

Complex Differentiation

We begin Part 1 with a short review chapter on the concepts of differentiation using field and metric properties in $\mathbb{R}$, and differentiation using vector space and metric properties in $\mathbb{R}^n$. In Chapter 2, we introduce *complex* differentiation using field and metric properties in $\mathbb{C}$, and compare this to the *real* differentiation arising in Chapter 1 from the identification of $\mathbb{C}$ and $\mathbb{R}^2$. The striking Cauchy-Goursat theorem on complex derivatives is proved, along with Cauchy's representation formula and the consequent power series expansions of holomorphic functions. Chapter 3 develops the elementary properties of holomorphic functions; zeroes and singularities, the maximum principle, uniform convergence and normal families, and the open mapping theorem. Chapter 4 uses the pivotal Schwarz lemma and the theory of simply connected domains to prove the Riemann Mapping Theorem. Finally, Chapter 5 uses the residue theorem and analytic continuation of the Riemann zeta function to prove the Prime Number Theorem.

Recall that the field of real numbers $\mathbb{R}$ can be constructed either from *Dedekind cuts* of rational numbers $\mathbb{Q}$, or from Weierstrass' *Cauchy sequences* of rational numbers. A Dedekind cut $\alpha \subset \mathbb{Q}$ is a "left infinite interval open on the right" of rational numbers that is associated with the "real number" on the number line that marks its right hand endpoint. More precisely, a cut α is a subset of $\mathbb{Q}$ satisfying (here p and q denote rational numbers)

$$\begin{aligned} \alpha &\neq \varnothing \text{ and } \alpha \neq \mathbb{Q}, \\ p &\in \alpha \text{ and } q < p \text{ implies } q \in \alpha, \\ p &\in \alpha \text{ implies there is } q \in \alpha \text{ with } p < q. \end{aligned}$$

One can define an ordered field structure on the set of cuts, and this is identified as the field $\mathbb{R}$ (see e.g. Chapter 1 of [**4**]). Alternatively, one can define an ordered field structure on the set of equivalence classes of Cauchy sequences in $\mathbb{Q}$, and this produces an ordered field isomorphic to $\mathbb{R}$.

The field of complex numbers can now be easily constructed from the real linear space $\mathbb{R}^2$ by defining the following multiplication on ordered pairs in $\mathbb{R}^2$:

$$(x,y)(u,v) = (xu - yv, xv + yu), \qquad (x,y),(u,v) \in \mathbb{R}^2.$$

It is a routine matter to verify that a field structure is defined on $\mathbb{R}^2$ with addition as usual in $\mathbb{R}^2$, multiplication as above, $(0,0)$ the additive identity, and $(1,0)$ the multiplicative identity. It is customary to introduce an abstract symbol i and write

$$\mathbb{R}^2 = \mathbb{R}(1,i) = \{x + iy : x, y \in \mathbb{R}\}.$$

Thus i represents the ordered pair $(0,1)$ and hence $i^2 = -1$. The field constructed above is now denoted $\mathbb{C}$ and an element $z \in \mathbb{C}$ has a unique representation as $z = x + iy$ where x and y are called the real and imaginary parts of z. As there is no square root of -1 in the ordered field $\mathbb{R}$, the number i was historically considered "not real, but imaginary", and this accounts for the terminology used today.

CHAPTER 1

The Real Field

We take as known the real field $\mathbb{R}$ and its association with the "points in a line". In addition to the field structure on $\mathbb{R}$, we also consider the metric topology given by the distance function

$$d(x,y) = |x-y|, \qquad x, y \in \mathbb{R}.$$

With these algebraic and topological definitions in hand, we can proceed with analysis, in particular the definition of the *derivative* f' of a function $f : \mathbb{R} \to \mathbb{R}$.

1. The derivative of a real-valued function on $\mathbb{R}$

Definition 1. *Given a function $f : \mathbb{R} \to \mathbb{R}$ and a point $x \in \mathbb{R}$, we define*

$$f'(x) = \lim_{y \to x} \frac{f(y) - f(x)}{y - x} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}, \tag{1.1}$$

for those $x \in \mathbb{R}$ for which the limit exists.

We can also rewrite the second form of the limit using Landau's "little oh" notation as

$$f(x+h) = f(x) + f'(x)\,h + o(h), \tag{1.2}$$

where $o(h)$ denotes a function of h satisfying $\frac{o(h)}{h} \to 0$ as $h \to 0$. We say that the function f is differentiable at x if (1.1) (equivalently (1.2)) holds. From (1.2) we note that if f is differentiable at x then it is also continuous at x. From these definitions it is easy to prove the standard "calculus" of derivatives. First we have the calculus of the field operations.

Proposition 1. *Suppose that f and g are both functions differentiable at the point $x \in \mathbb{R}$, and suppose that $c \in \mathbb{R}$ represents the constant function. Then we have*

(1) $(f+g)'(x) = f'(x) + g'(x)$,
(2) $(cf)'(x) = cf'(x)$,
(3) $(fg)'(x) = f'(x)\,g(x) + f(x)\,g'(x)$,
(4) $\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$ *provided* $g(x) \neq 0$.

For example, to prove (3) we use the formulation (1.1) of derivative and the corresponding properties of limits to obtain

$$\begin{aligned}
(fg)'(x) &= \lim_{y\to x}\frac{(fg)(y)-(fg)(x)}{y-x}\\
&= \lim_{y\to x}\left\{\frac{f(y)g(y)-f(x)g(y)}{y-x}+\frac{f(x)g(y)-f(x)g(x)}{y-x}\right\}\\
&= \lim_{y\to x}\frac{f(y)-f(x)}{y-x}\lim_{y\to x}g(y)+f(x)\lim_{y\to x}\frac{g(y)-g(x)}{y-x}\\
&= f'(x)g(x)+f(x)g'(x).
\end{aligned}$$

Second we have the calculus of composition of functions, the so-called "chain rule".

Proposition 2. *Suppose that f is differentiable at x and that g is differentiable at $y=f(x)$. Then*

(1) $(g\circ f)'(x)=g'(y)f'(x)=g'(f(x))f'(x)$,
(2) $\left(f^{-1}\right)'(y)=\frac{1}{f'(x)}=\frac{1}{f'(f^{-1}(y))}$ *provided f is invertible near x.*

For example, to prove (1) we use the Landau formulation (1.2) of derivative and the corresponding properties of limits as follows. Write

$$\begin{aligned}
f(x+h_1) &= f(x)+f'(x)h_1+o_1(h_1),\\
g(y+h_2) &= g(y)+g'(y)h_2+o_2(h_2),
\end{aligned}$$

and then with $h_2=f'(x)h_1+o_1(h_1)$ we have,

$$\begin{aligned}
(g\circ f)(x+h_1) &= g(f(x+h_1))\\
&= g(f(x)+f'(x)h_1+o_1(h_1))\\
&= g(y+h_2)\\
&= g(y)+g'(y)h_2+o_2(h_2)\\
&= (g\circ f)(x)+g'(y)\{f'(x)h_1+o_1(h_1)\}\\
&\quad +o_2(f'(x)h_1+o_1(h_1))\\
&= (g\circ f)(x)+g'(y)f'(x)h_1+o_3(h_1)
\end{aligned}$$

where it is easy to see that

$$\frac{o_3(h_1)}{h_1}\equiv\frac{g'(y)o_1(h_1)+o_2(f'(x)h_1+o_1(h_1))}{h_1}\to 0 \text{ as } h_1\to 0.$$

To prove (2) we first note that since f is continuous and one-to-one, it follows easily that f is open, hence f^{-1} is continuous. Then with obvious notation,

$$\frac{f^{-1}(y+k)-f^{-1}(y)}{k}=\frac{h}{f(x+h)-f(x)}\to\frac{1}{f'(x)}$$

as $k\to 0$.

Example 1. *There is a function $f:\mathbb{R}\to\mathbb{R}$ whose derivative $f':\mathbb{R}\to\mathbb{R}$ exists everywhere on the real line, but the derivative function f' is* not *itself differentiable at* 0. *For example*

$$f(x)=\begin{cases} x^2\sin\frac{1}{x} & \text{if} \quad x\neq 0\\ 0 & \text{if} \quad x=0\end{cases}$$

has these properties. Indeed,

$$f'(x) = \begin{cases} 2x \sin \frac{1}{x} - \cos \frac{1}{x} & \textit{if} \quad x \neq 0 \\ 0 & \textit{if} \quad x = 0 \end{cases}$$

fails even to be continuous at the origin.

1.1. Derivatives in Euclidean space. We can extend the definition of derivative to functions $f : \mathbb{R}^n \to \mathbb{R}^m$ from one Euclidean space to another using the Landau formulation (1.2) together with the vector space structure of $\mathbb{R}^n$ and the metric topology on $\mathbb{R}^n$ given by the distance function

$$d(x, y) = \sqrt{\sum_{k=1}^{n} |x_k - y_k|^2}, \qquad x = (x_k)_{k=1}^n, y = (y_k)_{k=1}^n.$$

More precisely, we say that a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x = (x_k)_{k=1}^n \in \mathbb{R}^n$ if there is an $m \times n$ matrix A (equivalently a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$) such that

$$f(x + h) = f(x) + Ah + o(h),$$

for all $h \in \mathbb{R}^n$ and where the function $o(h)$ taking values in $\mathbb{R}^m$ satisfies $\frac{|o(h)|}{|h|} \to 0$ as $h \to 0 = (0, ..., 0)$ in $\mathbb{R}^n$. When such a matrix A exists, it is easily seen to be unique and we define the derivative $Df(x)$ at x to be the linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ with matrix representation A. If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable on all of $\mathbb{R}^n$, then $Df : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ where $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ denotes the space of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$. When $Df(x)$ exists, we thus have

$$f(x + h) = f(x) + Df(x)h + o(h), \qquad h \in \mathbb{R}^n, \tag{1.3}$$

which can be interpreted as stating that the linear map $Df(x)$ is the "best linear approximation" to the nonlinear map $h \to f(x + h) - f(x)$. In paricular, f is continuous at x if it is differentiable at x. The calculus in Proposition 1 extends easily with obvious modifications to this setting with similar proofs.

Finally, we recall that the matrix representation $[Df(x)]$ of $f'(x)$ has entries given by the partial derivatives

$$\frac{\partial f_j}{\partial x_k}(x) = \lim_{\eta \to 0} \frac{f_j(x_1, ..., x_k + \eta, ..., x_n) - f_j(x_1, ..., x_n)}{\eta},$$

of the components f_j of f with respect to the k^{th} variable x_k:

$$[Df(x)] = \left[\frac{\partial f_j}{\partial x_k}(x)\right]_{\substack{1 \leq j \leq m \\ 1 \leq k \leq n}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_n}(x) \\ & \\ \frac{\partial f_m}{\partial x_n}(x) & \frac{\partial f_m}{\partial x_n}(x) \end{bmatrix}. \tag{1.4}$$

The chain rule, Proposition 2, here takes the form

$$D(g \circ f)(x) = Dg(y)\, Df(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at x and $g : \mathbb{R}^m \to \mathbb{R}^p$ is differentiable at $y = f(x)$, and where the matix representations of the derivatives satisfy

$$[D(g \circ f)(x)] = [Dg(y)]\,[Df(x)],$$

with the multiplication of $[Dg(y)]$ and $[Df(x)]$ on the right side being matrix multiplication of a $p \times m$ matrix times an $m \times n$ matrix. The proof is similar to that in the case $m = n = 1$ given above.

CHAPTER 2

The Complex Field

We also take as known the complex field $\mathbb{C}$ and its association with the "points in the plane $\mathbb{R}^2$" where $z = x + iy \in \mathbb{C}$ is associated with $(x, y) \in \mathbb{R}^2$. The field structure on $\mathbb{C}$ uses the multiplication rule

$$zw = (x + iy)(u + iv) = (xu - yv) + i(xv + yu),$$

where $z = x + iy$ and $w = u + iv$. If we associate $z = x + iy$ to the matrix $\begin{bmatrix} x & -y \\ y & x \end{bmatrix}$, then this multiplication corresponds to matrix multiplication:

$$\begin{aligned} [z][w] &= \begin{bmatrix} x & -y \\ y & x \end{bmatrix} \begin{bmatrix} u & -v \\ v & u \end{bmatrix} \\ &= \begin{bmatrix} xu - yv & -xv - yu \\ yu + xv & -yv + xu \end{bmatrix} = [zw]. \end{aligned} \tag{0.5}$$

Since the matrix

$$\begin{bmatrix} x & -y \\ y & x \end{bmatrix} = r \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

is dilation by the nonnegative number $r = \sqrt{x^2 + y^2}$ and rotation by the angle $\theta = \tan^{-1} \frac{y}{x}$ in the counterclockwise direction, we see that if z has polar coordinates (r, θ) and w has polar coordinates (s, ϕ), then zw has polar coordinates $(rs, \theta + \phi)$.

We also consider the metric topology on $\mathbb{C}$ given by the distance function

$$d(z, w) = \sqrt{|x - u|^2 + |y - v|^2}, \qquad z = x + iy, w = u + iv,$$

which coincides with the usual distance function in the plane $\mathbb{R}^2$. With these algebraic and topological definitions in hand, we can define derivatives $f'(x)$ of functions $f : \mathbb{C} \to \mathbb{C}$ at points x just as in (1.1) and (1.2), but where x, y, h now denote complex numbers. Then the analogues of the two calculus propositions above hold for the complex field as well.

Definition 2. *Given a function $f : \mathbb{C} \to \mathbb{C}$ and a point $z \in \mathbb{C}$, we define*

$$f'(z) = \lim_{w \to z} \frac{f(w) - f(z)}{w - z} = \lim_{h \to 0} \frac{f(z + h) - f(z)}{h},$$

for those $z \in \mathbb{C}$ for which the limit exists. Equivalently, using Landau's "little oh" notation this is

$$f(z + h) = f(z) + f'(z) h + o(h),$$

where $o(h)$ denotes a function of h satisfying $\frac{o(h)}{h} \to 0$ as $h \to 0$.

Analogues of Propositions 1 and 2 hold for complex derivatives. In particular, since a constant function $f(z) = c$ has derivative 0, and the identity function $f(z) =$

z has derivative 1, we see that the polynomial $f(z) = \sum_{n=0}^{N} a_n z^n$ is holomorphic in $\mathbb{C}$ and has derivative $f'(z) = \sum_{n=1}^{N} n a_n z^{n-1}$.

1. Two derivatives

We now have two *different* definitions of derivative of a function $f : \mathbb{C} \to \mathbb{C}$, namely the complex derivative $f'(z)$ using the field structure of $\mathbb{C}$, and also the real derivative $Df(x,y)$ given in (1.3) of Chapter 1 with $z = x + iy$ using the identification of $\mathbb{C}$ with $\mathbb{R}^2$. Note that if f has a complex derivative $f'(z)$ at $z = x + iy$, then (1.3) holds with the linear map $Df(x,y) \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$ given by

$$h = \alpha + i\beta \to f'(z)\, h, \qquad h \in \mathbb{C} = \mathbb{R}^2. \tag{1.1}$$

If we write the real and imaginary components of f as u and v, i.e. $f(w) = u(w) + iv(w)$, then (1.4) shows that

$$[Df(x,y)] = \begin{bmatrix} \frac{\partial u}{\partial x}(x,y) & \frac{\partial u}{\partial y}(x,y) \\ \frac{\partial v}{\partial x}(x,y) & \frac{\partial v}{\partial y}(x,y) \end{bmatrix}. \tag{1.2}$$

On the other hand, (0.5) shows that

$$[f'(z)] = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \tag{1.3}$$

for some real numbers a and b. Equating (1.2) and (1.3) gives the Cauchy-Riemann equations:

$$\begin{aligned} \frac{\partial u}{\partial x}(z) &= \frac{\partial v}{\partial y}(z), \\ \frac{\partial u}{\partial y}(z) &= -\frac{\partial v}{\partial x}(z). \end{aligned} \tag{1.4}$$

Conversely, it is easy to see that if $Df(x,y)$ exists and the Cauchy-Riemann equations (1.4) hold, then $f = u + iv$ has a complex derivative at $z = x + iy$. We summarize this discussion in the following proposition.

Proposition 3. *Suppose that $f : \mathbb{C} \to \mathbb{C}$ and that the real derivative $Df(x,y)$ exists at $z = x+iy$. Then $f = u+iv$ has a complex derivative $f'(z)$ at z if and only if the Cauchy-Riemann equations (1.4) hold, equivalently if and only if $Df(x,y)$ is the composition of a dilation and a rotation.*

However, magical properties arise from the definition of derivative in the complex field! For example, if $f : \mathbb{C} \to \mathbb{C}$ is complex differentiable at all points in $\mathbb{C}$, then the derivative function $f' : \mathbb{C} \to \mathbb{C}$ is *also complex differentiable at all points* in $\mathbb{C}$. As a consequence, the derivative f'' of f' exists and is differentiable in $\mathbb{C}$ and in fact f is complex differentiable of all orders. This is a far cry from the example given at the end of the first chapter. Note however that the function

$$f(z) = \begin{cases} z^2 \sin \frac{1}{z} & \text{if} \quad z \neq 0 \\ 0 & \text{if} \quad z = 0 \end{cases},$$

obtained by replacing x with z in Example 1, fails to be bounded in any neighbourhood of the origin in $\mathbb{C}$. Here $\sin z = \frac{e^{iz} - e^{-iz}}{2i}$ where e^z is defined in the next section.

Definition 3. *Let Ω be an open subset of the complex plane $\mathbb{C}$. We say that $f : \Omega \to \mathbb{C}$ is* holomorphic *in Ω if $f'(z)$ exists for all $z \in \Omega$.*

The proof that holomorphic functions are infinitely differentiable is not easy, and it will occupy the next few sections. But first we review the important exponential function, and then the useful complex partial derivatives $\frac{\partial}{\partial z}$ and $\frac{\partial}{\partial \bar{z}}$.

2. The exponential function

DEFINITION 4. *For $z \in \mathbb{C}$ define*

$$e^z \equiv \exp z \equiv \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

Note that by the root test for series, the series $\sum_{n=0}^{\infty} \frac{z^n}{n!}$ converges absolutely and uniformly in each compact subset of the complex plane. We have the exponent formula:

$$e^{z+w} = e^z e^w, \qquad z, w \in \mathbb{C}. \tag{2.1}$$

To see this we write

$$\begin{aligned} e^{z+w} &= \sum_{n=0}^{\infty} \frac{(z+w)^n}{n!} = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} z^k w^{n-k} \\ &= \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{1}{k!\,(n-k)!} z^k w^{n-k} \\ &= \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \frac{1}{k!j!} z^k w^j = \sum_{k=0}^{\infty} \frac{1}{k!} z^k \sum_{j=0}^{n} \frac{1}{j!} w^j = e^z e^w, \end{aligned}$$

where the absolute convergence of the exponential series justifies the substitution $j = n - k$ with n fixed.

We also compute that

$$\begin{aligned} e^{i\theta} &= \sum_{n=0}^{\infty} \frac{(i\theta)^n}{n!} \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \right) + i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \right) \\ &\equiv \cos\theta + i\sin\theta, \end{aligned} \tag{2.2}$$

and using the exponent formula we obtain de Moivre's formula:

$$\left(e^{i\theta}\right)^n = e^{in\theta} = \cos n\theta + i \sin n\theta.$$

Thus if $z = x + iy$ has polar coordinates (r, θ), then

$$z = r\cos\theta + ir\sin\theta = re^{i\theta}$$

and

$$z^n = r^n e^{in\theta}.$$

More generally, if $w = se^{i\phi}$, then

$$zw = rse^{i(\theta+\phi)},$$

so that as we observed earlier, the moduli r, s of z, w multiply as nonnegative real numbers, and the arguments θ, ϕ of z, w add as real numbers modulo 2π.

Finally, we observe that if $f:\mathbb{C}\to\mathbb{C}$ by $f(z)=e^z$, then

(2.3)
$$f'(z)=\lim_{h\to 0}\frac{e^{z+h}-e^z}{h}=\lim_{h\to 0}e^z\frac{e^h-1}{h}=e^z\lim_{h\to 0}\left(1+\frac{h}{2!}+\frac{h^2}{3!}+...\right)=e^z=f(z).$$

Thus we see that $e^z=e^xe^{iy}$ is a nonvanishing holomorphic function on $\mathbb{C}$. Further properties of the exponential function are

$$\begin{aligned}
\overline{e^{i\theta}} &= \overline{\cos\theta+i\sin\theta}=\cos\theta-i\sin\theta=\cos(-\theta)+i\sin(-\theta)=e^{-i\theta},\\
\left|e^{i\theta}\right|^2 &= e^{i\theta}\overline{e^{i\theta}}=e^{i\theta}e^{-i\theta}=e^{i\theta-i\theta}=e^0=1,\\
\cos^2\theta+\sin^2\theta &= |\cos\theta+i\sin\theta|^2=\left|e^{i\theta}\right|^2=1,\\
\frac{d}{d\theta}\cos\theta+i\frac{d}{d\theta}\sin\theta &= \frac{d}{d\theta}e^{i\theta}=ie^{i\theta}=i(\cos\theta+i\sin\theta)=-\sin\theta+i\cos\theta,\\
\frac{d}{d\theta}\cos\theta &= -\sin\theta \text{ and } \frac{d}{d\theta}\sin\theta=\cos\theta.
\end{aligned}$$

It now follows easily that if $\frac{\pi}{2}$ is defined to be the smallest positive root of $\cos\theta$ (which exists because of the intermediate value theorem), then e^z is $2\pi i$ periodic, maps the imaginary axis *onto* the unit circle, and has range $\mathbb{C}\setminus\{0\}$.

2.1. Power series. We were able to show in (2.3) that the exponential function had a complex derivative by exploiting the exponent formula (2.1). In fact *all power series* $\sum_{n=0}^{\infty}a_nz^n$ are holomorphic in their open disks of convergence. Note that by the root test, the *derived* series $\sum_{n=1}^{\infty}na_nz^{n-1}$ also has radius of convergence R.

THEOREM 1. *Suppose the power series* $\sum_{n=0}^{\infty}a_nz^n$ *has radius of convergence* $R>0$*, and define*

$$f(z)=\sum_{n=0}^{\infty}a_nz^n,\qquad z\in B(0,R).$$

Then f is holomorphic in $B(0,R)$ and

$$f'(z)=\sum_{n=1}^{\infty}na_nz^{n-1},\qquad z\in B(0,R). \tag{2.4}$$

Proof: To see (2.4), define $g(z)=\sum_{n=1}^{\infty}na_nz^{n-1}$ and use the identity

$$a^m-b^m=(a-b)\sum_{j=1}^{m}a^{m-j}b^{j-1}$$

twice to obtain:

$$\begin{aligned}
\frac{f(w)-f(z)}{w-z}-g(z) &= \sum_{n=1}^{\infty} a_n \left\{ \frac{w^n-z^n}{w-z} - nz^{n-1} \right\} \\
&= \sum_{n=2}^{\infty} a_n \left\{ \sum_{k=1}^{n} w^{n-k} z^{k-1} - nz^{n-1} \right\} \\
&= \sum_{n=2}^{\infty} a_n \left\{ \sum_{k=1}^{n} z^{k-1} \left(w^{n-k} - z^{n-k} \right) \right\} \\
&= \sum_{n=2}^{\infty} a_n \left\{ \sum_{k=1}^{n} z^{k-1} (w-z) \sum_{\ell=1}^{n-k} w^{n-k-\ell} z^{\ell-1} \right\} \\
&= (w-z) \sum_{n=2}^{\infty} a_n \left\{ \sum_{k=1}^{n-1} \sum_{\ell=1}^{n-k} w^{n-k-\ell} z^{k+\ell-2} \right\} \\
&= (w-z) \sum_{n=2}^{\infty} a_n \left\{ \sum_{j=2}^{n} (j-1) w^{n-j} z^{j-2} \right\}.
\end{aligned}$$

Thus if $|z|, |w| < \rho < R$, then

$$\left| \frac{f(w)-f(z)}{w-z} - g(z) \right| \leq |w-z| \sum_{n=2}^{\infty} |a_n| \frac{n(n-1)}{2} \rho^{n-2} \to 0$$

as $w \to z$ since $\sum_{n=2}^{\infty} |a_n| \frac{n(n-1)}{2} \rho^{n-2} < \infty$. This latter series is finite since the twice derived series $\sum_{n=2}^{\infty} a_n n(n-1) z^{n-2}$ has radius of convergence R and is absolutely convergent in its disk of convergence.

Of course the above applies equally well to a power series $\sum_{n=0}^{\infty} a_n (z-a)^n$ in the disk $B(a,R)$ where R is the radius of convergence. We will see later that every holomorphic function in a disk $B(a,R)$ has a power series representation in that disk.

3. Complex partial derivatives

First we recall a special case of Stokes' theorem in the plane. Suppose that Ω is a convex open subset of the plane whose boundary $\partial\Omega$ is an oriented piecewise continuously differentiable simple closed curve in the plane. Then if

$$F = A(x,y)\,dx + B(x,y)\,dy$$

is a continuously differentiable one-form on $\overline{\Omega}$, we have

$$\int_{\partial\Omega} F = \int_{\Omega} dF \tag{3.1}$$

where

$$\begin{aligned} dF &= dA(x,y)\wedge dx + dB(x,y)\wedge dy \\ &= \left\{\frac{\partial A}{\partial x}(x,y)\,dx + \frac{\partial A}{\partial y}(x,y)\,dy\right\}\wedge dx \\ &\quad + \left\{\frac{\partial B}{\partial x}(x,y)\,dx + \frac{\partial B}{\partial y}(x,y)\,dy\right\}\wedge dy \\ &= \left\{\frac{\partial B}{\partial x}(x,y) - \frac{\partial A}{\partial y}(x,y)\,dy\right\} dxdy \end{aligned}$$

is the exterior derivative of F. This version (3.1) of Stokes' theorem is usually called Green's theorem.

However, when dealing with complex differentiable functions, it is more convenient to use variables intimately connected with complex derivatives, namely

$$\begin{aligned} z &= x+iy \text{ and } \overline{z} = x - iy, \\ x &= \frac{1}{2}(z+\overline{z}) \text{ and } y = \frac{1}{2i}(z-\overline{z}). \end{aligned}$$

We also use the associated partial derivatives *defined* by

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} + \frac{1}{i}\frac{\partial}{\partial y}\right) \text{ and } \frac{\partial}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} - \frac{1}{i}\frac{\partial}{\partial y}\right).$$

Note that these definitions are suggested by a formal application of the chain rule:

$$\begin{aligned} \frac{\partial}{\partial z} &= \frac{\partial x}{\partial z}\frac{\partial}{\partial x} + \frac{\partial y}{\partial z}\frac{\partial}{\partial y} = \frac{1}{2}\frac{\partial}{\partial x} + \frac{1}{2i}\frac{\partial}{\partial y}, \\ \frac{\partial}{\partial \overline{z}} &= \frac{\partial x}{\partial \overline{z}}\frac{\partial}{\partial x} + \frac{\partial y}{\partial \overline{z}}\frac{\partial}{\partial y} = \frac{1}{2}\frac{\partial}{\partial x} - \frac{1}{2i}\frac{\partial}{\partial y}. \end{aligned}$$

Now we compute that

$$dz \wedge d\overline{z} = d(x+iy)\wedge d(x-iy) = -2idxdy,$$

and so Stokes' theorem (3.1) for the one-form $A(z)\,dz$ (the general one-form is given in this notation by $A(z)\,dz + B(z)\,d\overline{z}$) becomes

$$\begin{aligned} \int_{\partial\Omega} A(z)\,dz &= \int_\Omega d\{A(z)\,dz\} = \int_\Omega \left\{\frac{\partial A}{\partial z}(z)\,dz + \frac{\partial A}{\partial \overline{z}}(z)\,d\overline{z}\right\}\wedge dz \\ &= \int_\Omega \frac{\partial A}{\partial \overline{z}}(z)\,d\overline{z}\wedge dz = 2i\int_\Omega \frac{\partial A}{\partial \overline{z}}(z)\,dxdy. \end{aligned} \tag{3.2}$$

If $f = u + iv$, the equation

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

is equivalent to the Cauchy-Riemann equations (1.4). As a consequence we obtain from (3.2) that a special case of Cauchy's theorem holds for complex differentiable functions with a *continuous* derivative (only later will we see that complex differentiable functions f are automatically infinitely differentiable, so that Cauchy's theorem holds without the assumption of continuity on f').

THEOREM 2. *Let Ω be a convex open subset of the plane whose boundary $\partial\Omega$ is an oriented piecewise continuously differentiable simple closed curve in the plane.*

Suppose that f is holomorphic in Ω, and that f and f' extend continuously to $\overline{\Omega}$. Then

$$\int_{\partial\Omega} f(z)\,dz = 0.$$

Proof: Green's theorem yields

$$\int_{\partial\Omega} f(z)\,dz = 2i\int_{\Omega} \frac{\partial f}{\partial \overline{z}}(z)\,dxdy = 2i\int_{\Omega} 0dxdy = 0,$$

since $\frac{\partial f}{\partial \overline{z}}(z) = 0$ when f is complex differentiable at z.

4. The Cauchy-Goursat Theorem

In this section we obtain a variant of Theorem 2 *without* the assumption that f' is continuous in Ω.

THEOREM 3. *Let Ω be a convex open subset of the plane, and let γ be an oriented piecewise continuously differentiable simple closed curve in Ω. If f is holomorphic in Ω, then*

$$\int_{\gamma} f(z)\,dz = 0. \tag{4.1}$$

Proof: We proceed in a sequence of three steps as follows.

Step 1: If γ is a triangle, then (4.1) holds.

This is the main step in the proof. Set $\gamma_1^0 \equiv \gamma$ and let

$$A_1^0 = \int_{\gamma_1^0} f(z)\,dz.$$

We divide the triangle γ_1^0 into four congruent triangles $\gamma_1^1, \gamma_2^1, \gamma_3^1, \gamma_4^1$ by joining the midpoints of the sides of γ_1^0. Clearly

$$A_1^0 = \sum_{k=1}^{4} A_k^1$$

where

$$A_k^1 \equiv \int_{\gamma_k^1} f(z)\,dz, \qquad 1 \leq k \leq 4.$$

Now since $\left|A_1^0\right| \leq \sum_{k=1}^4 \left|A_k^1\right|$, we may assume by relabelling that $\left|A_1^1\right| \geq \frac{1}{4}\left|A_1^0\right|$. Now divide the triangle γ_1^1 into four congruent triangles $\gamma_1^2, \gamma_2^2, \gamma_3^2, \gamma_4^2$ by joining the midpoints of the sides of γ_1^1. With

$$A_k^2 \equiv \int_{\gamma_k^2} f(z)\,dz, \qquad 1 \leq k \leq 4,$$

we may again assume by relabelling that $\left|A_1^2\right| \geq \frac{1}{4}\left|A_1^1\right|$.

Continuing in this manner we obtain a sequence of triangles $\gamma_1^0 \supset \gamma_1^1 \supset \gamma_1^2 \supset \ldots\gamma_1^n \supset \ldots$ whose closed convex hulls H_1^n contain a unique point $z_0 \in \Omega$. We have

$$\begin{aligned}
diam\left(\gamma_1^n\right) &= 2^{-n} diam\left(\gamma_1^0\right) \\
length\left(\gamma_1^n\right) &= 2^{-n} length\left(\gamma_1^0\right), \\
area\left(H_1^n\right) &= 4^{-n} area\left(H_1^0\right), \\
\left|A_1^n\right| &\geq 4^{-n}\left|A_1^0\right|.
\end{aligned}$$

Now from Theorem 2 applied to the function $z \to f(z_0) + f'(z_0)(z - z_0)$ (since this function has an antiderivative, namely $f(z_0)z + f'(z_0)\frac{1}{2}(z - z_0)^2$, we could instead use the fundamental theorem of line integrals), we obtain

$$\int_{\gamma_1^n} \{f(z_0) + f'(z_0)(z - z_0)\} \, dz = 0, \qquad n \geq 1.$$

Thus since f is complex differentiable at z_0 we conclude with $h_n = diam(\gamma_1^n) = 2^{-n} diam(\gamma_1^0)$ that

$$\begin{aligned} |A_1^n| &= \left| \int_{\gamma_k^n} f(z) \, dz \right| = \left| \int_{\gamma_1^n} \{f(z) - f(z_0) - f'(z_0)(z - z_0)\} \, dz \right| \\ &\leq o(diam(\gamma_1^n)) \, length(\gamma_1^n) \\ &\leq 2^{-n} diam(\gamma_1^0) \frac{o(h_n)}{h_n} 2^{-n} length(\gamma_1^0) \\ &\leq |A_1^n| \frac{length(\gamma_1^0)}{|A_1^0|} \frac{o(h_n)}{h_n}, \end{aligned}$$

which implies that $|A_1^n| = 0$ as soon as $\frac{o(h_n)}{h_n} < \frac{|A_1^0|}{length(\gamma_1^0)}$. Thus $|A_1^0| \leq 4^n |A_1^n| = 0$ and (4.1) holds when γ is a triangle.

Remark 1. *In the special case where γ is a triangle, we need only assume that f is continuous in Ω, and holomorphic in $\Omega \setminus \{p\}$ for some exceptional point $p \in \Omega$. Indeed, Step 1 already proves the case when p is outside γ. If p is a vertex of $\gamma = \triangle(a, b, c)$, say $a = p$, we consider points x and y arbitrarily close to p on the sides $[a, b]$ and $[a, c]$ respectively of γ. Then (4.1) holds for γ replaced by either $\triangle(x, b, c)$ or $\triangle(x, y, c)$, and since the perimeter of $\triangle(x, y, a)$ is as small as we wish, we obtain (4.1). The cases where p is on an edge or in the interior of γ are now easy to establish by decomposing γ into appropriate subtriangles.*

Step 2: There is a holomorphic antiderivative F of f in Ω, i.e. $F'(z) = f(z)$ for $z \in \Omega$.

Fix $z_0 \in \Omega$ and define

$$F(z) = \int_{[z_0, z]} f(w) \, dw, \qquad z \in \Omega,$$

where $[z_0, z]$ denotes the curve γ whose image is the straight line segment joining z_0 to z, parameterized for example by $\gamma(t) = z_0 + t(z - z_0)$, $0 \leq t \leq 1$. By Step 1 we have

$$\int_{[z_0, z]} f(w) \, dw + \int_{[z, z+h]} f(w) \, dw + \int_{[z+h, z_0]} f(w) \, dw = 0,$$

and so

$$\begin{aligned} \frac{F(z+h) - F(z)}{h} &= \frac{1}{h} \left\{ \int_{[z_0, z+h]} f(w) \, dw - \int_{[z_0, z]} f(w) \, dw \right\} \\ &= \frac{1}{h} \int_{[z, z+h]} f(w) \, dw \\ &= \frac{1}{h} \int_0^1 f(z + th) \, h dt \to f(z) \end{aligned}$$

as $h \to 0$.

Step 3: $\int_\gamma f(z)\,dz = 0$.

Parameterize γ as $\gamma(t)$, $a \le t \le b$ (without loss of generality we may assume that γ is continuously differentiable in (a,b)). Let F be an antiderivative of f as in Step 2. The chain rule shows that $\frac{d}{dt}F(\gamma(t)) = F'(\gamma(t))\gamma'(t)$ (a formal and nonrigorous proof of this is that

$$\frac{F(\gamma(t+h)) - F(\gamma(t))}{h} = \frac{F(\gamma(t+h)) - F(\gamma(t))}{\gamma(t+h)-\gamma(t)}\frac{\gamma(t+h)-\gamma(t)}{h}$$

tends to $F'(\gamma(t))\gamma'(t)$ as $h \to 0$). We now compute that

$$\begin{aligned}\int_\gamma f(z)\,dz &= \int_\gamma F'(z)\,dz = \int_a^b F'(\gamma(t))\gamma'(t)\,dt \\ &= \int_a^b \frac{d}{dt}F(\gamma(t))\,dt = F(\gamma(b)) - F(\gamma(a)) = 0\end{aligned}$$

since $\gamma(b) = \gamma(a)$ when γ is a closed curve.

Corollary 1. *Theorem 3 holds if we merely assume that f is continuous in Ω, and holomorphic in $\Omega \setminus \{p\}$ for some exceptional point $p \in \Omega$.*

5. Cauchy's representation formula

A remarkable property of holomorphic functions f is that they are uniquely determined inside a "nice" domain by their boundary values, and moreover there is a simple formula for recovering the interior values $f(z)$ from the boundary values $f(w)$, namely

$$f(z) = \frac{1}{2\pi i}\int_{\partial\Omega}\frac{f(w)}{w-z}dw, \qquad z \in \Omega. \tag{5.1}$$

In order to state and prove a precise version of Cauchy's formula (5.1), we need the notion of the *index* of a curve γ in the complex plane relative to a point $z \in \mathbb{C}$.

Definition 5. *If γ is a closed piecewise continuously differentiable curve in the plane, and if $z \in \mathbb{C}\setminus\gamma^*$ (we denote by γ^* the image of γ in the plane), we define the* index *of γ about z to be*

$$Ind_\gamma(z) \equiv \frac{1}{2\pi i}\int_\gamma \frac{1}{w-z}dw.$$

Proposition 4. *Suppose that γ is a closed piecewise continuously differentiable curve in the plane. Then the function $Ind_\gamma(z)$ is integer-valued, constant in each component of $\mathbb{C}\setminus\gamma^*$, and vanishes in the unbounded component of $\mathbb{C}\setminus\gamma^*$.*

Proof: Let $\gamma(t)$, $a \le t \le b$, be a parameterization of γ and fix $z \in \mathbb{C}\setminus\gamma^*$. Heuristically, we expect the antiderivative of $\frac{\gamma'(t)}{\gamma(t)-z}$ to be $\log(\gamma(t)-z)$ where log is an inverse function to exp. Then we would get

$$Ind_\gamma(z) = \int_a^b \frac{\gamma'(t)}{\gamma(t)-z}dt = \log(\gamma(t)-z)\,|_a^b = \log\frac{\gamma(b)-z}{\gamma(a)-z} = \log 1.$$

Since exp is $2\pi i$ periodic and $e^0 = 1$, $\log 1 \in \{2\pi i n\}_{n\in\mathbb{Z}}$, and this would prove the proposition.

Unfortunately, we have not yet defined a logarithm function, so we proceed via a route that achieves the same goal without actually using a logarithm. We accomplish this with the function

$$\varphi(s)=\exp\left\{\int_a^s \frac{1}{\gamma(t)-z}\gamma'(t)\,dt\right\},\qquad a\le s\le b.$$

Since $Ind_\gamma(z)=\int_a^b \frac{1}{\gamma(t)-z}\gamma'(t)\,dt$, we see that $Ind_\gamma(z)$ is an integer *if and only if* $\varphi(b)=\exp\{Ind_\gamma(z)\}$ equals 1. To show that $\varphi(b)=1$ it suffices to show that

$$\frac{d}{ds}\frac{\varphi(s)}{\gamma(s)-z}=0,\qquad a\le s\le b, \tag{5.2}$$

since then

$$\frac{\varphi(b)}{\gamma(b)-z}=\frac{\varphi(a)}{\gamma(a)-z}=\frac{1}{\gamma(b)-z},$$

as $\varphi(a)=1$ and $\gamma(a)=\gamma(b)$ (since γ is a closed path). Actually, we can only show (5.2) at points s where $\gamma(s)$ is continuously differentiable, but this is easily seen to be enough. For this we compute

$$\varphi'(s)=\varphi(s)\frac{\gamma'(s)}{\gamma(s)-z},$$

so that

$$\frac{d}{ds}\frac{\varphi(s)}{\gamma(s)-z}=\frac{(\gamma(s)-z)\varphi'(s)-\varphi(s)\gamma'(s)}{(\gamma(s)-z)^2}=0.$$

Since $Ind_\gamma(z)$ is continuous in $\mathbb{C}\setminus\gamma^*$, we conclude that $Ind_\gamma(z)$ is a constant integer in each component of $\mathbb{C}\setminus\gamma^*$, and since $\lim_{z\to\infty} Ind_\gamma(z)=0$, we conclude that $Ind_\gamma(z)$ vanishes in the unbounded component of $\mathbb{C}\setminus\gamma^*$.

Corollary 2. *If γ is the positively oriented circle centered at a with radius r, then*

$$Ind_\gamma(z)=\begin{cases}1 & \text{if}\quad |z-a|<r\\ 0 & \text{if}\quad |z-a|>r\end{cases}.$$

Proof: If we parameterize γ by $\gamma(t)=e^{it}$, $0\le t\le 2\pi$, then

$$Ind_\gamma(0)=\frac{1}{2\pi i}\int_\gamma\frac{1}{w-0}dw=\frac{1}{2\pi i}\int_0^{2\pi}\frac{1}{e^{it}-0}ie^{it}dt=\frac{1}{2\pi i}\int_0^{2\pi} idt=1.$$

We can now give Cauchy's representation formula. For convenience we define a *path* to be a piecewise continuously differentiable curve in the plane.

Theorem 4. *Suppose that γ is a closed path in a convex set Ω. If f is holomorphic in Ω and $z\in\Omega\setminus\gamma^*$, then*

$$f(z)\,Ind_\gamma(z)=\frac{1}{2\pi i}\int_\gamma\frac{f(w)}{w-z}dw.$$

Proof: Fix $z\in\Omega\setminus\gamma^*$ and for $w\in\Omega$ define

$$g(w)=\begin{cases}\frac{f(w)-f(z)}{w-z} & \text{if}\quad w\ne z\\ f'(z) & \text{if}\quad w=z\end{cases}.$$

Then g is continuous in Ω, and holomorphic in $\Omega \setminus \{z\}$, and so Corollary 1 shows that

$$\frac{1}{2\pi i}\int_\gamma \frac{f(w)}{w-z}dw - f(z)\,Ind_\gamma(z) = \frac{1}{2\pi i}\int_\gamma \frac{f(w)-f(z)}{w-z}dw = \frac{1}{2\pi i}\int_\gamma g(w)\,dw = 0.$$

THEOREM 5. *Suppose that f is holomorphic in a disk $B(a,R)$. Then there is a power series $\sum_{n=0}^\infty a_n(z-a)^n$ having radius of convergence R such that*

$$f(z) = \sum_{n=0}^\infty a_n(z-a)^n, \qquad z \in B(a,R).$$

Moreover, f is infinitely complex differentiable and

$$f^{(m)}(z) = \sum_{n=m}^\infty a_n \frac{n!}{(n-m)!}(z-a)^{n-m}, \qquad z \in B(a,R).$$

In particular,

$$f^{(m)}(a) = m!a_m, \qquad m = 0,1,2,...$$

and we have Taylor's formula:

$$f(z) = \sum_{n=0}^\infty \frac{f^{(n)}(a)}{n!}(z-a)^n, \qquad z \in B(a,R).$$

Proof: We assume without loss of generality that $a = 0$. Then if γ is the positively oriented circle of radius $r < R$ about the origin, we have from Theorem 4 that for $z \in B(0,r)$:

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\int_\gamma \frac{f(w)}{w-z}dw = \frac{1}{2\pi i}\int_0^{2\pi} \frac{f\left(re^{i\theta}\right)}{re^{i\theta}-z} ire^{i\theta}d\theta \\ &= \frac{1}{2\pi}\int_0^{2\pi} \frac{f\left(re^{i\theta}\right)}{1-\frac{z}{re^{i\theta}}}d\theta = \frac{1}{2\pi}\int_0^{2\pi}\sum_{n=0}^\infty \left(\frac{z}{re^{i\theta}}\right)^n f\left(re^{i\theta}\right)d\theta \\ &= \sum_{n=0}^\infty \left\{\frac{1}{2\pi}\int_0^{2\pi}\frac{f\left(re^{i\theta}\right)}{r^n}e^{-in\theta}d\theta\right\} z^n = \sum_{n=0}^\infty a_n z^n. \end{aligned}$$

The above interchange of summation and integration follows from the uniform convergence of the geometric series in the unit disk. It follows that the power series $\sum_{n=0}^\infty a_n z^n$ has radius of convergence at least r. Moreover, by iterating (2.4) we see that f is infinitely differentiable in $B(0,r)$ and that $a_n = \frac{f^{(n)}(a)}{n!}$ for $n \geq 0$. Since r was any positive number less than R, the proof is complete.

DEFINITION 6. *Let Ω be an open set in the plane. A function $f : \Omega \to \mathbb{C}$ is* analytic *in Ω if f has a power series representation with radius of convergence R in each open disk $B(a,R)$ contained in Ω.*

Combining Theorems 1 and 5, we see that a function $f : \Omega \to \mathbb{C}$ is *holomorphic* in Ω if and only if it is *analytic* in Ω. Example 1 shows that this fails miserably for derivatives in the real field.

CHAPTER 3

Properties of holomorphic functions

In this chapter we develop some of the surprising properties of holomorphic functions, beginning with the dichotomy of zero sets, and the dichotomy of isolated singularities, of analytic functions. Then we investigate the maximum modulus theorem and some of its consequences, leading to the theory of normal families and the open mapping theorem. These tools will be used in the next chapter to prove the remarkable Riemann Mapping Theorem that characterizes the biholomorphic images of the open unit disk as comprising *all* open simply connected proper subsets of the complex plane.

1. Zeroes of analytic functions

Suppose a holomorphic function f is defined in a disk $B(a,R)$ and vanishes at the center a. The fact that f has a power series expansion $f(z)=\sum_{n=0}^{\infty}a_n(z-a)^n$ with radius of convergence at least R results in a dichotomy of just *two* possibilites:

(1) **either** all of the coefficients a_n are zero and f vanishes in the entire ball $B(a,R)$,

(2) **or** there is a first coefficient a_N that is non-zero and then

$$f(z)=(z-a)^N g(z) \tag{1.1}$$

where g is holomorphic in $B(a,R)$ and *nonvanishing* at a: indeed

$$\begin{aligned} f(z) &= \sum_{n=N}^{\infty} a_n(z-a)^n \\ &= (z-a)^N\left\{a_N+a_{N+1}(z-a)+a_{N+2}(z-a)^2+\ldots\right\} \\ &\equiv (z-a)^N g(z). \end{aligned}$$

The positive integer N is uniquely determined in the second possibility and is called the *order* of the zero of f at the point a. By convention we say that f has a zero of order 0 at a if $f(a)\neq 0$. In a connected open set, this phenomenon takes the following form.

THEOREM 6. *Suppose Ω is open and connected and $f\in H(\Omega)$. Then the zero set $Z=\{z\in\Omega : f(z)=0\}$ of f in Ω is either all of Ω or is a discrete subset of Ω (this means Z has no limit point in Ω). In the latter case there is associated to each point $a\in Z$ a unique positive integer N such that $f(z)=(z-a)^N g(z)$ where $g\in H(\Omega)$ and $g(a)\neq 0$.*

Proof: The set Z is closed in Ω since f is continuous. We claim that the interior $\overset{\circ}{Z}$ of Z is also a closed subset of Ω. Indeed, if $a\in\Omega\setminus\overset{\circ}{Z}$ is a limit point of $\overset{\circ}{Z}$, then $a\in Z$ and the first possibility of the dichotomy above fails, and so the second possibility holds. Then we conclude from the continuity of g in (1.1) that a is isolated in Z, contradicting the assumption that a is a limit point of $\overset{\circ}{Z}$. Thus $\overset{\circ}{Z}$ is both open and closed in Ω. Since Ω is connected, we conclude that either $\overset{\circ}{Z} = \Omega$ (in which case f is identically zero in Ω) or $\overset{\circ}{Z} = \phi$. In the latter case every point $a \in Z$ is an isolated point since only the second possibililty can hold

Corollary 3. *(The Coincidence Principle) Suppose that $f, g \in H(\Omega)$ where Ω is open and connected. If $f(z) = g(z)$ in some set E of points having a limit point in Ω, then $f = g$ in all of Ω.*

Proof: Since E is a subset of the zero set of $f - g$, the previous theorem shows that $E = \Omega$.

Example 2. *Define the holomorphic functions* sin *and* cos *in the complex plane by the series for* $\cos\theta$ *and* $\sin\theta$ *in (2.2), but replacing the real variable θ by the complex variable z, i.e.*

$$\begin{aligned} \cos z &= 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - ... \\ \sin z &= z - \frac{z^3}{3!} + \frac{z^5}{5!} - ... \end{aligned}$$

Then

$$\sin(z+w) = \sin z \cos w + \cos z \sin w, \qquad z, w \in \mathbb{C}. \tag{1.2}$$

To prove this we take as known that

$$\sin(x+y) = \sin x \cos y + \cos x \sin y, \qquad x, y \in \mathbb{R}, \tag{1.3}$$

By the coincidence principle we obtain that

$$\sin(z+y) = \sin z \cos y + \cos z \sin y, \qquad z \in \mathbb{C}, y \in \mathbb{R}, \tag{1.4}$$

since for fixed $y \in \mathbb{R}$, both sides of (1.4) are holomorphic functions of $z \in \mathbb{C}$ that coincide on the real line by (1.3). Now fix $z \in \mathbb{C}$ in (1.2), and note that both sides of (1.2) are holomorphic functions of $w \in \mathbb{C}$ that coincide on the real line by (1.4). Another application of the coincidence principle now proves (1.2).

2. Isolated singularities of analytic functions

Suppose a holomorphic function f is defined in a punctured disk

$$B'(a, R) = B(a, R) \setminus \{a\},$$

and has a singularity at the center a in the sense that there is no analytic function F in the entire ball $B(a, R)$ whose restriction to $B'(a, R)$ is f. In the case that such an F *does* exist, we say that f has a *removable singularity* at a. For an actual (nonremovable) singularity there is a dichotomy of just *two* possibilites:

(1) **either** $\lim_{z \to a} |f(z)| = \infty$ and there is a positive integer N such that

$$h(z) = (z-a)^N f(z), \qquad z \in B'(a, R),$$

has a removable singularity at a and $h(a) \neq 0$,

(2) **or** the image $f(B'(a, r))$ under f of each punctured disk $B'(a, r)$, $0 < r < R$, is dense in the complex plane $\mathbb{C}$. In this case $\lim_{z \to a} f(z)$ fails to exist in the most spectacular way possible, namely the cluster set of f at a is $\mathbb{C}$.

In possibility (1) we say that f has a *pole* of order N at a, and in possibility (2) we say that f has an *essential singularity* at a. In order to prove the dichotomy, we begin with Riemann's theorem on removable singularities, which can be restated as saying that if f has an actual isolated singularity at a, then f must be *unbounded* in every neighbourhood of a.

THEOREM 7. *Suppose f is holomorphic and bounded in a punctured disk $B'(a,R)$. Then f has a removable singularity at a.*

Proof: Let $h(z) = \begin{cases} (z-a)^2 f(z) & \text{if} \quad z \in B'(a,R) \\ 0 & \text{if} \quad z = a \end{cases}$. Then $h'(a) = 0$ and so $h \in H(B(a,R))$. By Theorem 5, h has a power series expansion in $B(a,R)$:

$$h(z) = \sum_{n=0}^{\infty} \frac{h^{(n)}(a)}{n!} (z-a)^n, \qquad z \in B(a,R).$$

Since $h(a) = h'(a) = 0$, we have

$$(z-a)^2 f(z) = h(z) = \sum_{n=2}^{\infty} \frac{h^{(n)}(a)}{n!} (z-a)^n = (z-a)^2 \sum_{n=0}^{\infty} \frac{h^{(n+2)}(a)}{(n+2)!} (z-a)^n,$$

for $z \in B'(a,R)$, and hence f is the restriction to $B'(a,R)$ of the holomorphic function $F(z) = \sum_{n=0}^{\infty} \frac{h^{(n+2)}(a)}{(n+2)!} (z-a)^n$, $z \in B(a,R)$.

Now we can prove the dichotomy of isolated singular points. Indeed, suppose that f has a (nonremovable) singularity at a and that possibility (2) fails. Then there is $0 < r < R$ and a disk $B(w,\varepsilon)$ in the plane such that

$$f(B'(a,r)) \cap B(w,\varepsilon) = \phi.$$

It follows that

$$g(z) = \frac{1}{f(z) - w}, \qquad z \in B'(a,r),$$

is bounded ($|g(z)| \leq \frac{1}{\varepsilon}$) and nonvanishing in the punctured disk $B'(a,r)$. By Riemann's theorem there is $G \in H(B(a,r))$ which restricts to g in $B'(a,r)$.

We must have $G(a) = 0$ since otherwise solving $G(z) = \frac{1}{f(z)-w}$ for f shows that

$$f(z) = w + \frac{1}{G(z)}, \qquad z \in B'(a,r),$$

has a removable singularity at a, a contradiction. It now follows from Theorem 6 that there is a unique positive integer N such that

$$G(z) = (z-a)^N H(z), \qquad z \in B(a,r),$$

with H holomorphic and nonvanishing in all of $B(a,r)$. Thus we have

$$f(z) = w + \frac{1}{(z-a)^N H(z)}, \qquad z \in B'(a,r).$$

It follows that

$$\lim_{z \to a} |f(z)| = \infty,$$

and that

$$h(z) = \begin{cases} (z-a)^N f(z) & \text{if} \quad z \in B'(a,R) \\ w(z-a)^N + \frac{1}{H(z)} & \text{if} \quad z \in B(a,r) \end{cases},$$

defines a holomorphic function in $B(a,R)$ with $h(a) = \frac{1}{H(a)} \neq 0$. This proves that possibility (1) holds.

We note that when possibility (1) holds, we can write

$$w(z-a)^N + \frac{1}{H(z)} = \sum_{n=0}^{\infty} a_n (z-a)^n$$

as a power series and obtain the representation,

$$\begin{aligned} f(z) &= (z-a)^{-N} \sum_{n=0}^{\infty} a_n (z-a)^n = \sum_{n=0}^{\infty} a_n (z-a)^{n-N} \\ &= \frac{a_0}{(z-a)^N} + \ldots + \frac{a_{N-1}}{z-a} + a_N + a_{N+1}(z-a) + \ldots \end{aligned}$$

which is usually written

$$f(z) = \sum_{n=-N}^{\infty} b_n (z-a)^n = \frac{b_{-N}}{(z-a)^N} + \ldots + \frac{b_{-1}}{z-a} + \sum_{n=0}^{\infty} b_n (z-a)^n,$$

for $z \in B'(a,R)$. This motivates calling such a singularity a *pole* of order N at a. The polynomial in $\frac{1}{z-a}$,

$$P_N(z) = \frac{b_{-N}}{(z-a)^N} + \ldots + \frac{b_{-1}}{z-a},$$

is called the *principal part* of f at the pole a.

3. Cauchy's estimates, maximum principle, and uniform convergence

THEOREM 8. *Suppose Ω is open and convex, and that γ is a closed path in Ω. If $f \in H(\Omega)$, then*

$$f^{(n)}(z)\, Ind_\gamma(z) = \frac{n!}{2\pi i} \int_\gamma \frac{f(w)}{(w-z)^{n+1}} dw, \qquad z \in \Omega \setminus \gamma^*, n \geq 0. \tag{3.1}$$

Proof: For $z, z+h$ in the same component of $\Omega \setminus \gamma^*$, Cauchy's formula shows that

$$\begin{aligned} \frac{f(z+h)-f(z)}{h} Ind_\gamma(z) &= \frac{1}{2\pi i} \int_\gamma \frac{1}{h} \left\{ \frac{1}{w-z-h} - \frac{1}{w-z} \right\} f(w)\, dw \\ &= \frac{1}{2\pi i} \int_\gamma \left\{ \frac{1}{(w-z-h)(w-z)} \right\} f(w)\, dw \\ &\to \frac{1}{2\pi i} \int_\gamma \frac{f(w)}{(w-z)^2} dw \end{aligned} \tag{3.2}$$

as $h \to 0$ since $\frac{1}{w-z-h} \to \frac{1}{w-z}$ uniformly on γ^*. Thus we have

$$f'(z)\, Ind_\gamma(z) = \frac{1}{2\pi i} \int_\gamma \frac{f(w)}{(w-z)^2} dw, \qquad z \in B(a,r),$$

and (3.1) now follows by repetition of the argument.

COROLLARY 4. *(Cauchy's estimates) If* $f \in H(B(a,R))$, *then for* $0 \le n < \infty$,

$$\begin{aligned}
\left|f^{(n)}(a)\right| &\le \frac{n!}{r^n}\left\{\frac{1}{2\pi}\int_0^{2\pi}\left|f\left(a+re^{i\theta}\right)\right|d\theta\right\}, \qquad 0<r<R, \qquad (3.3)\\
\left|f^{(n)}(a)\right| &\le \frac{n!}{R^n}\left\{\sup_{0<r<R}\frac{1}{2\pi}\int_0^{2\pi}\left|f\left(a+re^{i\theta}\right)\right|d\theta\right\}\\
&\le \frac{n!}{R^n}\sup_{z\in B(a,R)}|f(z)|.
\end{aligned}$$

Proof: For $0<r<R$ we have

$$\begin{aligned}
\left|f^{(n)}(a)\right| &= \left|\frac{n!}{2\pi i}\int_\gamma \frac{f(w)}{(w-a)^{n+1}}dw\right| = \left|\frac{n!}{2\pi i}\int_0^{2\pi}\frac{f\left(a+re^{i\theta}\right)}{\left(re^{i\theta}\right)^{n+1}}ire^{i\theta}d\theta\right|\\
&\le \frac{n!}{2\pi}\int_0^{2\pi}\frac{\left|f\left(a+re^{i\theta}\right)\right|}{r^n}d\theta.
\end{aligned}$$

The second inequality in (3.3) follows easily.

THEOREM 9. *(Maximum principle) Suppose that* Ω *is open and connected. If* $f \in H(\Omega)$ *then* f cannot *have a strict local maximum in* Ω. *If* f *has a local maximum, it must be constant in* Ω.

Proof: Suppose, in order to derive a contraction, that f has a strict local maximum at $a \in \Omega$. Then there is $B(a,R)\subset\Omega$ such that

$$|f(a)| > |f(z)|, \qquad z\in B'(a,R).$$

Then for every $0<r<R$ we have from the first inequality in (3.3),

$$|f(a)| \le \frac{1}{2\pi}\int_0^{2\pi}\left|f\left(a+re^{i\theta}\right)\right|d\theta < \frac{1}{2\pi}\int_0^{2\pi}|f(a)|\,d\theta = |f(a)|,$$

the desired contradiction. Now suppose only that f has a local maximum at $a\in\Omega$. Then there is $B(a,R)\subset\Omega$ such that

$$|f(a)| \ge |f(z)|, \qquad z\in B'(a,R).$$

Then for every $0<r<R$ we have from the first inequality in (3.3),

$$|f(a)| \le \frac{1}{2\pi}\int_0^{2\pi}\left|f\left(a+re^{i\theta}\right)\right|d\theta \le \frac{1}{2\pi}\int_0^{2\pi}|f(a)|\,d\theta = |f(a)|,$$

and it follows from the continuity of f that $\left|f\left(a+re^{i\theta}\right)\right| = |f(a)|$ for all $a+re^{i\theta}\in B(a,R)$.

Thus we have $|f(a)|^2 = |f(z)|^2$ for $z\in B(a,R)$. If $f(a)=0$ we have that f is the constant 0 in $B(a,R)$, and otherwise we have

$$0 = \frac{\partial}{\partial z}|f(a)|^2 = \frac{\partial}{\partial z}\left\{f(z)\overline{f(z)}\right\} = f'(z)\overline{f(z)}, \qquad z\in B(a,R),$$

since $\frac{\partial}{\partial z}\overline{f(z)} = \overline{\frac{\partial}{\partial \bar z}f(z)} = 0$ by the Cauchy-Riemann equations. Thus $f'(z)=0$ in $B(a,R)$ and so f is again constant in $B(a,R)$ since

$$f(z)-f(a) = \int_{[a,z]} f'(w)\,dw = 0, \qquad z\in B(a,R).$$

Finally, the coincidence principle implies that f is constant in Ω.

REMARK 2. *Theorem 9 fails to hold with "maximum" replaced by "minimum" as evidenced by the function* $f(z) = z$ *in* $\Omega = B(0,1)$. *However, if we assume in addition that* f *is nonvanishing in* Ω, *then the theorem holds with "maximum" replaced by "minimum" throughout (simply apply Theorem 9 to* $\frac{1}{f(z)}$*).*

COROLLARY 5. *Suppose* Ω *is bounded, open and connected. If* $f \in H(\Omega) \cap C(\overline{\Omega})$, *then*

$$|f(z)| \leq \sup_{w \in \partial\Omega} |f(w)|, \qquad z \in \Omega,$$

with strict *inequality unless* f *is constant in* Ω. *If in addition,* f *is nonvanishing in* Ω, *then*

$$|f(z)| \geq \inf_{w \in \partial\Omega} |f(w)|, \qquad z \in \Omega,$$

with strict *inequality unless* f *is constant in* Ω.

Here is a famous application of the second inequality in (3.3).

THEOREM 10. *(Liouville's theorem) If* $f : \mathbb{C} \to \mathbb{C}$ *is holomorphic and bounded, then* f *is constant.*

Proof: From the second inequality in (3.3) we obtain

$$|f'(a)| \leq \frac{1}{R^n} \sup_{z \in B(a,R)} |f(z)| \to 0 \text{ as } R \to \infty,$$

or $f'(a) = 0$ for all $a \in \mathbb{C}$. It follows that f is constant.

PROBLEM 1. *Suppose more generally that* $f : \mathbb{C} \to \mathbb{C}$ *is holomorphic and has polynomial growth at infinity, i.e.* $|f(z)| \leq C\left(1 + |z|^N\right)$ *for some positive constant* C *and positive integer* N. *Show that* $f(z)$ *is a polynomial of degree at most* N.

Hint*: The second inequality in (3.3) yields*

$$\left|f^{(n)}(a)\right| \leq \frac{n!}{R^n} \sup_{z \in B(a,R)} |f(z)| \leq \frac{n!}{R^n} C\left(1 + (|a| + R)^N\right).$$

Show that this vanishes in the limit as $R \to \infty$ *whenever* $n > N$.

Another consequence of the second inequality in (3.3) concerns uniform convergence of a sequence of holomorphic functions.

We take for granted that if γ is a simple closed path in the plane, then $\mathbb{C} \setminus \gamma^*$ has a single bounded component and $Ind_\gamma(z) = 1$ for $z \in D$ if γ surrounds D in the positive direction (the same assertion for a simple closed *curve* is the difficult Jordan Curve Theorem).

THEOREM 11. *(Uniform convergence theorem) Suppose that* Ω *is an open convex set in the plane and that* γ *is a simple closed path in* Ω. *If* $\{f_n\}_{n=1}^\infty$ *is a sequence of holomorphic functions in* Ω *that converge uniformly on* γ^*, *then* $\{f_n\}_{n=1}^\infty$ *converges uniformly on compact subsets of the bounded component* D *of* $\mathbb{C} \setminus \gamma^*$ *to a holomorphic function in* D. *Moreover,* $\{f_n'\}_{n=1}^\infty$ *converges uniformly on compact subsets of* D.

Proof: If K is a compact subset of D, and $\delta = dist\left(K, D^c\right)$, then $\delta > 0$ and for $z \in K$ we have

$$\begin{aligned} \left|f_m(z) - f_n(z)\right| &= \left|\frac{1}{2\pi i}\int_\gamma \frac{f_m(w) - f_n(w)}{w - z}dw\right| \\ &\leq \frac{1}{2\pi}\int_a^b \frac{\left|f_m(w) - f_n(w)\right|}{\delta}\left|\gamma'(t)\right|dt \\ &\leq \frac{1}{2\pi\delta} length\left(\gamma\right)\sup_{w\in\gamma^*}\left|f_m(w) - f_n(w)\right|. \end{aligned}$$

Thus $\{f_n\}_{n=1}^\infty$ converges uniformly on K to a continuous function f in D. But since we have

$$f(z) = \lim_{n\to\infty} f_n(z) = \lim_{n\to\infty}\frac{1}{2\pi i}\int_\gamma \frac{f_n(w)}{w - z}dw = \frac{1}{2\pi i}\int_\gamma \frac{f(w)}{w - z}dw,$$

for $z \in D$, the calculation in (3.2) shows that f is holomorphic in D, and moreover,

$$f'(z) = \frac{1}{2\pi i}\int_\gamma \frac{f(w)}{(w - z)^2}dw.$$

For convenience we repeat the argument here using $Ind_\gamma(z) = Ind_\gamma(z + h) = 1$ for $z, z + h \in D$:

$$\begin{aligned} f'(z) &= \lim_{h\to 0}\frac{f(z + h) - f(z)}{h} \\ &= \lim_{h\to 0}\frac{1}{2\pi i}\int_\gamma \frac{1}{h}\left\{\frac{1}{w - z - h} - \frac{1}{w - z}\right\}f(w)\,dw \\ &= \frac{1}{2\pi i}\int_\gamma \left\{\lim_{h\to 0}\frac{1}{(w - z - h)(w - z)}\right\}f(w)\,dw \\ &= \frac{1}{2\pi i}\int_\gamma \frac{f(w)}{(w - z)^2}dw. \end{aligned}$$

Since

$$f_n'(z) = \frac{1}{2\pi i}\int_\gamma \frac{f_n(w)}{(w - z)^2}dw,$$

the argument above using $f_n \to f$ uniformly on γ^*, shows that f_n' converges uniformly to f' on compact subsets of D.

4. Normal families

Here we show that every sequence of holomorphic functions on an open set Ω, that is uniformly bounded on compact subsets of Ω, has a subsequence that converges uniformly on compact subsets of Ω to a holomorphic function. The proof will use the above theorem on uniform convergence together with the Arzela-Ascoli theorem. We begin with the statement and proof of the Arzela-Ascoli theorem, one of the most useful real-variable theorems in analysis.

Suppose K is a compact metric space with metric d_K. We define the metric space $\mathcal{C}(K)$ of continuous complex-valued functions on K by

$$\mathcal{C}(K) = \{f : K \to \mathbb{C} : f \text{ is continuous on } K\}$$

with metric

$$d_{\mathcal{C}(K)}(f,g) \equiv \sup_{x\in K} |f(x) - g(x)|, \qquad f,g \in \mathcal{C}(K).$$

Recall that $\mathcal{C}(K)$ is a *complete* metric space (every Cauchy sequence in $\mathcal{C}(K)$ converges) and that a sequence $\{f_n\}_{n=1}^{\infty} \subset \mathcal{C}(K)$ converges in the metric space $\mathcal{C}(K)$ if and only if it converges uniformly on K. Recall also that a continuous function f on a *compact* metric space K is actually *uniformly continuous* on K: for every $\varepsilon > 0$ there is $\delta > 0$ such that

$$|f(x) - f(y)| < \varepsilon, \qquad \text{whenever } d_K(x,y) < \delta.$$

A family $\mathcal{F} \subset \mathcal{C}(K)$ of continuous functions on K is called *equicontinuous* if for every $\varepsilon > 0$ there is $\delta > 0$ such that

$$|f(x) - f(y)| < \varepsilon \qquad \text{whenever } d_K(x,y) < \delta \text{ and } f \in \mathcal{F}. \tag{4.1}$$

Finally, we say that a family $\mathcal{F} \subset \mathcal{C}(K)$ is pointwise bounded on K if

$$\sup_{f\in\mathcal{F}} |f(x)| < \infty \qquad \text{for each } x \in K.$$

THEOREM 12. *(Arzela-Ascoli theorem) Suppose that K is a compact metric space and that $\{f_n\}_{n=1}^{\infty} \subset \mathcal{C}(K)$ is a sequence of continuous functions on K. If the sequence $\{f_n\}_{n=1}^{\infty}$ is both pointwise bounded and equicontinuous on K, then:*

1. *The sequence $\{f_n\}_{n=1}^{\infty}$ is uniformly bounded.*
2. *There is a subsequence $\{f_{n_k}\}_{k=1}^{\infty}$ that converges in $\mathcal{C}(K)$.*

Proof: (1) We first use equicontinuity of $\{f_n\}_{n=1}^{\infty}$ and the compactness of K to improve the pointwise boundedness of $\{f_n\}_{n=1}^{\infty}$ to actual boundedness in the metric space $\mathcal{C}(K)$. By equicontinuity there is $\delta > 0$ so that $|f(x) - f(y)| < 1$ for $d_K(x,y) < \delta$. Now select a finite set of balls $\{B_K(x_k,1)\}_{k=1}^{N}$ that cover the compact set K (the collection of all balls with unit radius covers K). Then $M_{x_k} = \sup_n |f_n(x_k)| < \infty$ for each x_k by pointwise boundedness, and so $M = \max_{1\leq k\leq N} M_{x_k} < \infty$. But then we have for any $x \in K$, if $x \in B_K(x_k, 1)$,

$$|f(x)| \leq |f(x) - f(x_k)| + |f(x_k)| < 1 + M_{x_k} \leq 1 + M.$$

Thus $\{f_n\}_{n=1}^{\infty} \subset B_{\mathcal{C}(K)}(0, M+1)$.

(2) We proceed in three steps.

Step 1: K has a countable dense subset E.

For each $n \in \mathbb{N}$ there is a finite set of balls $\left\{B_K\left(x_k^n, \frac{1}{n}\right)\right\}_{k=1}^{N_n}$ that cover K. Clearly the set $E = \bigcup_{n=1}^{\infty} \{x_k^n\}_{k=1}^{N_n}$ is countable and dense in K.

Step 2: There is a subsequence $\{g_n\}_{n=1}^{\infty}$ of $\{f_n\}_{n=1}^{\infty}$ that converges on E.

Relabel E as $E = \{e_k\}_{k=1}^{\infty}$. There is a subsequence $\{f_{n_k}\}_{k=1}^{\infty}$ of $\{f_n\}_{n=1}^{\infty}$ such that $\{f_{n_k}(e_1)\}_{k=1}^{\infty}$ converges in $\mathbb{C}$. There is then a subsequence $\left\{f_{n_{k_j}}\right\}_{j=1}^{\infty}$ of $\{f_{n_k}\}_{k=1}^{\infty}$ such that $\left\{f_{n_{k_j}}(e_2)\right\}_{j=1}^{\infty}$ converges in $\mathbb{C}$, and of course $\left\{f_{n_{k_j}}(e_1)\right\}_{j=1}^{\infty}$ converges as well. Repeating this procedure we obtain for each $\ell \in \mathbb{N}$ sequences $s_\ell = \left\{f_n^\ell\right\}_{n=1}^{\infty}$ such that s_1 is a subsequence of $s_0 \equiv \{f_n\}_{n=1}^{\infty}$, and $s_{\ell+1}$ is a subsequence of s_ℓ for all $\ell \in \mathbb{N}$. We also have that $s_\ell(e_k)$ converges for $1 \leq k \leq \ell$. Now Cantor's diagonal sequence $\{g_n\}_{n=1}^{\infty} \equiv \{f_n^n\}_{n=1}^{\infty}$ converges at each e_k in E.

Step 3: $\{g_n\}_{n=1}^{\infty}$ converges uniformly on K.

Let $\varepsilon > 0$ and by equicontinuity choose $\delta > 0$ so that

$$|f_n(x) - f_n(y)| < \varepsilon_1 \qquad \text{whenever } d_K(x, y) < \delta \text{ and } n \geq 1,$$

where $\varepsilon_1 > 0$ will be chosen later. Now let $\{B_K(y_j, \delta)\}_{j=1}^{J}$ be a finite set of balls centered at points y_j in E and with radius δ that cover K. For each j choose N_j so that

$$|g_m(y_j) - g_n(y_j)| < \varepsilon_2, \qquad m, n \geq N_j,$$

where $\varepsilon_2 > 0$ will be chosen later. Now set $N = \max_{1 \leq j \leq J} N_j$. Then for $m, n \geq N$ and $x \in K$, with say $x \in B_K(y_j, \delta)$, we have

$$\begin{aligned} |g_m(x) - g_n(x)| &= |g_m(x) - g_m(y_j) + g_m(y_j) - g_n(y_j) + g_n(y_j) - g_n(x)| \\ &\leq |g_m(x) - g_m(y_j)| + |g_m(y_j) - g_n(y_j)| + |g_n(y_j) - g_n(x)| \\ &< \varepsilon_1 + \varepsilon_2 + \varepsilon_1 < \varepsilon \end{aligned}$$

provided we choose $2\varepsilon_1 + \varepsilon_2 < \varepsilon$. This shows that $\{g_n\}_{n=1}^{\infty}$ converges uniformly on K, and since $\mathcal{C}(K)$ is complete, $\{g_n\}_{n=1}^{\infty}$ converges in $\mathcal{C}(K)$.

Remark 3. *Using the above theorem, it can be shown that a subset $\mathcal{F}$ of $\mathcal{C}(K)$ is compact* if and only if *$\mathcal{F}$ is closed, bounded and* equicontinuous. *Indeed, a compact set in any metric space is easily shown to be closed and bounded. To see that $\mathcal{F}$ is also equicontinuous, let $\varepsilon > 0$ and select a finite collection $\left\{B_{\mathcal{C}(K)}(f_k, \varepsilon_1)\right\}_{k=1}^{N}$ of balls in $\mathcal{C}(K)$ centered at f_k with radius ε_1 that cover $\mathcal{F}$. Since f_k is uniformly continuous there is $\delta_k > 0$ such that*

$$|f_k(x) - f_k(y)| < \varepsilon_2 \qquad \textit{whenever } d_K(x, y) < \delta_k.$$

Set $\delta = \min_{1 \leq k \leq N} \delta_k > 0$. Then if $d_K(x, y) < \delta$ and $f \in \mathcal{F}$, say $f \in B_{\mathcal{C}(K)}(f_k, \varepsilon_1)$, then

$$\begin{aligned} |f(x) - f(y)| &= |f(x) - f_k(x) + f_k(x) - f_k(y) + f_k(y) - f(y)| \\ &\leq |f(x) - f_k(x)| + |f_k(x) - f_k(y)| + |f_k(y) - f(y)| \\ &< \varepsilon_1 + \varepsilon_2 + \varepsilon_1 < \varepsilon \end{aligned}$$

if we choose $2\varepsilon_1 + \varepsilon_2 < \varepsilon$. Thus (4.1) holds.

Conversely, the Arzela-Ascoli theorem shows that every infinite set E in $\mathcal{F}$ has a limit point f in $\mathcal{C}(K)$, and since $\mathcal{F}$ is closed, $f \in \mathcal{F}$. Now it is a general fact that a metric space X is compact if and only if every infinite subset of X has a limit point in X. So we are done by the "if" statement of this general fact. The proof of this statement is a bit delicate. Using that X is contained in a finite union of balls of radius $\frac{1}{n}$ for each $n \in \mathbb{N}$, one first shows that there is a countable dense set E in X. Then the collection of balls $\mathcal{B} = \{B(x, r) : x \in E, r \in \mathbb{Q} \cap (0, 1)\}$ is a countable base. Now suppose that $\{G_\alpha\}_{\alpha \in A}$ is an open cover of X. For each $x \in X$ there is an index $\alpha \in A$ and a ball $B_x \in \mathcal{B}$ such that

$$x \in B_x \subset G_\alpha. \tag{4.2}$$

Note that the axiom of choice is not needed here since $\mathcal{B}$ is countable, hence well-ordered. If we can show that the cover $\widetilde{\mathcal{B}} = \{B_x\}_{x \in X}$ has a finite subcover, then (4.2) shows that $\{G_\alpha\}_{\alpha \in A}$ has a finite subcover as well. So it remains to show that

$\widetilde{\mathcal{B}}$ has a finite subcover. Relabel the countable base $\widetilde{\mathcal{B}}$ as $\widetilde{\mathcal{B}} = \{B_n\}_{n=1}^{\infty}$. Assume, in order to derive a contradiction, that $\widetilde{\mathcal{B}}$ has no finite subcover. Then the sets

$$F_N = X \setminus \left(\bigcup_{k=1}^{N} B_n \right)$$

are nonempty *closed sets that are decreasing, i.e. $F_{N+1} \subset F_N$, and that have* empty *intersection. Thus if we choose $x_N \in F_N$ for each N, the set $E = \bigcup_{N=1}^{\infty} \{x_N\}$ must be an infinite set, and so has a limit point $x \in X$. But then the fact that the F_N are closed and decreasing implies that $x \in F_N$ for all N, the desired contradiction.*

DEFINITION 7. *A family $\mathcal{F} \subset H(\Omega)$ of holomorphic functions on an open set Ω is said to be* normal *if every sequence of functions from $\mathcal{F}$ has a subsequence that converges uniformly on compact subsets of Ω (but not necessarily to a function in $\mathcal{F}$). In other words, $\mathcal{F}$ is "sequentially precompact".*

THEOREM 13. *(Montel's theorem) If $\mathcal{F} \subset H(\Omega)$ is uniformly bounded on each compact subset of Ω, then $\mathcal{F}$ is a normal family.*

Proof: Let $\{K_n\}_{n=1}^{\infty}$ be a seqence of compact subsets of Ω satisfying

$$\begin{aligned} K_n &\subset \overset{\circ}{K}_{n+1} \subset K_{n+1}, \qquad n \geq 1, \\ \cup_{n=1}^{\infty} K_n &= \Omega. \end{aligned}$$

In view of the Arzela-Ascoli theorem and the uniform convergence theorem, it suffices to prove that the restriction of $\mathcal{F}$ to the compact subset K_n is equicontinuous (since this will show that every sequence in $\mathcal{F}$ has a subsequence that converges uniformly on K_n; then we apply Cantor's diagonal trick).

So to prove that $\mathcal{F}\mid_{K_n}$ is equicontinuous, let

$$\delta_n = dist\left(K_n, \partial K_{n+1}\right) > 0,$$

and cover K_n with finitely many disks $\left\{B\left(z_k, \frac{\delta_n}{4}\right)\right\}_{k=1}^{N_n}$. If $z, w \in B\left(z_k, \frac{\delta_n}{2}\right)$ and $f \in \mathcal{F}$, we have

$$\begin{aligned} |f(z) - f(w)| &= \left| \int_{[w,z]} f'(w)\, dw \right| \\ &\leq |z - w| \sup_{\zeta \in B\left(z_k, \frac{\delta_n}{2}\right)} |f'(\zeta)| \\ &\leq |z - w| \left\{ \frac{1}{\left(\frac{\delta_n}{2}\right)} \sup_{\zeta \in B(z_k, \delta_n)} |f(\zeta)| \right\}, \end{aligned}$$

by Cauchy's inequality since for each $\zeta \in B\left(z_k, \frac{\delta_n}{2}\right)$, $B\left(\zeta, \frac{\delta_n}{2}\right) \subset B\left(z_k, \delta_n\right)$. Now we note that $B\left(z_k, \delta_n\right) \subset \overset{\circ}{K}_{n+1} \subset K_{n+1}$ and by hypothesis, there is a constant M_{n+1} such that $\sup_{\zeta \in K_{n+1}} |f(\zeta)| \leq M_{n+1}$ for $f \in \mathcal{F}$. Thus we obtain

$$|f(z) - f(w)| \leq \frac{2M_{n+1}}{\delta_n} |z - w|, \qquad z, w \in B\left(z_k, \frac{\delta_n}{2}\right), f \in \mathcal{F}. \tag{4.3}$$

Finally, we take a pair of points $z, w \in K_n$ such that $|z - w| < \frac{\delta_n}{4}$. There is k so that $z \in B\left(z_k, \frac{\delta_n}{4}\right)$. Then $w \in B\left(z_k, \frac{\delta_n}{2}\right)$ and so (4.3) yields

$$|f(z) - f(w)| \leq \frac{2M_{n+1}}{\delta_n} |z - w|, \qquad f \in \mathcal{F},$$

which easily implies the equicontinuity of $\mathcal{F}\mid_{K_n}$.

5. The open mapping theorem

An important topological property of a holomorphic map $f:\Omega\to\mathbb{C}$ on a connected open set is that **either** f is constant **or** f is an *open* map, i. e. $f(G)$ is open for all open subests G of Ω. This should be compared with the Invariance of Domain theorem of Brouwer that can use homotopy to conclude that any *one-to-one* continuous map on an open subset of the plane is an open map. More generally this can be extended to $\mathbb{R}^n$, $n>2$, using homology (see e.g. page 172 of [**2**]).

Theorem 14. *(Open mapping theorem) If f is holomorphic on an open connected set Ω in the complex plane, then $f(\Omega)$ is either a single point or an open set.*

Proof: Suppose that f is not constant, and fix $a\in\Omega$. We may suppose that $f(a)=0$. By the coincidence principle, there is $\overline{B(a,R)}\subset\Omega$ such that $f(z)\neq 0$ for all $z\in\partial B(a,R)$. By continuity of f and compactness of $\partial B(a,R)$,

$$\delta\equiv\min_{z\in\partial B(a,R)}|f(z)|>0.$$

We claim that

$$B\left(0,\frac{\delta}{2}\right)\subset f(B(a,R))\subset f(\Omega),\tag{5.1}$$

which clearly completes the proof that $f(\Omega)$ is open.

To see (5.1), choose $w\in B\left(0,\frac{\delta}{2}\right)$ and note that for $z\in\partial B(a,R)$,

$$\delta\leq|f(z)|\leq|f(z)-w|+|w|<|f(z)-w|+\frac{\delta}{2}.$$

Thus

$$|f(a)-w|=|w|<\frac{\delta}{2}\leq\min_{z\in\partial B(a,R)}|f(z)-w|,$$

and now the mimimum principle, Remark 2, implies that $f(z)-w$ *cannot* be non-vanishing in any open neighbourhood of the closed ball $\overline{B(a,R)}$. It follows that $f(z_0)=w$ for some $z_0\in B(a,R)$ and this completes the proof that (5.1) holds.

5.1. Locally injective holomorphic functions.

If f is holomorphic in a neighbourhood of a point a where $f'(a)\neq 0$, then there is $R>0$ such that f is one-to-one in the disk $B(a,R)$. Indeed, choose R so small that $|f'(z)-f'(a)|<\frac{1}{2}|f'(a)|$ for $z\in B(a,R)$. Then we have for $w_0,w_1\in B(a,R)$,

$$f(w_1)-f(w_0)=\int_{[w_0,w_1]}f'(z)\,dz=\int_{[w_0,w_1]}f'(a)\,dz+\int_{[w_0,w_1]}[f'(z)-f'(a)]\,dz.$$

Now

$$\int_{[w_0,w_1]}f'(a)\,dz=f'(a)(w_1-w_0)$$

and

$$\left|\int_{[w_0,w_1]}[f'(z)-f'(a)]\,dz\right|\leq\frac{1}{2}|f'(a)|\,|w_1-w_0|,$$

and it follows that

$$|f(w_1)-f(w_0)|>\frac{1}{2}|f'(a)|\,|w_1-w_0|>0.$$

Combining this observation with the open mapping theorem we get:

PROPOSITION 5. *If f is holomorphic in a neighbourhood of a point a where $f'(a) \neq 0$, then there are open sets U and V such that $a \in U$ and $f : U \to V$ is one-to-one and onto.*

We will also need to know that if the derivative f' of a holomorphic function f vanishes at a point a, then f is *not* one-to-one in any neighbourhood of a. This can be proved in many ways. We will use the fact that the Jacobian $J_f(z)$ of a holomorphic function $f = u + iv$ at the point z is given by

$$\begin{aligned} J_f(z) &= \det \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} = u_x v_y - v_x u_y \\ &= u_x^2 + v_x^2 = \left| \frac{\partial}{\partial x} f(z) \right|^2 = |f'(z)|^2, \end{aligned}$$

where the first equality in the second line follows from the Cauchy-Riemann equations. The idea of the proof is to show that if $f'(a) = 0$, then the integral of the Jacobian J_f over a small disk $B(a, R)$ is too big to be the area of the image $f(B(a, R))$.

PROPOSITION 6. *Suppose that $f \in H(\Omega)$ and that $f'(a) = 0$ for some $a \in \Omega$. Then f is* not *one-to-one in* any *disk $B(a, R) \subset \Omega$.*

Proof: Suppose without loss of generality that $a = 0$ and that f is not identically zero near the origin. Then

$$g(z) = f(z) - f(0) = z^n h(z)$$

where $h \in H(\Omega)$ satisfies $h(0) \neq 0$. Since $f'(0) = 0$ we must have $n \geq 2$. Now assume in order to derive a contradiction that f is one-to-one in some disk $B(0, R) \subset \Omega$, hence in every disk $B(0, r)$ with $0 < r \leq R$. Then the change of variable formula for g implies that

$$area\,(g(B(0, r))) = \int_{B(0,r)} J_g dxdy, \qquad 0 < r \leq R.$$

However, given $\varepsilon > 0$,

$$g(B(0, r)) \subset B(0, r^n (|h(0)| + \varepsilon))$$

for r sufficiently small, and then

$$area\,(g(B(0, r))) \leq \pi r^{2n} (|h(0)| + \varepsilon)^2.$$

On the other hand,

$$\begin{aligned} J_g(z) &= |g'(z)|^2 = \left| n z^{n-1} h(z) + z^n h'(z) \right|^2 = |z|^{2n-2} |nh(z) + zh'(z)|^2 \\ &\geq |z|^{2n-2} n^2 (|h(0)| - \varepsilon)^2 \end{aligned}$$

for $|z|$ sufficiently small, and then

$$\begin{aligned} \int_{B(0,r)} J_g dxdy &\geq 2\pi \int_{B(0,r)} s^{2n-2} n^2 (|h(0)| - \varepsilon)^2 s ds \\ &= 2\pi n^2 (|h(0)| - \varepsilon)^2 \frac{r^{2n}}{2n}, \end{aligned}$$

for r sufficiently small. Altogether then, for r sufficiently small,

$$2 \leq n \leq \left(\frac{|h(0)| + \varepsilon}{|h(0)| - \varepsilon} \right)^2,$$

which is a contradiction if we take $\varepsilon > 0$ small enough.

Corollary 6. *A holomorphic function f defined in a neighbourhood of a point a is one-to-one in some neighbourhood of a if and only if $f'(a) \neq 0$.*

CHAPTER 4

The Riemann Mapping Theorem

The previous chapter established that holomorphic functions, those having a *complex* derivative on an open set in the complex plane $\mathbb{C}$, have many "magical" properties compared to those functions having merely a *real* derivative. The purpose of this chapter is to show that, on the other hand, holomorphic functions are numerous enough to supply homeomorphisms between arbitrary nontrivial connected and simply connected open subsets of the plane. We will follow for the most part the treatment in Chapter 14 of Rudin [**5**].

THEOREM 15. *(Riemann Mapping Theorem) A subset Ω of the complex plane is the image $f(\mathbb{D})$ of a one-to-one onto holomorphic map $f : \mathbb{D} \to \Omega$ if and only if*

(1) *Ω is open*
(2) *Ω is connected and simply connected*
(3) $\Omega \neq \mathbb{C}$

A one-to-one onto holomorphic map $f : \mathbb{D} \to \Omega$ is called a *Riemann mapping* for Ω. The open mapping theorem shows that $f^{-1} : \Omega \to \mathbb{D}$ is continuous, hence a homeomorphism between Ω and $\mathbb{D}$, and Proposition 6 shows that f' is nonvanishing. Thus f^{-1} is also holomorphic with derivative

$$\begin{aligned}\left(f^{-1}\right)'(w) &= \lim_{k\to 0} \frac{\left(f^{-1}\right)(w+k) - \left(f^{-1}\right)(w)}{k} \\ &= \lim_{h\to 0} \frac{h}{f(z+h) - f(z)} = \frac{1}{f'(z)},\end{aligned}$$

where $f(z) = w$ and $f(z+h) = w+k$. Thus a Riemann mapping is a *biholomorphic* map (meaning both the map and its inverse are holomorphic).

Now we can easily obtain the necessity of properties (1), (2) and (3) in Theorem 15. Indeed, (1) is necessary by the open mapping theorem, (2) follows from topology and the fact that f is a homeomorphism between $\mathbb{D}$ and Ω, and (3) now follows from an application of Liouville's theorem to the holomorphic inverse function $f^{-1} : \Omega \to \mathbb{D}$.

The question of characterizing *all* the Riemann maps f for Ω, in terms of one of them, is easily reduced to the special case $\Omega = \mathbb{D}$, to which we turn in the next section. The Riemann maps for $\mathbb{D}$ are the biholomorphic maps from $\mathbb{D}$ to itself, usually called the automorphisms of $\mathbb{D}$. These maps play an important role in the proof of the Riemann Mapping Theorem.

1. Automorphisms of the disk

We begin by demonstrating that the only automorphisms of the disk that fix the center 0, are the rotations $U_\theta(z) = e^{i\theta}z$, $z \in \mathbb{D}$. This will follow from the Schwarz lemma below. Part (2) of this lemma will prove to be a very powerful tool.

LEMMA 1. *(The Schwarz Lemma) Suppose that* $f : \mathbb{D} \to \overline{\mathbb{D}}$ *is holomorphic and* $f(0) = 0$.

(1) *Then*

$$\begin{aligned} |f(z)| &\leq |z|, \qquad z \in \mathbb{D}, \\ |f'(0)| &\leq 1. \end{aligned} \tag{1.1}$$

(2) *If equality holds for at least one* z *in the first line, or if equality holds in the second line, then there is* $\theta \in [0, 2\pi)$ *such that* f *is the rotation* $f(z) = e^{i\theta} z$, $z \in \mathbb{D}$.

Proof: The function $g(z) = \frac{f(z)}{z}$ has a removable singularity at the origin and by the maximum principle,

$$\sup_{z \in B(0,r)} |g(z)| \leq \sup_{z \in \partial B(0,r)} \frac{|f(z)|}{|z|} \leq \frac{1}{r}, \qquad 0 < r < 1.$$

Thus we obtain $\sup_{z \in \mathbb{D}} |g(z)| \leq 1$, and hence (1.1) since $g(0) = f'(0)$. If equality holds as in part (2) of the lemma, then $|g(z)|$ attains its maximum inside the disk, and so g is a constant of modulus one, hence f is a rotation.

COROLLARY 7. *The only automorphisms of* $\mathbb{D}$ *that fix* 0 *are the rotations.*

Proof: Let $f : \mathbb{D} \to \mathbb{D}$ be an automorphism. Then $f^{-1} : \mathbb{D} \to \mathbb{D}$ is also an automorphism, and by the chain rule,

$$1 = f'(0) \left(f^{-1}\right)'(0). \tag{1.2}$$

Now part (1) of the Schwarz lemma implies that both $|f'(0)| \leq 1$ and $\left|\left(f^{-1}\right)'(0)\right| \leq 1$, and it follows from (1.2) that we must have both $|f'(0)| = 1$ and $\left|\left(f^{-1}\right)'(0)\right| = 1$. Part (2) of the Schwarz lemma now shows that f is a rotation.

Another class of automorphisms are given by the involutions, which are special biholomorphic maps that interchange 0 with a point $w \in \mathbb{D}$. For $w \in \mathbb{D}$ define

$$\varphi_w(z) = \frac{w - z}{1 - \overline{w} z}, \qquad z \in \mathbb{C} \setminus \left\{\overline{w}^{-1}\right\}.$$

Then $\varphi_w \in H\left(\mathbb{C} \setminus \left\{\overline{w}^{-1}\right\}\right)$ and satisfies

$$\begin{aligned} \varphi_w(0) &= w \text{ and } \varphi_w(w) = 0, \\ \varphi_w'(0) &= 1 - |w|^2 \text{ and } \varphi_w'(w) = \frac{1}{1 - |w|^2}, \end{aligned}$$

and

$$\varphi_w \circ \varphi_w(z) = z \text{ for } z \in \mathbb{C} \setminus \left\{\overline{w}^{-1}\right\}, \tag{1.3}$$

i.e. φ_w is its own inverse! Indeed, since $|w|<1$,

$$\begin{aligned}\varphi_w \circ \varphi_w(z) &= \frac{w-\varphi_w(z)}{1-\overline{w}\varphi_w(z)} = \frac{w-\frac{w-z}{1-\overline{w}z}}{1-\overline{w}\frac{w-z}{1-\overline{w}z}} \\ &= \frac{(1-\overline{w}z)w-(w-z)}{(1-\overline{w}z)-\overline{w}(w-z)} \\ &= \frac{z\left(1-|w|^2\right)}{\left(1-|w|^2\right)} = z.\end{aligned}$$

Claim 1. *The functions φ_w satisfy the following bijections:*

$$\begin{aligned}\varphi_w &: \mathbb{C}\setminus\left\{\overline{w}^{-1}\right\} \to \mathbb{C}\setminus\left\{\overline{w}^{-1}\right\} \text{ is a bijection,} \\ \varphi_w &: \mathbb{T}\to\mathbb{T} \text{ is a bijection,} \\ \varphi_w &: \mathbb{D}\to\mathbb{D} \text{ is a bijection,} \\ \varphi_w &: \overline{\mathbb{D}}\to\overline{\mathbb{D}} \text{ is a bijection.}\end{aligned}$$

Proof: The first bijection follows from (1.3). If $|z|=1$, then $z^{-1}=\overline{z}$ since $z\overline{z}=\overline{z}z=|z|^2=1$. If $|w|<1$, then $1-\overline{w}z\neq 0$ and we have

$$\left|\frac{w-z}{1-\overline{w}z}\right| = \left|z\frac{w\overline{z}-1}{1-\overline{w}z}\right| = \frac{|w\overline{z}-1|}{|\overline{w}z-1|} = 1$$

since the numbers $w\overline{z}-1$ and $\overline{w}z-1$ are complex conjugates, and hence have the same modulus. Thus the holomorphic function φ_w maps the unit circle $\mathbb{T}$ into itself. In fact, φ_w maps $\mathbb{T}$ *onto* $\mathbb{T}$ since φ_w is its own inverse. Now $\varphi_w(0)=w$ lies in the open unit disk $\mathbb{D}$, and since $\varphi_w(\mathbb{D})$ is a connected subset of $\mathbb{C}\setminus\mathbb{T}$, we have that $\varphi_w(\mathbb{D})$ is contained in the component of $\mathbb{C}\setminus\mathbb{T}$ containing w, i.e. $\varphi_w(\mathbb{D})\subset\mathbb{D}$ (alternatively, we could appeal to the maximum principle to conclude that $\varphi_w(\mathbb{D})\subset\mathbb{D}$). Thus φ_w and $(\varphi_w)^{-1}=\varphi_w$ each map $\mathbb{D}$ into $\mathbb{D}$, and it follows that $\varphi_w:\mathbb{D}\to\mathbb{D}$ is a bijection. Since we also showed $\varphi_w:\mathbb{T}\to\mathbb{T}$ is a bijection, we have that $\varphi_w:\overline{\mathbb{D}}\to\overline{\mathbb{D}}$ is a bijection as well.

The group generated by rotations and involutions is the entire automorphism group. Moreover, we have the following two unique representations.

Proposition 7. *If φ is an autormorphism of the disk such that $\varphi(0)=a$ and $\varphi(b)=0$, then there are rotations U_θ and U_ϕ such that*

$$\varphi=\varphi_a\circ U_\theta=U_\phi\circ\varphi_b.$$

Proof: The automorphisms $\varphi_a\circ\varphi$ and $\varphi\circ\varphi_b$ each fix 0, and so are rotations U_θ and U_ϕ. But then $\varphi_a\circ\varphi=U_\theta$ implies

$$\varphi=\varphi_a\circ\varphi_a\circ\varphi=\varphi_a\circ U_\theta,$$

and $\varphi\circ\varphi_b=U_\phi$ implies

$$\varphi=\varphi\circ\varphi_b\circ\varphi_b=U_\phi\circ\varphi_b.$$

2. Simply connected subsets of the plane

We say that two closed curves $\gamma_0 : \mathbb{T} \to \Omega$ and $\gamma_1 : \mathbb{T} \to \Omega$ in an open set Ω of the complex plane are Ω-*homotopic* if there is a continuous map

$$\Gamma : \mathbb{T} \times [0,1] \to \mathbb{C}$$

such that $\gamma_0(\zeta) = \Gamma(\zeta, 0)$ and $\gamma_1(\zeta) = \Gamma(\zeta, 1)$ for $\zeta \in \mathbb{T}$.

An open subset Ω of the plane $\mathbb{C}$ is *simply connected* if every closed curve γ in Ω is Ω-homotopic to a constant curve in Ω.

To prove the theorems in this chapter we will use Proposition 8, the Jordan Curve Theorem for taxicab paths, along with a standard homotopy result, Proposition 9. See the appendix for a statement and proof of the general form of the Jordan Curve Theorem. A *taxicab path* is a finite concatenation of line segments that are each parallel to either the real or imaginary axis.

Proposition 8. *A* simple closed *taxicab path γ divides the plane into two connected components, one of which is bounded and in which Ind_γ takes either the value* 1 *there, or the value* -1 *there. More precisely, $\mathbb{C} \setminus \gamma^* = \mathcal{U} \cup \mathcal{B}$ where both $\mathcal{U}$ and $\mathcal{B}$ are connected open subsets of $\mathbb{C}$ with $\mathcal{U}$ unbounded and $\mathcal{B}$ bounded, and $Ind_\gamma(a)$ is the constant ± 1 for all $a \in \mathcal{B}$.*

We borrow an idea from [**6**] to prove Proposition 8 - see after Step 2 in the proof of Proposition 17 in the appendix. Here is the homotopy result.

Proposition 9. *Let γ_0 and γ_1 be two closed paths in an open set Ω of the complex plane. If γ_0 and γ_1 are Ω-homotopic, then $Ind_{\gamma_0}(a) = Ind_{\gamma_1}(a)$ for all $a \in \mathbb{C} \setminus \Omega$.*

The proof of Proposition 9 can also be found in the appendix.

We now use these two propositions to extend Cauchy's theorem to closed paths in simply connected domains Ω. We use the word "domain" to denote an open connected subset of the plane, so as to avoid cumbersome expressions like "a simply connected connected open set". Here we will use an idea in Part 6C of Chapter 1 of Boas [**1**].

Theorem 16. *Let Ω be a simply connected domain in the complex plane and suppose that γ is a closed path in Ω. Then for $f \in H(\Omega)$ we have*

$$\int_\gamma f(z)\, dz = 0.$$

Proof: By the fundamental theorem of line integrals, it suffices to construct a holomorphic antiderivative F of f in Ω. For this we pick a reference point $z_0 \in \Omega$ and define

$$F(z) = \int_\sigma f(z)\, dz,$$

where σ is any *simple* taxicab path in Ω joining z_0 to z.

We claim that the set

$$E = \{z \in \Omega : \text{there is a simple taxicab path joining } z_0 \text{ to } z\}$$

is both open and closed in Ω, and since Ω is connected, we then have $E = \Omega$. Indeed, if we remove the final segment from a simple taxicab path γ joining z_0 to z, and call the resulting path β, then there is a disk $B(z,r) \subset \Omega \setminus \beta^*$, and clearly we have $B(z,r) \subset E$. Conversely, suppose that $w \in \Omega$ is a limit point of E. Choose $\delta > 0$ such that $B(w,\delta) \subset \Omega$ and then choose $z \in E \cap B(w,\delta)$. Let γ be a simple taxicab path joining z_0 to z. If $w \in \gamma^*$, we clearly have $w \in E$. Otherwise, $r = dist(\gamma^*, w) > 0$ and there is a *first* point $z \in \gamma^*$ such that $|z-w| = r$. Now it is clear that $w \in E$.

We claim that if σ and τ are each simple taxicab paths in Ω joining z_0 to z, then

$$\int_\sigma f(z)\,dz = \int_\tau f(z)\,dz. \tag{2.1}$$

Indeed, if we denote by $\rho = \sigma - \tau$ the (not necessarily simple) taxicab path that follows σ from z_0 to z and then continues on by following τ backwards from z to z_0, we can write

$$\int_\rho f(z)\,dz = \sum_i \int_{\rho_i} f(z)\,dz \tag{2.2}$$

where each ρ_i is a *simple* closed taxicab path in Ω, and the sum is finite.

Indeed, we perform the following algorithm while proceeding in the forward direction along σ. If there is an initial taxicab segment along which σ and τ proceed in opposite directions, let z_1 be the first point at which the paths diverge, and discard the segment traversed. Otherwise, proceed along σ from z_0 until the first time σ^* encounters a point z_1 from τ^*. Then let ρ_1 be the closed taxicab path that follows σ from z_0 to z_1 and continues on by following τ backwards from z_1 to z_0. Clearly ρ_1 is simple. Now set aside ρ_1 and apply the same procedure starting at z_1. This produces either a taxicab segment along which σ and τ proceed in opposite directions, or a simple closed taxicab path ρ_2. Continue this algorithm until we reach z. Of course we can ignore those discarded taxicab segments along which σ and τ travel in opposite directions since the line integrals along these segments cancel each other. Collecting all the simple closed taxicab paths ρ_i that were set aside results in (2.2).

Now ρ_i is a simple closed taxicab path, and so by Propositions 8 and 9, the "inside" $\mathcal{B}_i$ of ρ_i (the bounded component $\mathcal{B}_i$ of $\mathbb{C} \setminus \rho_i^* = \mathcal{U}_i \cup \mathcal{B}_i$) is contained in Ω since Ω is connected and simply connected. Indeed, if there is a point $a \in \mathcal{B}_i \setminus \Omega$ then $Ind_{\rho_i}(a) = \pm 1$ by Proposition 8. On the other hand, since Ω is simply connected, ρ_i is Ω-homotopic to a constant map, and so $Ind_{\rho_i}(a) = Ind_{constant}(a) = 0$ by Proposition 9, the desired contradiction. Since ρ_i is a taxicab path, we can write

$$\int_{\rho_i} f(z)\,dz = \sum_j \int_{\partial R_j^i} f(z)\,dz$$

where R_j^i is a rectangle contained inside ρ_i (hence in Ω), ∂R_j^i has the same orientation as ρ_i, and the sum is finite for each i. Indeed, simply construct a grid of infinite lines in the plane, each passing through one of the segments in ρ_i. This creates a collection of minimal rectangles with sides that are segments of these lines. Then the inside of ρ_i is the union of all the minimal rectangles R_j^i that happen to lie inside ρ_i. Finally we know that $\int_{\partial R_j^i} f(z)\,dz = 0$ by Cauchy's theorem for a rectangle in a convex set, and summing over i and j proves (2.1).

It remains to prove that $F'(z)$ exists for each $z \in \Omega$. So fix z and a simple taxicab path γ joining z_0 to z. Let β denote the path obtained from γ by deleting the final segment, and let $r > 0$ be such that $B(z,r) \subset \Omega \setminus \beta^*$. For $|h| < r$, let γ_h be the taxicab path consisting of γ followed by a (possibly empty) horizontal segment and then a (possibly empty) vertical segment ending at $z+h$ (if this procedure results in doubling back along the final segment of γ, simply remove the cancelled portion). Using Cauchy's theorem for a triangle in a convex set, we see that for h small we have

$$F(z+h)-F(z)=\int_{\gamma_h} f(w)\,dw-\int_{\gamma} f(w)\,dw=\int_{[z,z+h]} f(w)\,dw,$$

and so

$$\lim_{h\to 0}\frac{F(z+h)-F(z)}{h}=\lim_{h\to 0}\frac{1}{h}\int_{[z,z+h]} f(w)\,dw=f(z).$$

The above proof yields the following technical consequence which will find application in proving the Jordan Curve Theorem below.

Porism 2: Suppose that Ω is a connected open subset of the complex plane satisfying

$$\int_{\gamma} f(z)\,dz=0,\qquad f\in H(\Omega),$$

for every simple closed taxicab path γ in Ω. Then the argument in the proof of Theorem 16 shows that every $f\in H(\Omega)$ has an antiderivative F in $H(\Omega)$.

The following corollary will be indispensible in our proof of the Riemann Mapping Theorem. It shows that nonvanishing holomorphic functions have logarithms in simply connected domains.

Corollary 8. *Suppose that Ω is a simply connected domain in $\mathbb{C}$. If $f\in H(\Omega)$ is nonvanishing in Ω, then there is $g\in H(\Omega)$ satisfying*

$$f(z)=e^{g(z)},\qquad z\in\Omega.$$

Proof: Let $h(z)=\frac{f'(z)}{f(z)}$ for $z\in\Omega$. Then there is $g\in H(\Omega)$ such that $g'(z)=h(z)$ for $z\in\Omega$. The function $f(z)\,e^{-g(z)}$ has derivative

$$f'(z)\,e^{-g(z)}-f(z)\,e^{-g(z)}g'(z)=0,$$

hence is a nonzero constant c in Ω, so that $f(z)=ce^{g(z)}$. We can incorporate the nonzero constant c into g to obtain the corollary.

3. Proof of the Riemann Mapping Theorem

As motivation for the proof of the Riemann Mapping Theorem, we consider the following extremal problem.

Problem 2. *Given $\alpha,\beta\in\mathbb{D}$, calculate the maximum value*

$$M(\alpha,\beta)\equiv\sup\{|f'(\alpha)|: f:\mathbb{D}\to\mathbb{D}\text{ is holomorphic and }\beta=f(\alpha)\},$$

and determine the extremal functions.

We use involutions to move the extremal problem to the origin. Let

$$g=\varphi_\beta\circ f\circ\varphi_\alpha:\mathbb{D}\to\mathbb{D}.$$

Then $g(0) = 0$ and so the Schwarz lemma and the chain rule yield

$$1 \geq |g'(0)| = \left|\varphi'_\beta(\beta) f'(\alpha) \varphi'_\alpha(0)\right| = \frac{1}{1 - |\beta|^2} |f'(\alpha)| \left(1 - |\alpha|^2\right).$$

Thus we have

$$|f'(\alpha)| \leq \frac{1 - |\beta|^2}{1 - |\alpha|^2} \equiv M(\alpha, \beta),$$

where by assertion (2) of the Schwarz lemma, equality is achieved precisely when $\varphi_\beta \circ f \circ \varphi_\alpha$ is a rotation R_θ, i.e.

$$f = \varphi_\beta \circ R_\theta \circ \varphi_\alpha, \qquad \theta \in [0, 2\pi).$$

Note: It is useful to note that for fixed $\alpha \in \mathbb{D}$, the solution to Problem 2 shows that the largest value of $|f'(\alpha)|$ for $f : \mathbb{D} \to \mathbb{D}$ holomorphic occurs when $f(\alpha) = \beta = 0$.

REMARK 4. *It is remarkable that the extremal solutions to Problem 2 are in fact autormorphisms of the disk* - rational *functions taking the disk one-to-one onto itself. We will obtain a Riemann mapping for a proper simply connected domain Ω in the complex plane by solving the analogous extremal problem for certain holomorphic maps from Ω to $\mathbb{D}$. However, the extremals for this problem will not be evident, and we will use Montel's theorem on a certain normal family to obtain the existence of an extremal, which turns out to be (the inverse of) a Riemann map for Ω.*

We present the proof of the Riemann Mapping Theorem in three steps. Suppose that Ω is a proper simply connected domain in $\mathbb{C}$. The only consequence of Ω being simply connected that we use is this: every nonvanishing holomorphic function $f \in H(\Omega)$ has a holomorphic square root $h \in H(\Omega)$. This follows immediately from Corollary 8 if we set $h(z) = e^{\frac{g(z)}{2}}$, $z \in \Omega$. For future reference we record this observation.

Porism 3: A subset Ω of the complex plane is the image $f(\mathbb{D})$ of a one-to-one onto holomorphic map $f : \mathbb{D} \to \Omega$ if and only if

(1) Ω is open,
(2) Ω is connected and every nonvanishing holomorphic function $f \in H(\Omega)$ has a holomorphic square root $h \in H(\Omega)$,
(3) $\Omega \neq \mathbb{C}$.

Let

$$\Sigma = \{f \in H(\Omega) : f \text{ is } 1 - 1 \text{ and into } \mathbb{D}\}.$$

Our task is to show that there is $f \in \Sigma$ that is *onto* the unit disk $\mathbb{D}$.

Step 1: $\Sigma \neq \phi$

Here is where we use that Σ is a proper subset of $\mathbb{C}$. Pick $a \in \mathbb{C} \setminus \Sigma$. Then $f(z) = \frac{1}{z-a}$ is holomorphic and one-to-one in Ω, but not necessarily bounded in Ω. To fix this, note that f is nonvanishing in Ω, so has a square root $h \in H(\Omega)$: $h(z)^2 = f(z)$ for $z \in \Omega$. By the open mapping theorem, there is a disk $B(w, r) \subset h(\Omega) \setminus \mathbb{R}$. But then

$$B(-w, r) \cap h(\Omega) = \phi. \tag{3.1}$$

To see this we argue by contradiction. If there is $w_1 \in B(-w, r) \cap h(\Omega)$, then $w_2 = -w_1 \in B(w, r) \subset h(\Omega)$ and $w_2 \neq w_1$. Thus there are *distinct* points z_1 and z_2 in Ω such that

$$h(z_1) = w_1 \text{ and } h(z_2) = w_2.$$

But then

$$\frac{1}{z_1 - a} = f(z_1) = h(z_1)^2 = w_1^2 = w_2^2 = h(z_2)^2 = f(z_2) = \frac{1}{z_2 - a}$$

implies that $z_1 = z_2$, a contradiction.

Now we take $g(z) = \frac{1}{h(z)+w}$ for $z \in \Omega$. Clearly (3.1) shows that $g \in H(\Omega)$ is bounded by $\frac{1}{r}$. We also have that g is one-to-one since

$$g(z) = g(z') \Longrightarrow h(z) = h(z') \Longrightarrow f(z) = f(z') \Longrightarrow z = z',$$

since f is one-to-one. Thus $g \in \Sigma$.

Step 2: Fix $z_0 \in \Omega$ and $f \in \Sigma$. If f is *not* onto the unit disk $\mathbb{D}$, then there exists $h \in \Sigma$ such that

$$|h'(z_0)| > |f'(z_0)|.$$

Suppose $f \in \Sigma$ and that f is not onto the disk, say $\alpha \in \mathbb{D} \setminus f(\Omega)$. Then $\varphi_\alpha \circ f \in \Sigma$ and $0 \notin \varphi_\alpha \circ f(\Omega)$. Thus $\varphi_\alpha \circ f$ has a holomorphic square root $g \in \Sigma$:

$$g(z)^2 = \varphi_\alpha \circ f(z), \qquad z \in \Omega.$$

We now compose g with an involution φ_β chosen so that $\varphi_\beta \circ g(z_0) = 0$, since we expect that this choice will maximize $\left|(\varphi_\beta \circ g)'(z_0)\right|$. This choice requires that we take $\beta = g(z_0)$, and we can now compute that the function

$$h = \varphi_\beta \circ g \in \Sigma$$

satisfies

$$\varphi_\alpha \circ S \circ \varphi_\beta \circ h = f,$$

where the function S is defined by $S(z) = z^2$.

Now S fails to be one-to-one on $\mathbb{D}$, and hence $\psi = \varphi_\alpha \circ S \circ \varphi_\beta$ fails to be one-to-one on $\mathbb{D}$ as well. The Note following Problem 2 thus implies that

$$\left|\psi'(0)\right| < 1.$$

Since $h(z_0) = 0$, the chain rule now gives

$$|f'(z_0)| = \left|\psi'(0)\, h'(z_0)\right| < |h'(z_0)|.$$

Step 3: Fix $z_0 \in \Omega$ and let $M = \sup_{f \in \Sigma} |f'(z_0)|$. Then $M > 0$ and there is an extremal function $h \in \Sigma$ satisfying $|h'(z_0)| = M$.

Step 1 and Proposition 6 shows that $M > 0$. Since $|f(z)| < 1$ for all $f \in \Sigma$ and $z \in \Omega$, Montel's Theorem 13 shows that Σ is a normal family. Thus there is a sequence $\{f_n\}_{n=1}^\infty \subset \Sigma$ that converges uniformly on compact subsets of Ω such that

$$\lim_{n\to\infty} \operatorname{Re} f_n'(z_0) = \lim_{n\to\infty} |f_n'(z_0)| = M.$$

Let $h = \lim_{n\to\infty} f_n$. By the Uniform Convergence Theorem 11, $h \in H(\Omega)$ and $\operatorname{Re} h'(z_0) = \lim_{n\to\infty} \operatorname{Re} f_n'(z_0) = M$. Since $|h(z)| = \lim_{n\to\infty} |f_n(z)| \le 1$, we have $h(\Omega) \subset \overline{\mathbb{D}}$, and the open mapping theorem now shows that $h(\Omega) \subset \mathbb{D}$ (h is not constant since $M > 0$). Thus the only thing remaining to verify in order to conclude that $h \in \Sigma$ is that h is one-to-one in Ω.

To prove that h is one-to-one in Ω, fix $z_1, z_2 \in \Omega$ with $z_1 \neq z_2$. Since h is not constant (it has nonzero derivative at z_0 since $M > 0$), the coincidence principle implies that there is $r > 0$ such that $z_1 \notin \overline{B(z_2, r)}$ and also

$$\inf_{z \in \partial B(z_2,r)} |h(z) - h(z_1)| > 0.$$

Now each f_n is one-to-one on Ω and it follows that $f_n(z) - f_n(z_1)$ is nonvanishing in a neighourhood of $\overline{B(z_2, r)}$. Thus uniform convergence of f_n to h on $\overline{B(z_2, r)}$, together with the minimum principle, yields

$$\begin{aligned} 0 &< \inf_{z \in \partial B(z_2,r)} |h(z) - h(z_1)| = \lim_{n\to\infty} \inf_{z \in \partial B(z_2,r)} |f_n(z) - f_n(z_1)| \\ &\leq \lim_{n\to\infty} |f_n(z_2) - f_n(z_1)| = |h(z_2) - h(z_1)|. \end{aligned}$$

Thus $h(z_2) \neq h(z_1)$ and this completes the proof that $h \in \Sigma$.

It is now easy to complete the proof of the Riemann Mapping Theorem. Indeed, by Step 2 the extremal function h in Step 3 must be onto the unit disk $\mathbb{D}$, and so the inverse $f = h^{-1} : \mathbb{D} \to \Omega$ of h is a Riemann map for Ω.

Note 1: The Riemann map $f = h^{-1}$ constructed in the proof above satisfies $f(0) = z_0$. Indeed, if $h(z_0) = \beta \neq 0$, then

$$\left|(\varphi_\beta \circ h)'(z_0)\right| = \left|\varphi_\beta'(\beta)\, h'(z_0)\right| = \frac{|h'(z_0)|}{1 - |\beta|^2} > |h'(z_0)|,$$

contradicting Step 3. All other Riemann maps g for Ω have the form $g = f \circ \varphi$ for some automorphism of the unit disk.

Note 2: Suppose $f(x + iy) = u + iv$ maps $\mathbb{D}$ one-to-one and onto Ω, and is continuously (real) differentiable, but not necessarily holomorphic. If $\nabla f = (f_x, f_y)$ and $Jf = \det \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix}$, then a calculation yields

$$\begin{aligned} \frac{1}{2} |\nabla f|^2 &= \left|\frac{\partial f}{\partial z}\right|^2 + \left|\frac{\partial f}{\partial \overline{z}}\right|^2, \\ Jf &= \left|\frac{\partial f}{\partial z}\right|^2 - \left|\frac{\partial f}{\partial \overline{z}}\right|^2, \end{aligned}$$

and if $Jf \geq 0$ we conclude that the Dirichlet (or energy) integral

$$\begin{aligned} \int_{\mathbb{D}} \frac{|\nabla f|^2}{2} dxdy &= \int_{\mathbb{D}} Jf\ dxdy + 2 \int_{\mathbb{D}} \left|\frac{\partial f}{\partial \overline{z}}\right|^2 dxdy \\ &= area(\Omega) + 2 \int_{\mathbb{D}} \left|\frac{\partial f}{\partial \overline{z}}\right|^2 dxdy \end{aligned}$$

achieves its minimum value $area(\Omega)$ if and only if f is a Riemann map.

CHAPTER 5

Contour integrals and the Prime Number Theorem

We will now develop the theory of contour integrals and the residue theorem, and apply these results to prove the Prime Number Theorem: if

$$\pi(x) \text{ equals the number of positive primes } \leq x,$$

then

$$\lim_{x\to\infty} \frac{\pi(x)}{\frac{x}{\ln x}} = 1. \tag{0.2}$$

We will follow for the most part the treatment in Chapter 7 of Stein and Shakarchi [**6**].

The basic connection between complex analysis and prime numbers is the following beautiful identity of Euler:

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p\in\mathcal{P}} \frac{1}{1-\frac{1}{p^s}}, \qquad \operatorname{Re} s > 1, \tag{0.3}$$

where $\mathcal{P} = \{2,3,5,7,11,13,...\}$ is the set of *prime* positive integers p; p is prime if it has no positive integer factors other than 1 and itself (we exclude the multiplicative identity 1 from $\mathcal{P}$). We will often adopt the convention, common in number theory, of simply writing $\prod_p$ or $\sum_p$ to denote the infinite product or sum over all prime numbers $\mathcal{P}$. It is also customary in number theory to write a complex variable as $s = \sigma + i\tau$. Finally, we point out that for $a > 0$ and $s \in \mathbb{C}$, we define $a^s = e^{s\ln a}$.

At the core of the proof of Euler's identity (0.3) is the Fundamental Theorem of Arithmetic: every positive integer $n > 1$ has a *unique* factorization into a finite product of primes,

$$n = p_1^{k_1} p_2^{k_2} ... p_m^{k_m}, \tag{0.4}$$

where $k_i \geq 1$ for $1 \leq i \leq m$. The uniqueness refers to the positive integer m and the numbers $p_i \in \mathcal{P}$ and their associated powers k_i for $1 \leq i \leq m$. If we insist that the primes p_i are taken in increasing order, then the entire factorization (0.4) is uniquely determined by n. The proof of these assertions uses the Euclidean algorithm for division of positive integers and will not be repeated here.

Here is a formal argument that (0.3) holds:

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n^s} &= \sum_{n=1}^{\infty} \sum_{p_1^{k_1} p_2^{k_2} ... p_m^{k_m} = n} \frac{1}{\left\{p_1^{k_1} p_2^{k_2} ... p_m^{k_m}\right\}^s} \\
&= \sum_{m=1}^{\infty} \sum_{\{p_1,...p_m\}\subset\mathcal{P}} \sum_{\{k_1,...k_m\}\subset\mathbb{N}} \prod_{i=1}^{m} \frac{1}{p_i^{k_i s}}
\end{aligned}$$

where the second equality holds by the Fundamental Theorem of Arithmetic, and where the notation $\{p_1, ...p_m\}$ implies that the p_i are distinct. Continuing, we rearrange the terms in the infinite sum into an infinite product to obtain

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p\in\mathcal{P}} \left\{ \prod_{k\in\mathbb{N}} \frac{1}{p^{ks}} \right\} = \prod_{p\in\mathcal{P}} \frac{1}{1-\frac{1}{p^s}},$$

upon summing the geometric series $\sum_{k\in\mathbb{N}} \frac{1}{p^{ks}} = \frac{1}{1-\frac{1}{p^s}}$.

We can make this argument rigorous when $s > 1$ as follows: whenever $1 < N < M$ the Fundamental Theorem of Arithmetic shows that

$$\sum_{n=1}^{N} \frac{1}{n^s} \leq \prod_{p\leq N} \left(1 + \frac{1}{p^s} + ... + \frac{1}{p^{sM}}\right) \leq \prod_{p} \frac{1}{1-\frac{1}{p^s}},$$

and conversely,

$$\prod_{p\leq N} \left(1 + \frac{1}{p^s} + ... + \frac{1}{p^{sM}}\right) \leq \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

Taking limits appropriately in M and N gives (0.3) for $s > 1$.

Now we extend Euler's identity (0.3) to all complex numbers s in the half-plane $\Omega_1 = \{s \in \mathbb{C} : \operatorname{Re} s > 1\}$. For this we observe that the series on the left side of (0.3) is absolutely convergent for $s \in \Omega_1$, and converges uniformly on compact subsets of Ω_1. Hence by the Uniform Convergence Theorem 11, the series $\sum_{n=1}^{\infty} \frac{1}{n^s}$ defines a holomorphic function of s in the half-plane Ω_1. Moreover, we claim that the infinite product $\prod_p \frac{1}{1-\frac{1}{p^s}}$ on the right side of (0.3) is also uniformly convergent in s on compact subsets of Ω_1, and hence defines a holomorphic function of $s \in \Omega_1$. This requires just a small amount of additional work involving infinite products, to which we now turn.

If $0 \leq u_n < 1$ and $0 \leq v_n < \infty$ then

$$\begin{aligned} \prod_{n=1}^{\infty} (1-u_n) &> 0 \text{ if and only if } \sum_{n=1}^{\infty} u_n < \infty, \\ \prod_{n=1}^{\infty} (1+v_n) &< \infty \text{ if and only if } \sum_{n=1}^{\infty} v_n < \infty. \end{aligned} \tag{0.5}$$

To see (0.5) we may assume $0 \leq u_n, v_n \leq \frac{1}{2}$, so that $e^{-u_n} \geq 1 - u_n \geq e^{-2u_n}$ and $e^{\frac{1}{2}v_n} \leq 1 + v_n \leq e^{v_n}$. For example, when $0 \leq x \leq \frac{1}{2}$, the alternating series estimate yields

$$e^{-2x} \leq 1 - 2x + \frac{(2x)^2}{2!} \leq 1 - x,$$

while the geometric series estimate yields

$$e^{\frac{1}{2}x} \leq 1 + \left(\frac{1}{2}x\right)\left\{1 + x + x^2 + ...\right\} \leq 1 + x.$$

Thus we have

$$
\begin{aligned}
\exp\left(-\sum_{n=1}^{\infty} u_n\right) &\geq \prod_{n=1}^{\infty}(1-u_n) \geq \exp\left(-2\sum_{n=1}^{\infty} u_n\right), \\
\exp\left(\frac{1}{2}\sum_{n=1}^{\infty} v_n\right) &\leq \prod_{n=1}^{\infty}(1+v_n) \leq \exp\left(\sum_{n=1}^{\infty} v_n\right).
\end{aligned}
\tag{0.6}
$$

Now suppose that $u_n \in H(\Omega)$ and that $\sum_{n=1}^{\infty} u_n(z)$ converges absolutely and uniformly on compact subsets of Ω.

Claim 2. *The infinite product* $\prod_{n=1}^{\infty}(1-u_n(z))$ *converges uniformly on compact subsets of* Ω *to a holomorphic function* $f(z)$ *in* Ω. *Furthermore,* $f(z)=0$ *if and only if there is* $n \geq 1$ *such that* $u_n(z)=0$.

Indeed, if we expand products, cancel the 1's and take absolute values inside, we see that

$$
|p_{M,N}(z)-1| \leq p^*_{M,N}(z)-1, \qquad 1 \leq M \leq N,
\tag{0.7}
$$

where $p_{M,N}(z) = \prod_{n=M}^{N}(1-u_n(z))$ and $p^*_{M,N}(z) = \prod_{n=M}^{N}(1+|u_n(z)|)$. Alternatively, the case $N=M$ is obvious and the general case follows by induction from

$$
P_{M,N+1}-1 = P_{M,N}(1-u_{N+1}(z))-1 = (P_{M,N}-1)(1-u_{N+1}(z))-u_{N+1}(z),
$$

since then

$$
|P_{M,N+1}-1| \leq \left(P^*_{M,N}-1\right)(1+|u_{N+1}(z)|)+|u_{N+1}(z)| = P^*_{M,N+1}-1.
$$

It now follows from (0.7), the second line in (0.6), and the Uniform Convergence Theorem 11, that the infinite product $\prod_{n=1}^{\infty}(1-u_n(z))$ converges uniformly on compact subsets of Ω to a holomophic function $f \in H(\Omega)$:

$$
\begin{aligned}
\left|\prod_{n=1}^{N}(1-u_n(z)) - \prod_{n=1}^{M}(1-u_n(z))\right| &= \left|\left\{\prod_{n=1}^{M-1}(1-u_n(z))\right\}\{P_{M,N+1}-1\}\right| \\
&\leq \prod_{n=1}^{M-1}(1+|u_n(z)|)\ |P_{M,N+1}-1| \\
&\leq e^{\sum_{n=1}^{\infty}|u_n(z)|}\left(p^*_{M,N}(z)-1\right) \to 0
\end{aligned}
$$

as $N > M \to \infty$ uniformly on compact subsets of Ω.

From the inequality

$$
\left|\prod_{n=1}^{\infty}(1-u_n(z))\right| \geq \prod_{n=1}^{\infty}(1-|u_n(z)|),
$$

and the first line in (0.5), we see that $f(z)$ vanishes at z if and only if one of the functions $1-u_n(z)$ vanishes at z. Indeed, if $u_n(z) \neq 1$ for all n, then we can discard the finitely many n for which $|u_n(z)| \geq 1$ so that (0.5) applies.

If we apply the above Claim to the holomorphic functions

$$u_n(s) = 1 - \frac{1}{1 - \frac{1}{p_n^s}} = -\frac{1}{p_n^s - 1}, \qquad s \in \Omega_1,$$

where $\mathcal{P} = \{p_n\}_{n=1}^{\infty}$, we see that the infinite product on the right side of (0.3) converges uniformly on compact subsets of Ω_1 to a holomorphic function there. By the coincidence principle, the two sides of (0.3) coincide in Ω_1, so Euler's identity holds for all $s \in \Omega_1$.

Definition 8. *The Riemann zeta function is the holomorphic function*

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \qquad s \in \Omega_1.$$

Note that letting $s \to 1$ in (0.3) shows that

$$\infty = \sum_{n=1}^{\infty} \frac{1}{n} = \prod_p \frac{1}{1 - \frac{1}{p}},$$

so that $\prod_p \left(1 - \frac{1}{p}\right) = 0$. Then (0.5) yields $\sum_p \frac{1}{p} = \infty$, which quantifies the infinitude of primes, and begs the question of their density in the positive integers.

In the next section we will extend $\zeta(s)$ to a holomorphic function in $\mathbb{C} \setminus \{1\}$, and show that the extension has a simple pole at 1. Note that an extension of a holomorphic function f from a domain Ω to a larger domain is unique by the coincidence principle. Nevertheless, some caution must be exercised as different extensions of f need not coincide at a common point *outside* of Ω. Indeed, for any odd integer k there is an extension of $\log z = \ln |z| + i\theta$ from the ball $B(1,1)$ to a domain (that winds around the origin k times) containing -1 in which $\log(-1) = k\pi i$. As a result, these extensions are usually referred to as *an analytic continuation* of f to a larger set.

1. Analytic continuation of $\zeta(s)$

We begin by comparing the series $\sum_{n=1}^{\infty} \frac{1}{n^s}$ for $\zeta(s)$ to its corresponding integral $\int_1^{\infty} \frac{dx}{x^s}$. Now the integral has a holomorphic continuation to $\mathbb{C} \setminus \{1\}$ given by $\frac{1}{s-1}$ since for $\operatorname{Re} s > 1$,

$$\int_1^{\infty} \frac{dx}{x^s} = \frac{1}{s-1}.$$

We will show that the difference between the series and the integral admits a holomorphic continuation to the larger half-plane $\Omega_0 = \{s \in \mathbb{C} : \operatorname{Re} s > 0\}$. This will then give a holomorphic continuation of $\zeta(s)$ to $\Omega_0 \setminus \{1\}$ with a simple pole at 1.

Lemma 2. *There is a sequence of* entire *functions* $\{h_n(s)\}_{n=1}^{\infty}$ *which decrease to zero at the rate*

$$|h_n(s)| \leq \frac{|s|}{n^{\sigma+1}}, \qquad s = \sigma + i\tau \in \mathbb{C}, \tag{1.1}$$

and satisfy the equation

$$\sum_{n=1}^{N-1} \frac{1}{n^s} - \int_1^N \frac{dx}{x^s} = \sum_{n=1}^{N-1} h_n(s), \qquad s \in \mathbb{C}, \tag{1.2}$$

whenever $N>1$ is an integer. In particular, the rate of decay in (1.1) shows that the right side of the equation

$$\zeta(s)=\frac{1}{s-1}+\sum_{n=1}^{\infty}h_n(s), \qquad s\in\Omega_1, \tag{1.3}$$

defines a holomorphic function for $s\in\Omega_0\backslash\{1\}$ since the series on the right converges uniformly on compact subsets of Ω_0.

Proof: Define

$$h_n(s)=\frac{1}{n^s}-\int_n^{n+1}\frac{dx}{x^s}=\int_n^{n+1}\left(\frac{1}{n^s}-\frac{1}{x^s}\right)dx.$$

By the submean value theorem applied on the interval $[n,x]$, we have

$$\left|\frac{1}{n^s}-\frac{1}{x^s}\right|\le\left|\frac{-s}{c_n^{s+1}}\right|\le\frac{|s|}{n^{\sigma+1}},$$

for some $c_n\in(n,x)$. Estimate (1.1) and equation (1.2) now follow.

For $s\in\Omega_1$, the integral $\int_1^N\frac{dx}{x^s}$ converges to $\frac{1}{s-1}$ so that (1.3) holds. The estimate (1.1) implies that the series $\sum_{n=1}^{\infty}h_n(s)$ on the right side of (1.3) converges uniformly on compact subsets of Ω_0.

We can continue this process and obtain an analytic continuation of $\zeta(s)$ to $\Omega_{-1}\setminus\{1\}$, then to $\Omega_{-2}\setminus\{1\}$, etc. until we have extended $\zeta(s)$ to all of $\mathbb{C}\setminus\{1\}$. We emphasize that for the proof of the Prime Number Theorem we only need the analytic continuation of $\zeta(s)$ to $\Omega_0\setminus\{1\}$ given in Lemma 2. However, for the sake of completeness, we sketch the details of the continuation to all of $\mathbb{C}\setminus\{1\}$.

In order to continue $\zeta(s)$ to $\Omega_{-1}\setminus\{1\}$, we define $\{x\}=x-n$ for $n\le x<n+1$ to be the fractional part of x, and begin by computing for $\operatorname{Re}s>1$,

$$\begin{aligned}
s\int_1^{\infty}\frac{\{x\}}{x^{s+1}}dx &= s\sum_{n=1}^{\infty}\int_n^{n+1}\frac{x-n}{x^{s+1}}dx=s\sum_{n=1}^{\infty}\left\{\int_n^{n+1}x^{-s}dx-n\int_n^{n+1}x^{-s-1}dx\right\}\\
&= \sum_{n=1}^{\infty}\left\{\frac{s}{1-s}\left[(n+1)^{1-s}-n^{1-s}\right]+n\left[(n+1)^{-s}-n^{-s}\right]\right\}\\
&= \frac{s}{s-1}\left[2^{-s}-1^{-s}\right]+2\left[3^{-s}-2^{-s}\right]+3\left[4^{-s}-3^{-s}\right]+...\\
&= \frac{s}{s-1}-\zeta(s).
\end{aligned}$$

Now each of the functions $\zeta(s)$ and $\frac{s}{s-1}-s\int_1^{\infty}\frac{\{x\}}{x^{s+1}}dx$ is holomorphic in $\Omega_0\setminus\{1\}$ and so the coincidence principle implies

$$\zeta(s)=\frac{s}{s-1}-s\int_1^{\infty}\frac{\{x\}}{x^{s+1}}dx, \qquad s\in\Omega_0\setminus\{1\}.$$

With $Q(x)=\{x\}-\frac{1}{2}$ the above identity becomes

$$\zeta(s)=\frac{s}{s-1}-\frac{1}{2}-s\int_1^{\infty}\frac{Q(x)}{x^{s+1}}dx, \qquad s\in\Omega_0\setminus\{1\}.$$

Note that $Q(x)$ is periodic on the real line with period 1 and $\int_0^1 Q(x)\,dx=0$. Thus any antiderivative of $Q(x)$ on the real line will also have period 1.

Now we set $Q_0(x) = Q(x)$ and define $Q_1(x)$ on the real line by

$$Q_1'(x) = Q_0(x) \text{ and } \int_0^1 Q_1(x)\,dx = 0.$$

Thus Q_1 is periodic on the real line with period 1, $Q_1(x) = \frac{1}{2}x^2 - \frac{1}{2}x + \frac{1}{12}$ for $0 \le x < 1$, Q_1 is continuous on $[1,\infty)$, and Q_1 is continuously differentiable on $[1,\infty) \setminus \mathbb{N}$. We can now integrate by parts to obtain for $s \in \Omega_0 \setminus \{1\}$,

$$\begin{aligned}\int_1^\infty \frac{Q_1'(x)}{x^{s+1}}dx &= \frac{Q_1(x)}{x^{s+1}}\Big|_1^\infty - (-s-1)\int_1^\infty \frac{Q_1(x)}{x^{s+2}}dx \\ &= -\frac{1}{12} + (s+1)\int_1^\infty \frac{Q_1(x)}{x^{s+2}}dx.\end{aligned}$$

Thus for $s \in \Omega_0 \setminus \{1\}$ we have

$$\zeta(s) = \frac{s}{s-1} - \frac{1}{2} - s\left\{-\frac{1}{12} + (s+1)\int_1^\infty \frac{Q_1(x)}{x^{s+2}}dx\right\}.$$

The right side defines a holomorphic function of s in $\Omega_{-1} \setminus \{1\}$ and provides an analytic continuation of $\zeta(s)$ to $\Omega_{-1} \setminus \{1\}$.

Now we recursively define $Q_k(x)$ on the real line for all $k \ge 1$ by

$$Q_k'(x) = Q_{k-1}(x) \text{ and } \int_0^1 Q_k(x)\,dx = 0.$$

Then by induction Q_k is periodic on the real line with period 1, Q_k is $k-1$ times continuously differentiable on $[1,\infty)$, and Q_k is k times continuously differentiable on $[1,\infty) \setminus \mathbb{N}$. Thus we can integrate by parts k times in

$$\zeta(s) = \frac{s}{s-1} - \frac{1}{2} - s\int_1^\infty \frac{\frac{d^k}{dx^k}Q_k(x)}{x^{s+1}}dx, \qquad s \in \Omega_0 \setminus \{1\},$$

(the case $k = 1$ was done above), and the resulting right hand side defines an analytic continuation of $\zeta(s)$ to $\Omega_{-k} \setminus \{1\}$. Indeed, the boundary terms lead to a polynomial of degree k in s, while the integral remaining is

$$-s(s+1)\dots(s+k)\int_1^\infty \frac{Q_k(x)}{x^{s+k+1}}dx.$$

1.1. Nonvanishing of the zeta function. The key to the proof we give of the Prime Number Theorem is the nonvanishing of the zeta function $\zeta(s)$ on the line $\operatorname{Re} s = 1$. For this purpose it will be convenient to define $\log\zeta(s)$ in the simply connected domain Ω_1. This is possible by Corollary 8 since Euler's identity (0.3) and Claim 2 show that $\zeta(s)$ is nonvanishing in Ω_1. Indeed, Claim 2 applies since $1 - u_p(z) = \frac{1}{1-\frac{1}{p^s}}$ implies $u_p(z) = \frac{1}{1-p^s}$ where $\sum_p \frac{1}{1-p^s}$ converges absolutely and uniformly on compact subsets of Ω_1. Among the possible choices for the function $\log\zeta(s)$ we fix the one satisfying

$$\log\zeta(s) = \ln\zeta(s) > 0, \qquad s > 1.$$

THEOREM 17. *The zeta function $\zeta(s)$ has no zeroes on the line* $\operatorname{Re} s = 1$.

Proof: Using the power series expansion for logarithm on the unit interval,

$$\ln\left(\frac{1}{1-x}\right) = \sum_{m=1}^\infty \frac{x^m}{m}, \qquad 0 \le x < 1,$$

together with Euler's formula (0.3), we have when $s > 1$:

$$\ln \zeta(s) = \ln \prod_p \frac{1}{1-\frac{1}{p^s}} = \sum_p \ln\left(\frac{1}{1-\frac{1}{p^s}}\right) = \sum_p \sum_{m=1}^{\infty} \frac{p^{-sm}}{m},$$

where the double series is absolutely convergent. In fact the double series is absolutely convergent for $\operatorname{Re} s > 1$, and hence defines a holomorphic function in Ω_1. The coincidence principle now gives the following useful formula (which we will have occasion to use later),

$$\log \zeta(s) = \sum_{p,m} \frac{p^{-sm}}{m}, \qquad s \in \Omega_1. \tag{1.4}$$

With $c_n = \begin{cases} \frac{1}{m} & \text{if} \quad n = p^m \\ 0 & \text{if} \quad \text{not} \end{cases}$, we can rewrite this formula as

$$\log \zeta(s) = \sum_{n=1}^{\infty} c_n \frac{1}{n^s}, \qquad s \in \Omega_1.$$

We now invoke the identity

$$3 + 4\cos\theta + \cos 2\theta = 2(1+\cos\theta)^2 \geq 0, \tag{1.5}$$

to conclude that for $s = \sigma + i\tau \in \Omega_1$, i.e. $\sigma > 1$ and $\tau \in \mathbb{R}$,

$$\ln \left| \zeta(\sigma)^3 \zeta(\sigma+i\tau)^4 \zeta(\sigma+i2\tau) \right| \geq 0. \tag{1.6}$$

Indeed, from

$$\operatorname{Re}\left(\frac{1}{n^s}\right) = \operatorname{Re}\left(e^{-(\sigma+i\tau)\ln n}\right) = \frac{1}{n^\sigma}\cos(\tau \ln n),$$

we obtain

$$\begin{aligned} & \ln \left| \zeta(\sigma)^3 \zeta(\sigma+i\tau)^4 \zeta(\sigma+i2\tau) \right| \\ = \; & 3\ln|\zeta(\sigma)| + 4\ln|\zeta(\sigma+i\tau)| + \ln|\zeta(\sigma+i2\tau)| \\ = \; & 3\operatorname{Re}\log\zeta(\sigma) + 4\operatorname{Re}\log\zeta(\sigma+i\tau) + \operatorname{Re}\log\zeta(\sigma+i2\tau) \\ = \; & \sum_{n=1}^{\infty} c_n \frac{1}{n^\sigma}\{3 + 4\cos(\tau\ln n) + \cos(2\tau\ln n)\}, \end{aligned}$$

which is nonnegative by (1.5) and $c_n \geq 0$.

We can now use (1.6) to derive a contradiction from the assumption that $\zeta(1+i\tau) = 0$ for some $\tau \in \mathbb{R}$. Indeed, since ζ has a *simple* pole at 1, we have

$$\left|\zeta(\sigma)^3\right| \leq C(\sigma-1)^{-3}, \qquad \sigma \in (1,2).$$

Also, $\tau \neq 0$ and ζ must have a zero of order at least one at $1+i\tau$, so that

$$\left|\zeta(\sigma+i\tau)^4\right| \leq C(\sigma-1)^4, \qquad \sigma \in (1,2).$$

Finally, since ζ is holomorphic in a neighbourhood of the segment $\{\sigma + i\tau : 1 \leq \sigma \leq 2\}$, we have

$$|\zeta(\sigma+i2\tau)| \leq C, \qquad \sigma \in (1,2).$$

Thus altogether we obtain

$$\left|\zeta(\sigma)^3 \zeta(\sigma+i\tau)^4 \zeta(\sigma+i2\tau)\right| \leq C(\sigma-1), \qquad \sigma \in (1,2),$$

which contradicts (1.6) when $C(\sigma - 1) < 1$.

2. The residue theorem

Definition 9. *A function f is said to be* meromorphic *in an open set Ω if there is a subset A of Ω such that*

(1) *A has no limit point in Ω,*
(2) *f is holomorphic in $\Omega \setminus A$,*
(3) *f has a pole at each point of A.*

Recall that if f has a pole of order $N \in \mathbb{N}$ at a, then there is $R > 0$ so that for $z \in B(a, R)$,

$$
\begin{aligned}
f(z) &= \sum_{n=-N}^{\infty} b_n (z-a)^n \\
&= \left\{ \frac{b_{-N}}{(z-a)^N} + \dots + \frac{b_{-1}}{z-a} \right\} + \sum_{n=0}^{\infty} b_n (z-a)^n \\
&= P_N(z) + h(z),
\end{aligned} \tag{2.1}
$$

where $h \in H(B(a,R))$, For $0 < r < R$ we compute using Cauchy's theorem on h,

$$
\begin{aligned}
\int_{\partial B(a,r)} f(z)\, dz &= \int_{\partial B(a,r)} P_N(z)\, dz + \int_{\partial B(a,R)} h(z)\, dz \\
&= \int_{\partial B(a,r)} \frac{b_{-N}}{(z-a)^N} dz + \dots + \int_{\partial B(a,r)} \frac{b_{-1}}{z-a} dz \\
&= \int_{\partial B(a,r)} \frac{b_{-1}}{z-a} dz = 2\pi i b_{-1},
\end{aligned}
$$

where we have used the fact that the function $\frac{1}{(z-a)^k}$ has antiderivative $\frac{1}{1-k} \frac{1}{(z-a)^{k-1}}$ in the punctured disk $B'(a,R)$ to conclude that $\int_{\partial B(a,r)} \frac{1}{(z-a)^k} dz = 0$ for $2 \le k \le N$. We suspect that this formula will persist with $\partial B(a,r)$ replaced by any appropriate simple path γ that contains a in its interior. This justifies the following definition.

Definition 10. *If f has a pole at a with principal part*

$$
P_N(z) = \frac{b_{-N}}{(z-a)^N} + \dots + \frac{b_{-1}}{z-a},
$$

we define the residue of f at a to be

$$
Res\ (f; a) \equiv b_{-1}.
$$

Theorem 18. *Suppose that f is meromorphic in a simply connected domain Ω, and that the set A of poles of f in Ω is finite. Then if γ is a closed path in $\Omega \setminus A$, we have*

$$
\frac{1}{2\pi i} \int_{\gamma} f(z)\, dz = \sum_{a \in A} Res\ (f; a)\ Ind_{\gamma}(a). \tag{2.2}
$$

Proof: Let $P_a(z)$ be the principal part of f at the pole a, and set

$$g(z) = f(z) - \sum_{a \in A} P_a(z).$$

Then g has a removable singularity at each point $a \in A$, and so $g \in H(\Omega)$. Since

$$\frac{1}{2\pi i} \int_\gamma P_a(z)\, dz = \frac{1}{2\pi i} \int_\gamma \frac{Res\ (f;a)}{z-a} dz = Res\ (f;a)\ Ind_\gamma(a),$$

we have

$$\begin{aligned} 0 &= \frac{1}{2\pi i} \int_\gamma g(z)\, dz = \frac{1}{2\pi i} \int_\gamma f(z)\, dz - \sum_{a \in A} \frac{1}{2\pi i} \int_\gamma P_a(z)\, dz \\ &= \frac{1}{2\pi i} \int_\gamma f(z)\, dz - \sum_{a \in A} Res\ (f;a)\ Ind_\gamma(a). \end{aligned}$$

The identity (2.2) in the residue theorem can be extended to arbitrary meromorphic functions f in an open set Ω (it turns out that the index $Ind_\gamma(a)$ is nonzero for only finitely many poles), but we will not pursue this here as it is not needed in the sequel.

2.1. Calculation of residues. If f has a simple pole at a, there is a useful formula for computing the residue that is immediate from an inspection of the case $N = 1$ of (2.1):

$$Res\,(f;a) = \lim_{z \to a} (z-a) f(z). \tag{2.3}$$

More generally, there is an analogous formula for the residue at a pole of order N,

$$Res\,(f;a) = \lim_{z \to a} \frac{1}{(N-1)!} \frac{d^{N-1}}{dz^{N-1}} \left\{ (z-a)^N f(z) \right\}, \tag{2.4}$$

that follows from multiplying (2.1) by $(z-a)^N$ and then using the formula from Theorem 5 to compute the $(N-1)^{st}$ coefficient of the holomorphic function $g(z) = (z-a)^N f(z)$. Note that if we *overestimate* the order of the pole f has at a point a, formula (2.4) still gives the correct answer, i.e. (2.4) holds provided f has a pole of order at most N at a. On the other hand, if we *underestimate* the order of the pole f has at a point a, the right side of formula (2.4) is infinite, thereby alerting us to our error.

An observation that is useful in a very special case is this: if f is meromorphic in a domain Ω that is symmetric about the real axis $\mathbb{R}$, and if $f(z)$ is real on $\Omega \cap \mathbb{R}$, then f has a pole of order N at $a \in \Omega$ if and only if f has a pole of order N at $\overline{a}$, and moreover the residues are conjugate:

$$Res\,(f;\overline{a}) = \overline{Res\,(f;a)}.$$

This follows from the Schwarz reflection principle: the function $h(z) = f(z) - \overline{f(\overline{z})}$ is holomorphic in Ω and vanishes on $\Omega \cap \mathbb{R}$, hence vanishes in Ω by the coincidence principle.

2.2. Counting zeroes and poles. If f has a zero of order N at a then $\frac{f'(z)}{f(z)}$ has a simple pole at a with

$$Res\left(\frac{f'}{f};a\right)=N. \tag{2.5}$$

Indeed, if $f(z)=(z-a)^N h(z)$ with $h(a)\neq 0$, then for $z\in B(a,r)$ with r small,

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

Similarly, if f has a pole of order N at a then $\frac{f'(z)}{f(z)}$ has a simple pole at a with

$$Res\left(\frac{f'}{f};a\right)=-N. \tag{2.6}$$

Indeed, if $f(z)=(z-a)^{-N} h(z)$ with $h(a)\neq 0$, then for $z\in B(a,r)$ with r small,

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

Of course, if f has neither a zero nor a pole at a, then $\frac{f'}{f}$ is holomorphic in a neighbourhood of a. It now follows easily from the residue theorem that $\frac{1}{2\pi i}\int_\gamma \frac{f'(z)}{f(z)}dz$ counts the number of zeroes minus the number of poles inside γ. Here is a precise statement that is not the most general possible.

Theorem 19. *Suppose that f is meromorphic in a simply connected domain Ω, and that both the set P of poles of f in Ω is finite, and the set Z of zeroes of f in Ω is finite. Then if γ is a* simple *closed path in $\Omega\setminus(Z\cup P)$, we have*

$$\begin{aligned}\frac{1}{2\pi i}\int_\gamma \frac{f'(z)}{f(z)}dz &= \sum_{z\in Z \text{ that are inside } \gamma} \text{order of the zero at } z \\ &- \sum_{p\in P \text{ that are inside } \gamma} \text{order of the pole at } p.\end{aligned} \tag{2.7}$$

Proof: Let $A=Z\cup P$. The residue theorem gives

$$\frac{1}{2\pi i}\int_\gamma \frac{f'(z)}{f(z)}dz=\sum_{a\in A} Res\left(\frac{f'}{f};a\right) Ind_\gamma(a),$$

and then the calculations above give (2.7).

3. Proof of the Prime Number Theorem

Now we can complete the proof of the Prime Number Theorem:

$$\lim_{x\to\infty}\frac{\pi(x)}{\frac{x}{\ln x}}=1.$$

In order to prove that $\pi(x)\sim\frac{x}{\ln x}$, it turns out to be more convenient to consider the asymptotic form $\pi(x)\ln x\sim x$. Now

$$\pi(x)\ln x=\sum_{p\le x}\ln x=\sum_{p\le x}\frac{\ln x}{\ln p}\ln p,$$

which is very close to *Tchebychev's ψ function*:

$$\psi(x) \equiv \sum_{p\leq x}\left[\frac{\ln x}{\ln p}\right]\ln p = \sum_{p^m\leq x}\ln p = \sum_{1\leq n\leq x}\Lambda(n), \tag{3.1}$$

where

$$\Lambda(n) = \begin{cases} \ln p & \text{if } n = p^m \text{ for some prime } p \text{ and some } m \geq 1 \\ 0 & \text{otherwise} \end{cases}. \tag{3.2}$$

The Prime Number Theorem is reduced to the following asymptotic estimate for ψ.

PROPOSITION 10. *If* $\lim_{x\to\infty}\frac{\psi(x)}{x} = 1$, *then* $\lim_{x\to\infty}\frac{\pi(x)}{\frac{x}{\ln x}} = 1$.

Proof: We clearly have

$$\frac{\psi(x)}{x} = \frac{\sum_{p\leq x}\left[\frac{\ln x}{\ln p}\right]\ln p}{x} \leq \frac{\sum_{p\leq x}\frac{\ln x}{\ln p}\ln p}{x} = \frac{\pi(x)}{\frac{x}{\ln x}},$$

which gives

$$1 \leq \liminf_{x\to\infty}\frac{\pi(x)}{\frac{x}{\ln x}}.$$

Conversely, fix $0<\alpha<1$. Then

$$\psi(x) \geq \sum_{x^\alpha<p\leq x}\ln p \geq \{\pi(x) - \pi(x^\alpha)\}\ln x^\alpha,$$

and so

$$\alpha\frac{\pi(x)}{\frac{x}{\ln x}} \leq \alpha\frac{\pi(x^\alpha)\ln x}{x} + \frac{\psi(x)}{x} \leq \alpha\frac{x^\alpha\ln x}{x} + \frac{\psi(x)}{x}.$$

Taking the limit superior as $x\to\infty$ we obtain

$$\alpha\limsup_{x\to\infty}\frac{\pi(x)}{\frac{x}{\ln x}} \leq \alpha\limsup_{x\to\infty}\frac{\ln x}{x^{1-\alpha}} + 1 = 1.$$

Since $0<\alpha<1$ is arbitrary, we conclude that $\limsup_{x\to\infty}\frac{\pi(x)}{\frac{x}{\ln x}} \leq 1$, and hence $\lim_{x\to\infty}\frac{\pi(x)}{\frac{x}{\ln x}} = 1$.

We expect that for suitable functions ψ, $\psi(x)\sim x$ if and only if $\int_1^x\psi(t)\,dt \sim \int_1^x t\,dt \sim \frac{x^2}{2}$. So we define

$$\psi_1(x) = \int_1^x \psi(t)\,dt, \qquad x>1,$$

where ψ is Tchebychev's ψ function.

PROPOSITION 11. *If* $\lim_{x\to\infty}\frac{\psi_1(x)}{\frac{x^2}{2}} = 1$, *then* $\lim_{x\to\infty}\frac{\psi(x)}{x} = 1$.

Proof: Since ψ is increasing we have for $0<\alpha<1<\beta<\infty$,

$$\frac{\int_{\alpha x}^x\psi(t)\,dt}{\int_{\alpha x}^x dt} \leq \psi(x) \leq \frac{\int_x^{\beta x}\psi(t)\,dt}{\int_x^{\beta x}dt}. \tag{3.3}$$

The second inequality in (3.3) yields

$$\frac{\psi(x)}{x} \leq \frac{1}{x} \frac{1}{(\beta-1)x} \{\psi_1(\beta x) - \psi_1(x)\} = \frac{1}{(\beta-1)} \left\{ \beta^2 \frac{\psi_1(\beta x)}{(\beta x)^2} - \frac{\psi_1(x)}{x^2} \right\},$$

and taking limit superior as $x \to \infty$ we obtain

$$\limsup_{x\to\infty} \frac{\psi(x)}{x} \leq \frac{1}{(\beta-1)} \left\{ \frac{1}{2}\beta^2 - \frac{1}{2} \right\} = \frac{\beta+1}{2}.$$

Letting $\beta \to 1$ we get $\limsup_{x\to\infty} \frac{\psi(x)}{x} \leq 1$. Arguing in similar fashion with the first inequality in (3.3), we get $1 \leq \liminf_{x\to\infty} \frac{\psi(x)}{x}$, and hence $\lim_{x\to\infty} \frac{\psi(x)}{x} = 1$.

The connection between $\psi_1(x)$ and $\zeta(s)$ is given in the following identity, which will ultimately yield the asymptotic $\psi_1(x) \sim \frac{x^2}{2}$.

PROPOSITION 12. *For all $c > 1$,*

$$\psi_1(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{x^{s+1}}{s(s+1)} \left(-\frac{\zeta'(s)}{\zeta(s)} \right) ds, \tag{3.4}$$

where $\int_{c-i\infty}^{c+i\infty}$ denotes integration along the vertical line $\mathrm{Re}\, s = c$ in the upward direction.

Proof: We first claim that the integral on the right side of (3.4) is absolutely convergent. Indeed, if we differentiate (1.4),

$$\log \zeta(s) = \sum_{p,m} \frac{p^{-sm}}{m}, \qquad s \in \Omega_1,$$

we obtain

$$\frac{\zeta'(s)}{\zeta(s)} = -\sum_{p,m} (\ln p)\, p^{-sm} = -\sum_{n=1}^{\infty} \Lambda(n) \frac{1}{n^s}, \qquad \mathrm{Re}\, s > 1, \tag{3.5}$$

where $\Lambda(n)$ is defined in (3.2). In particular this yields

$$\left| \frac{\zeta'(s)}{\zeta(s)} \right| \leq \sum_{n=1}^{\infty} \frac{\ln n}{n^{\mathrm{Re}\, s}} < \infty, \qquad s \in \Omega_1.$$

Since we also have

$$\left| \frac{x^{s+1}}{s(s+1)} \right| \leq \frac{x^{c+1}}{(1+|\tau|)^2}, \qquad s = c + i\tau \in \Omega_1,$$

we see that the modulus of the integrand on the right side of (3.4) is dominated by $C(1+|\tau|)^{-2}$, hence integrable.

Next we claim that the following identity holds:

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{y^s}{s(s+1)} ds = \left(1 - \frac{1}{y}\right)_+ = \begin{cases} 0 & \text{if} \quad 0 < y \leq 1 \\ 1 - \frac{1}{y} & \text{if} \quad 1 \leq y < \infty \end{cases}. \tag{3.6}$$

Note that the integral on the right side of (3.6) is absolutely convergent since $|y^s| = y^c$ is constant on the path of integration. Assuming (3.6) we can quickly complete the proof of Proposition 12. Using (3.1) we have

$$\begin{aligned}\psi_1(x) &= \int_1^x \psi(t)\,dt = \int_1^x \left(\sum_{1\le n\le t}\Lambda(n)\right)dt \\ &= \sum_{1\le n\le x}\Lambda(n)(x-n) = x\sum_{n=1}^{\infty}\Lambda(n)\left(1-\frac{n}{x}\right)_+,\end{aligned}$$

and so using (3.6) with $y=\frac{x}{n}$ we conclude that

$$\begin{aligned}\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)ds &= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{x^{s+1}}{s(s+1)}\left(\sum_{n=1}^{\infty}\Lambda(n)\frac{1}{n^s}\right)ds \\ &= \sum_{n=1}^{\infty}\Lambda(n)\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{x\left(\frac{x}{n}\right)^s}{s(s+1)}ds \\ &= x\sum_{n=1}^{\infty}\Lambda(n)\left(1-\frac{n}{x}\right)_+ = \psi_1(x).\end{aligned}$$

Finally we prove the identity (3.6) using the residue theorem. Suppose first that $1\le y<\infty$. For $R>0$, denote by $[c-iR,c+iR]$ the line segment joining $c-iR$ to $c+iR$ that is directed upward. Also denote by $C_{left}(R)$ the half circle of radius R that joins $c+iR$ to $c-iR$ by travelling down the *left* half of the circle centered at c with radius R. Now define

$$f(s)=\frac{y^s}{s(s+1)}=\frac{e^{s\ln y}}{s(s+1)},\qquad s\in\mathbb{C}.$$

Note that f is meromorphic in the plane with simple poles at 0 and -1. If R is chosen large enough that these two poles lie inside the closed path $[c-iR,c+iR]\cup C_{left}(R)$, then the residue theorem gives

$$\begin{aligned}&\frac{1}{2\pi i}\int_{[c-iR,c+iR]}f(s)\,ds+\frac{1}{2\pi i}\int_{C_{left}(R)}f(s)\,ds \\ &\quad= Res\ (f;0)+Res\ (f;-1) \\ &\quad= \frac{y^0}{(0+1)}+\frac{y^{-1}}{(-1)}=1-\frac{1}{y}.\end{aligned}$$

Now let $R\to\infty$ to obtain

$$\lim_{R\to\infty}\frac{1}{2\pi i}\int_{C_{left}(R)}f(s)\,ds=0,$$

and hence

$$\lim_{R\to\infty}\frac{1}{2\pi i}\int_{[c-iR,c+iR]}f(s)\,ds=1-\frac{1}{y}.$$

Indeed, for $R>2(c+1)$ and $\frac{\pi}{2}\le\theta\le\frac{3\pi}{2}$,

$$\left|f\left(c+Re^{i\theta}\right)\right|\le\frac{e^{(c+R\cos\theta)\ln y}}{(R-c)(R-c-1)}\le\frac{4e^{c\ln y}}{R^2}$$

since $\cos\theta \leq 0$ and $\ln y \geq 0$. Now the length of the path $C_{left}(R)$ is πR and so we obtain

$$\left|\int_{C_{left}(R)} f(s)\,ds\right| \leq \pi R \frac{4e^{c\ln y}}{R^2} \to 0 \text{ as } R \to \infty.$$

To prove (3.6) when $0 < y \leq 1$, we consider instead the closed path $[c - iR, c + iR] \cup C_{right}(R)$ and apply Cauchy's theorem (which is just the residue theorem when there are no poles inside the path) to the holomorphic function $f(s)$ in the half-plane $\operatorname{Re} s > 1$. We obtain

$$\frac{1}{2\pi i}\int_{[c-iR,c+iR]} f(s)\,ds + \frac{1}{2\pi i}\int_{C_{right}(R)} f(s)\,ds = 0,$$

and then use the argument above with $\cos\theta \geq 0$ and $\ln y \leq 0$ to show that

$$\lim_{R\to\infty} \frac{1}{2\pi i}\int_{C_{right}(R)} f(s)\,ds = 0.$$

This completes the proof of (3.6), and hence that of Proposition 12.

3.1. Proof of the asymptotic for $\psi_1(x)$. Here we prove the asymptotic formula

$$\lim_{x\to\infty} \frac{\psi_1(x)}{\frac{1}{2}x^2} = 1 \tag{3.7}$$

in three steps. Propositions 10 and 11 then complete the proof of the Prime Number Theorem (0.2).

We begin with the identity

$$\psi_1(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right) ds \tag{3.8}$$

in Proposition 12, and try to move the path of integration to $[1 - i\infty, 1 + i\infty]$ where $\left|x^{s+1}\right| = x^2$ is at least comparable to the asymptotic estimate $\frac{1}{2}x^2$. Of course this is impossible since $\zeta(s)$ has a pole at 1, but we can come close enough using the residue theorem. Let $R \geq 3$ and $c > 1$. For $S > R$ and $0 < \eta < 1$, define the closed taxicab path $\gamma_{R,S,c,\eta}$ to consist of the eight segments

$$\begin{gathered}[c - iS, c + iS], [c + iS, 1 + iS], [1 + iS, 1 + iR], [1 + iR, 1 - \eta + iR], \\ [1 - \eta + iR, 1 - \eta - iR], [1 - \eta - iR, 1 - iR], [1 - iR, 1 - iS], [1 - iS, c - iS]\end{gathered}$$

concatenated with the directions and in the sequence given. Since $\zeta(s)$ has a pole at 1 and is nonvanishing on the rest of the line $\operatorname{Re} s = 1$, there exists $0 < \eta(R) < 1$ depending on R such that $\zeta(s)$ has no zero on or inside the rectangle with sides

$$[1 - \eta - iR, 1 - iR],\ [1 - iR, 1 + iR],\ [1 + iR, 1 - \eta + iR],\ [1 - \eta + iR, 1 - \eta - iR].$$

We will always assume η is so small that $0 < \eta < \eta(R)$.

Step 1: For $S > R \geq 3$ and $0 < \eta < \eta(R) < 1 < c$ we have

$$\frac{1}{2\pi i}\int_{\gamma_{R,S,c,\eta}} \frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right) ds = \frac{x^2}{2}. \tag{3.9}$$

Since the integrand

$$g_x(s) = \frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right) \tag{3.10}$$

is meromorphic in Ω_0 with a single pole at $s=1$ inside the simple path $\gamma_{R,S,c,\eta}$, the residue theorem implies

$$\frac{1}{2\pi i}\int_{\gamma_{R,S,c,\eta}} g_x(s)\,ds = Res\ (g_x;1)\ Ind_{\gamma_{R,S,c,\eta}}(1) = \lim_{s\to1}(s-1)\,g_x(s).$$

Now by Lemma 2, $\zeta(s) = \frac{1}{s-1} + h(s)$, $h\in H(\Omega_0)$, and so

$$\zeta'(s) = -\frac{1}{(s-1)^2} + h'(s),$$

hence by (2.3),

$$\begin{aligned}
(s-1)\,g_x(s) &= (s-1)\frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)\\
&= (s-1)\frac{x^{s+1}}{s(s+1)}\left(-\frac{-\frac{1}{(s-1)^2}+h'(s)}{\frac{1}{s-1}+h(s)}\right)\\
&= \frac{x^{s+1}}{s(s+1)}\left(\frac{1+(s-1)h'(s)}{1+(s-1)h(s)}\right)\to\frac{x^2}{2}\ \text{as } s\to1.
\end{aligned}$$

Alternatively, (2.6) shows that

$$Res\left(\frac{\zeta'(s)}{\zeta(s)};1\right) = -1,$$

and then as above, (2.3) implies that

$$Res\ \left(\frac{x^{s+1}}{s(s+1)}\frac{\zeta'(s)}{\zeta(s)};1\right) = \left(\frac{x^{s+1}}{s(s+1)}\ |_{s=1}\right)(-1) = -\frac{x^2}{2}.$$

Let $\beta_{R,\eta}$ be the *infinite* taxicab path consisting of the five segments

$$\begin{aligned}
&[1-i\infty,1-iR],[1-iR,1-\eta-iR],[1-\eta-iR,1-\eta+iR],\\
&[1-\eta+iR,1+iR],[1+iR,1+i\infty],
\end{aligned}$$

concatenated with the directions and in the sequence given. Thus $\beta_{R,\eta}$ is the line $[1-i\infty,1+i\infty]$ but with a rectangular jog to the left around the pole at 1.

Step 2: For $R\geq3$ and $0<\eta<\eta(R)<1<c$ we have,

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)ds = \frac{x^2}{2}+\frac{1}{2\pi i}\int_{\beta_{R,\eta}}\frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)ds.$$

Let S tend to infinity in (3.9). With g_x as in (3.10) we claim that the integrals over the top and bottom horizontal segments,

$$\int_{[c+iS,1+iS]} g_x(s)\,ds \text{ and } \int_{[1-iS,c-iS]} g_x(s)\,ds, \tag{3.11}$$

each tend to zero as $S\to\infty$. For this we will need estimates on the growth of the function $\tau\to\left|\frac{\zeta'(1+i\tau)}{\zeta(1+i\tau)}\right|$. Unfortunately the formula (3.5) for $\frac{\zeta'(s)}{\zeta(s)}$ is not much help in this regard since the formula doesn't take into account cancellations as s nears 1.

Instead we will estimate $\left|\zeta'(1+i\tau)\right|$ and $\left|\frac{1}{\zeta(1+i\tau)}\right|$ separately using Lemma 2 and a refinement of the argument in the proof of Theorem 17.

In order to estimate $\left|\zeta'(1+i\tau)\right|$, recall from Lemma 2 that

$$\zeta(s)=\frac{1}{s-1}+\sum_{n=1}^{\infty}h_n(s), \qquad s\in\Omega_1,$$

where the entire function

$$h_n(s)=\int_n^{n+1}\left(\frac{1}{n^s}-\frac{1}{x^s}\right)dx$$

satisfies (1.1),

$$|h_n(s)|\leq\frac{|s|}{n^{\sigma+1}}, \qquad s=\sigma+i\tau\in\mathbb{C},$$

as well as the crude estimate

$$|h_n(s)|\leq\frac{2}{n^{\sigma}}, \qquad s=\sigma+i\tau\in\mathbb{C}.$$

Let $0<\delta,\varepsilon<1$. Taking an ε-skewed geometric mean of these estimates we have

$$|h_n(s)|\leq\left(\frac{|s|}{n^{\sigma+1}}\right)^{\varepsilon}\left(\frac{2}{n^{\sigma}}\right)^{1-\varepsilon}\leq C\frac{|s|^{\varepsilon}}{n^{\sigma+\varepsilon}}, \qquad s=\sigma+i\tau\in\mathbb{C},$$

and then for $\sigma+\varepsilon\geq 1+\delta$, we have

$$\begin{aligned}(3.12)\qquad |\zeta(s)| &\leq \frac{1}{|s-1|}+\sum_{n=1}^{\infty}|h_n(s)|\leq\frac{1}{|\tau|}+C\sum_{n=1}^{\infty}\frac{(\sigma+|\tau|)^{\varepsilon}}{n^{\sigma+\varepsilon}}\\ &\leq \frac{1}{|\tau|}+C\frac{(\sigma+|\tau|)^{\varepsilon}}{\sigma+\varepsilon-1}\leq C_{\delta}|\tau|^{\varepsilon},\end{aligned}$$

for $s=\sigma+i\tau$, $\sigma+\varepsilon\geq 1+\delta$, and $|\tau|\geq 1$.

Now with $\varepsilon=2\delta$ (which forces $\delta<\frac{1}{2}$), apply Cauchy's estimates to $\zeta(s)$ on the disk $B(1+i\tau,\delta)$ using (3.12) to conclude that for any $0<\delta<\frac{1}{2}$ (and with a constant C_{δ} that may change from one occurence to the next),

$$(3.13)\qquad \left|\zeta'(1+i\tau)\right|\leq\frac{C_{\delta}|\tau|^{\varepsilon}}{\delta}\leq C_{\delta}|\tau|^{2\delta}, \qquad |\tau|\geq 1.$$

In order to estimate $\left|\frac{1}{\zeta(1+i\tau)}\right|$ we recall from the proof of Theorem 17 that

$$\left|\zeta(\sigma)^3\zeta(\sigma+i\tau)^4\zeta(\sigma+i2\tau)\right|\geq 1, \qquad \sigma\geq 1,\tau\in\mathbb{R}\setminus\{0\}.$$

Thus

$$\left|\zeta(\sigma+i\tau)^4\right|\geq|\zeta(\sigma)|^{-3}|\zeta(\sigma+i2\tau)|^{-1}, \qquad \sigma\geq 1,$$

and from (3.12) we now conclude that for $\sigma\geq 1$,

$$|\zeta(\sigma+i\tau)|^4\geq c(\sigma-1)^3C_{\delta}^{-1}|\tau|^{-\varepsilon}, \qquad s=\sigma+i\tau,\sigma+\varepsilon\geq 1+\delta,|\tau|\geq 1.$$

Rewriting the inequality with $0<\varepsilon=\delta<1$ we get

$$|\zeta(\sigma+i\tau)|\geq C(\sigma-1)^{\frac{3}{4}}|\tau|^{-\frac{\delta}{4}}.$$

Now if $\sigma_\tau = 1 + \frac{1}{|\tau|^{13\delta}}$ (the number 13 is convenient for numerology but any number bigger than 9 would work here), then we have

$$|\zeta(\sigma_\tau + i\tau)| \geq C\left(|\tau|^{-13\delta}\right)^{\frac{3}{4}} |\tau|^{-\frac{\delta}{4}} = C\,|\tau|^{-10\delta},$$

and using (3.13), which requires $0 < \delta < \frac{1}{2}$, we get

$$\begin{aligned}|\zeta(\sigma_\tau + i\tau) - \zeta(1+i\tau)| &= \left|\int_1^{\sigma_\tau} \zeta'(u+i\tau)\,du\right| \leq \int_1^{\sigma_\tau} \left|\zeta'(u+i\tau)\right| du \\ &\leq \int_1^{1+\frac{1}{|\tau|^{13\delta}}} C_\delta\,|\tau|^{2\delta}\,du = C_\delta\,|\tau|^{-11\delta}.\end{aligned}$$

Thus for $|\tau|$ sufficiently large we have

$$|\zeta(\sigma_\tau + i\tau) - \zeta(1+i\tau)| \leq \frac{1}{2}\,|\zeta(\sigma_\tau + i\tau)|,$$

since $C_\delta\,|\tau|^{-11\delta} \leq \frac{1}{2}C\,|\tau|^{-10\delta}$ for $|\tau|$ sufficiently large. Consequently, we obtain

$$|\zeta(1+i\tau)| \geq \frac{1}{2}\,|\zeta(\sigma_\tau + i\tau)| \geq \frac{C}{2}\,|\tau|^{-10\delta}, \qquad 0 < \delta < \frac{1}{2},$$

for $|\tau|$ suficiently large.

Altogether we have

$$\left|\frac{\zeta'(1+i\tau)}{\zeta(1+i\tau)}\right| \leq c\left(C_\delta\,|\tau|^{2\delta}\right)\left(C_\delta\,|\tau|^{10\delta}\right) = C_\delta\,|\tau|^{12\delta}, \qquad 0 < \delta < \frac{1}{2}. \tag{3.14}$$

It now follows from (3.14) with any $0 < \delta < \frac{1}{2}$ that

$$|g_x(s)| = \left|\frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right)\right| \leq C_\delta \frac{x^{\sigma+1}}{|\tau|^2}\,|\tau|^{12\delta} = C_\delta x^{\sigma+1}\,|\tau|^{12\delta-2},$$

for $|\tau|$ suficiently large. Thus the integrals in (3.11) are dominated by $(c-1)\,S^{12\delta-2}$ for S suficiently large, which tends to 0 as $S \to \infty$ provided $0 < \delta < \frac{1}{6}$.

Step 2 now follows immediately from Step 1.

Step3: Given $\varepsilon > 0$ there is $3 \leq R < \infty$ and $0 < \eta < \eta(R) < 1$ so that for all x sufficiently large,

$$\left|\frac{1}{2\pi i}\int_{\beta_{R,\eta}} \frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right) ds\right| < \varepsilon\frac{x^2}{2}.$$

We may assume that $x \geq 1$. First we fix R so large that

$$\begin{aligned}&\left|\int_{1-i\infty}^{1-iR} \frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right) ds\right| + \left|\int_{1+iR}^{1+i\infty} \frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right) ds\right| \\ &+ \left|\int_{1-\eta-iR}^{1-iR} \frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right) ds\right| + \left|\int_{1-\eta+iR}^{1+iR} \frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right) ds\right| \\ &< \varepsilon\frac{x^2}{4}, \qquad 0 < \eta < 1,\end{aligned}$$

uniformly in $x \geq 1$. This is possible since (3.14) implies both

$$\left|\int_{1+iR}^{1+i\infty} \frac{x^{s+1}}{s(s+1)}\left(-\frac{\zeta'(s)}{\zeta(s)}\right) ds\right| \leq \int_R^\infty \frac{x^2}{\tau^2} C_\delta \tau^{12\delta} d\tau \leq \frac{C_\delta}{1-12\delta} R^{12\delta-1} x^2,$$

and

$$\left| \int_{1-\eta-iR}^{1-iR} \frac{x^{s+1}}{s(s+1)} \left(-\frac{\zeta'(s)}{\zeta(s)} \right) ds \right| + \left| \int_{1-\eta+iR}^{1+iR} \frac{x^{s+1}}{s(s+1)} \left(-\frac{\zeta'(s)}{\zeta(s)} \right) ds \right|$$
$$\leq 2 \int_{1-\eta}^{1} \frac{x^2}{R^2} C_\delta R^{12\delta} d\sigma \leq C_\delta x^2 R^{12\delta-2},$$

where both $R^{12\delta-1}$ and $R^{12\delta-2}$ tend to 0 as $R \to \infty$ if $0 < \delta < \frac{1}{12}$.

Recall that for $0 < \eta < \eta(R)$, $\zeta(s)$ has no zero on or inside the rectangle with sides

$$[1-\eta-iR, 1-iR], \ [1-iR, 1+iR], \ [1+iR, 1-\eta+iR], \ [1-\eta+iR, 1-\eta-iR].$$

Thus we estimate

$$\left| \int_{1-\eta-iR}^{1-\eta+iR} \frac{x^{s+1}}{s(s+1)} \left(-\frac{\zeta'(s)}{\zeta(s)} \right) ds \right|$$
$$\leq \quad x^{2-\eta} \int_{-R}^{R} \frac{1}{|1-\eta-i\tau| \, |2-\eta-i\tau|} \left| \frac{\zeta'(s)}{\zeta(s)} \right| d\tau = C_{R,\eta} x^{2-\eta}.$$

Altogether we then have

$$\left| \frac{1}{2\pi i} \int_{\beta_{R,\eta}} \frac{x^{s+1}}{s(s+1)} \left(-\frac{\zeta'(s)}{\zeta(s)} \right) ds \right| < \varepsilon \frac{x^2}{4} + C_{R,\eta} x^{2-\eta} = \left(\frac{\varepsilon}{2} + \frac{C_{R,\eta}}{x^\eta} \right) \frac{x^2}{2},$$

which completes the proof of Step 3.

The asymptotic estimate (3.7),

$$\lim_{x \to \infty} \frac{\psi_1(x)}{\frac{1}{2}x^2} = 1,$$

now follows immediately. Indeed, from the identity (3.8),

$$\psi_1(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{x^{s+1}}{s(s+1)} \left(-\frac{\zeta'(s)}{\zeta(s)} \right) ds,$$

and Steps 2 and 3 above, we have for any $\varepsilon > 0$,

$$\left| \frac{\psi_1(x)}{\frac{1}{2}x^2} - 1 \right| = \left| \frac{\psi_1(x) - \frac{1}{2}x^2}{\frac{1}{2}x^2} \right| < \left| \frac{\varepsilon \frac{x^2}{2}}{\frac{1}{2}x^2} \right| = \varepsilon,$$

for all sufficiently large x.

This completes the proof of the Prime Number Theorem.

Corollary 9. *If $\{p_n\}_{n=1}^\infty$ is the sequence of primes in increasing order, then*

$$\lim_{n \to \infty} \frac{p_n}{n \ln n} = 1.$$

Proof: Since $\ln x$ is continuous, the Prime Number Theorem yields

$$0 = \ln 1 = \lim_{x \to \infty} \ln \frac{\pi(x)}{\frac{x}{\ln x}} = \lim_{x \to \infty} \{\ln \pi(x) + \ln \ln x - \ln x\}.$$

If we now divide by $\ln x$ and use l'Hôspital's rule to get

$$\lim_{x \to \infty} \frac{\ln \ln x}{\ln x} = \lim_{x \to \infty} \frac{\frac{1}{x \ln x}}{\frac{1}{x}} = \lim_{x \to \infty} \frac{1}{\ln x} = 0,$$

we obtain $\lim_{n\to\infty} \frac{\ln \pi(x)}{\ln x} = 1$. This and another application of the Prime Number Theorem yield

$$\lim_{x\to\infty} \frac{\pi(x) \ln \pi(x)}{x} = \lim_{x\to\infty} \frac{\pi(x)}{\frac{x}{\ln x}} \lim_{x\to\infty} \frac{\ln \pi(x)}{\ln x} = 1.$$

If we now replace x by p_n and use $\pi(p_n) = n$, we obtain $\lim_{n\to\infty} \frac{n \ln n}{p_n} = 1$.

Part 2

Boundary behaviour of Riemann maps

In Part 2 of these notes we address Carathéodory's theorem on extending a Riemann map to a homeomorphism up to the boundary. Chapter 6 introduces the theory of the Poisson integral and establishes Fatou's theorem on radial limits, which is the key ingredient in extending a Riemann map to a simple boundary point. In Chapter 7 we use Lindelöf's theorem to help complete the proof of Carathéodory's theorem. Finally, Appendices A and B provide those aspects of topology and measure theory needed in Part 2.

Suppose that $f : \mathbb{D} \to \Omega$ is a Riemann map onto a *bounded* simply connected domain Ω. The question we consider here is this: when does the map f extend to a homeomorphism from the closure $\overline{\mathbb{D}}$ to the closure $\overline{\Omega}$? An obvious necessary condition is that $\partial\Omega$ is a Jordan curve since $f : \mathbb{T} \to \partial\Omega$ is continuous, one-to-one and onto. Carathéodory's Theorem shows that the converse holds. Moreover, we also give a topological test for $\partial\Omega$ to be a Jordan curve involving the following notion of a *simple* boundary point.

Definition 11. *A boundary point w of a simply connected domain Ω in the plane is called* simple *if for every sequence $\{z_n\}_{n=1}^{\infty} \subset \Omega$ with $\lim_{n\to\infty} z_n = w$, there is a (continuous) curve $\beta : [0,1) \to \Omega$ and an increasing sequence of parameter points $\{t_n\}_{n=1}^{\infty} \subset [0,1)$ with $\lim_{n\to\infty} t_n = 1$ such that $\beta(t_n) = z_n$ for all $n \geq 1$ and $\lim_{t\to 1} \beta(t) = w$.*

Roughly speaking the point $w \in \partial\Omega$ is simple if for every sequence in Ω that approaches w, there is a curve passing in order through the sequence that has limit w. For example, every boundary point of the unit disk $\mathbb{D}$ is simple, and we will show below that that every point in a Jordan curve is a simple boundary point of its bounded component. An example of a simply connected domain Ω that has nonsimple boundary points is the open square $(0,1)^2$ with the vertical line segments $I_n = \left\{\frac{1}{2^n}\right\} \times \left(0, 1 - \frac{1}{2^n}\right]$ removed for each $n \geq 1$. Each point on I_n other than the endpoint $\left(\frac{1}{2^n}, 1 - \frac{1}{2^n}\right)$ is a nonsimple boundary point (consider a sequence that alternates each side of I_n) and each point $(0, y)$ with $0 < y < 1$ is an especially "nonsimple" boundary point in that there is *not a single* curve in Ω with limit $(0, y)$.

Theorem 20. *(Carathéodory's Theorem) Suppose that Ω is a bounded simply connected domain Ω in the plane. Then the following conditions are equivalent:*

1. *Every Riemann map $f : \mathbb{D} \to \Omega$ extends to a homeomorphism from $\overline{\mathbb{D}}$ to $\overline{\Omega}$,*
2. *The boundary $\partial\Omega$ of Ω is a Jordan curve,*
3. *Every boundary point of Ω is simple.*

The most difficult implication is that (3) implies (1). For this we will need the Poisson representation of holomorphic functions along with Fatou's Theorem and Lindelöf's Theorem, which we take up in the next chapter.

CHAPTER 6

The Poisson representation

We begin by introducing one of the most famous Hilbert spaces in analysis, the Hardy space $H^2(\mathbb{D})$ on the unit disk $\mathbb{D}$. For $f(z)=\sum_{n=0}^{\infty}a_n z^n$, $z\in\mathbb{D}$, the orthogonality relations

$$\frac{1}{2\pi}\int_0^{2\pi}e^{i(n-m)\theta}d\theta=\begin{cases}1 & \text{if}\quad n=m\\ 0 & \text{if}\quad n\neq m\end{cases}\tag{0.15}$$

yield

$$\begin{aligned}\frac{1}{2\pi}\int_0^{2\pi}\left|\sum_{n=0}^{\infty}a_n\left(re^{i\theta}\right)^n\right|^2 d\theta &= \frac{1}{2\pi}\int_0^{2\pi}\sum_{n,m=0}^{\infty}a_n\left(re^{i\theta}\right)^n\overline{a_m\left(re^{i\theta}\right)^m}d\theta\\ &= \sum_{n,m=0}^{\infty}a_n\overline{a_m}r^{n+m}\frac{1}{2\pi}\int_0^{2\pi}e^{i(n-m)\theta}d\theta=\sum_{n=0}^{\infty}\left|a_n r^n\right|^2\end{aligned}$$

for $0<r<1$ by absolute convergence. Thus we have

$$\sum_{n=0}^{\infty}|a_n|^2=\sup_{0<r<1}\frac{1}{2\pi}\int_0^{2\pi}\left|\sum_{n=0}^{\infty}a_n\left(re^{i\theta}\right)^n\right|^2 d\theta=\sup_{0<r<1}\frac{1}{2\pi}\int_0^{2\pi}\left|f\left(re^{i\theta}\right)\right|^2 d\theta,\tag{0.16}$$

which we use to define the norm squared $\|f\|^2_{H^2(\mathbb{D})}$ for the Hardy space

$$H^2(\mathbb{D})=\left\{f\in H(\mathbb{D}):\sup_{0<r<1}\left(\frac{1}{2\pi}\int_0^{2\pi}\left|f\left(re^{i\theta}\right)\right|^2 d\theta\right)^{\frac{1}{2}}<\infty\right\}.$$

We digress for a moment to provide the proof of completeness of the Hilbert space $L^2(\mathbb{T})$ of square integrable Lebesgue measurable functions on the circle $\mathbb{T}=\partial\mathbb{D}$, i.e. those measurable $f\left(e^{i\theta}\right)$ satisfying

$$\|f\|_2\equiv\left(\frac{1}{2\pi}\int_0^{2\pi}\left|f\left(e^{i\theta}\right)\right|^2 d\theta\right)^{\frac{1}{2}}<\infty.$$

We equip $L^2(\mathbb{T})$ with the metric $d(f,g)=\|f-g\|_2$, $f,g\in L^2(\mathbb{T})$. It is here that Lebesgue integration is required, as the next result would fail with Riemann integration in place of Lebesgue integration. See the appendix for this and the basic theory of Lebesgue integration.

Proposition 13. *The metric space $L^2(\mathbb{T})$ is complete.*

Proof: Suppose that $\{f_n\}_{n=1}^{\infty}$ is a Cauchy sequence in $L^2(\mathbb{T})$. Choose a rapidly converging subsequence $\{f_{n_k}\}_{k=1}^{\infty}$, by which we mean $\sum_{k=1}^{\infty}\left\|f_{n_{k+1}}-f_{n_k}\right\|_2<\infty$.

This is easily accomplished inductively by choosing for example $\{n_k\}_{k=1}^{\infty}$ strictly increasing such that

$$\|f_n - f_{n_k}\|_2 < \frac{1}{2^k}, \qquad n \geq n_{k+1}.$$

Then set

$$g = |f_{n_1}| + \sum_{k=1}^{\infty} \left| f_{n_{k+1}} - f_{n_k} \right|.$$

By Minkowski's inequality we have

$$\|g\|_2 \leq \|f_{n_1}\|_2 + \sum_{k=1}^{\infty} \left\| f_{n_{k+1}} - f_{n_k} \right\|_2 < \infty,$$

and it follows that

$$0 \leq g\left(e^{i\theta}\right) = \left| f_{n_1}\left(e^{i\theta}\right) \right| + \sum_{k=1}^{\infty} \left| f_{n_{k+1}}\left(e^{i\theta}\right) - f_{n_k}\left(e^{i\theta}\right) \right| < \infty$$

for almost every $\theta \in [0, 2\pi)$. Thus the series

$$f_{n_1}\left(e^{i\theta}\right) + \sum_{k=1}^{\infty} \left\{ f_{n_{k+1}}\left(e^{i\theta}\right) - f_{n_k}\left(e^{i\theta}\right) \right\}$$

converges absolutely for almost every $\theta \in [0, 2\pi)$ to a Lebesgue measurable function $f\left(e^{i\theta}\right)$.

We claim that $f \in L^2(\mathbb{T})$ and that $\lim_{n\to\infty} f_n = f$ in $L^2(\mathbb{T})$. Indeed, Fatou's lemma gives

$$\begin{aligned}
\frac{1}{2\pi}\int_0^{2\pi} \left| f\left(e^{i\theta}\right) - f_{n_\ell}\left(e^{i\theta}\right) \right|^2 d\theta &= \frac{1}{2\pi}\int_0^{2\pi} \liminf_{k\to\infty} \left| f_{n_k}\left(e^{i\theta}\right) - f_{n_\ell}\left(e^{i\theta}\right) \right|^2 d\theta \\
&\leq \liminf_{k\to\infty} \frac{1}{2\pi}\int_0^{2\pi} \left| f_{n_k}\left(e^{i\theta}\right) - f_{n_\ell}\left(e^{i\theta}\right) \right|^2 d\theta \\
&= \liminf_{k\to\infty} \|f_{n_k} - f_{n_\ell}\|_2^2 \to 0
\end{aligned}$$

as $\ell \to \infty$ by the Cauchy condition. This shows that $f - f_{n_\ell} \in L^2(\mathbb{T})$, hence $f \in L^2(\mathbb{T})$, and also that $f_{n_\ell} \to f$ in $L^2(\mathbb{T})$ as $\ell \to \infty$. Finally, this together with the fact that $\{f_n\}_{n=1}^{\infty}$ is a Cauchy sequence, easily shows that $f_n \to f$ in $L^2(\mathbb{T})$ as $n \to \infty$.

Porism 1: If $\{f_n\}_{n=1}^{\infty}$ is a rapidly converging sequence in $L^2(\mathbb{T})$,

$$\sum_{n=1}^{\infty} \|f_{n+1} - f_n\|_2 < \infty,$$

then

$$\lim_{n\to\infty} f_n\left(e^{i\theta}\right) = f_1\left(e^{i\theta}\right) + \sum_{n=1}^{\infty} \left\{ f_{n+1}\left(e^{i\theta}\right) - f_n\left(e^{i\theta}\right) \right\}$$

exists for almost every $\theta \in [0, 2\pi)$.

Returning to the Hardy space $H^2(\mathbb{D})$ we note that (0.15) shows that $\{f_r\}_{0<r<1}$ is Cauchy in $L^2(\mathbb{T})$ as $r \to 1$ where $f_r(z) = f(rz)$ for $r < 1$ and $z \in \overline{\mathbb{D}}$:

$$\|f_r - f_s\|^2_{L^2(\mathbb{T})} = \frac{1}{2\pi}\int_0^{2\pi}\left|\sum_{n=0}^{\infty}(r^n - s^n)\, a_n e^{in\theta}\right|^2 d\theta = \sum_{n=0}^{\infty}|r^n - s^n|^2\,|a_n|^2 \to 0$$

as $r, s \to 1$ by the dominated convergence theorem for series since $\sum_{n=0}^{\infty}|a_n|^2 < \infty$ by (0.16). By Proposition 13, the completeness of $L^2(\mathbb{T})$, there is $f^* \in L^2(\mathbb{T})$ such that $f^* = \lim_{r\to 1} f_r$ in $L^2(\mathbb{T})$. We easily compute by taking limits that the Fourier coefficients of f^* satisfy

$$\begin{aligned} \widehat{f^*}(n) &\equiv \frac{1}{2\pi}\int_0^{2\pi} f^*\left(e^{i\theta}\right)\overline{e^{in\theta}}d\theta = \lim_{r\to 1}\frac{1}{2\pi}\int_0^{2\pi} f_r\left(e^{i\theta}\right)\overline{e^{in\theta}}d\theta \\ &= \left\{\begin{array}{ccc} a_n & \text{if} & n \geq 0 \\ 0 & \text{if} & n < 0 \end{array}\right., \end{aligned}$$

and that the inner product $\langle\cdot,\cdot\rangle$ on $H^2(\mathbb{D})$ satisfies

$$\langle f, g\rangle = \sum_{n=0}^{\infty} a_n\overline{b_n} = \frac{1}{2\pi}\int_0^{2\pi} f^*\left(e^{i\theta}\right)\overline{g^*\left(e^{i\theta}\right)}d\theta$$

where $f(z) = \sum_{n=0}^{\infty} a_n z^n$ and $g(z) = \sum_{n=0}^{\infty} b_n z^n$. We also have by taking limits the Cauchy formula

$$f(z) = \lim_{r\to 1} f_r(z) = \lim_{r\to 1}\frac{1}{2\pi i}\int_{\mathbb{T}}\frac{f_r(w)}{w - z}dw = \frac{1}{2\pi i}\int_{\mathbb{T}}\frac{f^*(w)}{w - z}dw, \tag{0.17}$$

which can also be expressed in terms of the inner product as

$$f(z) = \frac{1}{2\pi i}\int_{\mathbb{T}}\frac{f^*\left(e^{i\theta}\right)}{e^{i\theta} - z}ie^{i\theta}d\theta = \frac{1}{2\pi}\int_{\mathbb{T}}\frac{f^*\left(e^{i\theta}\right)}{1 - e^{-i\theta}z}d\theta = \langle f, k_z\rangle\,, \tag{0.18}$$

for $z \in \mathbb{D}$ where

$$k_z(w) = \frac{1}{1 - \overline{z}w} = \sum_{n=0}^{\infty}\overline{z}^n w^n \in H^2(\mathbb{D}) \tag{0.19}$$

(since $\|k_z\|_{H^2(\mathbb{D})} = \sqrt{\sum_{n=0}^{\infty}|z|^{2n}} = \frac{1}{\sqrt{1-|z|^2}}$) is the so-called reproducing kernel for $H^2(\mathbb{D})$.

We would now like to obtain a representation formula of Cauchy type, such as in (0.17), for the real part of $f(z)$. So let $f(z) = u(z) + iv(z)$ where u and v are real-valued functions in the disk. Unfortunately we cannot just take the real part of each side in the Cauchy representation

$$f(z) = \frac{1}{2\pi i}\int_{\mathbb{T}}\frac{f^*(w)}{w - z}dw = \frac{1}{2\pi}\int_0^{2\pi} f^*\left(e^{it}\right)\frac{e^{it}}{e^{it} - z}dt, \tag{0.20}$$

since both the function $f^*\left(e^{it}\right)$ and the kernel $\frac{e^{it}}{e^{it}-z}$ are complex-valued. If we could replace the kernel $\frac{e^{it}}{e^{it}-z}$ in (0.20) with a kernel that was *real-valued*, then we could indeed take real parts of both sides of (0.20) to obtain a representation formula for $u(z) = \operatorname{Re} f(z)$.

For this purpose we have at our disposal Cauchy's theorem which says that

$$\frac{1}{2\pi i}\int_{\mathbb{T}}\frac{f^*(w)}{w - \zeta}dw = \lim_{r\to 1}\frac{1}{2\pi i}\int_{\mathbb{T}}\frac{f_r(w)}{w - \zeta}dw = \lim_{r\to 1} 0 = 0$$

for any $\zeta \in \mathbb{C} \setminus \overline{\mathbb{D}}$ (since $\frac{f_r(w)}{w-\zeta}$ is holomorphic in a neighbourhood of the closed disk $\overline{\mathbb{D}}$). Motivated by the fact that $z + \overline{z}$ is real where $\overline{z}$ is the reflection of z across the real line $\mathbb{R}$, we consider the reflection z^* of z across the circle $\mathbb{T}$ given by

$$z^* = \frac{z}{|z|^2} = \frac{1}{\overline{z}}, \qquad z \neq 0.$$

If we let $\zeta = z^*$ we obtain

$$f(z) - 0 = \frac{1}{2\pi} \int_0^{2\pi} f^*\left(e^{it}\right) \left\{ \frac{e^{it}}{e^{it} - z} - \frac{e^{it}}{e^{it} - z^*} \right\} dt.$$

Now we compute that

$$\begin{aligned} \frac{e^{it}}{e^{it} - z} - \frac{e^{it}}{e^{it} - z^*} &= \frac{e^{it}}{e^{it} - z} - \frac{e^{it}}{e^{it} - \frac{1}{\overline{z}}} = \frac{e^{it}}{e^{it} - z} - \frac{\overline{z}}{\overline{z} - e^{-it}} \\ &= \frac{e^{it}}{e^{it} - z} + \frac{\overline{z}}{\overline{e^{it} - z}} = \frac{1 - |z|^2}{\left|z - e^{it}\right|^2}, \end{aligned}$$

which yields the *Poisson representation* of f:

$$f(z) = \frac{1}{2\pi} \int_0^{2\pi} f^*\left(e^{it}\right) \frac{1 - |z|^2}{\left|z - e^{it}\right|^2} dt, \qquad z \in \mathbb{D}. \tag{0.21}$$

Finally, we can take real parts of each side of this latter equation to obtain

$$u(z) = \frac{1}{2\pi} \int_0^{2\pi} u^*\left(e^{it}\right) \frac{1 - |z|^2}{\left|z - e^{it}\right|^2} dt, \qquad z \in \mathbb{D}, \tag{0.22}$$

where we write $f^* = u^* + iv^*$ with u^* and v^* real-valued functions on $\mathbb{T}$. This formula is valid for $u = \operatorname{Re} f$ where $f \in H^2(\mathbb{D})$.

Remark 5. *An important side benefit of the Poisson kernel is its small size in comparison to the Cauchy kernel:*

$$\frac{1}{2\pi} \int_0^{2\pi} \frac{1 - |z|^2}{\left|z - e^{it}\right|^2} dt = 1 \textit{ while } \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{\left|e^{it} - z\right|} dt \approx \ln\left(1 + \frac{1}{1 - |z|}\right)$$

for $z \in \mathbb{D}$. This small size will play a crucial role in the proof of Fatou's Theorem 22 below.

Remark 6. *The Poisson representation (0.21) can also be quickly obtained from the Cauchy representation (0.20) via the following trick. For $f \in H^2(\mathbb{D})$ and $z \in \mathbb{D}$, let $g(w) = \frac{f(w)}{1 - w\overline{z}}$. Then $g \in H^2(\mathbb{D})$, $g^*\left(e^{it}\right) = \frac{f^*\left(e^{it}\right)}{1 - e^{it}\overline{z}}$ and (0.20) yields*

$$\frac{f(z)}{1 - |z|^2} = g(z) = \frac{1}{2\pi} \int_0^{2\pi} \frac{f^*\left(e^{it}\right)}{1 - e^{it}\overline{z}} \frac{e^{it}}{e^{it} - z} dt = \frac{1}{2\pi} \int_0^{2\pi} \frac{f^*\left(e^{it}\right)}{\left|z - e^{it}\right|^2} dt.$$

We note in passing the following consequence of Porism 6: if $\lim_{n\to\infty} r_n = 1$ and $\{f_{r_n}\}_{n=1}^{\infty}$ is a rapidly converging sequence in $L^2(\mathbb{T})$, then

$$f^*\left(e^{it}\right) = \lim_{n\to\infty} f_{r_n}\left(e^{it}\right), \qquad a.e.\ e^{it} \in \mathbb{T}, \tag{0.23}$$

and an analogous statement holds for $u^*\left(e^{it}\right)$. For the proof of Carathéodory's Theorem 20 we will need the much stronger theorem of Fatou that says $f^*\left(e^{it}\right) = \lim_{r\to 1} f_r\left(e^{it}\right)$ for a.e. $e^{it} \in \mathbb{T}$. This will be proved below after a short detour to consider harmonic functions and the Dirichlet problem in the disk. But first we note the following uniqueness result for bounded holomorphic functions that will play a significant role in the proof of Theorem 20.

LEMMA 3. *If $f \in H^{\infty}(\mathbb{D})$ and $f^*\left(e^{it}\right) = 0$ for almost every e^{it} in some arc I of positive length in $\mathbb{T}$, then f is identically zero.*

Proof: Choose an integer N so that $|I| > \frac{2\pi}{N}$ and consider the function $F \in H^2(\mathbb{D})$ given by

$$F(z) = \prod_{k=1}^{N} f\left(e^{i\frac{2\pi k}{N}} z\right) \equiv \prod_{k=1}^{N} f^k(z), \qquad z \in \mathbb{D}.$$

The boundedness of f is used in concluding that the product F is actually a bounded holomophic function in the disk $\mathbb{D}$, hence is in $H^2(\mathbb{D})$. Next choose a sequence $\{r_n\}_{n=1}^{\infty}$ with $\lim_{n\to\infty} r_n = 1$ for which $\{F_{r_n}\}$ and $\left\{f_{r_n}^k\right\}_{n=1}^{\infty}$, $1 \le k \le N$, are each rapidly convergent sequences in $L^2(\mathbb{T})$. Then for almost every $t \in [0, 2\pi)$ we have

$$\begin{aligned} F^*\left(e^{it}\right) &= \lim_{n\to\infty} F_{r_n}\left(e^{it}\right) = \lim_{n\to\infty} \prod_{k=1}^{N} \left(f^k\right)_{r_n}\left(e^{it}\right) \\ &= \prod_{k=1}^{N} \lim_{n\to\infty} \left(f^k\right)_{r_n}\left(e^{it}\right) = \prod_{k=1}^{N} \left(f^k\right)^*\left(e^{it}\right). \end{aligned}$$

Since $\left(f^k\right)^*\left(e^{it}\right) = 0$ for a.e. $t \in e^{-i\frac{2\pi k}{N}} I$, and since $\bigcup_{k=1}^{N} e^{-i\frac{2\pi k}{N}} I = \mathbb{T}$, we conclude that $F^*\left(e^{it}\right) = 0$ for a.e. $t \in \mathbb{T}$. Now either the Cauchy representation (0.17) or the Poisson representation (0.21) shows that F is identically zero in the disk $\mathbb{D}$. Thus the zero set $Z(f)$ of f must be uncountable since $\mathbb{D} = Z(F) = \bigcup_{k=1}^{N} Z\left(f^k\right)$, and hence f is identically zero in the disk $\mathbb{D}$ by Theorem 6.

1. Harmonic functions

We pause at this point to note that we only use the Poisson representation for *holomorphic* functions (0.21) in our proof of Carathéodory's Theorem 20. However, since we used the real part representation (0.22) to motivate our calculation of the Poisson kernel $\frac{1-|z|^2}{|z-e^{it}|^2}$, we will spend some time using (0.22) to develop the most elementary facts in the theory of harmonic functions u, i.e. twice continuously differentiable solutions u to Laplace's equation

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

In particular we will prove formula (0.22) with $u^*\left(e^{it}\right) = u\left(e^{it}\right)$ for any $u \in C\left(\overline{\mathbb{D}}\right) \cap C^2(\mathbb{D})$ that is harmonic in $\mathbb{D}$.

We first observe that if $f = u + iv \in H(\mathbb{D})$, then the functions u and v are harmonic in $\mathbb{D}$:

$$\triangle u(z) = \triangle v(z) = 0, \qquad z \in \mathbb{D}. \tag{1.1}$$

Indeed, if $f = u + iv$ is a holomorphic function of $z = x + iy$ then $\frac{\partial}{\partial \overline{z}} f \equiv 0$ and so

$$\begin{aligned} 0 &= \frac{\partial}{\partial z} 0 = \frac{\partial}{\partial z} \frac{\partial}{\partial \overline{z}} f(z) = \frac{1}{2}\left(\frac{\partial}{\partial x} + \frac{1}{i}\frac{\partial}{\partial y}\right) \frac{1}{2}\left(\frac{\partial}{\partial x} - \frac{1}{i}\frac{\partial}{\partial y}\right) f(z) \\ &= \frac{1}{4}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)(u(z) + iv(z)) = \frac{1}{4} \triangle u(z) + \frac{i}{4} \triangle v(z). \end{aligned}$$

We next note the many faces of the Poisson kernel $\frac{1-|z|^2}{|z-e^{it}|^2}$.

Lemma 4. *For $z = re^{i\theta} \in \mathbb{D}$ and $e^{it} \in \mathbb{T}$ we have*

$$\frac{1-r^2}{1-2r\cos(\theta-t)+r^2} = \frac{1-|z|^2}{|z-e^{it}|^2} = \operatorname{Re}\left\{\frac{e^{it}+z}{e^{it}-z}\right\} = \sum_{n\in\mathbb{Z}} r^{|n|} e^{in(\theta-t)}. \tag{1.2}$$

Proof: We have

$$\begin{aligned}
\operatorname{Re}\left\{\frac{e^{it}+z}{e^{it}-z}\right\} &= \operatorname{Re}\left\{1+2\frac{z}{e^{it}-z}\right\} = 1+2\operatorname{Re}\frac{z}{e^{it}-z} \\
&= 1+\frac{z}{e^{it}-z}+\frac{\overline{z}}{\overline{e^{it}-z}} \\
&= \frac{e^{it}}{e^{it}-z}+\frac{\overline{z}}{\overline{e^{it}-z}} = \frac{1-|z|^2}{|z-e^{it}|^2},
\end{aligned}$$

and also

$$1+2\operatorname{Re}\frac{z}{e^{it}-z} = 1+2\operatorname{Re}\frac{ze^{-it}}{1-ze^{-it}} = 1+2\operatorname{Re}\sum_{n=1}^{\infty}\left(ze^{-it}\right)^n = \sum_{n\in\mathbb{Z}} r^{|n|}e^{in(\theta-t)}.$$

An inspection of the representation formula (0.22) together with the lemma reveals that (0.22) can be rewritten

$$\begin{aligned}
u(z) &= \frac{1}{2\pi}\int_0^{2\pi} u^*\left(e^{it}\right)\frac{1-r^2}{1-2r\cos(\theta-t)+r^2}dt \\
&= \frac{1}{2\pi}\int_0^{2\pi} u^*\left(e^{it}\right)P_r(\theta-t)\,dt \\
&= P_r * u^*(\theta), \qquad z = re^{i\theta}\in\mathbb{D},
\end{aligned}$$

where

$$P_r(\phi) = \frac{1-r^2}{1-2r\cos(\phi)+r^2}, \qquad 0\le r<1, 0\le\phi<2\pi,$$

and

$$(f*g)(\zeta) = \int_{\mathbb{T}} f\left(\zeta\eta^{-1}\right)g(\eta)\,d\sigma(\eta)$$

denotes convolution of f and g on the compact group $\mathbb{T}$ with Haar measure $d\sigma\left(e^{it}\right) = \frac{1}{2\pi}dt$.

A crucial observation at this juncture is that the function $P_r*\varphi(\theta)$ is a harmonic function of $z = re^{i\theta}\in\mathbb{D}$ for *any* integrable function φ on $\mathbb{T}$. Indeed, if φ is real-valued, then

$$\begin{aligned}
P_r*\varphi(\theta) &= \frac{1}{2\pi}\int_0^{2\pi}\varphi\left(e^{it}\right)P_r(\theta-t)\,dt \\
&= \frac{1}{2\pi}\int_0^{2\pi}\varphi\left(e^{it}\right)\operatorname{Re}\left\{\frac{e^{it}+z}{e^{it}-z}\right\}dt \\
&= \operatorname{Re}\frac{1}{2\pi}\int_0^{2\pi}\varphi\left(e^{it}\right)\left\{\frac{e^{it}+z}{e^{it}-z}\right\}dt.
\end{aligned}$$

Now the calculation in (3.2) shows that the integral

$$\int_0^{2\pi} \varphi\left(e^{it}\right)\left\{\frac{e^{it}+z}{e^{it}-z}\right\} dt$$

defines a holomorphic function of z in the disk $\mathbb{D}$, and hence its real part $P_r * \varphi(\theta)$ is harmonic by (1.1). In general we write φ as a sum of its real and imaginary parts and apply the above result to each part. This completes the proof that $P_r * \varphi(\theta)$ is harmonic for integrable φ.

The above observation justifies the following definition: for $\varphi \in L^1(\mathbb{T})$, the space of complex-valued Lebesgue integrable functions on $\mathbb{T}$, we set

$$\mathbb{P}\varphi(z) = P_r * \varphi(\theta), \qquad z = re^{i\theta} \in \mathbb{D},$$

and refer to $\mathbb{P}\varphi$ as the Poisson integral of φ. Note that φ is an integrable function defined on $\mathbb{T} = \partial\mathbb{D}$ the boundary of the disk $\mathbb{D}$, while $\mathbb{P}\varphi$ is a harmonic function defined on the disk $\mathbb{D}$ itself. A natural question now arises: if φ is *continuous* on the circle, does the Poisson integral $\mathbb{P}\varphi$ have limit φ at the boundary, i.e. is the function

$$\widetilde{\varphi}(z) = \begin{cases} \mathbb{P}\varphi(z) & \text{for} \quad z \in \mathbb{D} \\ \varphi(z) & \text{for} \quad z \in \mathbb{T} \end{cases}, \tag{1.3}$$

defined for z in the closed disk $\overline{\mathbb{D}}$, actually continuous on $\overline{\mathbb{D}}$? The answer turns out to be yes, and this justifies writing $\mathbb{P}\varphi$ for $\widetilde{\varphi}$ on the closed disk $\overline{\mathbb{D}}$, and referring to $\mathbb{P}\varphi$ as the Poisson extension of φ to the closed disk $\overline{\mathbb{D}}$.

PROPOSITION 14. *If $\varphi \in C(\mathbb{T})$, then $\widetilde{\varphi} \in C\left(\overline{\mathbb{D}}\right)$ where $\widetilde{\varphi}$ is the Poisson extension of φ to $\overline{\mathbb{D}}$ defined in (1.3).*

Proof: We have $P_r(\phi) = \frac{1-r^2}{|re^{i\theta}-1|^2}$ by (1.2) with $\theta = \phi$ and $t = 0$. We now claim that P_r has the following three properties of an approximate identity:

(1) $P_r(\phi) > 0$ for all $re^{i\phi} \in \mathbb{D}$,
(2) $\int_0^{2\pi} P_r(\phi)\, d\phi = 1$ for all $re^{i\phi} \in \mathbb{D}$,
(3) $\max_{|e^{i\phi}-1|\geq\delta} P_r(\phi)$ tends to 0 as $r \to 1$ for each fixed $0 < \delta < 2$.

The first property is obvious, the second property follows from (0.21) with $f \equiv 1$, and the third property follows from

$$\left|re^{i\phi} - 1\right| \geq \left|e^{i\phi} - 1\right| \geq \delta.$$

We now use a standard paradigm to show that $\mathbb{P}\varphi\left(re^{i\theta}\right) \to \varphi\left(e^{i\phi}\right)$ as $re^{i\theta} \to e^{i\phi}$. By rotation invariance we may assume that $\phi = 0$, i.e. $e^{i\phi} = 1$, and by property (2) we may subtract the constant $\varphi(1)$ which means we may assume that $\varphi(1) = 0$. Thus we must prove

$$\lim_{re^{i\theta}\to 1} \mathbb{P}\varphi\left(re^{i\theta}\right) = 0.$$

Let $\varepsilon > 0$ be given. By the continuity of φ there is $0 < \delta < 2$ so that

$$\left|\varphi\left(e^{it}\right)\right| < \frac{\varepsilon}{2}, \qquad \text{whenever } \left|e^{it} - 1\right| < \delta. \tag{1.4}$$

So using properties (1) and (2) and inequality (1.4) we estimate

$$\begin{aligned}\left|\mathbb{P}\varphi\left(re^{i\theta}\right)\right| &= \left|\frac{1}{2\pi}\int_0^{2\pi}\varphi\left(e^{it}\right)P_r\left(\theta-t\right)dt\right| \\ &\leq \frac{1}{2\pi}\left\{\int_{\left|e^{it}-1\right|<\delta}+\int_{\left|e^{it}-1\right|\geq\delta}\right\}\left|\varphi\left(e^{it}\right)\right|P_r\left(\theta-t\right)dt \\ &< \frac{\varepsilon}{2}+\max_{\zeta\in\mathbb{T}}\left|\varphi\left(\zeta\right)\right|\max_{\left|e^{i\phi}-1\right|\geq\frac{\delta}{2}}P_r\left(\phi\right),\end{aligned}$$

provided $\left|e^{i\theta}-1\right|<\frac{\delta}{2}$, since then

$$\left|e^{i(\theta-t)}-1\right|=\left|\left(e^{i(\theta-t)}-e^{i\theta}\right)+\left(e^{i\theta}-1\right)\right|\geq\left|e^{it}-1\right|-\left|e^{i\theta}-1\right|\geq\delta-\frac{\delta}{2}=\frac{\delta}{2}.$$

But now property (3) shows that for all r sufficiently close to 1 and $\left|e^{i\theta}-1\right|<\frac{\delta}{2}$ we have

$$\left|\mathbb{P}\varphi\left(re^{i\theta}\right)\right|<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.$$

Since this inequality holds also for $r=1$ and $\left|e^{i\theta}-1\right|<\frac{\delta}{2}$, the proof of Proposition 14 is complete.

Proposition 14 solves the Dirichlet problem for the Laplace operator in the unit disk. Namely, given continuous boundary data φ on $\mathbb{T}$, there is $u\in C\left(\overline{\mathbb{D}}\right)\cap C^2\left(\mathbb{D}\right)$ satisfying the boundary value problem:

$$\left\{\begin{array}{ccc}\triangle u=0 & \text{in} & \mathbb{D}\\ u=\varphi & \text{on} & \mathbb{T}=\partial\mathbb{D}\end{array}\right. . \tag{1.5}$$

Uniqueness of the solution u to the Dirichlet problem (1.5) will follow from the maximum principle for harmonic functions.

Proposition 15. *(Maximum principle for harmonic functions) Let Ω be a bounded domain in the plane. If $u\in C\left(\overline{\Omega}\right)\cap C^2\left(\Omega\right)$ is harmonic in Ω, then u achieves its maximum on the boundary:*

$$\sup_{z\in\overline{\Omega}}u\left(z\right)\leq\sup_{z\in\partial\Omega}u\left(z\right). \tag{1.6}$$

Proof: For $\varepsilon>0$ consider the function

$$u_\varepsilon\left(z\right)=u\left(z\right)+\varepsilon\left|z\right|^2,$$

for which we have

$$\triangle u_\varepsilon\left(z\right)=\triangle u\left(z\right)+\varepsilon\triangle\left(x^2+y^2\right)=4\varepsilon>0.$$

By Fermat's theorem, we have both $\frac{\partial^2 u_\varepsilon}{\partial x^2}\left(z\right)\leq 0$ and $\frac{\partial^2 u_\varepsilon}{\partial y^2}\left(z\right)\leq 0$ at a relative maximum z, and it follows that u_ε cannot have a relative maximum in Ω, and so must achieve its maximum on the boundary $\partial\Omega$:

$$\sup_{z\in\overline{\Omega}}u\left(z\right)\leq\sup_{z\in\overline{\Omega}}u_\varepsilon\left(z\right)\leq\sup_{z\in\partial\Omega}u_\varepsilon\left(z\right)\leq\sup_{z\in\partial\Omega}u\left(z\right)+\varepsilon\sup_{z\in\partial\Omega}\left|z\right|^2.$$

If we let $\varepsilon\to 0$ we obtain (1.6).

Now we prove the uniqueness of the solution u to the Dirichlet problem (1.5). If v is another solution, then $w=u-v$ is harmonic in $\mathbb{D}$ and vanishes on $\partial\mathbb{D}$. By the maximum principle Proposition 15, we conclude that $w \leq 0$ in $\mathbb{D}$. But $-w$ is also harmonic and vanishes on the boundary, so $-w \leq 0$ in $\mathbb{D}$ as well. Thus $u = v$.

Suppose now that $u \in C\left(\overline{\mathbb{D}}\right) \cap C^2(\mathbb{D})$ is harmonic in $\mathbb{D}$. Let $\mathbb{P}u$ be the Poisson integral of the restriction $u\,|_{\mathbb{T}}$ of u to the circle $\mathbb{T}$. Then both u and $\mathbb{P}u$ satisfy the Dirichlet problem (1.5) with $\varphi = u\,|_{\mathbb{T}}$. By uniqueness we then have

$$u(z) = \mathbb{P}u(z) = \frac{1}{2\pi}\int_0^{2\pi} u\left(e^{it}\right) P_r(\theta - t)\,dt, \qquad z = re^{i\theta} \in \mathbb{D}.$$

Thus u equals its Poisson integral for any $u \in C\left(\overline{\mathbb{D}}\right) \cap C^2(\mathbb{D})$ that is harmonic in $\mathbb{D}$. This proves (0.22) for such functions and moreover, by translating and rescaling formula (0.22) to arbitrary disks we obtain the following characterization of harmonic functions.

THEOREM 21. *Let Ω be an open subset of the plane. Then $u \in C^2(\Omega)$ is harmonic in Ω if and only if u is locally the real part of a holomorphic function, i.e. for any disk $B(a,R) \subset \Omega$, there is $f \in H(B(a,R))$ such that $u = \operatorname{Re} f$ in $B(a,R)$. Moreover, u equals its Poisson integral in any disk $B(a,R) \subset \Omega$:*

$$u\left(a + Rre^{i\theta}\right) = \frac{1}{2\pi}\int_0^{2\pi} u\left(a + Re^{it}\right) P_r(\theta - t)\,dt.$$

REMARK 7. *The above theorem is an example of a regularity theorem for a partial differential equation. It says that any C^2 solution u to Laplace's equation $\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) u = 0$ must be infinitely differentiable, in fact real-analytic. This type of phenomenon persists more generally for elliptic partial differential equations.*

2. Fatou's Theorem

Our purpose now is to extend the pointwise limit on certain radii in (0.23) to the full radial limit.

THEOREM 22. *(Fatou's Theorem) For $f \in H^2(\mathbb{D})$ we have*

$$\lim_{r\to 1} f\left(re^{it}\right) = f^*\left(e^{it}\right), \qquad a.e.\ e^{it} \in \mathbb{T}.$$

Proof: Due to (0.23), we see that if the radial limit $\lim_{r\to 1} f\left(re^{it}\right)$ exists, then it must equal $f^*\left(e^{it}\right)$ almost everywhere. So it suffices to prove that $\lim_{r\to 1} f\left(re^{it}\right)$ exists almost everywhere, equivalently that the oscillation $\omega\left(f; e^{it}\right)$ given by

$$\max\left\{\limsup_{r\to 1} u\left(re^{it}\right) - \liminf_{r\to 1} u\left(re^{it}\right), \limsup_{r\to 1} v\left(re^{it}\right) - \liminf_{r\to 1} v\left(re^{it}\right)\right\},$$

where $f = u + iv$, vanishes for almost every $e^{it} \in \mathbb{T}$. The two pieces of information at hand that we can exploit for this purpose are

- $\omega\left(f; e^{it}\right) \equiv 0$ if $f \in C\left(\overline{\mathbb{D}}\right) \cap H(\mathbb{D})$,
- $f_s \in C\left(\overline{\mathbb{D}}\right) \cap H(\mathbb{D})$ and $f_s^* \to f^*$ in $L^2(\mathbb{T})$ if $f \in H^2(\mathbb{D})$.

In order to take advantage of the second bullet item, we will need a way of controlling the oscillation of $f - f_s$ in terms of the L^2 norm $\|f - f_s\|_2$. This is accomplished with the aid of the *maximal function*: for $h \in L^1(\mathbb{T})$ we define

$$\mathcal{M}h\left(e^{i\theta}\right) \equiv \sup_{e^{i\theta} \in I \subset \mathbb{T}} \frac{1}{|I|}\int_I \left|h\left(e^{it}\right)\right| dt, \qquad e^{i\theta} \in \mathbb{T},$$

where the notation $\sup_{e^{i\theta}\in I\subset\mathbb{T}}$ means that the supremum is taken over all arcs I in the circle $\mathbb{T}$ that contain the point $e^{i\theta}$. It turns out to be easy to show that the maximal function dominates the Poisson integral in the sense that there is a positive constant C such that

$$\left|\mathbb{P}h\left(re^{i\theta}\right)\right| \leq C\mathcal{M}h\left(e^{i\theta}\right), \qquad 0 \leq r < 1. \tag{2.1}$$

Indeed, this follows readily from the following inequality with $\delta = 1 - r$ and C a positive constant;

$$P_r(\theta - t) = \frac{1-r^2}{\left|re^{i\theta} - e^{it}\right|^2} \leq C\frac{1}{2\delta}\chi_{[\theta-\delta,\theta+\delta]}(t) + C\sum_{k=1}^{\infty} 2^{-k}\frac{1}{2^{1-k}\delta}\chi_{[\theta-2^k\delta,\theta+2^k\delta]},$$

since we conclude from this that

$$\begin{aligned}
\left|\mathbb{P}h\left(re^{i\theta}\right)\right| &= \left|\frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(\theta-t)\, h\left(e^{it}\right) dt\right| \\
&\leq C\frac{1}{2\delta}\int_{\theta-\delta}^{\theta+\delta}\left|h\left(e^{it}\right)\right| dt + C\sum_{k=1}^{\infty}2^{-k}\frac{1}{2^{1-k}\delta}\int_{\theta-2^k\delta}^{\theta+2^k\delta}\left|h\left(e^{it}\right)\right| dt \\
&\leq C\mathcal{M}h\left(e^{i\theta}\right) + C\sum_{k=1}^{\infty}2^{-k}\mathcal{M}h\left(e^{i\theta}\right) \leq C\mathcal{M}h\left(e^{i\theta}\right).
\end{aligned}$$

The following Maximal Theorem is then decisive. We denote the rotation invariant probability measure σ of a subset E of $\mathbb{T} \equiv [0, 2\pi)$ by $|E|_\sigma = \int_E \frac{dt}{2\pi}$.

THEOREM 23. *(Maximal Theorem) For $h \in L^1(\mathbb{T})$ and $\lambda > 0$, we have*

$$\lambda\left|\left\{e^{i\theta}\in\mathbb{T} : \mathcal{M}h\left(e^{i\theta}\right) > \lambda\right\}\right|_\sigma \leq \frac{2}{\pi}\int_0^{2\pi}\left|h\left(e^{it}\right)\right| dt. \tag{2.2}$$

Let us assume the Maximal Theorem for the moment, and show how it and the above bullet items prove Fatou's Theorem. Fix $f \in H^2(\mathbb{D})$ and note that by the subadditivity of the oscillation ω, the bullet items and (2.1), we have

$$\begin{aligned}
\omega\left(f; e^{i\theta}\right) &\leq \omega\left(f - f_s; e^{i\theta}\right) + \omega\left(f_s; e^{i\theta}\right) = \omega\left(f - f_s; e^{i\theta}\right) \\
&\leq 2\limsup_{r\to1}\left|f - f_s\left(re^{i\theta}\right)\right| = 2\limsup_{r\to1}\left|\mathbb{P}\left(f - f_s\right)^*\left(re^{i\theta}\right)\right| \\
&\leq 2C\mathcal{M}\left(f - f_s\right)^*\left(e^{i\theta}\right) = 2C\mathcal{M}\left(f^* - f_s^*\right)\left(e^{i\theta}\right).
\end{aligned}$$

Applying the Maximal Theorem we have for any $\lambda > 0$,

$$\begin{aligned}
\left|\left\{e^{i\theta}\in\mathbb{T} : \omega\left(f; e^{i\theta}\right) > \lambda\right\}\right|_\sigma &\leq \left|\left\{e^{i\theta}\in\mathbb{T} : \mathcal{M}\left(f^* - f_s^*\right)\left(e^{i\theta}\right) > \frac{\lambda}{2C}\right\}\right|_\sigma \\
&\leq \frac{4C}{\pi\lambda}\int_0^{2\pi}\left|f^*\left(e^{it}\right) - f_s^*\left(e^{it}\right)\right| dt \\
&\leq \frac{4C\sqrt{2\pi}}{\pi\lambda}\left(\int_0^{2\pi}\left|f^*\left(e^{it}\right) - f_s^*\left(e^{it}\right)\right|^2 dt\right)^{\frac{1}{2}},
\end{aligned}$$

which tends to zero as $s \to 1$ by the second bullet item above. Thus

$$\left|\left\{e^{i\theta}\in\mathbb{T} : \omega\left(f; e^{i\theta}\right) > \lambda\right\}\right|_\sigma = 0$$

for all $\lambda > 0$ and so

$$\begin{aligned}\left|\left\{e^{i\theta} \in \mathbb{T} : \omega\left(f; e^{i\theta}\right) \neq 0\right\}\right|_\sigma &= \left|\bigcup_{n=1}^{\infty}\left\{e^{i\theta} \in \mathbb{T} : \omega\left(f; e^{i\theta}\right) > \frac{1}{n}\right\}\right|_\sigma \\ &\leq \sum_{n=1}^{\infty}\left|\left\{e^{i\theta} \in \mathbb{T} : \omega\left(f; e^{i\theta}\right) > \frac{1}{n}\right\}\right|_\sigma \\ &= \sum_{n=1}^{\infty} 0 = 0.\end{aligned}$$

This completes the proof that $\omega\left(f; e^{i\theta}\right) = 0$ for almost every $\theta \in \mathbb{T}$.

Proof of the Maximal Theorem 23: For $N \geq 1$, let $\mathcal{E}_N = \left\{\frac{j}{2^{N+2}}2\pi\right\}_{1 \leq j \leq 2^{N+2}}$ and denote by

$$\mathcal{D}_N = \{[a, b) : a, b \in \mathcal{E}_N\}$$

the collection of arcs in the circle having endpoints in the set $\mathcal{E}_N$. The point of introducing the collections $\mathcal{D}_N$ is that each one is a *finite* collection of arcs and the following swallowing property holds: given any arc $I \subset \mathbb{T}$ with $|I| \geq \frac{1}{2^N}2\pi$, there is an arc $J \in \mathcal{D}_N$ satisfying

$$I \subset J \text{ and } |J| \leq 2|I|. \tag{2.3}$$

This allows us to reduce the proof of (2.2) to the same inequality with $\mathcal{M}$ replaced by $\mathcal{M}_N$ where

$$\mathcal{M}_N h\left(e^{i\theta}\right) \equiv \sup_{e^{i\theta} \in I \in \mathcal{D}_N} \frac{1}{|I|}\int_I \left|h\left(e^{it}\right)\right| dt, \qquad e^{i\theta} \in \mathbb{T}.$$

Indeed, if we introduce the intermediate operator

$$\mathcal{M}_N^* h\left(e^{i\theta}\right) \equiv \sup_{e^{i\theta} \in I \subset \mathbb{T} : |I| \geq 2^{-N}} \frac{1}{|I|}\int_I \left|h\left(e^{it}\right)\right| dt, \qquad e^{i\theta} \in \mathbb{T},$$

then since the measure of an increasing union of sets is the limit of the measures of the sets, we have

$$\left|\left\{e^{i\theta} \in \mathbb{T} : \mathcal{M}h\left(e^{i\theta}\right) > \lambda\right\}\right|_\sigma = \lim_{N \to \infty}\left|\left\{e^{i\theta} \in \mathbb{T} : \mathcal{M}_N^* h\left(e^{i\theta}\right) > \lambda\right\}\right|_\sigma.$$

Moreover, the swallowing property (2.3) of $\mathcal{D}_N$ shows that

$$\left\{e^{i\theta} \in \mathbb{T} : \mathcal{M}_N^* h\left(e^{i\theta}\right) > \lambda\right\} \subset \left\{e^{i\theta} \in \mathbb{T} : \mathcal{M}_N h\left(e^{i\theta}\right) > \frac{\lambda}{2}\right\},$$

since if $\frac{1}{|I|}\int_I \left|h\left(e^{it}\right)\right| dt > \lambda$ where $|I| \geq 2^{-N}$, then there is $J \in \mathcal{D}_N$ satisfying (2.3), so that

$$\frac{1}{|J|}\int_J \left|h\left(e^{it}\right)\right| dt \geq \frac{|I|}{|J|}\frac{1}{|I|}\int_I \left|h\left(e^{it}\right)\right| dt > \frac{1}{2}\lambda.$$

Altogether then it suffices to prove

$$\lambda\left|\left\{e^{i\theta} \in \mathbb{T} : \mathcal{M}_N h\left(e^{i\theta}\right) > \lambda\right\}\right|_\sigma \leq \frac{1}{\pi}\int_0^{2\pi} \left|h\left(e^{it}\right)\right| dt. \tag{2.4}$$

The inequality (2.4) is proved by a simple "covering lemma": if $I = \bigcup_{m=1}^{M} I_m$, $I_m \in \mathcal{D}_N$, then there exists a subcollection $\{I_{m_a}\}_{a=1}^{A}$ satisfying

$$I = \bigcup_{a=1}^{A} I_{m_a} \text{ and } \sum_{a=1}^{A} \chi_{I_{m_a}} \leq 2, \tag{2.5}$$

i.e. the subcollection $\{I_{m_a}\}_{a=1}^{A}$ covers I with overlap at most 2. To see this let $I = [a, b)$ and pick $I_{m_1} = [a, b_{m_1})$ to be a largest interval with endpoint a. Then choose $I_{m_2} = [a_{m_2}, b_{m_2})$ to be an interval containing b_{m_1} with largest endpoint b_{m_2}. Inductively choose $I_{m_{a+1}} = \left[a_{m_{a+1}}, b_{m_{a+1}}\right)$ to be an interval containing b_{m_a} with largest endpoint $b_{m_{a+1}}$. This procedure ends in finitely many steps A and it is clear that (2.5) holds with overlap 2 since if $I_{m_{a-1}} \cap I_{m_{a+1}} \neq \phi$, then we would have chosen $I_{m_{a+1}}$ in place of I_{m_a} at the a^{th} inductive step.

Now we note that

$$\left\{e^{i\theta} \in \mathbb{T} : \mathcal{M}_N h\left(e^{i\theta}\right) > \lambda\right\} = \bigcup \left\{I \in \mathcal{D}_N : \frac{1}{|I|}\int_I \left|h\left(e^{it}\right)\right| dt > \lambda\right\} = \bigcup_{j=1}^{J} I^j,$$

where $I^j = [a_j, b_j)$ with $b_j < a_{j+1}$ and

$$I^j = \bigcup_{m=1}^{M_j} I^j_m \text{ where } \frac{1}{\left|I^j_m\right|}\int_{I^j_m} \left|h\left(e^{it}\right)\right| dt > \lambda.$$

Apply the covering lemma to I^j to obtain $I^j = \bigcup_{a=1}^{A_j} I^j_{m_a}$ satisfying (2.5). Then we have

$$\begin{aligned}
\left|\left\{e^{i\theta} \in \mathbb{T} : \mathcal{M}_N h\left(e^{i\theta}\right) > \lambda\right\}\right|_\sigma &= \frac{1}{2\pi}\sum_{j=1}^{J}\left|I^j\right| \leq \frac{1}{2\pi}\sum_{j=1}^{J}\sum_{a=1}^{A_j}\left|I^j_{m_a}\right| \\
&< \frac{1}{2\pi}\sum_{j=1}^{J}\sum_{a=1}^{A_j}\frac{1}{\lambda}\int_{I^j_{m_a}}\left|h\left(e^{it}\right)\right| dt \\
&= \frac{1}{2\pi}\sum_{j=1}^{J}\frac{1}{\lambda}\int\left(\sum_{a=1}^{A_j}\chi_{I^j_{m_a}}(t)\right)\left|h\left(e^{it}\right)\right| dt \\
&\leq \frac{1}{2\pi}\sum_{j=1}^{J}\frac{2}{\lambda}\int_{I^j}\left|h\left(e^{it}\right)\right| dt \leq \frac{1}{\pi\lambda}\int_0^{2\pi}\left|h\left(e^{it}\right)\right| dt.
\end{aligned}$$

This completes the proof of the Maximal Theorem 23 and with it, the proof of Fatou's Theorem 22.

CHAPTER 7

Extending Riemann maps

Fatou's theorem plays a crucial role below in extending a Riemann map $f : \Omega \to \mathbb{D}$ to a simple boundary point $w \in \partial\Omega$. However, we will also need to know that such extensions to *distinct* boundary points w_1, w_2 have *distinct* images $f(w_1), f(w_2)$. For this we need one final result concerning radial limits, namely Lindelöf's theorem.

1. Lindelöf's theorem

The next theorem shows that if a bounded holomorphic function g in the disk $\mathbb{D}$ has limit L along some curve ending at $e^{i\theta} \in \mathbb{T}$, then we can conclude that g has *radial* limit L at $e^{i\theta}$.

THEOREM 24. *(Lindelöf's Theorem) Suppose* $\Gamma : [0,1] \to \mathbb{D} \cup \{1\}$ *is a continuous curve such that* $|\Gamma(t)| < 1$ *if* $t < 1$ *and* $\Gamma(1) = 1$*. Then if* $g \in H^\infty(\mathbb{D})$ *satisfies*

$$\lim_{t \to 1} g(\Gamma(t)) = L,$$

it follows that g *has radial limit* L *at* 1*:*

$$\lim_{r \to 1} g(r) = L.$$

Proof: Without loss of generality we assume that $L = 0$ and $\|g\|_\infty < 1$. Let $\varepsilon > 0$ be given. We will show there is $r_0 < 1$ such that

$$|g(r)| \leq \sqrt[4]{\varepsilon}, \qquad r_0 < r < 1.$$

First we note that since $\lim_{t \to 1} g(\Gamma(t)) = 0$, there is $a < 1$ such that $\operatorname{Re}\Gamma(a) > \frac{1}{2}$ and $|g(\Gamma(t))| < \varepsilon$ whenever $a < t < 1$. Now let $r_0 = \operatorname{Re}\Gamma(a)$ and

$$t_0 = \sup\{0 \leq t < 1 : \operatorname{Re}\Gamma(t) = r_0\},$$

so that we have

$$|g(\Gamma(t))| < \varepsilon \text{ and } \operatorname{Re}\Gamma(t) > r_0 > \frac{1}{2}, \qquad t_0 < t < 1.$$

Fix $r \in (r_0, 1)$ for the moment and define a lens-shaped region Ω centered at r by

$$\Omega = \mathbb{D} \cap \{\mathbb{D} + 2r\} = B(0,1) \cap B(2r,1),$$

and a holomorphic function $h \in H(\Omega)$ by

$$h(z) = g(z)\,\overline{g(\overline{z})g(2r - \overline{z})}g(2r - z).$$

Note that h is a product of four holomorphic functions

$$h = g_1 g_2 g_3 g_4,$$

namely $g_1(z) = g(z)$, a "reflection" $g_2(z) = \overline{g(\overline{z})}$ of $g_1(z)$ across the real axis in the disk $\mathbb{D}$, a "reflection" $g_3(z) = g_2(2r - z)$ of $g_2(z)$ across the vertical axis $\operatorname{Re} z = r$

to the disk $\mathbb{D}+2r$, and a "reflection" $g_4(z)=\overline{g_3(\overline{z})}$ of $g_3(z)$ across the real axis in the disk $\mathbb{D}+2r$. Clearly we have $|h(z)|<1$ for $z\in\Omega$. Since $h(r)=|g(r)|^4$ we will be done once we show that

$$h(r)\le\varepsilon. \tag{1.1}$$

We will use the maximum modulus principle, together with the geometry of the curve Γ and its reflections and translations, to obtain (1.1). Let

$$t_1=\sup\{0\le t<1:\operatorname{Re}\Gamma(t)=r\}\in(t_0,1),$$

and define $E_1=\Gamma([t_1,1])$ to be the closed arc of Γ^* that connects $\Gamma(t_1)$ to 1 and has the property that E_1 minus its endpoints lies in the open right half of Ω. Let E_2 be the reflection of E_1 across the real axis, E_3 be the reflection of E_2 across the vertical axis $\operatorname{Re}z=r$ of symmetry of Ω, and finally let E_4 be the reflection of E_3 across the real axis, which coincides with the reflection of E_1 across the vertical axis $\operatorname{Re}z=r$. Note that E_1 and E_2 can have a complicated intersection, but that $E_1\cup E_2$ and $E_3\cup E_4$ have only the points $\Gamma(t_1)$ and $\overline{\Gamma(t_1)}$ in common. Then set

$$E=E_1\cup E_2\cup E_3\cup E_4.$$

We have $E\subset\Omega^*\equiv\Omega\cup\{1,2r-1\}$. A crucial property of the function h on the set E is

$$|h(z)|<\varepsilon,\qquad z\in E. \tag{1.2}$$

Indeed, for $z\in E$, we must have $z\in E_i$ for some i, and then the corresponding factor g_i satisfies $|g_i(z)|<\varepsilon$ while the other factors g_j satisfy $|g_j(z)|<1$. Now pick a small $\alpha>0$ and define

$$h_\alpha(z)=\begin{cases} h(z)(1-z)^\alpha(2r-1-z)^\alpha & \text{for} \quad z\in\Omega\\ 0 & \text{for} \quad z=1,2r-1\end{cases}.$$

The point of h_α is that it is holomorphic in Ω and continuous on $\Omega^*=\Omega\cup\{1,2r-1\}$ for $\alpha>0$.

Now let K be the union of the compact set E and the bounded components of the open set $\mathbb{C}\setminus E$. Then K is compact and h_α satisfies the following properties on K:

(1) h_α is continuous on K,
(2) h_α is holomorphic in the interior $\overset{\circ}{K}$ of K,
(3) $|h_\alpha|<\varepsilon$ on the boundary ∂K of K by (1.2).

The maximum principle now shows that $|h_\alpha|<\varepsilon$ on K since the boundary of any connected component of $\mathbb{C}\setminus E$ lies in E.

We claim that $r\in K$ by the construction of E. Indeed, if $r\in E$ we are done, so we assume r lies in the open set $\mathbb{C}\setminus E$. Now consider the closed curve β obtained by concatenating the curve $\Gamma_1:[t_1,1]\to\Omega^*$ with its three reflections Γ_2, Γ_3 and Γ_4, taken in the appropriate direction, about the real axis and the axis of symmetry $\operatorname{Re}z=r$. Then $\beta^*=E$ and we assume, in order to derive a contradiction, that r lies in the unbounded component of $\mathbb{C}\setminus\beta^*=\mathbb{C}\setminus E$. Then there is a path σ in $\mathbb{C}\setminus E$ joining r to the number 3, which is well outside the convex set Ω. Let $\delta>0$ satisfy both

$$B(r,2\delta)\subset\mathbb{C}\setminus E \text{ and } \delta<\frac{1}{2}dist(\sigma^*,E).$$

Now choose a closed polygonal path $\gamma(t)$ that joins consecutive (sufficiently close) points on β^* with line segments, so that

$$|\beta(t) - \gamma(t)| < \frac{\delta}{10}. \tag{1.3}$$

Then σ^* is disjoint from γ^* and r lies in the unbounded component of $\mathbb{C} \setminus \gamma^*$. Proposition 4 shows that

$$Ind_\gamma(r) = 0. \tag{1.4}$$

On the other hand, it is not hard to see that γ is $(\mathbb{C} \setminus \{r\})$-homotopic to $\partial B(r, \delta)$, taken in the positive direction, simply by following in sequence those portions $\gamma_1(t)$, $\gamma_2(t)$, $\gamma_3(t)$ and $\gamma_4(t)$ of $\gamma(t)$ when $\beta(t)$ is given by $\Gamma_1(t)$, $\Gamma_2(t)$, $\Gamma_3(t)$ and $\Gamma_4(t)$ respectively. In fact, if $\alpha_1(t)$, $\alpha_2(t)$, $\alpha_3(t)$ and $\alpha_4(t)$ denote the corresponding quarter arcs of $\partial B(r, \delta)$ taken in the same sequence, then one can use the homotopies $\mathcal{H}_i(t, \theta) = (1 - \theta)\gamma_i(t) + \theta\alpha_i(t)$ which avoid the point r since both γ_i^* and α_i^* lie in a common half plane that doesn't contain r. Thus Proposition 9 shows that

$$Ind_\gamma(r) = Ind_{\partial B(r,\delta)}(r) = 1,$$

which contradicts (1.4) and shows that r must lie in a *bounded* component of $\mathbb{C} \setminus E$, hence in K.

We thus conclude that $|h_\alpha(r)| < \varepsilon$. Now let $\alpha \to 0$ to obtain $|h(r)| \leq \varepsilon$, which is (1.1), and this completes the proof of Lindelöf's Theorem 24.

2. Proof of Carathéodory's Theorem

We can now prove Carathéodory's Theorem 20. First we will use Fatou's Theorem 22 to prove Lemma 5 below, and then we will use Lindelöf's Theorem 24 to prove Lemma 6 below.

LEMMA 5. *Let Ω be a bounded simply connected domain in the complex plane, and let $f : \Omega \to \mathbb{D}$ be holomorphic, one-to-one and onto. Suppose that w is a simple boundary point of Ω. Then f has a continuous extension $f : \Omega \cup \{w\} \to \overline{\mathbb{D}}$ and $f(w) \in \mathbb{T}$.*

LEMMA 6. *Let Ω be a bounded simply connected domain in the complex plane, and let $f : \Omega \to \mathbb{D}$ be holomorphic, one-to-one and onto. Suppose that w_1 and w_2 are* distinct *simple boundary points of Ω, and that $f : \Omega \cup \{w_1, w_2\} \to \overline{\mathbb{D}}$ is as in Lemma 5. Then $f(w_1) \neq f(w_2)$.*

Proof of Lemma 5: Let $g = f^{-1}$ so that $g : \mathbb{D} \to \Omega$ is one-to-one and onto, and $g \in H^\infty(\mathbb{D})$. Let $\{w_n\}_{n=0}^\infty \subset \Omega$ be a sequence with limit w and such that

$$\lim_{n\to\infty} f(w_{2n}) = \zeta_0 \text{ and } \lim_{n\to\infty} f(w_{2n+1}) = \zeta_1.$$

Suppose, in order to derive a contradiction, that $\zeta_0 \neq \zeta_1$. Since w is a simple boundary point of Ω, there is a curve $\gamma : [0, 1] \to \Omega \cup \{w\}$ as in Definition 11 that passes through the sequence $\{w_n\}_{n=0}^\infty$ and ends at w. Set

$$\Gamma(t) = f(\gamma(t)), \qquad 0 \leq t < 1.$$

Now for $0 < r < 1$, $g\left(r\overline{\mathbb{D}}\right)$ is a compact subset of Ω disjoint from w. Thus there is $t_r < 1$ depending on r such that $\gamma(t) \notin g\left(r\overline{\mathbb{D}}\right)$ if $t_r < t < 1$. It follows that

$|\Gamma(t)| > r$ for $t_r < t < 1$, and so $\lim_{t\to 1} |\Gamma(t)| = 1$. In particular, if $w_n = f(t_n)$, then both

$$|\zeta_0| = \lim_{t\to 1} |\Gamma(t_{2n})| = 1 \text{ and } |\zeta_1| = \lim_{t\to 1} |\Gamma(t_{2n+1})| = 1. \tag{2.1}$$

Let I_1 and I_2 be the two open arcs of $\mathbb{T}$ whose union is $\mathbb{T} \setminus \{\zeta_0, \zeta_1\}$. At least one of these arcs, call it J, has the property that every segment S_θ from the origin to a point $e^{i\theta}$ in J intersects the range of Γ in an infinite subset of S_θ that has $e^{i\theta}$ as a limit point. Indeed, if not, there would be two segments S_{θ_1} and S_{θ_2} ending in I_1 and I_2 respectively such that for a sufficiently large $T < 1$, $\Gamma(t)$ is disjoint from $S_{\theta_1} \cup S_{\theta_2}$ for all $T < t < 1$. But this contradicts the connectedness of $\Gamma((T,1))$ since both ζ_0 and ζ_1 lie in the closure of $\Gamma((T,1))$.

Thus for every $e^{i\theta} \in J$ at which g has a radial limit, we have

$$\lim_{r\to 1} g\left(re^{i\theta}\right) = w.$$

Now $g \in H^\infty(\mathbb{D})$ and so by Fatou's Theorem 22, $g^*\left(e^{i\theta}\right) = w$ for almost every $e^{i\theta} \in J$. The uniqueness result Lemma 3 now shows that $g - w \in H^\infty(\mathbb{D})$ is identically zero, the desired contradiction since g is one-to-one on $\mathbb{D}$.

Thus $\zeta_0 = \zeta_1 \in \mathbb{T}$, and we conclude that f has a continuous extension to $\Omega \cup \{w\}$, and that $f(w) \in \mathbb{T}$.

Proof of Lemma 6: We prove the contrapositive: $w_1 = w_2$ if $f(w_1) = f(w_2)$. We may suppose that $f(w_1) = f(w_2) = 1$. Since w_i is a simple boundary point of Ω, there is a curve $\gamma_i : [0,1] \to \Omega \cup \{w_i\}$ with $\gamma_i([0,1)) \subset \Omega$ and $\gamma_i(1) = w_i$. Set

$$\Gamma_i(t) = f(\gamma_i(t)), \qquad 0 \le t \le 1.$$

Then $\Gamma_i([0,1)) \subset \mathbb{D}$ and $\Gamma_i(1) = 1$. Since $g(\Gamma_i(t)) = \gamma_i(t)$ for $0 \le t < 1$ we have

$$\lim_{t\to 1} g(\Gamma_i(t)) = \lim_{t\to 1} \gamma_i(t) = w_i.$$

Thus Lindelof's Theorem 24 implies that

$$w_i = \lim_{r\to 1} g(r),$$

for both $i = 1$ and $i = 2$, hence $w_1 = w_2$.

The proof that (3) implies (1) in Theorem 20 is now easily accomplished in a few lines. Suppose (3) holds and that $f : \Omega \to \mathbb{D}$ is holomorphic and one-to-one. Lemma 5 shows that there is an extension $f : \overline{\Omega} \to \overline{\mathbb{D}}$ such that $f(w_n) \to f(w)$ whenever $\{w_n\}_{n=1}^\infty \subset \Omega$ is a sequence in Ω converging to w. If $\{z_n\}_{n=1}^\infty \subset \overline{\Omega}$ is a sequence in $\overline{\Omega}$ converging to w, then there exist points $w_n \in \Omega$ such that

$$|z_n - w_n| < \frac{1}{n} \text{ and } |f(z_n) - f(w_n)| < \frac{1}{n}.$$

Thus $\{w_n\}_{n=1}^\infty \subset \Omega$ is a sequence in Ω converging to w, and so $f(w_n) \to f(w)$. But then $f(z_n) \to f(w)$ as well. This proves that $f : \overline{\Omega} \to \overline{\mathbb{D}}$ is continuous.

Since $\mathbb{D} \subset f\left(\overline{\Omega}\right) \subset \overline{\mathbb{D}}$ and $f\left(\overline{\Omega}\right)$ is compact, hence closed, we have $f\left(\overline{\Omega}\right) = \overline{\mathbb{D}}$ and so $f : \overline{\Omega} \to \overline{\mathbb{D}}$ is onto.

Lemma 6 shows that $f : \overline{\Omega} \to \overline{\mathbb{D}}$ is one-to-one.

Finally, it is a standard result that $f : \overline{\Omega} \to \overline{\mathbb{D}}$ is now a homeomorphism. Indeed, $f^{-1} : \overline{\mathbb{D}} \to \overline{\Omega}$ is continuous since if G is open in $\overline{\Omega}$, then $\overline{\Omega} \setminus G$ is compact, so $f\left(\overline{\Omega} \setminus G\right) = \overline{\mathbb{D}} \setminus f(G)$ is compact and hence closed, so $f(G) = \left(f^{-1}\right)^{-1}(G)$ is open.

2.1. Simple boundary points and Jordan curves. Now it is obvious that (1) implies (2) in Carathéodory's Theorem 20, and so we turn our attention to the remaining implication, namely that if $\partial\Omega$ is a Jordan curve, then every boundary point of Ω is a simple boundary point. Our proof of this fact is largely topological, in contrast to the converse implication that we proved using the full force of the Riemann mapping theorem and its boundary behaviour.

So we suppose that $\partial\Omega = \beta^*$ where $\beta : \mathbb{T} \to \mathbb{C}$ is continuous and one-to-one. Since Ω is bounded and connected, we must have $\Omega = \mathcal{B}_\beta$ where $\mathcal{B}_\beta$ is the bounded component of $\mathbb{C}\backslash\beta^* = \mathbb{C}\backslash\partial\Omega$. Indeed, fix $z_0 \in \Omega$ and suppose that $\gamma : [0,1] \to \mathbb{C}\backslash\partial\Omega$ is a simple taxicab path joining z_0 to z. If $z \in \Omega^c \setminus \partial\Omega$, and $t_1 = \inf\{t \in [0,1] : \gamma(t) \in \Omega^c \setminus \partial\Omega\}$, then $\gamma(t_1) \in \partial\Omega$, a contradiction. So $z \in \Omega$. It now follows that Ω is a connected component of $\mathbb{C}\backslash\beta^*$, and hence must be the bounded component $\mathcal{B}_\beta$. Now it follows from Proposition 18 in the appendix below that every boundary point of $\partial\Omega$ is simple.

APPENDIX A

Topology

We collect here some background material requiring contributions from topology.

1. Homotopy and index

Here we prove Proposition 9 that says the index is unchanged by homotopy. We begin with a short lemma that gives a condition on two paths under which one of them cannot wrap around a point more often than the other. The reader who has walked a dog in a park will recognize this condition as shortening the leash near a pole or tree just enough to prevent the dog from winding around it.

Lemma 7. *(Dog leash lemma) Suppose that $\gamma_0 : \mathbb{T} \to \mathbb{C}$ and $\gamma_1 : \mathbb{T} \to \mathbb{C}$ are closed paths in the complex plane, and that a is a complex number such that*

$$\left|\gamma_1(\zeta) - \gamma_0(\zeta)\right| < \left|a - \gamma_0(\zeta)\right|, \qquad \zeta \in \mathbb{T}.$$

Then $Ind_{\gamma_0}(a) = Ind_{\gamma_1}(a)$.

Proof: Let $\mathbb{T} = [0, 2\pi]$ with 0 and 2π identified, and define

$$\gamma(t) = \frac{\gamma_1(t) - a}{\gamma_0(t) - a}, \qquad 0 \leq t \leq 2\pi.$$

Then

$$\left|\gamma(t) - 1\right| = \left|\frac{\gamma_1(t) - a}{\gamma_0(t) - a} - 1\right| = \left|\frac{\gamma_1(t) - \gamma_0(t)}{\gamma_0(t) - a}\right| < 1$$

implies that $\gamma^* \subset B(1,1)$ and so $Ind_\gamma(0) = 0$ by Cauchy's theorem. Thus we have

$$\begin{aligned}
0 &= Ind_\gamma(0) = \frac{1}{2\pi i}\int_\gamma \frac{1}{z-0}dz = \frac{1}{2\pi i}\int_0^{2\pi} \frac{1}{\gamma(t)}\gamma'(t)\,dt \\
&= \frac{1}{2\pi i}\int_0^{2\pi} \frac{\gamma_0(t) - a}{\gamma_1(t) - a}\frac{(\gamma_0(t) - a)\gamma_1'(t) - (\gamma_1(t) - a)\gamma_0'(t)}{(\gamma_0(t) - a)^2}dt \\
&= \frac{1}{2\pi i}\int_0^{2\pi} \frac{\gamma_1'(t)}{\gamma_1(t) - a}dt - \frac{1}{2\pi i}\int_0^{2\pi} \frac{\gamma_0'(t)}{\gamma_0(t) - a}dt \\
&= Ind_{\gamma_1}(a) - Ind_{\gamma_0}(a).
\end{aligned}$$

Now we recall Proposition 9.

Proposition 16. *Let γ_0 and γ_1 be two closed paths in an open set Ω of the complex plane. If γ_0 and γ_1 are Ω-homotopic, then $Ind_{\gamma_0}(a) = Ind_{\gamma_1}(a)$ for all $a \in \mathbb{C} \setminus \Omega$.*

Proof: There is a continuous map $\Gamma : \mathbb{T} \times [0,1] \rightarrow \Omega$ such that $\Gamma(\zeta, 0) = \gamma_0$ and $\Gamma(\zeta, 1) = \gamma_1$. Choose

$$0 < \varepsilon < \frac{1}{4} dist\,(a, \Omega)\,. \tag{1.1}$$

By uniform continuity of Γ on the compact set $\mathbb{T} \times [0,1]$ there is $\delta > 0$ such that

$$|\Gamma(\zeta_1, t_1) - \Gamma(\zeta_2, t_2)| < \varepsilon \text{ whenever } |(\zeta_1, t_1) - (\zeta_2, t_2)| < \delta.$$

If we now choose points $0 = t_1 < t_2 ... < t_N = 1$ with

$$\triangle t_k = t_k - t_{k-1} < \delta,$$

and if we define closed curves $\beta_k : \mathbb{T} \rightarrow \Omega$ by

$$\beta_k(\zeta) = \Gamma(\zeta, t_k)\,, \qquad 1 \leq k \leq N,$$

then β_1 is γ_0 and β_N is γ_1; moreover,

$$\left|\beta_{k+1}(\zeta) - \beta_k(\zeta)\right| = |\Gamma(\zeta, t_{k+1}) - \Gamma(\zeta, t_k)| < \varepsilon, \tag{1.2}$$

since $|(\zeta, t_{k+1}) - (\zeta, t_k)| = |t_{k+1} - t_k| < \delta$. Using (1.1) we then have

$$\left|\beta_{k+1}(\zeta) - \beta_k(\zeta)\right| < \frac{1}{4} dist\,(a, \Omega) < \frac{1}{4}\left|a - \beta_k(\zeta)\right|. \tag{1.3}$$

If it were the case that the curves β_k were actually *paths*, then Lemma 7 and inequality (1.3) (even without the factor $\frac{1}{4}$) would yield

$$Ind_{\beta_{k+1}}(a) = Ind_{\beta_k}(a)\,, \qquad 1 \leq k < N,$$

which would prove that

$$Ind_{\gamma_1}(a) = Ind_{\beta_N}(a) = Ind_{\beta_{N-1}}(a) = ... = Ind_{\beta_1}(a) = Ind_{\gamma_0}(a)\,.$$

Of course there is no reason to assume that the β_k are paths, so we must make a further approximation. Write $\mathbb{T} = [0, 2\pi]$ with 0 and 2π identified, and choose

$$0 = \theta_0 < \theta_1 < ... < \theta_M = 2\pi$$

so that

$$\triangle\theta_\ell = \theta_\ell - \theta_{\ell-1} < \delta.$$

For $1 \leq k \leq N$ we define $\alpha_k : \mathbb{T} \rightarrow \Omega$ to be the polygonal path obtained from β_k by joining each point $\beta_k(\theta_{\ell-1})$ to $\beta_k(\theta_\ell)$ by a straight line segment with the usual parameterization. Then for $\theta \in [\theta_{\ell-1}, \theta_\ell]$ and $1 \leq k \leq N$ we have

$$\begin{aligned}
&|\alpha_k(\theta) - \beta_k(\theta)| \\
= \;& \left|\frac{\theta_\ell - \theta}{\triangle\theta_\ell}\beta_k(\theta_{\ell-1}) + \frac{\theta - \theta_{\ell-1}}{\triangle\theta_\ell}\beta_k(\theta_\ell) - \beta_k(\theta)\right| \\
\leq \;& \frac{\theta_\ell - \theta}{\triangle\theta_\ell}\left|\beta_k(\theta_{\ell-1}) - \beta_k(\theta)\right| + \frac{\theta - \theta_{\ell-1}}{\triangle\theta_\ell}\left|\beta_k(\theta_\ell) - \beta_k(\theta)\right| \\
= \;& \frac{\theta_\ell - \theta}{\triangle\theta_\ell}\left|\Gamma(\theta_{\ell-1}, t_k) - \Gamma(\theta, t_k)\right| + \frac{\theta - \theta_{\ell-1}}{\triangle\theta_\ell}\left|\Gamma(\theta_\ell, t_k) - \Gamma(\theta, t_k)\right| \\
< \;& \frac{\theta_\ell - \theta}{\triangle\theta_\ell}\varepsilon + \frac{\theta - \theta_{\ell-1}}{\triangle\theta_\ell}\varepsilon = \varepsilon.
\end{aligned} \tag{1.4}$$

Altogether, from (1.2) and (1.4) we obtain for $1 \leq k < N$,

$$\begin{aligned}
|\alpha_{k+1}(\theta) - \alpha_k(\theta)| &\leq \left|\alpha_{k+1}(\theta) - \beta_{k+1}(\theta)\right| + \left|\beta_{k+1}(\theta) - \beta_k(\theta)\right| + |\beta_k(\theta) - \alpha_k(\theta)| \\
&< \varepsilon + \varepsilon + \varepsilon = 3\varepsilon < dist(a, \Omega) - \varepsilon \\
&< |a - \beta_k(\theta)| - \varepsilon \leq |a - \alpha_k(\theta)| + |\alpha_k(\theta) - \beta_k(\theta)| - \varepsilon \\
&\leq |a - \alpha_k(\theta)|.
\end{aligned}$$

Thus Lemma 7 applies to show that

$$Ind_{\alpha_{k+1}}(a) = Ind_{\alpha_k}(a), \qquad 1 \leq k < N,$$

as well as

$$Ind_{\alpha_1}(a) = Ind_{\beta_1}(a) \text{ and } Ind_{\alpha_N}(a) = Ind_{\beta_N}(a).$$

We conclude that

$$\begin{aligned}
Ind_{\gamma_1}(a) &= Ind_{\beta_N}(a) \\
&= Ind_{\alpha_N}(a) = Ind_{\alpha_{N-1}}(a) = ... = Ind_{\alpha_1}(a) \\
&= Ind_{\beta_1}(a) = Ind_{\gamma_0}(a).
\end{aligned}$$

2. The Jordan Curve Theorem

If $\beta : \mathbb{T} \to \mathbb{C}$ is a simple closed curve, by which we mean that β is continuous and one-to-one, the *Jordan Curve Theorem* says that the complement $\mathbb{C} \setminus \beta^*$ of the image β^* has exactly two connected components, one unbounded and the other bounded and simply connected.

THEOREM 25. *Suppose $\beta : \mathbb{T} \to \mathbb{C}$ is continuous and one-to-one. Then $\mathbb{C} \setminus \beta^* = \mathcal{U} \cup \mathcal{B}$ where*

(1) *$\mathcal{U}$ is unbounded and connected,*
(2) *$\mathcal{B}$ is bounded, connected and simply connected,*
(3) *$\mathcal{U} \cap \mathcal{B} = \phi$.*

We first establish this theorem for closed taxicab paths, where we follow the arguments for paths in [**6**], but with some simplifications permitted by the restriction to taxicab paths. The reader is referred to Maehara [**3**] for a different proof using the Brouwer fixed point theorem in the plane (for which see e.g. page 31 in [**2**]). However, our proof also yields that every point on a Jordan curve is a *simple* boundary point of each connected component of the complement. See Proposition 18 below.

PROPOSITION 17. *Theorem 25 holds if in addition β is assumed to be a taxicab path, i.e. β^* is a finite concatenation of line segments parallel to either the real axis or the imaginary axis. In this case Ind_β is either 1 or -1 throughout the bounded component $\mathcal{B}$.*

Proof: Let $\{S_1, S_2, ..., S_n\}$ be a list of the distinct line segments in β^* where $S_k = [a_k, b_k]$ is the closed line segment joining a_k to b_k, and suppose that

$$S_k \cap S_{k+1} = \{b_k\} = \{a_{k+1}\}, \qquad 1 \leq k \leq n,$$

where we define $S_{n+1} = S_1$. We proceed in three steps.

Step 1: $\mathbb{C} \setminus \beta^*$ is *not* connected.

Let a be the midpoint of the segment S_1 and choose $r > 0$ so small that $\overline{B(a,r)} \cap \{\bigcup_{k=2}^n S_k\} = \phi$. Now

- $\partial B(a,r) \cap S_1$ consists of two points b and c,
- $B(a,r) \cap S_1$ consists of an open line segment L endpoints b and c,
- $\partial B(a,r) \setminus S_1$ consists of two open semicircles C_1 and C_2 each of whose endpoints are b and c.

For $j=1,2$ define the path β_j to be the path β but with the open line segment L replaced by the open semicircle C_j. Then β_j is still a closed path since the endpoints of L and C_j coincide (the direction on C_j is chosen to match that on L). We then have

$$Ind_{\beta_1}(a) - Ind_{\beta_2}(a) = \pm Ind_{\partial B(a,r)}(a) = \pm 1, \tag{2.1}$$

where the sign $\pm$ is determined by the direction on L.

It now follows from Cauchy's theorem that Ind_β has the same value as Ind_{β_1} on the semicircle C_2, and that Ind_β has the same value as Ind_{β_2} on the semicircle C_1. Since these two values differ by exactly 1 or -1, it now follows from Proposition 4 that

$$\text{the semicircles } C_1 \text{and } C_2 \text{ lie in } \textit{different} \text{ components of } \mathbb{C} \setminus \beta^*, \tag{2.2}$$

and thus that there are *at least two* connected components in $\mathbb{C} \setminus \beta^*$.

Step 2: $\mathbb{C} \setminus \beta^*$ has *exactly two* connected components.

Let

$$\delta = \min \{ dist(S_i, S_j) : S_i \text{ and } S_j \text{ are not consecutive segments in } \beta^* \}.$$

Since β is continuous and one-to-one, we have $\delta > 0$. Now take $0 < t < \frac{\delta}{4}$ and consider the set

$$E_t = \{ z \in \mathbb{C} : d_{taxi}(z, \beta^*) = t \},$$

where $d_{taxi}(z,w) = \max\{|z_1 - w_1|, |z_2 - w_2|\}$ is the taxicab distance between z and w.

To fix notation let us parameterize β by arc length s (this is particularly easy for a taxicab path). Starting at the point a in Step 1, we (where we identify with the point $\beta(s)$ on the path β) begin travelling along S_1 in the positive direction (as determined by the map β) and note that, at least initially since $t < \frac{\delta}{4}$ and $\delta \le length(S_1)$, there are exactly two points in E_t that are at (Euclidean) distance t from $\beta(s)$, one to the "right" of $\beta(s)$, call it $\beta^t_{right}(s)$, and one to the "left" of $\beta(s)$, call it $\beta^t_{left}(s)$. As we approach to within distance t of the segment S_2, which we may assume veers to the right from the endpoint b_1 of the segment S_1, we *halt* the point $\beta^t_{right}(s)$, keeping it constant as we go around the right angle at b_1 traversing a distance t along S_1 to b_1 followed by a distance t from b_1 along S_2. As for the point $\beta^t_{left}(s)$, we continue on until $\beta(s)$ reaches b_1, and then we pause momentarily to let β^t_{left} continue on for a distance t and then turn right for a distance t, allowing it to "catch up" to $\beta^t_{left}(s)$ as $\beta(s)$ begins traversing S_2. A picture makes this quite transparent.

In this way we continue to construct β^t_{right} and β^t_{left} until $\beta(s)$ returns to the point $\beta(a)$ on the curve β at which we started. By construction both β^t_{left} and β^t_{right} are taxicab paths lying in the set E_t. If we parameterize each of these paths by arc length, and denote by L_{left} and L_{right} their respective lengths, then we have

$$\begin{aligned} \beta^t_{left} &: [0, L_{left}] \to E_t \subset \mathbb{C} \setminus \beta^*, \\ \beta^t_{right} &: [0, L_{right}] \to E_t \subset \mathbb{C} \setminus \beta^*. \end{aligned}$$

It is clear that

$$\left(\beta_{left}^{t}\right)^{*} \cup \left(\beta_{right}^{t}\right)^{*} = E_{t}, \tag{2.3}$$

and furthermore that only the first of the two possibilities listed below can occur when the paths β_{left}^{t} and β_{right}^{t} return:

Either: (railway track matchup) $\beta_{left}^{t}(L_{left}) = \beta_{left}^{t}(0)$ and $\beta_{right}^{t}(L_{right}) = \beta_{right}^{t}(0)$,

Or: (Möbius band matchup) $\beta_{left}^{t}(L_{left}) = \beta_{right}^{t}(0)$ and $\beta_{right}^{t}(L_{right}) = \beta_{left}^{t}(0)$.

The reason the first possibility occurs is that the path β_{right}^{t} always stays to the "right" of $\beta(s)$ as s moves in the positive direction. Alternatively, the Möbius band matchup *cannot* occur in the plane since otherwise it would follow that the two semicircles C_1 and C_2 in Step 1 (with $r = t$) would lie in the same connected component of $\mathbb{C} \setminus \beta^{*}$, contradicting (2.2).

Now we claim that $\mathbb{C}\setminus\beta^{*}$ is the union of the component containing the semicircle C_1 (in Step 1) and the component containing the semicircle C_2 (in Step 1). Indeed, if $z \in \mathbb{C}\setminus\beta^{*}$, pick a in one of the segments comprising β^{*} such that the line segment $[z,a]$ is neither horizontal nor vertical. Let $u \in (z,a] \cap \beta^{*}$ be the first point in β^{*} encountered by $[z,a]$ as it travels from z to a. It is now clear that $[z,u]$ must intersect E_t for a *sufficiently small* positive t. Indeed, if we zoom in at a, we just need to know that two non-parallel lines must intersect. This is one of Euclid's axioms. It is interesting to note that in the proof of the Jordan Curve Theorem in [**3**], the Brouwer fixed point theorem is used to prove the analogue of Euclid's axiom with *curves* in place of lines, a much more difficult task.

Now let $v \in [z,u) \cap E_t$ be the first time the line segment $[z,u)$ intersects E_t. By (2.3) and the railway track matchup, we must have either $v \in \left(\beta_{left}^{t}\right)^{*}$ or $v \in \left(\beta_{right}^{t}\right)^{*}$. If we take $r = t$ in Step 1, then z can be connected to either C_1 or C_2 by a path that lies entirely in $\mathbb{C} \setminus \beta^{*}$, namely $[z,v]$ followed by a portion of either β_{left}^{t} or β_{right}^{t}. In particular this shows that there are exactly two connected components in $\mathbb{C} \setminus \beta^{*}$, and completes the proof of Step 2.

Now if $B(0,R)$ is a large disk containing β^{*}, then the connected set $\overline{B(0,R)}^{c}$ is contained in one of the two components of $\mathbb{C} \setminus \beta^{*}$, namely the unbounded one $\mathcal{U}$. The other connected component $\mathcal{B}$ of $\mathbb{C} \setminus \beta^{*}$ is contained in $B(0,R)$, hence is bounded. Moreover, the argument in Step 1 shows that the value of Ind_{β} on opposite sides of β^{*} differs by ± 1. Thus Ind_{β} takes either the value 1 or -1 on the bounded component since $Ind_{\beta} = 0$ on the unbounded component by Proposition 4. At this point the proof of Proposition 8 is complete, and so both the Riemann Mapping Theorem 15 and its Porism 3 are at our disposal. We will use these observations in the proof of the next step.

Step 3: The bounded component $\mathcal{B}$ of $\mathbb{C} \setminus \beta^{*}$ is simply connected.

We will prove this using Porism 3 to the Riemann Mapping Theorem. Note that there is no circularity here since our proof of the Riemann Mapping Theorem and Porism 3 used only Propositions 9 and 8, the first of which was proved in the previous section, and the second of which is proved by Steps 1 and 2 above. In fact, once we show that $\mathcal{B}$ satisfies the hypotheses of Porism 3, the Riemann map from $\mathbb{D}$ to $\mathcal{B}$ shows that $\mathcal{B}$ is homeomorphic to the unit disk $\mathbb{D}$, thus simply connected. In order to verify the hypotheses of Porism 3, it remains to show that every nonvanishing $f \in H(\mathcal{B})$ has a holomorphic square root in $\mathcal{B}$, and this in turn will follow if we show that every $f \in H(\mathcal{B})$ has an antiderivative in $H(\mathcal{B})$. We now apply Porism 2. For this we need only show that

$$\int_\gamma f(z)\,dz = 0, \qquad f \in H(\mathcal{B}), \tag{2.4}$$

for all simple closed taxicab paths γ in $\mathcal{B}$.

Finally, to prove (2.4), we first note that it follows easily using the proof of Step 1 that

$$\mathbb{C} = \mathcal{B} \cup \beta^* \cup \mathcal{U} = \mathcal{B}_\gamma \cup \gamma^* \cup \mathcal{U}_\gamma, \tag{2.5}$$

where $\mathcal{B}_\gamma$ and $\mathcal{U}_\gamma$ are the bounded and unbounded components respectively of $\mathbb{C}\backslash\gamma^*$. Now $\gamma^* \subset \mathcal{B}$ implies $\gamma^* \cap \mathcal{U} = \phi$. Since $\mathcal{U}$ is connected and $\mathcal{U} \cap \mathcal{U}_\gamma \neq \phi$ (they both contain points near infinity), it follows that $\mathcal{U} \subset \mathcal{U}_\gamma$. This together with $\gamma^* \cap \beta^* = \phi$ and (2.5) show that $\mathcal{B}_\gamma \subset \mathcal{B}$. We can now follow an argument already used in the proof of Theorem 16. Indeed, we write

$$\int_\gamma f(z)\,dz = \sum_j \int_{\partial R_j^i} f(z)\,dz$$

where R_j^i is a rectangle contained inside γ (hence in $\mathcal{B}$), ∂R_j^i has the same orientation as γ, and the sum is finite for each i. For this we simply construct a grid of infinite lines in the plane, each passing through one of the segments in γ. This creates a collection of minimal rectangles with sides that are segments of these lines. Then the inside $\mathcal{B}_\gamma$ of γ is the union of all the minimal rectangles R_j^i that happen to lie inside γ. Finally we know that $\int_{\partial R_j^i} f(z)\,dz = 0$ by Cauchy's theorem for a rectangle in a convex subset of $\mathcal{B}$, and summing over i and j proves (2.4). This completes the proof of Proposition 17.

2.1. Proof of the Jordan Curve Theorem. We now turn our attention to the proof of Theorem 25, which we prove in a series of four lemmas. We begin with an approximation lemma which makes the connection with taxicab paths.

LEMMA 8. *Given a simple closed curve β in the plane and $\varepsilon > 0$, there is a simple closed taxicab path α such that*

$$\|\beta - \alpha\|_\infty \equiv \sup_{\zeta \in \mathbb{T}} |\beta(\zeta) - \alpha(\zeta)| < \varepsilon.$$

Proof: We define upper and lower moduli of continuity for $\beta : \mathbb{T} \to \mathbb{C}$ by

$$\begin{aligned} \omega^+(\delta) &= \sup_{|\zeta-\eta|\le\delta} |\beta(\zeta) - \beta(\eta)|, \\ \omega_-(\delta) &= \inf_{|\zeta-\eta|\ge\delta} |\beta(\zeta) - \beta(\eta)|, \end{aligned} \tag{2.6}$$

where ζ, η range over the unit circle $\mathbb{T}$. Since β is continuous and one-to-one we have

$$0 < \omega_-(\delta) \le \omega^+(\delta) \to 0 \text{ as } \delta \to 0.$$

Now choose $\delta > 0$ such that

$$\omega^+(\delta) < \frac{\varepsilon}{4}, \tag{2.7}$$

and then choose a closed, but not necessarily simple, taxicab path γ such that

$$\|\beta-\gamma\|_\infty<\frac{\omega_-(\delta)}{2}.$$

This is easily accomplished by first choosing a polygonal approximation with sufficiently small edges, and then replacing each edge with a vertical and horizontal segment.

We now use γ to construct a *simple* closed taxicab path α satisfying

$$\|\beta-\alpha\|_\infty\le 4\omega^+(\delta), \tag{2.8}$$

which by (2.7) proves Lemma 8. First we modify the path γ by halting it for short periods of time according to the following algorithm. We identify $\mathbb{T}$ with $[0,2\pi)$ under $\zeta=e^{i\theta}\to\theta$, and set $\rho(\theta_1,\theta_2)=\left|e^{i\theta_1}-e^{i\theta_2}\right|$. Let $\theta_1\ge 0$ be the *first* time $\gamma(\theta_1)$ intersects $\gamma((\theta_1,\theta_1+\pi])$. Let $t_1\in(\theta_1,\theta_1+\pi]$ be the *last* time for which $\gamma(t_1)=\gamma(\theta_1)$. Then we have

$$|\beta(t_1)-\beta(\theta_1)|\le|\beta(t_1)-\gamma(t_1)|+|\gamma(\theta_1)-\beta(\theta_1)|<\omega_-(\delta),$$

and it follows that $\rho(t_1,\theta_1)<\delta$. Now we begin defining a modification $\widetilde{\gamma}$ of γ by halting γ at $\gamma(\theta_1)$ for the interval $[\theta_1,t_1]$, i.e.

$$\widetilde{\gamma}(t)=\begin{cases}\gamma(t) & \text{for} \quad 0\le t\le\theta_1\\ \gamma(t_1) & \text{for} \quad \theta_1\le t\le t_1\end{cases}.$$

Now let $\theta_2\ge t_1$ be the *first* time $\gamma(\theta_2)$ intersects $\gamma((\theta_2,\theta_2+\pi])$, and let $t_2\in(\theta_2,\theta_2+\pi]$ be the *last* time for which $\gamma(t_2)=\gamma(\theta_2)$. It is easy to see using $\delta<\frac{\pi}{2}$ that in fact $\theta_2>t_1$. Moreover we have just as above that

$$\rho(t_2,\theta_2)<\delta.$$

Then we continue $\widetilde{\gamma}$ by again halting γ at θ_2:

$$\widetilde{\gamma}(t)=\begin{cases}\gamma(t) & \text{for} \quad 0\le t\le\theta_1\\ \gamma(\theta_1) & \text{for} \quad \theta_1\le t\le t_1\\ \gamma(t) & \text{for} \quad t_1\le t\le\theta_2\\ \gamma(\theta_2) & \text{for} \quad \theta_2\le t\le t_2\end{cases}.$$

We proceed in this way for finitely many steps until $\widetilde{\gamma}$ has been defined for all $t\in[0,2\pi)$. Now $\widetilde{\gamma}$ is a closed taxicab path that has no self-intersections apart from those arising from the finite number of intervals of constancy $[\theta_j,t_j]$ constructed in the above algorithm. A crucial observation is that for $t\in[\theta_j,t_j]$ we have $\rho(t,\theta_j)\le\rho(t_j,\theta_j)<\delta$ and so

$$\begin{aligned}|\widetilde{\gamma}(t)-\beta(t)| &= |\gamma(\theta_j)-\beta(t)|\le|\gamma(\theta_j)-\beta(\theta_j)|+|\beta(\theta_j)-\beta(t)|\\ &< \omega_-(\delta)+\omega^+(\delta)\le 2\omega^+(\delta).\end{aligned}$$

Thus we have $\|\beta-\widetilde{\gamma}\|_\infty\le 2\omega^+(\delta)$, and it is now a simple matter to modify the parameterization of $\widetilde{\gamma}$ near the intervals of constancy so as to produce a *simple* closed taxicab path α that satisfies (2.8). This completes the proof of Lemma 8.

The complement $\mathbb{C}\setminus\beta^*$ of the image of the curve β is a pairwise disjoint union of connected open sets called components.

LEMMA 9. *Let β be a simple closed curve in the plane. Then $\mathbb{C}\setminus\beta^*$ has exactly one unbounded component and at least one bounded component.*

Proof: Clearly there is exactly one unbounded component $\mathcal{U}_\beta$, the one containing the complement of $\overline{B(0,R)}$ for any R chosen large enough that $\beta^* \subset B(0,R)$. We now show using Lemma 8, Proposition 17 and an elementary argument in Maehara [**3**] that there is at least one bounded component. Indeed, we recall the situation depicted in Figure 1 on page 643 of [**3**]. The square $Q \equiv [-1,1] \times [-2,2]$ contains β^* and $\beta^* \cap \partial Q$ consists of just two points $(-1,0), (1,0)$ which we label $W = (-1,0)$ and $E = (1,0)$. We also label $N = (0,2)$ and $S = (0,-2)$. The curve β is divided into two closed arcs by the points W and E, and we label the upper and lower arcs β_{north} and β_{south}. The upper arc β_{north} is determined by passing through the highest point of the intersection of the vertical line segment $\overrightarrow{NS}$ joining N to S (such an intersection point exists by the connectedness of β^*). Denote by T_{north} and B_{north} the top and bottom points in $\beta^*_{north} \cap \overrightarrow{NS}$.

We claim that β^*_{south} must intersect the segment $\overrightarrow{B_{north}S}$. If not, let $\widetilde{T_{north}B_{north}}$ denote the arc of β_{north} that joins T_{north} to B_{north} and note that the simple closed curve

$$\sigma = \overrightarrow{NT_{north}} + \widetilde{T_{north}B_{north}} + \overrightarrow{B_{north}S} + \overrightarrow{S(-1,-2)} + \overrightarrow{(-1,-2)(-1,2)} + \overrightarrow{(-1,2)N}$$

is disjoint from the compact set β^*_{south} provided we modify the segment $\overrightarrow{(-1,-2)(-1,2)}$ to jut to the left around $W = (-1,0)$. Using Lemma 8, we can replace the curve σ with a nearby simple closed taxicab path α whose image α^* is still disjoint from β^*_{south}. Moreover, for $\|\sigma - \alpha\|_\infty$ small enough, there are portions of β^*_{south} in both the bounded and unbounded components of $\mathbb{C} \setminus \alpha^*$ (namely portions near W and portions near E respectively), contradicting the connectedness of β^*_{south}. This proves our claim. Now we denote by T_{south} and B_{south} the top and bottom points in $\beta^*_{south} \cap \overrightarrow{B_{north}S}$. Let Z_0 be the midpoint of $\overrightarrow{B_{north}T_{south}}$. Thus the labeled points running down the vertical segment $\overrightarrow{NS}$ are given in order by

$$N, T_{north}, B_{north}, Z_0, T_{south}, B_{south}, S.$$

Following [**3**] we claim that the component of $\mathbb{C} \setminus \beta^*$ containing Z_0 is *bounded*. If not, then there is a simple taxicab path μ in $\mathbb{C} \setminus \beta^*$ joining Z_0 to $(2,0)$. Let Z_1 be the first point where μ meets ∂Q, and let ν be the arc of μ joining Z_0 to Z_1. Suppose that Z_1 lies in the bottom half of ∂Q (otherwise we mirror the argument given here). Now consider the closed curve

$$\tau = \overrightarrow{NT_{north}} + \widetilde{T_{north}B_{north}} + \overrightarrow{B_{north}Z_0} + \nu + \psi,$$

where ψ is the simple path in ∂Q that starts at Z_1 and proceeds along ∂Q to S without passing through W or E, and then continues on along ∂Q through either W or E (but not both) before ending at N. Once again we modify ψ so as to jut around either W or E, whichever of these it initially passed through. The image τ^* of the curve τ doesn't intersect β^*_{south} and we can again replace τ by a simple closed taxicab path α whose image doesn't intersect β^*_{south} and yet contains portions of β^*_{south} in both the bounded and unbounded components of $\mathbb{C} \setminus \alpha^*$, contradicting the connectedness of β^*_{south}. This completes the proof of Lemma 9.

Lemma 10. *Suppose that β is a simple closed curve in the plane. Then every bounded component of $\mathbb{C} \setminus \beta^*$ is simply connected.*

Proof: If γ is a closed path in a bounded component $\mathcal{B}$ of $\mathbb{C} \setminus \beta^*$ then we can use Lemma 8 to find a simple closed taxicab path α such that γ is contained in the bounded component $\mathcal{B}_\alpha$ of $\mathbb{C} \setminus \alpha^*$. By Proposition 17 we have that $\mathcal{B}_\alpha$ is simply connected, so that γ is $\mathcal{B}_\alpha$-homotopic to a point in $\mathcal{B}_\alpha$. Since $\mathcal{B}_\alpha \subset \mathcal{B}$ we also have that γ is $\mathcal{B}$-homotopic to a point in $\mathcal{B}$. This proves that $\mathcal{B}$ is simply connected and completes the proof of Lemma 10.

Thus far we have shown that

(1) $\mathbb{C} \setminus \beta^*$ has one unbounded component $\mathcal{U}_\beta$,
(2) $\mathbb{C} \setminus \beta^*$ has at least one bounded component,
(3) every bounded component of $\mathbb{C} \setminus \beta^*$ is simply connected.

It remains only to show that there is *exactly* one bounded component in $\mathbb{C} \setminus \beta^*$.

LEMMA 11. *Suppose that β is a simple closed curve in the plane. Then $\mathbb{C} \setminus (\beta^* \cup \mathcal{U}_\beta)$ is connected.*

Suppose that two disks $B(z_1, r)$ and $B(z_2, r)$ are contained in $\mathbb{C} \setminus (\beta^* \cup \mathcal{U}_\beta)$ for some $r > 0$. It suffices to show there is a path in $\mathbb{C} \setminus (\beta^* \cup \mathcal{U}_\beta)$ joining z_1 to z_2. With moduli of continuity ω^+ and ω_- as in (2.6), choose $\delta > 0$ such that

$$\omega^+(\delta) < \frac{r}{2}, \tag{2.9}$$

and then choose a simple closed taxicab path γ so that both

$$\|\beta - \gamma\|_\infty < \frac{\omega_-(\delta)}{1000} \text{ and } B\left(z_j, \frac{3r}{4}\right) \subset \mathcal{B}_\gamma, \qquad j = 1, 2, \tag{2.10}$$

where $\mathcal{B}_\gamma$ is the bounded component of $\mathbb{C} \setminus \gamma^*$. Note that $Ind_\gamma(z_j) = 1$ for $j = 1, 2$ (if $Ind_\gamma(z_j) = -1$ we can reverse the direction of γ). Now consider a very fine grid of horizontal and vertical lines whose consecutive distances apart lie in a small interval $[\eta, 4\eta]$ with $\eta > 0$. With η small enough we can arrange to have all the segments in γ lie in grid lines, and furthermore we can assume that $\eta < \frac{\omega_-(\delta)}{1000}$. Let $\mathcal{R}$ be the collection of all minimal rectangles with edges in the grid lines. Define

$$E = \bigcup \left\{ R \in \mathcal{R} : R \subset \mathcal{B}_\gamma \text{ and } dist(R, \gamma) < \frac{\omega_-(\delta)}{100} \right\}.$$

Then E is a finite union of rectangles R from $\mathcal{R}$, and ∂E consists of finitely many simple closed taxicab paths, one of which is γ. Now it is easy to see that exactly one of the remaining taxicab paths, say ρ, includes z_1 in its bounded component $\mathcal{B}_\rho$. We claim that z_2 is included in $\mathcal{B}_\rho$ as well. To see this it suffices by Proposition 9 to show that the paths γ and ρ are Δ-homotopic to each other where $\Delta = \mathbb{C} \setminus \{z_2\}$. Indeed, we would then have

$$Ind_\rho(z_2) = Ind_\gamma(z_2) = 1, \tag{2.11}$$

and it would follow from Proposition 8 that z_2 lies in the bounded component $\mathcal{B}_\rho$ of ρ as well.

So pick a sequence $\{P_n\}_{n=1}^N$ of points on ρ that traverse ρ in the positive direction and that satisfy

$$\frac{\omega_-(\delta)}{10} < |P_n - P_{n+1}| < \frac{\omega_-(\delta)}{5}, \qquad 1 \le n \le N,$$

where $P_{N+1} = P_1$. Then pick a sequence $\{Q_n\}_{n=1}^N$ of points on γ satisfying

$$|P_n - Q_n| < \frac{\omega_-(\delta)}{100}. \tag{2.12}$$

Now we note that

$$|Q_n - Q_{n+1}| \geq |P_n - P_{n+1}| - \{|P_n - Q_n| + |P_{n+1} - Q_{n+1}|\} > \frac{\omega_-(\delta)}{20},$$

and thus that γ is divided into two taxicab arcs by Q_n and Q_{n+1}. One of these taxicab arcs joining Q_n and Q_{n+1} must have diameter at least $\frac{r}{2} > \omega^+(\delta)$ in order that $Ind_\gamma(z_1) = 1$ (use that every half-ray from z_1 must intersect γ^*), and it then follows from (2.6) and

$$|Q_n - Q_{n+1}| \leq |Q_n - P_n| + |P_n - P_{n+1}| + |P_{n+1} - Q_{n+1}| < \omega_-(\delta)$$

that the other taxicab arc joining Q_n and Q_{n+1} must have diameter at most $\omega^+(\delta) < \frac{r}{2}$. We denote this latter taxicab arc of γ by $\widetilde{Q_nQ_{n+1}}$. It also follows that the circular arcs $\beta^{-1}\left(\widetilde{Q_nQ_{n+1}}\right)$ have diameter at most δ, have pairwise disjoint interiors since $Ind_\rho(z_1) = 1$, and have union equal to $\mathbb{T}$. Thus we have

$$\gamma = \widetilde{Q_1Q_2} + ... + \widetilde{Q_{N-1}Q_N} + \widetilde{Q_NQ_1}. \tag{2.13}$$

We will show that γ and ρ are $\triangle$-homotopic by exhibiting intermediate paths which are successively $\triangle$-homotopic to one another by elementary homotopies such as "growing a finger" and "shrinking of a closed curve to a point by dilation". First we claim that γ is $\triangle$-homotopic to the path τ_1 constructed as follows:

- Start at P_2 and proceed along the segment $\overrightarrow{P_2P_1}$ (which will not normally lie on the path ρ), then along the segment $\overrightarrow{P_1Q_1}$, then along the arc $\widetilde{Q_1Q_2}$, then along the segment $\overrightarrow{Q_2P_2}$. This results in a closed path
$$\sigma_2 = \overrightarrow{P_2P_1} + \overrightarrow{P_1Q_1} + \widetilde{Q_1Q_2} + \overrightarrow{Q_2P_2}$$
through the point P_2 that is contained in the disk $B\left(P_2, \frac{2r}{3}\right)$ since $diam\left(\widetilde{Q_nQ_{n+1}}\right) < \frac{r}{2}$. Then proceed from P_2 to P_3 along the segment $\overrightarrow{P_2P_3}$.
- Recursively, for $n \geq 3$, start at P_n and follow the algorithm described in the first bullet item. This results in a closed path
$$\sigma_n = \overrightarrow{P_nP_{n-1}} + \overrightarrow{P_{n-1}Q_{n-1}} + \widetilde{Q_{n-1}Q_n} + \overrightarrow{Q_nP_n}$$
through the point P_n that is contained in the disk $B\left(P_n, \frac{2r}{3}\right)$, which is then followed by the segment $\overrightarrow{P_nP_{n+1}}$.

In this way we construct the path

$$\tau_1 = \sigma_2 + \overrightarrow{P_2P_3} + \sigma_3 + \overrightarrow{P_3P_4} + ... + \sigma_N + \overrightarrow{P_NP_1} + \sigma_1 + \overrightarrow{P_1P_2},$$

which is clearly $\triangle$-homotopic to

$$\widetilde{Q_1Q_2} + ... + \widetilde{Q_{N-1}Q_N} + \widetilde{Q_NQ_1},$$

since $\overrightarrow{Q_{n-1}P_{n-1}} + \overrightarrow{P_{n-1}P_n} + \overrightarrow{P_nP_{n-1}} + \overrightarrow{P_{n-1}Q_{n-1}}$ is a "finger" disjoint from z_2. But this latter path is γ by (2.13). We next claim that τ_1 is $\triangle$-homotopic to the path τ_2 given by

$$\tau_2 = \overrightarrow{P_1P_2} + \overrightarrow{P_2P_3} + ... + \overrightarrow{P_{N-1}P_N} + \overrightarrow{P_NP_1}.$$

Indeed, simply contract by dilation the closed path σ_n to the point P_n within the disk $B\left(P_n, \frac{2r}{3}\right)$. This clearly avoids the point z_2 since P_n lies on ρ and $B\left(z_2, \frac{3r}{4}\right)$ is disjoint from ρ by (2.10). Finally, it is an easy exercise to show that the taxicab path ρ is also homotopic to the path τ_2 (for example, one can argue as above with ρ in place of γ). Thus Proposition 9 shows that (2.11) holds, and then Proposition 8 shows that $z_2 \in \mathcal{B}_\rho$.

Thus both z_1 and z_2 lie in $\mathcal{B}_\rho$, which is simply connected by Proposition 17, and it follows that there is a path connecting z_1 to z_2 in $\mathcal{B}_\rho$. However, by construction and the definition of E, the path ρ is at a distance at least $\frac{\omega_-(\delta)}{200}$ from γ, while γ itself is at a distance at most $\frac{\omega_-(\delta)}{1000}$ from β. It follows that ρ is at a distance at least $\frac{\omega_-(\delta)}{300}$ from β, and so

$$\mathcal{B}_\rho \subset \mathbb{C} \setminus (\beta^* \cup \mathcal{U}_\beta) = \mathcal{B}_\beta. \tag{2.14}$$

This completes the proof of Lemma 11.

The proof of the Jordan Curve Theorem 25 follows immediately from the four lemmas above.

2.2. Simple boundary points. We now prove that every point w on β^* is not only a boundary point of the bounded component $\mathcal{B}_\beta$ of $\mathbb{C} \setminus \beta^*$, but is actually a *simple* boundary point of $\mathcal{B}_\beta$. We recall from Definition 11 that w is a simple boundary point of Ω if for every sequence $\{z_n\}_{n=1}^\infty \subset \Omega$ with $\lim_{n\to\infty} z_n = w$, there is a (continuous) curve $\Gamma : [0,1) \to \Omega$ and an increasing sequence of parameter points $\{t_n\}_{n=1}^\infty \subset [0,1)$ with $\lim_{n\to\infty} t_n = 1$ such that $\Gamma(t_n) = z_n$ for all $n \geq 1$ and $\lim_{t\to 1} \Gamma(t) = w$.

PROPOSITION 18. *Suppose $\beta : \mathbb{T} \to \mathbb{C}$ is continuous and one-to-one. If $w \in \beta^*$, then w is a simple boundary point of the bounded component $\mathcal{B}_\beta$ of $\mathbb{C} \setminus \beta^*$.*

Proof: We refer to the proof of Lemma 11 above. Consider points z_1 and z_2 as in the proof there. Select points A_1 and A_2 in β^* with

$$|z_i - A_i| = dist\,(z_i, \beta^*)\,, \qquad i = 1, 2,$$

and then select points B_1 and B_2 in ρ^* with

$$B_i \in B\,(z_i, dist\,(z_i, \beta^*)) \cap B\,(A_i, \omega_-(\delta))\,, \qquad i = 1, 2.$$

Let $\eta > 0$ satisfy

$$|A_1 - A_2| < \omega_-(\eta)\,.$$

Then one of the two arcs of β obtained by removing A_1 and A_2 from β^* must have diameter at most $\omega^+(\eta)$. It follows that one of the two arcs of ρ obtained by removing B_1 and B_2 from ρ^*, call it τ, must have diameter at most

$$\omega^+(\eta) + \omega_-(\delta)\,.$$

Now let σ be the path obtained by joining z_1 to B_1 with a line segment, then joining B_1 to B_2 with τ, and finally joining B_2 to z_2 with a line segment. We then have the estimate

$$diam\,(\sigma^*) < dist\,(z_1, \beta^*) + \omega^+(\eta) + \omega_-(\delta) + dist\,(z_2, \beta^*)\,.$$

CLAIM 3. *Given $\varepsilon > 0$, there exists $\lambda > 0$ such that whenever $z_1, z_2 \in \mathcal{B}_\beta$ with $|z_1 - z_2| < \lambda$, there is a path σ in $\mathcal{B}_\beta$ that joins z_1 to z_2 and has*

$$diam\,(\sigma^*) < \varepsilon.$$

To see this let $\varepsilon > 0$, and choose $\eta > 0$ so that $\omega^{+}(\eta) < \frac{\varepsilon}{3}$. Then if z_1 and z_2 satisfy

$$dist\,(z_1, \beta^*) + |z_1 - z_2| + dist\,(z_2, \beta^*) < \omega_-(\eta)\,, \tag{2.15}$$

it follows that

$$|A_1 - A_2| < \omega_-(\eta)\,,$$

and we showed above that there is in this case a path σ in $\mathcal{B}_\beta$ that joins z_1 to z_2 and satisfies

$$\begin{aligned} diam\,(\sigma^*) &< dist\,(z_1, \beta^*) + \omega^+(\eta) + \omega_-(\delta) + dist\,(z_2, \beta^*) \\ &< 2\omega_-(\eta) + \omega^+(\eta) \leq 3\omega^+(\eta) < \varepsilon, \end{aligned}$$

since

$$\omega_-(\delta) \leq \omega^+(\delta) \leq \frac{r}{2} < dist\,(z_1, \beta^*) < \omega_-(\eta) \leq \omega^+(\eta)\,.$$

On the other hand, if $|z_1 - z_2| < \frac{1}{4}\omega_-(\eta)$ and (2.15) fails, it follows that both $dist\,(z_1, \beta^*)$ and $dist\,(z_2, \beta^*)$ exceed $\frac{1}{4}\omega_-(\eta)$. Indeed, $dist\,(z_1, \beta^*) \leq \frac{1}{4}\omega_-(\eta)$ implies

$$dist\,(z_2, \beta^*) \leq |z_2 - z_1| + dist\,(z_1, \beta^*) \leq \frac{1}{2}\omega_-(\eta)\,,$$

which implies that (2.15) holds. It now follows that

$$B\,(z_1, dist\,(z_1, \beta^*)) \cap B\,(z_2, dist\,(z_2, \beta^*)) \neq \phi,$$

and so the line segment joining z_1 to z_2 lies in $\mathcal{B}_\beta$ and has length $\frac{1}{4}\omega_-(\eta) < \frac{1}{4}\omega^+(\eta) < \frac{\varepsilon}{12}$. Altogether this shows that the claim holds if we take $\lambda = \frac{1}{4}\omega_-(\eta)$.

It is now an easy matter to show that $w \in \partial\mathcal{B}_\beta \subset \beta^*$ is a simple boundary point of $\mathcal{B}_\beta$. Let $\{z_n\}_{n=1}^\infty \subset \mathcal{B}_\beta$ have limit w. By the above claim there is a decreasing sequence $\{\lambda_k\}_{k=1}^\infty$ of positive numbers with limit 0 as $k \to \infty$ such that if

$$|z_m - z_n| < \lambda_k, \tag{2.16}$$

there is a path $\sigma_{m,n}$ joining z_m to z_n in $\mathcal{B}_\beta$ with

$$diam\left(\sigma_{m,n}^*\right) < 2^{-k}.$$

Denote by $\widetilde{\sigma_{m,n}} = \sigma_{m,n} - \sigma_{m,n}$ the closed path that runs from z_m to z_n and back to z_m. Fix a subsequence $\{N_k\}_{k=1}^\infty$ of positive integers satisfying

$$\sup_{n \geq N_k} |z_{N_k} - z_n| < \lambda_k, \qquad k \geq 1.$$

Now let $\pi_k : [0,1) \to \mathcal{B}_\beta$ be the curve joining z_{N_k} to $z_{N_{k+1}}$ obtained by concatening in order the following paths:

$$\pi_k \equiv \widetilde{\sigma_{N_k, N_k+1}} + \widetilde{\sigma_{N_k, N_k+2}} + ...+, \widetilde{\sigma_{N_k, N_{k+1}-1}} +, \sigma_{N_k, N_{k+1}}.$$

Note that π_k does indeed pass through the points $z_{N_k}, z_{N_k+1}, ...z_{N_{k+1}}$ in order (we can ignore the fact that it also returns to z_{N_k} and possibly other of the points repeatedly). Now we let $\pi : [0,1) \to \mathcal{B}_\beta$ be the *curve* obtained by concatening in order the paths π_k:

$$\pi \equiv \pi_1 + \pi_2 + ... = \sum_{k=1}^\infty \pi_k.$$

By construction there is an increasing sequence $\{t_n\}_{n=1}^{\infty}$ in $[0,1)$ such that t_n is a parameter point for π_k if $N_k \leq n < N_{k+1}$, and such that $\sigma(t_n) = z_n$. Since

$$diam\,(\pi\,([t_n,1))) \leq \sum_{\ell=k}^{\infty} 2^{-\ell} = 2^{1-k},$$

if $n \geq N_k$, we conclude that

$$\lim_{t\to 1} \sigma(t) = \lim_{n\to\infty} \sigma(t_n) = \lim_{n\to\infty} z_n = w.$$

Finally, it remains to show that $\partial\mathcal{B}_\beta = \beta^*$. We continue to refer to the proof of Lemma 11 above. By (2.10) there is a point Q on γ^* such that $|w - Q| < \frac{\omega_-(\delta)}{1000}$. By (2.13) the point Q lies on one of the arcs $\widetilde{Q_n Q_{n+1}}$, which by construction has diameter at most $\omega^+(\delta)$. From (2.12) it follows that

$$\begin{aligned} |P_n - w| &\leq |P_n - Q_n| + |Q_n - Q| + |Q - w| \\ &< \frac{\omega_-(\delta)}{100} + \omega^+(\delta) + \frac{\omega_-(\delta)}{1000} \leq 2\omega^+(\delta). \end{aligned}$$

Now ρ is a simple closed taxicab path and it is clear that P_n is a boundary point of the bounded component $\mathcal{B}_\rho$ of $\mathbb{C} \setminus \rho^*$. Since $\mathcal{B}_\rho \subset \mathcal{B}_\beta$ by (2.14), it follows that there is a point in $\mathcal{B}_\beta$ within distance $2\omega^+(\delta)$ of the point w. Since $\lim_{\delta\to 0}\omega^+(\delta) = 0$ and δ can be chosen arbitrarily small in the proof of Lemma 11, we conclude that $w \in \overline{\mathcal{B}_\beta}$.

APPENDIX B

Lebesgue measure theory

Let $f : [0,1) \rightarrow [0,M)$ be a nonnegative bounded function on the half open unit interval $[0,1)$. In *Riemann's* theory of integration, we partition the *domain* $[0,1)$ of the function into finitely many disjoint subintervals

$$[0,1) = \bigcup_{n=1}^{N} [x_{n-1}, x_n),$$

and denote the parition by $\mathcal{P} = \{0 = x_0 < x_1 < ... < x_N = 1\}$ and the length of the subinterval $[x_{n-1}, x_n)$ by $\triangle x_n = x_n - x_{n-1} > 0$. Then we define *upper and lower Riemann sums* associated with the partition $\mathcal{P}$ by

$$\begin{aligned} U(f;\mathcal{P}) &= \sum_{n=1}^{N} \left(\sup_{[x_{n-1},x_n)} f \right) \triangle x_n, \\ L(f;\mathcal{P}) &= \sum_{n=1}^{N} \left(\inf_{[x_{n-1},x_n)} f \right) \triangle x_n. \end{aligned}$$

Then we define the *upper and lower Riemann integrals* of f on $[0,1)$ by

$$\mathcal{U}(f) = \inf_{\mathcal{P}} U(f;\mathcal{P}), \qquad \mathcal{L}(f) = \sup_{\mathcal{P}} L(f;\mathcal{P}).$$

Thus the upper Riemann integral $\mathcal{U}(f)$ is the "smallest" of all the upper sums, and the lower Riemann integral is the "largest" of all the lower sums. By considering the refinement $\mathcal{P}_1 \cup \mathcal{P}_2$ of two partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ it is easy to see that

$$U(f;\mathcal{P}_1) \geq U(f;\mathcal{P}_1 \cup \mathcal{P}_2) \geq L(f;\mathcal{P}_1 \cup \mathcal{P}_2) \geq L(f;\mathcal{P}_2).$$

Taking the infimum over $\mathcal{P}_1$ and the supremum over $\mathcal{P}_2$ shows that

$$\mathcal{U}(f) \geq \mathcal{L}(f).$$

Finally we say that f is Riemann integrable on $[0,1)$, written $f \in \mathcal{R}[0,1)$, if $\mathcal{U}(f) = \mathcal{L}(f)$, and we denote the common value by $\int_0^1 f$ or $\int_0^1 f(x)\,dx$.

This definition is simple and easy to work with and applies in particular to bounded continuous functions f on $[0,1)$ since it is not too hard to prove that $f \in \mathcal{R}[0,1)$ for such f. However, if we consider the vector space $L^2_{\mathcal{R}}([0,1))$ of Riemann integrable functions $f \in \mathcal{R}[0,1)$ endowed with the metric

$$d(f,g) = \left(\int_0^1 |f(x) - g(x)|^2 \, dx \right)^{\frac{1}{2}},$$

it turns out that while $L^2_{\mathcal{R}}([0,1))$ can indeed be proved a metric space, it fails to be *complete*. This is a serious shortfall of Riemann's theory of integration, and is our main motivation for considering the more complicated theory of Lebesgue below. We note that the immediate reason for the lack of completeness of $L^2_{\mathcal{R}}([0,1))$ is the inability of Riemann's theory to handle general unbounded functions. However, even locally there are problems. For example, once we have Lebesgue's theory in hand, we can construct a famous example of a Lebesgue measurable subset E of $[0,1)$ with the (somewhat surprising) property that

$$0<|E\cap(a,b)|<b-a,\qquad 0\leq a<b\leq 1,$$

where $|F|$ denotes the Lebesgue measure of a measurable set F (see Problem 5 below). It follows that the characteristic function χ_E is bounded and Lebesgue measurable, but that there is no Riemann integrable function f such that $f=\chi_E$ almost everywhere, since such an f would satisfy $\mathcal{U}(f)=1$ and $\mathcal{L}(f)=0$. Nevertheless, by Lusin's Theorem (see page 34 in [**6**] or page 55 in [**5**]) there is a sequence of compactly supported *continuous* functions (hence Riemann integrable) converging to χ_E almost everywhere.

On the other hand, in *Lebesgue's* theory of integration, we partition the *range* $[0,M)$ of the function into a homogeneous partition,

$$[0,M)=\bigcup_{n=1}^{N}\left[(n-1)\frac{M}{N},n\frac{M}{N}\right)\equiv\bigcup_{n=1}^{N}I_n,$$

and we consider the associated upper and lower *Lebesgue* sums of f on $[0,1)$ defined by

$$\begin{aligned}U^*(f;\mathcal{P})&=\sum_{n=1}^{N}\left(n\frac{M}{N}\right)\left|f^{-1}(I_n)\right|,\\ L^*(f;\mathcal{P})&=\sum_{n=1}^{N}\left((n-1)\frac{M}{N}\right)\left|f^{-1}(I_n)\right|,\end{aligned}$$

where of course

$$f^{-1}(I_n)=\left\{x\in[0,1):f(x)\in I_n=\left[(n-1)\frac{M}{N},n\frac{M}{N}\right)\right\},$$

and $|E|$ denotes the "measure" or "length" of the subset E of $[0,1)$.

Here there will be no problem obtaining that $U^*(f;\mathcal{P})-L^*(f;\mathcal{P})$ is small provided we can make sense of $\left|f^{-1}(I_n)\right|$. But this is precisely the difficulty with Lebesgue's approach - we need to define a notion of "measure" or "length" for subsets E of $[0,1)$. That this is not going to be as easy as we might hope is evidenced by the following negative result. Let $\mathcal{P}([0,1))$ denote the power set of $[0,1)$, i.e. the set of all subsets of $[0,1)$. For $x\in[0,1)$ and $E\in\mathcal{P}([0,1))$ we define the translation $E\oplus x$ of E by x to be the set in $\mathcal{P}([0,1))$ defined by

$$\begin{aligned}E\oplus x&=E+x\qquad(\mathrm{mod}\,1)\\&=\{z\in[0,1):\text{there is }y\in E\text{ with }y+x-z\in\mathbb{Z}\}.\end{aligned}$$

Theorem 26. *There is* no *map* $\mu:\mathcal{P}([0,1))\to[0,\infty)$ *satisfying the following three properties:*

(1) $\mu([0,1))=1$,
(2) $\mu\left(\dot{\bigcup}_{n=1}^{\infty}E_n\right)=\sum_{n=1}^{\infty}\mu(E_n)$ *whenever* $\{E_n\}_{n=1}^{\infty}$ *is a pairwise disjoint sequence of sets in* $\mathcal{P}([0,1))$,
(3) $\mu(E\oplus x)=\mu(E)$ *for all* $E\in\mathcal{P}([0,1))$.

REMARK 8. *All three of these properties are desirable for any notion of measure or length of subsets of* $[0,1)$. *The theorem suggests then that we should* not *demand that every subset of* $[0,1)$ *be "measurable". This will then restrict the functions* f *that we can integrate to those for which* $f^{-1}([a,b))$ *is "measurable" for all* $-\infty < a < b < \infty$.

Proof: Let $\{r_n\}_{n=1}^{\infty} = \mathbb{Q} \cap [0,1)$ be an enumeration of the rational numbers in $[0,1)$. Define an equivalence relation on $[0,1)$ by declaring that $x \sim y$ if $x - y \in \mathbb{Q}$. Let $\mathcal{A}$ be the set of equivalence classes. Use the *axiom of choice* to pick a representative $a = \langle A \rangle$ from each equivalence class A in $\mathcal{A}$. Finally, let $E = \{\langle A \rangle : A \in \mathcal{A}\}$ be the set consisting of these representatives a, one from each equivalence class A in $\mathcal{A}$.

Then we have

$$[0,1) = \dot{\bigcup}_{n=1}^{\infty} E \oplus r_n.$$

Indeed, if $x \in [0,1)$, then $x \in A$ for some $A \in \mathcal{A}$, and thus $x \sim a = \langle A \rangle$, i.e. $x - a \in \{r_n\}_{n=1}^{\infty}$. If $x \geq a$ then $x - a \in \mathbb{Q} \cap [0,1)$ and $x = a + r_m$ where $a \in E$ and $r_m \in \{r_n\}_{n=1}^{\infty}$. If $x < a$ then $x - a + 1 \in \mathbb{Q} \cap [0,1)$ and $x = a + (r_m \ominus 1)$ where $a \in E$ and $r_m \ominus 1 \in \{r_n\}_{n=1}^{\infty}$. Finally, if $a \oplus r_m = b \oplus r_n$, then $a \ominus b = r_n \ominus r_m \in \mathbb{Q}$ which implies that $a \sim b$ and then $r_n = r_m$.

Now by properties (1), (2) and (3) in succession we have

$$1 = \mu([0,1)) = \mu\left(\dot{\bigcup}_{n=1}^{\infty} E \oplus r_n\right) = \sum_{n=1}^{\infty} \mu(E \oplus r_n) = \sum_{n=1}^{\infty} \mu(E),$$

which is impossible since the infinite series $\sum_{n=1}^{\infty} \mu(E)$ is either ∞ if $\mu(E) > 0$ or 0 if $\mu(E) = 0$.

1. Lebesgue measure on the real line

In order to define a "measure" satisfying the three properties in Theorem 26, we must restrict the domain of definition of the set functional μ to a "suitable" proper subset of the power set $\mathcal{P}([0,1))$. A good notion of "suitable" is captured by the following definition where we expand our quest for measure to the entire real line.

DEFINITION 12. *A collection* $\mathcal{A} \subset \mathcal{P}(\mathbb{R})$ *of subsets of real numbers* $\mathbb{R}$ *is called a* σ-algebra *if the following properties are satisfied:*

1. $\phi \in \mathcal{A}$,
2. $A^c \in \mathcal{A}$ *whenever* $A \in \mathcal{A}$,
3. $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ *whenever* $A_n \in \mathcal{A}$ *for all* n.

Here is the theorem asserting the existence of "Lebesgue measure" on the real line.

THEOREM 27. *There is a* σ*-algebra* $\mathcal{L} \subset \mathcal{P}(\mathbb{R})$ *and a function* $\mu : \mathcal{L} \to [0,\infty]$ *such that*

1. $[a,b) \in \mathcal{L}$ *and* $\mu([a,b)) = b - a$ *for all* $-\infty < a < b < \infty$,
2. $\dot{\bigcup}_{n=1}^{\infty} E_n \in \mathcal{L}$ *and* $\mu\left(\dot{\bigcup}_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n)$ *whenever* $\{E_n\}_{n=1}^{\infty}$ *is a pairwise disjoint sequence of sets in* $\mathcal{L}$,
3. $E + x \in \mathcal{L}$ *and* $\mu(E + x) = \mu(E)$ *for all* $E \in \mathcal{L}$,

(4) $E \in \mathcal{L}$ *and* $\mu(E) = 0$ *whenever* $E \subset F$ *and* $F \in \mathcal{L}$ *with* $\mu(F) = 0$.

It turns out that both the σ-algebra $\mathcal{L}$ and the function μ are uniquely determined by these four properties, but we will only need the *existence* of such $\mathcal{L}$ and μ. The sets in the σ-algebra $\mathcal{L}$ are called *Lebesgue measurable* sets.

A pair $(\mathcal{L}, \mu)$ satisfying only property (2) is called a measure space. Property (1) says that the measure μ is an extension of the usual length function on intervals. Property (3) says that the measure is translation invariant, while property (4) says that the measure is complete.

From property (2) and the fact that μ is nonnegative, we easily obtain the following elementary consequences (where membership in $\mathcal{L}$ is implied by *context*):

$$\begin{aligned}
\phi &\in \mathcal{L} \text{ and } \mu(\phi) = 0, \\
E &\in \mathcal{L} \text{ for every open set } E \text{ in } \mathbb{R}, \\
\mu(I) &= b - a \text{ for } \textit{any} \text{ interval } I \text{ with endpoints } a \text{ and } b, \\
\mu(E) &= \sup_n \mu(E_n) = \lim_{n\to\infty} \mu(E_n) \text{ if } E_n \nearrow E, \\
\mu(E) &= \inf_n \mu(E_n) = \lim_{n\to\infty} \mu(E_n) \text{ if } E_n \searrow E \textbf{ and } \mu(E_1) < \infty.
\end{aligned} \tag{1.1}$$

For example, the fourth line follows from writing

$$E = E_1 \dot{\cup} \left\{ \dot{\bigcup}_{n=1}^{\infty} E_{n+1} \cap (E_n)^c \right\}$$

and then using property (2) of μ.

To prove Theorem 27 we follow the treatment in [**6**] with simplifications due to the fact that the connected open subsets of the real numbers $\mathbb{R}$ are just the open intervals (a, b). Define for *any* $E \in \mathcal{P}(\mathbb{R})$, the *outer Lebesgue measure* $\mu^*(E)$ of E by,

$$\mu^*(E) = \inf \left\{ \sum_{n=1}^{\infty} (b_n - a_n) : E \subset \dot{\bigcup}_{n=1}^{\infty} (a_n, b_n) \text{ and } -\infty \le a_n < b_n \le \infty \right\}.$$

It is immediate that μ^* is monotone,

$$\mu^*(E) \le \mu^*(F) \text{ if } E \subset F.$$

A little less obvious is countable subadditivity of μ^*.

LEMMA 12. μ^* *is countably subadditive:*

$$\mu^*\left(\bigcup_{n=1}^{\infty} E_n \right) \le \sum_{n=1}^{\infty} \mu^*(E_n), \qquad \{E_n\}_{n=1}^{\infty} \subset \mathcal{P}(\mathbb{R}).$$

Proof: Given $0 < \varepsilon < 1$, we have $E_n \subset \dot{\bigcup}_{k=1}^{\infty} (a_{k,n}, b_{k,n})$ with

$$\sum_{k=1}^{\infty} (b_{k,n} - a_{k,n}) < \mu^*(E_n) + \frac{\varepsilon}{2^n}, \qquad n \ge 1.$$

Now let

$$\bigcup_{n=1}^{\infty} \left(\dot{\bigcup}_{k=1}^{\infty} (a_{k,n}, b_{k,n}) \right) = \dot{\bigcup}_{m=1}^{M^*} (c_m, d_m),$$

where $M^* \in \mathbb{N} \cup \{\infty\}$. Then define disjoint sets of indices

$$\mathcal{I}_m = \{(k,n) : (a_{k,n}, b_{k,n}) \subset (c_m, d_m)\}.$$

In the case $c_m, d_m \in \mathbb{R}$, we can choose by compactness a finite subset $\mathcal{F}_m$ of $\mathcal{I}_m$ such that

$$\left[c_m + \frac{\varepsilon}{2}\delta_m, d_m - \frac{\varepsilon}{2}\delta_m\right] \subset \bigcup_{(k,n)\in\mathcal{F}_m}^{\infty} (a_{k,n}, b_{k,n}), \tag{1.2}$$

where $\delta_m = d_m - c_m$. Fix m and arrange the left endpoints $\{a_{k,n}\}_{(k,n)\in\mathcal{F}_m}$ in strictly increasing order $\{a_i\}_{i=1}^{I}$ and denote the corresponding right endpoints by b_i (if there is more than one interval (a_i, b_i) with the same left endpoint a_i, discard all but one of the largest of them). From (1.2) it now follows that $a_{i+1} \in (a_i, b_i)$ for $i < I$ since otherwise b_i would be in the left side of (1.2), but not in the right side, a contradiction. Thus $a_{i+1} - a_i \le b_i - a_i$ for $1 \le i < I$ and we have the inequality

$$\begin{aligned}
(1-\varepsilon)\,\delta_m &= \left(d_m - \frac{\varepsilon}{2}\delta_m\right) - \left(c_m + \frac{\varepsilon}{2}\delta_m\right) \\
&\le b_I - a_1 = (b_I - a_I) + \sum_{i=1}^{I-1} (a_{i+1} - a_i) \\
&\le \sum_{i=1}^{I} (b_i - a_i) = \sum_{(k,n)\in\mathcal{F}_m} (b_{k,n} - a_{k,n}) \\
&\le \sum_{(k,n)\in\mathcal{I}_m} (b_{k,n} - a_{k,n}).
\end{aligned}$$

We also observe that a similar argument shows that $\sum_{(k,n)\in\mathcal{I}_m} (b_{k,n} - a_{k,n}) = \infty$ if $\delta_m = \infty$. Then we have

$$\begin{aligned}
\mu^*(E) &\le \sum_{m=1}^{\infty} \delta_m \le \frac{1}{1-\varepsilon} \sum_{m=1}^{\infty} \sum_{(k,n)\in\mathcal{F}_m} (b_{k,n} - a_{k,n}) \\
&\le \frac{1}{1-\varepsilon} \sum_{k,n} (b_{k,n} - a_{k,n}) = \frac{1}{1-\varepsilon} \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} (b_{k,n} - a_{k,n}) \\
&< \frac{1}{1-\varepsilon} \sum_{n=1}^{\infty} \left(\mu^*(E_n) + \frac{\varepsilon}{2^n}\right) = \frac{1}{1-\varepsilon} \sum_{n=1}^{\infty} \mu^*(E_n) + \frac{\varepsilon}{1-\varepsilon}.
\end{aligned}$$

Let $\varepsilon \to 0$ to obtain the countable subadditivity of μ^*.

Now define the subset $\mathcal{L}$ of $\mathcal{P}(\mathbb{R})$ to consist of all subsets A of the real line such that for every $\varepsilon > 0$, there is an open set $G \supset A$ satisfying

$$\mu^*(G \setminus A) < \varepsilon. \tag{1.3}$$

REMARK 9. *Condition (1.3) says that A can be well approximated from the outside by open sets. The most difficult task we will face below in using this definition of $\mathcal{L}$ is to prove that such sets A can also be well approximated from the inside by closed sets.*

Set

$$\mu(A) = \mu^*(A), \qquad A \in \mathcal{L}.$$

Trivially, every open set and every interval is in $\mathcal{L}$. We will use the following two claims in the proof of Theorem 27.

CLAIM 4. *If G is open and $G = \overset{\cdot}{\bigcup}_{n=1}^{N^*} (a_n, b_n)$ (where $N^* \in \mathbb{N} \cup \{\infty\}$) is the decomposition of G into its connected components (a_n, b_n), then*

$$\mu(G) = \mu^*(G) = \sum_{n=1}^{N^*} (b_n - a_n).$$

We first prove Claim 4 when $N^* < \infty$. If $G \subset \overset{\cdot}{\bigcup}_{m=1}^{\infty} (c_m, d_m)$, then for each $1 \le n \le N^*$, $(a_n, b_n) \subset (c_m, d_m)$ for some m since (a_n, b_n) is connected. If

$$\mathcal{I}_m = \{n : (a_n, b_n) \subset (c_m, d_m)\},$$

it follows upon arranging the a_n in increasing order that

$$\sum_{n \in \mathcal{I}_m} (b_n - a_n) \le d_m - c_m,$$

since the intervals (a_n, b_n) are pairwise disjoint. We now conclude that

$$\begin{aligned} \mu^*(G) &= \inf \left\{ \sum_{m=1}^{\infty} (d_m - c_m) : G \subset \overset{\cdot}{\bigcup}_{m=1}^{\infty} (c_m, d_m) \right\} \\ &\ge \sum_{m=1}^{\infty} \sum_{n \in \mathcal{I}_m} (b_n - a_n) = \sum_{n=1}^{N^*} (b_n - a_n), \end{aligned}$$

and hence that $\mu^*(G) = \sum_{n=1}^{N^*} (b_n - a_n)$ by definition since $G \subset \overset{\cdot}{\bigcup}_{m=1}^{N^*} (a_n, b_n)$.

Finally, if $N^* = \infty$, then from what we just proved and monotonicity, we have

$$\mu^*(G) \ge \mu^* \left(\overset{\cdot}{\bigcup}_{m=1}^{N} (a_n, b_n) \right) = \sum_{n=1}^{N} (b_n - a_n)$$

for each $1 \le N < \infty$. Taking the supremum over N gives $\mu^*(G) \ge \sum_{n=1}^{\infty} (b_n - a_n)$, and then equality follows by definition since $G \subset \overset{\cdot}{\bigcup}_{n=1}^{\infty} (a_n, b_n)$.

CLAIM 5. *If A and B are disjoint compact subsets of $\mathbb{R}$, then*

$$\mu^*(A) + \mu^*(B) = \mu^*(A \cup B).$$

Since $\delta = dist(A, B) > 0$, we can find open sets U and V such that

$$A \subset U \text{ and } B \subset V \text{ and } U \cap V = \phi.$$

For example, $U = \bigcup_{x \in A} B\left(x, \frac{\delta}{2}\right)$ and $V = \bigcup_{x \in B} B\left(x, \frac{\delta}{2}\right)$ work. Now suppose that

$$A \cup B \subset G \equiv \overset{\cdot}{\bigcup}_{n=1}^{\infty} (a_n, b_n).$$

Then we have

$$A \subset U \cap G = \overset{\cdot}{\bigcup}_{k=1}^{K^*} (e_k, f_k) \text{ and } B \subset V \cap G = \overset{\cdot}{\bigcup}_{\ell=1}^{L^*} (g_\ell, h_\ell),$$

and then from Claim 4 and monotonicity of μ^* we obtain

$$\begin{aligned}
\mu^*(A)+\mu^*(B) &\leq \sum_{k=1}^{K^*}(f_k-e_k)+\sum_{\ell=1}^{L^*}(h_\ell-g_\ell) \\
&= \mu^*\left(\left(\dot{\bigcup}_{k=1}^{K^*}(e_k,f_k)\right)\dot{\cup}\left(\dot{\bigcup}_{\ell=1}^{L^*}(g_\ell,h_\ell)\right)\right) \\
&\leq \mu^*(G)=\sum_{n=1}^{\infty}(b_n-a_n).
\end{aligned}$$

Taking the infimum over such G gives $\mu^*(A)+\mu^*(B)\leq\mu^*(A\cup B)$, and subadditivity of μ^* now proves equality.

Proof (of Theorem 27): We now prove that $\mathcal{L}$ is a σ-algebra and that $\mathcal{L}$ and μ satisfy the four properties in the statement of Theorem 27. First we establish that $\mathcal{L}$ is a σ-algebra in four steps.

Step 1: $A\in\mathcal{L}$ if $\mu^*(A)=0$.

Given $\varepsilon>0$, there is an open $G\supset A$ with $\mu^*(G)<\varepsilon$. But then $\mu^*(G\setminus A)\leq\mu^*(G)<\varepsilon$ by monontonicity.

Step 2: $\bigcup_{n=1}^{\infty}A_n\in\mathcal{L}$ whenever $A_n\in\mathcal{L}$ for all n.

Given $\varepsilon>0$, there is an open $G_n\supset A_n$ with $\mu^*(G_n\setminus A_n)<\frac{\varepsilon}{2^n}$. Then $A\equiv\bigcup_{n=1}^{\infty}A_n$ is contained in the open set $G\equiv\bigcup_{n=1}^{\infty}G_n$, and since $G\setminus A$ is contained in $\bigcup_{n=1}^{\infty}(G_n\setminus A_n)$, monotonicity and subadditivity of μ^* yield

$$\mu^*(G\setminus A)\leq\mu^*\left(\bigcup_{n=1}^{\infty}(G_n\setminus A_n)\right)\leq\sum_{n=1}^{\infty}\mu^*(G_n\setminus A_n)<\sum_{n=1}^{\infty}\frac{\varepsilon}{2^n}=\varepsilon.$$

Step 3: $A\in\mathcal{L}$ if A is closed.

Suppose first that A is compact, and let $\varepsilon>0$. Then using Claim 4 there is $G=\dot{\bigcup}_{n=1}^{N^*}(a_n,b_n)$ containing A with

$$\mu^*(G)=\sum_{n=1}^{\infty}(b_n-a_n)\leq\mu^*(A)+\varepsilon<\infty.$$

Now $G\setminus A$ is open and so $G\setminus A=\dot{\bigcup}_{m=1}^{M^*}(c_m,d_m)$. We want to show that $\mu^*(G\setminus A)\leq\varepsilon$. Fix a finite $M\leq M^*$ and

$$0<\eta<\frac{1}{2}\min_{1\leq m\leq M}(d_m-c_m).$$

Then the compact set

$$K_\eta=\bigcup_{m=1}^{M}[c_m+\eta,d_m-\eta]$$

is disjoint from A, so by Claim 5 we have

$$\mu^*(A)+\mu^*(K_\eta)=\mu^*(A\cup K_\eta).$$

We conclude from subadditivity and $A \cup K_\eta \subset G$ that

$$\begin{aligned} \mu^*(A) + \sum_{m=1}^{M} (d_m - c_m - 2\eta) &= \mu^*(A) + \mu^*\left(\bigcup_{m=1}^{M} (c_m + \eta, d_m - \eta)\right) \\ &\le \mu^*(A) + \mu^*(K_\eta) \\ &= \mu^*(A \cup K_\eta) \\ &\le \mu^*(G) \le \mu^*(A) + \varepsilon. \end{aligned}$$

Since $\mu^*(A) < \infty$ for A compact, we thus have

$$\sum_{m=1}^{M} (d_m - c_m) \le \varepsilon + 2M\eta$$

for all $0 < \eta < \frac{1}{2} \min_{1 \le m \le M} (d_m - c_m)$. Hence $\sum_{m=1}^{M} (d_m - c_m) \le \varepsilon$ and taking the supremum in $M \le M^*$ we obtain from Claim 4 that

$$\mu^*(G \setminus A) = \sum_{m=1}^{M^*} (d_m - c_m) \le \varepsilon.$$

Finally, if A is closed, it is a countable union of compact sets $A = \bigcup_{n=1}^{\infty} ([-n, n] \cap A)$, and hence $A \in \mathcal{L}$ by Step 2.

Step 4: $A^c \in \mathcal{L}$ if $A \in \mathcal{L}$.

For each $n \ge 1$ there is by Claim 4 an open set $G_n \supset A$ such that $\mu^*(G_n \setminus A) < \frac{1}{n}$. Then $F_n \equiv G_n^c$ is closed and hence $F_n \in \mathcal{L}$ by Step 3. Thus

$$S \equiv \bigcup_{n=1}^{\infty} F_n \in \mathcal{L}, \quad S \subset A^c,$$

and $A^c \setminus S \subset G_n \setminus A$ for all n implies that

$$\mu^*(A^c \setminus S) \le \mu^*(G_n \setminus A) < \frac{1}{n}, \qquad n \ge 1.$$

Thus $\mu^*(A^c \setminus S) = 0$ and by Step 1 we have $A^c \setminus S \in \mathcal{L}$. Finally, Step 2 shows that

$$A^c = S \cup (A^c \setminus S) \in \mathcal{L}.$$

Thus far we have shown that $\mathcal{L}$ is a σ-algebra, and we now turn to proving that $\mathcal{L}$ and μ satisfy the four properties in Theorem 27. Property (1) is an easy exercise. Property (2) is the main event. Let $\{E_n\}_{n=1}^{\infty}$ be a pairwise disjoint sequence of sets in $\mathcal{L}$, and let $E = \dot{\bigcup}_{n=1}^{\infty} E_n$.

We will consider first the case where each of the sets E_n is bounded. Let $\varepsilon > 0$ be given. Then $E_n^c \in \mathcal{L}$ and so there are open sets $G_n \supset E_n^c$ such that

$$\mu^*(G_n \setminus E_n^c) < \frac{\varepsilon}{2^n}, \qquad n \ge 1.$$

Equivalently, with $F_n = G_n^c$, we have F_n closed, contained in E_n, and

$$\mu^*(E_n \setminus F_n) < \frac{\varepsilon}{2^n}, \qquad n \ge 1.$$

Thus the sets $\{F_n\}_{n=1}^{\infty}$ are compact and pairwise disjoint. Claim 5 and induction shows that

$$\sum_{n=1}^{N} \mu^* (F_n) = \mu^* \left(\bigcup_{n=1}^{N} F_n \right) \leq \mu^* (E), \qquad N \geq 1,$$

and taking the supremum over N yields

$$\sum_{n=1}^{\infty} \mu^* (F_n) \leq \mu^* (E).$$

Thus we have

$$\begin{aligned} \sum_{n=1}^{\infty} \mu^* (E_n) &\leq \sum_{n=1}^{\infty} \{\mu^* (E_n \setminus F_n) + \mu^* (F_n)\} \\ &\leq \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} + \sum_{n=1}^{\infty} \mu^* (F_n) \leq \varepsilon + \mu^* (E). \end{aligned}$$

Since $\varepsilon > 0$ we conclude that $\sum_{n=1}^{\infty} \mu^* (E_n) \leq \mu^* (E)$, and subadditivity of μ^* then proves equality.

In general, define $E_{n,k} = E_n \cap \{(-k-1, -k] \cup [k, k+1)\}$ for $k, n \geq 1$ so that

$$E = \dot{\bigcup}_{n=1}^{\infty} E_n = \dot{\bigcup}_{n,k=1}^{\infty} E_{n,k}.$$

Then from what we just proved we have

$$\mu^* (E) = \sum_{n,k=1}^{\infty} \mu^* (E_{n,k}) = \sum_{n=1}^{\infty} \left(\sum_{k=1}^{\infty} \mu^* (E_{n,k}) \right) = \sum_{n=1}^{\infty} \mu^* (E_n).$$

Finally, property (3) follows from the observation that $E \subset \dot{\bigcup}_{n=1}^{\infty} (a_n, b_n)$ if and only if $E + x \subset \dot{\bigcup}_{n=1}^{\infty} (a_n + x, b_n + x)$. It is then obvious that $\mu^* (E + x) = \mu^* (E)$ and that $E + x \in \mathcal{L}$ if $E \in \mathcal{L}$. Property (4) is immediate from Step 1 above. This completes the proof of Theorem 27.

2. Measurable functions and integration

Let $[-\infty, \infty] = \mathbb{R} \cup \{-\infty, \infty\}$ be the extended real numbers with order and (some) algebra operations defined by

$$\begin{aligned} -\infty &< x < \infty, \qquad x \in \mathbb{R}, \\ x + \infty &= \infty, \qquad x \in \mathbb{R}, \\ x - \infty &= -\infty, \qquad x \in \mathbb{R}, \\ x \cdot \infty &= \infty, \qquad x > 0, \\ x \cdot \infty &= -\infty, \qquad x < 0, \\ 0 \cdot \infty &= 0. \end{aligned}$$

The final assertion $0 \cdot \infty = 0$ is dictated by $\sum_{n=1}^{\infty} a_n = 0$ if all the $a_n = 0$. It turns out that these definitions give rise to a consistent theory of measure and integration of functions with values in the extended real number system.

Let $f : \mathbb{R} \to [-\infty, \infty]$. We say that f is (Lebesgue) measurable if

$$f^{-1} ([-\infty, x)) \in \mathcal{L}, \qquad x \in \mathbb{R}.$$

The simplest examples of measurable functions are the characteristic functions χ_E of measurable sets E. Indeed,

$$(\chi_E)^{-1}([-\infty, x)) = \begin{cases} \phi & \text{if} \quad x \leq 0 \\ E^c & \text{if} \quad 0 < x \leq 1 \\ \mathbb{R} & \text{if} \quad x > 1 \end{cases}.$$

It is then easy to see that finite linear combinations $s = \sum_{n=1}^{N} a_n \chi_{E_n}$ of such characteristic functions χ_{E_n}, called simple functions, are also measurable. Here $a_n \in \mathbb{R}$ and E_n is a measurable subset of $\mathbb{R}$. It turns out that if we define the integral of a simple function $s = \sum_{n=1}^{N} a_n \chi_{E_n}$ by

$$\int_{\mathbb{R}} s = \sum_{n=1}^{N} a_n \mu(E_n),$$

the value is independent of the representation of s as a simple function. Armed with this fact we can then extend the definition of integral $\int_{\mathbb{R}} f$ to functions f that are nonnegative on $\mathbb{R}$, and then to functions f such that $\int_{\mathbb{R}} |f| < \infty$.

At each stage one establishes the relevant properties of the integral along with the most useful theorems. For the most part these extensions are rather routine, the cleverness inherent in the theory being in the overarching organization of the concepts rather than in the details of the demonstrations. As a result, we will merely state the main results in logical order and sketch proofs when not simply routine. We will however give fairly detailed proofs of the three famous convergence theorems, the Monotone Convergence Theorem, Fatou's Lemma, and the Dominated Convergence Theorem. The reader is referred to the excellent exposition in [**6**] for the complete story including many additional fascinating insights.

2.1. Properties of measurable functions. From now on we denote the Lebesgue measure of a measurable subset E of $\mathbb{R}$ by $|E|$ rather than by $\mu(E)$ as in the previous sections. We say that two measurable functions $f, g : \mathbb{R} \to [-\infty, \infty]$ are equal *almost everywhere* (often abbreviated *a.e.*) if

$$|\{x \in \mathbb{R} : f(x) \neq g(x)\}| = 0.$$

We say that f is *finite-valued* if $f : \mathbb{R} \to \mathbb{R}$. We now collect a number of elementary properties of measurable functions.

LEMMA 13. *Suppose that $f, f_n, g : \mathbb{R} \to [-\infty, \infty]$ for $n \in \mathbb{N}$.*

(1) *If f is finite-valued, then f is measurable* if and only if *$f^{-1}(G) \in \mathcal{L}$ for all open sets $G \subset \mathbb{R}$* if and only if *$f^{-1}(F) \in \mathcal{L}$ for all closed sets $F \subset \mathbb{R}$.*
(2) *If f is finite-valued and continuous, then f is measurable.*
(3) *If f is finite-valued and measurable and $\Phi : \mathbb{R} \to \mathbb{R}$ is continuous, then $\Phi \circ f$ is measurable.*
(4) *If $\{f_n\}_{n=1}^{\infty}$ is a sequence of measurable functions, then the following functions are all measurable:*
$$\sup_n f_n(x), \quad \inf_n f_n(x), \ldots \limsup_{n\to\infty} f_n(x), \quad \liminf_{n\to\infty} f_n(x).$$
(5) *If $\{f_n\}_{n=1}^{\infty}$ is a sequence of measurable functions and $f(x) = \lim_{n\to\infty} f_n(x)$, then f is measurable.*
(6) *If f is measurable, so is f^n for $n \in \mathbb{N}$.*
(7) *If f and g are finite-valued and measurable, then so are $f + g$ and fg.*

(8) *If f is measurable and $f = g$ almost everywhere, then g is measurable.*

Comments: For property (1), first show that f is measurable if and only if $f^{-1}((a,b)) \in \mathcal{L}$ for all $-\infty < a < b < \infty$. For property (3) use $(\Phi \circ f)^{-1}(G) = f^{-1}\left(\Phi^{-1}(G)\right)$ and note that $\Phi^{-1}(G)$ is open if G is open. For property (7), use

$$\begin{aligned} \{f+g>a\} &= \bigcup_{r\in\mathbb{Q}} \left[\{f>a-r\} \cap \{g>r\}\right], \qquad a \in \mathbb{R}, \\ fg &= \frac{1}{4}\left[(f+g)^2 - (f-g)^2\right]. \end{aligned}$$

Recall that a measurable simple function φ (i.e. the range of φ is finite) has the form

$$\varphi = \sum_{k=1}^{N} \alpha_k \chi_{E_k}, \qquad \alpha_k \in \mathbb{R}, E_k \in \mathcal{L}.$$

Next we collect two approximation properties of simple functions.

Proposition 19. *Let $f : \mathbb{R} \to [-\infty, \infty]$ be measurable.*

(1) *If f is nonnegative there is an increasing sequence of nonnegative simple functions $\{\varphi_k\}_{k=1}^{\infty}$ that converges pointwise and monotonically to f:*

$$\varphi_k(x) \le \varphi_{k+1}(x) \text{ and } \lim_{k\to\infty} \varphi_k(x) = f(x), \qquad \text{for all } x \in \mathbb{R}.$$

(2) *There is a sequence of simple functions $\{\varphi_k\}_{k=1}^{\infty}$ satisfying*

$$|\varphi_k(x)| \le |\varphi_{k+1}(x)| \text{ and } \lim_{k\to\infty} \varphi_k(x) = f(x), \qquad \text{for all } x \in \mathbb{R}.$$

Comments: To prove (1) let $f_M = \min\{f, M\}$, and for $0 \le n < NM$ define

$$E_{n,N,M} = \left\{x \in \mathbb{R} : \frac{n}{N} < f_M(x) \le \frac{n+1}{N}\right\}.$$

Then $\varphi_k(x) = \sum_{n=1}^{2^k k} \frac{n}{2^k} \chi_{E_{n,2^k,k}}(x)$ works. Property (2) is routine given (1).

2.2. Properties of integration and convergence theorems. If φ is a measurable simple function (i.e. its range is a finite set), then φ has a unique canonical representation

$$\varphi = \sum_{k=1}^{N} \alpha_k \chi_{E_k},$$

where the real constants α_k are distinct and nonzero, and the measurable sets E_k are pairwise disjoint. We define the Lebesgue integral of φ by

$$\int \varphi(x)\,dx = \sum_{k=1}^{N} \alpha_k |E_k|.$$

If E is a measurable subset of $\mathbb{R}$ and φ is a measurable simple function, then so is $\chi_E \varphi$, and we define

$$\int_E \varphi(x)\,dx = \int (\chi_E \varphi)(x)\,dx.$$

Lemma 14. *Suppose that φ and ψ are measurable simple functions and that $E, F \in \mathcal{L}$.*

(1) *If* $\varphi = \sum_{k=1}^{M} \beta_k \chi_{F_k}$ *(not necessarily the canonical representation), then*

$$\int \varphi(x)\, dx = \sum_{k=1}^{M} \beta_k |F_k| .$$

(2) $\int (a\varphi + b\psi) = a \int \varphi + b \int \psi$ *for* $a, b \in \mathbb{C}$,
(3) $\int_{E\cup F} \varphi = \int_E \varphi + \int_F \varphi$ *if* $E \cap F = \phi$,
(4) $\int \varphi \leq \int \psi$ *if* $\varphi \leq \psi$,
(5) $\left|\int \varphi\right| \leq \int |\varphi|$.

Properties (2) - (5) are usually referred to as *linearity*, *additivity*, *monotonicity* and the *triangle inequality* respectively. The proofs are routine.

Now we turn to defining the integral of a nonnegative measurable function $f : \mathbb{R} \to [0, \infty]$. For such f we define

$$\int f(x)\, dx = \sup \left\{ \int g(x)\, dx : 0 \leq \varphi \leq f \text{ and } \varphi \text{ is simple} \right\}.$$

It is essential here that f be permitted to take on the value ∞, and that the supremum may be ∞ as well. We say that f is (Lebesgue) *integrable* if $\int f(x)\, dx < \infty$. For E measurable define

$$\int_E f(x)\, dx = \int (\chi_E f)(x)\, dx.$$

Here is an analogue of Lemma 14 whose proof is again routine.

Lemma 15. *Suppose that* $f, g : \mathbb{R} \to [0, \infty]$ *are nonnegative measurable functions and that* $E, F \in \mathcal{L}$.

(1) $\int (af + bg) = a \int f + b \int g$ *for* $a, b \in (0, \infty)$,
(2) $\int_{E\cup F} f = \int_E f + \int_F f$ *if* $E \cap F = \phi$,
(3) $\int f \leq \int g$ *if* $0 \leq f \leq g$,
(4) *If* $\int f < \infty$, *then* $f(x) < \infty$ *for a.e.* x,
(5) *If* $\int f = 0$, *then* $f(x) = 0$ *for a.e.* x.

Note that convergence of integrals does not always follow from pointwise convergence of the integrands. For example,

$$\lim_{n\to\infty} \int \chi_{[n,n+1]}(x)\, dx = 1 \neq 0 = \int \lim_{n\to\infty} \chi_{[n,n+1]}(x)\, dx,$$

and

$$\lim_{n\to\infty} \int n\chi_{\left(0,\frac{1}{n}\right)}(x)\, dx = 1 \neq 0 = \int \lim_{n\to\infty} n\chi_{\left[0,\frac{1}{n}\right]}(x)\, dx.$$

In each of these examples, the mass of the integrands "disappears" in the limit; at "infinity" in the first example and at the origin in the second example. Here are our first two classical convergence theorems giving conditions under which convergence *does* hold.

Theorem 28. *(Monotone Convergence Theorem) Suppose that* $\{f_n\}_{n=1}^{\infty}$ *is an* increasing *sequence of nonnegative measurable functions, i.e.* $f_n(x) \leq f_{n+1}(x)$, *and let*

$$f(x) = \sup_n f_n(x) = \lim_{n\to\infty} f_n(x).$$

Then f is nonegative and measurable and

$$\int f(x)\,dx = \lim_{n\to\infty}\int f_n(x)\,dx.$$

Proof: Since $\int f_n \le \int f_{n+1}$ we have $\lim_{n\to\infty}\int f_n = L \in [0,\infty]$. Now f is measurable and $f_n \le f$ implies $\int f_n \le \int f$ so that

$$L \le \int f.$$

To prove the opposite inequality, momentarily fix a simple function φ such that $0 \le \varphi \le f$. Choose $c < 1$ and define

$$E_n = \{x \in \mathbb{R} : f_n(x) \ge c\varphi(x)\}, \qquad n \ge 1.$$

Then E_n is an increasing sequence of measurable sets with $\bigcup_{n=1}^{\infty} E_n = \mathbb{R}$. We have

$$\int f_n \ge \int_{E_n} f_n \ge c\int_{E_n}\varphi, \qquad n \ge 1.$$

Now let $\varphi = \sum_{k=1}^{N}\alpha_k\chi_{F_k}$ be the canonical representation of φ. Then

$$\int_{E_n}\varphi = \sum_{k=1}^{N}\alpha_k\,|E_n\cap F_k|\,,$$

and since $\lim_{n\to\infty}|E_n\cap F_k| = |F_k|$ by the fourth line in (1.1), we obtain that

$$\int_{E_n}\varphi = \sum_{k=1}^{N}\alpha_k\,|E_n\cap F_k| \to \sum_{k=1}^{N}\alpha_k\,|F_k| = \int\varphi$$

as $n\to\infty$. Altogether then we have

$$L = \lim_{n\to\infty}\int f_n \ge c\int\varphi$$

for all $c < 1$, which implies $L \ge \int\varphi$ for all simple φ with $0 \le \varphi \le f$, which implies $L \ge \int f$ as required.

Corollary 10. *Suppose that $a_k(x) \ge 0$ is measurable for $k \ge 1$. Then*

$$\int\sum_{k=1}^{\infty}a_k(x)\,dx = \sum_{k=1}^{\infty}\int a_k(x)\,dx.$$

To prove the corollary apply the Monotone Convergence Theorem to the sequence of partial sums $f_n(x) = \sum_{k=1}^{n}a_k(x)$.

Lemma 16. *(Fatou's Lemma) If $\{f_n\}_{n=1}^{\infty}$ is a sequence of* nonnegative *measurable functions, then*

$$\int\liminf_{n\to\infty}f_n(x)\,dx \le \liminf_{n\to\infty}\int f_n(x)\,dx.$$

Proof: Let $g_n(x) = \inf_{k\ge n}f_k(x)$ so that $g_n \le f_n$ and $\int g_n \le \int f_n$. Then $\{g_n\}_{n=1}^{\infty}$ is an increasing sequence of nonnegative measurable functions that converges pointwise to $\liminf_{n\to\infty}f_n(x)$. So the Monotone Convergence Theorem yields

$$\int\liminf_{n\to\infty}f_n(x)\,dx = \lim_{n\to\infty}\int g_n(x)\,dx \le \liminf_{n\to\infty}\int f_n(x)\,dx.$$

Finally, we can give an unambiguous meaning to the integral $\int f(x)\,dx$ in the case when f is *integrable*, by which we mean that f is measurable and $\int |f(x)|\,dx < \infty$. To do this we introduce the positive and negative parts of f:

$$f^+(x) = \max\{f(x), 0\} \text{ and } f_-(x) = \max\{-f(x), 0\}.$$

Then both f^+ and f_- are nonnegative measurable functions with finite integral. We define

$$\int f(x)\,dx = \int f^+(x)\,dx - \int f_-(x)\,dx.$$

With this definition we have the usual elementary properties of linearity, additivity, monotonicity and the triangle inequality.

LEMMA 17. *Suppose that f, g are integrable and that $E, F \in \mathcal{L}$.*

1. $\int (af + bg) = a\int f + b\int g$ *for* $a, b \in \mathbb{R}$,
2. $\int_{E\cup F} f = \int_E f + \int_F f$ *if* $E \cap F = \phi$,
3. $\int f \leq \int g$ *if* $f \leq g$,
4. $\left|\int f\right| \leq \int |f|$.

Our final convergence theorem is one of the most useful in analysis.

THEOREM 29. *(Dominated Convergence Theorem) Let g be a nonnegative integrable function. Suppose that $\{f_n\}_{n=1}^\infty$ is a sequence of measurable functions satisfying*

$$\lim_{n\to\infty} f_n(x) = f(x), \qquad a.e.\ x,$$

and

$$|f_n(x)| \leq g(x), \qquad a.e.\ x.$$

Then

$$\lim_{n\to\infty} \int |f(x) - f_n(x)|\,dx = 0,$$

and hence

$$\int f(x)\,dx = \lim_{n\to\infty} \int f_n(x)\,dx.$$

Proof: Since $|f| \leq g$ and f is measurable, f is integrable. Since $|f - f_n| \leq 2g$, Fatou's Lemma can be applied to the sequence of functions $2g - |f - f_n|$ to obtain

$$\begin{aligned} \int 2g &\leq \liminf_{n\to\infty} \int (2g - |f - f_n|) \\ &= \int 2g + \liminf_{n\to\infty} \left(-\int |f - f_n|\right) \\ &= \int 2g - \limsup_{n\to\infty} \int |f - f_n|. \end{aligned}$$

Since $\int 2g < \infty$, we can subtract it from both sides to obtain

$$\limsup_{n\to\infty} \int |f - f_n| \leq 0,$$

which implies $\lim_{n\to\infty} \int |f - f_n| = 0$. Then $\int f = \lim_{n\to\infty} \int f_n$ follows from the triangle inequality $\left|\int (f - f_n)\right| \leq \int |f - f_n|$.

Finally, if $f(x) = u(x) + iv(x)$ is complex-valued where $u(x)$ and $v(x)$ are real-valued measurable functions such that

$$\int |f(x)|\, dx = \int \sqrt{u(x)^2 + v(x)^2}\, dx < \infty,$$

then we define

$$\int f(x)\, dx = \int u(x)\, dx + i \int v(x)\, dx.$$

The usual properties of linearity, additivity, monotonicity and the triangle inequality all hold for this definition as well.

2.3. Three famous measure problems. The following three problems are listed in order of increasing difficulty.

PROBLEM 3. *Suppose that* $E_1, ..., E_n$ *are* n *Lebesgue measurable subsets of* $[0,1]$ *such that each point* x *in* $[0,1]$ *lies in some* k *of these subsets. Prove that there is at least one set* E_j *with* $|E_j| \geq \frac{k}{n}$.

PROBLEM 4. *Suppose that* E *is a Lebesgue measurable set of positive measure. Prove that*

$$E - E = \{x - y : x, y \in E\}$$

contains a nontrivial open interval.

PROBLEM 5. *Construct a Lebesgue measurable subset of the real line such that*

$$0 < \frac{|E \cap I|}{|I|} < 1$$

for all nontrivial open intervals I.

To solve Problem 3, note that the hypothesis implies $k \leq \sum_{j=1}^n \chi_{E_j}(x)$ for $x \in [0,1]$. Now integrate to obtain

$$k = \int_0^1 k dx \leq \int_0^1 \left(\sum_{j=1}^n \chi_{E_j}(x) \right) dx = \sum_{j=1}^n \int_0^1 \chi_{E_j}(x)\, dx = \sum_{j=1}^n |E_j|,$$

which implies that $|E_j| \geq \frac{k}{n}$ for some j. The solution is much less elegant without recourse to integration.

To solve Problem 4, choose K compact contained in E such that $|K| > 0$. Then choose G open containing K such that $|G \setminus K| < |K|$. Let $\delta = dist(K, G^c) > 0$. It follows that $(-\delta, \delta) \subset K - K \subset E - E$. Indeed, if $x \in (-\delta, \delta)$ then $K - x \subset G$ and $K \cap (K - x) \neq \phi$ since otherwise we have a contradiction:

$$2|K| = |K| + |K - x| \leq |G| \leq |G \setminus K| + |K| < 2|K|.$$

Thus there are k_1 and k_2 in K such that $k_1 = k_2 - x$ and so

$$x = k_2 - k_1 \in K - K.$$

Problem 5 is most easily solved using generalized Cantor sets E_α. Let $0 < \alpha \leq 1$ and set $I_1^0 = [0,1]$. Remove the open interval of length $\frac{1}{3}\alpha$ centered in I_1^0 and denote the two remaining closed intervals by I_1^1 and I_2^1. Then remove the open interval of length $\frac{1}{3^2}\alpha$ centered in I_1^1 and denote the two remaining closed intervals by I_1^2 and I_2^2. Do the same for I_2^1 and denote the two remaining closed intervals by I_3^2 and I_4^2.

Continuing in this way, we obtain at the k^{th} generation, a collection $\left\{I_j^k\right\}_{j=1}^{2^k}$ of 2^k pairwise disjoint closed intervals of equal length. Let

$$E_\alpha = \bigcap_{k=1}^{\infty} \left(\bigcup_{j=1}^{2^k} I_j^k \right).$$

Then by summing the lengths of the removed open intervals, we obtain

$$|[0,1] \setminus E_\alpha| = \frac{1}{3}\alpha + \frac{2}{3^2}\alpha + \frac{2^2}{3^3}\alpha + ... = \alpha,$$

and it follows that E_α is compact and has Lebesgue measure $1-\alpha$. It is not hard to show that E_α is also nowhere dense. The case $\alpha = 1$ is particularly striking: E_1 is a compact, perfect and uncountable subset of $[0,1]$ having Lebesgue measure 0. This is the classical Cantor set.

In order to construct the set E in Problem 3, it suffices by taking unions of translates by integers, to construct a subset E of $[0,1]$ satisfying

$$0 < \frac{|E \cap I|}{|I|} < 1, \qquad \text{for all intervals } I \subset [0,1] \text{ of positive length.} \tag{2.1}$$

Fix $0 < \alpha_1 < 1$ and start by taking $E^1 = E_{\alpha_1}$. It is not hard to see that $\frac{|E^1 \cap I|}{|I|} < 1$ for all I, but the left hand inequality in (2.1) fails for $E = E^1$ whenever I is a subset of one of the component intervals in the open complement $[0,1] \setminus E^1$. To remedy this fix $0 < \alpha_2 < 1$ and for each component interval J of $[0,1] \setminus E^1$, translate and dilate E_{α_2} to fit snugly in the closure $\overline{J}$ of the component, and let E^2 be the union of E^1 and all these translates and dilates of E_{α_2}. Then again, $\frac{|E^2 \cap I|}{|I|} < 1$ for all I but the left hand inequality in (2.1) fails for $E = E^2$ whenever I is a subset of one of the component intervals in the open complement $[0,1] \setminus E^2$. Continue this process indefinitely with a sequence of numbers $\{\alpha_n\}_{n=1}^{\infty} \subset (0,1)$. We claim that $E = \bigcup_{n=1}^{\infty} E^n$ satisfies (2.1) if and only if

$$\sum_{n=1}^{\infty} (1 - \alpha_n) < \infty. \tag{2.2}$$

To see this, first note that no matter what sequence of numbers α_n *less than one* is used, we obtain that $0 < \frac{|E \cap I|}{|I|}$ for all intervals I of positive length. Indeed, each set E^n is easily seen to be compact and nowhere dense, and each component interval in the complement $[0,1] \setminus E^n$ has length at most

$$\frac{\alpha_1}{3} \frac{\alpha_2}{3} ... \frac{\alpha_n}{3} \leq 3^{-n}.$$

Thus given an interval I of positive length, there is n large enough such that I will contain one of the component intervals J of $[0,1] \setminus E^n$, and hence will contain the translated and dilated copy $\mathcal{C}\left(E_{\alpha_{n+1}}\right)$ of $E_{\alpha_{n+1}}$ that is fitted into J by construction. Since the dilation factor is the length $|J|$ of J, we have

$$|E \cap I| \geq \left|\mathcal{C}\left(E_{\alpha_{n+1}}\right)\right| = |J| \left|E_{\alpha_{n+1}}\right| = |J| \left(1 - \alpha_{n+1}\right) > 0,$$

since $\alpha_{n+1} < 1$.

It remains to show that $|E \cap I| < |I|$ for all intervals I of positive length in $[0,1]$, and it is here that we must use (2.2). Indeed, fix I and let J be a component interval of $[0,1] \setminus E^n$ (with n large) that is contained in I. Let $\mathcal{C}\left(E_{\alpha_{n+1}}\right)$ be the translated and dilated copy of $E_{\alpha_{n+1}}$ that is fitted into J by construction. We compute that

$$\begin{aligned}
|E \cap J| &= \left|\mathcal{C}\left(E_{\alpha_{n+1}}\right)\right| + (1-\alpha_{n+2})\left|J \setminus \mathcal{C}\left(E_{\alpha_{n+1}}\right)\right| + ... \\
&= (1-\alpha_{n+1})|J| + (1-\alpha_{n+2})(1-(1-\alpha_{n+1}))|J| \\
&\quad + (1-\alpha_{n+3})(1-(1-\alpha_{n+1}) - (1-\alpha_{n+2})(1-(1-\alpha_{n+1})))|J| + ... \\
&= \sum_{k=1}^{\infty} \beta_k^n |J|,
\end{aligned}$$

where by induction,

$$\beta_k^n = (1-\alpha_{n+k})\alpha_{n+k-1}...\alpha_{n+1}, \qquad k \geq 1.$$

Then we have

$$|E \cap J| = \left(\sum_{k=1}^{\infty} \beta_k^n\right)|J| < |J|,$$

and hence also $\frac{|E \cap I|}{|I|} < 1$, if we choose $\{\alpha_n\}_{n=1}^{\infty}$ so that $\sum_{k=1}^{\infty} \beta_k^n < 1$ for all n.

Now we have

$$\sum_{k=1}^{\infty} \beta_k^n = \sum_{k=1}^{\infty} (1-\alpha_{n+k})\alpha_{n+k-1}...\alpha_{n+1} = 1 - \prod_{k=1}^{\infty} \alpha_{n+k},$$

and by the first line in (0.5) of Chapter 5, this is strictly less than 1 *if and only if* $\sum_{k=1}^{\infty}(1-\alpha_{n+k}) < \infty$ for all n. Thus the set E constructed above satisfies (2.1) if and only if (2.2) holds.

Bibliography

[1] R. P. Boas, *Invitation to Complex Analysis*, Random House/Birkhäuser Mathematics Series, Random House, New York, 1987.

[2] A. Hatcher, *Algebraic Topology*, Cambridge University Press, 2002.

[3] R. Maehara, The Jordan curve theorem via the Brouwer fixed point theorem, Amer. Math. Monthly , (1984), 641-643.

[4] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, 3rd edition, 1976.

[5] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, 3rd edition, 1987.

[6] E. M. Stein and R. Shakarchi, *Complex Analysis, Princeton Lectures in Analysis* **II**, Princeton University Press, Princeton and Oxford, 2003.