GRADUATE STUDIES IN MATHEMATICS 202

# Introduction to Complex Analysis

Michael E. Taylor

American Mathematical Society

10.1090/gsm/202

GRADUATE STUDIES IN MATHEMATICS 202

# Introduction to Complex Analysis

Michael E. Taylor

American Mathematical Society
Providence, Rhode Island

2010 *Mathematics Subject Classification*. Primary 30-01.

For additional information and updates on this book, visit
**www.ams.org/bookpages/gsm-202**

**Library of Congress Cataloging-in-Publication Data**

Names: Taylor, Michael E., 1946– author.
Title: Introduction to complex analysis / Michael E. Taylor.
Description: Providence, Rhode Island: American Mathematical Society, [2019] | Series: Graduate studies in mathematics; volume 202 | Includes bibliographical references and index.
Identifiers: LCCN 2019019267 | ISBN 9781470452865 (alk. paper)
Subjects: LCSH: Functions of complex variables. | Numbers, Complex. | Mathematical analysis. | AMS: Functions of a complex variable – Instructional exposition (textbooks, tutorial papers, etc.). msc
Classification: LCC QA331.7 .T3925 2019 | DDC 515/.9–dc23
LC record available at https://lccn.loc.gov/2019019267



10 9 8 7 6 5 4 3 2 1 24 23 22 21 20 19

# Contents

# Preface

This text is designed for a first course in complex analysis for beginning graduate students or for well-prepared undergraduates whose background includes multivariable calculus, linear algebra, and advanced calculus. In this course the student will learn that all the basic functions that arise in calculus, first derived as functions of a real variable—such as powers and fractional powers, exponentials and logs, trigonometric functions and their inverses, as well as many new functions that the student will meet—are naturally defined for complex arguments. Furthermore, this expanded setting reveals a much richer understanding of such functions.

Care is taken to first introduce these basic functions in real settings. In the opening section on complex power series and exponentials, in Chapter 1, the exponential function is first introduced for real values of its argument as the solution to a differential equation. This is used to derive its power series, and from there extend it to complex argument. Similarly $\sin t$ and $\cos t$ are first given geometrical definitions for real angles and the Euler identity is established based on the geometrical fact that $e^{it}$ is a unit-speed curve on the unit circle for real $t$. Then one sees how to define $\sin z$ and $\cos z$ for complex $z$.

The central objects in complex analysis are functions that are complex-differentiable (i.e., holomorphic). One goal in the early part of the text is to establish an equivalence between being holomorphic and having a convergent power series expansion. Half of this equivalence, namely the holomorphy of convergent power series, is established in Chapter 1.

Chapter 2 starts with two major theoretical results: the Cauchy integral theorem and its corollary, the Cauchy integral formula. These theorems have a major impact on the rest of the text, including the demonstration that if a function $f(z)$ is holomorphic on a disk, then it is given by a convergent power series on that disk. A useful variant of such power series is the Laurent series for a function holomorphic on an annulus.

The text segues from Laurent series to Fourier series in Chapter 3 and from there to the Fourier transform and the Laplace transform. These three topics have many applications in analysis, such as constructing harmonic functions and providing other tools for differential equations. The Laplace transform of a function has the important property of being holomorphic on a half-space. It is convenient to have a treatment of the Laplace transform after the Fourier transform, since the Fourier inversion formula serves to motivate and provide a proof of the Laplace inversion formula.

Results on these transforms illuminate the material in Chapter 4. For example, these transforms are a major source of important definite integrals that one cannot evaluate by elementary means, but that are amenable to analysis by residue calculus, a key application of the Cauchy integral theorem. Chapter 4 starts with this and proceeds to the study of two important special functions: the Gamma function and the Riemann zeta function.

The Gamma function, which is the first "higher" transcendental function, is essentially a Laplace transform. The Riemann zeta function is a basic object of analytic number theory arising in the study of prime numbers. One sees in Chapter 4 the roles of Fourier analysis, residue calculus, and the Gamma function in the study of the zeta function. For example, a relation between Fourier series and the Fourier transform, known as the Poisson summation formula, plays an important role in its study.

In Chapter 5, the text takes a geometrical turn, viewing holomorphic functions as conformal maps. This notion is pursued not only for maps between planar domains but also for maps to surfaces in $\mathbb{R}^3$. The standard case is the unit sphere $S^2$ and the associated stereographic projection. The text also considers other surfaces. It constructs conformal maps from planar domains to general surfaces of revolution, deriving for the map a first-order differential equation, nonlinear but separable. These surfaces are discussed as examples of Riemann surfaces. The Riemann sphere $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ is also discussed as a Riemann surface, conformally equivalent to $S^2$. One sees the group of linear fractional transformations as a group of conformal automorphisms of $\widehat{\mathbb{C}}$ and certain subgroups as groups of conformal automorphisms of the unit disk and of the upper half-plane.

We also bring in the theory of normal families of holomorphic maps. We use this to prove the Riemann mapping theorem, which states that if $\Omega \subset \mathbb{C}$ is simply connected and $\Omega \neq \mathbb{C}$, then there is a holomorphic diffeomorphism $\Phi : \Omega \to D$, the unit disk. Application of this theorem to a special domain, together with a reflection argument, shows that there is a holomorphic covering of $\mathbb{C} \setminus \{0, 1\}$ by the unit disk. This leads to key results of Picard and Montel, applications to the behavior of iterations of holomorphic maps $R : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$, and the study of the associated Fatou and Julia sets, on which these iterates behave tamely and wildly, respectively.

The treatment of Riemann surfaces includes some differential geometric material. In an appendix to Chapter 5, we introduce the concept of a metric tensor and show how it is associated to a surface in Euclidean space and how the metric tensor behaves under smooth mappings, in particular how this behavior characterizes conformal mappings. We discuss the notion of metric tensors beyond the setting of metrics induced on surfaces in Euclidean space. In particular, we introduce a special metric on the unit disk, called the Poincaré metric, which has the property of being invariant under all conformal automorphisms of the disk. We show how the geometry of the Poincaré metric leads to another proof of Picard's theorem and also provides a different perspective on the proof of the Riemann mapping theorem.

The text next examines elliptic functions in Chapter 6. These are doubly periodic functions on $\mathbb{C}$, holomorphic except at poles (that is, meromorphic). Such a function can be regarded as a meromorphic function on the torus $\mathbb{T}_\Lambda = \mathbb{C}/\Lambda$, where $\Lambda \subset \mathbb{C}$ is a lattice. A prime example is the Weierstrass function $\wp_\Lambda(z)$, defined by a double series. Analysis shows that $\wp'_\Lambda(z)^2$ is a cubic polynomial in $\wp_\Lambda(z)$, so the Weierstrass function inverts an elliptic integral. Elliptic integrals arise in many situations in geometry and mechanics, including arclengths of ellipses and pendulum problems, to mention two basic cases. The analysis of general elliptic integrals leads to the problem of finding the lattice whose associated elliptic functions are related to these integrals. This is the Abel inversion problem. Section 6.5 of the text tackles this problem by constructing the Riemann surface associated to $\sqrt{p(z)}$, where $p(z)$ is a cubic or quartic polynomial.

Early in this text, the exponential function was defined by a differential equation and given a power series solution, and these two characterizations were used to develop its properties. Coming full circle, we devote Chapter 7 to other classes of differential equations and their solutions. We first study a special class of functions known as Bessel functions, characterized as solutions to Bessel equations. Part of the central importance of these functions arises from their role in producing solutions to partial differential equations in several variables, as explained in an appendix. The Bessel functions for real values of their arguments arise as solutions to wave equations, and for imaginary values of their arguments they arise as solutions to diffusion equations. Thus it is very useful that they can be understood as holomorphic functions of a complex variable. Next, Chapter 7 deals with more general differential equations on a complex domain. Results include constructing solutions as convergent power series and the analytic continuation of such solutions to larger domains. General results here are used to put the Bessel equations in a larger context. This includes a study of equations with "regular singular points." Other classes of equations with regular singular points are presented, particularly hypergeometric equations.

The text ends with a short collection of appendices. Some of these survey background material that the reader might have seen in an advanced calculus course, including material on convergence and compactness, and differential calculus of several variables. Others develop tools that prove useful in the text, such as the Laplace asymptotic method, the Stieltjes integral, and results on Abelian and Tauberian theorems. The last appendix shows how to solve cubic and quartic equations via radicals and introduces a special function, called the Bring radical, to treat quintic equations. (In Chapter 7 the Bring radical is shown to be given in terms of a generalized hypergeometric function.)

As indicated in the discussion above, while the first goal of this text is to present the beautiful theory of functions of a complex variable, we have the further objective of placing this study within a broader mathematical framework. Examples of how this text differs from many others in the area include the following.

1) A greater emphasis on Fourier analysis, both as an application of basic results in complex analysis and as a tool of more general applicability in analysis. We see the use of Fourier series in the study of harmonic functions. We see the influence of the Fourier transform on the study of the Laplace transform, and then the Laplace transform as a tool in the study of differential equations.

2) The use of geometrical techniques in complex analysis. This clarifies the study of conformal maps, extends the usual study to more general surfaces, and shows how geometrical concepts are effective in classical problems from the Riemann mapping theorem to Picard's theorem. An appendix discusses applications of the Poincaré metric on the disk.

3) Connections with differential equations. The use of techniques of complex analysis to study differential equations is a strong point of this text. This important area is frequently neglected in complex analysis texts, and the treatments one sees in many differential equations texts are often confined to solutions for real variables and may furthermore lack a complete analysis of crucial convergence issues. Material here also provides a more detailed study than one usually sees of significant examples, such as Bessel functions.

4) Special functions. In addition to material on the Gamma function and the Riemann zeta function, the text has a detailed study of elliptic functions

and Bessel functions and also material on Airy functions, Legendre functions, and hypergeometric functions.

We follow this introduction with a record of some standard notation that will be used throughout this text.

**Acknowledgments**
Thanks to Shrawan Kumar for testing this text in his Complex Analysis course, for pointing out corrections, and for other valuable advice.

During the preparation of this book, my research has been supported by a number of NSF grants, most recently DMS-1500817.

# Some basic notation

$\mathbb{R}$ is the set of real numbers.

$\mathbb{C}$ is the set of complex numbers.

$\mathbb{Z}$ is the set of integers.

$\mathbb{Z}^+$ is the set of integers $\geq 0$.

$\mathbb{N}$ is the set of integers $\geq 1$ (the "natural numbers").

$x \in \mathbb{R}$ means $x$ is an element of $\mathbb{R}$, i.e., $x$ is a real number.

$(a, b)$ denotes the set of $x \in \mathbb{R}$ such that $a < x < b$.

$[a, b]$ denotes the set of $x \in \mathbb{R}$ such that $a \leq x \leq b$.

$\{x \in \mathbb{R} : a \leq x \leq b\}$ denotes the set of $x$ in $\mathbb{R}$ such that $a \leq x \leq b$.

$[a, b) = \{x \in \mathbb{R} : a \leq x < b\}$ and $(a, b] = \{x \in \mathbb{R} : a < x \leq b\}$.

$\overline{z} = x - iy$ if $z = x + iy \in \mathbb{C},\ x, y \in \mathbb{R}$.

$\overline{\Omega}$ denotes the closure of the set $\Omega$.

$f : A \to B$ denotes that the function $f$ takes points in the set $A$ to points in $B$. One also says $f$ maps $A$ to $B$.

$x \to x_0$ means the variable $x$ tends to the limit $x_0$.

$f(x) = O(x)$ means $f(x)/x$ is bounded. Similarly $g(\varepsilon) = O(\varepsilon^k)$ means $g(\varepsilon)/\varepsilon^k$ is bounded.

$f(x) = o(x)$ as $x \to 0$ (resp., $x \to \infty$) means $f(x)/x \to 0$ as $x$ tends to the specified limit.

$S = \sup_n |a_n|$ means $S$ is the smallest real number that satisfies $S \geq |a_n|$ for all $n$. If there is no such real number then we take $S = +\infty$.

$\limsup_{k\to\infty} |a_k| = \lim_{n\to\infty} \big(\sup_{k\geq n} |a_k|\big)$.

# Complementary material

In addition to various appendices scattered through Chapters 1–7, we have six "global" appendices, collected here.

In Appendix A.1 we cover material on metric spaces and compactness, such as what one might find in a good advanced calculus course (cf. [**44**] and [**45**]). This material applies both to subsets of the complex plane and to various sets of functions. In the latter category, we have the Arzela-Ascoli theorem, which is an important ingredient in the theory of normal families. We also have the Contraction Mapping Theorem, of use in Appendix A.2.

In Appendix A.2 we discuss the derivative of a function of several real variables and prove the Inverse Function Theorem, in the real context, which is of use in §1.5 of Chapter 1 to get the Inverse Function Theorem for holomorphic functions on domains in $\mathbb{C}$. It is also useful for the treatment of surfaces in Chapter 5.

Appendix A.3 treats a method of analyzing an integral of the form

$$\int_{-\infty}^{\infty} e^{-t\varphi(x)} g(x)\, dx \tag{A.0.1}$$

for large $t$, known as the Laplace asymptotic method. This is applied here to analyze the behavior of $\Gamma(z)$ for large $z$ (Stirling's formula). Also, in §7.1 of Chapter 7, this method is applied to analyze the behavior of Bessel functions for a large argument.

Appendix A.4 provides some basic results on the Stieltjes integral

$$\int_a^b f(x)\,du(x). \tag{A.0.2}$$

We assume that $f \in C([a,b])$ and $u : [a,b] \to \mathbb{R}$ is increasing. Possibly $b = \infty$, and then there are restrictions on the behavior of $f$ and $u$ at infinity. The Stieltjes integral provides a convenient language to use to relate functions that count primes to the Riemann zeta function, and we make use of it in §4.4 of Chapter 4. It also provides a convenient setting for the material in Appendix A.5.

Appendix A.5 deals with Abelian theorems and Tauberian theorems. These are results to the effect that one sort of convergence implies another. In a certain sense, Tauberian theorems are partial converses to Abelian theorems. One source for such results is the following: in many proofs of the prime number theorem, including the one given in §4.4 of Chapter 4, the last step involves using a Tauberian theorem. The particular Tauberian theorem needed to end the analysis in §4.4 is given a short proof in Appendix A.5, as a consequence of a result of broad general use known as Karamata's Tauberian theorem.

In Appendix A.6 we show how the formula

$$\sin 3z = -4\sin^3 z + 3\sin z \tag{A.0.3}$$

enables one to solve cubic equations, and move on to seek formulas for solutions to quartic equations and quintic equations. In the latter case this cannot necessarily be done in terms of radicals, and this appendix introduces a special function, called the Bring radical, to treat quintic equations.

## A.1. Metric spaces, convergence, and compactness

A metric space is a set $X$, together with a distance function $d : X \times X \to [0,\infty)$, having the properties that

$$\begin{aligned} d(x,y) = 0 &\Longleftrightarrow x = y, \\ d(x,y) &= d(y,x), \\ d(x,y) &\le d(x,z) + d(y,z). \end{aligned} \tag{A.1.1}$$

The third of these properties is called the triangle inequality. An example of a metric space is the set of rational numbers $\mathbb{Q}$, with $d(x,y) = |x-y|$. Another example is $X = \mathbb{R}^n$, with

$$d(x,y) = \sqrt{(x_1-y_1)^2 + \cdots + (x_n-y_n)^2}. \tag{A.1.2}$$

If $(x_\nu)$ is a sequence in $X$, indexed by $\nu = 1,2,3,\dots$, i.e., by $\nu \in \mathbb{N}$, one says $x_\nu \to y$ if $d(x_\nu, y) \to 0$, as $\nu \to \infty$. One says $(x_\nu)$ is a Cauchy sequence if $d(x_\nu, x_\mu) \to 0$ as $\mu, \nu \to \infty$. One says $X$ is a *complete* metric space if every Cauchy sequence converges to a limit in $X$. Some metric spaces are not complete; for example, $\mathbb{Q}$ is not complete. You can take a sequence $(x_\nu)$ of rational numbers such that $x_\nu \to \sqrt{2}$, which is not rational. Then $(x_\nu)$ is Cauchy in $\mathbb{Q}$, but it has no limit in $\mathbb{Q}$.

If a metric space $X$ is not complete, one can construct its completion $\widehat{X}$ as follows. Let an element $\xi$ of $\widehat{X}$ consist of an *equivalence class* of Cauchy sequences in $X$, where we say $(x_\nu) \sim (y_\nu)$ provided $d(x_\nu, y_\nu) \to 0$. We write the equivalence class containing $(x_\nu)$ as $[x_\nu]$. If $\xi = [x_\nu]$ and $\eta = [y_\nu]$, we can set $d(\xi, \eta) = \lim_{\nu\to\infty} d(x_\nu, y_\nu)$, and verify that this is well defined and makes $\widehat{X}$ a complete metric space.

If the completion of $\mathbb{Q}$ is constructed by this process, you get $\mathbb{R}$, the set of real numbers. This construction provides a good way to develop the basic theory of the real numbers. A detailed construction of $\mathbb{R}$ using this method is given in Chapter 1 of [**44**].

There are a number of useful concepts related to the notion of closeness. We define some of them here. First, if $p$ is a point in a metric space $X$ and $r \in (0, \infty)$, the set

$$B_r(p) = \{x \in X : d(x, p) < r\} \tag{A.1.3}$$

is called the open ball about $p$ of radius $r$. Generally, a *neighborhood* of $p \in X$ is a set containing such a ball, for some $r > 0$.

A set $U \subset X$ is called *open* if it contains a neighborhood of each of its points. The complement of an open set is said to be *closed.* The following result characterizes closed sets.

**Proposition A.1.1.** *A subset $K \subset X$ of a metric space $X$ is closed if and only if*

$$x_j \in K, \; x_j \to p \in X \Longrightarrow p \in K. \tag{A.1.4}$$

**Proof.** Assume $K$ is closed, $x_j \in K$, $x_j \to p$. If $p \notin K$, then $p \in X \setminus K$, which is open, so some $B_\varepsilon(p) \subset X \setminus K$, and $d(x_j, p) \geq \varepsilon$ for all $j$. This contradiction implies $p \in K$.

Conversely, assume (A.1.4) holds, and let $q \in U = X \setminus K$. If $B_{1/n}(q)$ is not contained in $U$ for any $n$, then there exists $x_n \in K \cap B_{1/n}(q)$, and hence $x_n \to q$, contradicting (A.1.4). This completes the proof. ☐

The following is straightforward.

**Proposition A.1.2.** *If $U_\alpha$ is a family of open sets in $X$, then $\bigcup_\alpha U_\alpha$ is open. If $K_\alpha$ is a family of closed subsets of $X$, then $\bigcap_\alpha K_\alpha$ is closed.*

Given $S \subset X$, we denote by $\overline{S}$ (the *closure* of $S$) the smallest closed subset of $X$ containing $S$, i.e., the intersection of all the closed sets $K_\alpha \subset X$ containing $S$. The following result is straightforward.

**Proposition A.1.3.** *Given $S \subset X$, $p \in \overline{S}$ if and only if there exist $x_j \in S$ such that $x_j \to p$.*

Given $S \subset X$, $p \in X$, we say $p$ is an *accumulation point* of $S$ if and only if, for each $\varepsilon > 0$, there exists $q \in S \cap B_\varepsilon(p)$, $q \neq p$. It follows that $p$ is an accumulation point of $S$ if and only if each $B_\varepsilon(p)$, $\varepsilon > 0$, contains infinitely many points of $S$. One straightforward observation is that all points of $\overline{S} \setminus S$ are accumulation points of $S$.

The *interior* of a set $S \subset X$ is the largest open set contained in $S$, i.e., the union of all the open sets contained in $S$. Note that the complement of the interior of $S$ is equal to the closure of $X \setminus S$.

We now turn to the notion of compactness. We say a metric space $X$ is *compact* provided the following property holds:

$$\text{Each sequence } (x_k) \text{ in } X \text{ has a convergent subsequence.} \tag{A.1.5}$$

We will establish various properties of compact metric spaces and provide various equivalent characterizations. For example, it is easily seen that (A.1.5) is equivalent to:

$$\text{Each infinite subset } S \subset X \text{ has an accumulation point.} \tag{A.1.6}$$

The following property is known as total boundedness:

**Proposition A.1.4.** *If $X$ is a compact metric space, then*

$$\begin{aligned} &\textit{Given } \varepsilon > 0, \ \exists \ \textit{finite set } \{x_1, \dots, x_N\} \\ &\textit{such that } B_\varepsilon(x_1), \dots, B_\varepsilon(x_N) \ \textit{covers } X. \end{aligned} \tag{A.1.7}$$

**Proof.** Take $\varepsilon > 0$ and pick $x_1 \in X$. If $B_\varepsilon(x_1) = X$, we are done. If not, pick $x_2 \in X \setminus B_\varepsilon(x_1)$. If $B_\varepsilon(x_1) \cup B_\varepsilon(x_2) = X$, we are done. If not, pick $x_3 \in X \setminus [B_\varepsilon(x_1) \cup B_\varepsilon(x_2)]$. Continue, taking $x_{k+1} \in X \setminus [B_\varepsilon(x_1) \cup \cdots \cup B_\varepsilon(x_k)]$ if $B_\varepsilon(x_1) \cup \cdots \cup B_\varepsilon(x_k) \neq X$. Note that, for $1 \le i, j \le k$,

$$i \neq j \Longrightarrow d(x_i, x_j) \ge \varepsilon. \tag{A.1.8}$$

If one never covers $X$ this way, consider $S = \{x_j : j \in \mathbb{N}\}$. This is an infinite set with no accumulation point, so property (A.1.6) is contradicted. □

**Corollary A.1.5.** *If $X$ is a compact metric space, it has a countable dense subset.*

**Proof.** Given $\varepsilon = 2^{-n}$, let $S_n$ be a finite set of points $x_j$ such that $\{B_\varepsilon(x_j)\}$ covers $X$. Then $\mathcal{C} = \bigcup_n S_n$ is a countable dense subset of $X$. □

Here is another useful property of compact metric spaces, which will eventually be generalized even further, in (A.1.13) below.

**Proposition A.1.6.** *Let $X$ be a compact metric space. Assume $K_1 \supset K_2 \supset K_3 \supset \cdots$ form a decreasing sequence of closed subsets of $X$. If each $K_n \neq \emptyset$, then $\bigcap_n K_n \neq \emptyset$.*

**Proof.** Pick $x_n \in K_n$. If (A) holds, $(x_n)$ has a convergent subsequence, $x_{n_k} \to y$. Since $\{x_{n_k} : k \geq \ell\} \subset K_{n_\ell}$, which is closed, we have $y \in \bigcap_n K_n$. □

**Corollary A.1.7.** *Let $X$ be a compact metric space. Assume $U_1 \subset U_2 \subset U_3 \subset \cdots$ form an increasing sequence of open subsets of $X$. If $\bigcup_n U_n = X$, then $U_N = X$ for some $N$.*

**Proof.** Consider $K_n = X \setminus U_n$. □

The following is an important extension of Corollary A.1.7.

**Proposition A.1.8.** *If $X$ is a compact metric space, then it has the property:*

$$\text{(A.1.9)} \qquad \textit{Every open cover } \{U_\alpha : \alpha \in \mathcal{A}\} \textit{ of } X \textit{ has a finite subcover.}$$

**Proof.** Each $U_\alpha$ is a union of open balls, so it suffices to show that (A.1.5) implies the following:

$$\text{(A.1.10)} \qquad \begin{array}{l}\text{Every cover } \{B_\alpha : \alpha \in \mathcal{A}\} \text{ of } X \text{ by open balls} \\ \text{has a finite subcover.}\end{array}$$

Let $\mathcal{C} = \{z_j : j \in \mathbb{N}\} \subset X$ be a countable dense subset of $X$, as in Corollary A.1.7. Each $B_\alpha$ is a union of balls $B_{r_j}(z_j)$, with $z_j \in \mathcal{C} \cap B_\alpha$, $r_j$ rational. Thus it suffices to show that

$$\text{(A.1.11)} \qquad \begin{array}{l}\text{Every countable cover } \{B_j : j \in \mathbb{N}\} \text{ of } X \\ \text{by open balls has a finite subcover.}\end{array}$$

For this, we set

$$\text{(A.1.12)} \qquad U_n = B_1 \cup \cdots \cup B_n$$

and apply Corollary A.1.7. □

The following is a convenient alternative to property (A.1.9):

$$\text{(A.1.13)} \qquad \begin{array}{l}\text{If } K_\alpha \subset X \text{ are closed and } \displaystyle\bigcap_\alpha K_\alpha = \emptyset, \\ \text{then some finite intersection is empty.}\end{array}$$

Considering $U_\alpha = X \setminus K_\alpha$, we see that

$$\text{(A.1.14)} \qquad \text{(A.1.9)} \Longleftrightarrow \text{(A.1.13)}.$$

The following result completes Proposition A.1.8.

**Theorem A.1.9.** *For a metric space $X$,*

$$\text{(A.1.5)} \Longleftrightarrow \text{(A.1.9)}. \tag{A.1.15}$$

**Proof.** By Proposition A.1.8, (A.1.5) ⇒ (A.1.9). To prove the converse, it will suffice to show that (A.1.13) ⇒ (A.1.6). So let $S \subset X$ and assume $S$ has no accumulation point. We claim:

$$\text{Such } S \text{ must be closed.} \tag{A.1.16}$$

Indeed, if $z \in \overline{S}$ and $z \notin S$, then $z$ would have to be an accumulation point. Say $S = \{x_\alpha : \alpha \in \mathcal{A}\}$. Set $K_\alpha = S \setminus \{x_\alpha\}$. Then each $K_\alpha$ has no accumulation point, hence $K_\alpha \subset X$ is closed. Also $\bigcap_\alpha K_\alpha = \emptyset$. Hence there exists a finite set $\mathcal{F} \subset \mathcal{A}$ such that $\bigcap_{\alpha \in \mathcal{F}} K_\alpha = \emptyset$, if (A.1.13) holds. Hence $S = \bigcup_{\alpha \in \mathcal{F}} \{x_\alpha\}$ is finite, so indeed (A.1.13) ⇒ (A.1.6). ☐

Remark. So far we have that for every metric space $X$,

$$\text{(A.1.5)} \Longleftrightarrow \text{(A.1.6)} \Longleftrightarrow \text{(A.1.9)} \Longleftrightarrow \text{(A.1.13)} \Longrightarrow \text{(A.1.7)}. \tag{A.1.17}$$

We claim that (A.1.7) implies the other conditions if $X$ is *complete*. Of course, compactness implies completeness, but (A.1.7) may hold for incomplete $X$, e.g., $X = (0,1) \subset \mathbb{R}$.

**Proposition A.1.10.** *If $X$ is a complete metric space with property* (A.1.7), *then $X$ is compact.*

**Proof.** It suffices to show that (A.1.7) ⇒ (A.1.6) if $X$ is a complete metric space. So let $S \subset X$ be an infinite set. Cover $X$ by a finite number of balls, $B_{1/2}(x_1), \dots, B_{1/2}(x_N)$. One of these balls contains infinitely many points of $S$, and so does its closure, say $X_1 = \overline{B_{1/2}(y_1)}$. Now cover $X$ by finitely many balls of radius $1/4$; their intersection with $X_1$ provides a cover of $X_1$. One such set contains infinitely many points of $S$, and so does its closure $X_2 = \overline{B_{1/4}(y_2)} \cap X_1$. Continue in this fashion, obtaining

$$X_1 \supset X_2 \supset X_3 \supset \cdots \supset X_k \supset X_{k+1} \supset \cdots, \quad X_j \subset \overline{B_{2^{-j}}(y_j)}, \tag{A.1.18}$$

each containing infinitely many points of $S$. One sees that $(y_j)$ forms a Cauchy sequence. If $X$ is complete, it has a limit, $y_j \to z$, and $z$ is seen to be an accumulation point of $S$. ☐

If $X_j$, $1 \le j \le m$, is a finite collection of metric spaces, with metrics $d_j$, we can define a Cartesian product metric space

$$X = \prod_{j=1}^{m} X_j, \quad d(x,y) = d_1(x_1,y_1) + \cdots + d_m(x_m,y_m). \tag{A.1.19}$$

Another choice of metric is $\delta(x,y) = \sqrt{d_1(x_1,y_1)^2 + \cdots + d_m(x_m,y_m)^2}$. The metrics $d$ and $\delta$ are *equivalent*, i.e., there exist constants $C_0, C_1 \in (0,\infty)$ such that

$$C_0\delta(x,y) \le d(x,y) \le C_1\delta(x,y), \quad \forall\, x,y \in X. \tag{A.1.20}$$

A key example is $\mathbb{R}^m$, the Cartesian product of $m$ copies of the real line $\mathbb{R}$.

We describe some important classes of compact spaces.

**Proposition A.1.11.** *If $X_j$ are compact metric spaces, $1 \le j \le m$, so is $X = \prod_{j=1}^m X_j$.*

**Proof.** If $(x_\nu)$ is an infinite sequence of points in $X$, say $x_\nu = (x_{1\nu}, \dots, x_{m\nu})$, pick a convergent subsequence of $(x_{1\nu})$ in $X_1$, and consider the corresponding subsequence of $(x_\nu)$, which we relabel $(x_\nu)$. Using this, pick a convergent subsequence of $(x_{2\nu})$ in $X_2$. Continue. Having a subsequence such that $x_{j\nu} \to y_j$ in $X_j$ for each $j = 1, \dots, m$, we then have a convergent subsequence in $X$. □

The following result is useful for calculus on $\mathbb{R}^n$.

**Proposition A.1.12.** *If $K$ is a closed bounded subset of $\mathbb{R}^n$, then $K$ is compact.*

**Proof.** The discussion above reduces the problem to showing that any closed interval $I = [a,b]$ in $\mathbb{R}$ is compact. This compactness is a corollary of Proposition A.1.10. For pedagogical purposes, we redo the argument here, since in this concrete case it can be streamlined.

Suppose $S$ is a subset of $I$ with infinitely many elements. Divide $I$ into 2 equal subintervals, $I_1 = [a,b_1]$, $I_2 = [b_1,b]$, $b_1 = (a+b)/2$. Then either $I_1$ or $I_2$ must contain infinitely many elements of $S$. Say $I_j$ does. Let $x_1$ be any element of $S$ lying in $I_j$. Now divide $I_j$ in two equal pieces, $I_j = I_{j1} \cup I_{j2}$. One of these intervals (say $I_{jk}$) contains infinitely many points of $S$. Pick $x_2 \in I_{jk}$ to be one such point (different from $x_1$). Then subdivide $I_{jk}$ into two equal subintervals, and continue. We get an infinite sequence of distinct points $x_\nu \in S$, and $|x_\nu - x_{\nu+k}| \le 2^{-\nu}(b-a)$, for $k \ge 1$. Since $\mathbb{R}$ is complete, $(x_\nu)$ converges, say to $y \in I$. Any neighborhood of $y$ contains infinitely many points in $S$, so we are done. □

If $X$ and $Y$ are metric spaces, a function $f : X \to Y$ is said to be continuous provided $x_\nu \to x$ in $X$ implies $f(x_\nu) \to f(x)$ in $Y$. An equivalent condition, which the reader is invited to verify, is

$$U \text{ open in } Y \Longrightarrow f^{-1}(U) \text{ open in } X. \tag{A.1.21}$$

**Proposition A.1.13.** *If $X$ and $Y$ are metric spaces, $f : X \to Y$ continuous, and $K \subset X$ compact, then $f(K)$ is a compact subset of $Y$.*

**Proof.** If $(y_\nu)$ is an infinite sequence of points in $f(K)$, pick $x_\nu \in K$ such that $f(x_\nu) = y_\nu$. If $K$ is compact, we have a subsequence $x_{\nu_j} \to p$ in $X$, and then $y_{\nu_j} \to f(p)$ in $Y$. □

If $F : X \to \mathbb{R}$ is continuous, we say $f \in C(X)$. A useful corollary of Proposition A.1.13 is:

**Proposition A.1.14.** *If $X$ is a compact metric space and $f \in C(X)$, then $f$ assumes a maximum and a minimum value on $X$.*

**Proof.** We know from Proposition A.1.13 that $f(X)$ is a compact subset of $\mathbb{R}$. Hence $f(X)$ is bounded, say $f(X) \subset I = [a, b]$. Repeatedly subdividing $I$ into equal halves, as in the proof of Proposition A.1.12, at each stage throwing out intervals that do not intersect $f(X)$, and keeping only the leftmost and rightmost interval amongst those remaining, we obtain points $\alpha \in f(X)$ and $\beta \in f(X)$ such that $f(X) \subset [\alpha, \beta]$. Then $\alpha = f(x_0)$ for some $x_0 \in X$ is the minimum and $\beta = f(x_1)$ for some $x_1 \in X$ is the maximum. □

If $S \subset \mathbb{R}$ is a nonempty, bounded set, Proposition A.1.12 implies $\overline{S}$ is compact. The function $\eta : \overline{S} \to \mathbb{R}$, $\eta(x) = x$ is continuous, so by Proposition A.1.14 it assumes a maximum and a minimum on $\overline{S}$. We set

$$\sup S = \max_{s \in \overline{S}} x, \quad \inf S = \min_{x \in \overline{S}} x, \tag{A.1.22}$$

when $S$ is bounded. More generally, if $S \subset \mathbb{R}$ is nonempty and bounded from above, say $S \subset (-\infty, B]$, we can pick $A < B$ such that $S \cap [A, B]$ is nonempty, and set

$$\sup S = \sup S \cap [A, B]. \tag{A.1.23}$$

Similarly, if $S \subset \mathbb{R}$ is nonempty and bounded from below, say $S \subset [A, \infty)$, we can pick $B > A$ such that $S \cap [A, B]$ is nonempty, and set

$$\inf S = \inf S \cap [A, B]. \tag{A.1.24}$$

If $X$ is a nonempty set and $f : X \to \mathbb{R}$ is bounded from above, we set

$$\sup_{x \in X} f(x) = \sup f(X), \tag{A.1.25}$$

and if $f : X \to \mathbb{R}$ is bounded from below, we set

$$\inf_{x \in X} f(x) = \inf f(X). \tag{A.1.26}$$

If $f$ is not bounded from above, we set $\sup f = +\infty$, and if $f$ is not bounded from below, we set $\inf f = -\infty$.

Given a set $X$, $f : X \to \mathbb{R}$, and $x_n \to x$, we set

$$\limsup_{n \to \infty} f(x_n) = \lim_{n \to \infty} \Bigl( \sup_{k \geq n} f(x_k) \Bigr) \tag{A.1.27}$$

and

$$\liminf_{n\to\infty} f(x_n) = \lim_{n\to\infty} \Bigl(\inf_{k\ge n} f(x_k)\Bigr). \tag{A.1.28}$$

We return to the notion of continuity. A function $f \in C(X)$ is said to be *uniformly continuous* provided that, for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$x, y \in X,\ d(x,y) \le \delta \Longrightarrow |f(x) - f(y)| \le \varepsilon. \tag{A.1.29}$$

An equivalent condition is that $f$ have a *modulus of continuity*, i.e., a monotonic function $\omega : [0,1) \to [0,\infty)$ such that $\delta \searrow 0 \Rightarrow \omega(\delta) \searrow 0$, and such that

$$x, y \in X,\ d(x,y) \le \delta \le 1 \Longrightarrow |f(x) - f(y)| \le \omega(\delta). \tag{A.1.30}$$

Not all continuous functions are uniformly continuous. For example, if $X = (0,1) \subset \mathbb{R}$, then $f(x) = \sin 1/x$ is continuous, but not uniformly continuous, on $X$. The following result is useful, for example, in the development of the Riemann integral.

**Proposition A.1.15.** *If $X$ is a compact metric space and $f \in C(X)$, then $f$ is uniformly continuous.*

**Proof.** If not, there exist $x_\nu, y_\nu \in X$ and $\varepsilon > 0$ such that $d(x_\nu, y_\nu) \le 2^{-\nu}$ but

$$|f(x_\nu) - f(y_\nu)| \ge \varepsilon. \tag{A.1.31}$$

Taking a convergent subsequence $x_{\nu_j} \to p$, we also have $y_{\nu_j} \to p$. Now continuity of $f$ at $p$ implies $f(x_{\nu_j}) \to f(p)$ and $f(y_{\nu_j}) \to f(p)$, contradicting (A.1.31). □

If $X$ and $Y$ are metric spaces, the space $C(X,Y)$ of continuous maps $f : X \to Y$ has a natural metric structure, under some additional hypotheses. We use

$$D(f,g) = \sup_{x\in X} d\bigl(f(x), g(x)\bigr). \tag{A.1.32}$$

This sup exists provided $f(X)$ and $g(X)$ are *bounded* subsets of $Y$, where to say $B \subset Y$ is bounded is to say $d : B \times B \to [0,\infty)$ has a bounded image. In particular, this supremum exists if $X$ is compact. The following is a natural completeness result.

**Proposition A.1.16.** *If $X$ is a compact metric space and $Y$ is a complete metric space, then $C(X,Y)$, with the metric* (A.1.32)*, is complete.*

**Proof.** That $D(f,g)$ satisfies the conditions to define a metric on $C(X,Y)$ is straightforward. We check completeness. Suppose $(f_\nu)$ is a Cauchy sequence in $C(X,Y)$, so, as $\nu \to \infty$,

$$\sup_{k\ge 0}\sup_{x\in X} d\big(f_{\nu+k}(x), f_\nu(x)\big) \le \varepsilon_\nu \to 0. \tag{A.1.33}$$

Then in particular $(f_\nu(x))$ is a Cauchy sequence in $Y$ for each $x \in X$, so it converges, say to $g(x) \in Y$. It remains to show that $g \in C(X,Y)$ and that $f_\nu \to g$ in the metric (A.1.24).

In fact, taking $k \to \infty$ in the estimate above, we have

$$\sup_{x\in X} d\big(g(x), f_\nu(x)\big) \le \varepsilon_\nu \to 0, \tag{A.1.34}$$

i.e., $f_\nu \to g$ uniformly. It remains only to show that $g$ is continuous. For this, let $x_j \to x$ in $X$ and fix $\varepsilon > 0$. Pick $N$ so that $\varepsilon_N < \varepsilon$. Since $f_N$ is continuous, there exists $J$ such that $j \ge J \Rightarrow d(f_N(x_j), f_N(x)) < \varepsilon$. Hence

(A.1.35)
$$\begin{aligned} j \ge J \Rightarrow d\big(g(x_j), g(x)\big) &\le d\big(g(x_j), f_N(x_j)\big) \\ &\quad + d\big(f_N(x_j), f_N(x)\big) + d\big(f_N(x), g(x)\big) < 3\varepsilon. \end{aligned}$$

This completes the proof. ☐

In case $Y = \mathbb{R}$, $C(X,\mathbb{R}) = C(X)$, introduced earlier in this appendix. The distance function (A.1.32) can be written

$$D(f,g) = \|f-g\|_{\sup}, \quad \|f\|_{\sup} = \sup_{x\in X} |f(x)|. \tag{A.1.36}$$

$\|f\|_{\sup}$ is a *norm* on $C(X)$.

Generally, a norm on a vector space $V$ is an assignment $f \mapsto \|f\| \in [0,\infty)$, satisfying

$$\|f\| = 0 \Leftrightarrow f = 0, \quad \|af\| = |a|\,\|f\|, \quad \|f+g\| \le \|f\| + \|g\|, \tag{A.1.37}$$

given $f, g \in V$ and $a$ a scalar (in $\mathbb{R}$ or $\mathbb{C}$). A vector space equipped with a norm is called a normed vector space. It is then a metric space, with distance function $D(f,g) = \|f-g\|$. If the space is complete, one calls $V$ a *Banach space.*

In particular, by Proposition A.1.16, $C(X)$ is a Banach space when $X$ is a compact metric space.

We next give a couple of slightly more sophisticated results on compactness. The following extension of Proposition A.1.11 is a special case of Tychonov's Theorem.

**Proposition A.1.17.** *If $\{X_j : j \in \mathbb{Z}^+\}$ are compact metric spaces, so is $X = \prod_{j=1}^\infty X_j$.*

Here, we can make $X$ a metric space by setting

$$d(x,y) = \sum_{j=1}^{\infty} 2^{-j} \frac{d_j(p_j(x), p_j(y))}{1 + d_j(p_j(x), p_j(y))}, \tag{A.1.38}$$

where $p_j : X \to X_j$ is the projection onto the $j$th factor. It is easy to verify that, if $x_\nu \in X$, then $x_\nu \to y$ in $X$, as $\nu \to \infty$, if and only if, for each $j$, $p_j(x_\nu) \to p_j(y)$ in $X_j$.

**Proof.** Following the argument in Proposition A.1.11, if $(x_\nu)$ is an infinite sequence of points in $X$, we obtain a nested family of subsequences

$$(x_\nu) \supset (x^1{}_\nu) \supset (x^2{}_\nu) \supset \cdots \supset (x^j{}_\nu) \supset \cdots \tag{A.1.39}$$

such that $p_\ell(x^j{}_\nu)$ converges in $X_\ell$, for $1 \le \ell \le j$. The next step is a *diagonal construction.* We set

$$\xi_\nu = x^\nu{}_\nu \in X. \tag{A.1.40}$$

Then, for each $j$, after throwing away a finite number $N(j)$ of elements, one obtains from $(\xi_\nu)$ a subsequence of the sequence $(x^j{}_\nu)$ in (A.1.39), so $p_\ell(\xi_\nu)$ converges in $X_\ell$ for all $\ell$. Hence $(\xi_\nu)$ is a convergent subsequence of $(x_\nu)$. □

The next result is known as the Arzela-Ascoli Theorem. It is useful in the theory of normal families, developed in §5.2.

**Proposition A.1.18.** *Let $X$ and $Y$ be compact metric spaces, and fix a modulus of continuity $\omega(\delta)$. Then*

$$\mathcal{C}_\omega = \big\{ f \in C(X,Y) : d\big(f(x), f(x')\big) \le \omega\big(d(x,x')\big) \ \forall\, x, x' \in X \big\} \tag{A.1.41}$$

*is a compact subset of $C(X,Y)$.*

**Proof.** Let $(f_\nu)$ be a sequence in $\mathcal{C}_\omega$. Let $\Sigma$ be a countable dense subset of $X$, as in Corollary A.1.5. For each $x \in \Sigma$, $(f_\nu(x))$ is a sequence in $Y$, which hence has a convergent subsequence. Using a diagonal construction similar to that in the proof of Proposition A.1.17, we obtain a subsequence $(\varphi_\nu)$ of $(f_\nu)$ with the property that $\varphi_\nu(x)$ converges in $Y$, for *each* $x \in \Sigma$, say

$$\varphi_\nu(x) \to \psi(x), \tag{A.1.42}$$

for all $x \in \Sigma$, where $\psi : \Sigma \to Y$.

So far, we have not used (A.1.41). This hypothesis will now be used to show that $\varphi_\nu$ converges uniformly on $X$. Pick $\varepsilon > 0$. Then pick $\delta > 0$ such that $\omega(\delta) < \varepsilon/3$. Since $X$ is compact, we can cover $X$ by finitely many balls $B_\delta(x_j)$, $1 \le j \le N$, $x_j \in \Sigma$. Pick $M$ so large that $\varphi_\nu(x_j)$ is within $\varepsilon/3$ of its limit for all $\nu \geq M$ (when $1 \leq j \leq N$). Now, for any $x \in X$, picking $\ell \in \{1, \dots, N\}$ such that $d(x, x_\ell) \leq \delta$, we have, for $k \geq 0$, $\nu \geq M$,

$$\begin{aligned} d\big(\varphi_{\nu+k}(x), \varphi_\nu(x)\big) &\leq d\big(\varphi_{\nu+k}(x), \varphi_{\nu+k}(x_\ell)\big) + d\big(\varphi_{\nu+k}(x_\ell), \varphi_\nu(x_\ell)\big) \\ &\quad + d\big(\varphi_\nu(x_\ell), \varphi_\nu(x)\big) \\ &\leq \varepsilon/3 + \varepsilon/3 + \varepsilon/3. \end{aligned} \tag{A.1.43}$$

Thus $(\varphi_\nu(x))$ is Cauchy in $Y$ for all $x \in X$, and hence convergent. Call the limit $\psi(x)$, so we now have (A.1.42) for all $x \in X$. Letting $k \to \infty$ in (A.1.43) we have uniform convergence of $\varphi_\nu$ to $\psi$. Finally, passing to the limit $\nu \to \infty$ in

$$d(\varphi_\nu(x), \varphi_\nu(x')) \leq \omega(d(x, x')) \tag{A.1.44}$$

gives $\psi \in \mathcal{C}_\omega$. □

We want to re-state Proposition A.1.18, bringing in the notion of *equicontinuity*. Given metric spaces $X$ and $Y$, and a set of maps $\mathcal{F} \subset C(X, Y)$, we say $\mathcal{F}$ is equicontinuous at a point $x_0 \in X$ provided

$$\begin{aligned} &\forall\, \varepsilon > 0,\ \exists\, \delta > 0 \text{ such that } \forall\, x \in X,\ f \in \mathcal{F}, \\ &d_X(x, x_0) < \delta \Longrightarrow d_Y(f(x), f(x_0)) < \varepsilon. \end{aligned} \tag{A.1.45}$$

We say $\mathcal{F}$ is equicontinuous on $X$ if it is equicontinuous at each point of $X$. We say $\mathcal{F}$ is *uniformly equicontinuous* on $X$ provided

$$\begin{aligned} &\forall\, \varepsilon > 0,\ \exists\, \delta > 0 \text{ such that } \forall\, x, x' \in X,\ f \in \mathcal{F}, \\ &d_X(x, x') < \delta \Longrightarrow d_Y(f(x), f(x')) < \varepsilon. \end{aligned} \tag{A.1.46}$$

Note that (A.1.46) is equivalent to the existence of a modulus of continuity $\omega$ such that $\mathcal{F} \subset \mathcal{C}_\omega$, given by (A.1.41). It is useful to record the following result.

**Proposition A.1.19.** *Let $X$ and $Y$ be metric spaces, $\mathcal{F} \subset C(X, Y)$. Assume $X$ is compact. Then*

$$\mathcal{F} \text{ equicontinuous} \Longrightarrow \mathcal{F} \text{ is uniformly equicontinuous.} \tag{A.1.47}$$

**Proof.** The argument is a variant of the proof of Proposition A.1.15. In more detail, suppose there exist $x_\nu, x'_\nu \in X$, $\varepsilon > 0$, and $f_\nu \in \mathcal{F}$ such that $d(x_\nu, x'_\nu) \leq 2^{-\nu}$ but

$$d(f_\nu(x_\nu), f_\nu(x'_\nu)) \geq \varepsilon. \tag{A.1.48}$$

Taking a convergent subsequence $x_{\nu_j} \to p \in X$, we also have $x'_{\nu_j} \to p$. Now equicontinuity of $\mathcal{F}$ at $p$ implies that there exists $N < \infty$ such that

$$d(g(x_{\nu_j}), g(p)) < \frac{\varepsilon}{2}, \quad \forall\, j \geq N,\ g \in \mathcal{F}, \tag{A.1.49}$$

contradicting (A.1.48). □

Putting together Propositions A.1.18 and A.1.19 then gives the following.

**Proposition A.1.20.** *Let $X$ and $Y$ be compact metric spaces. If $\mathcal{F} \subset C(X, Y)$ is equicontinuous on $X$, then it has compact closure in $C(X, Y)$.*

We next define the notion of a *connected* space. A metric space $X$ is said to be connected provided that it cannot be written as the union of two disjoint nonempty open subsets. The following is a basic class of examples.

**Proposition A.1.21.** *Each interval $I$ in $\mathbb{R}$ is connected.*

**Proof.** Suppose $A \subset I$ is nonempty, with nonempty complement $B \subset I$, and both sets are open. Take $a \in A$, $b \in B$; we can assume $a < b$. Let $\xi = \sup\{x \in [a, b] : x \in A\}$. This exists as a consequence of the basic fact that $\mathbb{R}$ is complete.

Now we obtain a contradiction, as follows. Since $A$ is closed $\xi \in A$. But then, since $A$ is open, there must be a neighborhood $(\xi - \varepsilon, \xi + \varepsilon)$ contained in $A$; this is not possible. □

We say $X$ is path-connected if, given any $p, q \in X$, there is a continuous map $\gamma : [0, 1] \to X$ such that $\gamma(0) = p$ and $\gamma(1) = q$. It is an easy consequence of Proposition A.1.21 that $X$ is connected whenever it is path-connected.

The next result, known as the Intermediate Value Theorem, is frequently useful.

**Proposition A.1.22.** *Let $X$ be a connected metric space and $f : X \to \mathbb{R}$ continuous. Assume $p, q \in X$, and $f(p) = a < f(q) = b$. Then, given any $c \in (a, b)$, there exists $z \in X$ such that $f(z) = c$.*

**Proof.** Under the hypotheses, $A = \{x \in X : f(x) < c\}$ is open and contains $p$, while $B = \{x \in X : f(x) > c\}$ is open and contains $q$. If $X$ is connected, then $A \cup B$ cannot be all of $X$; so any point in its complement has the desired property. □

The next result is known as the Contraction Mapping Principle, and it has many uses in analysis. In particular, we will use it in the proof of the Inverse Function Theorem, in Appendix A.2.

**Theorem A.1.23.** *Let $X$ be a complete metric space, and let $T : X \to X$ satisfy*

$$d(Tx, Ty) \le r\, d(x, y), \tag{A.1.50}$$

*for some $r < 1$. (We say $T$ is a contraction.) Then $T$ has a unique fixed point $x$. For any $y_0 \in X$, $T^k y_0 \to x$ as $k \to \infty$.*

**Proof.** Pick $y_0 \in X$ and let $y_k = T^k y_0$. Then $d(y_k, y_{k+1}) \le r^k\, d(y_0, y_1)$, so

$$\begin{aligned} d(y_k, y_{k+m}) &\le d(y_k, y_{k+1}) + \cdots + d(y_{k+m-1}, y_{k+m}) \\ &\le (r^k + \cdots + r^{k+m-1})\, d(y_0, y_1) \\ &\le r^k (1-r)^{-1}\, d(y_0, y_1). \end{aligned} \tag{A.1.51}$$

It follows that $(y_k)$ is a Cauchy sequence, so it converges; $y_k \to x$. Since $Ty_k = y_{k+1}$ and $T$ is continuous, it follows that $Tx = x$, i.e., $x$ is a fixed point. Uniqueness of the fixed point is clear from the estimate $d(Tx, Tx') \le r\, d(x, x')$, which implies $d(x, x') = 0$ if $x$ and $x'$ are fixed points. This proves Theorem A.1.23. □

## Exercises

1. If $X$ is a metric space, with distance function $d$, show that

$$|d(x,y) - d(x',y')| \le d(x,x') + d(y,y'), \tag{A.1.52}$$

and hence

$$d : X \times X \longrightarrow [0, \infty) \ \text{ is continuous.} \tag{A.1.53}$$

2. Let $\varphi : [0,\infty) \to [0,\infty)$ be a $C^2$ function. Assume

$$\varphi(0) = 0, \quad \varphi' > 0, \quad \varphi'' < 0. \tag{A.1.54}$$

Prove that if $d(x,y)$ is symmetric and satisfies the triangle inequality, so does

$$\delta(x,y) = \varphi(d(x,y)). \tag{A.1.55}$$

*Hint.* Show that such $\varphi$ satisfies $\varphi(s+t) \le \varphi(s) + \varphi(t)$, for $s, t \in \mathbb{R}^+$.

3. Show that the function $d(x,y)$ defined by (A.1.38) satisfies (A.1.1).
*Hint.* Consider $\varphi(r) = r/(1+r)$.

4. Let $X$ be a compact metric space. Assume $f_j, f \in C(X)$ and

$$f_j(x) \nearrow f(x), \quad \forall\, x \in X. \tag{A.1.56}$$

Prove that $f_j \to f$ uniformly on $X$. (This result is called Dini's theorem.)
*Hint.* For $\varepsilon > 0$, let $K_j(\varepsilon) = \{x \in X : f(x) - f_j(x) \ge \varepsilon\}$. Note that $K_j(\varepsilon) \supset K_{j+1}(\varepsilon) \supset \cdots$.

5. In the setting of (A.1.19), let

$$\delta(x,y) = \Big\{ d_1(x_1,y_1)^2 + \cdots + d_m(x_m,y_m)^2 \Big\}^{1/2}. \tag{A.1.57}$$

Show that

$$\delta(x,y) \le d(x,y) \le \sqrt{m}\, \delta(x,y). \tag{A.1.58}$$

6. Let $X$ and $Y$ be compact metric spaces. Show that if $\mathcal{F} \subset C(X,Y)$ is compact, then $\mathcal{F}$ is equicontinuous. (This is a converse to Proposition A.1.20.)

7. Recall that a Banach space is a complete normed linear space. Consider $C^1(I)$, where $I = [0,1]$, with norm

$$\|f\|_{C^1} = \sup_I |f| + \sup_I |f'|. \tag{A.1.59}$$

Show that $C^1(I)$ is a Banach space.

8. Let $\mathcal{F} = \{f \in C^1(I) : \|f\|_{C^1} \le 1\}$. Show that $\mathcal{F}$ has compact closure in $C(I)$. Find a function in the closure of $\mathcal{F}$ that is not in $C^1(I)$.

## A.2. Derivatives and diffeomorphisms

To start this section off, we define the derivative and discuss some of its basic properties. Let $\mathcal{O}$ be an open subset of $\mathbb{R}^n$, and $F : \mathcal{O} \to \mathbb{R}^m$ a continuous function. We say $F$ is differentiable at a point $x \in \mathcal{O}$, with derivative $L$, if $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation such that, for $y \in \mathbb{R}^n$, small,

$$F(x+y) = F(x) + Ly + R(x,y) \tag{A.2.1}$$

with

$$\frac{\|R(x,y)\|}{\|y\|} \to 0 \quad \text{as} \quad y \to 0. \tag{A.2.2}$$

We denote the derivative at $x$ by $DF(x) = L$. With respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$, $DF(x)$ is simply the matrix of partial derivatives,

$$DF(x) = \left(\frac{\partial F_j}{\partial x_k}\right), \tag{A.2.3}$$

so that if $v = (v_1, \dots, v_n)^t$ (regarded as a column vector). then

$$DF(x)v = \Bigl(\sum_k \frac{\partial F_1}{\partial x_k} v_k, \dots, \sum_k \frac{\partial F_m}{\partial x_k} v_k\Bigr)^t. \tag{A.2.4}$$

It will be shown below that $F$ is differentiable whenever all the partial derivatives exist and are *continuous* on $\mathcal{O}$. In such a case we say $F$ is a $C^1$ function on $\mathcal{O}$. More generally, $F$ is said to be $C^k$ if all its partial derivatives of order $\le k$ exist and are continuous. If $F$ is $C^k$ for all $k$, we say $F$ is $C^\infty$.

In (A.2.2), we can use the *Euclidean* norm on $\mathbb{R}^n$ and $\mathbb{R}^m$. This norm is defined by

$$\|x\| = \bigl(x_1^2 + \cdots + x_n^2\bigr)^{1/2} \tag{A.2.5}$$

for $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Any other norm would do equally well.

We now derive the *chain rule* for the derivative. Let $F : \mathcal{O} \to \mathbb{R}^m$ be differentiable at $x \in \mathcal{O}$, as above, let $U$ be a neighborhood of $z = F(x)$ in

$\mathbb{R}^m$, and let $G : U \to \mathbb{R}^k$ be differentiable at $z$. Consider $H = G \circ F$. We have

$$\begin{aligned} H(x+y) &= G(F(x+y)) \\ &= G\big(F(x) + DF(x)y + R(x,y)\big) \\ &= G(z) + DG(z)\big(DF(x)y + R(x,y)\big) + R_1(x,y) \\ &= G(z) + DG(z)DF(x)y + R_2(x,y) \end{aligned} \tag{A.2.6}$$

with

$$\frac{\|R_2(x,y)\|}{\|y\|} \to 0 \quad \text{as} \quad y \to 0. \tag{A.2.7}$$

Thus $G \circ F$ is differentiable at $x$, and

$$D(G \circ F)(x) = DG(F(x)) \cdot DF(x). \tag{A.2.8}$$

Another useful remark is that, by the fundamental theorem of calculus, applied to $\varphi(t) = F(x+ty)$,

$$F(x+y) = F(x) + \int_0^1 DF(x+ty)y\,dt, \tag{A.2.9}$$

provided $F$ is $C^1$. A closely related application of the fundamental theorem of calculus is that if we assume $F : \mathcal{O} \to \mathbb{R}^m$ is differentiable in each variable separately, and that each $\partial F/\partial x_j$ is continuous on $\mathcal{O}$, then

(A.2.10)

$$F(x+y) = F(x) + \sum_{j=1}^n \big[F(x+z_j) - F(x+z_{j-1})\big] = F(x) + \sum_{j=1}^n A_j(x,y)y_j,$$

$$A_j(x,y) = \int_0^1 \frac{\partial F}{\partial x_j}\big(x + z_{j-1} + ty_je_j\big)\,dt,$$

where $z_0 = 0$, $z_j = (y_1, \dots, y_j, 0, \dots, 0)$, and $\{e_j\}$ is the standard basis of $\mathbb{R}^n$. Now (A.2.10) implies $F$ is differentiable on $\mathcal{O}$, as we stated below (A.2.4). Thus we have established the following.

**Proposition A.2.1.** *If $\mathcal{O}$ is an open subset of $\mathbb{R}^n$ and $F : \mathcal{O} \to \mathbb{R}^m$ is of class $C^1$, then $F$ is differentiable at each point $x \in \mathcal{O}$.*

As is shown in many calculus texts, one can use the mean value theorem instead of the fundamental theorem of calculus, and obtain a slightly sharper result.

For the study of higher order derivatives of a function, the following result is fundamental.

**Proposition A.2.2.** *Assume $F : \mathcal{O} \to \mathbb{R}^m$ is of class $C^2$, with $\mathcal{O}$ open in $\mathbb{R}^n$. Then, for each $x \in \mathcal{O}$, $1 \le j, k \le n$,*

$$\frac{\partial}{\partial x_j}\frac{\partial F}{\partial x_k}(x) = \frac{\partial}{\partial x_k}\frac{\partial F}{\partial x_j}(x). \tag{A.2.11}$$

To prove Proposition A.2.2, it suffices to treat real valued functions, so consider $f : \mathcal{O} \to \mathbb{R}$. For $1 \le j \le n$, set

$$\Delta_{j,h} f(x) = \frac{1}{h}\big(f(x + he_j) - f(x)\big), \tag{A.2.12}$$

where $\{e_1, \dots, e_n\}$ is the standard basis of $\mathbb{R}^n$. The mean value theorem (for functions of $x_j$ alone) implies that if $\partial_j f = \partial f/\partial x_j$ exists on $\mathcal{O}$, then, for $x \in \mathcal{O}$, $h > 0$ sufficiently small,

$$\Delta_{j,h} f(x) = \partial_j f(x + \alpha_j h e_j), \tag{A.2.13}$$

for some $\alpha_j \in (0, 1)$, depending on $x$ and $h$. Iterating this, if $\partial_j(\partial_k f)$ exists on $\mathcal{O}$, then, for $x \in \mathcal{O}$ and $h > 0$ sufficiently small,

$$\begin{aligned} \Delta_{k,h}\Delta_{j,h} f(x) &= \partial_k(\Delta_{j,h} f)(x + \alpha_k h e_k) \\ &= \Delta_{j,h}(\partial_k f)(x + \alpha_k h e_k) \\ &= \partial_j \partial_k f(x + \alpha_k h e_k + \alpha_j h e_j), \end{aligned} \tag{A.2.14}$$

with $\alpha_j, \alpha_k \in (0, 1)$. Here we have used the elementary result

$$\partial_k \Delta_{j,h} f = \Delta_{j,h}(\partial_k f). \tag{A.2.15}$$

We deduce the following.

**Proposition A.2.3.** *If $\partial_k f$ and $\partial_j \partial_k f$ exist on $\mathcal{O}$ and $\partial_j \partial_k f$ is continuous at $x_0 \in \mathcal{O}$, then*

$$\partial_j \partial_k f(x_0) = \lim_{h \to 0} \Delta_{k,h}\Delta_{j,h} f(x_0). \tag{A.2.16}$$

Clearly

$$\Delta_{k,h}\Delta_{j,h} f = \Delta_{j,h}\Delta_{k,h} f, \tag{A.2.17}$$

so we have the following, which easily implies Proposition A.2.2.

**Corollary A.2.4.** *In the setting of Proposition* A.2.3*, if also $\partial_j f$ and $\partial_k \partial_j f$ exist on $\mathcal{O}$ and $\partial_k \partial_j f$ is continuous at $x_0$, then*

$$\partial_j \partial_k f(x_0) = \partial_k \partial_j f(x_0). \tag{A.2.18}$$

If $U$ and $V$ are open subsets of $\mathbb{R}^n$ and $F : U \to V$ is a $C^1$ map, we say $F$ is a diffeomorphism of $U$ onto $V$ provided $F$ maps $U$ one-to-one and onto $V$, and its inverse $G = F^{-1}$ is a $C^1$ map. If $F$ is a diffeomorphism, it follows from the chain rule that $DF(x)$ is invertible for each $x \in U$. We now present a partial converse of this, the Inverse Function Theorem, which is a fundamental result in multivariable calculus.

**Theorem A.2.5.** *Let $F$ be a $C^k$ map from an open neighborhood $\Omega$ of $p_0 \in \mathbb{R}^n$ to $\mathbb{R}^n$, with $q_0 = F(p_0)$. Assume $k \geq 1$. Suppose the derivative $DF(p_0)$ is invertible. Then there is a neighborhood $U$ of $p_0$ and a neighborhood $V$ of $q_0$ such that $F : U \to V$ is one-to-one and onto, and $F^{-1} : V \to U$ is a $C^k$ map. (So $F : U \to V$ is a diffeomorphism.)*

First we show that $F$ is one-to-one on a neighborhood of $p_0$, under these hypotheses. In fact, we establish the following result, of interest in its own right.

**Proposition A.2.6.** *Assume $\Omega \subset \mathbb{R}^n$ is open and convex, and let $f : \Omega \to \mathbb{R}^n$ be $C^1$. Assume that the symmetric part of $Df(u)$ is positive-definite, for each $u \in \Omega$. Then $f$ is one-to-one on $\Omega$.*

**Proof.** Take distinct points $u_1,\ u_2 \in \Omega$, and set $u_2 - u_1 = w$. Consider $\varphi : [0,1] \to \mathbb{R}$, given by

$$\varphi(t) = w \cdot f(u_1 + tw). \tag{A.2.19}$$

Then $\varphi'(t) = w \cdot Df(u_1 + tw)w > 0$ for $t \in [0,1]$, so $\varphi(0) \neq \varphi(1)$. But $\varphi(0) = w \cdot f(u_1)$ and $\varphi(1) = w \cdot f(u_2)$, so $f(u_1) \neq f(u_2)$. □

To continue the proof of Theorem A.2.5, let us set

$$f(u) = A\big(F(p_0 + u) - q_0\big), \quad A = DF(p_0)^{-1}. \tag{A.2.20}$$

Then $f(0) = 0$ and $Df(0) = I$, the identity matrix. We show that $f$ maps a neighborhood of 0 one-to-one and onto some neighborhood of 0. We can write

$$f(u) = u + R(u), \quad R(0) = 0, \quad DR(0) = 0, \tag{A.2.21}$$

and $R$ is $C^1$. Pick $b > 0$ such that

$$\|u\| \leq 2b \Longrightarrow \|DR(u)\| \leq \frac{1}{2}. \tag{A.2.22}$$

Then $Df = I + DR$ has a positive definite symmetric part on

$$B_{2b}(0) = \{u \in \mathbb{R}^n : \|u\| < 2b\}, \tag{A.2.23}$$

so by Proposition A.2.6,

$$f : B_{2b}(0) \longrightarrow \mathbb{R}^n \text{ is one-to-one.} \tag{A.2.24}$$

We will show that the range $f(B_{2b}(0))$ contains $B_b(0)$, that is to say, we can solve

$$f(u) = v, \tag{A.2.25}$$

given $v \in B_b(0)$, for some (unique) $u \in B_{2b}(0)$. This is equivalent to $u + R(u) = v$.

To get this solution, we set

$$T_v(u) = v - R(u). \tag{A.2.26}$$

Then solving (A.2.25) is equivalent to solving

$$T_v(u) = u. \tag{A.2.27}$$

We look for a fixed point

$$u = K(v) = f^{-1}(v). \tag{A.2.28}$$

Also we want to show that $DK(0) = I$, i.e., that

$$K(v) = v + r(v), \quad r(v) = o(\|v\|). \tag{A.2.29}$$

If we succeed in doing this, it follows that, for $y$ close to $q_0$, $G(y) = F^{-1}(y)$ is defined. Also, taking

$$x = p_0 + u, \quad y = F(x), \quad v = f(u) = A(y - q_0), \tag{A.2.30}$$

as in (A.2.20), we have, via (A.2.29),

$$\begin{aligned} G(y) &= p_0 + u = p_0 + K(v) \\ &= p_0 + K(A(y - q_0)) \\ &= p_0 + A(y - q_0) + o(\|y - q_0\|). \end{aligned} \tag{A.2.31}$$

Hence $G$ is differentiable at $q_0$ and

$$DG(q_0) = A = DF(p_0)^{-1}. \tag{A.2.32}$$

A parallel argument, with $p_0$ replaced by a nearby $x$ and $y = F(x)$, gives

$$DG(y) = \Big(DF(G(y))\Big)^{-1}. \tag{A.2.33}$$

A tool we will use to solve (A.2.27) is the Contraction Mapping Principle, established in Appendix A.1, which states that if $X$ is a complete metric space, and if $T : X \to X$ satisfies

$$\operatorname{dist}(Tx, Ty) \le r \operatorname{dist}(x, y), \tag{A.2.34}$$

for some $r < 1$ (we say $T$ is a contraction), then $T$ has a unique fixed point $x \in X$.

We prepare to solve (A.2.27). Having $b$ as in (A.2.22), we claim that

$$\|v\| \le b \Longrightarrow T_v : X_v \to X_v, \tag{A.2.35}$$

where

$$\begin{aligned} X_v &= \{u \in \overline{B_{2b}(0)} : \|u - v\| \le A_v\}, \\ A_v &= \sup_{\|w\| \le 2\|v\|} \|R(w)\|. \end{aligned} \tag{A.2.36}$$

Note from (A.2.21)–(A.2.22) that

$$\|w\| \le 2b \Longrightarrow \|R(w)\| \le \frac{1}{2}\|w\| \ \text{ and } \|R(w)\| = o(\|w\|). \tag{A.2.37}$$

Hence

$$\|v\| \le b \Longrightarrow A_v \le \|v\| \ \text{ and } A_v = o(\|v\|). \tag{A.2.38}$$

Thus $\|u - v\| \le A_v \Rightarrow u \in X_v$. Also,

$$\begin{aligned} u \in X_v &\Longrightarrow \|u\| \le 2\|v\| \\ &\Longrightarrow \|R(u)\| \le A_v \\ &\Longrightarrow \|T_v(u) - v\| \le A_v, \end{aligned} \tag{A.2.39}$$

so (A.2.35) holds.

As for the contraction property, given $u_j \in X_v,\ \|v\| \le b$,

$$\begin{aligned} \|T_v(u_1) - T_v(u_2)\| &= \|R(u_2) - R(u_1)\| \\ &\le \frac{1}{2}\|u_1 - u_2\|, \end{aligned} \tag{A.2.40}$$

the last inequality by (A.2.22), so the map (A.2.35) is a contraction. Hence there is a unique fixed point, $u = K(v) \in X_v$. Also, since $u \in X_v$,

$$\|K(v) - v\| \le A_v = o(\|v\|). \tag{A.2.41}$$

Thus we have (A.2.29). This establishes the existence of the inverse function $G = F^{-1} : V \to U$, and we have the formula (A.2.33) for the derivative $DG$. Since $G$ is differentiable on $V$, it is certainly continuous, so (A.2.33) implies $DG$ is continuous, given $F \in C^1(U)$. An inductive argument then shows that $G$ is $C^k$ if $F$ is $C^k$.

This completes the proof of Theorem A.2.5.

Thus if $DF$ is invertible on the domain of $F$, $F$ is a local diffeomorphism. Stronger hypotheses are needed to guarantee that $F$ is a global diffeomorphism onto its range. Proposition A.2.6 provides one tool for doing this. Here is a slight strengthening.

**Corollary A.2.7.** *Assume $\Omega \subset \mathbb{R}^n$ is open and convex, and that $F : \Omega \to \mathbb{R}^n$ is $C^1$. Assume there exist $n \times n$ matrices $A$ and $B$ such that the symmetric part of $A\,DF(u)\,B$ is positive definite for each $u \in \Omega$. Then $F$ maps $\Omega$ diffeomorphically onto its image, an open set in $\mathbb{R}^n$.*

**Proof.** Exercise. □

## A.3. The Laplace asymptotic method and Stirling's formula

Recall that the Gamma function is given by

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1}\, dt, \tag{A.3.1}$$

for $\operatorname{Re} z > 0$. We aim to analyze its behavior for large $z$, particularly in a sector

$$A_\beta = \{re^{i\theta} : r > 0,\ |\theta| \le \beta\}, \tag{A.3.2}$$

for $\beta < \pi/2$. Let us first take $z > 0$, and set $t = sz$, and then $s = e^y$, to write

$$\begin{aligned} \Gamma(z) &= z^z \int_0^\infty e^{-z(s-\log s)} s^{-1}\, ds \\ &= z^z e^{-z} \int_{-\infty}^\infty e^{-z(e^y - y - 1)}\, dy. \end{aligned} \tag{A.3.3}$$

Having done this, we see that each side of (A.3.3) is holomorphic in the half plane $\operatorname{Re} z > 0$, so the identity holds for all such $z$. The last integral has the form

$$I(z) = \int_{-\infty}^\infty e^{-z\varphi(y)} A(y)\, dy, \tag{A.3.4}$$

with $A(y) \equiv 1$ in this case, and $\varphi(y) = e^y - y - 1$. Note that $\varphi(y)$ is real valued and has a nondegenerate minimum at $y = 0$,

$$\varphi(0) = 0, \quad \varphi'(0) = 0, \quad \varphi''(0) > 0. \tag{A.3.5}$$

Furthermore,

$$\begin{aligned} \varphi(y) \ge ay^2 \quad &\text{for} \quad |y| \le 1, \\ a \quad &\text{for} \quad |y| \ge 1, \end{aligned} \tag{A.3.6}$$

for some $a > 0$.

The Laplace asymptotic method analyzes the asymptotic behavior of such an integral, as $z \to \infty$ in a sector $A_{\pi/2-\delta}$. In addition to the hypotheses (A.3.5)–(A.3.6), we assume that $\varphi$ and $A$ are smooth, and we assume that, given $\alpha > 0$, there exists $\beta > 0$ such that

$$\Bigl| \int\limits_{|y|>\alpha} e^{-z\varphi(y)} A(y)\, dy \Bigr| \le Ce^{-\beta \operatorname{Re} z}, \quad \text{for} \quad \operatorname{Re} z \ge 1. \tag{A.3.7}$$

These hypotheses are readily verified for the integral that arises in (A.3.3).

Given these hypotheses, our first step to tackle (A.3.4) is to pick $b \in C^\infty(\mathbb{R})$ such that $b(y) = 1$ for $|y| \le \alpha$ and $b(y) = 0$ for $|y| \ge 2\alpha$, and set

$$A_0(y) = b(y)A(y), \quad A_1(y) = (1 - b(y))A(y), \tag{A.3.8}$$

so

$$\Big|\int_{-\infty}^{\infty} e^{-z\varphi(y)} A_1(y)\,dy\Big| \le Ce^{-\beta\,\mathrm{Re}\,z}, \tag{A.3.9}$$

for $\mathrm{Re}\,z \ge 1$. It remains to analyze

$$I_0(z) = \int_{-\infty}^{\infty} e^{-z\varphi(y)} A_0(y)\,dy. \tag{A.3.10}$$

Pick $\alpha$ sufficiently small that you can write

$$\varphi(y) = \xi(y)^2, \quad \text{for} \quad |y| \le 2\alpha, \tag{A.3.11}$$

where $\xi$ maps $[-2\alpha, 2\alpha]$ diffeomorphically onto an interval about 0 in $\mathbb{R}$. Then

$$I_0(z) = \int_{-\infty}^{\infty} e^{-z\xi^2} B_0(\xi)\,d\xi, \tag{A.3.12}$$

with $B_0(\xi) = A_0(y(\xi))y'(\xi)$, where $y(\xi)$ denotes the map inverse to $\xi(y)$. Hence $B_0 \in C_0^\infty(\mathbb{R})$.

To analyze (A.3.12), we use the Fourier transform

$$\widehat{B}_0(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} B_0(\xi) e^{-ix\xi}\,d\xi, \tag{A.3.13}$$

studied in §3.2. Arguing as in the calculation (3.2.13)–(3.2.14), we have, for $\mathrm{Re}\,z > 0$,

$$E_z(\xi) = e^{-z\xi^2} \Longrightarrow \widehat{E}_z(x) = \Big(\frac{1}{2z}\Big)^{1/2} e^{-x^2/4z}. \tag{A.3.14}$$

Hence, by Plancherel's theorem, for $\mathrm{Re}\,z > 0$,

$$\begin{aligned} I_0(z) &= (2\zeta)^{1/2} \int_{-\infty}^{\infty} e^{-\zeta x^2} \widehat{B}_0(x)\,dx \\ &= (2\zeta)^{1/2} \mathcal{I}_0(\zeta), \qquad \zeta = \frac{1}{4z}. \end{aligned} \tag{A.3.15}$$

Now, given $B_0 \in C_0^\infty(\mathbb{R})$, one has $\widehat{B}_0 \in C^\infty(\mathbb{R})$, and

$$\big|x^j \widehat{B}_0^{(k)}(x)\big| \le C_{jk}, \quad \forall x \in \mathbb{R}. \tag{A.3.16}$$

We say $\widehat{B}_0 \in \mathcal{S}(\mathbb{R})$. Using this, it follows that

$$\mathcal{I}_0(\zeta) = \int_{-\infty}^{\infty} e^{-\zeta x^2} \widehat{B}_0(x)\,dx \tag{A.3.17}$$

is holomorphic on $\{\zeta \in \mathbb{C} : \mathrm{Re}\,\zeta > 0\}$ and $C^\infty$ on $\{\zeta \in \mathbb{C} : \mathrm{Re}\,\zeta \ge 0\}$. We have

$$\mathcal{I}_0(0) = \int_{-\infty}^{\infty} \widehat{B}_0(x)\,dx = \sqrt{2\pi} B_0(0). \tag{A.3.18}$$

It follows that, for $\operatorname{Re}\zeta \geq 0$,

$$\mathcal{I}_0(\zeta) = \sqrt{2\pi} B_0(0) + O(|\zeta|), \quad \text{as } \zeta \to 0, \tag{A.3.19}$$

and hence, for $\operatorname{Re} z \geq 0$, $z \neq 0$,

$$I_0(z) = \Big(\frac{\pi}{z}\Big)^{1/2} B_0(0) + O(|z|^{-3/2}), \quad \text{as } z \to \infty. \tag{A.3.20}$$

If we apply this to (A.3.3)–(A.3.9), we obtain Stirling's formula,

$$\Gamma(z) = z^z e^{-z} \Big(\frac{2\pi}{z}\Big)^{1/2} \Big[1 + O(|z|^{-1})\Big], \tag{A.3.21}$$

for $z \in A_{\pi/2-\delta}$, taking into account that in this case $B_0(0) = \sqrt{2}$.

Asymptotic analysis of the Hankel function, done in §7.1, leads to an integral of the form (A.3.4) with

$$\varphi(y) = \frac{\sinh^2 y}{\cosh y}, \tag{A.3.22}$$

and

$$A(y) = e^{-\nu u(y)} u'(y), \quad u(y) = y + i \tan^{-1}(\sinh y), \tag{A.3.23}$$

so $u'(y) = 1 + i/\cosh y$. The conditions for applicability of (A.3.5)–(A.3.9) are readily verified for this case, yielding the asymptotic expansion in Proposition 7.1.3.

Returning to Stirling's formula, we mention another approach which gives more precise information. It involves the ingenious identity

$$\log \Gamma(z) = \Big(z - \frac{1}{2}\Big) \log z - z + \frac{1}{2} \log 2\pi + \omega(z), \tag{A.3.24}$$

with a convenient integral formula for $\omega(z)$, namely

$$\omega(z) = \int_0^\infty f(t) e^{-tz}\, dt, \tag{A.3.25}$$

for $\operatorname{Re} z > 0$, with

$$\begin{aligned} f(t) &= \Big(\frac{1}{2} - \frac{1}{t} + \frac{1}{e^t - 1}\Big)\frac{1}{t} \\ &= \frac{1}{t}\Big(\frac{1}{2}\,\frac{\cosh t/2}{\sinh t/2} - \frac{1}{t}\Big), \end{aligned} \tag{A.3.26}$$

which is a smooth, even function of $t$ on $\mathbb{R}$, asymptotic to $1/2t$ as $t \nearrow +\infty$. A proof of (A.3.24) can be found in §1.4 of [**27**].

We show how to derive a complete asymptotic expansion of the Laplace transform (A.3.25), valid for $z \to \infty$, $\operatorname{Re} z \geq 0$, just given that $f \in$

$C^\infty([0,\infty))$ and that $f^{(j)}$ is integrable on $[0,\infty)$ for each $j \ge 1$. To start, integration by parts yields

$$\begin{aligned} \int_0^\infty f(t)e^{-zt}\,dt &= -\frac{1}{z}\int_0^\infty f(t)\,\frac{d}{dt}e^{-zt}\,dt \\ &= \frac{1}{z}\int_0^\infty f'(t)e^{-zt}\,dt + \frac{f(0)}{z}, \end{aligned} \tag{A.3.27}$$

valid for $\operatorname{Re} z > 0$. We can iterate this argument to obtain

$$\omega(z) = \sum_{k=1}^{N} \frac{f^{(k-1)}(0)}{z^k} + \frac{1}{z^N}\int_0^\infty f^{(N)}(t)e^{-zt}\,dt \tag{A.3.28}$$

and

$$\Bigl|\int_0^\infty f^{(N)}(t)e^{-zt}\,dt\Bigr| \le \int_0^\infty |f^{(N)}(t)|\,dt < \infty, \quad \text{for } \ N \ge 1,\ \operatorname{Re} z \ge 0. \tag{A.3.29}$$

By (A.3.24), $\omega(z)$ is holomorphic on $\mathbb{C} \setminus (-\infty, 0]$. Meanwhile, the right side of (A.3.28) is continuous on $\{z \in \mathbb{C} : \operatorname{Re} z \ge 0,\ z \ne 0\}$, so equality in (A.3.28) holds on this region.

To carry on, we note that, for $|t| < 2\pi$,

$$\frac{1}{2} - \frac{1}{t} + \frac{1}{e^t - 1} = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{(2k)!} B_k t^{2k-1}, \tag{A.3.30}$$

where $B_k$ are the *Bernoulli numbers*, introduced in §2.8, Exercises 6–8, and related to $\zeta(2k)$ in §6.1. Hence, for $|t| < 2\pi$,

$$f(t) = \sum_{\ell=0}^{\infty} \frac{(-1)^\ell}{(2\ell+2)!} B_{\ell+1} t^{2\ell}. \tag{A.3.31}$$

Thus

$$\begin{aligned} f^{(j)}(0) = \quad & 0 & & j \text{ odd}, \\ & \frac{(-1)^\ell B_{\ell+1}}{(2\ell+1)(2\ell+2)} & & j = 2\ell, \end{aligned} \tag{A.3.32}$$

so

$$\omega(z) \sim \sum_{\ell \ge 0} \frac{(-1)^\ell B_{\ell+1}}{(2\ell+1)(2\ell+2)}\,\frac{1}{z^{2\ell+1}}, \quad z \to \infty,\ \operatorname{Re} z \ge 0. \tag{A.3.33}$$

Thus there are $A_k \in \mathbb{R}$ such that

$$e^{\omega(z)} \sim 1 + \sum_{k \ge 1} \frac{A_k}{z^k}, \quad z \to \infty,\ \operatorname{Re} z \ge 0. \tag{A.3.34}$$

This yields the following refinement of (A.3.21):

$$\Gamma(z) \sim z^z e^{-z}\Big(\frac{2\pi}{z}\Big)^{1/2}\Big[1+\sum_{k\ge 1} A_k z^{-k}\Big], \quad |z|\to\infty,\ \operatorname{Re} z\ge 0. \tag{A.3.35}$$

We can push the asymptotic analysis of $\Gamma(z)$ into the left half-plane, using the identity

$$\Gamma(-z)\sin \pi z = -\frac{\pi}{z\Gamma(z)} \tag{A.3.36}$$

to extend (A.3.24), i.e.,

$$\Gamma(z) = \Big(\frac{z}{e}\Big)^z \sqrt{\frac{2\pi}{z}}\, e^{\omega(z)}, \quad \text{for} \quad \operatorname{Re} z \ge 0,\ z\neq 0, \tag{A.3.37}$$

to the rest of $\mathbb{C}\setminus\mathbb{R}^-$. If we define $z^z$ and $\sqrt{z}$ in the standard fashion for $z\in(0,\infty)$ and to be holomorphic on $\mathbb{C}\setminus\mathbb{R}^-$, we get

$$\Gamma(z) = \frac{1}{1-e^{2\pi i z}}\Big(\frac{z}{e}\Big)^z \sqrt{\frac{2\pi}{z}}\, e^{-\omega(-z)}, \quad \text{for} \quad \operatorname{Re} z\le 0,\ \operatorname{Im} z>0, \tag{A.3.38}$$

and

$$\Gamma(z) = \frac{1}{1-e^{-2\pi i z}}\Big(\frac{z}{e}\Big)^z \sqrt{\frac{2\pi}{z}}\, e^{-\omega(-z)}, \quad \text{for} \quad \operatorname{Re} z\le 0,\ \operatorname{Im} z<0. \tag{A.3.39}$$

Comparing (A.3.37) and (A.3.38) for $z=iy,\ y>0$, we see that

$$e^{-\omega(-iy)} = (1-e^{-2\pi y})e^{\omega(iy)}, \quad y>0. \tag{A.3.40}$$

That $e^{-\omega(-iy)}$ and $e^{\omega(iy)}$ have the same asymptotic behavior as $y\to+\infty$ also follows from the fact that only odd powers of $z^{-1}$ appear in (A.3.33).

## A.4. The Stieltjes integral

Here we develop basic results on integrals of the form

$$\int_a^b f(x)\, du(x), \tag{A.4.1}$$

known as Stieltjes integrals. We assume that $f\in C([a,b])$ and that

$$u:[a,b]\longrightarrow\mathbb{R} \ \text{ is increasing}, \tag{A.4.2}$$

i.e., $x_1<x_2 \Rightarrow u(x_1)\le u(x_2)$. We also assume $u$ is right continuous, i.e.,

$$u(x) = \lim_{y\searrow x} u(y), \quad \forall\, x\in[a,b). \tag{A.4.3}$$

Note that (A.4.2) implies the existence for all $x\in[a,b]$ of

$$u^+(x) = \liminf_{y\searrow x} u(y), \quad u^-(x) = \limsup_{y\nearrow x} u(y) \tag{A.4.4}$$

(with the convention that $u^-(a) = u(a)$ and $u^+(b) = u(b)$). We have

$$u^-(x) \le u^+(x), \quad \forall\, x \in [a,b], \tag{A.4.5}$$

and (A.4.3) says $u(x) = u^+(x)$ for all $x$. Note that $u$ is continuous at $x$ if and only if $u^-(x) = u^+(x)$. If $u$ is not continuous at $x$, it has a jump discontinuity there, and it is easy to see that $u$ can have at most countably many such discontinuities.

We prepare to define the integral (A.4.1), mirroring a standard development of the Riemann integral when $u(x) = x$. For now, we allow $f$ to be any bounded, real-valued function on $[a,b]$, say $|f(x)| \le M$. To start, we partition $[a,b]$ into smaller intervals. A *partition* $\mathcal{P}$ of $[a,b]$ is a finite collection of subintervals $\{J_k : 0 \le k \le N-1\}$, disjoint except for their endpoints, whose union is $[a,b]$. We order the $J_k$ so that $J_k = [x_k, x_{k+1}]$, where

$$a = x_0 < x_1 < \cdots < x_N = b. \tag{A.4.6}$$

We call the points $x_k$ the endpoints of $\mathcal{P}$. We set

$$\begin{aligned} \overline{I}_{\mathcal{P}}(f\,du) &= \sum_{k=0}^{N-1} (\sup_{J_k} f)[u(x_{k+1}) - u(x_k)], \\ \underline{I}_{\mathcal{P}}(f\,du) &= \sum_{k=0}^{N-1} (\inf_{J_k} f)[u(x_{k+1}) - u(x_k)]. \end{aligned} \tag{A.4.7}$$

Note that $\underline{I}_{\mathcal{P}}(f\,du) \le \overline{I}_{\mathcal{P}}(f\,du)$. These quantities should be approximations to (A.4.1) if the partition $\mathcal{P}$ is sufficiently "fine."

To be more precise, if $\mathcal{P}$ and $\mathcal{Q}$ are two partitions of $[a,b]$, we say $\mathcal{P}$ *refines* $\mathcal{Q}$, and write $\mathcal{P} \succ \mathcal{Q}$, if $\mathcal{P}$ is formed by partitioning the intervals in $\mathcal{Q}$. Equivalently, $\mathcal{P} \succ \mathcal{Q}$ if and only if all the endpoints of $\mathcal{Q}$ are endpoints of $\mathcal{P}$. It is easy to see that any two partitions have a common refinement; just take the union of their endpoint, to form a new partition. Note that

$$\begin{aligned} \mathcal{P} \succ \mathcal{Q} \Rightarrow \overline{I}_{\mathcal{P}}(f\,du) &\le \overline{I}_{\mathcal{Q}}(f\,du) \ \text{ and} \\ \underline{I}_{\mathcal{P}}(f\,du) &\ge \underline{I}_{\mathcal{Q}}(f\,du). \end{aligned} \tag{A.4.8}$$

Consequently, if $\mathcal{P}_j$ are two partitions of $[a,b]$ and $\mathcal{Q}$ is a common refinement, we have

$$\underline{I}_{\mathcal{P}_1}(f\,du) \le \underline{I}_{\mathcal{Q}}(f\,du) \le \overline{I}_{\mathcal{Q}}(f\,du) \le \overline{I}_{\mathcal{P}_2}(f\,du). \tag{A.4.9}$$

Thus, whenever $f : [a, b] \to \mathbb{R}$ is bounded, the following quantities are well defined:

$$\begin{aligned} \overline{I}_a^b(f\,du) &= \inf_{\mathcal{P}\in\Pi[a,b]} \overline{I}_{\mathcal{P}}(f\,du), \\ \underline{I}_a^b(f\,du) &= \sup_{\mathcal{P}\in\Pi[a,b]} \underline{I}_{\mathcal{P}}(f\,du), \end{aligned} \tag{A.4.10}$$

where $\Pi[a, b]$ denotes the set of all partitions of $[a, b]$. Clearly, by (A.4.9),

$$\underline{I}_a^b(f\,du) \le \overline{I}_a^b(f\,du). \tag{A.4.11}$$

We say a bounded function $f : [a, b] \to \mathbb{R}$ is Riemann-Stieltjes integrable provided there is equality in (A.4.11). In such a case, we set

$$\int_a^b f(x)\,du(x) = \overline{I}_a^b(f\,du) = \underline{I}_a^b(f\,du), \tag{A.4.12}$$

and we write $f \in \mathcal{R}([a, b], du)$. Though we will not emphasize it, another notation for (A.4.12) is

$$\int_I f(x)\,du(x), \quad I = (a, b]. \tag{A.4.13}$$

Our first basic result is that each continuous function on $[a, b]$ is Riemann-Stieltjes integrable.

**Proposition A.4.1.** *If $f : [a, b] \to \mathbb{R}$ is continuous, then $f \in \mathcal{R}([a, b], du)$.*

**Proof.** Any continuous function on $[a, b]$ is uniformly continuous (cf. Appendix A.1). Thus there is a function $\omega(\delta)$ such that

$$|x - y| \le \delta \Rightarrow |f(x) - f(y)| \le \omega(\delta), \quad \omega(\delta) \to 0 \text{ as } \delta \to 0. \tag{A.4.14}$$

Given $J_k = [x_k, x_{k+1}]$, let us set $\ell(J_k) = x_{k+1} - x_k$, and, for the partition $\mathcal{P}$ with endpoints as in (A.4.6), set

$$\operatorname{maxsize}(\mathcal{P}) = \max_{0\le k\le N-1} \ell(J_k). \tag{A.4.15}$$

Then

$$\operatorname{maxsize}(\mathcal{P}) \le \delta \Longrightarrow \overline{I}_{\mathcal{P}}(f\,du) - \underline{I}_{\mathcal{P}}(f\,du) \le \omega(\delta)[u(b) - u(a)], \tag{A.4.16}$$

which yields the proposition. □

We will concentrate on (A.4.1) for continuous $f$, but there are a couple of results that are conveniently established for more general integrable $f$.

**Proposition A.4.2.** *If $f, g \in \mathcal{R}([a, b], du)$, then $f + g \in \mathcal{R}([a, b], du)$, and*

$$\int_a^b (f(x) + g(x))\,du(x) = \int_a^b f(x)\,du(x) + \int_a^b g(x)\,du(x). \tag{A.4.17}$$

**Proof.** If $J_k$ is any subinterval of $[a,b]$, then

$$\begin{aligned}\sup_{J_k}(f+g) &\le \sup_{J_k} f + \sup_{J_k} g \quad \text{and} \\ \inf_{J_k}(f+g) &\ge \inf_{J_k} f + \inf_{J_k} g,\end{aligned} \tag{A.4.18}$$

so, for any partition $\mathcal{P}$, we have $\overline{I}_{\mathcal{P}}(f+g)\,du) \le \overline{I}_{\mathcal{P}}(f\,du) + \overline{I}_{\mathcal{P}}(g\,du)$. Also, using a common refinement of partitions, we can *simultaneously* approximate $\overline{I}_a^b(f\,du)$ and $\overline{I}_a^b(g\,du)$ by $\overline{I}_{\mathcal{P}}(f\,du)$ and $\overline{I}_{\mathcal{P}}(g\,du)$, and likewise for $\overline{I}_a^b((f+g)\,du)$. Then the characterization (A.4.10) implies $\overline{I}_a^b((f+g)\,du) \le \overline{I}_a^b(f\,du) + \overline{I}_a^b(g\,du)$. A parallel argument implies $\underline{I}_a^b((f+g)\,du) \ge \underline{I}_a^b(f\,du) + \underline{I}_a^b(g\,du)$, and the proposition follows. □

Here is another useful additivity result.

**Proposition A.4.3.** *Let $a < b < c$, $f : [a,c] \to \mathbb{R}$, $f_1 = f|_{[a,b]}$, $f_2 = f|_{[b,c]}$. Assume $u : [a,c] \to \mathbb{R}$ is increasing and right continuous. Then*

$$f \in \mathcal{R}([a,c],du) \Leftrightarrow f_1 \in \mathcal{R}([a,b],du) \ \text{and}\ f_2 \in \mathcal{R}([b,c],du), \tag{A.4.19}$$

*and, if this holds,*

$$\int_a^c f(x)\,du(x) = \int_a^b f_1(x)\,du(x) + \int_b^c f_2(x)\,du(x). \tag{A.4.20}$$

**Proof.** Since any partition of $[a,c]$ has a refinement for which $b$ is an endpoint, we may as well consider a partition $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$, where $\mathcal{P}_1$ is a partition of $[a,b]$ and $\mathcal{P}_2$ is a partition of $[b,c]$. Then

$$\overline{I}_{\mathcal{P}}(f\,du) = \overline{I}_{\mathcal{P}_1}(f_1\,du) + \overline{I}_{\mathcal{P}_2}(f_2\,du), \tag{A.4.21}$$

with a parallel identity for $\underline{I}_{\mathcal{P}}(f\,du)$, so

$$\overline{I}_{\mathcal{P}}(f\,du) - \underline{I}_{\mathcal{P}}(f\,du) = \{\overline{I}_{\mathcal{P}_1}(f_1\,du) - \underline{I}_{\mathcal{P}_1}(f_1\,du)\} + \{\overline{I}_{\mathcal{P}_2}(f_2\,du) - \underline{I}_{\mathcal{P}_2}(f_2\,du)\}. \tag{A.4.22}$$

Since both terms in braces in (A.4.22) are $\ge 0$, we have the equivalence in (A.4.19). Then (A.4.20) follows from (A.4.21) upon taking sufficiently fine partitions. □

In the classical case $u(x) = x$, we denote $\mathcal{R}([a,b],du)$ by $\mathcal{R}([a,b])$, the space of Riemann integrable functions on $[a,b]$. We record a few standard results about the Riemann integral, whose proofs can be found in many texts, including [**44**], Chapter 4, and [**45**], §0.

**Proposition A.4.4.** *If $u : [a,b] \to \mathbb{R}$ is increasing and right continuous, then $u \in \mathcal{R}([a,b])$.*

**Proposition A.4.5.** *If $f, g \in \mathcal{R}([a,b])$, then $fg \in \mathcal{R}([a,b])$.*

The next result is known as the Darboux theorem for the Riemann integral.

**Proposition A.4.6.** *Let $\mathcal{P}_\nu$ be a sequence of partitions of $[a,b]$, into $\nu$ intervals $J_{\nu k}$, $1 \le k \le \nu$, such that*

$$\text{maxsize}(\mathcal{P}_\nu) \longrightarrow 0, \tag{A.4.23}$$

*and assume $f \in \mathcal{R}([a,b])$. Then*

$$\int_a^b f(x)\,dx = \lim_{\nu\to\infty} \sum_{k=1}^{\nu} f(\xi_{\nu k})\ell(J_{\nu k}), \tag{A.4.24}$$

*for arbitrary $\xi_{\nu k} \in J_{\nu k}$, where $\ell(J_{\nu k})$ is the length of the interval $J_{\nu k}$.*

We now present a very useful result, known as integration by parts for the Stieltjes integral.

**Proposition A.4.7.** *Let $u : [a,b] \to \mathbb{R}$ be increasing and right continuous, and let $f \in C^1([a,b])$, so $f' \in C([a,b])$. Then*

$$\int_a^b f(x)\,du(x) = fu\Big|_a^b - \int_a^b f'(x)u(x)\,dx, \tag{A.4.25}$$

*where*

$$fu\Big|_a^b = f(b)u(b) - f(a)u(a). \tag{A.4.26}$$

**Proof.** Pick a partition $\mathcal{P}$ of $[a,b]$ with endpoints $x_k$, $0 \le k \le N$, as in (A.4.6). Then

$$\int_a^b f(x)\,du(x) = \sum_{k=0}^{N-1} \int_{x_k}^{x_{k+1}} f(x)\,du(x). \tag{A.4.27}$$

Now, given $\varepsilon > 0$, pick $\delta > 0$ such that

$$\text{maxsize}(\mathcal{P}) \le \delta \Longrightarrow \sup_{\xi\in[x_k,x_{k+1}]} |f(\xi) - f(x_k)| \le \varepsilon. \tag{A.4.28}$$

Then

$$\int_a^b f(x)\,du(x) = \sum_{k=0}^{N-1} f(x_k)[u(x_{k+1}) - u(x_k)] + O(\varepsilon). \tag{A.4.29}$$

We can write this last sum as

$$\begin{aligned} &-f(x_0)u(x_0) + [f(x_0) - f(x_1)]u(x_1) + \cdots \\ &\qquad + [f(x_{N-1}) - f(x_N)]u(x_N) + f(x_N)u(x_N), \end{aligned} \tag{A.4.30}$$

so

$$\int_a^b f(x)\,du(x) = fu\Big|_a^b + \sum_{k=0}^{N-1} [f(x_k) - f(x_{k+1})]u(x_k) + O(\varepsilon). \tag{A.4.31}$$

Now the mean value theorem implies

$$f(x_k) - f(x_{k+1}) = -f'(\zeta_k)(x_{k+1} - x_k), \tag{A.4.32}$$

for some $\zeta_k \in (x_k, x_{k+1})$. Since $f' \in C([a,b])$, we have in addition to (A.4.28) that, after perhaps shrinking $\delta$,

$$\mathrm{maxsize}(\mathcal{P}) \le \delta \Rightarrow \sup_{\zeta \in [x_k, x_{k+1}]} |f'(\zeta) - f'(x_k)| \le \varepsilon. \tag{A.4.33}$$

Hence

$$\int_a^b f(x)\,du(x) = fu\Big|_a^b - \sum_{k=0}^{N-1} f'(x_k)u(x_k)(x_{k+1} - x_k) + O(\varepsilon). \tag{A.4.34}$$

Now Propositions A.4.4–A.4.5 imply $f'u \in \mathcal{R}([a,b])$, and then Proposition A.4.6, applied to $f'u$, implies that, in the limit as $\mathrm{maxsize}(\mathcal{P}) \to 0$, the sum on the right side of (A.4.34) tends to

$$\int_a^b f'(x)u(x)\,dx. \tag{A.4.35}$$

This proves (A.4.25). □

We discuss some natural extensions of the integral (A.4.1). For one, we can take $w = u - v$, where $v : [a,b] \to \mathbb{R}$ is also increasing and right continuous, and set

$$\int_a^b f(x)\,dw(x) = \int_a^b f(x)\,du(x) - \int_a^b f(x)\,dv(x). \tag{A.4.36}$$

Let us take $f \in C([a,b])$. To see that (A.4.36) is well defined, suppose that also $w = u_1 - v_1$, where $u_1$ and $v_1$ are also increasing and right continuous. The identity of the right side of (A.4.36) with

$$\int_a^b f(x)\,du_1(x) - \int_a^b f(x)\,dv_1(x) \tag{A.4.37}$$

is equivalent to the identity

$$\int_a^b f(x)\,du(x) + \int_a^b f(x)\,dv_1(x) = \int_a^b f(x)\,du_1(x) + \int_a^b f(x)\,dv(x), \tag{A.4.38}$$

and hence to

$$\int_a^b f(x)\,du(x) + \int_a^b f(x)\,dv_1(x) = \int_a^b f(x)\,d(u + v_1)(x), \tag{A.4.39}$$

which is readily established, via

$$\overline{I}_{\mathcal{P}}(f\,du) + \overline{I}_{\mathcal{P}}(f\,dv_1) = \overline{I}_{\mathcal{P}}(f\,d(u + v_1)) \tag{A.4.40}$$

and similar identities.

Another extension is to take $u : [0, \infty) \to \mathbb{R}$, increasing and right continuous, and define

$$\int_0^\infty f(x)\, du(x), \tag{A.4.41}$$

for a class of functions $f : [0, \infty) \to \mathbb{R}$ satisfying appropriate bounds at infinity. For example, we might take

$$\begin{aligned} u(x) &\le C_\varepsilon e^{\varepsilon x}, \quad \forall\, \varepsilon > 0, \\ |f(x)| &\le C e^{-ax}, \quad \text{for some } a > 0. \end{aligned} \tag{A.4.42}$$

There are many variants. One then sets

$$\int_0^\infty f(x)\, du(x) = \lim_{R\to\infty} \int_0^R f(x)\, du(x). \tag{A.4.43}$$

Extending the integration by parts formula (A.4.25), we have

$$\begin{aligned} \int_0^\infty f(x)\, du(x) &= \lim_{R\to\infty} fu\Big|_0^R - \int_0^R f'(x)u(x)\, dx \\ &= -f(0)u(0) - \int_0^\infty f'(x)u(x)\, dx, \end{aligned} \tag{A.4.44}$$

for $f \in C^1([0, \infty))$, under an appropriate additional condition on $f'(x)$, such as

$$|f'(x)| \le C e^{-ax}, \tag{A.4.45}$$

when (A.4.42) holds.

In addition, one can also have $v : [0, \infty) \to \mathbb{R}$, increasing and right continuous, set $w = u - v$, and define $\int_0^\infty f(x)\, dw(x)$, in a fashion parallel to (A.4.36). If, for example, (A.4.42) also holds with $u$ replaced by $v$, we can extend (A.4.44) to

$$\int_0^\infty f(x)\, dw(x) = -f(0)w(0) - \int_0^\infty f'(x)w(x)\, dx. \tag{A.4.46}$$

The material developed above is adequate for use in §4.4 and Appendix A.5, but we mention that further extension can be made to the Lebesgue-Stieltjes integral. In this set-up, one associates a "measure" $\mu$ on $[a, b]$ to the function $u$, and places the integral (A.4.1) within the framework of the Lebesgue integral with respect to a measure. Material on this can be found in many texts on measure theory, such as [**47**], Chapters 5 and 13. In this setting, the content of Proposition A.4.7 is that the measure $\mu$ is the "weak derivative" of $u$, and one can extend the identity (A.4.25) to a class of functions $f$ much more general than $f \in C^1([a, b])$.

## A.5. Abelian theorems and Tauberian theorems

Abelian theorems and Tauberian theorems are results to the effect that one sort of convergence leads to another. We start with the original Abelian theorem, due to Abel, and give some applications of that result, before moving on to other Abelian theorems and to Tauberian theorems.

**Proposition A.5.1.** *Assume we have a convergent series*

$$\sum_{k=0}^{\infty} a_k = A. \tag{A.5.1}$$

*Then*

$$f(r) = \sum_{k=0}^{\infty} a_k r^k \tag{A.5.2}$$

*converges uniformly on* $[0,1]$*, so* $f \in C([0,1])$*. In particular,* $f(r) \to A$ *as* $r \nearrow 1$*.*

As a warm up, we look at the following somewhat simpler result. Compare Propositions 3.1.2 and 2.11.4.

**Proposition A.5.2.** *Assume we have an absolutely convergent series*

$$\sum_{k=0}^{\infty} |a_k| < \infty. \tag{A.5.3}$$

*Then the series* (A.5.2) *converges uniformly on* $[-1,1]$*, so* $f \in C([-1,1])$*.*

**Proof.** Clearly

$$\Bigl|\sum_{k=m}^{m+n} a_k r^k\Bigr| \le \sum_{k=m}^{m+n} |a_k|, \tag{A.5.4}$$

for $r \in [-1,1]$, so if (A.5.3) holds, then (A.5.2) converges uniformly for $r \in [-1,1]$. Of course, a uniform limit of a sequence of continuous functions on $[-1,1]$ is also continuous on this set. □

Proposition A.5.1 is much more subtle than Proposition A.5.2. One ingredient in the proof is the following *summation by parts* formula.

**Proposition A.5.3.** *Let* $(a_j)$ *and* $(b_j)$ *be sequences, and let*

$$s_n = \sum_{j=0}^{n} a_j. \tag{A.5.5}$$

*If* $m > n$, *then*

$$\sum_{k=n+1}^{m} a_k b_k = (s_m b_m - s_n b_{n+1}) + \sum_{k=n+1}^{m-1} s_k (b_k - b_{k+1}). \tag{A.5.6}$$

**Proof.** Write the left side of (A.5.6) as

$$\sum_{k=n+1}^{m} (s_k - s_{k-1}) b_k. \tag{A.5.7}$$

It is then straightforward to obtain the right side. □

Before applying Proposition A.5.3 to the proof of Proposition A.5.1, we note that, by Proposition 1.1.4 and its proof, especially (1.1.40), the power series (A.5.2) converges uniformly on compact subsets of $(-1, 1)$ and defines $f \in C((-1,1))$. Our task here is to get uniform convergence up to $r = 1$.

To proceed, we apply (A.5.6) with $b_k = r^k$ and $n + 1 = 0, s_{-1} = 0$, to get

$$\sum_{k=0}^{m} a_k r^k = (1 - r) \sum_{k=0}^{m-1} s_k r^k + s_m r^m. \tag{A.5.8}$$

Now, we want to add and subtract a function $g_m(r)$, defined for $0 \le r < 1$ by

$$\begin{aligned} g_m(r) &= (1 - r) \sum_{k=m}^{\infty} s_k r^k \\ &= A r^m + (1 - r) \sum_{k=m}^{\infty} \sigma_k r^k, \end{aligned} \tag{A.5.9}$$

with $A$ as in (A.5.1) and

$$\sigma_k = s_k - A \longrightarrow 0, \quad \text{as} \quad k \to \infty. \tag{A.5.10}$$

Note that, for $0 \le r < 1$, $\mu \in \mathbb{N}$,

$$\begin{aligned} (1 - r) \Big| \sum_{k=\mu}^{\infty} \sigma_k r^k \Big| &\le \Big( \sup_{k \ge \mu} |\sigma_k| \Big) (1 - r) \sum_{k=\mu}^{\infty} r^k \\ &= \Big( \sup_{k \ge \mu} |\sigma_k| \Big) r^{\mu}. \end{aligned} \tag{A.5.11}$$

It follows that

$$g_m(r) = A r^m + h_m(r) \tag{A.5.12}$$

extends to be continuous on $[0, 1]$ and

$$|h_m(r)| \le \sup_{k \ge m} |\sigma_k|, \quad h_m(1) = 0. \tag{A.5.13}$$

Now adding and subtracting $g_m(r)$ in (A.5.8) gives

$$\sum_{k=0}^{m} a_k r^k = g_0(r) + (s_m - A)r^m - h_m(r), \tag{A.5.14}$$

and this converges uniformly for $r \in [0,1]$ to $g_0(r)$. We have Proposition A.5.1, with $f(r) = g_0(r)$.

Here is one illustration of Proposition A.5.1. Let $a_k = (-1)^{k-1}/k$, which produces a convergent series by the alternating series test (Section 1.1, Exercise 8). By (1.5.37),

$$\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} r^k = \log(1+r), \tag{A.5.15}$$

for $|r| < 1$. It follows from Proposition A.5.1 that this infinite series converges uniformly on $[0,1]$, and hence

$$\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} = \log 2. \tag{A.5.16}$$

See Exercise 2 in §1.5 for a more direct approach to (A.5.16), using the special behavior of alternating series. Here is a more subtle generalization, which we will establish below.

**Claim.** *For all* $\theta \in (0, 2\pi)$, *the series*

$$\sum_{k=1}^{\infty} \frac{e^{ik\theta}}{k} = S(\theta) \tag{A.5.17}$$

*converges.*

Given this claim, it follows from Proposition A.5.1 that

$$\lim_{r \nearrow 1} \sum_{k=1}^{\infty} \frac{e^{ik\theta}}{k} r^k = S(\theta), \quad \forall\, \theta \in (0, 2\pi). \tag{A.5.18}$$

Note that taking $\theta = \pi$ gives (A.5.16). We recall from §1.5 that the function $\log : (0, \infty) \to \mathbb{R}$ has a natural extension to

$$\log : \mathbb{C} \setminus (-\infty, 0] \longrightarrow \mathbb{C} \tag{A.5.19}$$

and

$$\sum_{k=1}^{\infty} \frac{1}{k} z^k = -\log(1-z), \quad \text{for } \ |z| < 1, \tag{A.5.20}$$

from which we deduce, via Proposition A.5.1, that $S(\theta)$ in (A.5.17) satisfies

$$S(\theta) = -\log(1 - e^{i\theta}), \quad 0 < \theta < 2\pi. \tag{A.5.21}$$

We want to establish the convergence of (A.5.17) for $\theta \in (0, 2\pi)$. In fact, we prove the following more general result.

**Proposition A.5.4.** *If $b_k \searrow 0$, then*

$$\sum_{k=1}^{\infty} b_k e^{ik\theta} = F(\theta) \tag{A.5.22}$$

*converges for all $\theta \in (0, 2\pi)$.*

Given Proposition A.5.4, it then follows from Proposition A.5.1 that

$$\lim_{r \nearrow 1} \sum_{k=1}^{\infty} b_k r^k e^{ik\theta} = F(\theta), \quad \forall\, \theta \in (0, 2\pi). \tag{A.5.23}$$

In turn, Proposition A.5.4 is a special case of the following more general result, known as the *Dirichlet test* for convergence of an infinite series.

**Proposition A.5.5.** *If $b_k \searrow 0,\ a_k \in \mathbb{C}$, and there exists $B < \infty$ such that*

$$s_k = \sum_{j=1}^{k} a_j \Longrightarrow |s_k| \le B, \quad \forall\, k \in \mathbb{N}, \tag{A.5.24}$$

*then*

$$\sum_{k=1}^{\infty} a_k b_k \quad \textit{converges.} \tag{A.5.25}$$

To apply Proposition A.5.5 to Proposition A.5.4, take $a_k = e^{ik\theta}$ and observe that

$$\sum_{j=1}^{k} e^{ij\theta} = \frac{1 - e^{ik\theta}}{1 - e^{i\theta}} e^{i\theta}, \tag{A.5.26}$$

which is uniformly bounded (in $k$) for each $\theta \in (0, 2\pi)$.

To prove Proposition A.5.5, we use summation by parts, Proposition A.5.3. We have, via (A.5.6) with $n = 0,\ s_0 = 0$,

$$\sum_{k=1}^{m} a_k b_k = s_m b_m + \sum_{k=1}^{m-1} s_k (b_k - b_{k+1}). \tag{A.5.27}$$

Now, if $|s_k| \le B$ for all $k$ and $b_k \searrow 0$, then

$$\sum_{k=1}^{\infty} |s_k (b_k - b_{k+1})| \le B \sum_{k=1}^{\infty} (b_k - b_{k+1}) = B b_1 < \infty, \tag{A.5.28}$$

so the infinite series

$$\sum_{k=1}^{\infty} s_k (b_k - b_{k+1}) \tag{A.5.29}$$

is absolutely convergent, and the convergence of the left side of (A.5.27) readily follows.

For a first generalization of Proposition A.5.1, let us make a change of variable, $r \mapsto e^{-s}$, so $r \nearrow 1 \Leftrightarrow s \searrow 0$. Also think of $\{k \in \mathbb{Z}^+\}$ as a discretization of $\{t \in \mathbb{R}^+\}$. To proceed, assume we have

$$u, v : [0, \infty) \longrightarrow [0, \infty), \quad \text{monotone increasing,} \tag{A.5.30}$$

e.g., $t_1 < t_2 \Rightarrow u(t_1) \le u(t_2)$, and right continuous. Also assume that $u(0) = v(0) = 0$, and that

$$u(t),\, v(t) \le C_\varepsilon e^{\varepsilon t}, \quad \forall\, \varepsilon > 0. \tag{A.5.31}$$

Now form

$$f(t) = u(t) - v(t). \tag{A.5.32}$$

An example would be a piecewise constant $f(t)$, with jumps $a_k$ at $t = k$. The following result generalizes Proposition A.5.1. We use the Stieltjes integral, discussed in Appendix A.4.

**Proposition A.5.6.** *Take $f$ as above, and assume*

$$f(t) \longrightarrow A, \quad \text{as} \;\; t \to \infty. \tag{A.5.33}$$

*Then*

$$\int_0^\infty e^{-st}\, df(t) \longrightarrow A, \quad \text{as} \;\; s \searrow 0. \tag{A.5.34}$$

**Proof.** The hypothesis (A.5.31) implies that the left side of (A.5.33) is an absolutely convergent integral for each $s > 0$. Replacing summation by parts by integration by parts in the Stieltjes integral, we have

$$\begin{aligned} \int_0^\infty e^{-st}\, df(t) &= s \int_0^\infty e^{-st} f(t)\, dt \\ &= A + s \int_0^\infty e^{-st} [f(t) - A]\, dt. \end{aligned} \tag{A.5.35}$$

Pick $\varepsilon > 0$, and then take $K < \infty$ such that

$$t \ge K \Longrightarrow |f(t) - A| \le \varepsilon. \tag{A.5.36}$$

Then

$$\begin{aligned} & s \int_0^\infty e^{-st} |f(t) - A|\, dt \\ & \le s \int_0^K e^{-st} |f(t) - A|\, dt + \varepsilon s \int_K^\infty e^{-st}\, dt \\ & \le \Bigl(\sup_{t \le K} |f(t) - A|\Bigr) K s + \varepsilon. \end{aligned} \tag{A.5.37}$$

Hence

$$\limsup_{s\searrow 0} \Big| \int_0^\infty e^{-st} df(t) - A \Big| \le \varepsilon, \quad \forall\, \varepsilon > 0, \tag{A.5.38}$$

and we have (A.5.34). □

We next replace the hypothesis (A.5.33) by

$$f(t) \sim At^\alpha, \quad \text{as} \ \ t \to \infty, \tag{A.5.39}$$

given $\alpha \ge 0$.

**Proposition A.5.7.** *In the setting of Proposition* A.5.6*, if the hypothesis* (A.5.33) *is replaced by* (A.5.39)*, with* $\alpha \ge 0$*, then*

$$\int_0^\infty e^{-st}\, df(t) \sim A\Gamma(\alpha+1) s^{-\alpha}, \quad \textit{as} \ \ s \searrow 0. \tag{A.5.40}$$

**Proof.** Noting that

$$\int_0^\infty e^{-st} t^\alpha\, dt = \Gamma(\alpha+1) s^{-\alpha-1}, \tag{A.5.41}$$

we have, in place of (A.5.35),

$$\begin{aligned} \int_0^\infty e^{-st}\, df(t) &= s \int_0^\infty e^{-st} f(t)\, dt \\ &= A\Gamma(\alpha+1)s^{-\alpha} + s \int_0^\infty e^{-st} [f(t) - At^\alpha]\, dt. \end{aligned} \tag{A.5.42}$$

Now, in place of (A.5.36), pick $\varepsilon > 0$ and take $K < \infty$ such that

$$t \ge K \Longrightarrow |f(t) - At^\alpha| \le \varepsilon t^\alpha. \tag{A.5.43}$$

We have

$$\begin{aligned} & s^{1+\alpha} \int_0^\infty e^{-st} |f(t) - At^\alpha|\, dt \\ & \le s^{1+\alpha} \int_0^K e^{-st} |f(t) - At^\alpha|\, dt + \varepsilon s^{1+\alpha} \int_K^\infty e^{-st} t^\alpha\, dt \\ & \le \Big( \sup_{t\le K} |f(t) - At^\alpha| \Big) K s^{1+\alpha} + \varepsilon \Gamma(\alpha+1). \end{aligned} \tag{A.5.44}$$

Hence

$$\limsup_{s\searrow 0} \Big| s^\alpha \int_0^\infty e^{-st}\, df(t) - A\Gamma(\alpha+1) \Big| \le \varepsilon\Gamma(\alpha+1), \quad \forall\, \varepsilon > 0, \tag{A.5.45}$$

and we have (A.5.40). □

In the next result, we weaken the hypothesis (A.5.39).

**Proposition A.5.8.** *Let $f$ be as in Proposition* A.5.7, *except that we replace hypothesis* (A.5.39) *by the hypothesis*

$$f_1(t) \sim Bt^{\alpha+1}, \quad \text{as} \ \ t \to \infty, \tag{A.5.46}$$

*where*

$$f_1(t) = \int_0^t f(\tau)\, d\tau. \tag{A.5.47}$$

*Then the conclusion* (A.5.40) *holds, with*

$$A = (\alpha+1)B. \tag{A.5.48}$$

**Proof.** In place of (A.5.42), write

$$\begin{aligned} \int_0^\infty e^{-st}\, df(t) &= s\int_0^\infty e^{-st} f(t)\, dt \\ &= s\int_0^\infty e^{-st}\, df_1(t). \end{aligned} \tag{A.5.49}$$

We apply Proposition A.5.7, with $f$ replaced by $f_1$ (and $At^\alpha$ replaced by $Bt^{\alpha+1}$), to deduce that

$$\int_0^\infty e^{-st}\, df_1(t) \sim B\Gamma(\alpha+2)s^{-\alpha-1}. \tag{A.5.50}$$

Multiplying both sides of (A.5.50) by $s$ and noting that $\Gamma(\alpha+2) = (\alpha+1)\Gamma(\alpha+1)$, we have (A.5.40). ☐

The Abelian theorems given above have been stated for real-valued $f$, but we can readily treat complex-valued $f$, simply by taking the real and imaginary parts.

Tauberian theorems are to some degree converse results to Abelian theorems. However, Tauberian theorems require some auxiliary structure on $f$, or, in the setting of Proposition A.5.1, on $\{a_k\}$. To see this, we bring in the geometric series

$$\sum_{k=0}^\infty z^k = \frac{1}{1-z}, \quad \text{for} \ \ |z| < 1. \tag{A.5.51}$$

If we take $a_k = e^{ik\theta}$, then

$$\sum_{k=0}^\infty a_k r^k = \frac{1}{1-re^{i\theta}} \longrightarrow \frac{1}{1-e^{i\theta}}, \quad \text{as} \ \ r \nearrow 1, \tag{A.5.52}$$

for $0 < \theta < 2\pi$. However, since $|a_k| \equiv 1$, the series $\sum_k a_k$ is certainly not convergent. A classical theorem of Littlewood does obtain the convergence (A.5.1) from convergence $f(r) \to A$ in (A.5.2), under the hypothesis that $|a_k| \le C/k$. One can see [**11**] for such a result.

The Tauberian theorems we concentrate on here require

$$f(t) = u(t) \nearrow, \tag{A.5.53}$$

or, in the setting of Proposition A.5.1, $a_k \geq 0$. In the latter case, it is clear that

$$\lim_{r\nearrow 1} \sum a_k r^k = A \Longrightarrow \sum a_k < \infty, \tag{A.5.54}$$

and then the "easy" result Proposition A.5.2 applies.

However, converses of results like Proposition A.5.7 when $\alpha > 0$ are not at all trivial. In particular, we have the following important result, known as Karamata's Tauberian theorem.

**Proposition A.5.9.** *Let* $u : [0,\infty) \to [0,\infty)$ *be an increasing, right-continuous function, as in* (A.5.30). *Take* $\alpha \in (0,\infty)$, *and assume*

$$\int_0^\infty e^{-st}\, du(t) \sim Bs^{-\alpha}, \quad \textit{as} \ \ s \searrow 0. \tag{A.5.55}$$

*Then*

$$u(t) \sim \frac{B}{\Gamma(\alpha+1)} t^\alpha, \quad \textit{as} \ \ t \nearrow \infty. \tag{A.5.56}$$

**Proof.** Let us phrase the hypothesis (A.5.55) as

$$\int_0^\infty e^{-st}\, du(t) \sim B\varphi(s), \tag{A.5.57}$$

where

$$\varphi(s) = s^{-\alpha} = \int_0^\infty e^{-st} v(t)\, dt, \quad v(t) = \frac{1}{\Gamma(\alpha)} t^{\alpha-1}. \tag{A.5.58}$$

Our goal in (A.5.56) is equivalent to showing that

$$\int_0^{1/s} du(t) = B \int_0^{1/s} v(t)\, dt + o(s^{-\alpha}), \quad s \searrow 0. \tag{A.5.59}$$

We tackle this problem in stages, examining when we can show that

$$\int_0^\infty F(st)\, du(t) = B \int_0^\infty F(st) v(t)\, dt + o(s^{-\alpha}), \tag{A.5.60}$$

for various functions $F(t)$, ultimately including

$$\begin{aligned} \chi_I(t) = 1, &\quad \text{for} \ \ 0 \leq t < 1, \\ 0, &\quad \text{for} \ \ t \geq 1. \end{aligned} \tag{A.5.61}$$

We start with the function space

$$\mathcal{E} = \Big\{ \sum_{k=1}^M \gamma_k e^{-kt} : \gamma_k \in \mathbb{R},\ M \in \mathbb{N} \Big\}. \tag{A.5.62}$$

As seen in Appendix 3.6, as a consequence of the Weierstrass approximation theorem, the space $\mathcal{E}$ is dense in

$$C_0([0,\infty)) = \{f \in C([0,\infty)) : \lim_{t\to\infty} f(t) = 0\}. \tag{A.5.63}$$

Now if $F \in \mathcal{E}$, say

$$F(t) = \sum_{k=1}^{M} \gamma_k e^{-kt}, \tag{A.5.64}$$

then (A.5.57) implies

$$\begin{aligned} \int_0^\infty F(st)\,du(t) &= \sum_{k=1}^{M} \gamma_k \int_0^\infty e^{-skt}\,du(t) \\ &= B\sum_{k=1}^{M} \gamma_k \varphi(ks) + o\Big(\sum_{k=1}^{M}(ks)^{-\alpha}\Big) \\ &= B\int_0^\infty F(st)v(t)\,dt + o(s^{-\alpha}). \end{aligned} \tag{A.5.65}$$

Hence (A.5.60) holds for all $F \in \mathcal{E}$. The following moves us along. ☐

**Proposition A.5.10.** *In the setting of Proposition* A.5.9*, the result* (A.5.60) *holds for all*

$$F \in C_0([0,\infty)) \quad \textit{such that} \quad e^t F \in C_0([0,\infty)). \tag{A.5.66}$$

**Proof.** Given such $F$ and given $\varepsilon > 0$, we take $H \in \mathcal{E}$ such that $\sup |H(t) - e^t F(t)| \le \varepsilon$, and set $G(t) = e^{-t}H(t)$, so

$$G \in \mathcal{E}, \quad |F(t) - G(t)| \le \varepsilon e^{-t}. \tag{A.5.67}$$

This implies

$$\int_0^\infty |F(st) - G(st)|\,du(t) \le \varepsilon \int_0^\infty e^{-st}\,du(t) \tag{A.5.68}$$

and

$$\int_0^\infty |F(st) - G(st)|v(t)\,dt \le \varepsilon \int_0^\infty e^{-st}v(t)\,dt. \tag{A.5.69}$$

The facts that the right sides of (A.5.68) and (A.5.69) are both $\le C\varepsilon\varphi(s)$ follow from (A.5.57) and (A.5.58), respectively. But we know that (A.5.60) holds with $G$ in place of $F$. Hence

$$\Big|\int_0^\infty F(st)\,du(t) - B\int_0^\infty F(st)v(t)\,dt\Big| \le 2C\varepsilon\varphi(s) + o(\varphi(s)), \tag{A.5.70}$$

for each $\varepsilon > 0$. Taking $\varepsilon \searrow 0$ yields the lemma. ☐

We now tackle (A.5.60) for $F = \chi_I$, given by (A.5.61). For each $\delta \in (0,1/2]$, take $f_\delta, g_\delta \in C_0([0,\infty))$ such that

$$0 \le f_\delta \le \chi_I \le g_\delta \le 1, \tag{A.5.71}$$

with

$$\begin{aligned} f_\delta(t) = 1 \quad &\text{for} \quad 0 \le t \le 1-\delta, \\ 0 \quad &\text{for} \quad t \ge 1 \end{aligned} \tag{A.5.72}$$

and

$$\begin{aligned} g_\delta(t) = 1 \quad &\text{for} \quad 0 \le t \le 1, \\ 0 \quad &\text{for} \quad t \ge 1+\delta. \end{aligned} \tag{A.5.73}$$

Note that Proposition A.5.10 is applicable to each $f_\delta$ and $g_\delta$. Hence

$$\begin{aligned} \int_0^\infty \chi_I(st)\,du(t) &\le \int_0^\infty g_\delta(st)\,du(t) \\ &= A\int_0^\infty g_\delta(st)v(t)\,dt + o(\varphi(s)) \end{aligned} \tag{A.5.74}$$

and

$$\begin{aligned} \int_0^\infty \chi_I(st)\,du(t) &\ge \int_0^\infty f_\delta(st)\,du(t) \\ &= A\int_0^\infty f_\delta(st)v(t)\,dt + o(\varphi(s)). \end{aligned} \tag{A.5.75}$$

Complementing the estimates (A.5.74)–(A.5.75), we have

$$\begin{aligned} &\int_0^\infty \big[g_\delta(st) - f_\delta(st)\big]v(t)\,dt \\ &\quad \le \int_{(1-\delta)/s}^{(1+\delta)/s} v(t)\,dt \\ &\quad \le \frac{2\delta}{s}\cdot\max\Big\{v(t) : \frac{1-\delta}{s} \le t \le \frac{1+\delta}{s}\Big\} \\ &\quad \le C\delta s^{-\alpha}. \end{aligned} \tag{A.5.76}$$

It then follows from (A.5.74)–(A.5.75) that

$$\begin{aligned} &\limsup_{s\searrow 0}\, s^\alpha\Big|\int_0^\infty \chi_I(st)\,du(t) - B\int_0^\infty \chi_I(st)v(t)\,dt\Big| \\ &\le \inf_{\delta\le 1/2} C\delta = 0. \end{aligned} \tag{A.5.77}$$

This yields (A.5.59) and hence completes the proof of Proposition A.5.9.

Arguments proving Proposition A.5.9 can also be used to establish variants of the implication (A.5.55) ⇒ (A.5.56), such as

$$\begin{aligned}(\text{A.5.78})\qquad &\int_0^\infty e^{-st}\,du(t) \sim A\Big(\log\frac{1}{s}\Big)s^{-\alpha}, \quad s\searrow 0,\\ &\Longrightarrow u(t) \sim \frac{A}{\Gamma(\alpha+1)}t^{\alpha}(\log t), \quad t\to\infty,\end{aligned}$$

provided $u:[0,\infty)\to[0,\infty)$ is increasing and $\alpha>0$. The reader might like to verify this. Hint: replace the calculation in (A.5.58) by the Laplace transform identity

$$(\text{A.5.79})\qquad \int_0^\infty e^{-st}t^{\alpha-1}(\log t)\,dt = \big(\Gamma'(\alpha)-\Gamma(\alpha)\log s\big)s^{-\alpha}.$$

See Exercise 3 in §4.3.

Putting together Propositions A.5.8 and A.5.9 yields the following result, of use in §4.4. In fact, Proposition A.5.11 below is equivalent to Proposition 4.4.14, which plays a role in the proof of the prime number theorem.

**Proposition A.5.11.** *Let $\psi:[0,\infty)\to[0,\infty)$ be an increasing function, and set $\psi_1(t)=\int_0^t\psi(\tau)\,d\tau$. Take $B\in(0,\infty)$, $\alpha\in[0,\infty)$. Then*

$$\begin{aligned}(\text{A.5.80})\qquad &\psi_1(t)\sim Bt^{\alpha+1}, \quad \text{as } t\to\infty\\ &\Longrightarrow \psi(t)\sim(\alpha+1)Bt^{\alpha}, \quad \text{as } t\to\infty.\end{aligned}$$

**Proof.** First, by Proposition A.5.8, the hypothesis on $\psi_1(t)$ in (A.5.80) implies

$$(\text{A.5.81})\qquad \int_0^\infty e^{-st}\,d\psi(t)\sim B\Gamma(\alpha+2)s^{-\alpha}, \quad s\searrow 0.$$

Then Karamata's Tauberian theorem applied to (A.5.81) yields the conclusion in (A.5.80), at least for $\alpha>0$. But such a conclusion for $\alpha=0$ is elementary. □

Karamata's Tauberian theorem is a widely applicable tool. In addition to the application we have made in the proof of the prime number theorem, it has uses in partial differential equations, which can be found in [**46**].

We mention another Tauberian theorem, known as Ikehara's Tauberian theorem.

**Proposition A.5.12.** *Let $u:[0,\infty)\to[0,\infty)$ be increasing, and consider*

$$(\text{A.5.82})\qquad F(s)=\int_0^\infty e^{-st}\,du(t).$$

*Assume the integral is absolutely convergent on $\{s\in\mathbb{C}:\operatorname{Re}s>1\}$ and that*

$$(\text{A.5.83})\qquad F(s)-\frac{A}{s-1}\quad \textit{is continuous on}\quad \{s\in\mathbb{C}:\operatorname{Re}s\ge 1\}.$$

*Then*

$$e^{-t}u(t) \longrightarrow A \quad \textit{as} \quad t \to \infty. \tag{A.5.84}$$

We refer to [**11**] for a proof of Proposition A.5.12. This result is applicable to (4.4.68),

$$-\frac{\zeta'(s)}{\zeta(s)} = \int_1^\infty x^{-s}\, d\psi(x), \tag{A.5.85}$$

with $\psi$ given by (4.4.66). In fact, setting $u(t) = \psi(e^t)$ gives

$$-\frac{\zeta'(s)}{\zeta(s)} = \int_0^\infty e^{-ts}\, du(t). \tag{A.5.86}$$

Then Propositions 4.4.2 and 4.4.4 imply (A.5.83), with $A = 1$, so (A.5.84) yields $e^{-t}u(t) \to 1$, and hence

$$\frac{\psi(x)}{x} \longrightarrow 1, \quad \text{as} \quad x \to \infty. \tag{A.5.87}$$

In this way, we get another proof of (4.4.70), which yields the prime number theorem. This proof requires less information on the Riemann zeta function than was used in the proof of Theorem 4.4.10. It requires Proposition 4.4.4, but not its refinement, Proposition 4.4.8 and Corollary 4.4.9. On the other hand, the proof of Ikehara's theorem is more subtle than that of Proposition A.5.11. This illustrates the advantage of obtaining more insight into the Riemann zeta function.

## A.6. Cubics, quartics, and quintics

We take up the problem of finding formulas for the roots of polynomials, i.e., elements $z \in \mathbb{C}$ such that

$$P(z) = z^n + a_{n-1}z^{n-1} + a_{n-2}z^{n-1} + \cdots + a_1 z + a_0 = 0, \tag{A.6.1}$$

given $a_j \in \mathbb{C}$, with emphasis on the cases $n = 3, 4$, and 5. We start with generalities, involving two elementary transformations. First, if $z = w - a_{n-1}/n$, then

$$P(z) = Q(w) = w^n + b_{n-2}w^{n-2} + \cdots + b_0, \tag{A.6.2}$$

with $b_j \in \mathbb{C}$. We have arranged that the coefficient of $w^{n-1}$ be zero. In case $n = 2$, we get

$$Q(w) = w^2 + b_0, \tag{A.6.3}$$

with roots $w = \pm\sqrt{-b_0}$, leading to the familiar quadratic formula.

For $n \geq 3$, the form of $Q(w)$ is more complicated. We next take $w = \gamma u$, so

$$Q(w) = \gamma^n R(u) = \gamma^n(u^n + c_{n-2}u^{n-2} + \cdots + c_0), \quad c_j = \gamma^{j-n}b_j. \tag{A.6.4}$$

In particular, $c_{n-2} = \gamma^{-2} b_{n-2}$. This has the following significance. If $b_{n-2} \neq 0$, we can preselect $c \in \mathbb{C} \setminus 0$ and choose $\gamma \in \mathbb{C}$ such that $\gamma^{-2} b_{n-2} = c$, i.e.,

$$\gamma = \big(c b_{n-2}^{-1}\big)^{1/2}, \tag{A.6.5}$$

and therefore achieve that $c$ is the coefficient of $u^{n-2}$ in $R(u)$.

In case $n = 3$, we get

$$R(u) = u^3 + cu + d, \quad d = \gamma^{-3} c_0. \tag{A.6.6}$$

Our task is to find a formula for the roots of $R(u)$, along the way making a convenient choice of $c$ to facilitate this task. One neat approach involves a trigonometric identity, expressing $\sin 3\zeta$ as a polynomial in $\sin\zeta$. Starting with

$$\sin(\zeta + 2\zeta) = \sin\zeta\, \cos 2\zeta + \cos\zeta\, \sin 2\zeta, \tag{A.6.7}$$

it is an exercise to obtain

$$\sin 3\zeta = -4\sin^3\zeta + 3\sin\zeta, \quad \forall\, \zeta \in \mathbb{C}. \tag{A.6.8}$$

Consequently, we see that the equation

$$4u^3 - 3u + 4d = 0 \tag{A.6.9}$$

is solved by

$$u = \sin\zeta, \quad \text{if } \ 4d = \sin 3\zeta. \tag{A.6.10}$$

Here we have taken $c = -3/4$ in (A.6.6). In this case, the other solutions to (A.6.9) are

$$u_2 = \sin\Big(\zeta + \frac{2\pi}{3}\Big), \quad u_3 = \sin\Big(\zeta - \frac{2\pi}{3}\Big). \tag{A.6.11}$$

REMARK. The polynomial (A.6.9) has a double root if and only if $4d = \pm 1$. In such a case, we can take $\zeta = \pi/6$ (respectively, $\zeta = -\pi/6$) and then $u = u_2$ (respectively, $u = u_3$).

Now (A.6.10)–(A.6.11) provide formulas for the solutions to (A.6.9), but they involve the transcendental functions sin and $\sin^{-1}$. We can obtain purely algebraic formulas as follows. If $4d = \sin 3\zeta$, as in (A.6.10),

$$\begin{aligned} e^{3i\zeta} = \eta &\Longrightarrow \eta - \eta^{-1} = 8id \\ &\Longrightarrow \eta = 4id \pm \sqrt{-(4d)^2 + 1}. \end{aligned} \tag{A.6.12}$$

Then

$$u = \sin\zeta = \frac{1}{2i}\big(\eta^{1/3} - \eta^{-1/3}\big). \tag{A.6.13}$$

Note that the two roots $\eta_\pm$ in (A.6.12) are related by $\eta_- = -1/\eta_+$, so they lead to the same quantity $\eta^{1/3} - \eta^{-1/3}$. In (A.6.13), the cube root is regarded as a multivalued function; for $a \in \mathbb{C}$,

$$a^{1/3} = \{b \in \mathbb{C} : b^3 = a\}. \tag{A.6.14}$$

Similarly, if $a \neq 0$, then

$$a^{1/3} - a^{-1/3} = \{b - b^{-1} : b^3 = a\}. \tag{A.6.15}$$

Taking the three cube roots of $\eta$ in (A.6.13) gives the three roots of $R(u)$.

We have obtained an algebraic formula for the roots of (A.6.6), with the help of the functions sin and $\sin^{-1}$. Now we will take an alternative route, avoiding explicit use of these functions. To get it, note that, with $v = e^{i\zeta}$, the identity (A.6.8) is equivalent to

$$v^3 - v^{-3} = (v - v^{-1})^3 + 3(v - v^{-1}), \tag{A.6.16}$$

which is also directly verifiable via the binomial formula. Thus, if we set

$$u = v - v^{-1}, \tag{A.6.17}$$

and take $c = 3$ in (A.6.6), we see that $R(u) = 0$ is equivalent to

$$v^3 - v^{-3} = -d. \tag{A.6.18}$$

This time, in place of (A.6.12), we have

$$\begin{aligned} v^3 = \eta &\Longrightarrow \eta - \eta^{-1} = -d \\ &\Longrightarrow \eta = -\frac{d}{2} \pm \frac{1}{2}\sqrt{d^2 + 4}. \end{aligned} \tag{A.6.19}$$

Then, in place of (A.6.13), we get

$$u = \eta^{1/3} - \eta^{-1/3}. \tag{A.6.20}$$

Again the two roots $\eta_\pm$ in (A.6.19) are related by $\eta_- = -1/\eta_+$, so they lead to the same quantity $\eta^{1/3} - \eta^{-1/3}$. Furthermore, taking the three cube roots of $\eta$ gives, via (A.6.20), the three roots of $R(u)$. The two formulas (A.6.13) and (A.6.20) have a different appearance simply because of the different choices of $c$: $c = -3/4$ for (A.6.13) and $c = 3$ for (A.6.20).

We move on to quartic polynomials,

$$P(z) = z^4 + a_3 z^3 + a_2 z^2 + a_1 z + a_0. \tag{A.6.21}$$

As before, setting $z = w - a_3/4$ yields

$$P(z) = Q(w) = w^4 + bw^2 + cw + d. \tag{A.6.22}$$

We seek a formula for the solutions to $Q(w) = 0$. We can rewrite this equation as

$$\Bigl(w^2 + \frac{b}{2}\Bigr)^2 = -cw - d + \frac{b^2}{4}. \tag{A.6.23}$$

The left side is a perfect square, but the right side is not, unless $c = 0$. We desire to add a certain quadratic polynomial in $w$ to both sides of (A.6.23) so that the resulting polynomials are both perfect squares. We aim for the new left side to have the form

$$\Big(w^2 + \frac{b}{2} + \alpha\Big)^2, \tag{A.6.24}$$

with $\alpha \in \mathbb{C}$ to be determined. This requires adding $2\alpha(w^2 + b/2) + \alpha^2$ to the left side of (A.6.23), and adding this to the right side of (A.6.23) yields

$$2\alpha w^2 - cw + \Big(\alpha^2 + b\alpha + \frac{b^2}{4} - d\Big). \tag{A.6.25}$$

We want this to be a perfect square. If it were, it would have to be

$$\Big(\sqrt{2\alpha}\, w - \frac{c}{2\sqrt{2\alpha}}\Big)^2. \tag{A.6.26}$$

This is equal to (A.6.25) if and only if

$$8\alpha^3 + 4b\alpha^2 + (2b^2 - 8d)\alpha - c = 0. \tag{A.6.27}$$

This is a *cubic* equation for $\alpha$, solvable by means discussed above. For (A.6.26) to work, we need $\alpha \neq 0$. If $\alpha = 0$ solves (A.6.27), this forces $c = 0$, and hence $Q(w) = w^4 + bw^2 + d$, which is a quadratic polynomial in $w^2$, solvable by elementary means. Even if $c = 0$, (A.6.27) has a nonzero root unless also $b = d = 0$, i.e., unless $Q(w) = w^4$.

Now, assuming $Q(w) \neq w^4$, we pick $\alpha$ to be *one* nonzero solution to (A.6.27). Then the solutions to $Q(w) = 0$ are given by

$$w^2 + \frac{b}{2} + \alpha = \pm\Big(\sqrt{2\alpha}\, w - \frac{c}{2\sqrt{2\alpha}}\Big). \tag{A.6.28}$$

This is a pair of quadratic equations. Each has two roots, and together they yield the four roots of $Q(w)$.

It is interesting to consider a particular quartic equation for which a different approach, not going through (A.6.22), is effective, namely

$$z^4 + z^3 + z^2 + z + 1 = 0, \tag{A.6.29}$$

which arises from factoring $z - 1$ out of $z^5 - 1$, therefore seeking the other fifth roots of unity. Let us multiply (A.6.29) by $z^{-2}$, obtaining

$$z^2 + z + 1 + z^{-1} + z^{-2} = 0. \tag{A.6.30}$$

The symmetric form of this equation suggests making the substitution

$$w = z + z^{-1}, \tag{A.6.31}$$

so

$$w^2 = z^2 + 2 + z^{-2}, \tag{A.6.32}$$

and (A.6.30) becomes

$$w^2 + w - 1 = 0, \tag{A.6.33}$$

a quadratic equation with solutions

$$w = -\frac{1}{2} \pm \frac{1}{2}\sqrt{5}. \tag{A.6.34}$$

Then (A.6.31) becomes a quadratic equation for $z$. We see that (A.6.29) is solvable in a fashion that requires no extraction of cube roots. Noting that the roots of (A.6.29) have absolute value 1, we see that $w = 2\Re z$, and (A.6.34) says

$$\cos\frac{2\pi}{5} = \frac{\sqrt{5}-1}{4}, \quad \cos\frac{4\pi}{5} = -\frac{\sqrt{5}+1}{4}. \tag{A.6.35}$$

Such a calculation allows one to construct a regular pentagon with compass and straightedge.

Let us extend the scope of this, and look at solutions to $z^7 - 1 = 0$, arising in the construction of a regular 7-gon. Factoring out $z - 1$ yields

$$z^6 + z^5 + z^4 + z^3 + z^2 + z + 1 = 0, \tag{A.6.36}$$

or equivalently

$$z^3 + z^2 + z + 1 + z^{-1} + z^{-2} + z^{-3} = 0. \tag{A.6.37}$$

Again we make the substitution (A.6.31). Complementing (A.6.32) with

$$w^3 = z^3 + 3z + 3z^{-1} + z^{-3}, \tag{A.6.38}$$

we see that (A.6.37) leads to the *cubic* equation

$$q(w) = w^3 + w^2 - 2w - 1 = 0. \tag{A.6.39}$$

Since $q(-1) > 0$ and $q(0) < 0$, we see that (A.6.39) has three real roots, satisfying

$$w_3 < w_2 < 0 < w_1, \tag{A.6.40}$$

and, parallel to (A.6.35), we have

$$\cos\frac{2\pi}{7} = \frac{w_1}{2}, \quad \cos\frac{4\pi}{7} = \frac{w_2}{2}, \quad \cos\frac{6\pi}{7} = \frac{w_3}{2}. \tag{A.6.41}$$

One major difference between (A.6.35) and (A.6.41) is that the computation of $w_j$ involves the extraction of cube roots. In the time of Euclid, the problems of whether one could construct cube roots or a regular 7-gon by compass and straightedge were regarded as major mysteries. Much later, a young Gauss proved that one could make such a construction of a regular $n$-gon if and only if $n$ is of the form $2^k$, perhaps times a product of distinct Fermat primes, i.e., primes of the form $p = 2^{2^j} + 1$. The smallest examples are $p = 2^1 + 1 = 3$, $p = 2^2 + 1 = 5$, and $p = 2^4 + 1 = 17$. Modern treatments of these problems cast them in the framework of Galois theory; see [**25**].

We now consider quintics, i.e., fifth degree polynomials,

$$P(z) = z^5 + a_4 z^4 + a_3 z^3 + a_2 z^2 + a_1 z + a_0. \tag{A.6.42}$$

The treatment of this differs markedly from that of equations of degree $\le 4$, in that one cannot find a formula for the roots in terms of radicals, i.e., involving a finite number of arithmetical operations and extraction of $n$th roots. That no such general formula exists was proved by Abel. Then Galois showed that specific equations, such as

$$z^5 - 16z + 2 = 0, \tag{A.6.43}$$

had roots that could not be obtained from the set $\mathbb{Q}$ of rational numbers via radicals. We will not delve into Galois theory here; see [**25**]. Rather, we will discuss how there are formulas for the roots of (A.6.42) that bring in other special functions.

In our analysis, we will find it convenient to assume the roots of $P(z)$ are distinct. Otherwise, any double root of $P(z)$ is also a root of $P'(z)$, which has degree 4. We also assume $z = 0$ is not a root, i.e., $a_0 \neq 0$.

A key tool in the analysis of (A.6.42) is the reduction to *Bring-Jerrard normal form*:

$$Q(w) = w^5 - w + a. \tag{A.6.44}$$

That is to say, given $a_j \in \mathbb{C}$, one can find $a \in \mathbb{C}$ such that the roots of $P(z)$ in (A.6.42) are related to the roots of $Q(w)$ in (A.6.44) by means of solving polynomial equations of degree $\le 4$. Going from (A.6.42) to (A.6.44) is done via a *Tschirnhaus transformation.* Generally, such a transformation takes $P(z)$ in (A.6.42) to a polynomial of the form

$$Q(w) = w^5 + b_4 w^4 + b_3 w^3 + b_2 w^2 + b_1 w + b_0, \tag{A.6.45}$$

in a way that the roots of $P(z)$ and of $Q(w)$ are related as described above. The ultimate goal is to produce a Tschirnhaus transformation that yields (A.6.45) with

$$b_4 = b_3 = b_2 = 0. \tag{A.6.46}$$

As we have seen, the linear change of variable $z = w - a_4/5$ achieves $b_4 = 0$, but here we want to achieve much more. This will involve a nonlinear change of variable.

Following [**3**], we give a matrix formulation of Tschirnhaus transformations. Relevant linear algebra background can be found in §§6–7 of [**49**].

To start, given (A.6.42), pick a matrix $A \in M(5, \mathbb{C})$ whose characteristic polynomial is $P(z)$,

$$P(z) = \det(zI - A). \tag{A.6.47}$$

For example, $A$ could be the companion matrix of $P$. Note that the set of eigenvalues of $A$,

$$\operatorname{Spec} A = \{z_j : 1 \le j \le 5\}, \tag{A.6.48}$$

is the set of roots of (A.6.42). The Cayley-Hamilton theorem implies

$$P(A) = A^5 + a_4A^4 + a_3A^3 + a_2A^2 + a_1A + a_0I = 0. \tag{A.6.49}$$

It follows that

$$\mathcal{A} = \operatorname{Span}\{I, A, A^2, A^3, A^4\} \tag{A.6.50}$$

is a commutative matrix algebra. The hypothesis that the roots of $A$ are disinct implies that $P$ is the minimal polynomial of $A$, so the 5 matrices listed in (A.6.50) form a basis of $\mathcal{A}$.

In this setting, a Tschirnhaus transformation is produced by taking

$$B = \beta(A) = \sum_{j=0}^{m} \beta_j A^j, \tag{A.6.51}$$

where $\beta(z)$ is a nonconstant polynomial of degree $m \le 4$. Then $B \in \mathcal{A}$, with characteristic polynomial

$$Q(w) = \det(wI - B), \tag{A.6.52}$$

of the form (A.6.45). The set of roots of $Q(w)$ forms

$$\operatorname{Spec} B = \{\beta(z_j) : z_j \in \operatorname{Spec} A\}. \tag{A.6.53}$$

We can entertain two possibilities, depending on the behavior of

$$\{I, B, B^2, B^3, B^4\}. \tag{A.6.54}$$

Case I. The set (A.6.54) is linearly dependent.

Then $q(B) = 0$ for some polynomial $q(w)$ of degree $\le 4$, so

$$\operatorname{Spec} B = \{w_j : 1 \le j \le 5\} \tag{A.6.55}$$

and each $w_j$ is a root of $q$. Methods described earlier in this appendix apply to solving for the roots of $q$, and to find the roots of $P(z)$, i.e., the elements of $\operatorname{Spec} A$, we solve

$$\beta(z_j) = w_j \tag{A.6.56}$$

for $z_j$. Since, for each $j$, (A.6.56) has $m$ solutions, this may produce solutions not in $\operatorname{Spec} A$, but one can test each solution $z_j$ to see if it is a root of $P(z)$.

Case II. The set (A.6.54) is linearly independent.

Then this set spans $\mathcal{A}$, so we can find $\gamma_j \in \mathbb{C}$ such that

$$A = \sum_{j=0}^{4} \gamma_j B^j = \Gamma(B). \tag{A.6.57}$$

It follows that

$$\operatorname{Spec} A = \{\Gamma(w_j) : w_j \in \operatorname{Spec} B\}. \tag{A.6.58}$$

It remains to find $\operatorname{Spec} B$, i.e., the set of roots of $Q(w)$ in (A.6.45). It is here that we want to implement (A.6.46). The following result is relevant to this endeavor.

**Lemma A.6.1.** *Let $Q(w)$, given by* (A.6.52)*, have the form* (A.6.45)*, and pick $\ell \in \{1, \dots, 5\}$. Then*

$$b_{5-j} = 0 \text{ for } 1 \le j \le \ell \Longleftrightarrow \operatorname{Tr} B^j = 0 \text{ for } 1 \le j \le \ell. \tag{A.6.59}$$

**Proof.** To start, we note that $b_4 = -\operatorname{Tr} B$. More generally, $b_{5-j}$ is given (up to a sign) as an elementary symmetric polynomial in the eigenvalues $\{w_1, \dots, w_5\}$ of $B$. The equivalence (A.6.54) follows from the classical Newton formula for these symmetric polynomials in terms of the polynomials $w_1^j + \cdots + w_5^j = \operatorname{Tr} B^j$. □

We illustrate the use of (A.6.51) to achieve (A.6.59) in case $\ell = 2$. In this case, we take $m = 2$ in (A.6.43), and set

$$B = \beta_0 I + \beta_1 A + \beta_2 A^2, \quad \beta_2 = 1. \tag{A.6.60}$$

Then

$$\begin{aligned} B^2 = \beta_0^2 I + 2\beta_0\beta_1 A + (2\beta_2 + \beta_1^2) A^2 \\ +2\beta_1\beta_2 A^3 + \beta_2^2 A^4. \end{aligned} \tag{A.6.61}$$

Then, if

$$\xi_j = \operatorname{Tr} A^j, \tag{A.6.62}$$

we obtain

$$\begin{aligned} \operatorname{Tr} B &= 5\beta_0 + \xi_1\beta_1 + \xi_2, \\ \operatorname{Tr} B^2 &= 5\beta_0^2 + 2\xi_1\beta_0\beta_1 + \xi_2(\beta_1^2 + 2) + 2\xi_3\beta_1 + \xi_4. \end{aligned} \tag{A.6.63}$$

Set $\operatorname{Tr} B = \operatorname{Tr} B^2 = 0$. Then the first identity in (A.6.58) yields

$$\beta_0 = -\frac{1}{5}(\xi_1\beta_1 + \xi_2), \tag{A.6.64}$$

and substituting this into the second identity of (A.6.63) gives

$$\frac{1}{5}(\xi_1\beta_1 + \xi_2)^2 - \frac{2}{5}\xi_1\beta_1(\xi_1\beta_1 + \xi_2) + \xi_2\beta_1^2 + 2\xi_3\beta_1 = -2\xi_2 - \xi_4, \tag{A.6.65}$$

a quadratic equation for $\beta_1$, with leading term $(\xi_2 - \xi_1^2/5)\beta_1^2$. We solve for $\beta_0$ and $\beta_1$, and hence obtain $B \in \mathcal{A}$ with characteristic polynomial $Q(w)$ satisfying

$$Q(w) = w^5 + b_2 w^2 + b_1 w + b_0. \tag{A.6.66}$$

This goes halfway from (A.6.2) (with $n = 5$) to (A.6.46).

Before discussing closing this gap, we make another comment on achieving (A.6.66). Namely, suppose $A$ has been prepped so that

$$\operatorname{Tr} A = 0, \tag{A.6.67}$$

i.e., $A$ is replaced by $A - (1/5)(\operatorname{Tr} A)I$. Apply (A.6.60) to this new $A$. Then $\xi_1 = 0$, so (A.6.64)–(A.6.65) simplify to

$$\beta_0 = -\frac{1}{5}\xi_2, \tag{A.6.68}$$

and

$$\xi_2 \beta_1^2 + 2\xi_3 \beta_1 = -\frac{1}{5}\xi_2^2 - 2\xi_2 - \xi_4. \tag{A.6.69}$$

This latter equation is a quadratic equation for $\beta_1$ if $\xi_2 = \operatorname{Tr} A^2 \neq 0$. Of course, if $\operatorname{Tr} A^2 = 0$, we have already achieved our goal (A.6.66), with $B = A$.

Moving forward, let us now assume we have

$$A \in M(5, \mathbb{C}), \quad \operatorname{Tr} A = \operatorname{Tr} A^2 = 0, \tag{A.6.70}$$

having minimal polynomial of the form (A.6.42) with $a_4 = a_3 = 0$, and we desire to construct $B$ as in (A.6.51), satisfying

$$\operatorname{Tr} B = \operatorname{Tr} B^2 = \operatorname{Tr} B^3 = 0. \tag{A.6.71}$$

At this point, a first try would take

$$B = \beta_0 I + \beta_1 A + \beta_2 A^2 + \beta_3 A^3, \quad \beta_3 = 1. \tag{A.6.72}$$

Calculations parallel to (A.6.60)–(A.6.69) first yield

$$\beta_0 = -\frac{1}{5}\xi_3 \tag{A.6.73}$$

and then a pair of polynomial equations for $(\beta_1, \beta_2)$, one of degree 2 and one of degree 3. However, this system is more complicated than the 5th degree equation we are trying to solve. Another attack is needed.

E. Bring, and, independently, G. Jerrard, succeeded in achieving (A.6.46) by using a quartic transformation. In the current setting, this involves replacing (A.6.72) by

$$B = \beta_0 I + \beta_1 A + \beta_2 A^2 + \beta_3 A^3 + \beta_4 A^4, \quad \beta_4 = 1. \tag{A.6.74}$$

The extra parameter permits one to achieve (A.6.71) with coefficients $\beta_0, \dots,$ $\beta_3$ determined by fourth order equations. The computations are lengthy, and we refer to [**21**] for more details.

Once one has $Q(w)$ satisfying (A.6.45)–(A.6.46), i.e.,

$$Q(w) = w^5 + b_1 w + b_0, \tag{A.6.75}$$

then, if $b_1 \neq 0$, one can take $w = \gamma u$ as in (A.6.4), and, by a parallel computation, write

$$Q(w) = \gamma^5 R(u), \quad R(u) = u^5 - u + a, \tag{A.6.76}$$

with

$$\gamma^4 = -b_1, \quad a = \gamma^{-5} b_0. \tag{A.6.77}$$

$R(u)$ thus has the Bring-Jerrard normal form (A.6.44). Solving $R(u) = 0$ is equivalent to solving

$$\Phi(z) = a, \quad \Phi(z) = z - z^5. \tag{A.6.78}$$

Consequently our current task is to study mapping properties of $\Phi : \mathbb{C} \to \mathbb{C}$ and its inverse $\Phi^{-1}$, a multi-valued function known as the *Bring radical.*

To start, note that

$$\Phi'(z) = 1 - 5z^4, \tag{A.6.79}$$

and hence

$$\Phi'(\zeta) = 0 \Longleftrightarrow \zeta \in \mathcal{C} = \{\pm 5^{-1/4}, \pm i 5^{-1/4}\}. \tag{A.6.80}$$

If $z_0 \in \mathbb{C} \setminus \mathcal{C}$, the Inverse Function Theorem, Theorem 1.5.2, implies that there exist neighborhoods $\mathcal{O}$ of $z_0$ and $U$ of $\Phi(z_0)$ such that $\Phi : \mathcal{O} \to U$ is one-to-one and onto, with holomorphic inverse. This observation applies in particular to $z_0 = 0$, since

$$\Phi(0) = 0, \quad \Phi'(0) = 1. \tag{A.6.81}$$

Note that, by (A.6.79),

$$\begin{aligned} |z| < 5^{-1/4} &\Longrightarrow |\Phi'(z) - 1| < 1 \\ &\Longrightarrow \operatorname{Re} \Phi'(z) > 0, \end{aligned} \tag{A.6.82}$$

so, by Proposition 1.5.3,

$$\Phi : D_{5^{-1/4}}(0) \longrightarrow \mathbb{C} \ \text{ is one-to-one,} \tag{A.6.83}$$

where, for $\rho \in (0, \infty)$, $z_0 \in \mathbb{C}$,

$$D_\rho(z_0) = \{z \in \mathbb{C} : |z - z_0| < \rho\}. \tag{A.6.84}$$

Also note that

$$|z| = 5^{-1/4} \Longrightarrow |\Phi(z)| = |z - z^5| \geq 5^{-1/4} - 5^{-5/4} = (4/5)5^{-1/4}. \tag{A.6.85}$$

Hence, via results of §4.2 on the argument principle,

$$\Phi(D_{5^{-1/4}}(0)) \supset D_{(4/5)5^{-1/4}}(0). \tag{A.6.86}$$

We deduce that the map (A.6.83) has holomorphic inverse

$$\Phi^{-1} : D_{(4/5)5^{-1/4}}(0) \longrightarrow D_{5^{-1/4}}(0) \subset \mathbb{C}, \tag{A.6.87}$$

satisfying $\Phi^{-1}(0) = 0$.

Note that

$$\Phi(ia) = i\Phi(a). \tag{A.6.88}$$

Hence we can write, for $|a| < (4/5)5^{-1/4}$,

$$\Phi^{-1}(a) = a\Psi(a^4), \tag{A.6.89}$$

with

$$\Psi(b) \ \text{ holomorphic in } \ |b| < \frac{4^4}{5^5}, \tag{A.6.90}$$

satisfying

$$a\Psi(a^4) - a^5\Psi(a^4) = a, \quad \Psi(0) = 1, \tag{A.6.91}$$

and hence

$$\Psi(b) = 1 + b\Psi(b)^5, \quad \Psi(0) = 1. \tag{A.6.92}$$

Using (A.6.92), we can work out the power series

$$\Psi(b) = \sum_{k=0}^{\infty} \psi_k b^k, \quad \psi_0 = 1, \tag{A.6.93}$$

as follows. First, (A.6.93) yields

$$\begin{aligned} \Psi(b)^5 &= \prod_{\nu=1}^{5} \sum_{\ell_\nu=0}^{\infty} \psi_{\ell_\nu} b^{\ell_\nu} \\ &= \sum_{k=0}^{\infty} \sum_{\ell\geq 0, |\ell|=k} \psi_{\ell_1} \cdots \psi_{\ell_5} b^k, \end{aligned} \tag{A.6.94}$$

where $\ell = (\ell_1, \dots, \ell_5)$, $|\ell| = \ell_1 + \cdots + \ell_5$. Then (A.6.92) yields

$$\begin{aligned} b\sum_{k\geq 0} \psi_{k+1} b^k &= b\Psi(b)^5 \\ &= b\sum_{k\geq 0} \sum_{\ell\geq 0, |\ell|=k} \psi_{\ell_1} \cdots \psi_{\ell_5} b^k, \end{aligned} \tag{A.6.95}$$

and hence

$$\psi_{k+1} = \sum_{\ell \ge 0, |\ell|=k} \psi_{\ell_1} \cdots \psi_{\ell_5}, \quad \text{for} \quad k \ge 0. \tag{A.6.96}$$

While this recursive formula is pretty, it is desirable to have an explicit power series formula. Indeed, one has the following.

**Proposition A.6.2.** *In the setting of* (A.6.87),

$$\Phi^{-1}(a) = \sum_{j=0}^{\infty} \binom{5j}{j} \frac{a^{4j+1}}{4j+1}, \quad \textit{for} \quad |a| < \frac{4}{5} 5^{-1/4}. \tag{A.6.97}$$

**Proof.** By Proposition 2.1.6, we have from (A.6.83)–(A.6.87) that

$$\Phi^{-1}(a) = \frac{1}{2\pi i} \int_{\partial\mathcal{O}} \frac{z\Phi'(z)}{\Phi(z) - a}\, dz, \tag{A.6.98}$$

with $\mathcal{O} = D_{5^{-1/4}}(0)$, $|a| < (4/5)5^{-1/4}$. We then have the convergent power series

$$\begin{aligned} \frac{1}{\Phi(z) - a} &= \frac{1}{\Phi(z)} \frac{1}{1 - a/\Phi(z)} \\ &= \sum_{k \ge 0} \frac{a^k}{\Phi(z)^{k+1}}, \end{aligned} \tag{A.6.99}$$

given $z \in \partial\mathcal{O}$, $|a| < (4/5)5^{-1/4}$. Hence

$$\Phi^{-1}(a) = \frac{1}{2\pi i} \sum_{k \ge 0} \Bigl( \int_{\partial\mathcal{O}} \frac{z\Phi'(z)}{\Phi(z)^{k+1}}\, dz \Bigr) a^k. \tag{A.6.100}$$

Since $\Phi'(z) = 1 - 5z^4$ and $\Phi(z) = z(1 - z^4)$, the coefficient of $a^k$ is

$$\frac{1}{2\pi i} \int_{\partial\mathcal{O}} \frac{1 - 5z^4}{(1 - z^4)^{k+1}} \frac{dz}{z^k}, \tag{A.6.101}$$

or equivalently, the coefficient of $a^k$ is equal to the coefficient of $z^{k-1}$ in the power series expansion of $(1 - 5z^4)/(1 - z^4)^{k+1}$. This requires $k = 4j + 1$ for some $j \in \mathbb{Z}^+$, and we then seek

$$\text{the coefficient of } \zeta^j \text{ in } \frac{1 - 5\zeta}{(1 - \zeta)^{k+1}}, \quad k = 4j + 1. \tag{A.6.102}$$

We have

$$(1 - \zeta)^{-(k+1)} = \sum_{j=0}^{\infty} \binom{k+j}{j} \zeta^j, \tag{A.6.103}$$

and hence

$$\begin{aligned}-5\zeta(1-\zeta)^{-(k+1)} &= -5\sum_{\ell=0}^{\infty}\binom{k+\ell}{\ell}\zeta^{\ell+1} \\ &= -5\sum_{j=1}^{\infty}\binom{k+j-1}{j-1}\zeta^{j}.\end{aligned} \tag{A.6.104}$$

Thus the coefficient specified in (A.6.102) is 1 for $j = 0$, and, for $j \geq 1$, it is

$$\begin{aligned}&\binom{k+j}{j} - 5\binom{k+j-1}{j-1}, \quad \text{with} \quad k = 4j+1, \\ &= \binom{5j+1}{j} - 5\binom{5j}{j-1} \\ &= \frac{1}{4j+1}\binom{5j}{j},\end{aligned} \tag{A.6.105}$$

giving (A.6.97). □

We next discuss some global aspects of the map $\Phi : \mathbb{C} \to \mathbb{C}$, making use of results on covering maps from §5.6. To state the result, let us take $\mathcal{C} = \{\pm 5^{-1/4}, \pm i5^{-1/4}\}$, as in (A.6.80), and set

$$\mathcal{V} = \Phi(\mathcal{C}) = \frac{4}{5}\mathcal{C}, \quad \widetilde{\mathcal{C}} = \Phi^{-1}(\mathcal{V}). \tag{A.6.106}$$

**Lemma A.6.3.** *The map*

$$\Phi : \mathbb{C} \setminus \widetilde{\mathcal{C}} \longrightarrow \mathbb{C} \setminus \mathcal{V} \tag{A.6.107}$$

*is a 5-fold covering map.*

Note that (A.6.85) implies

$$D_{(4/5)5^{-1/4}}(0) \subset \mathbb{C} \setminus \mathcal{V}. \tag{A.6.108}$$

The following is a consequence of Proposition 5.6.5.

**Proposition A.6.4.** *Assume $\Omega$ is an open, connected, simply connected set satisfying*

$$\Omega \subset \mathbb{C} \setminus \mathcal{V}, \quad \Omega \supset D_{(4/5)5^{-1/4}}(0). \tag{A.6.109}$$

*Then $\Phi^{-1}$ in (A.6.87) has a unique extension to a holomorphic map*

$$\Phi^{-1} : \Omega \longrightarrow \mathbb{C} \setminus \widetilde{\mathcal{C}}. \tag{A.6.110}$$

An alternative path from Proposition A.6.2 to Proposition A.6.4 is provided by recognizing the power series (A.6.97) as representing $Q^{-1}(a)$ in terms of a generalized hypergeometric function, namely

$$\Phi^{-1}(a) = a\,{}_4F_3\Big(\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}; \frac{1}{2}, \frac{3}{4}, \frac{5}{4}; 5\Big(\frac{5a}{4}\Big)^4\Big). \tag{A.6.111}$$

See §7.2 for a discussion of hypergeometric functions, their differential equations and analytic continuations, and a proof of (A.6.111).

The analysis of the Bring radical was carried further by C. Hermite, who produced a formula for $\Phi^{-1}$ in terms of elliptic functions and their associated theta functions. This was also pursued by L. Kronecker and F. Klein. For more on this, we refer to [**22**] and to Chapter 5 of [**29**]. Work on the application of theta functions to higher degree polynomials is given in [**51**].

# Bibliography


[1] Mark J. Ablowitz and Athanassios S. Fokas, *Complex variables: introduction and applications*, 2nd ed., Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 2003. MR1989049

[2] Lars V. Ahlfors, *Complex analysis: An introduction of the theory of analytic functions of one complex variable*, Second edition, McGraw-Hill Book Co., New York-Toronto-London, 1966. MR0188405

[3] R. Andriambololoma, H. Rakotoson, R. Gilbert, and R. Tokiniaina, *Method to reduce quintic equations using linear algebra*, HEP-MAD 09, Aniananarivo (Madagascar), 21-28th Aug., 2009.

[4] Tom M. Apostol, *Introduction to analytic number theory*, Undergraduate Texts in Mathematics, Springer-Verlag, New York-Heidelberg, 1976. MR0434929

[5] W. N. Bailey, *Generalized hypergeometric series*, Cambridge Tracts in Mathematics and Mathematical Physics, No. 32, Stechert-Hafner, Inc., New York, 1964. MR0185155

[6] Joseph Bak and Donald J. Newman, *Complex analysis*, Undergraduate Texts in Mathematics, Springer-Verlag, New York-Berlin, 1982. MR671250

[7] Jonathan M. Borwein and Peter B. Borwein, *Pi and the AGM*, Canadian Mathematical Society Series of Monographs and Advanced Texts, John Wiley & Sons, Inc., New York, 1987. A study in analytic number theory and computational complexity; A Wiley-Interscience Publication. MR877728

[8] Lennart Carleson and Theodore W. Gamelin, *Complex dynamics*, Universitext: Tracts in Mathematics, Springer-Verlag, New York, 1993. MR1230383

[9] C. Herbert Clemens, *A scrapbook of complex curve theory*, The University Series in Mathematics, Plenum Press, New York-London, 1980. MR614289

[10] John B. Conway, *Functions of one complex variable*, 2nd ed., Graduate Texts in Mathematics, vol. 11, Springer-Verlag, New York-Berlin, 1978. MR503901

[11] William F. Donoghue Jr., *Distributions and Fourier transforms*, Pure and Applied Mathematics, vol. 32, Academic Press, New York, 1969. MR3363413

[12] H. M. Edwards, *Riemann's zeta function*, Dover Publications, Inc., Mineola, NY, 2001. Reprint of the 1974 original [Academic Press, New York; MR0466039 (57 #5922)]. MR1854455

[13] Hershel M. Farkas and Irwin Kra, *Riemann surfaces*, Graduate Texts in Mathematics, vol. 71, Springer-Verlag, New York-Berlin, 1980. MR583745

[14] Otto Forster, *Lectures on Riemann surfaces*, Graduate Texts in Mathematics, vol. 81, Springer-Verlag, New York-Berlin, 1981. Translated from the German by Bruce Gilligan. MR648106

[15] Theodore W. Gamelin, *Complex analysis*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 2001. MR1830078

[16] Theodore W. Gamelin, *Uniform algebras*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1969. MR0410387

[17] Marvin J. Greenberg, *Lectures on algebraic topology*, W. A. Benjamin, Inc., New York-Amsterdam, 1967. MR0215295

[18] Robert E. Greene and Steven G. Krantz, *Function theory of one complex variable*, 3rd ed., Graduate Studies in Mathematics, vol. 40, American Mathematical Society, Providence, RI, 2006. MR2215872

[19] E. Hille, *Analytic Function Theory, Vols. 1–2*, Chelsea, New York, 1962.

[20] E. Hille, *Ordinary Differential Equations in the Complex Plane*, Wiley-Interscience, New York, 1976.

[21] R. Bruce King, *Beyond the quartic equation*, Birkhäuser Boston, Inc., Boston, MA, 1996. MR1401346

[22] Felix Klein, *Lectures on the icosahedron and the solution of equations of the fifth degree*, Second and revised edition, Dover Publications, Inc., New York, N.Y., 1956. Translated into English by George Gavin Morrice. MR0080930

[23] Steven G. Krantz, *Complex analysis: the geometric viewpoint*, Carus Mathematical Monographs, vol. 23, Mathematical Association of America, Washington, DC, 1990. MR1074176

[24] Steven G. Krantz, *Function theory of several complex variables*, Pure and Applied Mathematics; A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York, 1982. MR635928

[25] Serge Lang, *Algebra*, Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965. MR0197234

[26] Derek F. Lawden, *Elliptic functions and applications*, Applied Mathematical Sciences, vol. 80, Springer-Verlag, New York, 1989. MR1007595

[27] N. N. Lebedev, *Special functions and their applications*, Dover Publications, Inc., New York, 1972. Revised edition, translated from the Russian and edited by Richard A. Silverman; Unabridged and corrected republication. MR0350075

[28] Rafe Mazzeo and Michael Taylor, *Curvature and uniformization*, Israel J. Math. **130** (2002), 323–346, DOI 10.1007/BF02764082. MR1919383

[29] Henry McKean and Victor Moll, *Elliptic curves*: *Function theory, geometry, arithmetic*, Cambridge University Press, Cambridge, 1997. MR1471703

[30] John Milnor, *Dynamics in one complex variable*, Introductory lectures, Friedr. Vieweg & Sohn, Braunschweig, 1999. MR1721240

[31] James R. Munkres, *Topology: a first course*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975. MR0464128

[32] Zeev Nehari, *Conformal mapping*, Dover Publications, Inc., New York, 1975. Reprinting of the 1952 edition. MR0377031

[33] D. J. Newman, *Simple analytic proof of the prime number theorem*, Amer. Math. Monthly **87** (1980), no. 9, 693–696, DOI 10.2307/2321853. MR602825

[34] Ivan Niven, *A simple proof that $\pi$ is irrational*, Bull. Amer. Math. Soc. **53** (1947), 509, DOI 10.1090/S0002-9904-1947-08821-2. MR0021013

[35] F. W. J. Olver, *Asymptotics and special functions*, Computer Science and Applied Mathematics, Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1974. MR0435697

[36] S. J. Patterson, *Eisenstein and the quintic equation* (English, with French and German summaries), Historia Math. **17** (1990), no. 2, 132–140, DOI 10.1016/0315-0860(90)90049-J. MR1057021

[37] Walter Rudin, *Real and complex analysis*, McGraw-Hill Book Co., New York-Toronto, Ont.-London, 1966. MR0210528

[38] Donald Sarason, *Complex function theory*, 2nd ed., American Mathematical Society, Providence, RI, 2007. MR2368077

[39] Joel L. Schiff, *Normal families*, Universitext, Springer-Verlag, New York, 1993. MR1211641

[40] Barry Simon, *Basic complex analysis*, A Comprehensive Course in Analysis, Part 2A, American Mathematical Society, Providence, RI, 2015. MR3443339

[41] Barry Simon, *Advanced complex analysis*, A Comprehensive Course in Analysis, Part 2B, American Mathematical Society, Providence, RI, 2015. MR3364090

[42] Michael Spivak, *A comprehensive introduction to differential geometry. Vol. V*, Publish or Perish, Inc., Boston, Mass., 1975. MR0394453

[43] Joseph L. Taylor, *Complex variables*, Pure and Applied Undergraduate Texts, vol. 16, American Mathematical Society, Providence, RI, 2011. MR2839384

[44] M. Taylor, *Introduction to Analysis in One Variable*, Preprint, 2019.

[45] M. Taylor, *Introduction to Analysis in Several Variables*, Preprint, 2019.

[46] Michael E. Taylor, *Partial differential equations*, Basic theory, Texts in Applied Mathematics, vol. 23, Springer-Verlag, New York, 1996. MR1395147

[47] Michael E. Taylor, *Measure theory and integration*, Graduate Studies in Mathematics, vol. 76, American Mathematical Society, Providence, RI, 2006. MR2245472

[48] M. Taylor, *Introduction to Differential Equations*, American Math. Soc., Providence RI, 2011.

[49] M. Taylor, *Linear Algebra*, Preprint, 2019.

[50] M. Tsuji, *Potential theory in modern function theory*, Chelsea Publishing Co., New York, 1975. Reprinting of the 1959 original. MR0414898

[51] H. Umemura, Resolution of algebraic equations by theta constants, pp. 261–270 in *Tata Lectures on Theta II* (D. Mumford), Birkhauser, Boston, 2007.

[52] G. N. Watson, *A treatise on the theory of Bessel functions*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1995. Reprint of the second (1944) edition. MR1349110

[53] E. Whittaker and G. Watson, *Modern Analysis* (4th ed.), Cambridge Univ. Press, Cambridge UK, 1927.

[54] Norbert Wiener, *Tauberian theorems*, Ann. of Math. (2) **33** (1932), no. 1, 1–100, DOI 10.2307/1968102. MR1503035

# Index

SELECTED PUBLISHED TITLES IN THIS SERIES

202 **Michael E. Taylor,** Introduction to Complex Analysis, 2019
199 **Weinan E, Tiejun Li, and Eric Vanden-Eijnden,** Applied Stochastic Analysis, 2019
198 **Robert L. Benedetto,** Dynamics in One Non-Archimedean Variable, 2019
197 **Walter Craig,** A Course on Partial Differential Equations, 2018
196 **Martin Stynes and David Stynes,** Convection-Diffusion Problems, 2018
195 **Matthias Beck and Raman Sanyal,** Combinatorial Reciprocity Theorems, 2018
194 **Seth Sullivant,** Algebraic Statistics, 2018
193 **Martin Lorenz,** A Tour of Representation Theory, 2018
192 **Tai-Peng Tsai,** Lectures on Navier-Stokes Equations, 2018
191 **Theo Bühler and Dietmar A. Salamon,** Functional Analysis, 2018
190 **Xiang-dong Hou,** Lectures on Finite Fields, 2018
189 **I. Martin Isaacs,** Characters of Solvable Groups, 2018
188 **Steven Dale Cutkosky,** Introduction to Algebraic Geometry, 2018
187 **John Douglas Moore,** Introduction to Global Analysis, 2017
186 **Bjorn Poonen,** Rational Points on Varieties, 2017
185 **Douglas J. LaFountain and William W. Menasco,** Braid Foliations in Low-Dimensional Topology, 2017
184 **Harm Derksen and Jerzy Weyman,** An Introduction to Quiver Representations, 2017
183 **Timothy J. Ford,** Separable Algebras, 2017
182 **Guido Schneider and Hannes Uecker,** Nonlinear PDEs, 2017
181 **Giovanni Leoni,** A First Course in Sobolev Spaces, Second Edition, 2017
180 **Joseph J. Rotman,** Advanced Modern Algebra: Third Edition, Part 2, 2017
179 **Henri Cohen and Fredrik Strömberg,** Modular Forms, 2017
178 **Jeanne N. Clelland,** From Frenet to Cartan: The Method of Moving Frames, 2017
177 **Jacques Sauloy,** Differential Galois Theory through Riemann-Hilbert Correspondence, 2016
176 **Adam Clay and Dale Rolfsen,** Ordered Groups and Topology, 2016
175 **Thomas A. Ivey and Joseph M. Landsberg,** Cartan for Beginners: Differential Geometry via Moving Frames and Exterior Differential Systems, Second Edition, 2016
174 **Alexander Kirillov Jr.,** Quiver Representations and Quiver Varieties, 2016
173 **Lan Wen,** Differentiable Dynamical Systems, 2016
172 **Jinho Baik, Percy Deift, and Toufic Suidan,** Combinatorics and Random Matrix Theory, 2016
171 **Qing Han,** Nonlinear Elliptic Equations of the Second Order, 2016
170 **Donald Yau,** Colored Operads, 2016
169 **András Vasy,** Partial Differential Equations, 2015
168 **Michael Aizenman and Simone Warzel,** Random Operators, 2015
167 **John C. Neu,** Singular Perturbation in the Physical Sciences, 2015
166 **Alberto Torchinsky,** Problems in Real and Functional Analysis, 2015
165 **Joseph J. Rotman,** Advanced Modern Algebra: Third Edition, Part 1, 2015
164 **Terence Tao,** Expansion in Finite Simple Groups of Lie Type, 2015
163 **Gérald Tenenbaum,** Introduction to Analytic and Probabilistic Number Theory, Third Edition, 2015
162 **Firas Rassoul-Agha and Timo Seppäläinen,** A Course on Large Deviations with an Introduction to Gibbs Measures, 2015

For a complete list of titles in this series, visit the
AMS Bookstore at **www.ams.org/bookstore/gsmseries/**.

In this text, the reader will learn that all the basic functions that arise in calculus—such as powers and fractional powers, exponentials and logs, trigonometric functions and their inverses, as well as many new functions that the reader will meet—are naturally defined for complex arguments. Furthermore, this expanded setting leads to a much richer understanding of such functions than one could glean by merely considering them in the real domain. For example, understanding the exponential function in the complex domain via its differential equation provides a clean path to Euler's formula and hence to a self-contained treatment of the trigonometric functions. Complex analysis, developed in partnership with Fourier analysis, differential equations, and geometrical techniques, leads to the development of a cornucopia of functions of use in number theory, wave motion, conformal mapping, and other mathematical phenomena, which the reader can learn about from material presented here.

This book could serve for either a one-semester course or a two-semester course in complex analysis for beginning graduate students or for well-prepared undergraduates whose background includes multivariable calculus, linear algebra, and advanced calculus.

ISBN 978-1-4704-5286-5

9 781470 452865

GSM/202

For additional information and updates on this book, visit

**www.ams.org/bookpages/gsm-202**

AMS
www.ams.org