Anthony W. Knapp

Basic Real Analysis

Along with a companion volume
Advanced Real Analysis

Selected Pages from Chapter I: pp. 1–26, 53–81

Birkhäuser
Boston • Basel • Berlin

Anthony W. Knapp
81 Upper Sheep Pasture Road
East Setauket, NY 11733-1729
U.S.A.
e-mail to: aknapp@math.sunysb.edu
http://www.math.sunysb.edu/~ aknapp/books/basic.html

Cover design by Mary Burgess.

Mathematics Subject Classicification (2000): 28-01, 26-01, 42-01, 54-01, 34-01

Library of Congress Cataloging-in-Publication Data
Knapp, Anthony W.
Basic real analysis: along with a companion volume Advanced real analysis / Anthony W. Knapp
p. cm. – (Cornerstones)
Includes bibliographical references and index.
ISBN 0-8176-3250-6 (alk. paper)
1. Mathematical analysis. I. Title. II. Cornerstones (Birkhäuser)

QA300.K56 2005
515–dc22 2005048070

ISBN-10 0-8176-3250-6 eISBN 0-8176-4441-5 Printed on acid-free paper.
ISBN-13 978-0-8176-3250-2

Advanced Real Analysis ISBN 0-8176-4382-6
Basic Real Analysis and *Advanced Real Analysis* (Set) ISBN 0-8176-4407-5

Printed in the United States of America. (MP)

9 8 7 6 5 4 3 2 1 SPIN 10934074

Birkhäuser is a part of *Springer Science+Business Media*

www.birkhauser.com

CHAPTER I

Theory of Calculus in One Real Variable

Abstract. This chapter, beginning with Section 2, develops the topic of sequences and series of functions, especially of functions of one variable. An important part of the treatment is an introduction to the problem of interchange of limits, both theoretically and practically. This problem plays a role repeatedly in real analysis, but its visibility decreases as more and more results are developed for handling it in various situations. Fourier series are introduced in this chapter and are carried along throughout the book as a motivating example for a number of problems in real analysis.

Section 1 makes contact with the core of a first undergraduate course in real-variable theory. Some material from such a course is repeated here in order to establish notation and a point of view. Omitted material is summarized at the end of the section, and some of it is discussed in a little more detail in an appendix at the end of the book. The point of view being established is the use of defining properties of the real number system to prove the Bolzano–Weierstrass Theorem, followed by the use of that theorem to prove some of the difficult theorems that are usually assumed in a one-variable calculus course. The treatment makes use of the extended real-number system, in order to allow sup and inf to be defined for any nonempty set of reals and to allow lim sup and lim inf to be meaningful for any sequence.

Sections 2–3 introduce the problem of interchange of limits. They show how certain concrete problems can be viewed in this way, and they give a way of thinking about all such interchanges in a common framework. A positive result affirms such an interchange under suitable hypotheses of monotonicity. This is by way of introduction to the topic in Section 3 of uniform convergence and the role of uniform convergence in continuity and differentiation.

Section 4 gives a careful development of the Riemann integral for real-valued functions of one variable, establishing existence of Riemann integrals for bounded functions that are discontinuous at only finitely many points, basic properties of the integral, the Fundamental Theorem of Calculus for continuous integrands, the change-of-variables formula, and other results. Section 5 examines complex-valued functions, pointing out the extent to which the results for real-valued functions in the first four sections extend to complex-valued functions.

Section 6 is a short treatment of the version of Taylor's Theorem in which the remainder is given by an integral. Section 7 takes up power series and uses them to define the elementary transcendental functions and establish their properties. The power series expansion of $(1+x)^p$ for arbitrary complex p is studied carefully. Section 8 introduces Cesàro and Abel summability, which play a role in the subject of Fourier series. A converse theorem to Abel's theorem is used to exhibit the function $|x|$ as the uniform limit of polynomials on $[-1, 1]$. The Weierstrass Approximation Theorem of Section 9 generalizes this example and establishes that every continuous complex-valued function on a closed bounded interval is the uniform limit of polynomials.

Section 10 introduces Fourier series in one variable in the context of the Riemann integral. The main theorems of the section are a convergence result for continuously differentiable functions, Bessel's inequality, the Riemann–Lebesgue Lemma, Fejér's Theorem, and Parseval's Theorem.

1. Review of Real Numbers, Sequences, Continuity

This section reviews some material that is normally in an undergraduate course in real analysis. The emphasis will be on a rigorous proof of the Bolzano–Weierstrass Theorem and its use to prove some of the difficult theorems that are usually assumed in a one-variable calculus course. We shall skip over some easier aspects of an undergraduate course in real analysis that fit logically at the end of this section. A list of such topics appears at the end of the section.

The system of real numbers $\mathbb{R}$ may be constructed out of the system of rational numbers $\mathbb{Q}$, and we take this construction as known. The formal definition is that a real number is a **cut** of rational numbers, i.e., a subset of rational numbers that is neither $\mathbb{Q}$ nor the empty set, has no largest element, and contains all rational numbers less than any rational that it contains. The idea of the construction is as follows: Each rational number q determines a cut q^*, namely the set of all rationals less than q. Under the identification of $\mathbb{Q}$ with a subset of $\mathbb{R}$, the cut defining a real number consists of all rational numbers less than the given real number.

The set of cuts gets a natural ordering, given by inclusion. In place of $\subseteq$, we write $\leq$. For any two cuts r and s, we have $r \leq s$ or $s \leq r$, and if both occur, then $r = s$. We can then define $<$, $\geq$, and $>$ in the expected way. The positive cuts r are those with $0^* < r$, and the negative cuts are those with $r < 0^*$.

Once cuts and their ordering are in place, one can go about defining the usual operations of arithmetic and proving that $\mathbb{R}$ with these operations satisfies the familiar associative, commutative, and distributive laws, and that these interact with inequalities in the usual ways. The definitions of addition and subtraction are easy: the sum or difference of two cuts is simply the set of sums or differences of the rationals from the respective cuts. For multiplication and reciprocals one has to take signs into account. For example, the product of two positive cuts consists of all products of positive rationals from the two cuts, as well as 0 and all negative rationals. After these definitions and the proofs of the usual arithmetic operations are complete, it is customary to write 0 and 1 in place of 0^* and 1^*.

An **upper bound** for a nonempty subset E of $\mathbb{R}$ is a real number M such that $x \leq M$ for all x in E. If the nonempty set E has an upper bound, we can take the cuts that E consists of and form their union. This turns out to be a cut, it is an upper bound for E, and it is $\leq$ all upper bounds for E. We can summarize this result as a theorem.

Theorem 1.1. Any nonempty subset E of $\mathbb{R}$ with an upper bound has a least upper bound.

The least upper bound is necessarily unique, and the notation for it is $\sup_{x \in E} x$ or $\sup \{x \mid x \in E\}$, "sup" being an abbreviation for the Latin word "supremum," the largest. Of course, the least upper bound for a set E with an upper bound need not be in E; for example, the supremum of the negative rationals is 0, which is not negative.

A **lower bound** for a nonempty set E of $\mathbb{R}$ is a real number m such that $x \geq m$ for all $x \in E$. If m is a lower bound for E, then $-m$ is an upper bound for the set $-E$ of negatives of members of E. Thus $-E$ has an upper bound, and Theorem 1.1 shows that it has a least upper bound $\sup_{x \in -E} x$. Then $-x$ is a greatest lower bound for E. This greatest lower bound is denoted by $\inf_{y \in E} y$ or $\inf\{y \mid y \in E\}$, "inf" being an abbreviation for "infimum." We can summarize as follows.

Corollary 1.2. Any nonempty subset E of $\mathbb{R}$ with a lower bound has a greatest lower bound.

A subset of $\mathbb{R}$ is said to be **bounded** if it has an upper bound and a lower bound. Let us introduce notation and terminology for **intervals** of $\mathbb{R}$, first treating the bounded ones.[1] Let a and b be real numbers with $a \leq b$. The **open interval** from a to b is the set $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$, the **closed interval** is the set $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$, and the **half-open intervals** are the sets $[a, b) = \{x \in \mathbb{R} \mid a \leq x < b\}$ and $(a, b] = \{x \in \mathbb{R} \mid a < x \leq b\}$. Each of the above intervals is indeed bounded, having a as a lower bound and b as an upper bound. These intervals are nonempty when $a < b$ or when the interval is $[a, b]$ with $a = b$, and in these cases the least upper bound is b and the greatest lower bound is a.

Open sets in $\mathbb{R}$ are defined to be arbitrary unions of open bounded intervals, and a **closed set** is any set whose complement in $\mathbb{R}$ is open. A set E is open if and only if for each $x \in E$, there is an open interval (a, b) such that $x \in (a, b) \subseteq E$. In this case we of course have $a < x < b$. If we put $\epsilon = \min\{x - a, b - x\}$, then we see that x lies in the subset $(x - \epsilon, x + \epsilon)$ of (a, b). The open interval $(x - \epsilon, x + \epsilon)$ equals $\{y \in \mathbb{R} \bigm| |y - x| < \epsilon\}$. Thus an open set in $\mathbb{R}$ is any set E such that for each $x \in E$, there is a number $\epsilon > 0$ such that $\{y \in \mathbb{R} \bigm| |y - x| < \epsilon\}$ lies in E. A **limit point** x of a subset F of $\mathbb{R}$ is a point of $\mathbb{R}$ such that any open interval containing x meets F in a point other than x. For example, the set $[a, b) \cup \{b + 1\}$ has $[a, b]$ as its set of limit points. A subset of $\mathbb{R}$ is closed if and only if it contains all its limit points.

Now let us turn to unbounded intervals. To provide notation for these, we shall make use of two symbols $+\infty$ and $-\infty$ that will shortly be defined to be "extended real numbers." If a is in $\mathbb{R}$, then the subsets $(a, +\infty) = \{x \in \mathbb{R} \mid a < x\}$, $(-\infty, a) = \{x \in \mathbb{R} \mid x < a\}$, $(-\infty, +\infty) = \mathbb{R}$, $[a, +\infty) = \{x \in \mathbb{R} \mid a \leq x\}$, and $(-\infty, a] = \{x \in \mathbb{R} \mid x \leq a\}$ are defined to be **intervals**, and they are all unbounded. The first three are open sets of $\mathbb{R}$ and are considered to be open

[1] Bounded intervals are called "finite intervals" by some authors.

intervals, while the last three are closed sets and are considered to be closed intervals. Specifically the middle set $\mathbb{R}$ is both open and closed.

One important consequence of Theorem 1.1 is the **archimedean property** of $\mathbb{R}$, as follows.

Corollary 1.3. If a and b are real numbers with $a > 0$, then there exists an integer n with $na > b$.

PROOF. If, on the contrary, $na \le b$ for all integers n, then b is an upper bound for the set of all na. Let M be the least upper bound of the set $\{na \mid n \text{ is an integer}\}$. Using that a is positive, we find that $a^{-1}M$ is a least upper bound for the integers. Thus $n \le a^{-1}M$ for all integers n, and there is no smaller upper bound. However, the smaller number $a^{-1}M - 1$ must be an upper bound, since saying $n \le a^{-1}M$ for all integers is the same as saying $n - 1 \le a^{-1}M - 1$ for all integers. We arrive at a contradiction, and we conclude that there is some integer n with $na > b$.

The archimedean property enables one to see, for example, that any two distinct real numbers have a rational number lying between them. We prove this consequence as Corollary 1.5 after isolating one step as Corollary 1.4.

Corollary 1.4. If c is a real number, then there exists an integer n such that $n \le c < n + 1$.

PROOF. Corollary 1.3 with $a = 1$ and $b = c$ shows that there is an integer M with $M > c$, and Corollary 1.3 with $a = 1$ and $b = -c$ shows that there is an integer m with $m > -c$. Then $-m < c < M$, and it follows that there exists a greatest integer n with $n \le c$. This n must have the property that $c < n + 1$, and the corollary follows.

Corollary 1.5. If x and y are real numbers with $x < y$, then there exists a rational number r with $x < r < y$.

PROOF. By Corollary 1.3 with $a = y - x$ and $b = 1$, there is an integer N such that $N(y - x) > 1$. This integer N has to be positive. Then $\frac{1}{N} < y - x$. By Corollary 1.4 with $c = Nx$, there exists an integer n with $n \le Nx < n + 1$, hence with $\frac{n}{N} \le x < \frac{n+1}{N}$. Adding the inequalities $\frac{n}{N} \le x$ and $\frac{1}{N} < y - x$ yields $\frac{n+1}{N} < y$. Thus $x \le \frac{n}{N} < \frac{n+1}{N} < y$. Since $\frac{n}{N} < \frac{2n+1}{2N} < \frac{n+1}{N}$, the rational number $r = \frac{2n+1}{2N}$ has the required properties.

A **sequence** in a set S is a function from a certain kind of subset of integers into S. It will be assumed that the set of integers is nonempty, consists of consecutive integers, and contains no largest integer. In particular the domain of any sequence is infinite. Usually the set of integers is either all nonnegative integers or all positive integers. Sometimes the set of integers is all integers, and the sequence in this case is often called "doubly infinite." The value of a sequence f at the integer n is normally written f_n rather than $f(n)$, and the sequence itself may be denoted by an expression like $\{f_n\}_{n\geq 1}$, in which the outer subscript indicates the domain.

A **subsequence** of a sequence f with domain $\{m, m+1, \dots\}$ is a composition $f \circ n$, where f is a sequence and n is a sequence in the domain of f such that $n_k < n_{k+1}$ for all k. For example, if $\{a_n\}_{n\geq 1}$ is a sequence, then $\{a_{2k}\}_{k\geq 1}$ is the subsequence in which the function n is given by $n_k = 2k$. The domain of a subsequence, by our definition, is always infinite.

A sequence a_n in $\mathbb{R}$ is **convergent**, or convergent in $\mathbb{R}$, if there exists a real number a such that for each $\epsilon > 0$, there is an integer N with $|a_n - a| < \epsilon$ for all $n \geq N$. The number a is necessarily unique and is called the **limit** of the sequence. Depending on how much information about the sequence is unambiguous, we may write $\lim_{n\to\infty} a_n = a$ or $\lim_n a_n = a$ or $\lim a_n = a$ or $a_n \to a$. We also say a_n **tends** to a as n tends to **infinity** or ∞.

A sequence in $\mathbb{R}$ is called **monotone increasing** if $a_n \leq a_{n+1}$ for all n in the domain, **monotone decreasing** if $a_n \geq a_{n+1}$ for all n in the domain, **monotone** if it is monotone increasing or monotone decreasing.

Corollary 1.6. Any bounded monotone sequence in $\mathbb{R}$ converges. If the sequence is monotone increasing, then the limit is the least upper bound of the image in $\mathbb{R}$ of the sequence. If the sequence is monotone decreasing, the limit is the greatest lower bound of the image.

REMARK. Often it is Corollary 1.6, rather than the existence of least upper bounds, that is taken for granted in an elementary calculus course. The reason is that the statement of Corollary 1.6 tends for calculus students to be easier to understand than the statement of the least upper bound property. Problem 1 at the end of the chapter asks for a derivation of the least-upper-bound property from Corollary 1.6.

PROOF. Suppose that $\{a_n\}$ is monotone increasing and bounded. Let $a = \sup_n a_n$, the existence of the supremum being ensured by Theorem 1.1, and let $\epsilon > 0$ be given. If there were no integer N with $a_N > a - \epsilon$, then $a - \epsilon$ would be a smaller upper bound, contradiction. Thus such an N exists. For that N, $n \geq N$ implies $a - \epsilon < a_N \leq a_n \leq a < a + \epsilon$. Thus $n \geq N$ implies $|a_n - a| < \epsilon$. Since ϵ is arbitrary, $\lim_{n\to\infty} a_n = a$. If the given sequence $\{a_n\}$ is monotone decreasing, we argue similarly with $a = \inf_n a_n$.

In working with sup and inf, it will be quite convenient to use the notation $\sup_{x\in E} x$ even when E is nonempty but not bounded above, and to use the notation

$\inf_{x\in E} x$ even when E is nonempty but not bounded below. We introduce symbols $+\infty$ and $-\infty$, plus and minus **infinity**, for this purpose and extend the definitions of $\sup_{x\in E} x$ and $\inf_{x\in E} x$ to all nonempty subsets E of $\mathbb{R}$ by taking

$$\sup_{x\in E} x = +\infty \qquad \text{if } E \text{ has no upper bound},$$
$$\inf_{x\in E} x = -\infty \qquad \text{if } E \text{ has no lower bound}.$$

To work effectively with these new pieces of notation, we shall enlarge $\mathbb{R}$ to a set $\mathbb{R}^*$ called the **extended real numbers** by defining

$$\mathbb{R}^* = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}.$$

An ordering on $\mathbb{R}^*$ is defined by taking $-\infty < r < +\infty$ for every member r of $\mathbb{R}$ and by retaining the usual ordering within $\mathbb{R}$. It is immediate from this definition that

$$\inf_{x\in E} x \le \sup_{x\in E} x$$

if E is any nonempty subset of $\mathbb{R}$. In fact, we can enlarge the definitions of $\inf_{x\in E} x$ and $\sup_{x\in E} x$ in obvious fashion to include the case that E is any nonempty subset of $\mathbb{R}^*$, and we still have inf $\le$ sup. With the ordering in place, we can unambiguously speak of **open intervals** (a, b), **closed intervals** $[a, b]$, and **half-open intervals** $[a, b)$ and $(a, b]$ in $\mathbb{R}^*$ even if a or b is infinite. Under our definitions the intervals of $\mathbb{R}$ are the intervals of $\mathbb{R}^*$ that are subsets of $\mathbb{R}$, even if a or b is infinite. If no special mention is made whether an interval lies in $\mathbb{R}$ or $\mathbb{R}^*$, it is usually assumed to lie in $\mathbb{R}$.

The next step is to extend the operations of arithmetic to $\mathbb{R}^*$. It is important not to try to make such operations be everywhere defined, lest the distributive laws fail. Letting r denote any member of $\mathbb{R}$ and a and b be any members of $\mathbb{R}^*$, we make the following new definitions:

Multiplication:

$$r(+\infty) = (+\infty)r = \begin{cases} +\infty & \text{if } r > 0, \\ 0 & \text{if } r = 0, \\ -\infty & \text{if } r < 0, \end{cases}$$

$$r(-\infty) = (-\infty)r = \begin{cases} -\infty & \text{if } r > 0, \\ 0 & \text{if } r = 0, \\ +\infty & \text{if } r < 0, \end{cases}$$

$$(+\infty)(+\infty) = (-\infty)(-\infty) = +\infty,$$
$$(+\infty)(-\infty) = (-\infty)(+\infty) = -\infty.$$

Addition:
$$r + (+\infty) = (+\infty) + r = +\infty,$$
$$r + (-\infty) = (-\infty) + r = -\infty,$$
$$(+\infty) + (+\infty) = +\infty,$$
$$(-\infty) + (-\infty) = -\infty.$$

Subtraction: $a - b = a + (-b)$ whenever the right side is defined.

Division: $a/b = 0$ if $a \in \mathbb{R}$ and b is $\pm\infty$,

$a/b = b^{-1}a$ if $b \in \mathbb{R}$ with $b \neq 0$ and a is $\pm\infty$.

The only surprise in the list is that 0 times anything is 0. This definition will be important to us when we get to measure theory, starting in Chapter V.

It is now a simple matter to define convergence of a sequence in $\mathbb{R}^*$. The cases that need addressing are that the sequence is in $\mathbb{R}$ and that the limit is $+\infty$ or $-\infty$. We say that a sequence $\{a_n\}$ in $\mathbb{R}$ tends to $+\infty$ if for any positive number M, there exists an integer N such that $a_n \geq M$ for all $n \geq N$. The sequence tends to $-\infty$ if for any negative number $-M$, there exists an integer N such that $a_n \leq -M$ for all $n \geq N$. It is important to indicate whether convergence/divergence of a sequence is being discussed in $\mathbb{R}$ or in $\mathbb{R}^*$. The default setting is $\mathbb{R}$, in keeping with standard terminology in calculus. Thus, for example, we say that the sequence $\{n\}_{n\geq 1}$ diverges, but it converges in $\mathbb{R}^*$ (to $+\infty$).

With our new definitions every monotone sequence converges in $\mathbb{R}^*$.

For a sequence $\{a_n\}$ in $\mathbb{R}$ or even in $\mathbb{R}^*$, we now introduce members $\limsup_n a_n$ and $\liminf_n a_n$ of $\mathbb{R}^*$. These will always be defined, and thus we can apply the operations lim sup and lim inf to any sequence in $\mathbb{R}^*$. For the case of lim sup we define $b_n = \sup_{k\geq n} a_k$ as a sequence in $\mathbb{R}^*$. The sequence $\{b_n\}$ is monotone decreasing. Thus it converges to $\inf_n b_n$ in $\mathbb{R}^*$. We define[2]

$$\limsup_n a_n = \inf_n \sup_{k\geq n} a_k$$

as a member of $\mathbb{R}^*$, and we define

$$\liminf_n a_n = \sup_n \inf_{k\geq n} a_k$$

as a member of $\mathbb{R}^*$. Let us underscore that $\limsup a_n$ and $\liminf a_n$ always exist. However, one or both may be $\pm\infty$ even if a_n is in $\mathbb{R}$ for every n.

Proposition 1.7. The operations lim sup and lim inf on sequences $\{a_n\}$ and $\{b_n\}$ in $\mathbb{R}^*$ have the following properties:

(a) if $a_n \leq b_n$ for all n, then $\limsup a_n \leq \limsup b_n$ and $\liminf a_n \leq \liminf b_n$,

[2]The notation $\overline{\lim}$ was at one time used for lim sup, and $\underline{\lim}$ was used for lim inf.

(b) $\liminf a_n \le \limsup a_n$,
(c) $\{a_n\}$ has a subsequence converging in $\mathbb{R}^*$ to $\limsup a_n$ and another subsequence converging in $\mathbb{R}^*$ to $\liminf a_n$,
(d) $\limsup a_n$ is the supremum of all subsequential limits of $\{a_n\}$ in $\mathbb{R}^*$, and $\liminf_n$ is the infimum of all subsequential limits of $\{a_n\}$ in $\mathbb{R}^*$,
(e) if $\limsup a_n < +\infty$, then $\limsup a_n$ is the infimum of all extended real numbers a such that $a_n \ge a$ for only finitely many n, and if $\liminf a_n > -\infty$, then $\liminf a_n$ is the supremum of all extended real numbers a such that $a_n \le a$ for only finitely many n,
(f) the sequence $\{a_n\}$ in $\mathbb{R}^*$ converges in $\mathbb{R}^*$ if and only if $\liminf a_n = \limsup a_n$, and in this case the limit is the common value of $\liminf a_n$ and $\limsup a_n$.

REMARK. It is enough to prove the results about lim sup, since $\liminf a_n = -\limsup(-a_n)$.

PROOFS FOR lim sup.

(a) From $a_l \le b_l$ for all l, we have $a_l \le \sup_{k\ge n} b_k$ if $l \ge n$. Hence $\sup_{l\ge n} a_l \le \sup_{k\ge n} b_k$. Then (a) follows by taking the limit on n.

(b) This follows by taking the limit on n of the inequality $\inf_{k\ge n} a_k \le \sup_{k\ge n} a_k$.

(c) We divide matters into cases. The main case is that $a = \limsup a_n$ is in $\mathbb{R}$. Inductively, for each $l \ge 1$, choose $N \ge n_{l-1}$ such that $|\sup_{k\ge N} a_k - a| < l^{-1}$. Then choose $n_l > n_{l-1}$ such that $|a_{n_l} - \sup_{k\ge N} a_k| < l^{-1}$. Together these inequalities imply $|a_{n_l} - a| < 2l^{-1}$ for all l, and thus $\lim_{l\to\infty} a_{n_l} = a$. The second case is that $a = \limsup a_n$ equals $+\infty$. Since $\sup_{k\ge n} a_k$ is monotone decreasing in n, we must have $\sup_{k\ge n} a_k = +\infty$ for all n. Inductively for $l \ge 1$, we can choose $n_l > n_{l-1}$ such that $a_{n_l} \ge l$. Then $\lim_{l\to\infty} a_{n_l} = +\infty$. The third case is that $a = \limsup a_n$ equals $-\infty$. The sequence $b_n = \sup_{k\ge n} a_k$ is monotone decreasing to $-\infty$. Inductively for $l \ge 1$, choose $n_l > n_{l-1}$ such that $b_{n_l} \le -l$. Then $a_{n_l} \le b_{n_l} \le -l$, and $\lim_{l\to\infty} a_{n_l} = -\infty$.

(d) By (c), $\limsup a_n$ is one subsequential limit. Let $a = \lim_{k\to\infty} a_{n_k}$ be another subsequential limit. Put $b_n = \sup_{l\ge n} a_l$. Then $\{b_n\}$ converges to $\limsup a_n$ in $\mathbb{R}^*$, and the same thing is true of every subsequence. Since $a_{n_k} \le \sup_{l\ge n_k} a_l = b_{n_k}$ for all k, we can let k tend to infinity and obtain $a = \lim_{k\to\infty} a_{n_k} \le \lim_{k\to\infty} b_{n_k} = \limsup a_n$.

(e) Since $\limsup a_n < +\infty$, we have $\sup_{k\ge n} a_k < +\infty$ for n greater than or equal to some N. For this N and any $a > \sup_{k\ge N} a_k$, we then have $a_n \ge a$ only finitely often. Thus there exists $a \in \mathbb{R}$ such that $a_n \ge a$ for only finitely many n. On the other hand, if a' is a real number $< \limsup a_n$, then (c) shows that $a_n \ge a'$ for infinitely many n. Hence

$$\limsup a_n \le \inf\{a \mid a_n \ge a \text{ for only finitely many } a\}.$$

Arguing by contradiction, suppose that $<$ holds in this inequality, and let a'' be a real number strictly in between the two sides of the inequality. Then $\sup_{k\geq n} a_k < a''$ for n large enough, and so $a_n \geq a''$ only finitely often. But then a'' is in the set

$$\{a \mid a_n \geq a \text{ for only finitely many } a\},$$

and the statement that a'' is less than the infimum of this set gives a contradiction.

(f) If $\{a_n\}$ converges in $\mathbb{R}^*$, then (c) forces $\liminf a_n = \limsup a_n$. Conversely suppose $\liminf a_n = \limsup a_n$, and let a be the common value of $\liminf a_n$ and $\limsup a_n$. The main case is that a is in $\mathbb{R}$. Let $\epsilon > 0$ be given. By (e), $a_n \geq a+\epsilon$ only finitely often, and $a_n \leq a - \epsilon$ only finitely often. Thus $|a_n - a| < \epsilon$ for all n sufficiently large. In other words, $\lim a_n = a$ as asserted. The other cases are that $a = +\infty$ or $a = -\infty$, and they are completely analogous to each other. Suppose for definiteness that $a = +\infty$. Since $\liminf a_n = +\infty$, the monotone increasing sequence $b_n = \inf_{k\geq n} a_k$ converges in $\mathbb{R}^*$ to $+\infty$. Given M, choose N such that $b_n \geq M$ for $n \geq N$. Then also $a_n \geq M$ for $n \geq N$, and a_n converges in $\mathbb{R}^*$ to $+\infty$. This completes the proof.

With Proposition 1.7 as a tool, we can now prove the Bolzano–Weierstrass Theorem. The remainder of the section will consist of applications of this theorem, showing that Cauchy sequences in $\mathbb{R}$ converge in $\mathbb{R}$, that continuous functions on closed bounded intervals of $\mathbb{R}$ are uniformly continuous, that continuous functions on closed bounded intervals are bounded and assume their maximum and minimum values, and that continuous functions on closed intervals take on all intermediate values.

Theorem 1.8 (Bolzano–Weierstrass). Every bounded sequence in $\mathbb{R}$ has a convergent subsequence with limit in $\mathbb{R}$.

PROOF. If the given bounded sequence is $\{a_n\}$, form the subsequence noted in Proposition 1.7c that converges in $\mathbb{R}^*$ to $\limsup a_n$. All quantities arising in the formation of $\limsup a_n$ are in $\mathbb{R}$, since $\{a_n\}$ is bounded, and thus the limit is in $\mathbb{R}$.

A sequence $\{a_n\}$ in $\mathbb{R}$ is called a **Cauchy sequence** if for any $\epsilon > 0$, there exists an N such that $|a_n - a_m| < \epsilon$ for all n and m that are $\geq N$.

EXAMPLE. Every convergent sequence in $\mathbb{R}$ with limit in $\mathbb{R}$ is Cauchy. In fact, let $a = \lim a_n$, and let $\epsilon > 0$ be given. Choose N such that $n \geq N$ implies $|a_n - a| < \epsilon$. Then $n, m \geq N$ implies

$$|a_n - a_m| \leq |a_n - a| + |a - a_m| < \epsilon + \epsilon = 2\epsilon.$$

Hence the sequence is Cauchy.

In the above example and elsewhere in this book, we allow ourselves the luxury of having our final bound come out as a fixed multiple $M\epsilon$ of ϵ, rather than ϵ itself. Strictly speaking, we should have introduced $\epsilon' = \epsilon/M$ and aimed for ϵ' rather than ϵ. Then our final bound would have been $M\epsilon' = \epsilon$. Since the technique for adjusting a proof in this way is always the same, we shall not add these extra steps in the future unless there would otherwise be a possibility of confusion.

This convention suggests a handy piece of terminology—that a proof as in the above example, in which $M = 2$, is a "2ϵ proof." That name conveys a great deal of information about the proof, saying that one should expect two contributions to the final estimate and that the final bound will be 2ϵ.

Theorem 1.9 (Cauchy criterion). Every Cauchy sequence in $\mathbb{R}$ converges to a limit in $\mathbb{R}$.

PROOF. Let $\{a_n\}$ be Cauchy in $\mathbb{R}$. First let us see that $\{a_n\}$ is bounded. In fact, for $\epsilon = 1$, choose N such that $n, m \geq N$ implies $|a_n - a_m| < 1$. Then $|a_m| \leq |a_N| + 1$ for $m \geq N$, and $M = \max\{|a_1|, \ldots, |a_{N-1}|, |a_N| + 1\}$ is a common bound for all $|a_n|$.

Since $\{a_n\}$ is bounded, it has a convergent subsequence $\{a_{n_k}\}$, say with limit a, by the Bolzano–Weierstrass Theorem. The subsequential limit has to satisfy $|a| \leq M$ within $\mathbb{R}^*$, and thus a is in $\mathbb{R}$.

Finally let us see that $\lim a_n = a$. In fact, if $\epsilon > 0$ is given, choose N such that $n_k \geq N$ implies $|a_{n_k} - a| < \epsilon$. Also, choose $N' \geq N$ such that $n, m \geq N'$ implies $|a_n - a_m| < \epsilon$. If $n \geq N'$, then any $n_k \geq N'$ has $|a_n - a_{n_k}| < \epsilon$, and hence

$$|a_n - a| \leq |a_n - a_{n_k}| + |a_{n_k} - a| < \epsilon + \epsilon = 2\epsilon.$$

This completes the proof.

Let f be a function with domain an interval and with range in $\mathbb{R}$. The interval is allowed to be unbounded, but it is required to be a subset of $\mathbb{R}$. We say that f is **continuous at** a point x_0 of the domain of f within $\mathbb{R}$ if for each $\epsilon > 0$, there is some $\delta > 0$ such that all x in the domain of f that satisfy $|x - x_0| < \delta$ have $|f(x) - f(x_0)| < \epsilon$. This notion is sometimes abbreviated as $\lim_{x\to x_0} f(x) = f(x_0)$. Alternatively, one may say that $f(x)$ tends to $f(x_0)$ as x tends to x_0, and one may write $f(x) \to f(x_0)$ as $x \to x_0$.

A mathematically equivalent definition is that f is continuous at x_0 if whenever a sequence has $x_n \to x_0$ within the domain interval, then $f(x_n) \to f(x_0)$. This latter version of continuity will be shown in Section II.4 to be equivalent to the former version, given in terms of continuous limits, in greater generality than just for $\mathbb{R}$, and thus we shall not stop to prove the equivalence now. We say that f is **continuous** if it is continuous at all points of its domain.

We say that the a function f as above is **uniformly continuous** on its domain if for any $\epsilon > 0$, there is some $\delta > 0$ such that $|f(x) - f(x_0)| < \epsilon$ whenever x and x_0 are in the domain interval and $|x - x_0| < \delta$. (In other words, the condition for the continuity to be uniform is that δ can always be chosen independently of x_0.)

EXAMPLE. The function $f(x) = x^2$ is continuous on $(-\infty, +\infty)$, but it is not uniformly continuous. In fact, it is not uniformly continuous on $[1, +\infty)$. Assuming the contrary, choose δ for $\epsilon = 1$. Then we must have $\left|(x+\frac{\delta}{2})^2 - x^2\right| < 1$ for all $x \geq 1$. But $\left|(x + \frac{\delta}{2})^2 - x^2\right| = \delta x + \frac{\delta^2}{4} \geq \delta x$, and this is ≥ 1 for $x \geq \delta^{-1}$.

Theorem 1.10. A continuous function f from a closed bounded interval $[a, b]$ into $\mathbb{R}$ is uniformly continuous.

PROOF. Fix $\epsilon > 0$. For x_0 in the domain of f, the continuity of f at x_0 means that it makes sense to define

$$\delta_{x_0}(\epsilon) = \min\left\{1,\ \sup\left\{\delta' > 0 \;\middle|\; \begin{array}{l} |x - x_0| < \delta' \text{ and } x \text{ in the domain} \\ \text{of } f \text{ imply } |f(x) - f(x_0)| < \epsilon \end{array}\right\}\right\}.$$

If $|x - x_0| < \delta_{x_0}(\epsilon)$, then $|f(x) - f(x_0)| < \epsilon$. Put $\delta(\epsilon) = \inf_{x_0 \in [a,b]} \delta_{x_0}(\epsilon)$. Let us see that it is enough to prove that $\delta(\epsilon) > 0$. If x and y are in $[a, b]$ with $|x - y| < \delta(\epsilon)$, then $|x - y| < \delta(\epsilon) \leq \delta_y(\epsilon)$. Hence $|f(x) - f(y)| < \epsilon$ as required.

Thus we are to prove that $\delta(\epsilon) > 0$. If $\delta(\epsilon) = 0$, then, for each integer $n > 0$, we can choose x_n such that $\delta_{x_n}(\epsilon) < \frac{1}{n}$. By the Bolzano–Weierstrass Theorem, there is a convergent subsequence, say with $x_{n_k} \to x'$. Along this subsequence, $\delta_{x_{n_k}}(\epsilon) \to 0$. Fix k large enough so that $|x_{n_k} - x'| < \frac{1}{2}\delta_{x'}(\frac{\epsilon}{2})$. Then $|f(x_{n_k}) - f(x')| < \frac{\epsilon}{2}$. Also, $|x - x_{n_k}| < \frac{1}{2}\delta_{x'}(\frac{\epsilon}{2})$ implies

$$|x - x'| \leq |x - x_{n_k}| + |x_{n_k} - x'| < \tfrac{1}{2}\delta_{x'}(\tfrac{\epsilon}{2}) + \tfrac{1}{2}\delta_{x'}(\tfrac{\epsilon}{2}) = \delta_{x'}(\tfrac{\epsilon}{2}),$$

so that $|f(x) - f(x')| < \frac{\epsilon}{2}$ and

$$|f(x_{n_k}) - f(x)| \leq |f(x_{n_k}) - f(x')| + |f(x') - f(x)| < \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2} = \epsilon.$$

Consequently our arbitrary large fixed k has $\delta_{x_{n_k}} \geq \frac{1}{2}\delta_{x'}(\frac{\epsilon}{2})$, and the sequence $\{\delta_{x_{n_k}}(\epsilon)\}$ cannot be tending to 0.

Theorem 1.11. A continuous function f from a closed bounded interval $[a, b]$ into $\mathbb{R}$ is bounded and takes on maximum and minimum values.

PROOF. Let $c = \sup_{x \in [a,b]} f(x)$ in $\mathbb{R}^*$. Choose a sequence x_n in $[a, b]$ with $f(x_n)$ increasing to c. By the Bolzano–Weierstrass Theorem, $\{x_n\}$ has a convergent subsequence, say $x_{n_k} \to x'$. By continuity, $f(x_{n_k}) \to f(x')$. Then $f(x') = c$, and c is a finite maximum. The proof for a finite minimum is similar.

Theorem 1.12 (Intermediate Value Theorem). Let $a < b$ be real numbers, and let $f : [a, b] \to \mathbb{R}$ be continuous. Then f, in the interval $[a, b]$, takes on all values between $f(a)$ and $f(b)$.

REMARK. The proof below, which uses the Bolzano–Weierstrass Theorem, does not make absolutely clear what aspects of the structure of $\mathbb{R}$ are essential to the argument. A conceptually clearer proof will be given in Section II.8 and will bring out that the essential property of the interval $[a, b]$ is its "connectedness" in a sense to be defined in that section.

PROOF. Let $f(a) = \alpha$ and $f(b) = \beta$, and let γ be between α and β. We may assume that γ is in fact strictly between α and β. Possibly by replacing f by $-f$, we may assume that also $\alpha < \beta$. Let

$$A = \{x \in [a, b] \mid f(x) \le \gamma\} \qquad \text{and} \qquad B = \{x \in [a, b] \mid f(x) \ge \gamma\}.$$

These sets are nonempty, since a is in A and b is in B, and f is bounded as a result of Theorem 1.11. Thus the numbers $\gamma_1 = \sup\{f(x) \mid x \in A\}$ and $\gamma_2 = \inf\{f(x) \mid x \in B\}$ are well defined and have $\gamma_1 \le \gamma \le \gamma_2$.

If $\gamma_1 = \gamma$, then we can find a sequence $\{x_n\}$ in A such that $f(x_n)$ converges to γ. Using the Bolzano–Weierstrass Theorem, we can find a convergent subsequence $\{x_{n_k}\}$ of $\{x_n\}$, say with limit x_0. By continuity of f, $\{f(x_{n_k})\}$ converges to $f(x_0)$. Then $f(x_0) = \gamma_1 = \gamma$, and we are done. Arguing by contradiction, we may therefore assume that $\gamma_1 < \gamma$. Similarly we may assume that $\gamma < \gamma_2$, but we do not need to do so.

Let $\epsilon = \gamma_2 - \gamma_1$, and choose, by Theorem 1.10 and uniform continuity, $\delta > 0$ such that $|x_1 - x_2| < \delta$ implies $|f(x_1) - f(x_2)| < \epsilon$ whenever x_1 and x_2 both lie in $[a, b]$. Then choose an integer n such that $2^{-n}(b - a) < \delta$, and consider the value of f at the points $p_k = a + k2^{-n}(b - a)$ for $0 \le k \le 2^{-n}$. Since $p_{k+1} - p_k = 2^{-n}(b - a) < \delta$, we have $|f(p_{k+1}) - f(p_k)| < \epsilon = \gamma_2 - \gamma_1$. Consequently if $f(p_k) \le \gamma_1$, then

$$f(p_{k+1}) \le f(p_k) + |f(p_{k+1}) - f(p_k)| < \gamma_1 + (\gamma_2 - \gamma_1) = \gamma_2,$$

and hence $f(p_{k+1}) \le \gamma_1$. Now $f(p_0) = f(a) = \alpha \le \gamma_1$. Thus induction shows that $f(p_k) \le \gamma_1$ for all $k \le 2^{-n}$. However, for $k = 2^n$, we have $p_{2^{-n}} = b$, and $f(b) = \beta \ge \gamma > \gamma_1$, and we have arrived at a contradiction.

Further topics. Here a number of other topics of an undergraduate course in real-variable theory fit well logically. Among these are countable vs. uncountable sets, infinite series and tests for their convergence, the fact that every rearrangement of an infinite series of positive terms has the same sum, special sequences, derivatives, the Mean Value Theorem as in Section A2 of the appendix, and continuity and differentiability of inverse functions as in Section A3 of the appendix. We shall not stop here to review these topics, which are treated in many books. One such book is Rudin's *Principles of Mathematical Analysis*, the relevant chapters being 1 to 5. In Chapter 2 of that book, only the first few pages are needed; they are the ones where countable and uncountable sets are discussed.

2. Interchange of Limits

Let $\{b_{ij}\}$ be a doubly indexed sequence of real numbers. It is natural to ask for the extent to which

$$\lim_i \lim_j b_{ij} = \lim_j \lim_i b_{ij},$$

more specifically to ask how to tell, in an expression involving iterated limits, whether we can interchange the order of the two limit operations. We can view matters conveniently in terms of an infinite matrix

$$\begin{pmatrix} b_{11} & b_{12} & \cdots \\ b_{21} & b_{22} & \\ \vdots & & \ddots \end{pmatrix}.$$

The left-hand iterated limit, namely $\lim_i \lim_j b_{ij}$, is obtained by forming the limit of each row, assembling the results, and then taking the limit of the row limits down through the rows. The right-hand iterated limit, namely $\lim_j \lim_i b_{ij}$, is obtained by forming the limit of each column, assembling the results, and then taking the limit of the column limits through the columns. If we use the particular infinite matrix

$$\begin{pmatrix} 1 & 1 & 1 & 1 & \cdots \\ 0 & 1 & 1 & 1 & \cdots \\ 0 & 0 & 1 & 1 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \vdots & & & & \ddots \end{pmatrix},$$

then we see that the first iterated limit depends only on the part of the matrix above the main diagonal, while the second iterated limit depends only on the part of the matrix below the main diagonal. Thus the two iterated limits in general have no reason at all to be related. In the specific matrix that we have just considered, they are 1 and 0, respectively. Let us consider some examples along the same lines but with an analytic flavor.

EXAMPLES.

(1) Let $b_{ij} = \dfrac{j}{i+j}$. Then $\lim_i \lim_j b_{ij} = 1$, while $\lim_j \lim_i b_{ij} = 0$.

(2) Let F_n be a continuous real-valued function on $\mathbb{R}$, and suppose that $F(x) = \lim F_n(x)$ exists for every x. Is F continuous? This is the same kind of question. It asks whether $\lim_{t\to x} F(t) \stackrel{?}{=} F(x)$, hence whether

$$\lim_{t\to x} \lim_{n\to\infty} F_n(t) \stackrel{?}{=} \lim_{n\to\infty} \lim_{t\to x} F_n(t).$$

If we take $f_k(x) = \dfrac{x^2}{(1+x^2)^k}$ for $k \geq 0$ and define $F_n(x) = \sum_{k=0}^n f_k(x)$, then each F_n is continuous. The sequence of functions $\{F_n\}$ has a pointwise limit $F(x) = \sum_{k=0}^\infty \dfrac{x^2}{(1+x^2)^k}$. The series is a geometric series, and we can easily calculate explicitly the partial sums and the limit function. The latter is

$$F(x) = \begin{cases} 0 & \text{if } x = 0 \\ 1 + x^2 & \text{if } x \neq 0. \end{cases}$$

It is apparent that the limit function is discontinuous.

(3) Let $\{f_n\}$ be a sequence of differentiable functions, and suppose that $f(x) = \lim f_n(x)$ exists for every x and is differentiable. Is $\lim f_n'(x) = f'(x)$? This question comes down to whether

$$\lim_{n\to\infty} \lim_{t\to x} \frac{f_n(t) - f_n(x)}{t - x} \stackrel{?}{=} \lim_{t\to x} \lim_{n\to\infty} \frac{f_n(t) - f_n(x)}{t - x}.$$

An example where the answer is negative uses the sine and cosine functions, which are undefined in the rigorous development until Section 7 on power series. The example has $f_n(x) = \dfrac{\sin nx}{\sqrt{n}}$ for $n \geq 1$. Then $\lim_n f_n(x) = 0$, so that $f(x) = 0$ and $f'(x) = 0$. Also, $f_n'(x) = \sqrt{n} \cos nx$, so that $f_n'(0) = \sqrt{n}$ does not tend to $0 = f'(0)$.

Yet we know many examples from calculus where an interchange of limits is valid. For example, in calculus of two variables, the first partial derivatives of nice functions—polynomials, for example—can be computed in either order with the same result, and double integrals of continuous functions over a rectangle can be calculated as iterated integrals in either order with the same result. Positive theorems about interchanging limits are usually based on some kind of uniform behavior, in a sense that we take up in the next section. A number of positive results of this kind ultimately come down to the following general theorem about doubly indexed sequences that are monotone increasing in each variable. In Section 3 we shall examine the mechanism of this theorem closely: the proof shows that the equality in question is $\sup_i \sup_j b_{ij} = \sup_j \sup_i b_{ij}$ and that it holds because both sides equal $\sup_{i,j} b_{ij}$.

Theorem 1.13. Let b_{ij} be members of $\mathbb{R}^*$ that are ≥ 0 for all i and j. Suppose that b_{ij} is monotone increasing in i, for each j, and is monotone increasing in j, for each i. Then

$$\lim_i \lim_j b_{ij} = \lim_j \lim_i b_{ij},$$

with all the indicated limits existing in $\mathbb{R}^*$.

PROOF. Put $L_i = \lim_j b_{ij}$ and $L'_j = \lim_i b_{ij}$. These limits exist in $\mathbb{R}^*$, since the sequences in question are monotone. Then $L_i \le L_{i+1}$ and $L'_j \le L'_{j+1}$, and thus

$$L = \lim_i L_i \qquad \text{and} \qquad L' = \lim_j L'_j$$

both exist in $\mathbb{R}^*$. Arguing by contradiction, suppose that $L < L'$. Then we can choose j_0 such that $L'_{j_0} > L$. Since $L'_{j_0} = \lim_i b_{ij_0}$, we can choose i_0 such that $b_{i_0 j_0} > L$. Then we have $L < b_{i_0 j_0} \le L_{i_0} \le L$, contradiction. Similarly the assumption $L' < L$ leads to a contradiction. We conclude that $L = L'$.

Corollary 1.14. If a_{lj} are members of $\mathbb{R}^*$ that are ≥ 0 and are monotone increasing in j for each l, then

$$\lim_j \sum_l a_{lj} = \sum_l \lim_j a_{lj}$$

in $\mathbb{R}^*$, the limits existing.

REMARK. This result will be generalized by the Monotone Convergence Theorem when we study abstract measure theory in Chapter V.

PROOF. Put $b_{ij} = \sum_{l=1}^{i} a_{lj}$ in Theorem 1.13.

Corollary 1.15. If c_{ij} are members of $\mathbb{R}^*$ that are ≥ 0 for all i and j, then

$$\sum_i \sum_j c_{ij} = \sum_j \sum_i c_{ij}$$

in $\mathbb{R}^*$, the limits existing.

REMARK. This result will be generalized by Fubini's Theorem when we study abstract measure theory in Chapter V.

PROOF. This follows from Corollary 1.14.

3. Uniform Convergence

Let us examine more closely what is happening in the proof of Theorem 1.13, in which it is proved that iterated limits can be interchanged under certain hypotheses of monotonicity. One of the iterated limits is $L = \lim_i \lim_j b_{ij}$, and the claim is that L is approached as i and j tend to infinity jointly. In terms of a matrix whose entries are the various b_{ij}'s, the pictorial assertion is that all the terms far down and to the right are close to L:

$$\left(\begin{array}{cc|l} \cdots & & \cdots \\ \cline{3-3} \cdots & & \text{All terms here} \\ & & \text{are close to } L \end{array}\right).$$

To see this claim, let us choose a row limit L_{i_0} that is close to L and then take an entry $b_{i_0 j_0}$ that is close to L_{i_0}. Then $b_{i_0 j_0}$ is close to L, and all terms down and to the right from there are even closer because of the hypothesis of monotonicity.

To relate this behavior to something uniform, suppose that $L < +\infty$, and let some $\epsilon > 0$ be given. We have just seen that we can arrange to have $|L - b_{ij}| < \epsilon$ whenever $i \geq i_0$ and $j \geq j_0$. Then $|L_i - b_{ij}| < \epsilon$ whenever $i \geq i_0$, provided $j \geq j_0$. Also, we have $\lim_j b_{ij} = L_i$ for $i = 1, 2, \ldots, i_0 - 1$. Thus $|L_i - b_{ij}| < \epsilon$ for all i, provided $j \geq j_0'$, where j_0' is some larger index than j_0. This is the notion of uniform convergence that we shall define precisely in a moment: an expression with a parameter (j in our case) has a limit (on the variable i in our case) with an estimate independent of the parameter. We can visualize matters as in the following matrix:

$$\begin{array}{c} \\ i \end{array} \begin{array}{c} \qquad j \qquad\quad j_0' \\ \left(\begin{array}{c|l} \cdots & \text{All terms here} \\ & \text{are close to } L_i \\ & \text{on all rows.} \end{array}\right). \end{array}$$

The vertical dividing line occurs when the column index j is equal to j_0', and all terms to the right of this line are close to their respective row limits L_i.

Let us see the effect of this situation on the problem of interchange of limits. The above diagram forces all the terms in the shaded part of $\left(\begin{array}{cc|l} \cdots & & \cdots \\ \cline{3-3} \cdots & & ////// \end{array}\right)$ to be close to one number if $\lim L_i$ exists, i.e., if the row limits are tending to a limit. If the other iterated limit exists, then it must be this same number. Thus the interchange of limits is valid under these circumstances.

Actually, we can get by with less. If, in the displayed diagram above, we assume that all the column limits L_j' exist, then it appears that all the column limits with $j \geq j_0'$ have to be close to the L_i's. From this we can deduce that the column limits have a limit L' and that the row limits L_i must tend to the limit of the column limits. In other words, the convergence of the rows in a suitable uniform fashion and the convergence of the columns together imply that both iterated limits exist and they are equal. We shall state this result rigorously as Proposition 1.16, which will become a prototype for applications later in this section.

Let S be a nonempty set, and let f and f_n, for integers $n \geq 1$, be functions from S to $\mathbb{R}$. We say that $f_n(x)$ **converges** to $f(x)$ **uniformly** for x in S if for any $\epsilon > 0$, there is an integer N such that $n \geq N$ implies $|f_n(x) - f(x)| < \epsilon$ for all x in S. It is equivalent to say that $\sup_{x \in E} |f_n(x) - f(x)|$ tends to 0 as n tends to infinity.

Proposition 1.16. Let b_{ij} be real numbers for $i \geq 1$ and $j \geq 1$. Suppose that

(i) $L_i = \lim_j b_{ij}$ exists in $\mathbb{R}$ uniformly in i, and
(ii) $L'_j = \lim_i b_{ij}$ exists in $\mathbb{R}$ for each j.

Then

(a) $L = \lim_i L_i$ exists in $\mathbb{R}$,
(b) $L' = \lim_j L'_j$ exists in $\mathbb{R}$,
(c) $L = L'$,
(d) the double limit on i and j of b_{ij} exists and equals the common value of the iterated limits L and L', i.e., for each $\epsilon > 0$, there exist i_0 and j_0 such that $|b_{ij} - L| < \epsilon$ whenever $i \geq i_0$ and $j \geq j_0$,
(e) $L'_j = \lim_i b_{ij}$ exists in $\mathbb{R}$ uniformly in j.

REMARK. In applications we shall sometimes have additional information, typically the validity of (a) or (b). According to the statement of the proposition, however, the conclusions are valid without taking this extra information as an additional hypothesis.

PROOF. Let $\epsilon > 0$ be given. By (i), choose j_0 such that $|b_{ij} - L_i| < \epsilon$ for all i whenever $j \geq j_0$. With $j \geq j_0$ fixed, (ii) says that $|b_{ij} - L'_j| < \epsilon$ whenever i is $\geq$ some $i_0 = i_0(j)$. For $j \geq j_0$ and $i \geq i_0(j)$, we then have

$$|L_i - L'_j| \leq |L_i - b_{ij}| + |b_{ij} - L'_j| < \epsilon + \epsilon = 2\epsilon.$$

If $j' \geq j_0$ and $i \geq i_0(j')$, we similarly have $|L_i - L'_{j'}| < 2\epsilon$. Hence if $j \geq j_0$, $j' \geq j_0$, and $i \geq \max\{i_0(j), i_0(j')\}$, then

$$|L'_j - L'_{j'}| \leq |L'_j - L_i| + |L_i - L'_{j'}| < 2\epsilon + 2\epsilon = 4\epsilon.$$

In other words, $\{L'_j\}$ is a Cauchy sequence. By Theorem 1.9, $L' = \lim_j L'_j$ exists in $\mathbb{R}$. This proves (b).

Passing to the limit in our inequality, we have $|L'_j - L'| \leq 4\epsilon$ when $j \geq j_0$ and in particular when $j = j_0$. If $i \geq i_0(j_0)$, then we saw that $|b_{ij_0} - L_i| < \epsilon$ and $|b_{ij_0} - L'_{j_0}| < \epsilon$. Hence $i \geq i_0(j_0)$ implies

$$|L_i - L'| \leq |L_i - b_{ij_0}| + |b_{ij_0} - L'_{j_0}| + |L'_{j_0} - L'| < \epsilon + \epsilon + 4\epsilon = 6\epsilon.$$

Since ϵ is arbitrary, $L = \lim_i L_i$ exists and equals L'. This proves (a) and (c).

Since $\lim_i L_i = L$, choose i_1 such that $|L_i - L| < \epsilon$ whenever $i \geq i_1$. If $i \geq i_1$ and $j \geq j_0$, we then have

$$|b_{ij} - L| \leq |b_{ij} - L_i| + |L_i - L| < \epsilon + \epsilon = 2\epsilon.$$

This proves (d).

Let i_1 and j_0 be as in the previous paragraph. We have seen that $|L'_j - L'_{j'}| < 4\epsilon$ for $j \geq j_0$. By (b), $|L'_j - L'| \leq 4\epsilon$ whenever $j \geq j_0$. Hence (c) and the inequality of the previous paragraph give

$$|b_{ij} - L'_j| \leq |b_{ij} - L| + |L - L'| + |L' - L'_j| < 2\epsilon + 0 + 4\epsilon = 6\epsilon$$

whenever $i \geq i_1$ and $j \geq j_0$. By (b), choose $j_1 \geq j_0$ such that $|b_{ij} - L'_j| < 6\epsilon$ whenever $i \in \{1, \ldots, i_1 - 1\}$ and $j \geq j_1$. Then $j \geq j_1$ implies $|b_{ij} - L'_j| < 6\epsilon$ for all i whenever $j \geq j_1$. This proves (e).

In checking for uniform convergence, we often do not have access to explicit expressions for limiting values. One device for dealing with the problem is a uniform version of the Cauchy criterion. Let S be a nonempty set, and let $\{f_n\}_{n\geq 1}$ be a sequence of functions from S to $\mathbb{R}$. We say that $\{f_n(x)\}$ is **uniformly Cauchy** for $x \in S$ if for any $\epsilon > 0$, there is an integer N such that $n \geq N$ and $m \geq N$ together imply $|f_n(x) - f_m(x)| < \epsilon$ for all x in S.

Proposition 1.17 (uniform Cauchy criterion). A sequence $\{f_n\}$ of functions from a nonempty set S to $\mathbb{R}$ is uniformly Cauchy if and only if it is uniformly convergent.

PROOF. If $\{f_n\}$ is uniformly convergent to f, we use a 2ϵ argument, just as in the example before Theorem 1.9: Given $\epsilon > 0$, choose N such that $n \geq N$ implies $|f_n(x) - f(x)| < \epsilon$. Then $n \geq N$ and $m \geq N$ together imply

$$|f_n(x) - f_m(x)| \leq |f_n(x) - f(x)| + |f(x) - f_m(x)| < \epsilon + \epsilon = 2\epsilon.$$

Thus $\{f_n\}$ is uniformly Cauchy.

Conversely suppose that $\{f_n\}$ is uniformly Cauchy. Then $\{f_n(x)\}$ is Cauchy for each x. Theorem 1.9 therefore shows that there exists a function $f : S \to \mathbb{R}$ such that $\lim_n f_n(x) = f(x)$ for each x. We prove that the convergence is uniform. Given $\epsilon > 0$, choose N, as is possible since $\{f_n\}$ is uniformly Cauchy, such that $n \geq N$ and $m \geq N$ together imply $|f_n(x) - f_m(x)| < \epsilon$. Letting m tend to ∞ shows that $|f_n(x) - f(x)| \leq \epsilon$ for $n \geq N$. Hence $\lim_n f_n(x) = f(x)$ uniformly for x in S.

In practice, uniform convergence often arises with infinite series of functions, and then the definition and results about uniform convergence are to be applied to the sequence of partial sums. If the series is $\sum_{k=1}^{\infty} a_k(x)$, one wants $\big|\sum_{k=m}^{n} a_k(x)\big|$ to be small for all m and n sufficiently large. Some of the standard tests for convergence of series of numbers yield tests for uniform convergence of series of functions just by introducing a parameter and ensuring that the estimates do not depend on the parameter. We give two clear-cut examples. One is the uniform **alternating series test** or **Leibniz test**, given in Corollary 1.18. A generalization is the handy test given in Corollary 1.19.

Corollary 1.18. If for each x in a nonempty set S, $\{a_n(x)\}_{n\geq 1}$ is a monotone decreasing sequence of nonnegative real numbers such that $\lim_n a_n(x) = 0$ uniformly in x, then $\sum_{n=1}^{\infty} (-1)^n a_n(x)$ converges uniformly.

PROOF. The hypotheses are such that $\big|\sum_{k=m}^{n} (-1)^k a_k(x)\big| \leq \sup_x |a_m(x)|$ whenever $n \geq m$, and the uniform convergence is immediate from the uniform Cauchy criterion.

Corollary 1.19. If for each x in a nonempty set S, $\{a_n(x)\}_{n\geq 1}$ is a monotone decreasing sequence of nonnegative real numbers such that $\lim_n a_n(x) = 0$ uniformly in x and if $\{b_n(x)\}_{n\geq 1}$ is a sequence of real-valued functions on S whose partial sums $B_n(x) = \sum_{k=1}^{n} b_k(x)$ have $|B_n(x)| \leq M$ for some M and all n and x, then $\sum_{n=1}^{\infty} a_n(x)b_n(x)$ converges uniformly.

PROOF. If $n \geq m$, summation by parts gives

$$\sum_{k=m}^{n} a_k(x)b_k(x) = \sum_{k=m}^{n-1} B_k(x)(a_k(x) - a_{k+1}(x)) + B_n(x)a_n(x) - B_{m-1}(x)a_m(x).$$

Let $\epsilon > 0$ be given, and choose N such that $a_k(x) \leq \epsilon$ for all x whenever $k \geq N$. If $n \geq m \geq N$, then

$$\begin{aligned} \Big|\sum_{k=m}^{n} a_k(x)b_k(x)\Big| &\leq \sum_{k=m}^{n-1} |B_k(x)|(a_k(x) - a_{k+1}(x)) + M\epsilon + M\epsilon \\ &\leq M\sum_{k=m}^{n-1}(a_k(x) - a_{k+1}(x)) + 2M\epsilon \\ &\leq Ma_m(x) + 2M\epsilon \\ &\leq 3M\epsilon, \end{aligned}$$

and the uniform convergence is immediate from the uniform Cauchy criterion.

A third consequence can be considered as a uniform version of the result that absolute convergence implies convergence. In practice it tends to be fairly easy to apply, but it applies only in the simplest situations.

Proposition 1.20 (Weierstrass M test). Let S be a nonempty set, and let $\{f_n\}$ be a sequence of real-valued functions on S such that $|f_n(x)| \le M_n$ for all x in S. Suppose that $\sum_n M_n < +\infty$. Then $\sum_{n=1}^\infty f_n(x)$ converges uniformly for x in S.

PROOF. If $n \ge m \ge N$, then $\big|\sum_{k=m+1}^n f_k(x)\big| \le \sum_{k=m}^n |f_k(x)| \le \sum_{k=m}^n M_k$, and the right side tends to 0 uniformly in x as N tends to infinity. Therefore the result follows from the uniform Cauchy criterion.

EXAMPLES.

(1) The series

$$\sum_{n=1}^\infty \frac{1}{n^2}\, x^n$$

converges uniformly for $-1 \le x \le 1$ by the Weierstrass M test with $M_n = 1/n^2$.

(2) The series

$$\sum_{n=1}^\infty (-1)^n\, \frac{x^2+n}{n^2}$$

converges uniformly for $-1 \le x \le 1$, but the M test does not apply. To see that the M test does not apply, we use the smallest possible M_n, which is $M_n = \sup_x \big|(-1)^n \frac{x^2+n}{n^2}\big| = \frac{n+1}{n^2}$. The series $\sum \frac{n+1}{n^2}$ diverges, and hence the M test cannot apply for any choice of the numbers M_n. To see the uniform convergence of the given series, we observe that the terms strictly alternate in sign. Also,

$$\frac{x^2+n}{n^2} \ge \frac{x^2+(n+1)}{(n+1)^2} \quad \text{because} \quad \frac{x^2}{n^2} \ge \frac{x^2}{(n+1)^2} \quad \text{and} \quad \frac{1}{n} \ge \frac{1}{n+1}.$$

Finally

$$\frac{x^2+n}{n^2} \le \frac{n+1}{n^2} \to 0$$

uniformly for $-1 \le x \le 1$. Hence the series converges uniformly by the uniform Leibniz test (Corollary 1.18).

Having developed some tools for proving uniform convergence, let us apply the notion of uniform convergence to interchanges of limits involving functions of a real variable. For a point of reference, recall the diagrams of interchanges of limits at the beginning of the section. We take the column index to be n and think of the row index as a variable t, which is tending to x. We make assumptions that correspond to (i) and (ii) in Proposition 1.16, namely that $\{f_n(t)\}$ converges uniformly in t as n tends to infinity, say to $f(t)$, and that $f_n(t)$ converges to some limit $f_n(x)$ as t tends to x. With $f_n(x)$ defined as this limit, f_n is continuous at x. In other words, the assumptions are that the sequence $\{f_n\}$ is uniformly convergent to f and each f_n is continuous.

Theorem 1.21. If $\{f_n\}$ is a sequence of real-valued functions on $[a, b]$ that are continuous at x and if $\{f_n\}$ converges to f uniformly, then f is continuous at x.

REMARKS. This is really a consequence of Proposition 1.16 except that one of the indices, namely t, is regarded as continuous and not discrete. Actually, there is a subtle simplification here, by comparison with Proposition 1.16, in that $\{f_n(x)\}$ at the limiting parameter x is being assumed to tend to $f(x)$. This corresponds to assuming (b) in the proposition, as well as (i) and (ii). Consequently the proof of the theorem will be considerably simpler than the proof of Proposition 1.16. In fact, the proof will be our first example of a 3ϵ proof. In many applications of Theorem 1.21, the given sequence $\{f_n\}$ is continuous at every x, and then the conclusion is that f is continuous at every x.

PROOF. We write

$$|f(t) - f(x)| \le |f(t) - f_n(t)| + |f_n(t) - f_n(x)| + |f_n(x) - f(x)|.$$

Given $\epsilon > 0$, choose N large enough so that $|f_n(t) - f(t)| < \epsilon$ for all t whenever $n \ge N$. With such an n fixed, choose some δ of continuity for the function f_n, the point x, and the number ϵ. Each term above is then $< \epsilon$, and hence $|f(t) - f(x)| < 3\epsilon$. Since ϵ is arbitrary, f is continuous at x.

Theorem 1.21 in effect uses only conclusion (c) of Proposition 1.16, which concerns the equality of the two iterated limits. Conclusion (d) gives a stronger result, namely that the double limit exists and equals each iterated limit. The strengthened version of Theorem 1.21 is as follows.

Theorem 1.21′. If $\{f_n\}$ is a sequence of real-valued functions on $[a, b]$ that are continuous at x and if $\{f_n\}$ converges to f uniformly, then for each $\epsilon > 0$, there exist an integer N and a number $\delta > 0$ such that

$$|f_n(t) - f(x)| < \epsilon$$

whenever $n \ge N$ and $|t - x| < \delta$.

PROOF. If $\epsilon > 0$ is given, choose N such that $|f_n(t) - f(t)| < \epsilon/2$ for all t whenever $n \ge N$, and choose δ in the conclusion of Theorem 1.21 such that $|t - x| < \delta$ implies $|f(t) - f(x)| < \epsilon/2$. Then

$$|f_n(t) - f(x)| \le |f_n(t) - f(t)| + |f(t) - f(x)| < \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2} = \epsilon$$

whenever $n \ge N$ and $|t - x| < \delta$. Theorem 1.21′ follows.

In interpreting our diagrams of interchanges of limits to get at the statement of Theorem 1.21, we took the column index to be n and thought of the row index as a variable t, which was tending to x. It is instructive to see what happens when the roles of n and t are reversed, i.e., when the row index is n and the column index is the variable t, which is tending to x. Again we have $f_n(t)$ converging to $f(t)$ and $\lim_{t\to x} f_n(t) = f_n(x)$, but the uniformity is different. This time we want the uniformity to be in n as t tends to x. This means that the δ of continuity that corresponds to ϵ can be taken independent of n. This is the notion of "equicontinuity," and there is a classical theorem about it. The theorem is actually stronger than Proposition 1.16 suggests, since the theorem assumes less than that $f_n(t)$ converges to $f(t)$ for all t.

Let $\mathcal{F} = \{f_\alpha \mid \alpha \in A\}$ be a set of real-valued functions on a bounded interval $[a, b]$. We say that $\mathcal{F}$ is **equicontinuous** at $x \in [a, b]$ if for each $\epsilon > 0$, there is some $\delta > 0$ such that $|t-x| < \delta$ implies $|f(t) - f(x)| < \epsilon$ for all $f \in \mathcal{F}$. The set $\mathcal{F}$ of functions is **pointwise bounded** if for each $t \in [a, b]$, there exists a number M_t such that $|f(t)| \le M_t$ for all $f \in \mathcal{F}$. The set is **uniformly equicontinuous** on $[a, b]$ if it is equicontinuous at each point x and if the δ can be taken independent of x. The set is **uniformly bounded** on $[a, b]$ if it is pointwise bounded at each $t \in [a, b]$ and the bound M_t can be taken independent of t.

Theorem 1.22 (Ascoli's Theorem). If $\{f_n\}$ is a sequence of real-valued functions on a closed bounded interval $[a, b]$ that is equicontinuous at each point of $[a, b]$ and pointwise bounded on $[a, b]$, then

(a) $\{f_n\}$ is uniformly equicontinuous and uniformly bounded on $[a, b]$,
(b) $\{f_n\}$ has a uniformly convergent subsequence.

PROOF. Since each f_n is continuous at each point, we know from Theorems 1.10 and 1.11 that each f_n is uniformly continuous and bounded. The proof of (a) amounts to an argument that the estimates in those theorems can be arranged to apply simultaneously for all n.

First consider the question of uniform boundedness. Choose, by Theorem 1.11, some x_n in $[a, b]$ with $|f_n(x_n)|$ equal to $K_n = \sup_{x\in[a,b]} |f_n(x)|$. Then choose a subsequence on which the numbers K_n tend to $\sup_n K_n$ in $\mathbb{R}^*$. There will be no loss of generality in assuming that this subsequence is our whole sequence. Apply the Bolzano–Weierstrass Theorem to find a convergent subsequence $\{x_{n_k}\}$ of $\{x_n\}$, say with limit x_0. By pointwise boundedness, find M_{x_0} with $|f_n(x_0)| \le M_{x_0}$ for all n. Then choose some δ of equicontinuity at x_0 for $\epsilon = 1$. As soon as k is large enough so that $|x_{n_k} - x_0| < \delta$, we have

$$K_{n_k} = |f_{n_k}(x_{n_k})| \le |f_{n_k}(x_{n_k}) - f_{n_k}(x_0)| + |f_{n_k}(x_0)| < 1 + M_{x_0}.$$

Thus $1 + M_{x_0}$ is a uniform bound for the functions f_n.

The proof of uniform equicontinuity proceeds in the same spirit but takes longer to write out. Fix $\epsilon > 0$. The uniform continuity of f_n for each n means that it makes sense to define

$$\delta_n(\epsilon) = \min\left\{1,\ \sup\left\{\delta' > 0 \;\middle|\; \begin{array}{l} |f(x) - f(y)| < \epsilon \text{ whenever } |x-y| < \delta' \\ \text{and } x \text{ and } y \text{ are in the domain of } f \end{array}\right\}\right\}.$$

If $|x - y| < \delta_n(\epsilon)$, then $|f_n(x) - f_n(y)| < \epsilon$. Put $\delta(\epsilon) = \inf_n \delta_n(\epsilon)$. Let us see that it is enough to prove that $\delta(\epsilon) > 0$: If x and y are in $[a, b]$ with $|x-y| < \delta(\epsilon)$, then $|x - y| < \delta(\epsilon) \le \delta_n(\epsilon)$. Hence $|f_n(x) - f_n(y)| < \epsilon$ as required.

Thus we are to prove that $\delta(\epsilon) > 0$. If $\delta(\epsilon) = 0$, then we first choose an increasing sequence $\{n_k\}$ of positive integers such that $\delta_{n_k}(\epsilon) < \frac{1}{k}$, and we next choose x_k and y_k in $[a, b]$ with $|x_k - y_k| < \delta_{n_k}(\epsilon)$ and $|f_k(x_k) - f_k(y_k)| \ge \epsilon$. Applying the Bolzano–Weierstrass Theorem, we obtain a subsequence $\{x_{k_l}\}$ of $\{x_k\}$ such that $\{x_{k_l}\}$ converges, say to x_0. Then

$$\limsup_l |y_{k_l} - x_0| \le \limsup_l |y_{k_l} - x_{k_l}| + \limsup_l |x_{k_l} - x_0| = 0 + 0 = 0,$$

so that $\{y_{k_l}\}$ converges to x_0. Now choose, by equicontinuity at x_0, a number $\delta' > 0$ such that $|f_n(x) - f_n(x_0)| < \frac{\epsilon}{2}$ for all n whenever $|x - x_0| < \delta'$. The convergence of $\{x_{k_l}\}$ and $\{y_{k_l}\}$ to x_0 implies that for large enough l, we have $|x_{k_l} - x_0| < \delta'/2$ and $|y_{k_l} - x_0| < \delta'/2$. Therefore $|f_{k_l}(x_{k_l}) - f_{k_l}(x_0)| < \frac{\epsilon}{2}$ and $|f_{k_l}(y_{k_l}) - f_{k_l}(x_0)| < \frac{\epsilon}{2}$, from which we conclude that $|f_{k_l}(x_{k_l}) - f_{k_l}(y_{k_l})| < \epsilon$. But we saw that $|f_k(x_k) - f_k(y_k)| \ge \epsilon$ for all k, and thus we have arrived at a contradiction. This proves the uniform equicontinuity and completes the proof of (a).

To prove (b), we first construct a subsequence of $\{f_n\}$ that is convergent at every rational point in $[a, b]$. We enumerate the rationals, say as $x_1, x_2, \ldots$. By the Bolzano–Weierstrass Theorem and the pointwise boundedness, we can find a subsequence of $\{f_n\}$ that is convergent at x_1, a subsequence of the result that is convergent at x_2, a subsequence of the result that is convergent at x_3, and so on. The trouble with this process is that each term of our original sequence may disappear at some stage, and then we are left with no terms that address all the rationals. The trick is to form the subsequence $\{f_{n_k}\}$ of the given $\{f_n\}$ whose k^{th} term is the k^{th} term of the k^{th} subsequence we constructed. Then the k^{th}, $(k+1)^{\text{st}}$, $(k+2)^{\text{nd}}, \ldots$ terms of $\{f_{n_k}\}$ all lie in our k^{th} constructed subsequence, and hence $\{f_{n_k}\}$ converges at the first k points $x_1, \ldots, x_k$. Since k is arbitrary, $\{f_{n_k}\}$ converges at every rational point.

Let us prove that $\{f_{n_k}\}$ is uniformly Cauchy. Let $\epsilon > 0$ be given, let δ be some corresponding number exhibiting equicontinuity, and choose finitely many rationals $r_1, \ldots, r_l$ in $[a, b]$ such that any member of $[a, b]$ is within δ of at least one of these rationals. Then choose N such that $|f_n(r_j) - f_m(r_j)| < \epsilon$ for

$1 \le j \le l$ whenever n and m are $\ge N$. If x is in $[a, b]$, let $r(x)$ be an r_j with $|x - r(x)| < \delta$. Whenever n and m are $\ge N$, we then have

$$\begin{aligned} &|f_n(x) - f_m(x)| \\ &\qquad \le |f_n(x) - f_n(r(x))| + |f_n(r(x)) - f_m(r(x))| + |f_m(r(x)) - f_m(x)| \\ &\qquad < \epsilon + \epsilon + \epsilon = 3\epsilon. \end{aligned}$$

Hence $\{f_{n_k}\}$ is uniformly Cauchy, and (b) follows from Proposition 1.17.

REMARK. The construction of the subsequence for which countably many convergence conditions were all satisfied is an important one and is often referred to as a **diagonal process** or as the **Cantor diagonal process**.

EXAMPLE. Let K and M be positive constants, and let $\mathcal{F}$ be the set of continuous real-valued functions f on $[a, b]$ such that $|f(t)| \le K$ for $a \le t \le b$ and such that the derivative $f'(t)$ exists for $a < t < b$ and satisfies $|f'(t)| \le M$ there. This set of functions is certainly uniformly bounded by K, and we show that it is also uniformly equicontinuous. To see the latter, we use the Mean Value Theorem. If x is in the closed interval $[a, b]$ and t is in the open interval (a, b), then there exists ξ depending on t and x such that

$$|f(t) - f(x)| = |f'(\xi)||t - x| \le M|t - x|.$$

From this inequality it follows that the number δ of uniform equicontinuity for ϵ and $\mathcal{F}$ can be taken to be ϵ/M. The hypotheses of Ascoli's Theorem are satisfied, and it follows that any sequence of functions in $\mathcal{F}$ has a uniformly convergent subsequence. The estimate of δ is independent of the uniform bound K, yet Ascoli's Theorem breaks down if there is no bound at all; for example, the sequence of constant functions with $f_n(x) = n$ is uniformly equicontinuous but has no convergent subsequence.

We turn now to the problem of interchange of derivative and limit. The two indices again will be an integer n that is tending to infinity and a parameter t that is tending to x. Proposition 1.16 takes away all the surprise in the statement of the theorem, and it tells us the steps to follow in a proof. What the proposition suggests is that the general entry in our interchange diagram should be whatever quantity we want to take an iterated limit of in either order. Thus we expect not a theorem about a general entry $f_n(t)$, but instead a theorem about a general entry $\dfrac{f_n(t) - f_n(x)}{t - x}$. The limit on n gives us $\dfrac{f(t) - f(x)}{t - x}$ for a limiting function f, and then the limit as $t \to x$ gives us $f'(x)$. In the other order the limit as $t \to x$ gives us $f_n'(x)$, and then we are to consider the limit on n. If Proposition 1.16 is to be a guide, we are to assume that the convergence in one variable is uniform in the other. The proposition also suggests that if we have existence of each row limit and each column limit, then uniform convergence when one variable occurs first is equivalent to uniform convergence when the other variable occurs first. Thus we should assume whichever is easier to verify.

Theorem 1.23. Suppose that $\{f_n\}$ is a sequence of real-valued functions continuous for $a \le t \le b$ and differentiable for $a < t < b$ such that $\{f_n'\}$ converges uniformly for $a < t < b$ and $\{f_n(x_0)\}$ converges in $\mathbb{R}$ for some x_0 with $a \le x_0 \le b$. Then $\{f_n\}$ converges uniformly for $a \le t \le b$ to a function f, and $f'(x) = \lim_n f_n'(x)$ for $a < x < b$, with the derivative and the limit existing.

REMARKS. The convergence of $\{f(x_0)\}$ cannot be dropped completely as a hypothesis because $f_n(t) = n$ would otherwise provide a counterexample. In practice, $\{f_n\}$ will be known in advance to be uniformly convergent. However, uniform convergence of $\{f_n\}$ is not enough by itself, as was shown by the example $f_n(x) = \dfrac{\sin nx}{\sqrt{n}}$ in Section 2.

PROOF. The first step is to apply the Mean Value Theorem to $f_n - f_m$, estimate $f_n' - f_m'$, and use the convergence of $\{f_n(x_0)\}$ to obtain the existence of the limit function f. The Mean Value Theorem produces some ξ strictly between t and x_0 such that

$$f_n(t) - f_m(t) = \big(f_n(x_0) - f_m(x_0)\big) + (t - x_0)\big(f_n'(\xi) - f_m'(\xi)\big).$$

Our hypotheses allow us to conclude that $\{f_n(t)\}$ is uniformly Cauchy, and thus $\{f_n\}$ converges uniformly to a limit function f by Proposition 1.17.

The second step is to apply the Mean Value Theorem again to $f_n - f_m$, this time to see that

$$\varphi_n(t) = \frac{f_n(t) - f_n(x)}{t - x}$$

converges uniformly in t (for $t \ne x$) as n tends to infinity, the limit being $\varphi(t) = \dfrac{f(t) - f(x)}{t - x}$. In fact, the Mean Value Theorem produces some ξ strictly between t and x such that

$$\varphi_n(t) - \varphi_m(t) = \frac{[f_n(t) - f_m(t)] - [f_n(x) - f_m(x)]}{t - x} = f_n'(\xi) - f_m'(\xi),$$

and the right side tends to 0 uniformly as n and m tend to infinity. Therefore $\{\varphi_n(t)\}$ is uniformly Cauchy for $t \ne x$, and Proposition 1.17 shows that it is uniformly convergent.

The third step is to extend the definition of φ to x by $\varphi_n(x) = f_n'(x)$ and then to see that φ_n is continuous at x and Theorem 1.21 applies. In fact, the definition of $\varphi_n(t)$ is as the difference quotient for the derivative of f_n at x, and thus $\varphi_n(t) \to f_n'(x) = \varphi_n(x)$. Hence φ_n is continuous at x. We saw in the second step that $\varphi_n(t)$ is uniformly convergent for $t \neq x$, and we are given that $\varphi_n(x) = f_n'(x)$ is convergent. Therefore $\varphi_n(t)$ is uniformly convergent for all t with

$$\lim \varphi_n(t) = \begin{cases} \dfrac{f(t) - f(x)}{t - x} & \text{for } t \neq x, \\ \lim f_n'(x) & \text{for } t = x. \end{cases}$$

Theorem 1.21 says that the limiting function $\lim \varphi_n(t)$ is continuous at x. Thus

$$\lim_{t \to x} \frac{f(t) - f(x)}{t - x} = \lim_n f_n'(x).$$

In other words, f is differentiable at x and $f'(x) = \lim_n f_n'(x)$.

4. Riemann Integral

This section contains a careful but limited development of the Riemann integral in one variable. The reader is assumed to have a familiarity with Riemann sums at the level of a calculus course. The objective in this section is to prove that bounded functions with only finitely many discontinuities are Riemann integrable, to address the interchange-of-limits problem that arises with a sequence of functions and an integration, to prove the Fundamental Theorem of Calculus in the case of continuous integrand, to prove a change-of-variables formula, and to relate Riemann integrals to general Riemann sums. The Riemann integral in several variables will be treated in Chapter III, and some of the theorems to be proved in the several-variable case at that time will be results that have not been proved here in the one-variable case. In Chapters VI and VII, in the context of the Lebesgue integral, we shall prove a much more sweeping version of the Fundamental Theorem of Calculus.

First we give the relevant definitions. We work with a function $f : [a, b] \to \mathbb{R}$ with $a \leq b$ in $\mathbb{R}$, and we always assume that f is bounded. A **partition** P of $[a, b]$ is a subdivision of the interval $[a, b]$ into subintervals, and we write such a partition as

$$a = x_0 \leq x_1 \leq \cdots \leq x_n = b.$$

The points x_j will be called the **subdivision points** of the partition, and we may abbreviate the partition as $P = \{x_i\}_{i=0}^n$. In order to permit integration over an interval of zero length, we allow partitions in which two consecutive x_j's are

Pages 27–52 do not appear in this file.

estimate, and thus the relationship between the sum $f(x)$ and the original function $(1+x)^p$ is not immediately apparent. However, we can use Corollary 1.38 to obtain

$$f'(x) = \sum_{n=1}^{\infty} \frac{np(p-1)\cdots(p-n+1)}{n!} x^{n-1} = \sum_{n=0}^{\infty} \frac{p(p-1)\cdots(p-n)}{n!} x^n$$

for $|x| < 1$. We compute $(1+x)f'(x)$ by multiplying the first series by x, the second series by 1, and adding. If we write the constant term separately, the result is

$$(1+x)f'(x) = p + \sum_{n=1}^{\infty} \frac{p(p-1)\cdots(p-n+1)[n+(p-n)]}{n!} x^n = pf(x).$$

Therefore

$$\begin{aligned} \frac{d}{dx}\big[(1+x)^{-p} f(x)\big] &= -p(1+x)^{-p-1} f(x) + (1+x)^{-p} f'(x) \\ &= (1+x)^{-p-1}[-pf(x) + (1+x)f'(x)] = 0, \end{aligned}$$

and $(1+x)^{-p} f(x)$ has to be constant for $|x| < 1$. From the series whose sum is $f(x)$, we see that $f(0) = 1$, and hence the constant is 1. Thus $f(x) = (1+x)^p$, and we have established the binomial series expansion

$$(1+x)^p = \sum_{n=0}^{\infty} \frac{p(p-1)\cdots(p-n+1)}{n!} x^n$$

for $-1 < x < 1$.

8. Summability

Summability refers to an operation on a sequence of complex numbers to make it more likely that the sequence will converge. The subject is of interest particularly with Fourier series, where the ordinary partial sums may not converge even at points where the given function is continuous.

Let $\{s_n\}_{n\geq 0}$ be a sequence in $\mathbb{C}$, and define its sequence of **Cesàro sums**, or arithmetic means, to be given by

$$\sigma_n = \frac{s_0 + s_1 + \cdots + s_n}{n+1}$$

for $n \geq 0$. If $\lim_n \sigma_n = \sigma$ exists in $\mathbb{C}$, we say that $\{s_n\}$ is **Cesàro summable** to the limit σ. For example the sequence with $s_n = (-1)^n$ for $n \geq 0$ is not convergent, but it is Cesàro summable to the limit 0 because σ_n is 0 for all odd n and is $\frac{1}{n+1}$ for all even n.

Theorem 1.47. If a complex sequence $\{s_n\}_{n\geq 0}$ is convergent in $\mathbb{C}$ to the limit s, then $\{s_n\}$ is Cesàro summable to the limit s.

REMARK. The argument is a 2ϵ proof, and two things are affecting σ_n. For k small and fixed, the contribution of s_k to σ_n is $s_k/(n+1)$ and is tending to 0. For k large, any s_k is close to s, and the average of such terms is close to s.

PROOF. Let $\epsilon > 0$ be given, and choose N_1 such that $k \geq N_1$ implies $|s_k - s| < \epsilon$. If $n \geq N_1$, then

$$\sigma_n - s = \frac{(s_0 - s) + \cdots + (s_{N_1} - s)}{n+1} + \frac{(s_{N_1+1} - s) + \cdots + (s_n - s)}{n+1},$$

so that

$$\begin{aligned} |\sigma_n - s| &\leq \frac{|s_0| + \cdots + |s_{N_1}| + (N_1+1)|s|}{n+1} + \frac{n - N_1}{n+1}\epsilon \\ &\leq \frac{|s_0| + \cdots + |s_{N_1}| + (N_1+1)|s|}{n+1} + \epsilon. \end{aligned}$$

The numerator of the first term is fixed, and thus we can choose $N \geq N_1$ large enough so that the first term is $< \epsilon$ whenever $n \geq N$. If $n \geq N$, then we see that $|\sigma_n - s| < 2\epsilon$. Since ϵ is arbitrary, the theorem follows.

Next let $\{a_n\}_{n\geq 0}$ be a complex sequence, and let $\{s_n\}_{n\geq 0}$ be the sequence of partial sums with $s_n = \sum_{k=0}^n a_k$. Form the power series $\sigma_r = \sum_{n=0}^\infty a_n r^n$. We say that the sequence $\{s_n\}$ of partial sums is **Abel summable** to the limit s in $\mathbb{C}$ if $\lim_{r\uparrow 1} \sigma_r = s$, i.e., if for each $\epsilon > 0$, there is some r_0 such that $r_0 \leq r < 1$ implies that $|\sigma_r - s| < \epsilon$. For example, take $a_k = (-1)^k$, so that s_n equals 1 if n is even and equals 0 if n is odd. The sequence $\{s_n\}$ of partial sums is divergent. The r^{th} Abel sum σ_r is given by the geometric series $\sum_{k=0}^\infty (-1)^k r^k$ with sum $1/(1+r)$. Letting r increase to 1, we see that $\{s_n\}$ is Abel summable with limit $\frac{1}{2}$.

Theorem 1.48 (Abel's Theorem). Let $\{a_n\}_{n\geq 0}$ be a complex sequence, and let $\{s_n\}_{n\geq 0}$ be the sequence of partial sums with $s_n = \sum_{k=0}^n a_k$. If $\{s_n\}_{n\geq 0}$ is convergent in $\mathbb{C}$ to the limit s, then $\{s_n\}$ is Abel summable to the limit s.

REMARK. The proof will proceed along the same lines as in the previous case. It is first necessary to express the Abel sums σ_r in terms of the s_k's.

PROOF. Since $\{s_n\}$ converges, $\{s_n\}$ and $\{a_n\}$ are bounded, and thus $\sum_{n=0}^\infty s_n r^n$ and $\sum_{k=0}^\infty a_k r^k$ are absolutely convergent for $0 \leq r < 1$. With $s_{-1} = 0$, write

$$\begin{aligned} \sigma_r &= \sum_{n=0}^\infty a_n r^n = \sum_{n=0}^\infty (s_n - s_{n-1}) r^n = \sum_{n=0}^\infty s_n r^n - \sum_{n=0}^\infty s_n r^{n+1} \\ &= (1-r)\sum_{n=0}^\infty s_n r^n = (1-r)\sum_{k=0}^N r^k s_k + \sum_{k=N+1}^\infty (1-r) r^k s_k. \end{aligned}$$

Let $\epsilon > 0$ be given, and choose N such that $k \geq N$ implies $|s_k - s| < \epsilon$. Then

$$\begin{aligned}|\sigma_r - s| &\leq (1-r)\sum_{k=0}^{N}(|s_k| + |s|) + \sum_{k=N+1}^{\infty}(1-r)r^k|s_k - s| \\ &\leq (1-r)\sum_{k=0}^{N}(|s_k| + |s|) + \Big((1-r)\sum_{k=N+1}^{\infty} r^k\Big)\epsilon \\ &\leq (1-r)\sum_{k=0}^{N}(|s_k| + |s|) + \epsilon.\end{aligned}$$

With N fixed, the coefficient of $(1-r)$ in the first term is fixed, and thus we can choose r_0 close enough to 1 so that the first term is $< \epsilon$ whenever $r_0 \leq r < 1$. If $r_0 \leq r < 1$, we see that $|\sigma_r - s| < 2\epsilon$. Since ϵ is arbitrary, the theorem follows.

EXAMPLE. For $|x| < 1$, the geometric series $\sum_{n=0}^{\infty}(-1)^n x^n$ converges and has sum $(1+x)^{-1}$. The Fundamental Theorem of Calculus gives $\log(1+t) = \int_0^x \frac{1}{1+t}\,dt = \int_0^x \sum_{n=0}^{\infty}(-1)^n t^n\,dt$ for $|x| < 1$, and Theorem 1.31 allows us to interchange sum and integral as long as $|x| < 1$. Consequently

$$\log(1+x) = \sum_{n=0}^{\infty}\frac{(-1)^n x^{n+1}}{n+1}$$

for $|x| < 1$. The sequence of partial sums on the right converges for $x = 1$ by the Leibniz test, and Theorem 1.48 says that the Abel sums must converge to the same limit. But the Abel sums have limit $\lim_{x\uparrow 1}\log(1+x) = \log 2$, since $\log(1+x)$ is continuous for $x > 0$. Thus Abel's Theorem has given us a rigorous proof of the familiar identity

$$\sum_{n=0}^{\infty}\frac{(-1)^n}{n+1} = \log 2.$$

Theorems 1.47 and 1.48, which say that one kind of convergence always implies another, are called **Abelian theorems**. Converse results, saying that the second kind of convergence implies the first under an additional hypothesis, are called **Tauberian theorems**. These tend to be harder to prove. We give two examples of Tauberian theorems; the first one will be applied immediately to yield an important special case of the main theorem of Section 9; the second one will be used in Chapter VI to prove a deep theorem about pointwise convergence of Fourier series.

Proposition 1.49. Let $\{a_n\}_{n\geq 0}$ be a complex sequence with all terms ≥ 0, and let $\{s_n\}_{n\geq 0}$ be the sequence of partial sums with $s_n = \sum_{k=0}^n a_k$. If $\{s_n\}_{n\geq 0}$ is Abel summable in $\mathbb{C}$ to the limit s, then $\{s_n\}$ is convergent to the limit s.

PROOF. Let $\{r_j\}_{j\geq 0}$ be a sequence increasing to the limit 1. Since $a_n r_j^n \geq 0$ is nonnegative and since it is monotone increasing in j for each n, Corollary 1.14 applies and gives $\lim_j \sum_{n=0}^\infty a_n r_j^n = \sum_{n=0}^\infty \lim_j a_n r_j^n$, the limits existing in $\mathbb{R}^*$. The left side is the (finite) limit s of the Abel sums, and the right side is $\lim s_n$, which Corollary 1.14 is asserting exists.

EXAMPLE. The binomial series expansion in Section 7 shows, for any complex p, that $(1-r)^p$ is given for $-1 < r < 1$ by the absolutely convergent series

$$(1-r)^p = 1 + \sum_{n=1}^\infty (-1)^n \, \frac{p(p-1)\cdots(p-n+1)}{n!}\, r^n.$$

For p real with $0 < p < 1$, inspection shows that all the coefficients in the sum on the right are ≤ 0. Therefore

$$1-(1-r)^p = \sum_{n=1}^\infty (-1)^{n+1} \, \frac{p(p-1)\cdots(p-n+1)}{n!}\, r^n \qquad (*)$$

has all coefficients ≥ 0 if $0 < p < 1$. For $0 \leq r < 1$, the sum of the series is $1-(1-r)^p$ and is ≥ 0. The fact that $\lim_{r\uparrow 1}[1-(1-r)^p] = 1$ means that the sequence of partial sums $s_k = \sum_{k=1}^\infty (-1)^{k+1} \frac{p(p-1)\cdots(p-k+1)}{k!}$ is Abel summable to 1. Proposition 1.49 shows that the series $(*)$ is convergent at $r = 1$, and the Weierstrass M test shows that $(*)$ converges uniformly for $-1 \leq r \leq 1$ to $1-(1-r)^p$. If we now take $p = \frac{1}{2}$, we have

$$\begin{aligned}
(1-r)^{1/2} &= 1 - \sum_{n=1}^\infty (-1)^{n+1} \, \frac{\frac{1}{2}(-\frac{1}{2})(-\frac{3}{2})\cdots(\frac{3}{2}-n)}{n!}\, r^n \\
&= \sum_{n=1}^\infty (-1)^{n+1} \, \frac{\frac{1}{2}(-\frac{1}{2})(-\frac{3}{2})\cdots(\frac{3}{2}-n)}{n!} \\
&\quad - \sum_{n=1}^\infty (-1)^{n+1} \, \frac{\frac{1}{2}(-\frac{1}{2})(-\frac{3}{2})\cdots(\frac{3}{2}-n)}{n!}\, r^n \\
&= \sum_{n=1}^\infty (-1)^{n+1} \, \frac{\frac{1}{2}(-\frac{1}{2})(-\frac{3}{2})\cdots(\frac{3}{2}-n)}{n!}\, (1-r^n),
\end{aligned}$$

the series on the right being uniformly convergent for $-1 \leq r \leq 1$. Putting $r = 1-x^2$ therefore gives

$$|x| = \sqrt{x^2} = \sum_{n=1}^\infty (-1)^{n+1} \, \frac{\frac{1}{2}(-\frac{1}{2})(-\frac{3}{2})\cdots(\frac{3}{2}-n)}{n!}\, \big(1-(1-x^2)^n\big),$$

the series on the right being uniformly convergent for $-1 \le x \le 1$. Consequently $|x|$ is the uniform limit of a sequence of polynomials on $[-1, 1]$, and all these polynomials are in fact 0 at $x = 0$.

Proposition 1.50. Let $\{a_n\}_{n\ge0}$ be a complex sequence, and let $\{s_n\}_{n\ge0}$ be the sequence of partial sums with $s_n = \sum_{k=0}^n a_k$. If $\{s_n\}$ is Cesàro summable to the limit s in $\mathbb{C}$ and if the sequence $\{na_n\}$ is bounded, then $\{s_n\}$ is convergent and the limit is s. The rate of convergence depends only on the bound for $\{na_n\}$ and the rate of convergence of the Cesàro sums.

REMARK. In our application in Chapter VI to pointwise convergence of Fourier series, the sequence of partial sums will be of the form $\{s_n(x)\}$, depending on a parameter x, and the statement about the rate of convergence will enable us to see that the convergence of $\{s_n(x)\}$ is uniform in x under suitable hypotheses.

PROOF. Let $\{s_n\}$ be the sequence of partial sums of $\{a_n\}$, and choose M such that $|na_n| \le M$ for all n. The first step is to establish a useful formula for $s_n - \sigma_n$. Let m be any integer with $0 \le m < n$. We start from the trivial identity $-(n-m)\sigma_n = (m+1)\sigma_n - (n+1)\sigma_n$, add $(n-m)s_n$ to both sides, and regroup as

$$\begin{aligned}(n-m)(s_n - \sigma_n) &= (m+1)\sigma_n - s_0 - \cdots - s_m + (n-m)s_n - s_{m+1} - \cdots - s_n \\ &= (m+1)(\sigma_n - \sigma_m) + \sum_{j=m+1}^{n} (s_n - s_j).\end{aligned}$$

Dividing by $(n-m)$ yields

$$s_n - \sigma_n = \frac{m+1}{n-m}(\sigma_n - \sigma_m) + \frac{1}{n-m}\sum_{j=m+1}^{n}(s_n - s_j),$$

which is the identity from which the estimates begin.

For $m + 1 \le j \le n$, we have

$$\begin{aligned}|s_n - s_j| &\le |a_n| + |a_{n-1}| + \cdots + |a_{j+1}| \le \frac{M}{n} + \frac{M}{n-1} + \cdots + \frac{M}{j+1} \\ &\le \frac{M}{j+1} + \frac{M}{j+1} + \cdots + \frac{M}{j+1} = \frac{(n-j)M}{j+1} \le \frac{(n-m-1)M}{m+2}.\end{aligned}$$

Substituting into our identity yields

$$|s_n - \sigma_n| \le \frac{m+1}{n-m}\left|\sigma_n - \sigma_m\right| + \frac{(n-m-1)M}{m+2}.$$

Let $\epsilon > 0$ be given, and choose N such that $|\sigma_k - s| \le \epsilon^2$ whenever $k \ge N$. We may assume that $\epsilon < \frac{1}{2}$ and $N \ge 4$. With ϵ fixed and with n fixed to be $\ge 2N$, define m to be the unique integer with

$$m \le \frac{n-\epsilon}{1+\epsilon} < m+1.$$

Then $0 \le m < n$, and our inequality for $|s_n - \sigma_n|$ applies. From the left inequality $m \le \frac{n-\epsilon}{1+\epsilon}$ defining m, we obtain $m + m\epsilon \le n - \epsilon$ and hence $(m+1)\epsilon \le n - m$ and $\frac{m+1}{n-m} \le \epsilon^{-1}$. From the right inequality $\frac{n-\epsilon}{1+\epsilon} < m+1$ defining m, we obtain $n - \epsilon < m + 1 + \epsilon m + \epsilon$ and hence $n - m - 1 < \epsilon(m+2)$ and $\frac{n-m-1}{m+2} < \epsilon$. Thus our main inequality becomes

$$|s_n - \sigma_n| \le \epsilon^{-1}|\sigma_n - \sigma_m| + M\epsilon.$$

To handle σ_m, we need to bound m below. We have seen that $n - m - 1 < \epsilon(m+2)$, and we have assumed that $\epsilon < \frac{1}{2}$. Then $n - m - 1 < \frac{1}{2}(m+2)$, and this simplifies to $m > \frac{2n}{3} - \frac{4}{3}$, which is $\ge \frac{n}{2}$ if $n \ge 8$, thus certainly if $N \ge 4$. In other words, $N \ge 4$ and $n \ge 2N$ makes $m \ge \frac{n}{2} \ge N$. Therefore $|\sigma_m - s| < \epsilon^2$, and $|\sigma_n - \sigma_m| < 2\epsilon^2$. Substituting into our main inequality, we obtain

$$|s_n - \sigma_n| < \epsilon^{-1} 2\epsilon^2 + M\epsilon = (M+2)\epsilon.$$

Since ϵ is arbitrary, the proof is complete.

9. Weierstrass Approximation Theorem

We saw as an application of Proposition 1.49 that the function $|x|$ on $[-1, 1]$ is the uniform limit of an explicit sequence $\{P_n\}$ of polynomials with $P_n(0) = 0$. This is a special case of a theorem of Weierstrass that any continuous complex-valued function on a bounded interval is the uniform limit of polynomials on the interval.

The device for proving the Weierstrass theorem for a general continuous complex-valued function is to construct the approximating polynomials as the result of a smoothing process, known as the use of an "approximate identity." The idea of an approximate identity is an important one in analysis and will occur several times in this book. If f is the given function, the smoothing is achieved by "convolution"

$$\int f(x-t)\varphi(t)\,dt$$

of f with some function φ, the integrals being taken over some particular intervals. The resulting function of x from the convolution turns out to be as "smooth" as the smoother of f and φ. In the case of the Weierstrass theorem, the function

φ will be a polynomial, and we shall arrange parameters so that the convolution will automatically be a polynomial.

To see how a polynomial $\int f(x-t)\varphi(t)\,dt$ might approximate f, one can think of φ as some kind of mass distribution; the mass is all nonnegative if $\varphi \geq 0$. The integration produces a function of x that is the "average" of translates $x \mapsto f(x-t)$ of f, the average being computed according to the mass distribution φ. If φ has total mass 1, i.e., total integral 1, and most of the mass is concentrated near $t = 0$, then f is being replaced essentially by an average of its translates, most of the translates being rather close to f, and we can expect the result to be close to f.

For the Weierstrass theorem, we use a single starting φ_1 at stage 1, namely $c_1(1-x^2)$ on $[-1, 1]$ with c_1 chosen so that the total integral is 1. The graph of φ_1 is a familiar inverted parabola, with the appearance of a bump centered at the origin. The function at stage n is $c_n(1-x^2)^n$, with c_n chosen so that the total integral is 1. Graphs for $n = 3$ and $n = 30$ appear in Figure 1.1. The bump near the origin appears to be more pronounced at n increases, and what we need to do is to translate the above motivation into a proof.

Lemma 1.51. If c_n is chosen so that $c_n \int_{-1}^{1} (1-x^2)^n\,dx = 1$, then $c_n \leq e\sqrt{n}$ for n sufficiently large.

PROOF. We have

$$c_n^{-1} = \int_{-1}^{1} (1-x^2)^n\,dx \geq \int_{-1/\sqrt{n}}^{1/\sqrt{n}} (1-x^2)\,dx = 2\int_0^{1/\sqrt{n}} (1-x^2)^n\,dx$$
$$\geq 2\int_0^{1/\sqrt{n}} (1-\tfrac{1}{n})^n\,dx = 2(1-\tfrac{1}{n})^n \Big/ \sqrt{n}.$$

Since $(1-\frac{1}{n})^n \to e^{-1}$, we have $(1-\frac{1}{n})^n \geq \frac{1}{2}e^{-1}$ for n large enough (actually for $n \geq 2$). Therefore $c_n^{-1} \geq e^{-1}/\sqrt{n}$ for n large enough, and hence $c_n \leq e\sqrt{n}$ for n large enough. This proves the lemma.

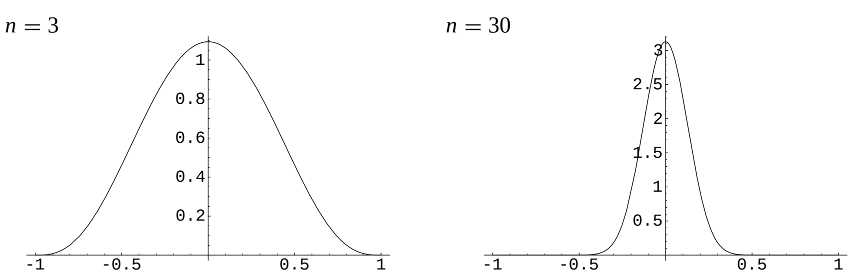

FIGURE 1.1. Approximate identity. Graphs of $c_n \int_{-1}^{1} (1-x^2)^n\,dx$ for for $n = 3$ and $n = 30$ with different scales used on the vertical axes.

Let $\varphi_n(x) = c_n(1-x^2)^n$ on $[-1, 1]$, with c_n as in the lemma. The polynomials φ_n have the following properties:

(i) $\varphi_n(x) \geq 0$,
(ii) $\int_{-1}^{1} \varphi_n(x)\,dx = 1$,
(iii) for any $\delta > 0$, $\sup\{\varphi_n(x) \mid \delta \leq x \leq 1\}$ tends to 0 as n tends to infinity.

Lemma 1.51 is used to verify (iii): the quantity

$$\sup\{\varphi_n(x) \mid \delta \leq x \leq 1\} = c_n(1-\delta^2)^n$$

tends to 0 because $\lim_n \sqrt{n}(1-\delta^2)^n = 0$. A function with the above three properties will be called an **approximate identity** on $[-1, 1]$.

Theorem 1.52 (Weierstrass Approximation Theorem). Any complex-valued continuous function on a bounded interval $[a, b]$ is the uniform limit of a sequence of polynomials.

PROOF. In order to arrange for the convolution to be a polynomial, we need to make some preliminary normalizations. Approximating $f(x)$ on $[a, b]$ by $P(x)$ uniformly within ϵ is the same as approximating $f(x+a)$ on $[0, b-a]$ by $P(x+a)$ uniformly within ϵ, and approximating $g(x)$ on $[0, c]$ uniformly by $Q(x)$ is the same as approximating $g(cx)$ uniformly by $Q(cx)$. Thus we may assume without loss of generality that $[a, b] = [0, 1]$.

If $h : [0, 1] \to \mathbb{C}$ is continuous and if r is the function defined by $r(x) = h(x) - h(0) - [h(1) - h(0)]x$, then r is continuous with $r(0) = r(1) = 0$. Approximating $h(x)$ on $[0, 1]$ uniformly by $R(x)$ is the same as approximating $r(x)$ on $[0, 1]$ uniformly by $R(x) - h(0) - [h(1) - h(0)]x$. Thus we may assume without loss of generality that the function to be approximated has value 0 at 0 and 1.

Let $f : [0, 1] \to \mathbb{C}$ be a given continuous function with $f(0) = f(1) = 0$; the function f is uniformly continuous by Theorem 1.10. We extend f to the whole line by making it be 0 outside $[0, 1]$, and the uniform continuity is maintained. Now let φ_n be the polynomial above, and put $P_n(x) = \int_{-1}^{1} f(x-t)\varphi_n(t)\,dt$.

Let us see that P_n is a polynomial. By our definition of the extended f, the integrand is 0 for a particular $x \in [0, 1]$ unless t is in $[x-1, x]$ as well as $[-1, 1]$. We change variables, replacing t by $s + x$ and making use of Theorem 1.34, and the integral becomes $P_n(x) = \int f(-s)\varphi_n(s+x)\,ds$, the integral being taken for s in $[-1, 0] \cap [-1-x, 1-x]$. Since x is in $[0, 1]$, the condition on s is that s is in $[-1, 0]$. Thus $P_n(x) = \int_{-1}^{0} f(-s)\varphi_n(s+x)\,ds$. In this integral, $\varphi_n(x)$ is a linear combination of monomials x^k, and x^k itself contributes $\int_{-1}^{0} f(-s)(x+s)^k\,ds$, which expands out to be a polynomial in x. Thus $P_n(x)$ is a polynomial in x.

By property (ii) of φ_n, we have

$$P_n(x) - f(x) = \int_{-1}^{1} f(x-t)\varphi_n(t)\,dt - f(x) = \int_{-1}^{1} [f(x-t) - f(x)]\,\varphi_n(t)\,dt.$$

Then property (i) gives

$$\begin{aligned}|P_n(x) - f(x)| &\le \int_{-1}^{1} |f(x-t) - f(x)|\,\varphi_n(t)\,dt \\ &= \int_{-\delta}^{\delta} |f(x-t) - f(x)|\,\varphi_n(t)\,dt + \Big(\int_{-1}^{-\delta} + \int_{\delta}^{1}\Big)|f(x-t) - f(x)|\,\varphi_n(t)\,dt,\end{aligned}$$

and two further uses of property (ii) show that this is

$$\le \sup_{|t|\le\delta} |f(x-t) - f(x)| + 4\Big(\sup_{y\in[0,1]} |f(y)|\Big)\Big(\sup_{\delta\le|t|\le1} \varphi_n(t)\Big).$$

Given $\epsilon > 0$, we choose some δ of uniform continuity for f and ϵ, and then the first term is $\le \epsilon$. With δ fixed, we use property (iii) of φ_n and the boundedness of f, given by Theorem 1.11, to produce an integer N such that the second term is $< \epsilon$ for $n \ge N$. Then $n \ge N$ implies that the displayed expression is $< 2\epsilon$. Since ϵ is arbitrary, P_n converges uniformly to f.

10. Fourier Series

A **trigonometric series** is a series of the form $\sum_{n=-\infty}^{\infty} c_n e^{inx}$ with complex coefficients. The individual terms of the series thus form a doubly infinite sequence, but the sequence of partial sums is always understood to be the sequence $\{s_N\}_{N=0}^{\infty}$ with $s_N(x) = \sum_{n=-N}^{N} c_n e^{inx}$. Such a series may also be written as

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

by putting

$$\left.\begin{aligned} e^{inx} &= \cos nx + i \sin nx \\ e^{-inx} &= \cos nx - i \sin nx \end{aligned}\right\} \quad \text{for } n > 0,$$

$$c_0 = \tfrac{1}{2}a_0, \quad c_n = \tfrac{1}{2}(a_n - ib_n), \quad \text{and} \quad c_{-n} = \tfrac{1}{2}(a_n + ib_n) \quad \text{for } n > 0.$$

Historically the notation with the a_n's and b_n's was introduced first, but the use of complex exponentials has become quite common. Nowadays the notation with

a_n's and b_n's tends to be used only when a function f under investigation is real-valued or when all the cosine terms are absent (i.e., f is even) or all the sine terms are absent (i.e., f is odd).

Power series enable us to enlarge our repertory of explicit functions, and the same thing is true of trigonometric series. Just as the coefficients of a power series whose sum is a function f have to be those arising from Taylor's formula for f, the coefficients of a trigonometric series formed from a function have to arise from specific formulas. Let us run through the relevant formal computation: First we observe that the partial sums have to be periodic with period 2π. The question then is the extent to which a complex-valued periodic function f on the real line can be given by a trigonometric series. Suppose that

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx}.$$

Multiply by e^{-ikx} and integrate to get

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-ikx}\,dx = \frac{1}{2\pi}\int_{-\pi}^{\pi} \sum_{n=-\infty}^{\infty} c_n e^{inx}e^{-ikx}\,dx.$$

If we can interchange the order of the integration and the infinite sum, e.g., if the trigonometric series is uniformly convergent to $f(x)$, the right side is

$$= \sum_{n=-\infty}^{\infty} c_n \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{inx}e^{-ikx}\,dx = \sum_{n=-\infty}^{\infty} c_n \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(n-k)x}\,dx = c_k$$

because

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{imx}\,dx = \begin{cases} 1 & \text{if } m = 0, \\ 0 & \text{if } m \neq 0. \end{cases}$$

Let f be Riemann integrable on $[-\pi, \pi]$, and regard f as periodic on $\mathbb{R}$. The trigonometric series $\sum_{n=-\infty}^{\infty} c_n e^{inx}$ with

$$c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}\,dx$$

is called the **Fourier series** of f. We write

$$f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx} \qquad \text{and} \qquad s_N(f;x) = \sum_{n=-N}^{N} c_n e^{inx}.$$

The numbers c_n are the **Fourier coefficients** of f, and the functions $s_N(f;x)$ are the partial sums of the Fourier series. The symbol $\sim$ is to be read as "has Fourier series," nothing more, at least initially. The formulas for the coefficients when the Fourier series is written with sines and cosines are

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx \qquad \text{for } n \ge 0,$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx \qquad \text{for } n \ge 1.$$

In applications one encounters periodic functions of periods other than 2π. If f is periodic of period $2l$, then the Fourier series of f is $f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{in\pi x/l}$ with $c_n = (2l)^{-1}\int_{-l}^{l} f(x)e^{-in\pi x/l}\,dx$. The formula for the series written with sines and cosines is $f(x) \sim a_0/2 + \sum_{n=1}^{\infty}(a_n\cos(n\pi x/l) + b_n\sin(n\pi x/l))$ with $a_n = l^{-1}\int_{-l}^{l} f(x)\cos(n\pi x/l)\,dx$ and $b_n = l^{-1}\int_{-l}^{l} f(x)\sin(n\pi x/l)\,dx$. In the present section of the text, we shall assume that our periodic functions have period 2π.

The result implicit in the formal computation above is that if $f(x)$ is the sum of a uniformly convergent trigonometric series, then the trigonometric series is the Fourier series of f, by Theorem 1.31.

We ask two questions: When does a general Fourier series converge? If the Fourier series converges, to what extent does the sum represent f? We begin with an illuminating example that brings together a number of techniques from this chapter.

EXAMPLE. As in the example following Theorem 1.48, we have

$$\log\left(\frac{1}{1-x}\right) = x + \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 + \cdots \qquad \text{for } -1 < x < 1.$$

We would like to extend this identity to complex z with $|z| < 1$ but do not want to attack the problem of making sense out of log as a function of a complex variable. What we do is apply exp to both sides and obtain an identity for which both sides make sense when the real x is replaced by a complex z:

$$\exp\left(z + \tfrac{1}{2}z^2 + \tfrac{1}{3}z^3 + \cdots\right) = \frac{1}{1-z} \qquad \text{for } |z| < 1.$$

In fact, this identity is valid for z complex with $|z| < 1$, and Problems 30–35 at the end of the chapter lead to a proof of it. Corollary 1.45 allows us to write $z = re^{i\theta}$ and $z + \frac{1}{2}z^2 + \frac{1}{3}z^3 + \cdots = \rho e^{i\varphi}$. Equating real and imaginary parts of the latter equation gives us

$$\rho\cos\varphi = \sum_{n=1}^{\infty}\frac{r^n\cos n\theta}{n} \qquad \text{and} \qquad \rho\sin\varphi = \sum_{n=1}^{\infty}\frac{r^n\sin n\theta}{n}.$$

We shall compute the left sides of these displayed equations in another way. We have

$$e^{\rho\cos\varphi}e^{i\rho\sin\varphi} = \exp(\rho\cos\varphi + i\rho\sin\varphi) = \exp(\rho e^{i\varphi}) = (1-z)^{-1}$$

and therefore also $e^{\rho\cos\varphi}e^{-i\rho\sin\varphi} = (1-\bar{z})^{-1}$. Thus

$$e^{2\rho\cos\varphi} = (1-z)^{-1}(1-\bar{z})^{-1} = (1-re^{i\theta})^{-1}(1-re^{-i\theta})^{-1} = (1-2r\cos\theta+r^2)^{-1}.$$

Taking log of both sides gives $2\rho\cos\varphi = \log\big((1-2r\cos\theta+r^2)^{-1}\big)$, and thus we have

$$\tfrac{1}{2}\log\Big(\frac{1}{1-2r\cos\theta+r^2}\Big) = \sum_{n=1}^{\infty}\frac{r^n\cos n\theta}{n}. \qquad (*)$$

Handling $\rho\sin\varphi$ is a little harder. From $e^{\rho\cos\varphi}e^{i\rho\sin\varphi} = (1-z)^{-1}$, we have $e^{i\rho\sin\varphi} = (1-z)^{-1}/|1-z|^{-1} = (1-\bar{z})/|1-z| = \frac{1-r\cos\theta}{|1-z|} + i\frac{r\sin\theta}{|1-z|}$, and hence

$$\cos(\rho\sin\varphi) = (1-r\cos\theta)/|1-z| \quad \text{and} \quad \sin(\rho\sin\varphi) = (r\sin\theta)/|1-z|.$$

Thus $\tan(\rho\sin\varphi) = r\sin\theta/(1-r\cos\theta)$. Since $1-r\cos\theta$ is >0, $\cos(\rho\sin\varphi)$ is >0, and $\rho\sin\varphi = \arctan\big((r\sin\theta)/(1-r\cos\theta)\big) + 2\pi N(r,\theta)$ for some integer $N(r,\theta)$ depending on r and θ. Hence

$$\arctan\big((r\sin\theta)/(1-r\cos\theta)\big) + 2\pi N(r,\theta) = \sum_{n=1}^{\infty}\frac{r^n\sin n\theta}{n}.$$

For fixed r, the first term on the left is continuous in θ, and the series on the right is uniformly convergent by the Weierstrass M test. By Theorem 1.21 the right side is continuous in θ. Thus $N(r,\theta)$ is continuous in θ for fixed r; since $N(r,0)=0$, $N(r,\theta)=0$ for all r and θ. We conclude that

$$\arctan\Big(\frac{r\sin\theta}{1-r\cos\theta}\Big) = \sum_{n=1}^{\infty}\frac{r^n\sin n\theta}{n}. \qquad (**)$$

Problem 15 at the end of the chapter observes that the partial sums $\sum_{n=1}^{N}\cos n\theta$ and $\sum_{n=1}^{N}\sin n\theta$ are uniformly bounded on any set $\epsilon \le \theta < \pi-\epsilon$ if $\epsilon > 0$. Corollary 1.19 therefore shows that the series

$$\sum_{n=1}^{\infty}\frac{\cos n\theta}{n} \qquad \text{and} \qquad \sum_{n=1}^{\infty}\frac{\sin n\theta}{n}$$

are uniformly convergent for $\epsilon \le \theta < \pi-\epsilon$ if $\epsilon > 0$. Abel's Theorem (Theorem 1.48) shows that each of these series is therefore Abel summable with the same limit. We can tell what the latter limits are from $(*)$ and $(**)$, and thus we conclude that

$$\tfrac{1}{2}\log\left(\frac{1}{2-2\cos\theta}\right)=\sum_{n=1}^{\infty}\frac{\cos n\theta}{n}$$

and

$$\arctan\left(\frac{\sin\theta}{1-\cos\theta}\right)=\sum_{n=1}^{\infty}\frac{\sin n\theta}{n},$$

The sum of the series with the cosine terms is unbounded near $\theta = 0$, and Riemann integration is not meaningful with it. We shall not be able to analyze this series further until we can treat the left side in Chapter VI by means of Lebesgue integration. The sum of the series with the sine terms is written in a way that stresses its periodicity. On the interval $[-\pi, \pi]$, we can rewrite its left side as $\frac{1}{2}(-\pi-\theta)$ for $-\pi \le \theta < 0$, 0 for $\theta = 0$, and $\frac{1}{2}(\pi-\theta)$ for $0 < \theta \le \pi$. The expression for the left side is nicer on the interval $(0, 2\pi)$, and there we have

$$\tfrac{1}{2}(\pi-\theta)=\sum_{n=1}^{\infty}\frac{\sin n\theta}{n}\qquad \text{for } 0<\theta<2\pi.$$

The function $\frac{1}{2}(\pi-\theta)$ is bounded on $(0, 2\pi)$, and we can readily compute its Fourier coefficients from the formula $b_n = \pi^{-1}\int_0^{2\pi}\frac{1}{2}(\pi-\theta)\sin n\theta\, d\theta$, using integration by parts (Corollary 1.33). The result is that $b_n = 1/n$. Hence the displayed series is the Fourier series. Graphs of some of the partial sums appear in Figure 1.2.

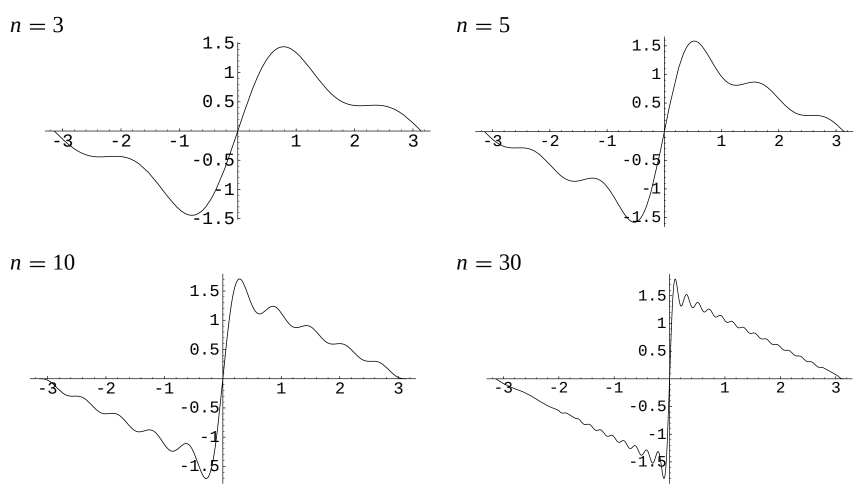

FIGURE 1.2. Fourier series of sawtooth function. Graphs of $\sum_{n=1}^{N}(\sin nx)/n$ for $n = 3, 5, 10, 30$.

The sawtooth function in the above example has a discontinuity, and yet its Fourier series converges to it pointwise. The recognition of the remarkable potential that Fourier series have for representing discontinuous functions dates to Joseph Fourier himself and caused many of Fourier's contemporaries to doubt the validity of his work.

Although the above Fourier series converges to the function, it cannot do so uniformly, as a consequence of Theorem 1.21. In any such situation the Fourier coefficients cannot decrease rapidly, and a decrease of order $1/n$ is the best that one gets for a nice function with a jump discontinuity.

This example points to a general heuristic principle contrasting how power series and trigonometric series behave: whereas Taylor series converge very rapidly and may not converge to the function, Fourier series are inclined to converge rather slowly and they are more likely to converge to the function.

We come to convergence results in a moment. First we establish some elementary properties of them. Taking the absolute value of c_n in the definition of Fourier coefficient, we obtain the trivial bound $|c_n| \le \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|\,dx$.

Theorem 1.53. Let f be in $\mathcal{R}[-\pi,\pi]$. Among all choices of $d_{-N},\dots,d_N$, the expression

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\Big|f(x)-\sum_{n=-N}^{N} d_n e^{inx}\Big|^2\,dx$$

is minimized uniquely by choosing d_n, for all n with $|n|\le N$, to be the Fourier coefficient $c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}\,dx$. The minimum value is

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|f(x)|^2\,dx - \sum_{n=-N}^{N}|c_n|^2.$$

PROOF. Put $d_n = c_n + \varepsilon_n$. Then

$$\begin{aligned}
&\tfrac{1}{2\pi}\textstyle\int_{-\pi}^{\pi}\big|f(x)-\sum_{n=-N}^{N} d_n e^{inx}\big|^2\,dx\\
&\qquad= \tfrac{1}{2\pi}\textstyle\int_{-\pi}^{\pi}|f(x)|^2\,dx - \tfrac{1}{2\pi}\,2\,\mathrm{Re}\sum_{n=-N}^{N}\overline{d_n}\int_{-\pi}^{\pi} f(x)e^{-inx}\,dx\\
&\qquad\quad+\tfrac{1}{2\pi}\textstyle\int_{-\pi}^{\pi}\sum_{m,n=-N}^{N} d_m\overline{d_n}e^{i(m-n)x}\,dx\\
&\qquad= \tfrac{1}{2\pi}\textstyle\int_{-\pi}^{\pi}|f(x)|^2\,dx - 2\,\mathrm{Re}\sum_{n=-N}^{N} c_n\overline{d_n} + \sum_{n=-N}^{N}|d_n|^2\\
&\qquad= \big(\tfrac{1}{2\pi}\textstyle\int_{-\pi}^{\pi}|f(x)|^2\,dx\big) - \big(2\sum_{n=-N}^{N}|c_n|^2 + 2\,\mathrm{Re}\sum_{n=-N}^{N} c_n\overline{\varepsilon_n}\big)\\
&\qquad\quad+\big(\textstyle\sum_{n=-N}^{N}|c_n|^2 + 2\,\mathrm{Re}\sum_{n=-N}^{N} c_n\overline{\varepsilon_n} + \sum_{n=-N}^{N}|\varepsilon_n|^2\big)\\
&\qquad= \tfrac{1}{2\pi}\textstyle\int_{-\pi}^{\pi}|f(x)|^2 - \sum_{n=-N}^{N}|c_n|^2 + \sum_{n=-N}^{N}|\varepsilon_n|^2.
\end{aligned}$$

The result follows.

Corollary 1.54 (Bessel's inequality). Let f be in $\mathcal{R}[-\pi, \pi]$, and let $f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{inx}$. Then

$$\sum_{n=-\infty}^{\infty} |c_n|^2 \le \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 \, dx.$$

In particular, $\sum_{n=-\infty}^{\infty} |c_n|^2$ is finite.

REMARK. In terms of the coefficients a_n and b_n, the corresponding result is

$$\frac{|a_0|^2}{2} + \sum_{n=1}^{\infty} \big(|a_n|^2 + |b_n|^2\big) \le \frac{1}{\pi} \int_{-\pi}^{\pi} |f(x)|^2 \, dx.$$

PROOF. The theorem shows that the minimum value of a certain nonnegative quantity depending on N is $\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 \, dx - \sum_{n=-N}^{N} |c_n|^2$. Thus, for any N, $\sum_{n=-N}^{N} |c_n|^2 \le \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 \, dx$. Letting N tend to infinity, we obtain the corollary.

Corollary 1.55 (Riemann–Lebesgue Lemma). If f is in $\mathcal{R}[-\pi, \pi]$ and has Fourier coefficients $\{c_n\}_{n=-\infty}^{\infty}$, then $\lim_{|n|\to\infty} c_n = 0$.

REMARK. This improves on the inequality $|c_n| \le \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)| \, dx$ observed above, which shows, by means of an explicit estimate, that $\{c_n\}$ is a bounded sequence.

PROOF. This is immediate from Corollary 1.54.

We now turn to convergence results. First it is necessary to clarify terms like "continuous" and "differentiable" in the context of Fourier series of functions on $[-\pi, \pi]$. Each term of a Fourier series is defined on all of $\mathbb{R}$ and is periodic with period 2π and is really given as the restriction to $[-\pi, \pi]$ of this periodic function. Thus it makes sense to regard a general function in the same way if one wants to form its Fourier series: a function f is extended to all of $\mathbb{R}$ so as to be periodic with period 2π, and if we consider f on $[-\pi, \pi]$, it is really the restriction to $[-\pi, \pi]$ that we are considering.

In particular, it makes sense to insist that $f(-\pi) = f(\pi)$; if f does not have this property initially, one or both of these endpoint values will have to be adjusted, but that adjustment will not affect any Fourier coefficients. Similarly continuity of f will refer to continuity of the extended function on all of $\mathbb{R}$, and similarly for differentiability.

That being said, let us take up the matter of integration by parts for the functions we are considering. The scope of integration by parts in Corollary 1.33 was limited to a pair of functions f and g that have a continuous first derivative. In the context of Fourier series, it is the periodic extensions that are to have these properties, and then the integration-by-parts formula simplifies. Namely,

$$\begin{aligned}\int_{-\pi}^{\pi} f(x)g'(x)\,dx &= \Big[f(x)g(x)\Big]_{-\pi}^{\pi} - \int_{-\pi}^{\pi} f'(x)g(x)\,dx \\ &= -\int_{-\pi}^{\pi} f'(x)g(x)\,dx,\end{aligned}$$

i.e., the integrated term drops out because of the assumed periodicity.

The simplest convergence result for Fourier series is that a periodic function (of period 2π) with two continuous derivatives has a uniformly convergent Fourier series. To prove this, we take $n \neq 0$ and use the above integration-by-parts formula twice to obtain

$$\begin{aligned}c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}\,dx &= -\frac{1}{2\pi}\Big(\frac{1}{-in}\Big)\int_{-\pi}^{\pi} f'(x)e^{-inx}\,dx \\ &= \frac{1}{2\pi}\Big(\frac{1}{-in}\Big)^2\int_{-\pi}^{\pi} f''(x)e^{-inx}\,dx.\end{aligned}$$

Then $|c_n e^{inx}| = |c_n| \leq C/n^2$, where $C = \frac{1}{2\pi}\int_{-\pi}^{\pi} |f''(x)|\,dx$, and the Fourier series converges uniformly by the Weierstrass M test. The argument does not say that the convergence is to f, but that fact will be proved in Theorem 1.57 below.

Adjusting the proof just given, we can prove a sharper convergence result.

Proposition 1.56 . If f is periodic (of period 2π) and has one continuous derivative, then the Fourier series of f converges uniformly.

PROOF. As in the above argument, $c_n = -\frac{1}{2\pi}\big(\frac{1}{-in}\big)\int_{-\pi}^{\pi} f'(x)e^{-inx}\,dx$, and this equals $\frac{1}{in}d_n$, where d_n is the n^{th} Fourier coefficient of the continuous function f'. In the computation that follows, we use the classical Schwarz inequality (as in Section A5 of the appendix) for finite sums and pass to the limit in order to get the first inequality, and then we use Bessel's inequality (Corollary 1.54) to get the second inequality:

$$\begin{aligned}\sum_{n\neq 0}|c_n| = \sum_{n\neq 0}|inc_n|\,\frac{1}{|n|} = \sum_{n\neq 0}\frac{1}{|n|}\,|d_n| &\leq \Big(\sum_{n\neq 0}\frac{1}{n^2}\Big)^{1/2}\Big(\sum_{n\neq 0}|d_n|^2\Big)^{1/2} \\ &\leq \Big(\sum_{n\neq 0}\frac{1}{n^2}\Big)^{1/2}\Big(\frac{1}{2\pi}\int_{-\pi}^{\pi}|f'(x)|^2\,dx\Big)^{1/2}.\end{aligned}$$

The right side is finite, and the proposition follows from the Weierstrass M test.

The fact that the convergence in Proposition 1.56 is actually to f will follow from Dini's test, which is Theorem 1.57 below. We first derive some simple formulas. The **Dirichlet kernel** is the periodic function of period 2π defined by

$$D_N(x) = \sum_{n=-N}^{N} e^{inx} = \frac{\sin\left((N+\frac{1}{2})x\right)}{\sin\frac{1}{2}x},$$

the second equality following from the formula for the sum of a geometric series. For a periodic function f of period 2π, the partial sums of the Fourier series of f are given by

$$\begin{aligned} s_N(f;x) &= \sum_{n=-N}^{N} \Big(\frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}\,dt\Big)e^{inx} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\sum_{n=-N}^{N} e^{in(x-t)}\,dt \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)D_N(x-t)\,dt \\ &= \frac{1}{2\pi}\int_{x-\pi}^{x+\pi} f(x-s)D_N(s)\,ds \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-t)D_N(t)\,dt, \end{aligned}$$

the last two steps following from the changes of variables $t \mapsto x+s$ (Theorem 1.34) and $s \mapsto -s$ (Proposition 1.30h) and from the periodicity of f and D_N.

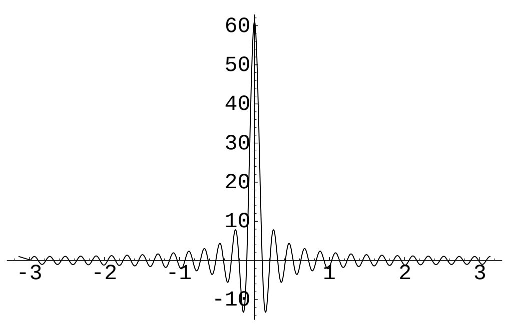

FIGURE 1.3. Dirichlet kernel. Graph of D_N for $N = 30$.

This is the kind of convolution integral that occurred in the previous section. Term-by-term integration shows that $\frac{1}{2\pi}\int_{-\pi}^{\pi} D_N(x)\,dx = 1$. However, D_N is not an approximate identity, not being everywhere ≥ 0. Figure 1.3 shows the graph of D_N for $N = 30$. Although $D_N(x)$ looks small in the graph away from $x = 0$, it is small only as a percentage of $D_N(0)$; $D_N(x)$ does not have $\lim_N D_N(x)$ equal to 0 for $x \neq 0$. Thus $D_N(x)$ fails in a second way to be an approximate identity. The failure of D_N to be an approximate identity is what makes the subject of convergence of Fourier series so subtle.

Theorem 1.57 (Dini's test). Let $f : \mathbb{R} \to \mathbb{C}$ be periodic of period 2π and Riemann integrable on $[-\pi, \pi]$. Fix x in $[-\pi, \pi]$. If there are constants $\delta > 0$ and $M < +\infty$ such that

$$|f(x+t) - f(x)| \leq M|t| \qquad \text{for } |t| < \delta,$$

then $\lim_N s_n(f; x) = f(x)$.

REMARK. This condition is satisfied if f is differentiable at x. Thus the convergence of the Fourier series in Proposition 1.56 is to the original function f. By contrast, the Dini condition is not satisfied at $x = 0$ for the continuous periodic extension of the function $f(x) = |x|^{1/2}$ defined on $(-\pi, \pi]$.

PROOF. With x fixed, let

$$g(t) = \begin{cases} \dfrac{f(x-t) - f(x)}{\sin t/2} & \text{for } 0 < |t| \leq \pi, \\ 0 & \text{for } t = 0. \end{cases}$$

Proposition 1.30d shows that $(\sin t/2)^{-1}$ is Riemann integrable on $\epsilon \leq |t| \leq \pi$ for any $\epsilon > 0$, and hence so is $g(t)$. Since $g(t)$ is bounded near $t = 0$, Lemma 1.28 shows that $g(t)$ is Riemann integrable on $[-\pi, \pi]$. Since $\int_{-\pi}^{\pi} D_N(x)\,dx = 1$, we have

$$\begin{aligned} s_N(f;x) - f(x) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-t)\,\frac{\sin\big((N+\frac{1}{2})t\big)}{\sin\frac{1}{2}t}\,dt - \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,\frac{\sin\big((N+\frac{1}{2})t\big)}{\sin\frac{1}{2}t}\,dt \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} g(t)\sin\big((N+\tfrac{1}{2})t\big)\,dt \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \big[g(t)\cos\tfrac{t}{2}\big]\sin Nt\,dt + \frac{1}{2\pi}\int_{-\pi}^{\pi} \big[g(t)\sin\tfrac{t}{2}\big]\cos Nt\,dt, \end{aligned}$$

and both terms on the right side tend to 0 with N by the Riemann–Lebesgue Lemma (Corollary 1.55).

Dini's test (Theorem 1.57) has implications for "localization" of the convergence of Fourier series. Suppose that $f = g$ on an open interval I, and suppose that the Fourier series of f converges to f on I. Then Dini's test shows that the Fourier series of $f - g$ converges to 0 on I, and hence the Fourier series of g converges to g on I. For example, f could be a function with a continuous derivative everywhere, and g could have discontinuities outside the open interval I. For f, the proof of Proposition 1.56 shows that $\sum |c_n| < +\infty$. But for g, the Fourier series cannot converge so rapidly because the sum of a uniformly convergent series of continuous functions has to be continuous. Thus the two series locally have the same sum, but their qualitative behavior is quite different.

Next let us address the question of the extent to which the Fourier series of f uniquely determines f. Our first result in this direction will be that if f and g are Riemann integrable and have the same respective Fourier coefficients, then $f(x) = g(x)$ at every point of continuity of both f and g. It may look as if some sharpening of Dini's test might apply just under the assumption of continuity of the function, and then this uniqueness result would be trivial. However, as we shall see in Chapter XII, the Fourier series of a continuous function need not converge to the function at particular points, and there can be no such sharpening of Dini's test. Instead, we shall handle the uniqueness question in a more indirect fashion.

The technique is to use an approximate identity, as in the proof of the Weierstrass Approximation Theorem in Section 9. Although the partial sums of the Fourier series of a continuous function need not converge at every point, the Cesàro sums do converge. To get at this fact, we shall examine the **Fejér kernel**

$$K_N(x) = \frac{1}{N+1}\sum_{n=0}^{N} D_n(x).$$

The N^{th} Cesàro sum of $s_n(f;x)$ is given by $\frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(x-t)\,f(t)\,dt$ because

$$\begin{aligned}\frac{1}{N+1}\sum_{n=0}^{N} s_n(f;x) &= \frac{1}{N+1}\sum_{n=0}^{N}\frac{1}{2\pi}\int_{-\pi}^{\pi} D_n(x-t)\,f(t)\,dt\\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(x-t)\,f(t)\,dt.\end{aligned}$$

The remarkable fact is that the Fejér kernel is an approximate identity even though the Dirichlet kernel is not, and the result will be that the Cesàro sums of a Fourier series converge in every way that they have any hope of converging.

Lemma 1.58. The Fejér kernel is given by

$$K_N(x) = \frac{1}{N+1}\,\frac{1-\cos(N+1)x}{1-\cos x}.$$

PROOF. We show by induction on N that the values of $K_N(x)$ in the definition and in the lemma are equal. For $N = 0$, we have $K_0(x) = D_0(x) = 1 = \frac{1-\cos 1x}{1-\cos x}$ as required. Assume the equality for $N - 1$. Then

$$\begin{aligned}
(N+1)K_N(x) &= \sum_{n=0}^{N} D_n(x) = NK_{N-1}(x) + D_N(x) \\
&= \frac{1-\cos Nx}{1-\cos x} + \frac{\sin\big((N+\frac{1}{2})x\big)}{\sin\frac{1}{2}x}\cdot\frac{\sin\frac{1}{2}x}{\sin\frac{1}{2}x} \qquad \text{by induction} \\
&= \frac{1-\cos Nx + 2\sin\big((N+\frac{1}{2})x\big)\sin\frac{1}{2}x}{1-\cos x} \\
&= \frac{1-\cos Nx - \big[\cos\big((N+\frac{1}{2})x+\frac{1}{2}x\big) - \cos\big((N+\frac{1}{2})x-\frac{1}{2}x\big)\big]}{1-\cos x} \\
&= \frac{1-\cos(N+1)x}{1-\cos x},
\end{aligned}$$

as required.

In line with the definition of approximate identity in Section 9, we are to show that $K_N(x)$ has the following properties:

(i) $K_N(x) \geq 0$,
(ii) $\frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(x)\,dx = 1$,
(iii) for any $\delta > 0$, $\sup_{\delta\leq|x|\leq\pi} K_N(x)$ tends to 0 as n tends to infinity.

Property (i) follows from the definition of $K_N(x)$, since $\cos x \leq 1$ everywhere; (ii) follows from the definition of $K_N(x)$ and the linearity of the integral, since $\frac{1}{2\pi}\int_{-\pi}^{\pi} D_n(x)\,dx = 1$ for all n; and (iii) follows from Lemma 1.58, since $1 - \cos(N+1)x \leq 2$ everywhere and $1 - \cos x \geq 1 - \cos\delta$ if $\delta \leq |x| \leq \pi$.

Theorem 1.59 (Fejér's Theorem). Let $f : \mathbb{R} \to \mathbb{C}$ be periodic of period 2π and Riemann integrable on $[-\pi, \pi]$. If f is continuous at a point x_0 in $[-\pi, \pi]$, then

$$\lim_{N\to\infty} \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)K_N(x_0 - x)\,dx = f(x_0).$$

If f is uniformly continuous on a subset E of $[-\pi, \pi]$, then the convergence is uniform for x_0 in E.

PROOF. Choose M such that $|f(x)| \le M$ for all x. By (ii) and then (i),

$$
\begin{aligned}
\Big|\frac{1}{2\pi}\int_{-\pi}^{\pi} & f(x)K_N(x_0-x)\,dx - f(x_0)\Big| \\
&= \Big|\frac{1}{2\pi}\int_{-\pi}^{\pi}[f(x)-f(x_0)]K_N(x_0-x)\,dx\Big| \\
&\le \frac{1}{2\pi}\int_{-\pi}^{\pi}|f(x)-f(x_0)|K_N(x_0-x)\,dx \\
&\le \frac{1}{2\pi}\int_{|x-x_0|\le\delta}|f(x)-f(x_0)|K_N(x_0-x)\,dx \\
&\quad + \frac{1}{2\pi}\int_{\delta\le|x-x_0|\le\pi} 2M\Big(\sup_{\delta\le|t|\le\pi}K_N(t)\Big)\,dx \\
&\le \frac{1}{2\pi}\int_{|x-x_0|\le\delta}|f(x)-f(x_0)|K_N(x_0-x)\,dx + 2M\sup_{\delta\le|t|\le\pi}K_N(t).
\end{aligned}
$$

Given $\epsilon > 0$, choose some δ for ϵ and continuity of f at x_0 or for ϵ and uniform continuity of f on E. In the first term on the right side, we then have $|f(x)-f(x_0)| \le \epsilon$ on the set where $|x-x_0| \le \delta$. Thus a second use of (i) shows that the above expression is

$$\le \epsilon + 2M\sup_{\delta\le|t|\le\pi}K_N(t).$$

With δ fixed, property (iii) shows that the right side is $< 2\epsilon$ if N is sufficiently large, and the theorem follows.

Corollary 1.60 (uniqueness theorem). Let $f : \mathbb{R} \to \mathbb{C}$ and $g : \mathbb{R} \to \mathbb{C}$ be periodic of period 2π and Riemann integrable on $[-\pi, \pi]$. If f and g have the same respective Fourier coefficients, then $f(x) = g(x)$ at every point of continuity of both f and g.

REMARK. The fact that f and g have the same Fourier coefficients means that $s_n(f; x) = s_n(g; x)$ for all n, hence that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} D_n(x-t)f(t)\,dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} D_n(x-t)g(t)\,dt$$

for all n. Then the same formula applies with D_n replaced by its Cesàro sums K_N.

PROOF. Apply Theorem 1.59 to $f - g$ at a point x_0 of continuity of both f and g.

Our second result about uniqueness will improve on Corollary 1.60, saying that any Riemann integrable function with all Fourier coefficients 0 is basically the 0 function—at least in the sense that any definite integral in which it is a factor of the integrand is 0. We shall prove this improved result as a consequence of Parseval's Theorem, which says that equality holds in Bessel's inequality. The proof of Parseval's Theorem will be preceded by an example and some lemmas.

Theorem 1.61 (Parseval's Theorem). Let $f : \mathbb{R} \to \mathbb{C}$ be periodic of period 2π and Riemann integrable on $[-\pi, \pi]$. If $f(x) \sim \sum_{-\infty}^{\infty} c_n e^{inx}$, then

$$\lim_{N\to\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x) - s_N(f; x)|^2 \, dx = 0$$

and

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 \, dx = \sum_{n=-\infty}^{\infty} |c_n|^2.$$

REMARK. In terms of the coefficients a_n and b_n, the corresponding result is

$$\frac{1}{\pi} \int_{-\pi}^{\pi} |f(x)|^2 \, dx = \frac{|a_0|^2}{2} + \sum_{n=1}^{\infty} \big(|a_n|^2 + |b_n|^2\big).$$

EXAMPLE. We saw near the beginning of this section that the periodic function f given by $f(x) = \frac{1}{2}(\pi - x)$ on $(0, 2\pi)$ has $f(x) \sim \sum_{n=1}^{\infty} \frac{\sin nx}{n}$. The formulation of Parseval's Theorem as in the remark, but with the interval $(0, 2\pi)$ replacing the interval $(-\pi, \pi)$, says that $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{1}{\pi} \int_0^{2\pi} \left|\frac{1}{2}(\pi - x)\right|^2 dx$. The right side is $= \frac{1}{4\pi} \int_{-\pi}^{\pi} x^2 \, dx = \frac{2\pi^3/3}{4\pi} = \frac{\pi^2}{6}$. Thus

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

This formula was discovered by Euler by other means before the work of Fourier.

For the purposes of the lemmas and the proof of Parseval's Theorem, let us introduce a "Hermitian inner product"[3] on $\mathcal{R}[-\pi, \pi]$ by the definition

$$(f, g)_2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\overline{g(x)} \, dx,$$

[3]The term "Hermitian inner product" will be defined precisely in Section II.1. The form $(f, g)_2$ comes close to being one, but it fails to meet all the conditions because $(f, f)_2 = 0$ is possible without $f = 0$.

as well as a "norm" defined by

$$\|f\|_2 = (f, f)_2^{1/2} = \Big(\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|^2\,dx\Big)^{1/2}$$

and a "distance function" defined by

$$d_2(f, g) = \|f - g\|_2 = \Big(\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x) - g(x)|^2\,dx\Big)^{1/2}.$$

The role of the function d_2 will become clearer in Chapter II, where "distance functions" of this kind will be studied extensively.

Lemma 1.62. If f is in $\mathcal{R}[-\pi, \pi]$ and $\int_{-\pi}^{\pi} |f(x)|^2\,dx = 0$, then $\int_{-\pi}^{\pi} |f(x)|\,dx = 0$ and also $\int_{-\pi}^{\pi} f(x)\overline{g(x)}\,dx = 0$ for all $g \in \mathcal{R}[-\pi, \pi]$.

PROOF. Write $M = \sup_{x\in[-\pi,\pi]} |f(x)|$, and let $\epsilon > 0$ be given. Choose a partition $P = \{x_i\}_{i=0}^n$ with $U(P, |f|^2) < \epsilon^3$, i.e.,

$$\sum_{i=1}^{n} \Big(\sup_{x\in[x_{i-1},x_i]} |f(x)|^2\Big)\Delta x_i \le \epsilon^3.$$

Divide the indices from 1 to n into two subsets, A and B, with

$$A = \Big\{i \;\Big|\; \sup_{x\in[x_{i-1},x_i]} |f(x)| \ge \epsilon\Big\} \qquad \text{and} \qquad B = \Big\{i \;\Big|\; \sup_{x\in[x_{i-1},x_i]} |f(x)| < \epsilon\Big\}.$$

The sum of the contributions from indices $i \in A$ to $U\big(P, |f|^2\big)$ is $\ge \epsilon^2 \sum_{i\in A} \Delta x_i$, and thus $\sum_{i\in A} \Delta x_i \le \epsilon$. Hence $\sum_{i\in A} \big(\sup_{x\in[x_{i-1},x_i]} |f(x)|\big)\Delta x_i \le M\epsilon$. Also, $\sum_{i\in B} \big(\sup_{x\in[x_{i-1},x_i]} |f(x)|\big)\Delta x_i \le 2\pi\epsilon$. Therefore $U(P, |f|) \le (2\pi + M)\epsilon$. Since ϵ is arbitrary, $\int_{-\pi}^{\pi} |f(x)|\,dx = 0$. This proves the first conclusion.

For the second conclusion it follows from the boundedness of $|g|$, say by M', that $\big|\int_{-\pi}^{\pi} f(x)\overline{g(x)}\,dx\big| \le \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)||g(x)|\,dx \le \frac{M'}{2\pi}\int_{-\pi}^{\pi} |f(x)|\,dx = 0$.

Lemma 1.63 (Schwarz inequality). If f and g are in $\mathcal{R}[-\pi, \pi]$, then

$$|(f, g)_2| \le \|f\|_2\|g\|_2.$$

REMARK. Compare this result with the version of the Schwarz inequality in Section A5 of the appendix. This kind of inequality is put into a broader setting in Section II.1.

PROOF. If $\|g\|_2 = 0$, then Lemma 1.62 shows that $(f, g)_2 = 0$ for all f. Thus the lemma is valid in this case. If $\|g\|_2 \ne 0$, then we have

$$\begin{aligned} 0 &\le \big\| f - \|g\|_2^{-2}(f, g)_2\, g\big\|_2^2 = \big(f - \|g\|_2^{-2}(f, g)_2\, g,\; f - \|g\|_2^{-2}(f, g)_2\, g\big)_2 \\ &= \|f\|_2^2 - 2\|g\|_2^{-2}|(f, g)_2|^2 + \|g\|_2^{-4}|(f, g)_2|^2\,\|g\|_2^2 = \|f\|_2^2 - \|g\|_2^{-2}|(f, g)_2|^2, \end{aligned}$$

and the lemma follows in this case as well.

Lemma 1.64 (triangle inequality). If f, g, and h are in $\mathcal{R}[-\pi,\pi]$, then $d_2(f,h) \le d_2(f,g) + d_2(g,h)$.

PROOF. For any two such functions F and G, Lemma 1.63 gives

$$\begin{aligned}\|F+G\|_2^2 &= (F+G,F+G)_2 = (F,F)_2 + (F,G)_2 + (G,F)_2 + (G,G)_2\\ &= \|F\|_2^2 + 2\,\mathrm{Re}(F,G)_2 + \|G\|_2^2\\ &\le \|F\|_2^2 + 2\|F\|_2\|G\|_2 + \|G\|_2^2 = (\|F\|_2 + \|G\|_2)^2.\end{aligned}$$

Taking the square root of both sides and substituting $F = f-g$ and $G = g-h$, we obtain the lemma.

Lemma 1.65. Let $f : \mathbb{R} \to \mathbb{C}$ be periodic of period 2π and Riemann integrable on $[-\pi,\pi]$, and let $\epsilon > 0$ be given. Then there exists a continuous periodic $g : \mathbb{R} \to \mathbb{C}$ of period 2π such that $\|f-g\|_2 < \epsilon$.

PROOF. Because of Lemma 1.64, we may assume that f is real-valued and is not identically 0. Define $M = \sup_{t\in[-\pi,\pi]} |f(t)| > 0$, let $\epsilon > 0$ be given, and let $P = \{x_i\}_{i=0}^n$ be a partition to be specified. Using P, we form the function g defined by

$$g(t) = \frac{x_i - t}{\Delta x_i}\, f(x_{i-1}) + \frac{t - x_{i-1}}{\Delta x_i}\, f(x_i) \quad \text{for } x_{i-1} \le t \le x_i.$$

The graph of g interpolates the points $(x_i, f(x_i))$, $0 \le i \le n$, by line segments. Fix attention on a particular $[x_{i-1},x_i]$, and let $I = \inf_{t\in[x_{i-1},x_i]} f(t)$ and $S = \sup_{t\in[x_{i-1},x_i]} f(t)$. For $t \in [x_{i-1},x_i]$, we have $I \le g(t) \le S$. At a single point t in this interval, $f(t) \ge g(t)$ implies $I \le g(t) \le f(t) \le S$, while $g(t) \ge f(t)$ implies $I \le g(t) \le f(t) \le S$. Thus in either case we have $|f(t) - g(t)| \le S - I$. Taking the supremum over t in the interval and summing on i, we obtain $U(P, |f-g|) \le U(P,f) - L(P,f)$.

Since $|f-g|^2 = |f-g||f+g|$, we have

$$\begin{aligned}\sup_{t\in[x_{i-1},x_i]} |f(t)-g(t)|^2 &\le \sup_{t\in[x_{i-1},x_i]} |f(t)-g(t)| \sup_{t\in[x_{i-1},x_i]} |f(t)+g(t)|\\ &\le 2M \sup_{t\in[x_{i-1},x_i]} |f(t)-g(t)|\end{aligned}$$

for $1 \le i \le n$. Summing on i gives $U\big(P, |f-g|^2\big) \le 2M(U(P,f) - L(P,f))$.

Now we can specify P; it is to be any partition for which $U(P,f) - L(P,f) \le \epsilon^2/(2M)$ and no Δx_i is 0. Then

$$\begin{aligned}0 &\le \tfrac{1}{2\pi}\int_{-\pi}^{\pi} |f(t)-g(t)|^2\,dt \le \tfrac{1}{2\pi}\, U\big(P, |f-g|^2\big)\\ &\le \tfrac{2M}{2\pi}\,(U(P,f) - L(P,f)) \le \epsilon^2/(2\pi) < \epsilon^2,\end{aligned}$$

as required.

PROOF OF THEOREM 1.61. Given $\epsilon > 0$, choose by Lemma 1.65 a continuous periodic g with $\|f - g\|_2 < \epsilon$. Write $g(x) \sim \sum_{n=-\infty}^{\infty} c_n' e^{inx}$, and put $g_N(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(x-t)g(t)\,dt$, where K_N is the Fejér kernel. Fejér's Theorem (Theorem 1.59) gives $\sup_{x\in[-\pi,\pi]} |g(x) - g_N(x)| < \epsilon$ for N sufficiently large. Since any Riemann integrable h has $\|h\|_2 \le \sup_{x\in[-\pi,\pi]} |h(x)|$, we obtain $\|g - g_N\|_2 < \epsilon$ for N sufficiently large. Fixing such an N and substituting from the definition of K_N, we have

$$\begin{aligned} g_N(x) &= \frac{1}{N+1}\sum_{n=0}^{N}\frac{1}{2\pi}\int_{-\pi}^{\pi} D_n(x-t)g(t)\,dt \\ &= \frac{1}{N+1}\sum_{n=0}^{N}\sum_{k=-n}^{n} c_k' e^{ikx} = \sum_{n=-N}^{N} d_n e^{inx} \end{aligned}$$

for suitable constants d_n. Theorem 1.53 and Lemma 1.64 then give

$$\begin{aligned} \Big(\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|^2\,dx - \sum_{n=-N}^{N} |c_n|^2\Big)^{1/2} &= \Big\| f - \sum_{n=-N}^{N} c_n e^{inx}\Big\|_2 \\ &\le \Big\| f - \sum_{n=-N}^{N} d_n e^{inx}\Big\|_2 = \|f - g_N\|_2 \\ &\le \|f - g\|_2 + \|g - g_N\|_2 < \epsilon + \epsilon = 2\epsilon, \end{aligned}$$

and the result follows.

Corollary 1.66 (uniqueness theorem). Let $f : \mathbb{R} \to \mathbb{C}$ be periodic of period 2π and Riemann integrable on $[-\pi, \pi]$. If f has all Fourier coefficients 0, then $\int_{-\pi}^{\pi} |f(x)|\,dx = 0$ and $\int_{-\pi}^{\pi} f(x)\overline{g(x)}\,dx = 0$ for every member g of $\mathcal{R}[-\pi, \pi]$.

PROOF. If f has all Fourier coefficients 0, then $\int_{-\pi}^{\pi} |f(x)|^2\,dx = 0$ by Theorem 1.61. Application of Lemma 1.62 completes the proof of the corollary.

It is natural to ask which sequences $\{c_n\}$ with $\sum |c_n|^2$ finite are the sequences of Fourier coefficients of some $f \in \mathcal{R}[-\pi, \pi]$. To see that this is a difficult question, one has only to compare the two series $\sum_{n=1}^{\infty} n^{-1}\sin x$ and $\sum_{n=1}^{\infty} n^{-1}\cos x$ studied at the beginning of this section. The first series comes from a function in $\mathcal{R}[-\pi, \pi]$, but a little argument shows that the second does not. It was an early triumph of Lebesgue integration that this question has a elegant answer when the Riemann integral is replaced by the Lebesgue integral: the answer when the Lebesgue integral is used is given by the Riesz–Fischer Theorem in Chapter VI, namely, *any* sequence with $\sum |c_n|^2$ finite is the sequence of Fourier coefficients of a square-integrable function.

11. Problems

1. Derive the least-upper-bound property (Theorem 1.1) from the convergence of bounded monotone increasing sequences (Corollary 1.6).

2. According to **Newton's method**, to find numerical approximations to $\sqrt{a}$ when $a > 0$, one can set $x_0 = 1$ and define $x_{n+1} = \frac{1}{2}(x_n^2 + a)/x_n$ for $n \geq 0$. Prove that $\{x_n\}$ converges and that the limit is $\sqrt{a}$.

3. Find $\limsup a_n$ and $\liminf a_n$ when a_n is defined by $a_1 = 0$, $a_{2n} = \frac{1}{2}a_{2n-1}$, $a_{2n+1} = \frac{1}{2} + a_{2n}$. Prove that your answers are correct.

4. For any two sequences $\{a_n\}$ and $\{b_n\}$ in $\mathbb{R}$, prove that $\limsup(a_n + b_n) \leq \limsup a_n + \limsup b_n$, provided the two terms on the right side are not $+\infty$ and $-\infty$ in some order.

5. Which of the following limits exist uniformly for $0 \leq x \leq 1$: (i) $\lim_{n\to\infty} x^n$, (ii) $\lim_{n\to\infty} x^n/n$, (iii) $\lim_{n\to\infty} \sum_{k=1}^n x^k/k$? Supply proofs for those that do converge uniformly. For the other ones, prove anyway that there is uniform convergence on any interval $0 \leq x \leq 1 - \epsilon$, where $\epsilon > 0$.

6. Let $a_n(x) = (-1)^n x^n(1-x)$ on $[0, 1]$. Show that $\sum_{n=0}^\infty a_n(x)$ converges uniformly and that $\sum_{n=0}^\infty |a_n(x)|$ converges pointwise but not uniformly.

7. **(Dini's Theorem)** Suppose that $f_n : [a, b] \to \mathbb{R}$ is continuous and that $f_1 \leq f_2 \leq f_3 \leq \cdots$. Suppose also that $f(x) = \lim f_n(x)$ is continuous and is nowhere $+\infty$. Use the Bolzano–Weierstrass Theorem (Theorem 1.8) to prove that f_n converges to f uniformly for $a \leq x \leq b$.

8. Prove that

$$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \frac{x^{11}}{11!} + \frac{x^{13}}{13!} - \frac{x^{15}}{15!} < \sin x$$

for all $x > 0$.

9. Let $f : (-\infty, +\infty) \to \mathbb{R}$ be infinitely differentiable with $|f^{(n)}(x)| \leq 1$ for all n and x. Use Taylor's Theorem (Theorem 1.36) to prove that

$$f(x) = \sum_{n=0}^\infty \frac{f^{(n)}(0)}{n!}\, x^n$$

for all x.

10. **(Helly's Selection Principle)** Suppose that $\{F_n\}$ is a sequence of nondecreasing functions on $[-1, 1]$ with $0 \leq F_n(x) \leq 1$ for all n and x. Using a diagonal process twice, prove that there is a subsequence $\{F_{n_k}\}$ that converges pointwise on $[-1, 1]$.

11. Prove that the radius of convergence of $\sum_{n=0}^\infty a_n x^n$ is $1/\limsup \sqrt[n]{|a_n|}$.

12. Find a power series expansion for each of the following functions, and find the radius of convergence:
    (a) $1/(1-x)^2 = \frac{d}{dx}(1-x)^{-1}$,
    (b) $\log(1-x) = -\int_1^x \frac{dt}{1-t}$,
    (c) $1/(1+x^2)$,
    (d) $\arctan x = \int_0^x \frac{dt}{1+t^2}$.

13. Prove, along the lines of the proof of Corollary 1.46a, that $\cos x$ has an inverse function $\arccos x$ defined for $0 < x < \pi$ and that the inverse function is differentiable. Find an explicit formula for the derivative of $\arccos x$. Relate $\arccos x$ to $\arcsin x$ when $0 < x < \pi/2$.

14. State and prove uniform versions of Abel's Theorem (Theorem 1.48) and of the corresponding theorem about Cesàro sums (Theorem 1.47), the uniformity being with respect to a parameter x.

15. Prove that the partial sums $\sum_{n=1}^N \cos n\theta$ and $\sum_{n=1}^N \sin n\theta$ are uniformly bounded on any set $\epsilon \le \theta < 2\pi - \epsilon$ if $\epsilon > 0$.

16. Verify the following calculations of Fourier series:
    (a) $f(x) = \left\{ \begin{matrix} +1 & \text{for } 0 < x < \pi \\ -1 & \text{for } -\pi < x < 0 \end{matrix} \right\}$ has $f(x) \sim \dfrac{4}{\pi} \displaystyle\sum_{n=1}^{\infty} \dfrac{\sin(2n-1)x}{2n-1}$.
    (b) $f(x) = e^{-i\alpha x}$ on $(0, 2\pi)$ has $f(x) \sim \dfrac{e^{-i\pi\alpha}\sin \pi\alpha}{\pi} \displaystyle\sum_{n=-\infty}^{\infty} \dfrac{e^{inx}}{n+\alpha}$, provided α is not an integer.

17. Combining Parseval's Theorem (Theorem 1.61) with the results of Problem 16, prove the following identities:
    (a) $\displaystyle\sum_{n=1}^{\infty} \frac{1}{(2n-1)^2} = \frac{\pi^2}{8}$, (b) $\displaystyle\sum_{n=-\infty}^{\infty} \frac{1}{|n+\alpha|^2} = \frac{\pi^2}{\sin^2 \pi\alpha}$.

Problems 18–19 identify the continuous functions $f : \mathbb{R} \to \mathbb{C}$ with $f(x)f(y) = f(x+y)$ for all x and y as the 0 function and the functions $f(x) = e^{cx}$, using two different kinds of techniques from the chapter.

18. Put $F(x) = \int_0^x f(t)\,dt$. Find an equation satisfied by F, and use it to show that f is differentiable everywhere. Then show that $f'(y) = f'(0)f(y)$, and deduce the form of f.

19. Proceed without using integration. Using continuity, find $x_0 > 0$ such that the expression $|f(x) - 1|$ is suitably small when $|x| \le |x_0|$. Show that $f(2^{-k}x_0)$ is then uniquely determined in terms of $f(x_0)$ for all $k \ge 0$. If f is not identically 0, use x_0 to define c. Then verify that $f(x) = e^{cx}$ for all c.

Problems 20–22 construct a nonzero infinitely differentiable function $f : \mathbb{R} \to \mathbb{R}$ having all derivatives equal to 0 at one point.

20. Let $P(x)$ and $Q(x)$ be two polynomials with Q not the zero polynomial. Prove that
$$\lim_{x\to 0} \tfrac{P(x)}{Q(x)}\, e^{-1/x^2} = 0.$$

21. With P and Q as in the previous problem, use the Mean Value Theorem to prove that the function $g : \mathbb{R} \to \mathbb{R}$ with
$$g(x) = \begin{cases} \tfrac{P(x)}{Q(x)}\, e^{-1/x^2} & \text{for } x \neq 0, \\ 0 & \text{for } x = 0, \end{cases}$$
has $g'(x) = 0$ and that g' is continuous.

22. Prove that the function $f : \mathbb{R} \to \mathbb{R}$ with
$$f(x) = \begin{cases} e^{-1/x^2} & \text{for } x \neq 0, \\ 0 & \text{for } x = 0, \end{cases}$$
is infinitely differentiable with derivatives of all orders equal to 0 at $x = 0$.

Problems 23–26 concern a generalization of Cesàro and Abel summability. A **Silverman–Toeplitz summability method** refers to the following construction: One starts with a system $\{M_{ij}\}_{i,j\geq 0}$ of nonnegative real numbers with the two properties that (i) $\sum_j M_{ij} = 1$ for all i and (ii) $\lim_{i\to\infty} M_{ij} = 0$ for all j. The method associates to a complex sequence $\{s_n\}_{n\geq 0}$ the complex sequence $\{t_n\}_{n\geq 0}$ with $t_i = \sum_{j\geq 0} M_{ij} s_j$ as if the process were multiplication by the infinite square matrix $\{M_{ij}\}$ on infinite column vectors.

23. Prove that if $\{s_n\}$ is a convergent sequence with limit s, then the corresponding sequence $\{t_n\}$ produced by a Silverman–Toeplitz summability method converges and has limit s.

24. Exhibit specific matrices $\{M_{ij}\}$ that produce the effects of Cesàro and Abel summability, the latter along a sequence r_i increasing to 1.

25. Let r_i be a sequence increasing to 1, and define $M_{ij} = (j+1)(r_i)^j(1-r_i)^2$. Show that $\{M_{ij}\}$ defines a Silverman–Toeplitz summability method.

26. Using the system $\{M_{ij}\}$ in the previous problem, prove the following: if a bounded sequence $\{s_n\}$ is not necessarily convergent but is Cesàro summable to a limit σ, then $\{s_n\}$ is Abel summable to the same limit σ.

Problems 27–29 concern the Poisson kernel, which plays the same role for Abel sums of Fourier series that the Fejér kernel plays for Cesàro sums. For $0 \leq r < 1$, define the **Poisson kernel** $P_r(\theta)$ to be the r^{th} Abel sum of the Dirichlet kernel $D_n(\theta) = 1 + \sum_{k=1}^n (e^{ik\theta} + e^{-ik\theta})$. In the terminology of Section 8 this means that $a_0 = 1$ and $a_k = e^{ik\theta} + e^{-ik\theta}$ for $k \geq 0$, so that the sequence of partial sums $\sum_{k=0}^n a_k$ is exactly the sequence whose n^{th} term is $D_n(\theta)$. The r^{th} Abel sum $\sum_{n=0}^{\infty} a_n r^n$ is therefore the expression

$$P_r(\theta) = \sum_{n=-\infty}^{\infty} r^{|n|} e^{in\theta}.$$

27. For f in $\mathcal{R}[-\pi, \pi]$, verify that the r^{th} Abel sum of $s_n(f; x)$ is given by the expression $\frac{1}{2\pi} \int_{-\pi}^{\pi} P_r(\theta - \varphi) f(\varphi)\, d\varphi$.

28. Verify that $P_r(\theta) = \dfrac{1 - r^2}{1 - 2r\cos\theta + r^2}$. Deduce that $P_r(\theta)$ has the following properties:
    (i) $P_r(\theta) \geq 0$,
    (ii) $\frac{1}{2\pi} \int_{-\pi}^{\pi} P_r(\theta)\, d\theta = 1$,
    (iii) for any $\delta > 0$, $\sup_{\delta \leq |\theta| \leq \pi} P_r(\theta)$ tends to 0 as r increases to 1.

29. Let $f : \mathbb{R} \to \mathbb{C}$ be periodic of period 2π and Riemann integrable on $[-\pi, \pi]$.
    (a) Prove that if f is continuous at a point θ_0 in $[-\pi, \pi]$, then

$$\lim_{r \uparrow 1} \frac{1}{2\pi} \int_{-\pi}^{\pi} P_r(\theta_0 - \theta) f(\theta)\, d\theta = f(\theta_0).$$

    (b) Prove that if f is uniformly continuous on a subset E of $[-\pi, \pi]$, then the convergence in (a) is uniform for θ_0 in E.

Problems 30–35 lead to a proof without complex-variable theory (and in particular without the complex logarithm) that $\exp\left(z + \frac{1}{2}z^2 + \frac{1}{3}z^3 + \cdots\right) = 1/(1 - z)$ for all complex z with $|z| < 1$.

30. Suppose that $R > 0$, that $f_k(x) = \sum_{n=0}^{\infty} c_{n,k} x^n$ is convergent for $|x| < R$, that $c_{n,k} \geq 0$ for all n and k, and that $\lim_{k \to \infty} f_k(x) = f(x)$ uniformly for $|x| \leq r$ whenever $r < R$. Prove for each $r < R$ that some subsequence $\{f_{k_l}\}$ of $\{f_k\}$ has $\lim_{l \to \infty} f'_{k_l}(x)$ existing uniformly for $|x| \leq r$.

31. In the setting of the previous problem, prove that f is infinitely differentiable for $|x| < R$.

32. In the setting of the previous two problems, use Taylor's Theorem to show that $f(x)$ is the sum of its infinite Taylor series for $|x| < R$.

33. If $0 \leq r < 1$, prove for $|z| \leq r$ that $\left|\frac{1}{N} z^N + \frac{1}{N+1} z^{N+1} + \cdots\right| \leq r^N/(1 - r)$, and deduce that $\exp\left(\frac{1}{N} z^N + \frac{1}{N+1} z^{N+1} + \cdots\right)$ converges to 1 uniformly for $|z| \leq r$.

34. Why is it true that if a power series $\sum_{n=0}^{\infty} c_n z^n$ with complex coefficients sums to 0 for all real z with $|z| < R$, then it sums to 0 for all complex z with $|z| < R$?

35. Prove that $\exp\left(z + \frac{1}{2}z^2 + \frac{1}{3}z^3 + \cdots\right) = 1/(1 - z)$ for all complex z with $|z| < 1$.