CONVEXITY AND OPTIMIZATION IN $\mathbb{R}^n$

CONVEXITY AND OPTIMIZATION IN $\mathbb{R}^n$

LEONARD D. BERKOVITZ
Purdue University

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.

This book is printed on acid-free paper. ∞

For ordering and customer service, call 1-800-CALL-WILEY.

Library of Congress Cataloging-in-Publication Data:

Berkovitz, Leonard David, 1924–
Convexity and optimization in R [superscript n] / Leonard D. Berkovitz.
p. cm.--(Pure and applied mathematics)
"A Wiley-Interscience publication."
Includes bibliographical references and index.
ISBN 0-471-35281-0 (cloth : alk. paper)
1. Convex sets. 2. Mathematical optimization. I. Title. II. Pure and applied mathematics (John Wiley & Sons : Unnumbered)

QA640.B46 2001
516′.08--dc21

2001045391

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To my wife, Anna

CONTENTS

PREFACE

This book presents the mathematics of finite-dimensional optimization, featuring those aspects of convexity that are useful in this context. It provides a basis for the further study of convexity, of more general optimization problems, and of numerical algorithms for the solution of finite-dimensional optimization problems. The intended audience consists of beginning graduate students in engineering, economics, operations research, and mathematics and strong undergraduates in mathematics. This was the audience in a one-semester course at Purdue, MA 521, from which this book evolved.

Ideally, the prerequisites for reading this book are good introductory courses in real analysis and linear algebra. In teaching MA 521, I found that while the mathematics students had the real analysis prerequisites, many of the other students who took the course because of their interest in optimization did not have this prerequisite. Chapter I is for those students and readers who do not have the real analysis prerequisite; in it I present those concepts and results from real analysis that are needed. Except for the Weierstrass theorem on the existence of a minimum, the "heavy" or "deep" theorems are stated without proof. Students without the real variables prerequisite found the material difficult at first, but most managed to assimilate it at a satisfactory level. The advice to readers for whom this is the first encounter with the material in Chapter I is to make a serious effort to master it and to return to it as it is used in the sequel.

To address as wide an audience as possible, I have not always presented the most general result or argument. Thus, in Chapter II I chose the "closest point" approach to separation theorems, rather than more generally valid arguments, because I believe it to be more intuitive and straightforward for the intended audience. Readers who wish to get the best possible separation theorem in finite dimensions should read Sections 6 and 7 of Chapter II. In proving the Fritz John Theorem, I used a penalty function argument due to McShane rather than more technical arguments involving linearizations. I limited the discussion of duality to Lagrangian duality and did not consider Fenchel duality, since the latter would require the development of more mathematical machinery.

The numbering system and reference system for theorems, lemmas, remarks, and corollaries is the following. Within a given chapter, theorems, lemmas, and remarks are numbered consecutively in each section, preceded by the section number. Thus, the first theorem of Section 1 is Theorem 1.1, the second, Theorem 1.2, and so on. The same applies to lemmas and remarks. Corollaries are numbered consecutively within each section without a reference to the section number. Reference to a theorem in the same chapter is given by the theorem number. Reference to a theorem in a chapter different from the current one is given by the theorem number preceded by the chapter number in Roman numerals. Thus, a reference in Chapter IV to Theorem 4.1 in Chapter II would be Theorem II.4.1. References to lemmas and remarks are similar. References to corollaries within the same section are given by the number of the corollary. References to corollaries in a different section of the same chapter are given by prefixing the section number to the corollary number; references in a different chapter are given by prefixing the chapter number in Roman numerals to the preceding.

I thank Rita Saerens and John Gregory for reading the course notes version of this book and for their corrections and suggestions for improvement. I thank Terry Combs for preparing the figures. I also thank Betty Gick for typing seemingly innumerable versions and revisions of the notes for MA 521. Her skill and cooperation contributed greatly to the success of this project.

LEONARD D. BERKOVITZ

West Lafayette, Indiana

Convexity and Optimization in $\mathbb{R}^n$. Leonard D. Berkovitz
ISBN: 0-471-35281-0

I

TOPICS IN REAL ANALYSIS

1. INTRODUCTION

The serious study of convexity and optimization problems in $\mathbb{R}^n$ requires some background in real analysis and in linear algebra. In teaching a course based on notes from which this text evolved, the author and his colleagues assumed that the students had an undergraduate course in linear algebra but did not necessarily have a background in real analysis. The purpose of this chapter is to provide the reader with most of the necessary background in analysis. Not all statements will be proved. The reader, however, is expected to understand the definitions and the theorems and is expected to follow the proofs of statements whenever the proofs are given. The bad news is that many readers will find this chapter to be the most difficult one in the text. The good news is that careful study of this material will provide background for many other courses and that subsequent chapters should be easier. If necessary, the reader should return to this chapter when encountering this material later on.

2. VECTORS IN $\mathbb{R}^n$

By euclidean n-space, or $\mathbb{R}^n$, we mean the set of all n-tuples $\mathbf{x} = (x_1, \ldots, x_n)$, where the x_i, $i = 1, \ldots, n$ are real numbers. Thus, $\mathbb{R}^n$ is a generalization of the familiar two- and three-dimensional spaces $\mathbb{R}^2$ and $\mathbb{R}^3$. The elements $\mathbf{x}$ of $\mathbb{R}^n$ are called *vectors* or *points*. We will often identify the vector $\mathbf{x} = (x_1, \ldots, x_n)$ with the $n \times 1$ matrix

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

and abuse the use of the equal sign to write

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

In this case we shall call $\mathbf{x}$ a column vector. We shall also identify $\mathbf{x}$ with the $1 \times n$ matrix $(x_1, \ldots, x_n)$ and write $\mathbf{x} = (x_1, \ldots, x_n)$. In this case we shall call $\mathbf{x}$ a row vector. It will usually be clear from the context whether we consider $\mathbf{x}$ to be a row vector or a column vector. When there is danger of confusion, we will identify $\mathbf{x}$ with the column vector and use the transpose symbol, which is a superscript t, to denote the row vector. Thus $\mathbf{x}^t = (x_1, \ldots, x_n)$.

We define two operations, vector addition and multiplication by a scalar. If $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ are two vectors, we define their sum $\mathbf{x} + \mathbf{y}$ to be

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, \ldots, x_n + y_n).$$

For any scalar, or real number, α we define $\alpha\mathbf{x}$ to be

$$\alpha\mathbf{x} = (\alpha x_1, \ldots, \alpha x_n).$$

We assume that the reader is familiar with the properties of these operations and knows that under these operations $\mathbb{R}^n$ is a vector space over the reals.

Another important operation is the *inner product*, or *dot product*, of two vectors, denoted by $\langle \mathbf{x}, \mathbf{y} \rangle$ or $\mathbf{x}^t\mathbf{y}$ and defined by

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i.$$

Again, we assume that the reader is familiar with the properties of the inner product. We use the inner product to define the norm $\|\cdot\|$ or length of a vector $\mathbf{x}$ as follows:

$$\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2} = \left[\sum_{i=1}^{n} x_i^2 \right]^{1/2}.$$

This norm is called the *euclidean norm*. In $\mathbb{R}^2$ and $\mathbb{R}^3$ the euclidean norm reduces to the familiar length. It is straightforward to show that the norm has the following properties:

$$\|\mathbf{x}\| \geqslant 0 \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^n, \tag{1}$$

$$\|\mathbf{x}\| = 0 \quad \text{if and only if } \mathbf{x} \text{ is the zero vector } \mathbf{0} \text{ in } \mathbb{R}^n, \tag{2}$$

$$\|\alpha\mathbf{x}\| = |\alpha| \, \|\mathbf{x}\| \quad \text{for all real numbers } \alpha \text{ and vectors } \mathbf{x} \text{ in } \mathbb{R}^n. \tag{3}$$

The norm has two additional properties, which we will prove:

For all vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leqslant \|\mathbf{x}\| \, \|\mathbf{y}\| \tag{4}$$

with equality holding if and only if one of the vectors is a scalar multiple of the other.

For all vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$

$$\|\mathbf{x} + \mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|, \tag{5}$$

with equality holding if and only if one of the vectors is a *nonnegative* scalar multiple of the other.

The inequality (4) is known as the Cauchy–Schwarz inequality. The reader should not be discouraged by the slickness of the proof that we will give or by the slickness of other proofs. Usually, the first time that a mathematical result is discovered, the proof is rather complicated. Often, many very smart people then look at it and constantly improve the proof until a relatively simple and clever argument is found. Therefore, do not be intimidated by clever arguments. Now to the proof.

Let $\mathbf{x}$ and $\mathbf{y}$ be any pair of fixed vectors in $\mathbb{R}^n$. Let λ be any real scalar. Then

$$\begin{aligned} 0 \leqslant \|\mathbf{x} + \lambda\mathbf{y}\|^2 &= \langle \mathbf{x} + \lambda\mathbf{y}, \mathbf{x} + \lambda\mathbf{y} \rangle \\ &= \|\mathbf{x}\|^2 + 2\lambda(\mathbf{x}, \mathbf{y}\rangle + \|\mathbf{y}\|^2\lambda^2 \quad \text{for all } \lambda. \end{aligned} \tag{6}$$

This says that the quadratic in λ on the right either has a double real root or has no real roots. Therefore, by the quadratic formula, its discriminant must satisfy

$$4\langle \mathbf{x}, \mathbf{y} \rangle^2 - 4\|\mathbf{x}\|^2\|\mathbf{y}\|^2 \leqslant 0.$$

Transposing and taking square roots give

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leqslant \|\mathbf{x}\| \, \|\mathbf{y}\|.$$

Equality holds if and only if the quadratic has a double real root, say $\lambda = \kappa$. But then, from (6), $\|\mathbf{x} + \kappa\mathbf{y}\| = 0$. Hence by (2) $\mathbf{x} + \kappa\mathbf{y} = 0$, and $\mathbf{x} = -\kappa\mathbf{y}$.

To prove (5), we write

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^2 = \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle &= \|\mathbf{x}\|^2 + 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2 \\ &\leqslant \|\mathbf{x}\|^2 + 2\|\mathbf{x}\| \, \|\mathbf{y}\| + \|\mathbf{y}\|^2 = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2, \end{aligned} \tag{7}$$

where the inequality follows from the Cauchy–Schwarz inequality. If we now take square roots, we get the triangle inequality. We obtain equality in (7), and hence in the triangle inequality, if and only if $\langle \mathbf{x}, \mathbf{y} \rangle$ is nonnegative and either $\mathbf{y}$ is a scalar multiple of $\mathbf{x}$ or $\mathbf{x}$ is a scalar multiple of $\mathbf{y}$. But under these circumstances, for $\langle \mathbf{x}, \mathbf{y} \rangle$ to be nonnegative, the multiple must be nonnegative. The inequality (5) is called the triangle inequality because when it is applied to vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ it says that the length of a third side of a triangle is less than or equal to the sum of the lengths of the other two sides.

Exercise 1.1. For any vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, show that $\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2$. Interpret this relation as a statement about parallelograms in $\mathbb{R}^2$ and $\mathbb{R}^3$.

In elementary courses it is shown that in $\mathbb{R}^2$ and $\mathbb{R}^3$ the cosine of the angle θ, $0 \leqslant \theta \leqslant \pi$, between two nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ is given by the formula

$$\cos\theta = \frac{\langle \mathbf{x}, \mathbf{y}\rangle}{\|\mathbf{x}\|\,\|\mathbf{y}\|}.$$

Thus, $\mathbf{x}$ and $\mathbf{y}$ are orthogonal if and only if $\langle \mathbf{x}, \mathbf{y}\rangle = 0$. In $\mathbb{R}^n$, the Cauchy–Schwarz inequality allows us to give meaning to the notion of angle between nonzero vectors as follows. Since $|\langle \mathbf{x}, \mathbf{y}\rangle\| \leqslant \|\mathbf{x}\|\,\|\mathbf{y}\|$, the absolute value of the quotient

$$\frac{\langle \mathbf{x}, \mathbf{y}\rangle}{\|\mathbf{x}\|\,\|\mathbf{y}\|} \tag{8}$$

is less than or equal to 1, and thus (8) is the cosine of an angle between zero and π. We define this angle to be the angle between $\mathbf{x}$ and $\mathbf{y}$. We say that $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal* if $\langle \mathbf{x}, \mathbf{y}\rangle = 0$.

3. ALGEBRA OF SETS

Given two sets A and B, the set A is said to be a *subset* of B, or to be *contained* in B, and written $A \subseteq B$, if every element of A is also an element of B. The set A is said to be a *proper subset* of B if A is a subset of B and there exists at least one element of B that is not in A. This is written as $A \subset B$. Two sets A and B are said to be *equal*, written $A = B$, if $A \subseteq B$ and $B \subseteq A$. The *empty set*, or *null* set, is the set that has no members. The empty set will be denoted by $\varnothing$. The notation $x \in S$ will mean "x belongs to the set S" or "x is an element of the set S."

Let $\{S_\alpha\}_{\alpha\in A}$ be a collection of sets indexed by an index set A. By the *union* of the sets $\{S_\alpha\}_{\alpha\in A}$, denoted by $\bigcup_{\alpha\in A} S_\alpha$, we mean the set consisting of all elements that belong to at least one of the S_α. Note that if not all of the S_α are empty, then $\bigcup_{\alpha\in A} S_\alpha$ is not empty. By the *intersection* of the sets $\{S_\alpha\}_{\alpha\in A}$, denoted by $\bigcap_{\alpha\in A} S_\alpha$, we mean the set of points s with the property that s belongs to *every* set S_α. Note that for a collection $\{S_\alpha\}_{\alpha\in A}$ of nonempty sets, the intersection can be empty.

Let $\mathfrak{X}$ be a set, say $\mathbb{R}^n$ for example, and let S be a subset of $\mathfrak{X}$. By the *complement* of S (relative to $\mathfrak{X}$), written cS, we mean the set consisting of those points in $\mathfrak{X}$ that do not belong to S. Note that $c\mathfrak{X}$ is the empty set. By convention, we take the complement of the empty set to be $\mathfrak{X}$. We thus always have $c(cS) = S$. Also, if $A \subset B$, then $cA \supset cB$. The reader should convince himself or herself of the truth of the following lemma, known as *de Morgan's* *law*, either by writing out a proof or drawing Venn diagrams for some simple cases.

LEMMA 3.1. *Let* $\{S_\alpha\}_{\alpha\in A}$ *be a collection of subsets of a set* $\mathfrak{X}$. *Then*

$$\bigcup_{\alpha\in A} S_\alpha = c\left[\bigcap_{\alpha\in A}(cS_\alpha)\right],$$

$$\bigcap_{\alpha\in A} S_\alpha = c\left[\bigcup_{\alpha\in A}(cS_\alpha)\right].$$

4. METRIC TOPOLOGY OF $\mathbb{R}^n$

Let $\mathfrak{X}$ be a set. A function d that assigns a real number to each pair of points (x, y) with $x\in\mathfrak{X}$ any $y\in\mathfrak{X}$ is said to be a *metric*, or *distance function*, on $\mathfrak{X}$ if it satisfies the following:

$$d(x, y) \geqslant 0, \quad \text{with equality holding if and only if } x = y, \tag{1}$$

$$d(x, y) = d(y, x), \tag{2}$$

$$d(x, y) \leqslant d(x, z) + d(z, y) \quad \text{for all } x, y, z \text{ in } \mathfrak{X}. \tag{3}$$

The last inequality is called the triangle inequality.

In $\mathbb{R}^n$ we define a function d on pairs of points $\mathbf{x}, \mathbf{y}$ by the formula

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \left(\sum_{i=1}^{n}(x_i - y_i)^2\right)^{1/2}. \tag{4}$$

Exercise 4.1. Use the properties of the norm to show that the function d defined by (4) is a metric, or distance function, on $\mathbb{R}^n$.

A set $\mathfrak{X}$ with metric d is called a *metric space*. Although metric spaces occur in many areas of mathematics and applications, we shall confine ourselves to $\mathbb{R}^n$. In $\mathbb{R}^2$ and $\mathbb{R}^3$ the function defined in (4) is the ordinary euclidean distance.

We now present some important definitions to the reader, who should master them and try to improve his or her understanding by drawing pictures in $\mathbb{R}^2$.

The (open) *ball* centered at $\mathbf{x}$ with radius $r > 0$, denoted by $B(\mathbf{x}, r)$, is defined to be the set of points $\mathbf{y}$ whose distance from $\mathbf{x}$ is *less than* r. We write this in symbols as

$$B(\mathbf{x}, r) = \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\| < r\}.$$

The *closed ball* $\overline{B(\mathbf{x}, r)}$ with center at $\mathbf{x}$ and radius $r > 0$ is defined by

$$\overline{B(\mathbf{x}, r)} = \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\| \leqslant r\}.$$

Let $S \subseteq \mathbb{R}^n$. A point $\mathbf{x}$ is an *interior point* of S if there exists an $r > 0$ such that $B(\mathbf{x}, r) \subset S$. A set S need not have any interior points. For example, consider the set of points in the plane of the form $(x_1, 0)$ with $0 < x_1 < 1$. This is the interval $0 < x_1 < 1$ on the x_1-axis. It has no interior points, considered as a set in $\mathbb{R}^2$. On the other hand, if we consider this as a set in $\mathbb{R}^1$, every point is an interior point. This leads to the important observation that for a given set, whether or not it has interior points may depend on in which euclidean space $\mathbb{R}^n$ we take the set to be lying, or *embedded*.

If the set of interior points of a set S is not empty, then we call the set of interior points the *interior* of S and denote it by $\text{int}(S)$.

A set S is said to be *open* if all points of S are interior points. Thus an equivalent definition of an open set is that it is equal to its interior. If we go back to the definition of interior point, we can restate the definition of an open set as follows. A set S is open if for every point $\mathbf{x}$ in S there exists a positive number $r > 0$, which may depend on $\mathbf{x}$, such that the ball $B(\mathbf{x}, r)$ is contained in S. The last definition, while being wordy, is the one that the reader should picture mentally.

A point $\mathbf{x}$ is said to be a *limit point* of a set S if for every $\varepsilon > 0$ there exists a point $\mathbf{x}_\varepsilon \neq \mathbf{x}$ such that $\mathbf{x}_\varepsilon$ belongs to S and $\mathbf{x}_\varepsilon \in B(\mathbf{x}, \varepsilon)$. The point $\mathbf{x}_\varepsilon$ will in general depend on ε. A set need not have any limit points, and a limit point of a set need not belong to the set. An example of a set without a limit point is the set $\mathbb{N}$ of positive integers, considered as a set in $\mathbb{R}^1$. For an example of a limit point that does not belong to a set, consider the set $S \equiv \{x : x = 1/n, n = 1, 2, 3, \ldots\}$ in $\mathbb{R}^1$. Zero is a limit point of the set, yet zero does not belong to S. We shall denote the set of limit points of a set S by S'.

Exercise 4.2. (a) Sketch the graph of $y = \sin(1/x)$, $x > 0$.

(b) Consider the graph as a set in $\mathbb{R}^2$ and find the limit points of this set.

The *closure* of a set S, denoted by $\bar{S}$, is defined to be the set $S \cup S'$; that is, $\bar{S} = S \cup S'$. A set S is said to be *closed* if S contains all of its limit points; that is, $S' \subset S$. A set S is closed if and only if $S = \bar{S}$. To see this, note that $S = \bar{S}$ and the definitions $\bar{S} = S \cup S'$ imply that $S = S \cup S'$. Hence $S' \subset S$. On the other hand, if $S' \subset S$, then $S \cup S' = S$, and so $\bar{S} = S \cup S' = S$.

A set can be neither open nor closed. In $\mathbb{R}^2$ consider $B(\mathbf{0}, 1)$, the ball with center at the origin and radius 1. It should be intuitively clear that all points $\mathbf{x}$ in $\mathbb{R}^2$ with $\|\mathbf{x}\| = 1$ are limit points of $B(\mathbf{0}, 1)$. Now consider the set

$$S = B(\mathbf{0}, 1) \cup \{\mathbf{x} = (x_1, x_2) : \|\mathbf{x}\| = 1, x_1 \geqslant 0\}.$$

(The reader should sketch this set.) Points $\mathbf{x} = (x_1, x_2)$ with $\|\mathbf{x}\| = 1$ and $x_1 \geqslant 0$ are not interior points of S since for such an $\mathbf{x}$, no matter how small we choose $\varepsilon > 0$, the ball $B(\mathbf{x}, \varepsilon)$ will not belong to S. Hence S is not open. Also, S is not closed since points $\mathbf{x} = (x_1, x_2)$ with $\|\mathbf{x}\| = 1$ and $x_1 < 0$ are limit points of S, yet they do not belong to S.

It follows from our definitions that $\mathbb{R}^n$ itself is both open and closed. By convention, we will also take the empty set to be open and closed.

THEOREM 4.1. *Complements of closed sets are open. Complements of open sets are closed.*

Proof. If $S = \varnothing$, where $\varnothing$ denotes the empty set, or if $S = \mathbb{R}^n$, then there is nothing to prove. Let $S \neq \varnothing$, $S \neq \mathbb{R}^n$ and let S be closed. Let $\mathbf{x}$ be any point in cS. Then $\mathbf{x} \notin S$ and $\mathbf{x}$ is not a limit point of S. If we recall the definition of "$\mathbf{x}$ is a limit point of S," we see that the statement "$\mathbf{x}$ is not a limit point of S" means that there is *an* $\varepsilon_0 > 0$ such that all points $\mathbf{y} \neq \mathbf{x}$ with $\mathbf{y} \in B(\mathbf{x}, \varepsilon_0)$ are in cS. Since $\mathbf{x} \in cS$, this means that $B(\mathbf{x}, \varepsilon_0) \subset cS$, and so cS is open. Now let $S = \varnothing$, $S \neq \mathbb{R}^n$ and let S be open. Let $\mathbf{x}$ be a limit point of cS. Then for *every* $\varepsilon > 0$ there exists a point $\mathbf{x}_\varepsilon \neq \mathbf{x}$ in cS such that $\mathbf{x}_\varepsilon \in B(\mathbf{x}, \varepsilon)$. Since $\mathbf{x}_\varepsilon \in cS$, the ball $B(\mathbf{x}, \varepsilon)$ does not belong to S. Thus $\mathbf{x}$ cannot belong to the open set S, for then we would be able to find an ε_0 such that $B(\mathbf{x}, \varepsilon_0)$ did belong to S. Hence $\mathbf{x} \in cS$, so cS is closed.

Exercise 4.3. Show that for $\mathbf{x} \in \mathbb{R}^n$ and $r > 0$ the set $B(\mathbf{x}, r)$ is open; that is, show that an open ball is open.

Exercise 4.4. Show that for $\mathbf{x} \in \mathbb{R}^n$ and $r > 0$ the closed ball $\overline{B(\mathbf{x}, r)}$ is closed.

Exercise 4.5. Show that any finite set of points $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in $\mathbb{R}^n$ is closed.

Exercise 4.6. Show that in $\mathbb{R}^n$ no point $\mathbf{x}$ with $\|\mathbf{x}\| = 1$ is an interior point of $\overline{B(\mathbf{0}, 1)}$.

Exercise 4.7. Show that for any set S in $\mathbb{R}^n$ the set $\bar{S}$ is closed.

Exercise 4.8. Show that for any set S the closure $\bar{S}$ is equal to the intersection of all closed sets containing S.

THEOREM 4.2. (*i*) *Let $\{O_\alpha\}_{\alpha \in A}$ be a collection of open sets. Then $\bigcup_{\alpha \in A} O_\alpha$ is open.*
(*ii*) *Let $O_1, \ldots, O_n$ be a finite collection of open sets. Then $\bigcap_{i=1}^n O_i$ is open.*
(*iii*) *Let $\{F_\alpha\}_{\alpha \in A}$ be a collection of closed sets. Then $\bigcap_{\alpha \in A} F_\alpha$ is closed.*
(*iv*) *Let $F_1, \ldots, F_n$ be a finite collection of closed sets. Then $\bigcup_{i=1}^n F_i$ is closed.*

Note that an infinite collection of open sets need not have an intersection that is open. For example, in $\mathbb{R}^1$, for each $n = 1, 2, 3, \ldots$, let

$$O_n = (1 - 1/n,\ 1 + 1/n) = \{x \in \mathbb{R}^1 : 1 - 1/n < x < 1 + 1/n\}.$$

In $\mathbb{R}^1$ each O_n is open, and $\bigcap_{n=1}^\infty O_n = \{1\}$, a point. By Exercise 4.5, the set consisting of the point $x = 1$ is closed. Similarly, the union of an infinite number of closed sets need not be closed. To see this, for each $n = 1, 2, 3, \ldots$ let $F_n = [0, 1 - 1/(n+1)] = \{x \in \mathbb{R}^1 : 0 \leqslant x \leqslant 1 - 1/(n+1)\}$. Each F_n is closed, and

$\bigcup_{n=1}^{\infty} F_n = [0, 1) = \{x \in \mathbb{R}^1 : 0 \leqslant x < 1\}$, which is not closed since it does not contain the limit point $x = 1$.

We now prove the theorem. To establish (i), we take an arbitrary element $\mathbf{x}$ in $\bigcup_{\alpha \in A} O_\alpha$. Since $\mathbf{x}$ is in the union, it must belong to at least one of the sets $\{O_\alpha\}_{\alpha \in A}$, say O_{α_0}. Since O_{α_0} is open, there exists an $r_0 > 0$ such that $B(\mathbf{x}, r_0) \subseteq O_{\alpha_0}$. Hence $B(\mathbf{x}, r_0) \subseteq \bigcup_{\alpha \in A} O_\alpha$, and thus the union is open. To establish (ii), let $\mathbf{x}$ be an arbitrary point in $\bigcap_{i=1}^{n} O_i$. Then $\mathbf{x} \in O_i$ for each $i = 1, \ldots, n$. Since each O_i is open, for each $i = 1, \ldots, n$ there exists an $r_i > 0$ such that $B(\mathbf{x}, r_i) \subseteq O_i$. Let $r = \min\{r_i : i = 1, \ldots, n\}$. Then $B(\mathbf{x}, r) \subseteq B(\mathbf{x}, r_i) \subseteq O_i$ for each $i = 1, \ldots, n$. Hence $B(\mathbf{x}, r) \subset \bigcap_{i=1}^{n} O_i$, and thus the intersection is open.

The most efficient way to establish (iii) is to write

$$\bigcap_{\alpha \in A} F_\alpha = \bigcap_{\alpha \in A} c(cF_\alpha) = c\left[\bigcup_{\alpha \in A} (cF_\alpha)\right],$$

where the second equality follows from Lemma 3.1. Since F_α is closed, cF_α is open and so is $\bigcup_{\alpha \in A} (cF_\alpha)$. Hence the complement of this set is closed, and therefore so is $\bigcap_{\alpha \in A} F_a$. To establish (iv), write

$$\bigcup_{i=1}^{n} F_i = \bigcup_{i=1}^{n} c(cF_i) = c\left[\bigcap_{i=1}^{n} (cF_i)\right].$$

Again, cF_i is open for each $i = 1, \ldots, n$. Hence the intersection $\bigcap_{i=1}^{n} (cF_i)$ is open and the complement of the intersection is closed.

The reader who wishes to test his or her mastery of the relevant definitions should prove statements (iii) and (iv) of the theorem directly.

5. LIMITS AND CONTINUITY

A *sequence* in $\mathbb{R}^n$ is a function that assigns to each positive integer k a vector or point $\mathbf{x}_k$ in $\mathbb{R}^n$. We usually write the sequence as $\{\mathbf{x}_k\}_{k=1}^{\infty}$ or $\{\mathbf{x}_k\}$, rather than in the usual function notation. Examples of sequences in $\mathbb{R}^2$ are

$$\begin{aligned}
&\text{(i)} \quad \{\mathbf{x}_k\} = \{(k, k)\},\\
&\text{(ii)} \quad \{\mathbf{x}_k\} = \left\{\left(\cos\frac{k\pi}{2}, \sin\frac{k\pi}{2}\right)\right\},\\
&\text{(iii)} \quad \{\mathbf{x}_k\} = \left\{\left(\frac{(-1)^k}{2^k}, \frac{1}{2^k}\right)\right\},\\
&\text{(iv)} \quad \{\mathbf{x}_k\} = \left\{\left((-1)^k - \frac{1}{k}, (-1)^k - \frac{1}{k}\right)\right\}.
\end{aligned} \tag{1}$$

The reader should plot the points in these sequences.

A sequence $\{\mathbf{x}_k\}$ is said to have a *limit* $\mathbf{y}$ or to *converge* to $\mathbf{y}$ if for *every* $\varepsilon > 0$ there is a positive integer $K(\varepsilon)$ such that whenever $k > K(\varepsilon)$, $\mathbf{x}_k \in B(\mathbf{y}, \varepsilon)$. We write

$$\lim_{k\to\infty} \mathbf{x}_k = \mathbf{y} \quad \text{or} \quad \mathbf{x}_k \to \mathbf{y}.$$

The sequences (i), (ii), and (iv) in (1) do not converge, while the sequence (iii) converges to $\mathbf{0} = (0, 0)$. Sequences that converge are said to be *convergent*; sequences that do not converge are said to be *divergent*. A sequence $\{\mathbf{x}_k\}$ is said to be *bounded* if there exists an $M > 0$ such that $\|\mathbf{x}_k\| < M$ for all k. A sequence that is not bounded is said to be unbounded. The sequence (i) is unbounded; the sequences (ii), (iii), and (iv) are bounded.

The following theorem lists some basic properties of convergent sequences.

THEOREM 5.1. (*i*) *The limit of a convergent sequence is unique.*

(*ii*) *Let* $\lim_{k\to\infty} \mathbf{x}_k = \mathbf{x}_0$, *let* $\lim_{k\to\infty} \mathbf{y}_k = \mathbf{y}_0$, *and let* $\lim_{k\to\infty} \alpha_k = \alpha$, *where* $\{\alpha_k\}$ *is a sequence of scalars. Then* $\lim_{k\to\infty} (\mathbf{x}_k + \mathbf{y}_k)$ *exists and equals* $\mathbf{x}_0 + \mathbf{y}_0$, *and* $\lim_{k\to\infty} \alpha_k \mathbf{x}_k$ *exists and equals* $\alpha \mathbf{x}_0$.

(*iii*) *A convergent sequence is bounded.*

Exercise 5.1. Prove Theorem 5.1.

Let $\{\mathbf{x}_k\}$ be a sequence in $\mathbb{R}^n$ and let $r_1 < r_2 < r_3 < \cdots < r_m < \cdots$ be a strictly increasing sequence of positive integers. Then the sequence $\{\mathbf{x}_{r_m}\}_{m=1}^{\infty}$ is called a *subsequence* of $\{\mathbf{x}_k\}$. Thus, $\{\mathbf{x}_{2k}\}_{k=1}^{\infty} = \{(2k, 2k)\}$ is a subsequence of (i). The sequence $\{\mathbf{x}_{2k}\}_{k=1}^{\infty} = \{(\cos k\pi, \sin k\pi)\}_{k=1}^{\infty}$ is a subsequence of (ii). Each of the sequences $\{\mathbf{x}_{2k}\} = \{((-1)^{2k} - 1/2k, (-1)^{2k} - 1/2k)\} = \{(1 - 1/2k, 1 - 1/2k)\}$ and $\{\mathbf{x}_{2k+1}\} = \{(-1 - 1/(2k+1), -1 - 1/(2k+1))\}$ is a subsequence of (iv). The reader should plot the various subsequences.

We can now formulate a very useful criterion for a point to be a limit point of a set. (The reader should not confuse the notions of limit and limit point. They are different.)

LEMMA 5.1. *A point* $\mathbf{x}$ *is a limit point of a set* S *if and only if there exists a sequence* $\{\mathbf{x}_k\}$ *of points in* S *such that, for each* k, $\mathbf{x}_k \neq \mathbf{x}$ *and* $\mathbf{x}_k \to \mathbf{x}$.

Proof. Let $\mathbf{x}$ be a limit point of S. Then for every positive integer k there is a point $\mathbf{x}_k$ in S such that $\mathbf{x}_k \neq \mathbf{x}$ and $\mathbf{x}_k \in B(\mathbf{x}, 1/k)$. For every $\varepsilon > 0$, there is a positive integer $K(\varepsilon)$ satisfying $1/K(\varepsilon) < \varepsilon$. Hence for $k > K(\varepsilon)$, we have $1/k < \varepsilon$ and $B(\mathbf{x}, 1/k) \subset B(\mathbf{x}, \varepsilon)$. Hence the sequence $\{\mathbf{x}_k\}$ converges to $\mathbf{x}$. Conversely, let there exist a sequence $\{\mathbf{x}_k\}$ of points in S with $\mathbf{x}_k \neq \mathbf{x}$ and $\mathbf{x}_k \to \mathbf{x}$. Let $\varepsilon > 0$ be arbitrary. Since $\mathbf{x}_k \to \mathbf{x}$, there exists a positive integer $K(\varepsilon)$ such that, for $k > K(\varepsilon)$, $\mathbf{x}_k \in B(\mathbf{x}, \varepsilon)$. Since by hypothesis $\mathbf{x}_k \neq \mathbf{x}$ for all k, it follows that we have satisfied the requirements of the definition that $\mathbf{x}$ is a limit point of S.

Remark 5.1. Let $\{\mathbf{x}_k\}$ be a sequence of points belonging to a set S and converging to a point $\mathbf{x}$. Then $\mathbf{x}$ must belong to $\bar{S}$. If $\mathbf{x}$ is in S, there is nothing to prove. If $\mathbf{x}$ were not in S, then since $\mathbf{x}_k \in S$ for each k, it follows that $\mathbf{x}_k \neq \mathbf{x}$. Lemma 5.1 then implies that $\mathbf{x}$ is a limit point of S. Thus $\mathbf{x} \in \bar{S}$. If S is a closed set, then since $S = \bar{S}$, the point $\mathbf{x}$ belongs to S.

Remark 5.2. Let S be a set in $\mathbb{R}^n$. If $\mathbf{x}$ belongs to $\bar{S}$, then there is a sequence of points $\{\mathbf{x}_k\}$ in S such that $\mathbf{x}_k \to \mathbf{x}$. To see this, note that if $\mathbf{x}$ is in $\bar{S}$, then $\mathbf{x}$ belongs either to S or to S'. If $\mathbf{x} \in S'$, the assertion follows from Lemma 5.1. If $\mathbf{x} \in S$ and $\mathbf{x} \notin S'$, we may take $\mathbf{x}_k = \mathbf{x}$ for all positive integers k.

Let S be a subset (proper or improper) of $\mathbb{R}^n$. Let $\mathbf{f}$ be a function with domain S and range contained in $\mathbb{R}^m$. We denote this by $\mathbf{f}: S \to \mathbb{R}^m$. Thus, $\mathbf{f} = (f_1, \ldots, f_m)$, where each f_i is a real-valued function.

Let $\mathbf{s}$ be a limit point of S. We say that "*the limit of* $\mathbf{f}(\mathbf{x})$ *as* $\mathbf{x}$ *approaches* $\mathbf{s}$ *exists and equals* $\mathbf{L}$" if for every $\varepsilon > 0$ there exists a $\delta > 0$, which may depend on ε, such that for all $\mathbf{x} \in B(\mathbf{s}, \delta) \cap S$, $\mathbf{x} \neq \mathbf{s}$, we have $\mathbf{f}(\mathbf{x}) \in B(\mathbf{L}, \varepsilon)$. We write

$$\lim_{\mathbf{x} \to \mathbf{s}} \mathbf{f}(\mathbf{x}) = \mathbf{L}.$$

It is not hard to show that, if $\mathbf{L} = (L_1, \ldots, L_m)$, $\lim_{\mathbf{x}\to\mathbf{s}} \mathbf{f}(\mathbf{x}) = \mathbf{L}$ if and only if, for each $i = 1, \ldots, m$, $\lim_{\mathbf{x}\to\mathbf{s}} f_i(\mathbf{x}) = L_i$.

It is also not hard to show that *if a limit exists it is unique* and that the algebra of limits, as stated in the next theorem, holds.

THEOREM 5.2. *Let* $\lim_{\mathbf{x}\to\mathbf{s}} \mathbf{f}(\mathbf{x}) = \mathbf{L}$ *and let* $\lim_{\mathbf{x}\to\mathbf{s}} \mathbf{g}(\mathbf{x}) = \mathbf{M}$. *Then*

$$\lim_{\mathbf{x} \to \mathbf{s}} [\mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})]$$

exists and equals $\mathbf{L} + \mathbf{M}$. *Also* $\lim_{\mathbf{x}\to\mathbf{s}} [\alpha\mathbf{f}(\mathbf{x})]$ *exists and equals* $\alpha\mathbf{L}$.

A useful sequential criterion for the existence of a limit is now given.

THEOREM 5.3. $\text{Lim}_{\mathbf{x}\to\mathbf{s}} \mathbf{f}(\mathbf{x}) = \mathbf{L}$ *if and only if for every sequence* $\{\mathbf{x}_k\}$ *of points in* S *such that* $\mathbf{x}_k \neq \mathbf{s}$ *for all* k *and* $\mathbf{x}_k \to \mathbf{s}$ *it is true that* $\lim_{k\to\infty} \mathbf{f}(\mathbf{x}_k) = \mathbf{L}$.

Proof. Let $\lim_{\mathbf{x}\to\mathbf{s}} \mathbf{f}(\mathbf{x}) = \mathbf{L}$ and let $\{\mathbf{x}_k\}$ be an arbitrary sequence of points in S with $\mathbf{x}_k \neq \mathbf{s}$ for all k and with $\mathbf{x}_k \to \mathbf{s}$. Let $\varepsilon > 0$ be given. Since $\lim_{\mathbf{x}\to\mathbf{s}} \mathbf{f}(\mathbf{x}) = \mathbf{L}$, there exists a $\delta(\varepsilon) > 0$ such that, for all $\mathbf{x} \neq \mathbf{s}$ and in $B(\mathbf{s}, \delta) \cap S$, $\mathbf{f}(\mathbf{x}) \in B(\mathbf{L}, \varepsilon)$. Since $\mathbf{x}_k \to \mathbf{s}$, there exists a positive integer $K(\delta(\varepsilon))$ such that whenever $k > K(\delta(\varepsilon))$, $\mathbf{x}_k \in B(\mathbf{s}, \delta)$. Thus, for $k > K(\delta(\varepsilon))$, $\mathbf{f}(\mathbf{x}_k) \in B(\mathbf{L}, \varepsilon)$, and so $\mathbf{f}(\mathbf{x}_k) \to \mathbf{L}$.

We now prove the implication in the other direction. Suppose that $\lim_{\mathbf{x}\to\mathbf{s}} \mathbf{f}(\mathbf{x})$ does not equal $\mathbf{L}$. Then there exists an $\varepsilon_0 > 0$ such that for every $\delta > 0$ there exists a point $\mathbf{x}_\delta$ in S satisying the conditions $\mathbf{x}_\delta \neq \mathbf{s}$, $\mathbf{x}_\delta \in B(\mathbf{s}, \delta)$, and $\mathbf{f}(\mathbf{x}_\delta) \notin B(\mathbf{L}, \varepsilon_0)$. Take δ through the sequence of values $\delta = 1/k$, where k runs through the positive integers. We then get a sequence $\{\mathbf{x}_k\}$, with $\mathbf{x}_k \in S$, $\mathbf{x}_k \neq \mathbf{s}$, and $\mathbf{x}_k \to \mathbf{s}$, with the further property that $\mathbf{f}(\mathbf{x}_k) \notin B(\mathbf{L}, \varepsilon_0)$. Thus $\mathbf{f}(\mathbf{x}_k)$ does not converge to $\mathbf{L}$, and the theorem is proved.

Let $\mathbf{x}_0$ be a point of S that is also a limit point of S; that is, $\mathbf{x}_0 \in S \cap S'$. We say that $\mathbf{f}$ is *continuous* at $\mathbf{x}_0$ if

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}_0).$$

If $\mathbf{x}_0 \in S$ and $\mathbf{x}_0$ is not a limit point of S, we will take $\mathbf{f}$ to be continuous at $\mathbf{x}_0$ by convention.

From the corresponding property of limits it follows that $\mathbf{f}$ is continuous at $\mathbf{x}_0$ if and only if each of the real-valued functions $f_1, \ldots, f_m$ is continuous at $\mathbf{x}_0$. It follows from Theorem 5.2 that if $\mathbf{f}$ and $\mathbf{g}$ are continuous at $\mathbf{s}$, then so are $\mathbf{f} + \mathbf{g}$ and $\alpha\mathbf{f}$.

The function $\mathbf{f}$ is said to be *continuous on* S if it is continuous at every point of S.

Exercise 5.2. Show that if $\lim_{\mathbf{x} \to \mathbf{s}} \mathbf{f}(\mathbf{x}) = \mathbf{L}$, then $\mathbf{L}$ is unique.

Exercise 5.3. Prove Theorem 5.2.

Exercise 5.4. Let f be a real-valued function defined on a set S in $\mathbb{R}^n$. Show that if f is continuous at a point $\mathbf{x}_0$ in S and if $f(\mathbf{x}_0) < 0$, then there exists a $\delta > 0$ such that $f(\mathbf{x}) < 0$ for all $\mathbf{x}$ in $B(\mathbf{x}_0, \delta) \cap S$.

Exercise 5.5. Show that the real-valued function $f : \mathbb{R}^n \to \mathbb{R}^1$ defined by $f(\mathbf{x}) = \|\mathbf{x}\|$ is continuous on $\mathbb{R}^n$ (i.e., show that the norm is a continuous function). *Hint:* Use the triangle inequality to show that, for any pair of vectors $\mathbf{x}$ and $\mathbf{y}$, $|\,\|\mathbf{x}\| - \|\mathbf{y}\|\,| \leqslant \|\mathbf{x} - \mathbf{y}\|$.

6. BASIC PROPERTY OF REAL NUMBERS

A set S in $\mathbb{R}^n$ is said to be *bounded* if there exists a positive number M such that $\|\mathbf{x}\| \leqslant M$ for all $\mathbf{x} \in S$. Another way of stating this is $S \subset B(\mathbf{0}, M)$. The number M is said to be a *bound* for the set. Note that if a set is bounded, then there are infinitely many bounds.

We now shall deal with sets in $\mathbb{R}^1$ only. In $\mathbb{R}^1$ a set S being bounded means that, for every x in S, $|x| \leqslant M$, or $-M \leqslant x \leqslant M$. A set S in $\mathbb{R}^1$ is said to be *bounded above* if there exists a real number A such that $x \leqslant A$ for all x in S. A set S is *bounded below* if there exists a real number B such that $x \geqslant B$ for all x in S. The number A is said to be an *upper bound* of S, the number B a *lower bound*.

A number U is said to be a *least upper bound* (l.u.b.), or *supremum* (sup), of a set S

(i) if U is an upper bound of S and

(ii) if U' is another upper bound of S and then $U' \geqslant U$.

If a set has a least upper bound, then the least upper bound is unique. To see this, let U_1 and U_2 be two least upper bounds of a set S. Then since U_2 is an upper bound and U_1 is a least upper bound, $U_2 \geqslant U_1$. Similarly, $U_1 \geqslant U_2$, and so $U_1 = U_2$.

Condition (ii) states that no number $A < U$ can be an upper bound of S. Therefore condition (ii) can be replaced by the equivalent statement:

(ii′) For every $\varepsilon > 0$ there is a number x_ε in S with $U \geqslant x_\varepsilon > U - \varepsilon$.

Let S be the set of numbers $1 - 1/n$, $n = 1, 2, 3, \ldots$. Then one is the l.u.b. of this set. Note that one does not belong to S. Now consider the set $S_2 = S \cup \{2\}$. The number 2 is the l.u.b. of this set and belongs to S_2. The reader should be clear about the distinction between an upper bound and a supremum, or l.u.b.

A number L is said to be a *greatest lower bound* (g.l.b.), or *infimum* (inf), of a set S

(i) if L is a lower bound of S and

(ii) if L' is another lower bound of S and then $L' \leqslant L$.

Observations analogous to those made following the definition of l.u.b. hold for the definition of g.l.b. We leave their formulation to the reader.

Exercise 6.1. Let $L =$ g.l.b. of a set S. Show that there exists a sequence of points $\{x_k\}$ in S such that $x_k \to L$. Show that the sequence $\{x_k\}$ can be taken to be nonincreasing, that is, $x_{k+1} \leqslant x_k$ for every k. Does L have to be a limit point of S?

Exercise 6.2. Let S be a set in $\mathbb{R}^1$. We define $-S$ to be $\{x : -x \in S\}$. Thus $-S$ is the set that we obtain by replacing each element x in S by the element $-x$. Show that S is bounded below if and only if $-S$ is bounded above. Show that α is the g.l.b. of S if and only if $-\alpha$ is the l.u.b. of $-S$.

We can now state the basic property of the real numbers $\mathbb{R}^1$, which is sometimes called the *completeness property*.