Michael Spivak

CALCULUS ON MANIFOLDS

Michael Spivak

CALCULUS ON MANIFOLDS

Michael Spivak
Brandeis University

Calculus on Manifolds

A MODERN APPROACH TO CLASSICAL THEOREMS
OF ADVANCED CALCULUS

The Advanced Book Program

Westview
PRESS
A Member of the Perseus Books Group

Table of Contents

MATHEMATICS MONOGRAPH SERIES

Frederick J. Almgren, Jr., *Institute for Advanced Study*

PLATEAU'S PROBLEM: AN INVITATION TO VARIFOLD GEOMETRY

Robert T. Seeley, *Brandeis University*

AN INTRODUCTION TO FOURIER SERIES

Michael Spivak, *Brandeis University*

CALCULUS ON MANIFOLDS: A MODERN APPROACH TO CLASSICAL

THEOREMS OF ADVANCED CALCULUS

Editors' Foreword

Mathematics has been expanding in all directions at a fabulous rate during the past half century. Ne fields have emerged, the diffusion into other disciplines has proceeded apace, and our knowledge the classical areas has grown ever more profound. At the same time, one of the most striking trends modern mathematics is the constantly increasing interrelationship between its various branches. Thu the present-day students of mathematics are faced with an immense mountain of material. In additio to the traditional areas of mathematics as presented in the traditional manner—and these presentatio do abound—there are the new and often enlightening ways of looking at these traditional areas, ar also the vast new areas teeming with potentialities. Much of this new material is scattered indigestib throughout the research journals, and frequently coherently organized only in the minds unpublished notes of the working mathematicians. And students desperately need to learn more ar more of this material.

This series of brief topical booklets has been conceived as a possible means to tackle and hopeful to alleviate some of these pedagogical problems. They are being written by active researc mathematicians, who can look at the latest developments, who can use these developments to clari and condense the required material, who know what ideas to underscore and what techniques to stres We hope that they will also serve to present to the able undergraduate an introduction to contempora research and problems in mathematics, and that they will be sufficiently informal that the person tastes and attitudes of the leaders in modern mathematics will shine through clearly to the readers.

The area of differential geometry is one in which recent developments have effected great change That part of differential geometry centered about Stokes' Theorem, sometimes called the fundament theorem of multivariate calculus, is traditionally taught in advanced calculus courses (second or thi year) and is essential in engineering and physics as well as in several current and important branch of mathematics. However, the teaching of this material has been relatively little affected by thes modern developments; so the mathematicians must relearn the material in graduate school, and oth scientists are frequently altogether deprived of it. Dr. Spivak's book should be a help to those wh wish to see Stoke's Theorem as the modern working mathematician sees it. A student with a goo course in calculus and linear algebra behind him should find this book quite accessible.

Robert Gunnin

Hugo Ros

Princeton, New Jersey
Waltham, Massachusetts
August 1965

Preface

This little book is especially concerned with those portions of “advanced calculus” in which th subtlety of the concepts and methods makes rigor difficult to attain at an elementary level. Th approach taken here uses elementary versions of modern methods found in sophisticated mathematic The formal prerequisites include only a term of linear algebra, a nodding acquaintance with th notation of set theory, and a respectable first-year calculus course (one which at least mentions th least upper bound (sup) and greatest lower bound (inf) of a set of real numbers). Beyond this a certa (perhaps latent) rapport with abstract mathematics will be found almost essential.

The first half of the book covers that simple part of advanced calculus which generaliz elementary calculus to higher dimensions. Chapter 1 contains preliminaries, and Chapters 2 and treat differentiation and integration.

The remainder of the book is devoted to the study of curves, surfaces, and higher-dimension analogues. Here the modern and classical treatments pursue quite different routes; there are, of cours many points of contact, and a significant encounter occurs in the last section. The very classic equation reproduced on the cover appears also as the last theorem of the book. This theorem (Stoke Theorem) has had a curious history and has undergone a striking metamorphosis.

The first statement of the Theorem appears as a postscript to a letter, dated July 2, 1850, from S William Thomson (Lord Kelvin) to Stokes. It appeared publicly as question 8 on the Smith’s Priz Examination for 1854. This competitive examination, which was taken annually by the be mathematics students at Cambridge University, was set from 1849 to 1882 by Professor Stokes; by th time of his death the result was known universally as Stokes’ Theorem. At least three proofs wer given by his contemporaries: Thomson published one, another appeared in Thomson and Tait *Treatise on Natural Philosophy,* and Maxwell provided another in *Electricity and Magnetism* [13 Since this time the name of Stokes has been applied to much more general results, which have figure so prominently in the development of certain parts of mathematics that Stokes’ Theorem may b considered a case study in the value of generalization.

In this book there are three forms of Stokes’ Theorem. The version known to Stokes appears in th last section, along with its inseparable companions, Green’s Theorem and the Divergence Theorer These three theorems, the classical theorems of the subtitle, are derived quite easily from a mode Stokes’ Theorem which appears earlier in Chapter 5. What the classical theorems state for curves ar surfaces, this theorem states for the higher-dimensional analogues (manifolds) which are studie thoroughly in the first part of Chapter

5. This study of manifolds, which could be justified solely on the basis of their importance modern mathematics, actually involves no more effort than a careful study of curves and surfac alone would require.

The reader probably suspects that the modern Stokes’ Theorem is at least as difficult as th classical theorems derived from it. On the contrary, it is a very simple consequence of yet anoth version of Stokes’ Theorem; this very abstract version is the final and main result of Chapter 4. It entirely reasonable to suppose that the difficulties so far avoided must be hidden here. Yet the proof this theorem is, in the mathematician’s sense, an utter triviality—a straight-forward computation. C the other hand, even the statement of this triviality cannot be understood without a horde of difficu definitions from Chapter 4. There are good reasons why the theorems should all be easy and th definitions hard. As the evolution of Stokes' Theorem revealed, a single simple principle, ca masquerade as several difficult results; the proofs of many theorems involve merely stripping awa the disguise. The definitions, on the other hand, serve a twofold purpose: they are rigorou replacements for vague notions, and machinery for elegant proofs. The first two sections of Chapter define precisely, and prove the rules for manipulating, what are classically described as "expressio of the form" $P\,dx + Q\,dy + R\,dz$, or $P\,dx\,dy + Q\,dy\,dz + R\,dz\,dx$. Chains, defined in the third sectio and partitions of unity (already introduced in Chapter 3) free our proofs from the necessity chopping manifolds up into small pieces; they reduce questions about manifolds, where everythin seems hard, to questions about Euclidean space, where everything is easy.

Concentrating the depth of a subject in the definitions is undeniably economical, but it is bound produce some difficulties for the student. I hope the reader will be encouraged to learn Chapter thoroughly by the assurance that the results will justify the effort: the classical theorems of the la section represent only a few, and by no means the most important, applications of Chapter 4; man others appear as problems, and further developments will be found by exploring the bibliography.

The problems and the bibliography both deserve a few words. Problems appear after every sectio and are numbered (like the theorems) within chapters. I have starred those problems whose results a used in the text, but this precaution should be unnecessary—the problems are the most important pa of the book, and the reader should at least attempt them all. It was necessary to make the bibliograph either very incomplete or unwieldy, since half the major branches of mathematics could legitimate be recommended as reasonable continuations of the material in the book. I have tried to make incomplete but tempting.

Many criticisms and suggestions were offered during the writing of this book. I am particular grateful to Richard Palais, Hugo Rossi, Robert Seeley, and Charles Stenard for their many helpf comments.

I have used this printing as an opportunity to correct many misprints and minor errors pointed o to me by indulgent readers. In addition, the material following Theorem 3-11 has been complete revised and corrected. Other important changes, which could not be incorporated in the text witho excessive alteration, are listed in the Addenda at the end of the book.

Michael Spiv

Waltham, Massachusetts
March 1968

Functions on Euclidean Space

NORM AND INNER PRODUCT

Euclidean *n*-space $\mathbf{R}^n$ is defined as the set of all n-tuples $(x^1, \ldots, x^n)$ of real numbers x^i (a "1-tuple numbers" is just a number and $\mathbf{R}^1 = \mathbf{R}$, the set of all real numbers). An element of $\mathbf{R}^n$ is often called point in $\mathbf{R}^n$, and $\mathbf{R}^1$ $\mathbf{R}^2$, $\mathbf{R}^3$ are often called the line, the plane, and space, respectively. If x denotes a element of $\mathbf{R}^n$, then x is an n-tuple of numbers, the ith one of which is denoted x^i; thus we can write

$$x = (x^1, \ldots, x^n).$$

A point in $\mathbf{R}^n$ is frequently also called a vector in $\mathbf{R}^n$, because $\mathbf{R}^n$, with $x + y = (x^1 + y^1, \ldots, x^n + y$ and $ax = (ax^1, \ldots, ax^n)$, as operations, is a vector space (over the real numbers, of dimension n). this vector space there is the notion of the length of a vector x, usually called the **norm** $|x|$ of x ar defined by $|x| = \sqrt{(x^1)^2 + \cdots + (x^n)^2}$. If $n = 1$, then $|x|$ is the usual absolute value of x. The relatio between the norm and the vector space structure of $\mathbf{R}^n$ is very important.

1-1 Theorem. *If* $x, y \in \mathbf{R}^n$ *and* $a \in \mathbf{R}$, *then*

1. $|x| \geq 0$, *and* $|x| = 0$ *if and only if* $x = 0$.
2. $\left|\sum_{i=1}^n x^i y^i\right| \leq |x| \cdot |y|$; *equality holds if and only if x and y are linearly dependent.*
3. $|x + y| \leq |x| + |y|$.
4. $|ax| = |a| \cdot |x|$.

Proof

1. is left to the reader.
2. If x and y are linearly dependent, equality clearly holds. If not, then $\lambda y — x \neq 0$ for all $\lambda \in \mathbf{R}$, s

$$\begin{aligned} 0 < |\lambda y - x|^2 &= \sum_{i=1}^n (\lambda y^i - x^i)^2 \\ &= \lambda^2 \sum_{i=1}^n (y^i)^2 - 2\lambda \sum_{i=1}^n x^i y^i + \sum_{i=1}^n (x^i)^2. \end{aligned}$$

Therefore the right side is a quadratic equation in λ with no real solution, and its discriminant must be negative. Thus

$$4\left(\sum_{i=1}^n x^i y^i\right)^2 - 4\sum_{i=1}^n (x^i)^2 \cdot \sum_{i=1}^n (y^i)^2 < 0.$$

3. $$\begin{aligned} |x+y|^2 &= \Sigma_{i=1}^n (x^i + y^i)^2 \\ &= \Sigma_{i=1}^n (x^i)^2 + \Sigma_{i=1}^n (y^i)^2 + 2\Sigma_{i=1}^n x^i y^i \\ &\leq |x|^2 + |y|^2 + 2|x| \cdot |y| \qquad \text{by (2)} \\ &= (|x| + |y|)^2. \end{aligned}$$

4. $|ax| = \sqrt{\Sigma_{i=1}^n (ax^i)^2} = \sqrt{a^2 \Sigma_{i=1}^n (x^i)^2} = |a| \cdot |x|$.

The quantity $\Sigma_{i=1}^n x^i y^i$ which appears in (2) is called the **inner product** of x and y and denoted $\langle x,y$ The most important properties of the inner product are the following.

1-2 Theorem. *If x, x_1, x_2 and y, y_1, y_2 are vectors in $\mathbf{R}^n$ and $a \in \mathbf{R}$, then*

1. $\langle x,y \rangle = \langle y,x \rangle$ *(symmetry*
2. $\langle ax,y \rangle = \langle x,ay \rangle = a\langle x,y \rangle$ *(bilinearity*
   $\langle x_1 + x_2, y \rangle = \langle x_1,y \rangle + \langle x_2,y \rangle$ $\langle x, y_1 + y_2 \rangle = \langle x,y_1 \rangle + \langle x,y_2 \rangle$
3. $\langle x,x \rangle \geq 0$, *and* $\langle x,x \rangle = 0$ *if and* *(positive definiteness*
   only if $x = 0$
4. $|x| = \sqrt{\langle x,x \rangle}$.
5. $\langle x,y \rangle = \dfrac{|x+y|^2 - |x-y|^2}{4}$ *(polarization identity*

Proof

1. $\langle x,y \rangle = \Sigma_{i=1}^n x^i y^i = \Sigma_{i=1}^n y^i x^i = \langle y,x \rangle$.
2. By (1) it suffices to prove

$$\begin{aligned} \langle ax,y \rangle &= a\langle x,y \rangle, \\ \langle x_1 + x_2, y \rangle &= \langle x_1,y \rangle + \langle x_2,y \rangle. \end{aligned}$$

   These follow from the equations

$$\begin{aligned} \langle ax,y \rangle &= \sum_{i=1}^n (ax^i) y^i = a \sum_{i=1}^n x^i y^i = a\langle x,y \rangle, \\ \langle x_1 + x_2, y \rangle &= \sum_{i=1}^n ({x_1}^i + {x_2}^i) y^i = \sum_{i=1}^n {x_1}^i y^i + \sum_{i=1}^n {x_2}^i y^i \\ &= \langle x_1,y \rangle + \langle x_2,y \rangle. \end{aligned}$$

3. and
4. are left to the reader.

5. $\frac{|x+y|^2 - |x-y|^2}{4}$

$$\begin{aligned} &= \tfrac{1}{4}[\langle x+y, x+y\rangle - \langle x-y, x-y\rangle] \qquad \text{by (4)} \\ &= \tfrac{1}{4}[\langle x,x\rangle + 2\langle x,y\rangle + \langle y,y\rangle - (\langle x,x\rangle - 2\langle x,y\rangle + \langle y,y\rangle)] \\ &= \langle x,y\rangle. \quad \blacksquare \end{aligned}$$

We conclude this section with some important remarks about notation. The vector (0, . . . ,0) wi usually be denoted simply 0. The **usual basis** of $\mathbf{R}^n$ is $e_1, \ldots, e_n$, where $e_i = (0, \ldots, 1, \ldots, 0)$, with th 1 in the *i*th place. If $T\colon \mathbf{R}^n \to \mathbf{R}^m$ is a linear transformation, the matrix of T with respect to the usu bases of $\mathbf{R}^n$ and $\mathbf{R}^m$ is the $m \times n$ matrix $A = (a_{ij})$, where $T(e_i) = \sum_{j=1}^m a_{ji} e_j$—the coefficients of $T(e_i)$ appe in the *i*th *column* of the matrix. If $S\colon \mathbf{R}^m \to \mathbf{R}^p$ has the $p \times m$ matrix B, then $S \circ T$ has the $p \times n$ matr BA [here $S \circ T(x) = S(T(x))$; most books on linear algebra denote $S \circ T$ simply ST]. To find $T(x)$ or computes the $m \times 1$ matrix

$$\begin{pmatrix} y^1 \\ \cdot \\ \cdot \\ \cdot \\ y^m \end{pmatrix} = \begin{pmatrix} a_{11}, & \ldots & , a_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{m1}, & \ldots & , a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x^1 \\ \cdot \\ \cdot \\ \cdot \\ x^n \end{pmatrix};$$

then $T(x) = (y^1, \ldots, y^m)$. One notational convention greatly simplifies many formulas: if $x \in \mathbf{R}^n$ an $y \in \mathbf{R}^m$, then (x,y) denotes

$$(x^1, \ldots, x^n, y^1, \ldots, y^m) \in \mathbf{R}^{n+m}.$$

Problems. **1-1.*** Prove that $|x| \le \sum_{i=1}^n |x^i|$

1-2. When does equality hold in Theorem 1-1(3)? *Hint:* Re-examine the proof; the answer is not "when x and y are linearly dependent."

1-3. Prove that $|x - y| \le |x| + |y|$. When does equality hold?

1-4. Prove that $\big|\, |x| - |y| \,\big| \le |x - y|$.

1-5. The quantity $|y - x|$ is called the **distance** between x and y. Prove and interpret geometricall the "triangle inequality": $|z - x| \le |z - y| + |y - x|$.

1-6. Let f and g be integrable on [a,b].

(a) Prove that $\left|\int_a^b f \cdot g\right| \le \left(\int_a^b f^2\right)^{\frac12} \cdot \left(\int_a^b g^2\right)^{\frac12}$. *Hint:* Consider separately the cases $0 = \int_a^b (f - \lambda g)^2$ for some $\lambda \in \mathbf{R}$ and $0 < \int_a^b (f - \lambda g)^2$ for all $\lambda \in \mathbf{R}$

(b) If equality holds, must $f = \lambda g$ for some $\lambda \in \mathbf{R}$? What if f and g are continuous?

(c) Show that Theorem 1-1(2) is a special case of (a).

1-7. A linear transformation $T\colon \mathbf{R}^n \to \mathbf{R}^n$ is **norm preserving** if $|T(x)| = |x|$, and **inner product preserving** if $\langle Tx, Ty\rangle = \langle x,y\rangle$.

(a) Prove that T is norm preserving if and only if T is innerproduct preserving.

(b) Prove that such a linear transformation T is 1-1 and T^{-1} is of the same sort.

1-8. If $x,y \in \mathbf{R}^n$ are non-zero, the **angle** between x and y, denoted $\angle(x,y)$, is defined as arccos $(\langle x,y\rangle / |x| \cdot |y|)$, which makes sense by Theorem 1-1 (2). The linear transformation T is **angle**

preserving if T is 1-1, and for $x,y \neq 0$ we have $\angle(Tx,Ty) = \angle(x,y)$.

(a) Prove that if T is norm preserving, then T is angle preserving.

(b) If there is a basis $x_1, \ldots, x_n$ of $\mathbf{R}^n$ and numbers $\lambda_1, \ldots, \lambda_n$ such that $Tx_i = \lambda_i x_i$, prove th T is angle preserving if and only if all $|\lambda_i|$ are equal.

(c) What are all angle preserving $T: \mathbf{R}^n \to \mathbf{R}^n$?

1-9. If $0 \leq \theta < \pi$, let $T: \mathbf{R}^2 \to \mathbf{R}^2$ have the matrix $\begin{pmatrix} \cos\theta, & \sin\theta \\ -\sin\theta, & \cos\theta \end{pmatrix}$. Show that T is angle preserving and if $x \neq 0$, then $\angle(x,Tx) = \theta$.

1-10.* If $T: \mathbf{R}^m \to \mathbf{R}^n$ is a linear transformation, show that there is a number M such that $|T(h)| \leq M|h|$ for $h \in \mathbf{R}^m$. *Hint:* Estimate $|T(h)|$ in terms of $|h|$ and the entries in the matrix of T.

1-11. If $x,y \in \mathbf{R}^n$ and $z,w \in \mathbf{R}^m$, show that $\langle(x,z),(y,w)\rangle = \langle x,y\rangle + (z,w)$ and $|(x,z)| = \sqrt{|x|^2 + |z|^2}$. Not that (x,z) and (y,w) denote points in $\mathbf{R}^{n+m}$.

1-12. * Let $(\mathbf{R}^n)^*$ denote the dual space of the vector space $\mathbf{R}^n$. If $x \in \mathbf{R}^n$, define $\phi_x \in (\mathbf{R}^n)^*$ by $\phi_x(y) = \langle x,y\rangle$. Define $T: \mathbf{R}^n \to (\mathbf{R}^n)^*$ by $T(x) = \phi_x$. Show that T is a 1-1 linear transformation and conclude that every $\phi \in (\mathbf{R}^n)^*$ is ϕ_x for a unique $x \in \mathbf{R}^n$.

1-13.* If $x,y \in \mathbf{R}^n$, then x and y are called **perpendicular** (or **orthogonal**) if $\langle x,y\rangle = 0$. If x and y are perpendicular, prove that $|x + y|^2 = |x|^2 + |y|^2$.

SUBSETS OF EUCLIDEAN SPACE

The closed interval [a,b] has a natural analogue in $\mathbf{R}^2$. This is the **closed rectangle** $[a,b] \times [c,c$ defined as the collection of all pairs (x,y) with $x \in [a,b]$ and $y \in [c,d]$. More generally, if $A \subset \mathbf{R}^m$ ar $B \subset \mathbf{R}^n$, then $A \times B \subset \mathbf{R}^{m+n}$ is defined as the set of all $(x,y) \in \mathbf{R}^{m+n}$ with $x \in A$ and $y \in B$. particular, $\mathbf{R}^{m+n} = \mathbf{R}^m \times \mathbf{R}^n$. If $A \subset \mathbf{R}^m$, $B \subset \mathbf{R}^n$, and $C \subset \mathbf{R}^p$, then $(A \times B) \times C = A \times (B \times C)$, and bo of these are denoted simply $A \times B \times C$; this convention is extended to the product of any number sets. The set $[a_1,b_1] \times \cdots \times [a_n,b_n] \subset \mathbf{R}^n$ is called **a closed rectangle** in $\mathbf{R}^n$, while the set $(a_1,b_1) \times \cdot \times (a_n,b_n) \subset \mathbf{R}^n$ is called an **open rectangle.** More generally a set $U \subset \mathbf{R}^n$ is called **open** (Figure 1-1) for each $x \in U$ there is an open rectangle A such that $x \in A \subset U$.

A subset C of $\mathbf{R}^n$ is **closed** if $\mathbf{R}^n$—C is open. For example, if C contains only finitely many point then C is closed.

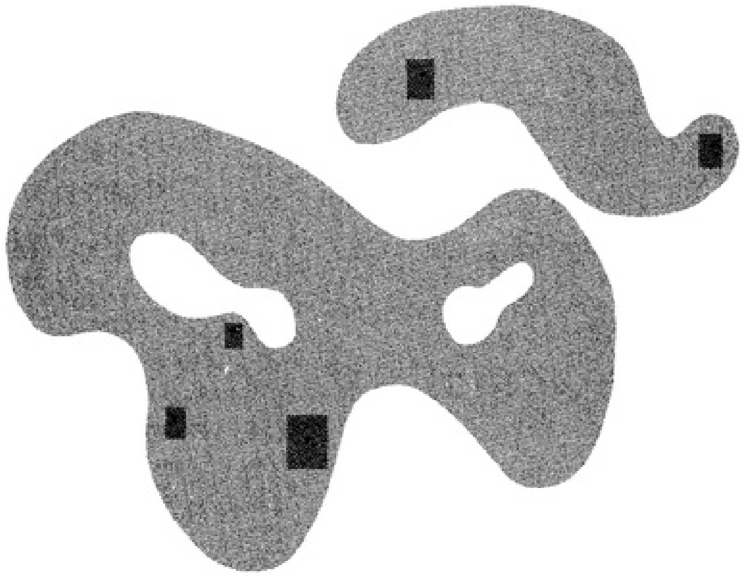

FIGURE 1-1

The reader should supply the proof that a closed rectangle in $\mathbf{R}^n$ is indeed a closed set.

If $A \subset \mathbf{R}^n$ and $x \in \mathbf{R}^n$**,** then one of three possibilities must hold (Figure 1-2):

1. There is an open rectangle B such that $x \in B \subset A$.
2. There is an open rectangle B such that $x \in B \subset \mathbf{R}^n$—$A$.
3. If B is any open rectangle with $x \in B$, then B contains points of both A and $\mathbf{R}^n$—A.

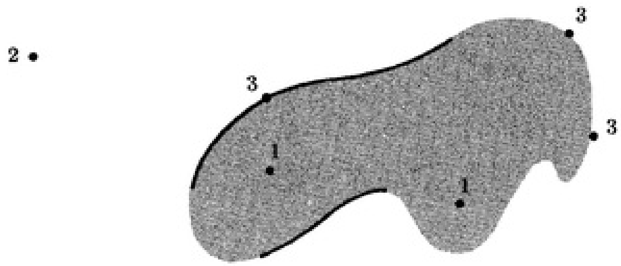

FIGURE 1-2

Those points satisfying (1) constitute the **interior** of A, those satisfying (2) the **exterior** of A, an
those satisfying (3) the **boundary** of A. Problems 1-16 to 1-18 show that these terms may sometim
have unexpected meanings.

It is not hard to see that the interior of any set A is open, and the same is true for the exterior of A
which is, in fact, the interior of $\mathbf{R}^n$ — A. Thus (Problem 1-14) their union is open, and what remair
the boundary, must be closed.

A collection O of open sets is an **open cover** of A (or, briefly, **covers** A) if every point $x \in A$ is
some open set in the collection O. For example, if O is the collection of all open intervals (a, a +
for $a \in \mathbf{R}$, then O is a cover of $\mathbf{R}$. Clearly no finite number of the open sets in O will cover $\mathbf{R}$ or, fo
that matter, any unbounded subset of $\mathbf{R}$. A similar situation can also occur for bounded sets. If O is th
collection of all open intervals $(1/n, 1 - 1/n)$ for all integers $n > 1$, then O is an open cover of (0,1), b
again no finite collection of sets in O will cover (0,1). Although this phenomenon may not appe

particularly scandalous, sets for which this state of affairs cannot occur are of such importance th
they have received a special designation: a set A is called **compact** if every open cover O contains
finite subcollection of open sets which also covers A.

A set with only finitely many points is obviously compact and so is the infinite set A which contai
0 and the numbers $1/n$ for all integers n (reason: if O is a cover, then $0 \in U$ for some open set U in
there are only finitely many other points of A not in U, each requiring at most one more open set).

Recognizing compact sets is greatly simplified by the following results, of which only the first h
any depth (i.e., uses any facts about the real numbers).

1-3 Theorem (***Heine-Borel***). *The closed interval* $[a,b]$ *is compact.*

Proof. If O is an open cover of $[a,b]$, let
$A = \{x: a \le x \le b$ and $[a,x]$ is covered by some finite number of open sets in O$\}$.

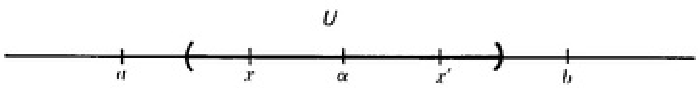

FIGURE 1-3

Note that $a \in A$ and that A is clearly bounded above (by b). We would like to show that $b \in A$. This
done by proving two things about α = least upper bound of A; namely, (1) $\alpha \in$ and (2) $b = \alpha$.

Since O is a cover, $\alpha \in U$ for some U in O. Then all points in some interval to the left of a are al
in U (see Figure 1-3). Since α is the least upper bound of A, there is an x in this interval such that x
A. Thus $[a,x]$ is covered by some finite number of open sets of O, while $[x,\alpha]$ is covered by the sing
set U. Hence $[a,\alpha]$ is covered by a finite number of open sets of O, and $\alpha \in A$. This proves (1).

To prove that (2) is true, suppose instead that $\alpha < b$. Then there is a point x' between α and b su
that $[\alpha,x'] \subset U$. Since $\alpha \in A$, the interval $[a,\alpha]$ is covered by finitely many open sets of 0, while $[\alpha,x$
is covered by U. Hence $x' \in A$, contradicting the fact that α is an upper bound of A.

If $B \subset \mathbf{R}^m$ is compact and $x \in \mathbf{R}^n$, it is easy to see that $\{x\} \times B \subset \mathbf{R}^{n+m}$ is compact. However,
much stronger assertion can be made.

1-4 Theorem. *If B is compact and O is an open cover of $\{x\} \times B$, then there is an open set $U \subset$*
containing x such that $U \times B$ is covered by a finite number of sets in O.

Proof. Since $\{x\} \times B$ is compact, we can assume at the outset that O is finite, and we need only fi
the open set U such that $U \times B$ is covered by O.

For each $y \in B$ the point (x,y) is in some open set W in O. Since W is open, we have $(x,y) \in U_y \times$
$\subset W$ for some open rectangle $U_y \times V_y$. The sets V_y cover the compact set B, so a finite number V_{y1}, .
,V_{yk} also cover B. Let $U = U_{y1} \cap \cdots \cap U_{yk}$. Then if $(x',y') \in U \times B$, we have $y' \in V_{yi}$ for some i (Figu
1-4), and certainly $x' \in U_{yi}$. Hence $(x',y') \in U_{yi} \times V_{yi}$, which is contained in some W in O.

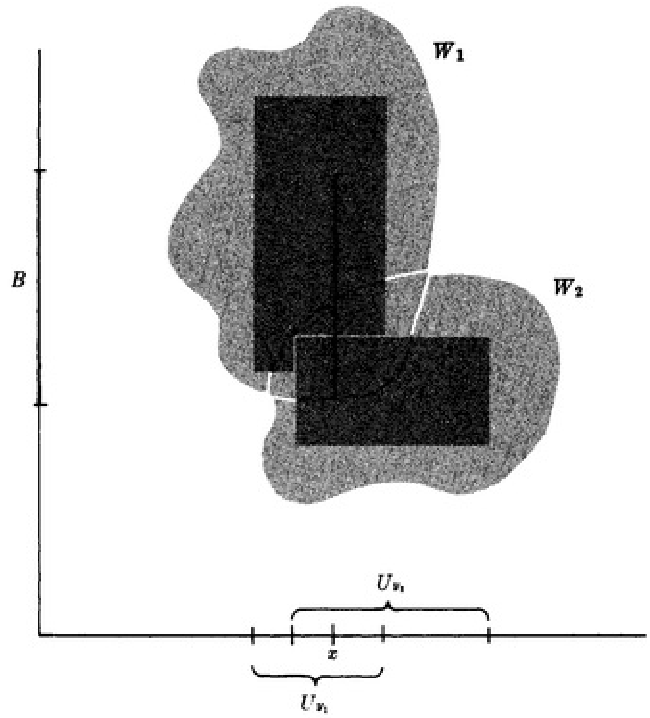

FIGURE 1-4

1-5 Corollary. *If $A \subset \mathbf{R}^n$ and $B \subset \mathbf{R}^m$ are compact, then $A \times B \subset \mathbf{R}^{n+m}$ is compact.*
Proof. If O is an open cover of $A \times B$, then O covers $\{x\} \times B$ for each $x \in A$. By Theorem 1-4 there an open set U_x containing x such that $U_x \times B$ is covered by finitely many sets in O. Since A is compac a finite number $U_{x1}, \ldots, U_{xn}$ of the U_x cover A. Since finitely many sets in O cover each $U_{xi} \times$ finitely many cover all of $A \times B$.

1-6 Corollary. *$A_1 \times \cdots \times A_k$ is compact if each A_i is. In particular, a closed rectangle in $\mathbf{R}^k$ is compact*
1-7 Corollary. ***A*** *closed bounded subset of* $\mathbf{R}^n$ *is compact*. (The converse is also true (Problem 1-20).)
Proof. If $A \subset \mathbf{R}^n$ is closed and bounded, then $A \subset B$ for some closed rectangle B. If O is an open cov of A, then O together with $\mathbf{R}^n - A$ is an open cover of B. Hence a finite number $U_1, \ldots, U_n$ of sets O, together with $\mathbf{R}^n$—A perhaps, cover B. Then $U_1, \ldots, U_n$ cover A.

Problems. **1-14.*** Prove that the union of any (even infinite) number of open sets is ope Prove that the intersection of two (and hence of finitely many) open sets is open. Give counterexample for infinitely many open sets.

1-15. Prove that $\{x \in \mathbf{R}^n: |x - a| < r\}$ is open (see also Problem 1-27).
1-16. Find the interior, exterior, and boundary of the sets

$\{x \in R^n: |x| \leq 1\}$
$\{x \in R^n: |x| = 1\}$
$\{x \in \mathbf{R}^n:$ each x^i is rational$\}$.

1-17. Construct a set $A \subset [0,1] \times [0,1]$ such that A contains at most one point on each horizontal and each vertical line but boundary $A = [0,1] \times [0,1]$. *Hint:* It suffices to ensure that A contain points in each quarter of the square $[0,1] \times [0,1]$ and also in each sixteenth, etc.

1-18. If $A \subset [0,1]$ is the union of open intervals (a_i,b_i) such that each rational number in $(0,1)$ is contained in some (a_i,b_i), show that boundary $A = [0,1] - A$.

1-19.* If A is a closed set that contains every rational number $r \in [0,1]$, show that $[0,1] \subset A$.

1-20. Prove the converse of Corollary 1-7: A compact subset of $\mathbf{R}^n$ is closed and bounded (see also Problem 1-28).

1-21.* (a) If A is closed and $x \notin A$, prove that there is a number $d > 0$ such that $|y-x| \geq d$ for all $y \in A$.

(b) If A is closed, B is compact, and $A \cap B = \emptyset$, prove that there is $d > 0$ such that $|y-x| \geq d$ for all $y \in A$ and $x \in B$. *Hint:* For each $b \in B$ find an open set U containing b such that this relation holds for $x \in U \cap B$.

(c) Give a counterexample in $\mathbf{R}^2$ if A and B are closed but neither is compact.

1-22.* If U is open and $C \subset U$ is compact, show that there is a compact set D such that $C \subset$ interior D and $D \subset U$.

FUNCTIONS AND CONTINUITY

A **function** from $\mathbf{R}^n$ to $\mathbf{R}^m$ (sometimes called a (vector-valued) function of n variables) is a rule which associates to each point in $\mathbf{R}^n$ some point in $\mathbf{R}^m$; the point a function f associates to x is denoted $f(x)$. We write $f: \mathbf{R}^n \to \mathbf{R}^m$ (read "f takes $\mathbf{R}^n$ into $\mathbf{R}^m$" or "f, taking $\mathbf{R}^n$ into $\mathbf{R}^m$," depending on context) to indicate that $f(x) \in \mathbf{R}^m$ is defined for $x \in \mathbf{R}^n$. The notation $f : A \to \mathbf{R}^m$ indicates that $f(x)$ is defined only for x in the set A, which is called the **domain** of f. If $B \subset A$, we define $f(B)$ as the set of all $f(x)$ for $x \in B$, and if $C \subset \mathbf{R}^m$ we define $f^{-1}(C) = \{x \in A : f(x) \in C\}$. The notation $f: A \to B$ indicates that $f(A) \subset B$.

A convenient representation of a function $f: \mathbf{R}^2 \to \mathbf{R}$ may be obtained by drawing a picture of its graph, the set of all 3-tuples of the form $(x,y,f(x,y))$, which is actually a figure in 3-space (see, e.g., Figures 2-1 and 2-2 of Chapter 2).

If $f,g: \mathbf{R}^n \to \mathbf{R}$, the functions $f + g$, $f-g$, $f \cdot g$, and f/g are defined precisely as in the one-variable case. If $f: A \to \mathbf{R}^m$ and $g: B \to \mathbf{R}^p$, where $B \subset \mathbf{R}^m$, then the **composition** $g \circ f$ is defined by $g \circ f(x) = g(f(x))$; the domain of $g \circ f$ is $A \cap f^{-1}(B)$. If $f: A \to \mathbf{R}^m$ is 1-1, that is, if $f(x) \neq f(y)$ when $x \neq y$, we define $f^{-1}: f(A) \to \mathbf{R}^n$ by the requirement that $f^{-1}(z)$ is the unique $x \in A$ with $f(x) = z$.

A function $f : A \to \mathbf{R}^m$ determines m **component functions** $f^1, \ldots, f^m: A \to \mathbf{R}$ by $f(x) = (f^1(x), \ldots, f^m(x))$. If conversely, m functions $g_1, \ldots, g_m: A \to \mathbf{R}$ are given, there is a unique function $f: A \to \mathbf{R}^m$ such that $f^i = g_i$, namely $f(x) = (g_1(x), \ldots, g_m(x))$. This function f will be denoted $(g_1, \ldots, g_m)$, so that we always have $f = (f^1, \ldots, f^m)$. If $\pi: \mathbf{R}^n \to \mathbf{R}^n$ is the identity function, $\pi(x) = x$, then $\pi^i(x) = x^i$; the function π^i is called the ith **projection function.**

The notation $\lim_{x \to a} f(x) = b$ means, as in the one-variable case, that we can get $f(x)$ as close to b as desired, by choosing x sufficiently close to, but not equal to, a. In mathematical terms this means that for every number $\varepsilon > 0$ there is a number $\delta > 0$ such that $|f(x) - b| < \varepsilon$ for all x in the domain of f which satisfy $0 < |x - a| < \delta$. A function $f: A \to \mathbf{R}^m$ is called **continuous** at $a \in A$ if $\lim_{x \to a} f(x) = f(a)$, and f

simply called continuous if it is continuous at each $a \in A$. One of the pleasant surprises about th concept of continuity is that it can be defined without using limits. It follows from the next theore that $f\colon \mathbf{R}^n \to \mathbf{R}^m$ is continuous if and only if $f^{-1}(U)$ is open whenever $U \subset \mathbf{R}^m$ is open; if the domain f is not all of $\mathbf{R}^n$, a slightly more complicated condition is needed.

1-8 Theorem. *If $A \subset \mathbf{R}^n$, a function $f\colon A \to \mathbf{R}^m$ is continuous if and only if for every open set $U \subset$ R there is some open set $V \subset \mathbf{R}^n$ such that $f^{-1}(U) = V \cap A$.*

Proof. Suppose f is continuous. If $a \in f^{-1}(U)$, then $f(a) \in U$. Since U is open, there is an ope rectangle B with $f(a) \in B \subset U$. Since f is continuous at a, we can ensure that $f(x) \in B$, provided w choose x in some sufficiently small rectangle C containing a. Do this for each $a \in f^{-1}(U)$ and let V b the union of all such C. Clearly $f^{-1}(U) = V \cap A$. The converse is similar and is left to the reader.

The following consequence of Theorem 1-8 is of great importance.

1-9 Theorem. If f*: A* $\to \mathbf{R}^m$ *is continuous, where* $A \subset \mathbf{R}^n$, *and A is compact, then $f(A) \subset \mathbf{R}^m$ compact*.

Proof. Let O be an open cover of f(A). For each open set U in O there is an open set V_U such th $f^{-1}(U) = V_U \cap A$. The collection of all V_U is an open cover of A. Since A is compact, a finite numbe $V_{U_1}, \dots, V_{U_n}$ cover A. Then $U_1, \dots, U_n$ cover $f(A)$.

If $f\colon A \to \mathbf{R}$ is bounded, the extent to which f fails to be continuous at $a \in A$ can be measured in precise way. For $\delta > 0$ let

$$M(a,f,\delta) = \sup\{f(x) : x \in A \text{ and } |x-a| < \delta\},$$
$$m(a,f,\delta) = \inf\{f(x) : x \in A \text{ and } |x-a| < \delta\}.$$

The **oscillation** $o(f,a)$ of f at a is defined by $o(f,a) = \lim_{\delta \to 0}[M(a,f,\delta) - m(a,f,\delta)]$. This limit always exist since $M(a,f,\delta) - m(a,f,\delta)$ decreases as δ decreases. There are two important facts about $o(f,a)$.

1-10 Theorem. *The bounded function f is continuous at a if and only if* $o(f,a) = 0$.

Proof. Let f be continuous at a. For every number $\varepsilon > 0$ we can choose a number $\delta > 0$ so that $|f($ $-f(a)| < \varepsilon$ for all $x \in A$ with $|x-a| < \delta$; thus $M(a,f,\delta) - m(a,f,\delta) \le 2\varepsilon$. Since this is true for every ε, w have $o(f,a) = 0$. The converse is similar and is left to the reader.

1-11 Theorem. *Let $A \subset \mathbf{R}^n$ be closed. If $f\colon A \to \mathbf{R}$ is any bounded function, and $\varepsilon > 0$, then $\{x \in$ $o(f,x) \ge \varepsilon\}$ is closed.*

Proof. Let ***B*** = $\{x \in A : o(f,x) \geq \varepsilon\}$. We wish to show that $\mathbf{R}^n - B$ is open. If $x \in \mathbf{R}^n - B$, then either *x* *A* or else $x \in A$ and $o(f,x) < \varepsilon$. In the first case, since *A* is closed, there is an open rectangle containing *x* such that $C \subset \mathbf{R}^n - A \subset \mathbf{R}^n - B$. In the second case there is a $\delta > 0$ such that $M(x,f,\delta)$ $m(x,f,\delta) < \varepsilon$. Let *C* be an open rectangle containing x such that $|x—y| < \delta$ for all $y \in C$. Then if $y \in$ there is a δ_1 such that $|x - z| < \delta$ for all *z* satisfying $|z - y| < \delta_1$. Thus $M(y,f,\delta_1) - m(y,f,\delta_1) <$ e, a consequently $o(y,f) < \varepsilon$. Therefore $C \subset \mathbf{R}^n$ – B.

Problems. **1-23**. If $f: A \to \mathbf{R}^m$ and $a \in A$, show that $\lim_{x\to a} f(x) = b$ if and only if $\lim_{x\to a} f^i(x) = b^i$ for $i = 1, . . ,m$.

1-24. Prove that $f: A \to \mathbf{R}^m$ is continuous at *a* if and only if each f^i is.

1-25. Prove that a linear transformation $T: \mathbf{R}^n \to \mathbf{R}^m$ is continuous. *Hint*: Use Problem 1-10.

1-26. Let $A = \{(x,y) \in \mathbf{R}^2: x > 0 \text{ and } 0 < y < x^2\}$.

(a) Show that every straight line through (0,0) contains an interval around (0,0) which is $\mathbf{R}^2 - A$.

(b) Define $f: \mathbf{R}^2 \to \mathrm{R}$ by $f(x) = 0$ if $x \notin$ A and $f(x) = 1$ if $x \in A$. For $h \in \mathbf{R}^2$ define $g_h: \mathbf{R} \to$ by $g_h(t) = f(th)$. Show that each g_h is continuous at 0, but *f* is not continuous at (0,0).

1-27. Prove that $\{x \in \mathbf{R}^n: |x - a| < r\}$ is open by considering the function $f: \mathbf{R}^n \to \mathbf{R}$ with $f(x) = |$ $- a|$.

1-28. If $A \subset \mathbf{R}^n$ is not closed, show that there is a continuous function $f: A \to \mathbf{R}$ which is unbounded. ***Hint:*** If $x \in \mathbf{R}^n - A$ but $x \notin$ interior $(\mathbf{R}^n - A)$, let $f(y) = 1/|y - x|$.

1-29. If *A* is compact, prove that every continuous function $f: A \to \mathbf{R}$ takes on a maximum and a minimum value.

1-30. Let $f: [a,b] \to \mathbf{R}$ be an increasing function. If $x_1, . . . ,x_n \in [a,b]$ are distinct, show that $\sum_{i=1}^{n} o(f,x_i) < f(b) - f(a)$.

2

Differentiation

BASIC DEFINITIONS

Recall that a function f: **R** → **R** is differentiable at $a \in$ **R** if there is a number $f'(a)$ such that

$$(1)\quad \lim_{h\to 0}\frac{f(a+h)-f(a)}{h}=f'(a).$$

This equation certainly makes no sense in the general case of a function f: $\mathbf{R}^n \to \mathbf{R}^m$, but can b reformulated in a way that does. If λ: **R** → **R** is the linear transformation defined by $\lambda(h) = f'(a) \cdot$ then equation (1) is equivalent to

$$(2)\quad \lim_{h\to 0}\frac{f(a+h)-f(a)-\lambda(h)}{h}=0.$$

Equation (2) is often interpreted as saying that $\lambda + f(a)$ is a good approximation to f at a (see Proble 2-9). Henceforth we focus our attention on the linear transformation λ and reformulate the definitio of differentiability as follows.

A function f : **R** → **R** is differentiable at $a \in$ **R** if there is a linear transformation λ : **R** → **R** suc that

$$\lim_{h\to 0}\frac{f(a+h)-f(a)-\lambda(h)}{h}=0.$$

In this form the definition has a simple generalization to higher dimensions:

A function $f : \mathbf{R}^n \to \mathbf{R}^m$ is **differentiable** at $a \in \mathbf{R}^n$ if there is a linear transformation $\lambda : \mathbf{R}^n \to$ R such that

$$\lim_{h\to 0}\frac{|f(a+h)-f(a)-\lambda(h)|}{|h|}=0.$$

Note that h is a point of $\mathbf{R}^n$ and $f(a + h)—f(a)—\lambda(h)$ a point of $\mathbf{R}^m$, so the norm signs are essentia The linear transformation λ is denoted $Df(a)$ and called the **derivative** of f at a. The justification f the phrase “*the* linear transformation λ” is