CALCULUS OF SEVERAL VARIABLES

SERGE LANG
Yale University

ADDISON-WESLEY PUBLISHING COMPANY
Reading, Massachusetts · Menlo Park, California · London · Don Mills, Ontario

This book is in the
ADDISON-WESLEY SERIES IN MATHEMATICS

LYNN H. LOOMIS
Consulting Editor

Cover photograph of a thunderstorm by Ernst Haas. Photograph appears in *The Creation* published by The Viking Press, Inc., 1971.

Printed in the United States of America. Published simultaneously in Canada. Library of Congress Catalog Card No. 74-183671

MA

Foreword

The present course on calculus of several variables is meant as a text, either for one semester following the *First Course in Calculus*, or for a longer period if the calculus sequence is so structured.

In a one-semester course, I suggest covering most of the first part, omitting Chapter II, §3 and omitting some material from the chapter on Taylor's formula in several variables, to suit the taste of the instructor and the class. One can then jump directly to the chapter on double and triple integrals, which could in fact be treated immediately after Chapter I. If time allows, one can also cover the first section in the chapter on Green's theorem, which gives a neat application of the techniques of double integrals and curve integrals. Joining them in this fashion will make the student learn both techniques better for having used them in a significant context.

The first part has considerable unity of style. Essentially all the results are immediately corollaries of the chain rule. The main idea is that given a function of several variables, if we want to look at its values at two points P and Q, we join these points by a curve (often a straight line), and then look at the values of the function on that curve. By this device, we are able to reduce a large number of problems in several variables to problems and techniques in one variable. For instance, the directional derivative, the law of conservation of energy, and Taylor's formula, are handled in this manner.

I have included only that part of linear algebra which is immediately useful for the applications to calculus. My *Introduction to Linear Algebra* provides an appropriate text when a whole semester is devoted to the subject. Many courses are still structured to give primary emphasis to

the analytic aspects, and only a few notions involving matrices and linear maps are needed to cover, say, the chain rule for mappings of one space into another, and to emphasize the importance of linear approximations.

The last chapter on surface integrals and Stokes' theorem could essentially be covered after Green's theorem and multiple integrals. The chapter on the change of variables formula in multiple integration is the most expendable one, and can be omitted altogether without affecting the understanding of the rest of the book. Each instructor will adapt the material to the needs of any given class.

New Haven, Connecticut
November 1972

SERGE LANG

Contents

Part One
Mappings from Numbers to Vectors and Vectors to Numbers

Chapter I

Vectors

Chapter II

Differentiation of Vectors

Chapter III

Functions of Several Variables

Chapter IV

The Chain Rule and the Gradient

CHAPTER V

Potential Functions and Curve Integrals

CHAPTER VI

Higher Derivatives

CHAPTER VII

Maximum and Minimum

Part Two
Matrices, Linear Maps, and Determinants

CHAPTER VIII

Matrices

Chapter IX

Linear Mappings

Chapter X

Determinants

Part Three
Mappings from Vectors to Vectors

Chapter XI

Applications to Functions of Several Variables

Part Four
Multiple Integration

Chapter XII

Multiple Integrals

CHAPTER XIII

The Change of Variables Formula

CHAPTER XIV

Green's Theorem

CHAPTER XV

Surface Integrals

APPENDIX

Fourier Series

PART ONE

MAPPINGS FROM NUMBERS TO VECTORS AND VECTORS TO NUMBERS

In dealing with higher dimensional space, we can often reduce certain problems to 1-dimensional ones by using the following idea. We can join two points in space by a line segment. If we have a function defined in some region in space containing the points, and we want to analyze the behavior of the function at these points, then we can look at the induced function on the line segment. This yields a function of one variable.

Dealing with a segment between two points amounts to dealing with a mapping from numbers to higher dimensional space, parametrizing the segment. On the other hand, a function defined on a region in space takes on values in the real numbers. These two cases are important in themselves, and are also used later in the general situation where we consider mappings from one space into another.

CHAPTER I

Vectors

The concept of a vector is basic for the study of functions of several variables. It provides geometric motivation for everything that follows. Hence the properties of vectors, both algebraic and geometric, will be discussed in full.

One significant feature of all the statements and proofs of this part is that they are neither easier nor harder to prove in 3- or n-space than they are in 2-space.

§1. *Definition of points in n-space*

We know that a number can be used to represent a point on a line, once a unit length is selected.

A pair of numbers (i.e. a couple of numbers) (x, y) can be used to represent a point in the plane.

These can be pictured as follows:

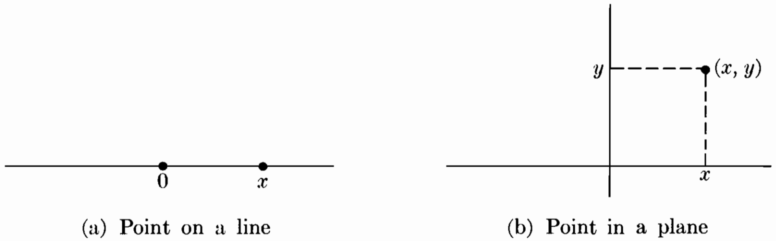


Figure 1

We now observe that a triple of numbers (x, y, z) can be used to represent a point in space, that is 3-dimensional space, or 3-space. We simply introduce one more axis.

The picture on the next page illustrates this.

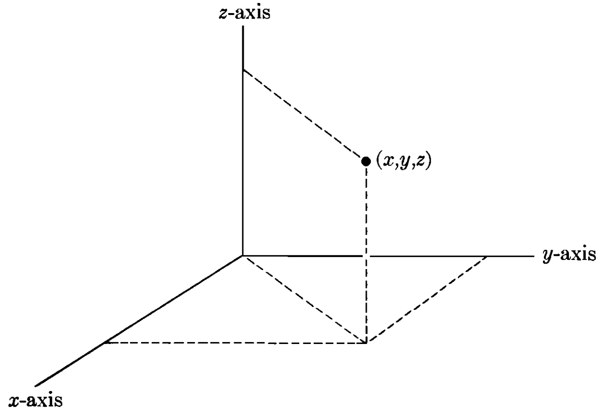


Figure 2

Instead of using x, y, z we could also use (x_1, x_2, x_3). The line could be called 1-space, and the plane could be called 2-space.

Thus we can say that a single number represents a point in 1-space. A couple represents a point in 2-space. A triple represents a point in 3-space.

Although we cannot draw a picture to go further, there is nothing to prevent us from considering a quadruple of numbers

$$(x_1, x_2, x_3, x_4)$$

and decreeing that this is a point in 4-space. A quintuple would be a point in 5-space, then would come a sextuple, septuple, octuple, ...

We let ourselves be carried away and **define a point in n-space** to be an n-tuple of numbers

$$(x_1, x_2, \ldots, x_n),$$

if n is a positive integer. We shall denote such an n-tuple by a capital letter X, and try to keep small letters for numbers and capital letters for points. We call the numbers $x_1, \ldots, x_n$ the **coordinates** of the point X. For example, in 3-space, 2 is the first coordinate of the point $(2, 3, -4)$, and -4 is its third coordinate.

Most of our examples will take place when $n = 2$ or $n = 3$. Thus the reader may visualize either of these two cases throughout the book. However, two comments must be made: First, practically no formula or theorem is simpler by making such assumptions on n. Second, the case $n = 4$ does occur in physics, and the case $n = n$ occurs often enough in practice or theory to warrant its treatment here. Furthermore, part of our purpose is in fact to show that the general case is always similar to the case when $n = 2$ or $n = 3$.

Examples. One classical example of 3-space is of course the space we live in. After we have selected an origin and a coordinate system, we can describe the position of a point (body, particle, etc.) by 3 coordinates. Furthermore, as was known long ago, it is convenient to extend this space to a 4-dimensional space, with the fourth coordinate as time, the time origin being selected, say, as the birth of Christ—although this is purely arbitrary (it might be more convenient to select the birth of the solar system, or the birth of the earth as the origin, if we could determine these accurately). Then a point with negative time coordinate is a BC point, and a point with positive time coordinate is an AD point.

Don't get the idea that "time is *the* fourth dimension", however. The above 4-dimensional space is only one possible example. In economics, for instance, one uses a very different space, taking for coordinates, say, the number of dollars expended in an industry. For instance, we could deal with a 7-dimensional space with coordinates corresponding to the following industries:

1. Steel	2. Auto	3. Farm products	4. Fish
5. Chemicals	6. Clothing	7. Transportation	

We agree that a megabuck per year is the unit of measurement. Then a point

$$(1{,}000, 800, 550, 300, 700, 200, 900)$$

in this 7-space would mean that the steel industry spent one billion dollars in the given year, and that the chemical industry spent 700 million dollars in that year.

We shall now define how to add points. If A, B are two points, say

$$A = (a_1, \dots, a_n), \qquad B = (b_1, \dots, b_n),$$

then we define $A + B$ to be the point whose coordinates are

$$(a_1 + b_1, \dots, a_n + b_n).$$

For example, in the plane, if $A = (1, 2)$ and $B = (-3, 5)$, then

$$A + B = (-2, 7).$$

In 3-space, if $A = (-1, \pi, 3)$ and $B = (\sqrt{2}, 7, -2)$, then

$$A + B = (\sqrt{2} - 1, \pi + 7, 1).$$

Furthermore, if c is any number, we **define** cA to be the point whose coordinates are

$$(ca_1, \dots, ca_n).$$

If $A = (2, -1, 5)$ and $c = 7$, then $cA = (14, -7, 35)$.

We observe that the following rules are satisfied:

(1) $(A + B) + C = A + (B + C)$.

(2) $A + B = B + A$.

(3) $c(A + B) = cA + cB$.

(4) If c_1, c_2 are numbers, then

$$(c_1 + c_2)A = c_1A + c_2A \qquad \text{and} \qquad (c_1c_2)A = c_1(c_2A).$$

(5) If we let $O = (0, \ldots, 0)$ be the point all of whose coordinates are 0, then $O + A = A + O = A$ for all A.

(6) $1 \cdot A = A$, and if we denote by $-A$ the n-tuple $(-1)A$, then

$$A + (-A) = O.$$

[Instead of writing $A + (-B)$, we shall frequently write $A - B$.]

All these properties are very simple to prove, and we suggest that you verify them on some examples. We shall give in detail the proof of property (3). Let $A = (a_1, \ldots, a_n)$ and $B = (b_1, \ldots, b_n)$. Then

$$A + B = (a_1 + b_1, \ldots, a_n + b_n)$$

and

$$\begin{aligned} c(A + B) &= \big(c(a_1 + b_1), \ldots, c(a_n + b_n)\big) \\ &= (ca_1 + cb_1, \ldots, ca_n + cb_n) \\ &= cA + cB, \end{aligned}$$

this last step being true by definition of addition of n-tuples.

The other proofs are left as exercises.

Note. Do not confuse the number 0 and the n-tuple $(0, \ldots, 0)$. We usually denote this n-tuple by O, and also call it zero, because no difficulty can occur in practice.

We shall now interpret addition and multiplication by numbers geometrically in the plane (you can visualize simultaneously what happens in 3-space).

Take an example. Let $A = (2, 3)$ and $B = (-1, 1)$. Then

$$A + B = (1, 4).$$

The figure looks like a parallelogram (Fig. 3).

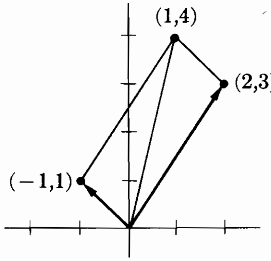


Figure 3

Take another example. Let $A = (3, 1)$ and $B = (1, 2)$. Then

$$A + B = (4, 3).$$

We see again that the geometric representation of our addition looks like a parallelogram (Fig. 4).

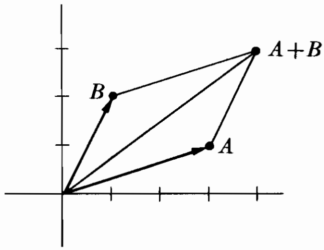
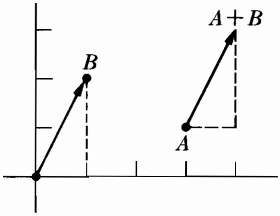


Figure 4

The reason why the figure looks like a parallelogram can be given in terms of plane geometry as follows. We obtain $B = (1, 2)$ by starting from the origin $O = (0, 0)$, and moving 1 unit to the right and 2 up. To get $A + B$, we start from A, and again move 1 unit to the right and 2 up. Thus the line segments between O and B, and between A and $A + B$ are the hypotenuses of right triangles whose corresponding legs are of the same length, and parallel. The above segments are therefore parallel and of the same length, as illustrated on the following figure.



What is the representation of multiplication by a number? Let $A = (1, 2)$ and $c = 3$. Then $cA = (3, 6)$ as in Fig. 5(a).

Multiplication by 3 amounts to stretching A by 3. Similarly, $\frac{1}{2}A$ amounts to stretching A by $\frac{1}{2}$, i.e. shrinking A to half its size. In general, if t is a number, $t > 0$, we interpret tA as a point in the same direction as A from the origin, but t times the distance.

Multiplication by a negative number reverses the direction. Thus $-3A$ would be represented as in Fig. 5(b).

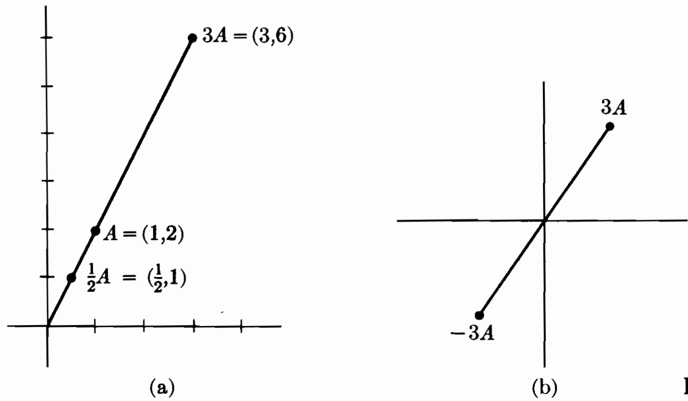
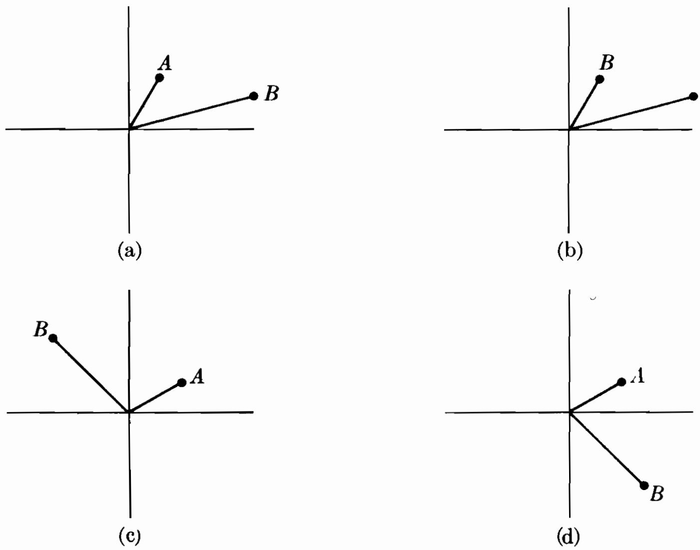


Figure 5

Exercises

Find $A + B$, $A - B$, $3A$, $-2B$ in each of the following cases. Draw the points of Exercises 1 and 2 on a sheet of graph paper.

1. $A = (2, -1)$, $B = (-1, 1)$
2. $A = (-1, 3)$, $B = (0, 4)$
3. $A = (2, -1, 5)$, $B = (-1, 1, 1)$
4. $A = (-1, -2, 3)$, $B = (-1, 3, -4)$
5. $A = (\pi, 3, -1)$, $B = (2\pi, -3, 7)$
6. $A = (15, -2, 4)$, $B = (\pi, 3, -1)$
7. Let $A = (1, 2)$ and $B = (3, 1)$. Draw $A + B$, $A + 2B$, $A + 3B$, $A - B$, $A - 2B$, $A - 3B$ on a sheet of graph paper.
8. Let A, B be as in Exercise 1. Draw the points $A + 2B$, $A + 3B$, $A - 2B$, $A - 3B$, $A + \frac{1}{2}B$ on a sheet of graph paper.
9. Let A and B be as drawn in the following figures. Draw the point $A - B$.



§2. *Located vectors*

We define a **located vector** to be an ordered pair of points which we write $\overrightarrow{AB}$. (This is *not* a product.) We visualize this as an arrow between A and B. We call A the **beginning point** and B the **end point** of the located vector (Fig. 6).

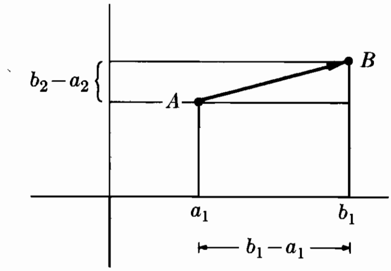


Figure 6

How are the coordinates of B obtained from those of A? We observe that in the plane,

$$b_1 = a_1 + (b_1 - a_1).$$

Similarly,

$$b_2 = a_2 + (b_2 - a_2).$$

This means that

$$B = A + (B - A).$$

Let $\overrightarrow{AB}$ and $\overrightarrow{CD}$ be two located vectors. We shall say that they are **equivalent** if $B - A = D - C$. Every located vector $\overrightarrow{AB}$ is equivalent to one whose beginning point is the origin, because $\overrightarrow{AB}$ is equivalent to $\overrightarrow{O(B - A)}$. Clearly this is the only located vector whose beginning point is the origin and which is equivalent to $\overrightarrow{AB}$. If you visualize the parallelogram law in the plane, then it is clear that equivalence of two located vectors can be interpreted geometrically by saying that the lengths of the line segments determined by the pair of points are equal, and that the "directions" in which they point are the same.

In the next figures, we have drawn the located vectors $\overrightarrow{O(B - A)}$, $\overrightarrow{AB}$, and $\overrightarrow{O(A - B)}$, $\overrightarrow{BA}$.

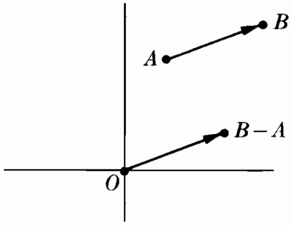


Figure 7

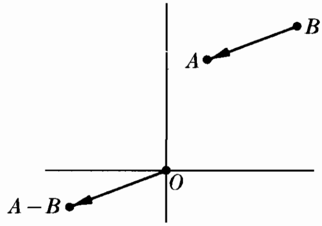


Figure 8

Given a located vector $\overrightarrow{OC}$ whose beginning point is the origin, we shall say that it is **located at the origin.** Given any located vector $\overrightarrow{AB}$, we shall say that it is **located at** A.

A located vector at the origin is entirely determined by its end point. In view of this, we shall call an n-tuple either a point or a vector, depending on the interpretation which we have in mind.

Two located vectors $\overrightarrow{AB}$ and $\overrightarrow{PQ}$ are said to be **parallel** if there is a number $c \neq 0$ such that $B - A = c(Q - P)$. They are said to have the **same direction** if there is a number $c > 0$ such that $B - A = c(Q - P)$, and to have **opposite direction** if there is a number $c < 0$ such that $B - A = c(Q - P)$. In the next pictures, we illustrate parallel located vectors.

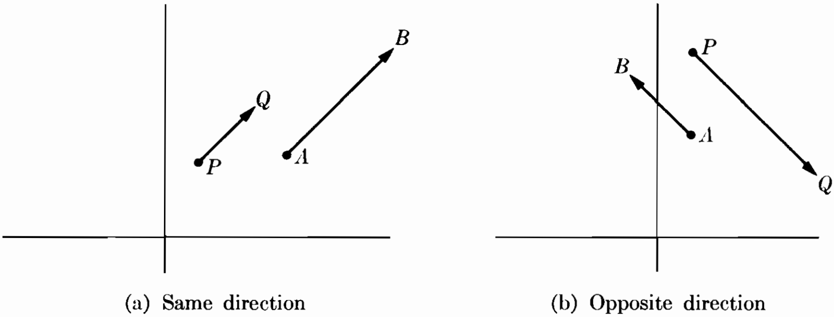


Figure 9

In a similar manner, any definition made concerning n-tuples can be carried over to located vectors. For instance, in the next section, we shall define what it means for n-tuples to be perpendicular. Then we can say that two located vectors $\overrightarrow{AB}$ and $\overrightarrow{PQ}$ are perpendicular if $B - A$ is perpendicular to $Q - P$. In the next figure, we have drawn a picture of such vectors in the plane.

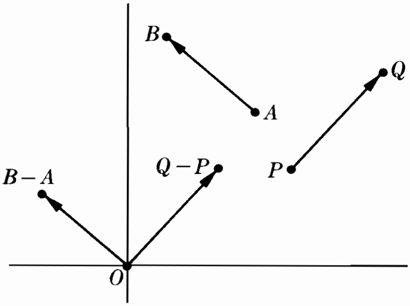


Figure 10

Example 1. Let $P = (1, -1, 3)$ and $Q = (2, 4, 1)$. Then $\overrightarrow{PQ}$ is equivalent to $\overrightarrow{OC}$, where $C = Q - P = (1, 5, -2)$. If $A = (4, -2, 5)$ and $B = (5, 3, 3)$, then $\overrightarrow{PQ}$ is equivalent to $\overrightarrow{AB}$ because

$$Q - P = B - A = (1, 5, -2).$$

Example 2. Let $P = (3, 7)$ and $Q = (-4, 2)$. Let $A = (5, 1)$ and $B = (-16, -14)$. Then

$$Q - P = (-7, -5) \quad \text{and} \quad B - A = (-21, -15).$$

Hence $\overrightarrow{PQ}$ is parallel to $\overrightarrow{AB}$, because $B - A = 3(Q - P)$. Since $3 > 0$, we even see that $\overrightarrow{PQ}$ and $\overrightarrow{AB}$ have the same direction.

Exercises

In each case, determine which located vectors $\overrightarrow{PQ}$ and $\overrightarrow{AB}$ are equivalent.

1. $P = (1, -1)$, $Q = (4, 3)$, $A = (-1, 5)$, $B = (5, 2)$.
2. $P = (1, 4)$, $Q = (-3, 5)$, $A = (5, 7)$, $B = (1, 8)$.
3. $P = (1, -1, 5)$, $Q = (-2, 3, -4)$, $A = (3, 1, 1)$, $B = (0, 5, 10)$.
4. $P = (2, 3, -4)$, $Q = (-1, 3, 5)$, $A = (-2, 3, -1)$, $B = (-5, 3, 8)$.

In each case, determine which located vectors $\overrightarrow{PQ}$ and $\overrightarrow{AB}$ are parallel.

5. $P = (1, -1)$, $Q = (4, 3)$, $A = (-1, 5)$, $B = (7, 1)$.
6. $P = (1, 4)$, $Q = (-3, 5)$, $A = (5, 7)$, $B = (9, 6)$.
7. $P = (1, -1, 5)$, $Q = (-2, 3, -4)$, $A = (3, 1, 1)$, $B = (-3, 9, -17)$.
8. $P = (2, 3, -4)$, $Q = (-1, 3, 5)$, $A = (-2, 3, -1)$, $B = (-11, 3, -28)$.
9. Draw the located vectors of Exercises 1, 2, 5, and 6 on a sheet of paper to illustrate these exercises. Also draw the located vectors $\overrightarrow{QP}$ and $\overrightarrow{BA}$. Draw the points $Q - P$, $B - A$, $P - Q$, and $A - B$.

§3. *Scalar product*

It is understood that throughout a discussion we select vectors always in the same n-dimensional space.

Let $A = (a_1, \ldots, a_n)$ and $B = (b_1, \ldots, b_n)$ be two vectors. We define their **scalar** or **dot product** $A \cdot B$ to be

$$a_1b_1 + \cdots + a_nb_n.$$

This product is a **number**. For instance, if

$$A = (1, 3, -2) \quad \text{and} \quad B = (-1, 4, -3)$$

then

$$A \cdot B = -1 + 12 + 6 = 17.$$

For the moment, we do not give a geometric interpretation to this scalar product. We shall do this later. We derive first some important properties. The basic ones are:

SP 1. *We have* $A \cdot B = B \cdot A$.

SP 2. *If* A, B, C *are three vectors then*

$$A \cdot (B + C) = A \cdot B + A \cdot C = (B + C) \cdot A.$$

SP 3. *If* x *is a number, then*

$$(xA) \cdot B = x(A \cdot B) \qquad \text{and} \qquad A \cdot (xB) = x(A \cdot B).$$

SP 4. *If* $A = O$ *is the zero vector, then* $A \cdot A = 0$, *and otherwise* $A \cdot A > 0$.

We shall now prove these properties.

Concerning the first, we have

$$a_1b_1 + \cdots + a_nb_n = b_1a_1 + \cdots + b_na_n,$$

because for any two numbers a, b, we have $ab = ba$. This proves the first property.

For **SP 2**, let $C = (c_1, \ldots, c_n)$. Then

$$B + C = (b_1 + c_1, \ldots, b_n + c_n)$$

and

$$\begin{aligned} A \cdot (B + C) &= a_1(b_1 + c_1) + \cdots + a_n(b_n + c_n) \\ &= a_1b_1 + a_1c_1 + \cdots + a_nb_n + a_nc_n. \end{aligned}$$

Reordering the terms yields

$$a_1b_1 + \cdots + a_nb_n + a_1c_1 + \cdots + a_nc_n,$$

which is none other than $A \cdot B + A \cdot C$. This proves what we wanted.

We leave property **SP 3** as an exercise.

Finally, for **SP 4**, we observe that if one coordinate a_i of A is not equal to 0, then there is a term $a_i^2 \neq 0$ and $a_i^2 > 0$ in the scalar product

$$A \cdot A = a_1^2 + \cdots + a_n^2.$$

Since every term is $\geqq 0$, it follows that the sum is > 0, as was to be shown.

In much of the work which we shall do concerning vectors, we shall use only the ordinary properties of addition, multiplication by numbers, and the four properties of the scalar product. We shall give a formal discussion

of these later. For the moment, observe that there are other objects with which you are familiar and which can be added, subtracted, and multiplied by numbers, for instance the continuous functions on an interval $[a, b]$ (cf. Exercise 6).

Instead of writing $A \cdot A$ for the scalar product of a vector with itself, it will be convenient to write also A^2. (This is the only instance when we allow ourselves such a notation. Thus A^3 has no meaning.) As an exercise, verify the following identities:

$$(A + B)^2 = A^2 + 2A \cdot B + B^2,$$
$$(A - B)^2 = A^2 - 2A \cdot B + B^2.$$

A dot product $A \cdot B$ may very well be equal to 0 without either A or B being the zero vector. For instance, let $A = (1, 2, 3)$ and $B = (2, 1, -\frac{4}{3})$. Then $A \cdot B = 0$.

We define two vectors A, B to be **perpendicular** (or as we shall also say, **orthogonal**) if $A \cdot B = 0$. For the moment, it is not clear that in the plane, this definition coincides with our intuitive geometric notion of perpendicularity. We shall convince you that it does in the next section. Here we merely note an example. Say in $\mathbf{R}^3$, let

$$E_1 = (1, 0, 0), \qquad E_2 = (0, 1, 0), \qquad E_3 = (0, 0, 1)$$

be the three unit vectors, as shown on the diagram (Fig. 11).

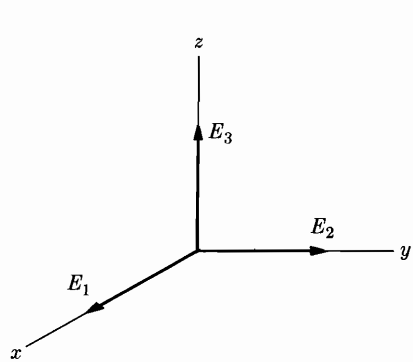


Figure 11

Then we see that $E_1 \cdot E_2 = 0$, and similarly $E_i \cdot E_j = 0$ if $i \neq j$. And these vectors look perpendicular. If $A = (a_1, a_2, a_3)$, then we observe that the i-th component of A, namely

$$a_i = A \cdot E_i$$

is the dot product of A with the i-th unit vector. We see that A is perpendicular to E_i (according to our definition of perpendicularity with the dot product) if and only if its i-th component is equal to 0.

Exercises

1. Find $A \cdot A$ for each of the following n-tuples.
   (a) $A = (2, -1), B = (-1, 1)$ (b) $A = (-1, 3), B = (0, 4)$
   (c) $A = (2, -1, 5), B = (-1, 1, 1)$ (d) $A = (-1, -2, 3), B = (-1, 3, -4)$
   (e) $A = (\pi, 3, -1), B = (2\pi, -3, 7)$ (f) $A = (15, -2, 4), B = (\pi, 3, -1)$
2. Find $A \cdot B$ for each of the above n-tuples.
3. Using only the four properties of the scalar product, verify in detail the identities given in the text for $(A + B)^2$ and $(A - B)^2$.
4. Which of the following pairs of vectors are perpendicular?
   (a) $(1, -1, 1)$ and $(2, 1, 5)$ (b) $(1, -1, 1)$ and $(2, 3, 1)$
   (c) $(-5, 2, 7)$ and $(3, -1, 2)$ (d) $(\pi, 2, 1)$ and $(2, -\pi, 0)$
5. Let A be a vector perpendicular to every vector X. Show that $A = O$.

Scalar product for functions.

6. Consider continuous functions on the interval $[-1, 1]$. Define the scalar product of two such functions f, g to be

$$\int_{-1}^{+1} f(x)g(x)\,dx.$$

   We denote this integral also by $\langle f, g\rangle$. Verify that the four rules for a scalar product are satisfied, in other words, show that:

   SP 1. $\langle f, g\rangle = \langle g, f\rangle$.

   SP 2. $\langle f, g + h\rangle = \langle f, g\rangle + \langle f, h\rangle$.

   SP 3. $\langle cf, g\rangle = c\langle f, g\rangle$.

   SP 4. *If* $f = 0$ *then* $\langle f, f\rangle = 0$ *and if* $f \neq 0$ *then* $\langle f, f\rangle > 0$.

7. If $f(x) = x$ and $g(x) = x^2$, what are $\langle f, f\rangle$, $\langle g, g\rangle$, and $\langle f, g\rangle$?
8. Consider continuous functions on the interval $[-\pi, \pi]$. Define a scalar product similar to the above for this interval. Show that the functions $\sin nx$ and $\cos mx$ are orthogonal for this scalar product (m, n being integers).

§4. *The norm of a vector*

We define the **norm**, or **length**, of a vector A, and denote by $\|A\|$, the number

$$\|A\| = \sqrt{A \cdot A}.$$

Since $A \cdot A \geqq 0$, we can take the square root.

In terms of coordinates, we see that

$$\|A\| = \sqrt{a_1^2 + \cdots + a_n^2},$$

and therefore that when $n = 2$ or $n = 3$, this coincides with our intuitive notion (derived from the Pythagoras theorem) of length. Indeed, when $n = 2$ and say $A = (a, b)$, then the norm of A is

$$\|A\| = \sqrt{a^2 + b^2},$$

as in the following picture.

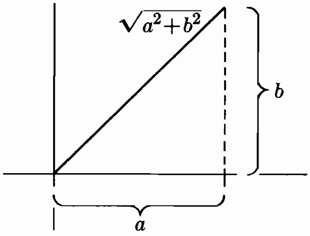


Figure 12

For example, if $A = (1, 2)$, then

$$\|A\| = \sqrt{1 + 4} = \sqrt{5}.$$

If $B = (-1, 2, 3)$, then

$$\|B\| = \sqrt{1 + 4 + 9} = \sqrt{14}.$$

If $n = 3$, then the picture looks like Fig. 13, with $A = (x, y, z)$.

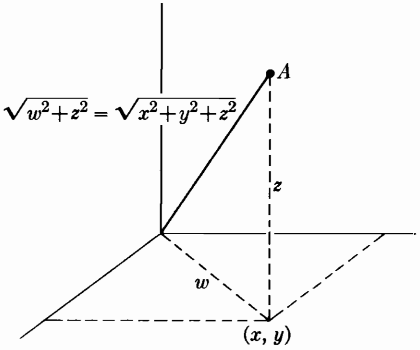


Figure 13

If we first look at the two components (x, y), then the length of the segment between $(0, 0)$ and (x, y) is equal to $w = \sqrt{x^2 + y^2}$, as indicated.

Then again the length of A by the Pythagoras theorem would be

$$\sqrt{w^2 + z^2} = \sqrt{x^2 + y^2 + z^2}.$$

Thus when $n = 3$, our definition of length is compatible with the geometry of the Pythagoras theorem.

If $A = (a_1, \dots, a_n)$ and $A \neq O$, then $\|A\| \neq 0$ because some coordinate $a_i \neq 0$, so that $a_i^2 > 0$, and hence $a_1^2 + \cdots + a_n^2 > 0$, so $\|A\| \neq 0$.

Observe that for any vector A we have

$$\|A\| = \|-A\|.$$

This is due to the fact that

$$(-a_1)^2 + \cdots + (-a_n)^2 = a_1^2 + \cdots + a_n^2,$$

because $(-1)^2 = 1$. Of course, this is as it should be from the picture:

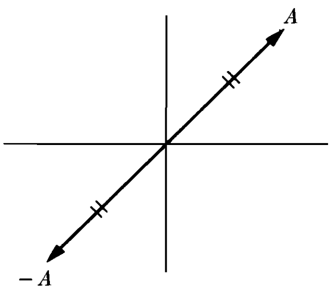


Figure 14

From the geometry of the situation, it is also reasonable to expect that if $c > 0$, then $\|cA\| = c\|A\|$, i.e. if we stretch a vector A by multiplying by a positive number c, then the length stretches also by that amount. We verify this formally using our definition of the length.

Theorem 1. *Let x be a number. Then*

$$\|xA\| = |x|\,\|A\|$$

(*absolute value of x times the length of A*).

Proof. By definition, we have

$$\|xA\|^2 = (xA) \cdot (xA),$$

which is equal to

$$x^2(A \cdot A)$$

by the properties of the scalar product. Taking the square root now yields what we want.

We shall say that a vector E is a **unit** vector if $\|E\| = 1$. Given any vector A, let $a = \|A\|$. If $a \neq 0$ then

$$\frac{1}{a}A$$

is a unit vector, because

$$\left\|\frac{1}{a}A\right\| = \frac{1}{a}a = 1.$$

We shall say that two vectors A, B (neither of which is O) have the **same direction** if there is a number $c > 0$ such that $cA = B$. In view of this definition, we see that the vector

$$\frac{1}{\|A\|}A$$

is a unit vector in the direction of A (provided $A \neq O$).

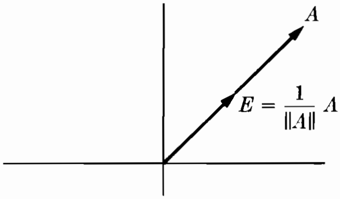


Figure 15

If E is the unit vector in the direction of A, and $\|A\| = a$, then

$$A = aE.$$

Example 1. Let $A = (1, 2, -3)$. Then $\|A\| = \sqrt{14}$. Hence the unit vector in the direction of A is the vector

$$E = \left(\frac{1}{\sqrt{14}}, \frac{2}{\sqrt{14}}, \frac{-3}{\sqrt{14}}\right).$$

We mention in passing that two vectors A, B (neither of which is O) have **opposite directions** if there is a number $c < 0$ such that $cA = B$.

Let A, B be two n-tuples. We define the **distance** between A and B to be

$$\|A - B\| = \sqrt{(A - B)\cdot(A - B)}.$$

This definition coincides with our geometric intuition when A, B are points in the plane (Fig. 16). It is the same thing as the length of the located vector $\overrightarrow{AB}$ or the located vector $\overrightarrow{BA}$.

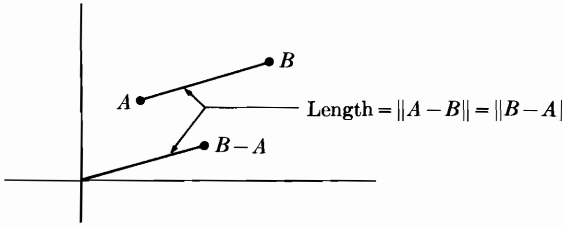


Figure 16

Example 2. Let $A = (-1, 2)$ and $B = (3, 4)$. Then the length of the located vector $\overrightarrow{AB}$ is $\|B - A\|$. But $B - A = (4, 2)$. Thus

$$\|B - A\| = \sqrt{16 + 4} = \sqrt{20}.$$

In the picture, we see that the horizontal side has length 4 and the vertical side has length 2. Thus our definitions reflect our geometric intuition derived from Pythagoras.

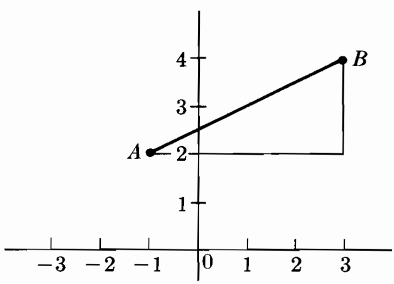


Figure 17

We are also in the position to justify our definition of perpendicularity. Given A, B in the plane, the condition that

$$\|A + B\| = \|A - B\|$$

(illustrated in Fig. 18(b)) coincides with the geometric property that A should be perpendicular to B.

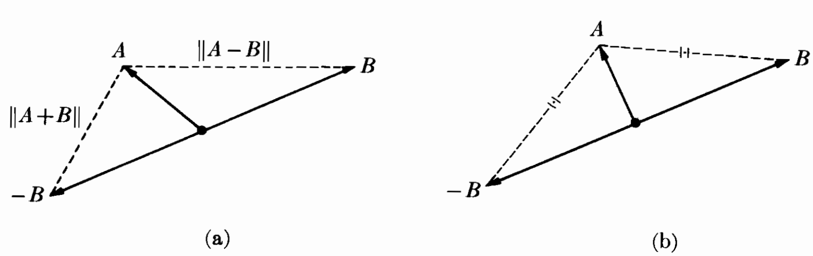


Figure 18

Taking the square of each side, we see that this condition is equivalent with

$$(A + B)\cdot(A + B) = (A - B)\cdot(A - B)$$

and expanding out, this equality is equivalent with

$$A \cdot A + 2A \cdot B + B \cdot B = A \cdot A - 2A \cdot B + B \cdot B.$$

Making cancellations, we obtain the equivalent condition

$$4A \cdot B = 0$$

or

$$A \cdot B = 0.$$

This achieves what we wanted to show, namely that

$$\|A - B\| = \|A + B\| \quad \text{if and only if} \quad A \cdot B = 0.$$

Observe that we have the general **Pythagoras theorem**; *If* A, B *are perpendicular, then*

$$\|A + B\|^2 = \|A\|^2 + \|B\|^2.$$

The theorem is illustrated on Fig. 19.

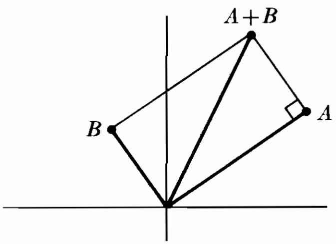


Figure 19

To prove this, we use the definitions, namely

$$\begin{aligned}\|A + B\|^2 &= (A + B)\cdot(A + B) = A^2 + 2A \cdot B + B^2\\ &= \|A\|^2 + \|B\|^2,\end{aligned}$$

because $A \cdot B = 0$, and $A \cdot A = \|A\|^2$, $B \cdot B = \|B\|^2$ by definition.

Remark. If A is perpendicular to B, and x is any number, then A is also perpendicular to xB because

$$A \cdot xB = xA \cdot B = 0.$$

We shall now use the notion of perpendicularity to derive the notion of projection. Let A, B be two vectors and $B \neq O$. We wish to define the projection of A along B, which will be a vector P as shown in the picture.

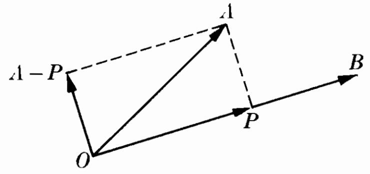


Figure 20

We seek a vector P such that $A - P$ is perpendicular to B, and such that P can be written in the form $P = cB$ for some number c. Suppose that we can find such a number c, namely one satisfying

$$(A - cB) \cdot B = 0.$$

We then obtain

$$A \cdot B = cB \cdot B,$$

and therefore

$$c = \frac{A \cdot B}{B \cdot B}.$$

We see that such a number c is uniquely determined by our condition of perpendicularity. Conversely, if we let c have the above value, then we have

$$(A - cB) \cdot B = A \cdot B - cB \cdot B = 0.$$

Thus this value of c satisfies our requirement.

We now define the vector cB to be the **projection** of A along B, if c is the number

$$c = \frac{A \cdot B}{B \cdot B},$$

and we define c to be the **component** of A along B. *If B is a unit vector, then we have simply*

$$c = A \cdot B.$$

Example. Let $A = (1, 2, -3)$ and $B = (1, 1, 2)$. Then the component of A along B is the **number**

$$c = \frac{A \cdot B}{B \cdot B} = \frac{-3}{6} = -\frac{1}{2}.$$

Hence the projection of A along B is the **vector**

$$cB = (-\tfrac{1}{2}, -\tfrac{1}{2}, -1).$$

Our construction has an immediate interpretation in the plane, which gives us a geometric interpretation for the scalar product. Namely, assume $A \neq O$ and look at the angle θ between A and B (Fig. 21). Then from plane geometry we see that

$$\cos\theta = \frac{c\|B\|}{\|A\|},$$

or substituting the value for c obtained above,

$$\boxed{A \cdot B = \|A\|\,\|B\|\cos\theta.}$$

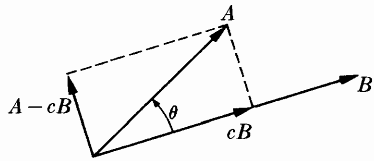


Figure 21

In some treatments of vectors, one takes the relation

$$A \cdot B = \|A\|\,\|B\|\cos\theta$$

as definition of the scalar product. This is subject to the following disadvantages, not to say objections:

(a) The four properties of the scalar product **SP 1** through **SP 4** are then by no means obvious.

(b) Even in 3-space, one has to rely on geometric intuition to obtain the cosine of the angle between A and B, and this intuition is less clear than in the plane. In higher dimensional space, it fails even more.

(c) It is extremely hard to work with such a definition to obtain further properties of the scalar product.

Thus we prefer to lay obvious algebraic foundations, and then recover very simply all the properties. Aside from that, in analysis, one uses scalar products in the context of functions, where $\cos\theta$ becomes completely meaningless, for instance in Exercise 5 of §3, which is the starting point of the theory of Fourier series.

We shall prove further properties of the norm and scalar product using our results on perpendicularity. First note a special case. If

$$E_i = (0, \ldots, 0, 1, 0, \ldots, 0)$$

is the i-th unit vector $\mathbf{R}^n$, and

$$A = (a_1, \ldots, a_n),$$

then

$$A \cdot E_i = a_i$$

is the i-th component of A, i.e. the component of A along E_i. We have

$$|a_i| = \sqrt{a_i^2} \leqq \sqrt{a_1^2 + \cdots + a_n^2} = \|A\|,$$

so that the absolute value of each component of A is at most equal to the length of A.

We don't have to deal only with the special unit vector as above. Let E be any unit vector, that is a vector of length 1. Let c be the component of A along E. We saw that

$$c = A \cdot E.$$

Then $A - cE$ is perpendicular to E, and

$$A = A - cE + cE.$$

Then $A - cE$ is also perpendicular to cE, and by the Pythagoras theorem, we find

$$\|A\|^2 = \|A - cE\|^2 + \|cE\|^2 = \|A - cE\|^2 + c^2.$$

Thus we have the inequality $c^2 \leqq \|A\|^2$, and

$$|c| \leqq \|A\|.$$

In the next theorem, we generalize this inequality to a dot product $A \cdot B$ when B is not necessarily a unit vector.

Theorem 2. *Let A, B be two vectors in $\mathbf{R}^n$. Then*

$$|A \cdot B| \leqq \|A\| \, \|B\|.$$

Proof. If $B = O$, then both sides of the inequality are equal to 0, and so our assertion is obvious. Suppose that $B \neq O$. Let E be the unit vector in the direction of B, so that

$$E = \frac{B}{\|B\|} \cdot$$

We use the result just derived, namely $|A \cdot E| \leqq \|A\|$, and find

$$\frac{|A \cdot B|}{\|B\|} \leqq \|A\|.$$

Multiplying by $\|B\|$ yields the proof of our theorem.

In view of Theorem 2, we see that for vectors A, B in n-space, the number

$$\frac{A \cdot B}{\|A\| \, \|B\|}$$

has absolute value $\leqq 1$. Consequently,

$$-1 \leqq \frac{A \cdot B}{\|A\| \, \|B\|} \leqq 1,$$

and there exists a unique angle θ such that $0 \leqq \theta \leqq \pi$, and such that

$$\cos \theta = \frac{A \cdot B}{\|A\| \, \|B\|}.$$

We define this angle to be the **angle between A and B.**

Example. Let $A = (1, 2, -3)$ and $B = (2, 1, 5)$. Find the cosine of the angle θ between A and B.

By definition, we must have

$$\cos \theta = \frac{A \cdot B}{\|A\| \, \|B\|} = \frac{2 + 2 - 15}{\sqrt{14}\sqrt{30}} = \frac{-11}{\sqrt{420}}.$$

The inequality of Theorem 2 is known as the **Schwarz inequality.**

Theorem 3. *Let A, B be vectors. Then*

$$\|A + B\| \leqq \|A\| + \|B\|.$$

Proof. Both sides of this inequality are positive or 0. Hence it will suffice to prove that their squares satisfy the desired inequality, in other words,

$$(A + B) \cdot (A + B) \leqq (\|A\| + \|B\|)^2.$$

To do this, we consider

$$(A + B) \cdot (A + B) = A \cdot A + 2A \cdot B + B \cdot B.$$

In view of our previous result, this satisfies the inequality

$$\leqq \|A\|^2 + 2\|A\| \, \|B\| + \|B\|^2,$$

and the right-hand side is none other than

$$(\|A\| + \|B\|)^2.$$

Our theorem is proved.

Theorem 3 is known as the **triangle inequality.** The reason for this is that if we draw a triangle as in Fig. 22, then Theorem 3 expresses the fact

that the length of one side is $\leqq$ the sum of the lengths of the other two sides.

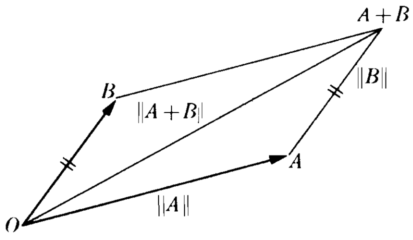


Figure 22

Exercises

1. Find the length of the vector A in the following cases.
   (a) $A = (2, -1), B = (-1, 1)$ (b) $A = (-1, 3), B = (0, 4)$
   (c) $A = (2, -1, 5), B = (-1, 1, 1)$ (d) $A = (-1, -2, 3), B = (-1, 3, -4)$
   (e) $A = (\pi, 3, -1), B = (2\pi, -3, 7)$ (f) $A = (15, -2, 4), B = (\pi, 3, -1)$
2. Find the length of vector B in the above cases.
3. Find the projection of A along B in the above cases.
4. Find the projection of B along A in the above cases.
5. Determine the cosine of the angles of the triangle whose vertices are
   (a) $(2, -1, 1)$, $(1, -3, -5)$, $(3, -4, -4)$.
   (b) $(3, 1, 1)$, $(-1, 2, 1)$, $(2, -2, 5)$.
6. Let $A_1, \dots, A_r$ be non-zero vectors which are mutually perpendicular, in other words $A_i \cdot A_j = 0$ if $i \neq j$. Let $c_1, \dots, c_r$ be numbers such that

$$c_1A_1 + \cdots + c_rA_r = 0.$$

   Show that all $c_i = 0$.
7. If A, B are two vectors in n-space, denote by $d(A, B)$ the distance between A and B, that is $d(A, B) = \|B - A\|$. Show that

$$d(A, B) = d(B, A),$$

   and that for any three vectors A, B, C we have

$$d(A, B) \leqq d(A, C) + d(B, C).$$

8. For any vectors A, B in n-space, prove the following relations:
   (a) $\|A + B\|^2 + \|A - B\|^2 = 2\|A\|^2 + 2\|B\|^2$.
   (b) $\|A + B\|^2 = \|A\|^2 + \|B\|^2 + 2A \cdot B$.
   (c) $\|A + B\|^2 - \|A - B\|^2 = 4A \cdot B$.
   Interpret (a) as a "parallelogram law".

9. Show that if θ is the angle between A and B, then

$$\|A - B\|^2 = \|A\|^2 + \|B\|^2 - 2\|A\|\,\|B\|\cos\theta.$$

10. Let A, B, C be three non-zero vectors. If $A \cdot B = A \cdot C$, show by an example that we do not necessarily have $B = C$.

11. Let A, B be non-zero vectors, mutually perpendicular. Show that for any number c we have $\|A + cB\| \geqq \|A\|$.

12. Let A, B be non-zero vectors. Assume that $\|A + cB\| \geqq \|A\|$ for all numbers c. Show that A, B are perpendicular.

13. Let $f(x) = x$ and $g(x) = x^2$. Using the scalar product

$$\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)\,dx,$$

find the projection of f along g and the projection of g along f, using the same definition of projection that has been given in the text, and did not refer to coordinates.

14. For this same scalar product, the norm of a function f is $\sqrt{\langle f, f \rangle}$. Find the norm of the constant function 1.

15. Consider now functions on the interval $[-\pi, \pi]$. Define the scalar product by

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)g(x)\,dx.$$

Find the norm of the functions $\sin 3x$ and $\cos x$.

16. Find the norm of the constant function 1 for the scalar product of Exercise 15.

17. In general, find the norm of the functions $\sin nx$ and $\cos mx$, where m, n are positive integers.

§5. *Lines and planes*

We define the parametric equation of a straight line passing through a point P in the direction of a vector $A \neq O$ to be

$$X = P + tA,$$

where t runs through all numbers (Fig. 23).

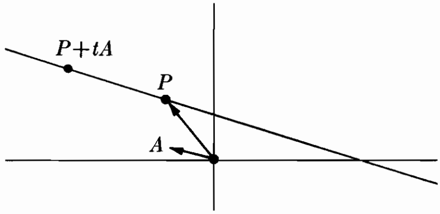


Figure 23

Suppose that we work in the plane, and write the coordinates of a point X as (x, y). Let $P = (p, q)$ and $A = (a, b)$. Then in terms of the coordinates, we can write

$$x = p + ta, \qquad y = q + tb.$$

We can then eliminate t and obtain the usual equation relating x and y.

For example, let $P = (2, 1)$ and $A = (-1, 5)$. Then the parametric equation of the line through P in the direction of A gives us

$$(*) \qquad x = 2 - t, \qquad y = 1 + 5t.$$

Multiplying the first equation by 5 and adding yields

$$(**) \qquad 5x + y = 11,$$

which is familiar.

This elimination of t shows that every pair (x, y) which satisfies the parametric equation (*) for some value of t also satisfies equation (**). Conversely, suppose we have a pair of numbers (x, y) satisfying (**). Let $t = 2 - x$. Then

$$y = 11 - 5x = 11 - 5(2 - t) = 1 + 5t.$$

Hence there exists some value of t which satisfies equation (*). Thus we have proved that the pairs (x, y) which are solutions of (**) are exactly the same pairs of numbers as those obtained by giving arbitrary values for t in (*). Thus the straight line can be described parametrically as in (*) or in terms of its usual equation (**). Starting with the ordinary equation

$$5x + y = 11,$$

we let $t = 2 - x$ in order to recover the specific parametrization of (*).

When we parametrize a straight line in the form

$$X = P + tA,$$

we have of course infinitely many choices for P on the line, and also infinitely many choices for A, differing by a scalar multiple. We can always select at least one. Namely, given an equation

$$ax + by = c$$

with numbers a, b, c, suppose that $a \neq 0$. We use y as parameter, and let

$$y = t.$$

Then we can solve for x, namely

$$x = \frac{c}{a} - \frac{b}{a}t.$$

Let $P = (c/a, 0)$ and $A = (-b/a, 1)$. We see that an arbitrary point (x, y) satisfying the equation

$$ax + by = c$$

can be expressed parametrically, namely

$$(x, y) = P + tA.$$

In higher dimension, starting with a parametric equation

$$X = P + tA,$$

we cannot eliminate t, and thus the parametric equation is the only one available to describe a straight line.

However, we can describe planes by an equation analogous to the single equation of the line. We proceed as follows.

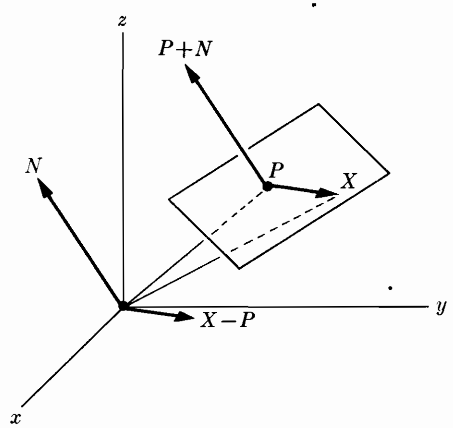


Figure 24

Let P be a point in 3-space and consider a located vector $\overrightarrow{ON}$. We define the **plane** passing through P perpendicular to $\overrightarrow{ON}$ to be the collection of all points X such that the located vector $\overrightarrow{PX}$ is perpendicular to $\overrightarrow{ON}$. According to our definitions, this amounts to the condition

$$(X - P) \cdot N = 0,$$

which can also be written as

$$X \cdot N = P \cdot N.$$

We shall also say that this plane is the one perpendicular to N, and consists of all vectors X such that $X - P$ is perpendicular to N. We have drawn a typical situation in 3-space in Fig. 24.

Instead of saying that N is perpendicular to the plane, one also says that N is **normal** to the plane.

Let t be a number $\neq 0$. Then the set of points X such that

$$(X - P) \cdot N = 0$$

coincides with the set of points X such that

$$(X - P) \cdot tN = 0.$$

Thus we may say that our plane is the plane passing through P and perpendicular to the **line** in the direction of N. To find the equation of the plane, we could use any vector tN (with $t \neq 0$) instead of N.

In 3-space, we get an ordinary plane. For example, let $P = (2, 1, -1)$ and $N = (-1, 1, 3)$. Then the equation of the plane passing through P and perpendicular to N is

$$-x + y + 3z = -2 + 1 - 3$$

or

$$-x + y + 3z = -4.$$

Observe that in 2-space, with $X = (x, y)$, the formulas lead to the equation of the line in the ordinary sense. For example, the equation of the line passing through $(4, -3)$ and perpendicular to $(-5, 2)$ is

$$-5x + 2y = -20 - 6 = -26.$$

We are now in position to interpret the coefficients $(-5, 2)$ of x and y in this equation. They give rise to a vector perpendicular to the line. In any equation

$$ax + by = c$$

the vector (a, b) is perpendicular to the line determined by the equation. Similarly, in 3-space, the vector (a, b, c) is perpendicular to the plane determined by the equation

$$ax + by + cz = d.$$

For example, the plane determined by the equation

$$2x - y + 3z = 5$$

is perpendicular to the vector $(2, -1, 3)$. If we want to find a point in that plane, we of course have many choices. We can give arbitrary values to x and y, and then solve for z. To get a concrete point, let $x = 1, y = 1$. Then we solve for z, namely

$$3z = 5 - 2 + 1 = 4,$$

so that $z = \frac{4}{3}$. Thus

$$(1, 1, \tfrac{4}{3})$$

is a point in the plane.

In n-space, the equation $X \cdot N = P \cdot N$ is said to be the equation of a **hyperplane.** For example,

$$3x - y + z + 2w = 5$$

is the equation of a hyperplane in 4-space, perpendicular to $(3, -1, 1, 2)$.

Two vectors A, B are said to be parallel if there exists a number $c \neq 0$ such that $cA = B$. Two lines are said to be **parallel** if, given two distinct points P_1, Q_1 on the first line and P_2, Q_2 on the second, the vectors

$$P_1 - Q_1$$

and

$$P_2 - Q_2$$

are parallel.

Two planes are said to be **parallel** (in 3-space) if their normal vectors are parallel. They are said to be **perpendicular** if their normal vectors are perpendicular. The **angle** between two planes is defined to be the angle between their normal vectors.

Example 1. Find the cosine of the angle between the planes

$$\begin{aligned} 2x - y + z &= 0, \\ x + 2y - z &= 1. \end{aligned}$$

This cosine is the cosine of the angle between the vectors

$$A = (2, -1, 1) \qquad \text{and} \qquad B = (1, 2, -1).$$

It is therefore equal to

$$\frac{A \cdot B}{\|A\|\,\|B\|} = -\frac{1}{6}.$$

Example 2. Let

$$Q = (1, 1, 1) \qquad \text{and} \qquad P = (1, -1, 2).$$

Let

$$N = (1, 2, 3).$$

Find the point of intersection of the line through P in the direction of N, and the plane through Q perpendicular to N.

The parametric equation of the line through P in the direction of N is

(1) $$X = P + tN.$$

The equation of the plane through Q perpendicular to N is

(2) $$(X - Q) \cdot N = 0.$$

We visualize the line and plane as follows:

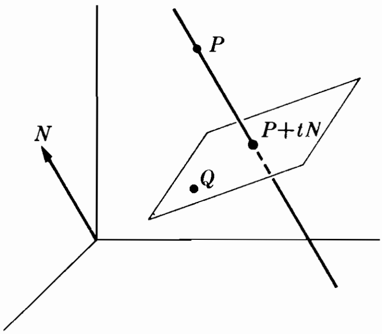


Figure 25

We must find the value of t such that the vector X in (1) also satisfies (2), that is

$$(P + tN - Q) \cdot N = 0,$$

or after using the rules of the dot product,

$$(P - Q) \cdot N + tN \cdot N = 0.$$

Solving for t yields

$$t = \frac{(Q - P) \cdot N}{N \cdot N} = \frac{1}{14}.$$

Thus the desired point of intersection is

$$P + tN = (1, -1, 2) + \tfrac{1}{14}(1, 2, 3) = (\tfrac{15}{14}, -\tfrac{12}{14}, \tfrac{31}{14}).$$

Example 3. Find the equation of the plane passing through the three points

$$P_1 = (1, 2, -1), \qquad P_2 = (-1, 1, 4), \qquad P_3 = (1, 3, -2).$$

We visualize schematically the three points as follows:

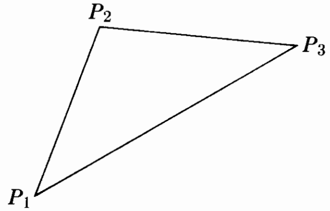


Figure 26

Then we find a vector N perpendicular to $\overrightarrow{P_1P_2}$ and $\overrightarrow{P_1P_3}$, or in other

words, perpendicular to $P_2 - P_1$ and $P_3 - P_1$. We have

$$P_2 - P_1 = (-2, -1, +5),$$
$$P_3 - P_1 = (0, 1, -1).$$

Let $N = (a, b, c)$. We must solve:

$$-2a - b + 5c = 0,$$
$$b - c = 0.$$

We take $b = c = 1$ and solve for a, getting $a = 2$. Then

$$N = (2, 1, 1)$$

satisfies our requirements. The plane perpendicular to N, passing through P_1 is the desired plane. Its equation is therefore

$$2x + y + z = 2 + 2 - 1 = 3.$$

Exercises

Find a parametric equation for the line passing through the following points.

1. $(1, 1, -1)$ and $(-2, 1, 3)$
2. $(-1, 5, 2)$ and $(3, -4, 1)$

Find the equation of the line in 2-space, perpendicular to A and passing through P, for the following values of A and P.

3. $A = (1, -1)$, $P = (-5, 3)$
4. $A = (-5, 4)$, $P = (3, 2)$
5. Show that the lines

$$3x - 5y = 1, \qquad 2x + 3y = 5$$

are not perpendicular.

6. Which of the following pairs of lines are perpendicular?

(a) $3x - 5y = 1$ and $2x + y = 2$
(b) $2x + 7y = 1$ and $x - y = 5$
(c) $3x - 5y = 1$ and $5x + 3y = 7$
(d) $-x + y = 2$ and $x + y = 9$

7. Find the equation of the plane perpendicular to the given vector N and passing through the given point P.

(a) $N = (1, -1, 3)$, $P = (4, 2, -1)$
(b) $N = (-3, -2, 4)$, $P = (2, \pi, -5)$
(c) $N = (-1, 0, 5)$, $P = (2, 3, 7)$

8. Find the equation of the plane passing through the following three points.

(a) $(2, 1, 1)$, $(3, -1, 1)$, $(4, 1, -1)$
(b) $(-2, 3, -1)$, $(2, 2, 3)$, $(-4, -1, 1)$
(c) $(-5, -1, 2)$, $(1, 2, -1)$, $(3, -1, 2)$

9. Find a vector perpendicular to $(1, 2, -3)$ and $(2, -1, 3)$, and another vector perpendicular to $(-1, 3, 2)$ and $(2, 1, 1)$.

10. Let P be the point $(1, 2, 3, 4)$ and Q the point $(4, 3, 2, 1)$. Let A be the vector $(1, 1, 1, 1)$. Let L be the line passing through P and parallel to A.
    (a) Given a point X on the line L, compute the distance between Q and X (as a function of the parameter t).
    (b) Show that there is precisely one point X_0 on the line such that this distance achieves a minimum, and that this minimum is $2\sqrt{5}$.
    (c) Show that $X_0 - Q$ is perpendicular to the line.

11. Let P be the point $(1, -1, 3, 1)$ and Q the point $(1, 1, -1, 2)$. Let A be the vector $(1, -3, 2, 1)$. Solve the same questions as in the preceding problem, except that in this case the minimum distance is $\sqrt{146/15}$.

12. Find a vector parallel to the line of intersection of the two planes

$$2x - y + z = 1, \qquad 3x + y + z = 2.$$

13. Same question for the planes,

$$2x + y + 5z = 2, \qquad 3x - 2y + z = 3.$$

14. Find a parametric equation for the line of intersection of the planes of Exercises 12 and 13.

15. Find the cosine of the angle between the following planes:

(a) $x + y + z = 1$, $x - y - z = 5$

(b) $2x + 3y - z = 2$, $x - y + z = 1$

(c) $x + 2y - z = 1$, $-x + 3y + z = 2$

(d) $2x + y + z = 3$, $-x - y + z = \pi$

16. (a) Let $P = (1, 3, 5)$ and $A = (-2, 1, 1)$. Find the intersection of the line through P in the direction of A, and the plane $2x + 3y - z = 1$.
    (b) Let $P = (1, 2, -1)$. Find the point of intersection of the plane

$$3x - 4y + z = 2,$$

with the line through P, perpendicular to that plane.

17. Let $Q = (1, -1, 2)$, $P = (1, 3, -2)$, and $N = (1, 2, 2)$. Find the point of the intersection of the line through P in the direction of N, and the plane through Q perpendicular to N.

18. Let P, Q be two points and N a vector in 3-space. Let P' be the point of intersection of the line through P, in the direction of N, and the plane through Q, perpendicular to N. We define the **distance** from P to that plane to be the distance between P and P'. Find the distance when

$$P = (1, 3, 5), \qquad Q = (-1, 1, 7), \qquad N = (-1, 1, -1).$$

19. In the notation of Exercise 18, show that the general formula for the distance is given by

$$\frac{|(Q - P) \cdot N|}{\|N\|}.$$

20. Find the distance between the indicated point and plane.
    (a) $(1, 1, 2)$ and $3x + y - 5z = 2$
    (b) $(-1, 3, 2)$ and $2x - 4 + z = 1$

21. Let $P = (1, 3, -1)$ and $Q = (-4, 5, 2)$. Determine the coordinates of the following points:
    (a) The midpoint of the line segment between P and Q.
    (b) The two points on this line segment lying one-third and two-thirds of the way from P to Q.
    (c) The point lying one-fifth of the way from P to Q.
    (d) The point lying two-fifths of the way from P to Q.

22. If P, Q are two arbitrary points in n-space, give the general formula for the midpoint of the line segment between P and Q.

§6. *The cross product*

You may omit this section and all references to it until you reach Chapter XV, where it will be used in an essential way.

This section applies only in 3-space!

Let $A = (a_1, a_2, a_3)$ and $B = (b_1, b_2, b_3)$ be two vectors in 3-space. We define their *cross product*

$$A \times B = (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1).$$

For instance, if $A = (2, 3, -1)$ and $B = (-1, 1, 5)$,

then

$$A \times B = (16, -9, 5).$$

We leave the following assertions as exercises:

CP 1. $A \times B = -(B \times A)$.

CP 2. $A \times (B + C) = (A \times B) + (A \times C)$,

and

$$(B + C) \times A = B \times A + C \times A.$$

CP 3. *For any number a, we have*

$$(aA) \times B = a(A \times B) = A \times (aB).$$

CP 4. $(A \times B) \times C = (A \cdot C)B - (B \cdot C)A$.

CP 5. *$A \times B$ is perpendicular to both A and B.*

As an example, we carry out this computation. We have

$$\begin{aligned} A \cdot (A \times B) &= a_1(a_2b_3 - a_3b_2) + a_2(a_3b_1 - a_1b_3) + a_3(a_1b_2 - a_2b_1) \\ &= 0 \end{aligned}$$

because all terms cancel. Similarly for $B \cdot (A \times B)$. This perpendicularity may be drawn as follows.

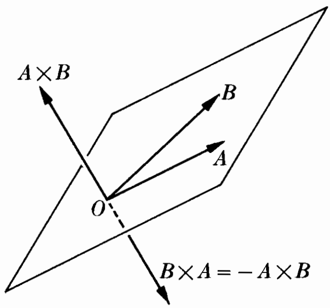


Figure 27

The vector $A \times B$ is perpendicular to the plane spanned by A and B. So is $B \times A$, but $B \times A$ points in the opposite direction.

Finally, as a last property, we have

CP 6. $(A \times B)^2 = (A \cdot A)(B \cdot B) - (A \cdot B)^2$.

Again, this can be verified by a computation on the coordinates. Namely, we have

$$(A \times B) \cdot (A \times B) = (a_2b_3 - a_3b_2)^2 + (a_3b_1 - a_1b_3)^2 + (a_1b_2 - a_2b_1)^2,$$

$$(A \cdot A)(B \cdot B) - (A \cdot B)^2 = (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1b_1 + a_2b_2 + a_3b_3)^2.$$

Expanding everything out, we find that **CP 6** drops out.

From our interpretation of the dot product, and the definition of the norm, we can rewrite **CP 6** in the form

$$\|A \times B\|^2 = \|A\|^2\|B\|^2 - \|A\|^2\|B\|^2 \cos^2 \theta,$$

where θ is the angle between A and B. Hence we obtain

$$\|A \times B\|^2 = \|A\|^2\|B\|^2 \sin^2 \theta$$

or

$$\boxed{\|A \times B\| = \|A\| \, \|B\| \, |\sin \theta|.}$$

This is analogous to the formula which gave us the absolute value of $A \cdot B$.

This formula can be used to make another interpretation of the cross product. Indeed, we see that $\|A \times B\|$ is the area of the parallelogram spanned by A and B, as shown on Fig. 28.

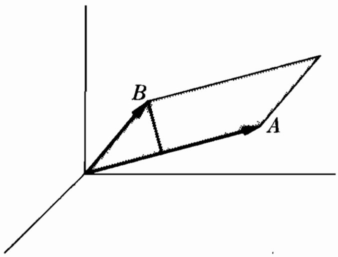


Figure 28

If we consider the plane containing the located vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$, then the picture looks like that in Fig. 29, and our assertion amounts simply to the statement that the area of a parallelogram is equal to the base times the altitude.

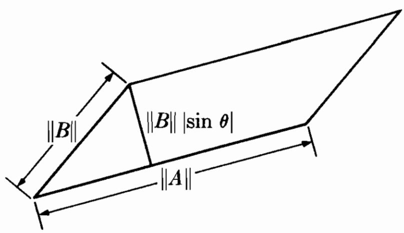


Figure 29

Example. Let $A = (3, 1, 4)$ and $B = (-2, 5, 3)$. Then the area of the parallelogram spanned by A and B is easily computed. First we get the cross product,

$$A \times B = (3 - 20, -8 - 9, 15 + 2) = (-17, -17, 17).$$

The area of the parallelogram spanned by A and B is therefore equal to the norm of this vector, and that is

$$\|A \times B\| = \sqrt{3 \cdot 17^2} = 17\sqrt{3}.$$

These considerations will be used especially in Chapter XV, when we discuss surface area, and in Chapter XIII, when we deal with the change of variables formula.

Exercises

Find $A \times B$ for the following vectors.

1. $A = (1, -1, 1)$ and $B = (-2, 3, 1)$
2. $A = (-1, 1, 2)$ and $B = (1, 0, -1)$

3. $A = (1, 1, -3)$ and $B = (-1, -2, -3)$
4. Find $A \times A$ and $B \times B$, in Exercises 1 through 3.
5. Let $E_1 = (1, 0, 0)$, $E_2 = (0, 1, 0)$, and $E_3 = (0, 0, 1)$. Find $E_1 \times E_2$, $E_2 \times E_3$, $E_3 \times E_1$.
6. Show that for any vector A in 3-space we have $A \times A = O$.
7. Compute $E_1 \times (E_1 \times E_2)$ and $(E_1 \times E_1) \times E_2$. Are these vectors equal to each other?
8. Carry out the proofs of **CP 1** through **CP 4.**
9. Compute the area of the parallelogram spanned by the following vectors.
   (a) $A = (3, -2, 4)$ and $B = (5, 1, 1)$
   (b) $A = (3, 1, 2)$ and $B = (-1, 2, 4)$
   (c) $A = (4, -2, 5)$ and $B = (3, 1, -1)$
   (d) $A = (-2, 1, 3)$ and $B = (2, -3, 4)$

CHAPTER II

Differentiation of Vectors

We begin to acquire the flavor of the mixture of algebra, geometry, and differentiation. Each gains in appeal from being mixed with the other two.

The chain rule especially leads into the classical theory of curves. As you will see, the chain rule in its various aspects occurs very frequently in this book, and forms almost as basic a tool as the algebra of vectors, with which it will in fact be intimately mixed.

§1. *Derivative*

Let I be an interval. A parametrized **curve** (defined on this interval) is an association which to each point of I associates a vector. If X denotes a curve defined on I, and t is a point of I, then $X(t)$ denotes the vector associated to t by X. We often write the association $t \mapsto X(t)$ as an arrow

$$X: I \to \mathbf{R}^n.$$

Each vector $X(t)$ can be written in terms of coordinates,

$$X(t) = (x_1(t), \ldots, x_n(t)),$$

each $x_i(t)$ being a function of t. We say that this curve is **differentiable** if each function $x_i(t)$ is a differentiable function of t.

For instance, the curve defined by

$$X(t) = (\cos t, \sin t, t)$$

is a spiral (Fig. 1). Here we have

$$x(t) = \cos t,$$

$$y(t) = \sin t,$$

$$z(t) = t.$$

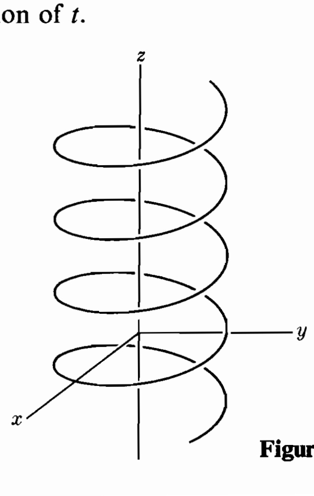


Figure 1

Remark. We take the intervals of definition for our curves to be open, closed, or also half-open or half-closed. When we define the derivative of a curve, it is understood that the interval of definition contains more than one point. In that case, at an end point the usual limit of

$$\frac{f(a+h)-f(a)}{h}$$

is taken for those h such that the quotient makes sense, i.e. $a+h$ lies in the interval. If a is a left end point, the quotient is considered only for $h > 0$. If a is a right end point, the quotient is considered only for $h < 0$. Then the usual rules for differentiation of functions are true in this greater generality, and thus Rules 1 through 4 below, and the chain rule of §2 remain true also. [An example of a statement which is not always true for curves defined over closed intervals is given in Exercise 11(b).]

Let us try to differentiate vectors using a Newton quotient. We consider

$$\frac{X(t+h)-X(t)}{h} = \left(\frac{x_1(t+h)-x_1(t)}{h}, \dots, \frac{x_n(t+h)-x_n(t)}{h}\right)$$

and see that each component is a Newton quotient for the corresponding coordinate. If each $x_i(t)$ is differentiable, then each quotient

$$\frac{x_i(t+h)-x_i(t)}{h}$$

approaches the derivative dx_i/dt. For this reason, we define the **derivative** dX/dt to be

$$\frac{dX}{dt} = \left(\frac{dx_1}{dt}, \dots, \frac{dx_n}{dt}\right).$$

In fact, we could also say that the vector

$$\left(\frac{dx_1}{dt}, \dots, \frac{dx_n}{dt}\right)$$

is the limit of the Newton quotient

$$\frac{X(t+h)-X(t)}{h}$$

as h approaches 0. Indeed, as h approaches 0, each component

$$\frac{x_i(t+h)-x_i(t)}{h}$$

approaches dx_i/dt. Hence the Newton quotient approaches the vector

$$\left(\frac{dx_1}{dt}, \dots, \frac{dx_n}{dt}\right).$$

For example, if $X(t) = (\cos t, \sin t, t)$ then

$$\frac{dX}{dt} = (-\sin t, \cos t, 1).$$

Physicists often denote dX/dt by $\dot{X}$; thus in the previous example, we could also write

$$\dot{X}(t) = (-\sin t, \cos t, 1) = X'(t).$$

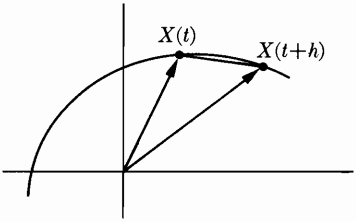


Figure 2

We define the **velocity vector** of the curve at time t to be the vector $X'(t)$. In our previous example, when

$$X(t) = (\cos t, \sin t, t),$$

the velocity vector at $t = \pi$ is

$$X'(\pi) = (0, -1, 1),$$

and for $t = \pi/4$ we get

$$X'(\pi/4) = (-1/\sqrt{2}, 1/\sqrt{2}, 1).$$

The velocity vector is located at the origin, but when we translate it to the point $X(t)$, then we visualize it as tangent to the curve, as in the next picture.

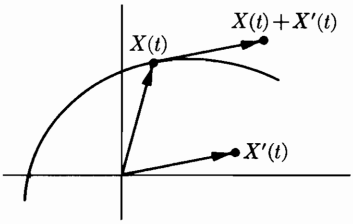


Figure 3

We define the **tangent line** to a curve X at time t to be the line passing through $X(t)$ in the direction of $X'(t)$, provided that $X'(t) \neq O$. Otherwise, we don't define a tangent line.

Example 1. Find a parametric equation of the tangent line to the curve $X(t) = (\sin t, \cos t)$ at $t = \pi/3$.

We have

$$X'(\pi/3) = (\tfrac{1}{2}, -\sqrt{3}/2) \qquad \text{and} \qquad X(\pi/3) = (\sqrt{3}/2, \tfrac{1}{2}).$$

Let $P = X(\pi/3)$ and $A = X'(\pi/3)$. Then a parametric equation of the tangent line at the required point is

$$L(t) = P + tA = \left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right) + \left(\frac{1}{2}, -\frac{\sqrt{3}}{2}\right) t.$$

(We use another letter L because X is already occupied.) In terms of the coordinates $L(t) = \big(x(t), y(t)\big)$, we can write the tangent line as

$$x(t) = \frac{\sqrt{3}}{2} + \frac{1}{2}t,$$

$$y(t) = \frac{1}{2} - \frac{\sqrt{3}}{2}t.$$

We define the **speed** of the curve $X(t)$ to be the length of the velocity vector. If we denote the speed by $v(t)$, then by definition we have

$$v(t) = \|X'(t)\|,$$

and thus

$$v(t)^2 = X'(t)^2 = X'(t) \cdot X'(t).$$

We can also omit the t from the notation, and write

$$v = X' \cdot X' = X'^2.$$

We define the **acceleration vector** to be the derivative dX'/dt, provided of course that X' is differentiable. We shall also denote the acceleration vector by X''. We define the acceleration scalar to be the length of the acceleration vector, and denote it by $a(t)$.

In the example given by $X(t) = (\cos t, \sin t, t)$ we find that

$$X''(t) = (-\cos t, -\sin t, 0).$$

Therefore $\|X''(t)\| = 1$ and we see that the spiral has a constant acceleration scalar, but not a constant acceleration vector.

Warning. $a(t)$ is not necessarily the derivative of $v(t)$. Almost any example shows this. For instance, let

$$X(t) = (\sin t, \cos t).$$

Then $v(t) = \|X(t)\| = 1$ so that $dv/dt = 0$. However, a simple computation shows that $X''(t) = (\cos t, -\sin t)$ and hence $a(t) = 1$.

We shall list the rules for differentiation. These will concern sums, products, and the chain rule which is postponed to the next section. We make a remark concerning products. If X is a curve and f a function, defined on the same interval I, then for each t in this interval we can take the product

$$f(t)X(t)$$

of the **numb**[redacted]t) by the **vector** $X(t)$. Thus if

$$X(t) = \big(x_1(t), \ldots, x_n(t)\big)$$

then

$$f(t)X(t) = \big(f(t)x_1(t), \ldots, f(t)x_n(t)\big).$$

For instance, if $X(t) = (\cos t, \sin t, t)$ and $f(t) = e^t$, then

$$f(t)X(t) = (e^t \cos t, e^t \sin t, e^t t),$$

and

$$f(\pi)X(\pi) = \big(e^\pi(-1), e^\pi(0), e^\pi\pi\big) = (-e^\pi, 0, e^\pi\pi).$$

The derivative of a curve is defined componentwise. Thus the rules for the derivative will be very similar to the rules for differentiating functions.

Rule 1. *Let $X(t)$ and $Y(t)$ be two differentiable curves (defined for the same values of t). Then the sum $X(t) + Y(t)$ is differentiable, and*

$$\frac{d\big(X(t) + Y(t)\big)}{dt} = \frac{dX}{dt} + \frac{dY}{dt}.$$

Rule 2. *Let c be a number, and let $X(t)$ be differentiable. Then $cX(t)$ is differentiable, and*

$$\frac{d\big(cX(t)\big)}{dt} = c\frac{dX}{dt}.$$

Rule 3. *Let $f(t)$ be a differentiable function, and $X(t)$ a differentiable curve (defined for the same values of t). Then $f(t)X(t)$ is differentiable, and*

$$\frac{d(fX)}{dt} = f(t)\frac{dX}{dt} + \frac{df}{dt}X(t).$$

Rule 4. *Let $X(t)$ and $Y(t)$ be two differentiable curves (defined for the same values of t). Then $X(t) \cdot Y(t)$ is a differentiable function whose derivative is*

$$\frac{d}{dt}[X(t) \cdot Y(t)] = X'(t) \cdot Y(t) + X(t) \cdot Y'(t).$$

(This is formally analogous to the derivative of a product of functions, namely the first times the derivative of the second plus the second times the derivative of the first, except that the product is now a scalar product.)

As an example of the proofs we shall give the third one in detail, and leave the others to you as exercises.

Let $X(t) = \big(x_1(t), \ldots, x_n(t)\big)$, and let $f = f(t)$ be a function. Then by definition

$$f(t)X(t) = \big(f(t)x_1(t), \ldots, f(t)x_n(t)\big).$$

We take the derivative of each component and apply the rule for the

derivative of a product of functions. We obtain:

$$\frac{d(fX)}{dt} = \left(f(t)\frac{dx_1}{dt} + \frac{df}{dt}x_1(t), \ldots, f(t)\frac{dx_n}{dt} + \frac{df}{dt}x_n(t)\right).$$

Using the rule for the sum of two vectors, we see that the expression on the right is equal to

$$\left(f(t)\frac{dx_1}{dt}, \ldots, f(t)\frac{dx_n}{dt}\right) + \left(\frac{df}{dt}x_1(t), \ldots, \frac{df}{dt}x_n(t)\right).$$

We can take f out of the vector on the left and df/dt out of the vector on the right to obtain

$$f(t)\frac{dX}{dt} + \frac{df}{dt}X(t),$$

as desired.

Example 2. Let A be a fixed vector, and let f be an ordinary differentiable function of one variable. Let $F(t) = f(t)A$. Then $F'(t) = f'(t)A$. For instance, if $F(t) = (\cos t)A$ and $A = (a, b)$ where a, b are fixed numbers, then $F(t) = (a \cos t, b \cos t)$ and thus

$$F'(t) = (-a \sin t, -b \sin t) = (-\sin t)A.$$

Similarly, if A, B are fixed vectors, and

$$G(t) = (\cos t)A + (\sin t)B,$$

then

$$G'(t) = (-\sin t)A + (\cos t)B.$$

One can also give a proof for the derivative of a product which does not use coordinates and is similar to the proof for the derivative of a product of functions. We carry this proof out. We must consider the Newton quotient

$$\frac{X(t+h)\cdot Y(t+h) - X(t)\cdot Y(t)}{h}$$

$$= \frac{X(t+h)\cdot Y(t+h) - X(t)\cdot Y(t+h) + X(t)\cdot Y(t+h) - X(t)\cdot Y(t)}{h}$$

$$= \frac{X(t+h) - X(t)}{h}\cdot Y(t+h) + X(t)\cdot\frac{Y(t+h) - Y(t)}{h}.$$

Taking the limit as $h \to 0$, we find

$$X'(t)\cdot Y(t) + X(t)\cdot Y'(t)$$

as desired.

Note that this type of proof applies without change if we replace the dot product by, say, the cross product. A coordinate proof for the derivative of the cross product can also be given (cf. Exercise 25).

Exercises

Find the velocity vector of the following curves.

1. $(e^t, \cos t, \sin t)$
2. $(\sin 2t, \log(1+t), t)$
3. $(\cos t, \sin t)$
4. $(\cos 3t, \sin 3t)$
5. In Exercises 3 and 4, show that the velocity vector is perpendicular to the position vector. Is this also the case in Exercises 1 and 2?
6. In Exercises 3 and 4, show that the acceleration vector is in the opposite direction from the position vector.
7. Let A, B be two constant vectors. What is the velocity vector of the curve $X = A + tB$?
8. Let $X(t)$ be a differentiable curve. A plane or line which is perpendicular to the velocity vector $X'(t)$ at the point $X(t)$ is said to be **normal** to the curve at the point t or also at the point $X(t)$. Find the equation of a line normal to the curves of Exercises 3 and 4 at the point $\pi/3$.
9. Find the equation of a plane normal to the curve
$$(e^t, t, t^2)$$
at the point $t = 1$.
10. Same question at the point $t = 0$.
11. Let $X(t)$ be a differentiable curve defined on an open interval. Let Q be a point which is not on the curve.
    (a) Write down the formula for the distance between Q and an arbitrary point on the curve.
    (b) If t_0 is a value of t such that the distance between Q and $X(t_0)$ is at a minimum, show that the vector $Q - X(t_0)$ is normal to the curve, at the point $X(t_0)$. [*Hint:* Investigate the minimum of the square of the distance.]
    (c) If $X(t)$ is the parametric equation of a straight line, show that there exists a unique value t_0 to such that the distance between Q and $X(t_0)$ is a minimum.
12. Assume that the differentiable curve $X(t)$ lies on the sphere of radius 1. Show that the velocity vector is perpendicular to the position vector. [*Hint:* Start from the condition $X(t)^2 = 1$.]
13. Let A be a non-zero vector, c a number, and Q a point. Let P_0 be the point of intersection of the line passing through Q, in the direction of A, and the plane $X \cdot A = c$. Show that for all points P of the plane, we have
$$\|Q - P_0\| \leqq \|Q - P\|.$$
[*Hint:* If $P \neq P_0$, consider the straight line passing through P_0 and P, and use Exercise 11(c).]

14. Prove that if the acceleration of a curve is always perpendicular to its velocity, then its speed is constant.

15. Let B be a non-zero vector, and let $X(t)$ be such that $X(t) \cdot B = t$ for all t. Assume also that the angle between $X'(t)$ and B is constant. Show that $X''(t)$ is perpendicular to $X'(t)$.

16. Write a parametric equation for the tangent line to the given curve at the given point in each of the following cases.
    (a) $(\cos 4t, \sin 4t, t)$ at the point $t = \pi/8$
    (b) $(t, 2t, t^2)$ at the point $(1, 2, 1)$
    (c) $(e^{3t}, e^{-3t}, 3\sqrt{2}t)$ at $t = 1$
    (d) (t, t^3, t^4) at the point $(1, 1, 1)$

17. Let A, B be fixed non-zero vectors. Let

$$X(t) = e^{2t}A + e^{-2t}B.$$

Show that $X''(t)$ has the same direction as $X(t)$.

18. Show that the two curves $(e^t, e^{2t}, 1 - e^{-t})$ and $(1 - \theta, \cos\theta, \sin\theta)$ intersect at the point $(1, 1, 0)$. What is the angle between their tangents at that point?

19. At what points does the curve $(2t^2, 1 - t, 3 + t^2)$ intersect the plane $3x - 14y + z - 10 = 0$?

20. Let $X(t)$ be a differentiable curve and suppose that $X'(t) = O$ for all t throughout its interval of definition I. What can you say about the curve? Suppose $X'(t) \neq O$ but $X''(t) = O$ for all t in the interval. What can you say about the curve?

21. Let $X(t) = (a\cos t, a\sin t, bt)$, where a, b are constant. Let $\theta(t)$ be the angle which the tangent line at a given point of the curve makes with the z-axis. Show that $\cos\theta(t)$ is the constant $b/\sqrt{a^2 + b^2}$.

22. Show that the velocity and acceleration vectors of the curve in Exercise 21 have constant lengths.

23. Let B be a fixed unit vector, and let $X(t)$ be a curve such that $X(t) \cdot B = e^{2t}$ for all t. Assume also that the velocity vector of the curve has a constant angle θ with the vector B, with $0 < \theta < \pi/2$.
    (a) Show that the speed is $2e^{2t}/\cos\theta$.
    (b) Determine the dot product $X'(t) \cdot X''(t)$ in terms of t and θ.

24. Let

$$X(t) = \left(\frac{2t}{1 + t^2}, \frac{1 - t^2}{1 + t^2}, 1\right).$$

Show that the cosine of the angle between $X(t)$ and $X'(t)$ is constant.

25. Using the definition of the cross product by coordinates given in Chapter I, prove that if $X(t)$ and $Y(t)$ are two differentiable curves (defined for the

same values of t), then

$$\frac{d[X(t) \times Y(t)]}{dt} = X(t) \times \frac{dY(t)}{dt} + \frac{dX(t)}{dt} \times Y(t).$$

26. Show that

$$\frac{d}{dt}[X(t) \times X'(t)] = X(t) \times X''(t).$$

27. Let $Y(t) = X(t) \times X'(t)$. Show that $Y'(t) = X(t) \times X''(t)$.
28. Let $Y(t) = X(t) \cdot (X'(t) \times X''(t))$. Show that $Y' = X \cdot (X' \times X''')$.

§2. *Length of curves*

We define the **length** of a curve X between two values a, b of t ($a \leqq b$) in the interval of definition of the curve to be the integral of the speed:

$$\int_a^b v(t)\,dt = \int_a^b \|X'(t)\|\,dt.$$

By definition, we can rewrite this integral in the form

$$\int_a^b \sqrt{\left(\frac{dx_1}{dt}\right)^2 + \cdots + \left(\frac{dx_n}{dt}\right)^2}\,dt.$$

When $n = 2$, this is the same formula for the length which we gave in an earlier course. Thus the formula in dimension n is a very natural generalization of the formula in dimension 2. Namely, when

$$X(t) = (x(t), y(t))$$

is given by two coordinates, then the length of the curve between a and b is equal to

$$\int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\,dt.$$

Example. Let the curve be defined by

$$X(t) = (\sin t, \cos t).$$

Then $X'(t) = (\cos t, -\sin t)$ and $v(t) = \sqrt{\cos^2 t + \sin^2 t} = 1$. Hence the length of the curve between $t = 0$ and $t = 1$ is

$$\int_0^1 v(t)\,dt = t\Big|_0^1 = 1.$$

In this case, of course, the integral is easy to evaluate. There is no reason why this should always be the case.

Example. Set up the integral for the length of the curve

$$X(t) = (e^t, \sin t, t)$$

between $t = 1$ and $t = \pi$.

We have $X'(t) = (e^t, \cos t, 1)$. Hence the desired integral is

$$\int_1^\pi \sqrt{e^{2t} + \cos^2 t + 1}\, dt.$$

In this case, there is no easy formula for the integral. In the exercises, however, the functions are adjusted in such a way that the integral can be evaluated by elementary techniques of integration. Don't expect this to be the case in real life, though.

Exercises

1. Find the length of the spiral $(\cos t, \sin t, t)$ between $t = 0$ and $t = 1$.
2. Find the length of the spiral $(\cos 2t, \sin 2t, 3t)$ between $t = 1$ and $t = 3$.
3. Find the length of the indicated curve for the given interval:
   (a) $(\cos 4t, \sin 4t, t)$ between $t = 0$ and $t = \pi/8$.
   (b) $(t, 2t, t^2)$ between $t = 1$ and $t = 3$.
   (c) $(e^{3t}, e^{-3t}, 3\sqrt{2}t)$ between $t = 0$ and $t = \frac{1}{3}$.
4. Find the length of the curve defined by

$$X(t) = (t - \sin, t, 1 - \cos t)$$

   between (a) $t = 0$ and $t = 2\pi$, (b) $t = 0$ and $t = \pi/2$.
5. Find the length of the curve $X(t) = (t, \log t)$ between (a) $t = 1$ and $t = 2$, (b) $t = 3$ and $t = 5$.
   [*Hint:* Substitute $u^2 = 1 + t^2$ to evaluate the integral.]
6. Find the length of the curve defined by $X(t) = (t, \log \cos t)$ between $t = 0$ and $t = \pi/4$.

§3. *The chain rule and applications*

This section may be omitted if the course is pressed for time or other topics.

Let X be a vector and c a number. As a matter of notation it will be convenient to define Xc to be cX, in other words, we allow ourselves to multiply vectors by numbers on the right. If we have a curve $X(t)$ defined

for some interval, and a function $g(t)$ defined on the same interval, then we let

$$X(t)g(t) = g(t)X(t).$$

Let $X = X(t)$ be a differentiable curve. Let f be a function defined on some interval, such that the values of f lie in the domain of definition of the curve $X(t)$. Then we may form the composite curve $X \circ f$. If s is a number at which f is defined, we let the value of $X \circ f$ at s be

$$(X \circ f)(s) = X\big(f(s)\big).$$

For example, let $X(t) = (t^2, e^t)$ and let $f(s) = \sin s$. Then

$$X\big(f(s)\big) = (\sin^2 s, e^{\sin s}).$$

Each component of $X\big(f(s)\big)$ becomes a function of s, just as when we studied the chain rule for functions.

Chain Rule. *If X is a differentiable curve and f is a differentiable function defined on some interval, whose values are contained in the interval of definition of the curve, then the composite curve $X \circ f$ is differentiable, and*

$$(X \circ f)'(s) = X'\big(f(s)\big)f'(s).$$

The expression on the right can also be written $f'(s)X'\big(f(s)\big)$. It is the product of the function f' times the vector X'.

In another notation, if we let $t = f(s)$, then we can write the above formula in the form

$$\frac{d(X \circ f)}{ds} = \frac{dX}{dt}\frac{dt}{ds}.$$

The proof of the chain rule is trivial, using the chain rule for functions. Indeed, let $Y(s) = X\big(f(s)\big)$. Then

$$Y(s) = \big(x_1(f(s)), \ldots, x_n(f(s))\big).$$

Taking the derivative term by term, we find:

$$Y'(s) = \big(x_1'(f(s))f'(s), \ldots, x_n'(f(s))f'(s)\big).$$

We can take $f'(s)$ outside the vector, and get

$$Y'(s) = X'\big(f(s)\big)f'(s),$$

which is precisely what we want.

The change of variables from t to s is also called a change of parametrization of the curve. Under certain changes of parametrization, certain formulas involving the velocity and acceleration of the curve become simpler and reflect geometric properties more clearly. We shall see examples of this in a moment.

Let us now assume that all the functions with which we dealt above have second derivatives. Using the chain rule, and the rule for the derivative of a product, we obtain the following two formulas:

$$Y'(s) = f'(s)X'(f(s)), \tag{1}$$

$$Y''(s) = f''(s)X'(f(s)) + (f'(s))^2X''(f(s)). \tag{2}$$

We shall consider an important special case of these formulas.

We have defined

$$v(t) = \|\dot{X}(t)\|$$

to be the speed. Let us now assume that each coordinate function of $X'(t)$ is continuous. In that case, we say that $X'(t)$ is **continuous.** Then $v(t)$ is a continuous function of t. *We shall assume throughout that $v(t) \neq 0$ for any value of t in the interval of definition of our curve.* Then $v(t) > 0$ for all such values of t. We let

$$s(t) = \int v(t)\,dt$$

be a fixed indefinite integral of $v(t)$ over our interval. (For instance, if a is a point of the interval, we could let

$$s(t) = \int_a^t v(u)\,du.$$

We know that any two indefinite integrals of v over the interval differ by a constant.) Then

$$\frac{ds}{dt} = v(t) > 0$$

for all values of t, and hence s is a strictly increasing function. Consequently, the inverse function exists. Call it

$$t = f(s).$$

We can then write

$$X(t) = X(f(s)) = Y(s).$$

Thus we are in the situation described above.

The velocity vectors of the curve depending on the two different parametrizations are related as in formula (1). From the theory of derivatives of inverse functions, we know that

$$f'(s) = \frac{df}{ds} = \left(\frac{ds}{dt}\right)^{-1}.$$

Hence $f'(s)$ is always positive. This means that in the present case, $Y'(s)$ and $X'(t)$ have the same direction when $t = f(s)$.

A curve $Y: J \to \mathbf{R}^n$ is said to be **parametrized by arc length** if $\|Y'(s)\| = 1$ for all s in the interval of definition J. The reason for this is contained in the next theorem.

Theorem 1. *Let $X: I \to \mathbf{R}^n$ be a curve whose speed $v(t)$ is > 0 for all t in the interval of definition. Let*

$$s(t) = \int_a^t v(u)\,du$$

and $t = f(s)$ be the inverse function. Then the curve given by

$$s \mapsto Y(s) = X(f(s))$$

is parametrized by arc length, and $Y'(s)$ is perpendicular to $Y''(s)$ for each value of s.

Proof. From formula (1), we get

$$\|Y'(s)\| = |f'(s)|\,\|X'(t)\| = \frac{df}{ds}\frac{ds}{dt}.$$

By what we just saw above, this last expression is equal to 1. Thus $Y'(s)$ is a vector of length 1, a unit vector, in the same direction as $X'(t)$. Thus the velocity vector of the curve Y has constant length.

In particular, we have $Y'(s)^2 = 1$. Differentiating with respect to s, we get

$$2Y' \cdot Y'' = 0.$$

Hence $Y'(s)$ is perpendicular to $Y''(s)$ for each value of s. This proves the theorem.

From (2), we see that the acceleration $Y''(s)$ has two components. First a tangential component

$$f''(s)X'(t)$$

parallel to $X'(t)$, which involves the naive notion of scalar acceleration, namely the second derivative $f''(s)$. Second, another component in the direction of $X''(t)$, with a coefficient

$$(f'(s))^2$$

which is positive. [We assume of course that $X''(t) \neq O$.]

For a given value of t, let us assume that $X'(t) \neq O$ and $X''(t) \neq O$, and also that $X'(t)$ and $X''(t)$ do not lie on the same straight line. Then the plane passing through $X(t)$, parallel to $X'(t)$ and $X''(t)$ is called the **osculating plane** of the curve at time t, or also at the point $X(t)$. [Actually, it is more accurate to say at time t, because there may be two numbers t_1, t_2 in the interval of definition of the curve such that $X(t_1) = X(t_2)$.]

Example 1. Let $X(t) = (\sin t, \cos t, t)$. Find the osculating plane to this curve at $t = \pi/2$.

We have

$$X'(\pi/2) = (0, -1, 1)$$

and

$$X''(\pi/2) = (-1, 0, 0).$$

We find first a vector perpendicular to $X'(\pi/2)$ and $X''(\pi/2)$. For instance, $N = (0, 1, 1)$ is such a vector. Furthermore, let $P = X(\pi/2) = (1, 0, \pi/2)$. Then the osculating plane at $t = \pi/2$ is the plane passing through P, perpendicular to N, and its equation is therefore

$$y + z = \pi/2.$$

In case of parametrization by arc length, or in fact in any other parametrization such that $f'(s) \neq 0$, we see from formulas (1) and (2) that the plane parallel to $X'(t)$ and $X''(t)$ is the same as the plane parallel to $Y'(s)$ and $Y''(s)$ because from these formulas, we can solve back for $X'(t)$ and $X''(t)$ in terms of this other pair of vectors. Thus the osculating plane does not depend on a change of parametrization $t = f(s)$ such that $f'(s) \neq 0$.

Let us assume that a curve is parametrized by arc length. Thus we write the curve as $Y(s)$, and by Theorem 1, we have $\|Y'(s)\| = 1$ and

$$Y'(s) \cdot Y''(s) = 0.$$

Then $Y'(s)$ and $Y''(s)$ look like this:

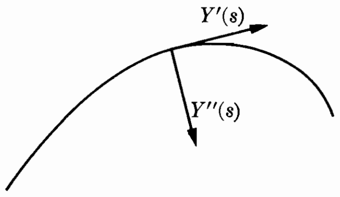


Figure 4

Example 2. Let R be a number > 0. A parametrization for the circle of radius R by arc length is given by

$$Y(s) = \left(R\cos\frac{s}{R}, R\sin\frac{s}{R}\right),$$

as one sees immediately, because $\|Y'(s)\| = 1$.

Differentiating twice shows that

$$Y''(s) = -\frac{1}{R}\left(\cos\frac{s}{R}, \sin\frac{s}{R}\right)$$

and hence that

$$\|Y''(s)\| = \frac{1}{R} \quad \text{or} \quad R = \frac{1}{\|Y''(s)\|}.$$

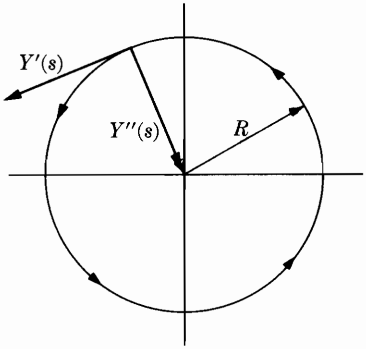


Figure 5

For an arbitrary curve Y parametrized by arc length, it is customary to make a definition which is motivated by the geometry of the special example just discussed, namely we define the **radius of curvature** $R(s)$ to be

$$R(s) = \frac{1}{\|Y''(s)\|}$$

at all points such that $\|Y''(s)\| \neq 0$. (Note that if $Y''(s) = O$ on some interval, then $Y(s) = As + B$ for suitable vectors A, B, and thus Y parametrizes a straight line. Thus intuitively, it is reasonable to view its radius of curvature as infinity.)

The same motivation as above leads us to define the **curvature** itself to be $\|Y''(s)\|$. The curvature is usually denoted by k.

Most curves are not usually given parametrized by arc length, and thus it is useful to have a formula which gives the curvature in terms of the given parameter t. This comes immediately from the chain rule. Indeed, keeping our notation $X(t)$ and $Y(s)$ with $ds/dt = v(t)$, we have the formula

$$\boxed{Y''(s) = \frac{1}{v(t)} \frac{d}{dt}\left(\frac{1}{v(t)} X'(t)\right)}$$

where $v(t) = \|X'(t)\|$ is the length of the velocity vector $X'(t)$.

Proof. From formula (1), we know that

$$Y'(s) = \frac{dt}{ds} X'(t) = \frac{1}{v(t)} X'(t).$$

By the chain rule,

$$Y''(s) = \frac{d\big(Y'(s(t))\big)}{dt} \frac{dt}{ds},$$

which yields precisely the formula in the box.

The curvature is then equal to the length of the vector in the box, that is:

$$k = \left\| \frac{1}{v(t)} \frac{d}{dt} \left(\frac{1}{v(t)} X'(t) \right) \right\| .$$

Example 3. Find the curvature of the curve given by

$$X(t) = (\cos t, \sin t, t).$$

We have $X'(t) = (-\sin t, \cos t, 1)$ and $v(t) = \sqrt{2}$ is constant. Then $X''(t) = (-\cos t, -\sin t, 0)$, and from the formula for the curvature we find

$$k(t) = \frac{1}{\sqrt{2}} \left\| \frac{1}{\sqrt{2}} X''(t) \right\| = \frac{1}{2}.$$

We see in particular that the curve has constant curvature.

Exercises

1. Find the equations of the osculating planes for each of the following curves at the given point.
   (a) $(\cos 4t, \sin 4t, t)$ at the point $t = \pi/8$
   (b) $(t, 2t, t^2)$ at the point $(1, 2, 1)$
   (c) $(e^{3t}, e^{-3t}, 3\sqrt{2}t)$ at $t = 1$
   (d) (t, t^3, t^4) at the point $(1, 1, 1)$
2. Prove formula (2) from formula (1) in detail.
3. Let r be a fixed number > 0, let $c > 0$, and let

$$X(t) = (r \cos t, r \sin t, ct).$$

   Find the curvature as a function of t.
4. Find the curvature of the curve

$$X(t) = (t, t^2, t^3)$$

   at (a) $t = 1$, (b) $t = 0$, (c) $t = -1$.

5. Let the plane curve be defined by $X(t) = \big(x(t), y(t)\big)$. Show that the curvature is given by

$$k(t) = \frac{|x'(t)y''(t) - x''(t)y'(t)|}{\big(x'^2(t) + y'^2(t)\big)^{3/2}}.$$

6. If a curve is parametrized by $x = t$, $y = f(t)$ (the natural parametrization arising from a function $y = f(x)$), find a simplification for the curvature given in the preceding exercise.

7. Find the radius of curvature of the curve $X(t) = (t, \log t)$. For which t is the radius of curvature a minimum?

8. Find the curvatures of the curves

(a) $X(t) = (t, \sin t)$,
(b) $X(t) = (\sin 3t, \cos 3t)$,
(c) $X(t) = (\sin 3t, \cos 3t, t)$.

9. Find the radius of curvature of the parabola $y = x^2$.

10. Find the radius of curvature of the ellipse given by

$$X(t) = (a \cos t, b \sin t),$$

where a, b are constants.

11. Find the curvature of the curve defined by

$$x(t) = \int_0^t \cos \frac{\pi u^2}{2}\, du,$$

$$y(t) = \int_0^t \sin \frac{\pi u^2}{2}\, du.$$

12. Find the curvature of the curve defined by

$$x(t) = \int_0^t \frac{\cos u}{\sqrt{u}}\, du,$$

$$y(t) = \int_0^t \frac{\sin u}{\sqrt{u}}\, du$$

in terms of the arc length s.

13. Show that the curvature of the curve defined by

$$X(t) = (e^t, e^{-t}, \sqrt{2}\, t)$$

is equal to $\sqrt{2}/(e^t - e^{-t})^2$.

14. If a curve has constant velocity and acceleration, show that the curvature is constant. Express the curvature in terms of the lengths of the velocity and acceleration vectors.

CHAPTER III

Functions of Several Variables

We view functions of several variables as functions of points in space. This appeals to our geometric intuition, and also relates such functions more easily with the theory of vectors. The gradient will appear as a natural generalization of the derivative. In this chapter we are mainly concerned with basic definitions and notions. We postpone the important theorems to the next chapter.

§1. *Graphs and level curves*

In order to conform with usual terminology, and for the sake of brevity, a collection of objects will simply be called a **set**. In this chapter, we are mostly concerned with sets of points in space.

Let S be a set of points in n-space. A **function** (defined on S) is an association which to each element of S associates a number.

In practice, we sometimes omit mentioning explicitly the set S, since the context usually makes it clear for which points the function is defined.

Example 1. In 2-space (the plane) we can define a function f by the rule

$$f(x, y) = x^2 + y^2.$$

It is defined for all points (x, y) and can be interpreted geometrically as the square of the distance between the origin and the point.

Example 2. Again in 2-space, let

$$f(x, y) = \frac{x^2 - y^2}{x^2 + y^2}$$

be defined for all

$$(x, y) \neq (0, 0).$$

We do not define f at $(0, 0)$ (also written O).

Example 3. In 3-space, we can define a function f by the rule

$$f(x, y, z) = x^2 - \sin(xyz) + yz^3.$$

Since a point and a vector are represented by the same thing (namely an n-tuple), we can think of a function such as the above also as a function of vectors. When we do not want to write the coordinates, we write $f(X)$ instead of $f(x_1, \ldots, x_n)$. As with numbers, we call $f(X)$ the **value** of f at the point (or vector) X.

Just as with functions of one variable, one can define the **graph** of a function f of n variables $x_1, \ldots, x_n$ to be the set of points in $(n + 1)$-space of the form

$$(x_1, \ldots, x_n, f(x_1, \ldots, x_n)),$$

the $(x_1, \ldots, x_n)$ being in the domain of definition of f. Thus when $n = 1$, the graph of a function f is a set of points $(x, f(x))$. When $n = 2$, the graph of a function f is the set of points $(x, y, f(x, y))$. When $n = 2$, it is already difficult to draw the graph since it involves a figure in 3-space. The graph of a function of two variables may look like this:

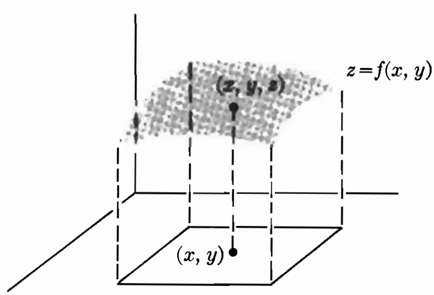


Figure 1

When we get to the graph of a function of three variables, it is of course impossible to draw it, since it exists in 4-space. However, we shall describe another means of visualizing the function.

For each number c, the equation $f(x, y) = c$ is the equation of a curve in the plane. We have considerable experience in drawing the graphs of such curves, and we may therefore assume that we know how to draw this graph in principle. This curve is called the **level curve** of f at c. It gives us the set of points (x, y) where f takes on the value c. By drawing a number of such level curves, we can get a good description of the function.

Example 1 (continued). The level curves are described by equations

$$x^2 + y^2 = c.$$

These have a solution only when $c \geqq 0$. In that case, they are circles

(unless $c = 0$ in which case the circle of radius 0 is simply the origin). In Fig. 2, we have drawn the level curves for $c = 1$ and 4.

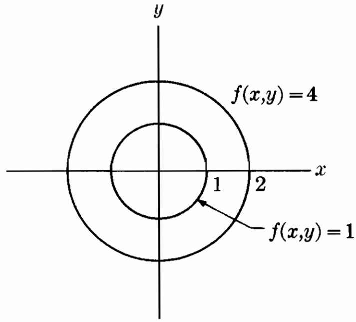


Figure 2

The graph of the function $z = f(x, y) = x^2 + y^2$ is then a figure in 3-space, which we may represent as follows.

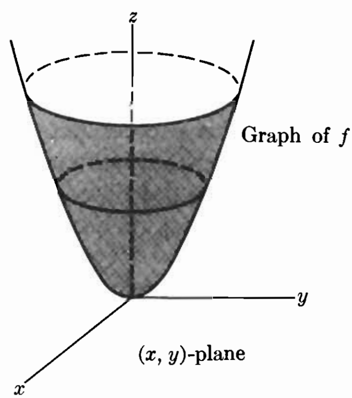


Figure 3

Example 2 (continued). To find the level curves in Example 2, we have to determine the values (x, y) such that

$$x^2 - y^2 = c(x^2 + y^2)$$

for a given number c. This amounts to solving $x^2(1 - c) = y^2(1 + c)$. If $x = 0$, then $f(0, y) = -1$. Thus on the y-axis our function has the constant value -1. If $x \neq 0$, then we can divide by x in the above equality, and we obtain (for $c \neq -1$)

$$\frac{y^2}{x^2} = \frac{1 - c}{1 + c}.$$

Taking the square root, we obtain two level lines, namely

$$y = ax \quad \text{and} \quad y = -ax, \quad \text{where} \quad a = \sqrt{\frac{1 - c}{1 + c}}.$$

Thus the level curves are straight lines (excluding the origin). We have drawn some of them in Fig. 4. (The numbers indicate the value of the function on the corresponding line.)

It would of course be technically much more disagreeable to draw the level lines in Example 3, and we shall not do so.

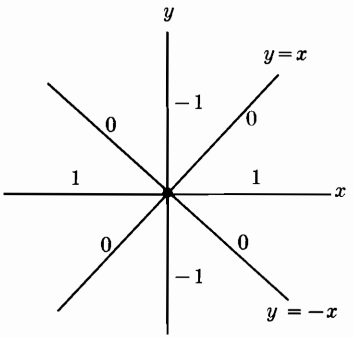


Figure 4

We see that the level lines are based on the same principle as the contour lines of a map. Each line describes, so to speak, the altitude of the function. If the graph is interpreted as a mountainous region, then each level curve gives the set of points of constant altitude. In Example 1, a person wanting to stay at a given altitude need but walk around in circles. In Example 2, such a person should walk on a straight line towards or away from the origin.

If we deal with a function of three variables, say $f(x, y, z)$, then $(x, y, z) = X$ is a point in 3-space. In that case, the set of points satisfying the equation

$$f(x, y, z) = c$$

for some constant c is a surface. The notion analogous to that of level curve is that of level surface.

In physics, a function f might be a potential function, giving the value of the potential energy at each point of space. The level surfaces are then sometimes called surfaces of **equipotential.** The function f might also give a temperature distribution (i.e. its value at a point X is the temperature at X). In that case, the level surfaces are called **isothermal** surfaces.

Exercises

Sketch the level curves for the functions $z = f(x, y)$, where $f(x, y)$ is given by the following expressions.

1. $x^2 + 2y^2$
2. $y - x^2$
3. $y - 3x^2$
4. $x - y^2$
5. $3x^2 + 3y^2$
6. xy

7. $(x - 1)(y - 2)$ 8. $(x + 1)(y + 3)$ 9. $\dfrac{x^2}{4} + \dfrac{y^2}{16}$

10. $2x - 3y$ 11. $\dfrac{xy}{x^2 + y^2}$ 12. $\dfrac{xy^2}{x^2 + y^4}$

13. $\dfrac{4xy(x^2 - y^2)}{x^2 + y^2}$ (try polar coordinates)

14. $\dfrac{x + y}{x - y}$ 15. $\dfrac{x^2 + y^2}{x^2 - y^2}$

(In Exercises 11, 12, and 13, the function is not defined at $(0, 0)$. In Exercise 14, it is not defined for $y = x$, and in Exercise 15 it is not defined for $y = x$ or $y = -x$.)

16. $(x - 1)^2 + (y + 3)^2$ 17. $x^2 - y^2$

§2. *Partial derivatives*

In this ~~section~~ and the next, we discuss the notion of differentiability for functions of several variables. When we discussed the derivative of functions of one variable, we assumed that such a function was defined on an interval. We shall have to make a similar assumption in the case of several variables, and for this we need to introduce a new notion.

Let P be a point in n-space, and let a be a number > 0. The set of points X such that

$$\|X - P\| < a$$

will be called the **open ball** of radius a and center P. The set of points X such that

$$\|X - P\| \leqq a$$

will be called the **closed ball** of radius a and center P. The set of points X such that

$$\|X - P\| = a$$

will be called the **sphere** of radius a and center P.

Thus when $n = 1$, we are in 1-space, and the open ball of radius a is the open interval centered at P. The sphere of radius a and center P consists only of two points.

When $n = 2$, the open ball of radius a and center P is also called the open **disc.** The sphere is the **circle.**

When $n = 3$, then our terminology coincides with the obvious interpretation we might want to place on the words.

The following are the pictures of the spheres of radius 1 in 2-space and 3-space respectively centered at the origin.

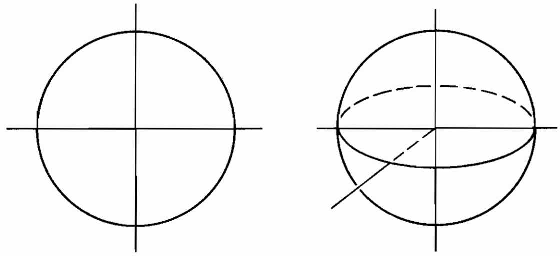


Figure 5

Let S_1 be the sphere of radius 1, centered at the origin. Let a be a number > 0. If X is a point of the sphere S_1, then aX is a point of the sphere of radius a, because

$$\|aX\| = a\|X\| = a.$$

In this manner, we get all points of the sphere of radius a. (Proof?) Thus the sphere of radius a is obtained by stretching the sphere of radius 1, through multiplication by a.

A similar remark applies to the open and closed balls of radius a, they being obtained from the open and closed balls of radius 1 through multiplication by a. (Prove this as an exercise.)

Let U be a set of points in n-space. We shall say that U is an **open set** in n-space if the following condition is satisfied: Given any point P in U, there exists an open ball B of radius $a > 0$ which is centered at P and such that B is contained in U.

Example 1. In the plane, the set consisting of the first quadrant, excluding the x- and y-axes, is an open set.

The x-axis is not open in the plane (i.e. in 2-space). Given a point on the x-axis, we cannot find an open disc centered at the point and contained in the x-axis.

On the other hand, if we view the x-axis as the set of points in 1-space, then it is open in 1-space. Similarly, the interval

$$-1 < x < 1$$

is open in 1-space, but not open in 2-space, or n-space for $n > 1$.

Example 2. Let U be the open ball of radius $a > 0$ centered at the origin. Then U is an open set. To prove this, let P be a point of this ball, so $\|P\| < a$. Say $\|P\| = b$. Let $c = a - b$. If X is a point such that $\|X - P\| < c$, then

$$\|X\| \leqq \|X - P\| + \|P\| < a - b + b = a.$$

Hence the open ball of radius c centered at P is contained in U. Hence U is open.

In the next picture we have drawn an open set in the plane, consisting of the region inside the curve, but not containing any point of the boundary. We have also drawn a point P in U, and a sphere (disc) around P contained in U.

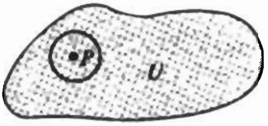


Figure 6

When we defined the derivative as a limit of

$$\frac{f(x+h) - f(x)}{h},$$

we needed the function f to be defined in some open interval around the point x.

Now let f be a function of n variables, defined on an open set U. Then for any point X in U, the function f is also defined at all points which are close to X, namely all points which are contained in an open ball centered at X and contained in U.

For small values of h, the point

$$(x_1 + h, x_2, \ldots, x_n)$$

is contained in such an open ball. Hence the function is defined at that point, and we may form the quotient

$$\frac{f(x_1 + h, x_2, \ldots, x_n) - f(x_1, \ldots, x_n)}{h}.$$

If the limit exists as h tends to 0, then we call it the **first partial derivative** of f and denote it by $D_1 f(x_1, \ldots, x_n)$, or $D_1 f(X)$, or also by

$$\frac{\partial f}{\partial x_1}.$$

Similarly, we let

$$\begin{aligned} D_i f(X) &= \frac{\partial f}{\partial x_i} \\ &= \lim_{h \to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_n)}{h} \end{aligned}$$

if it exists, and call it the i-th partial derivative.

When $n = 2$ and we work with variables (x, y), then the first and second partials are also noted

$$\frac{\partial f}{\partial x} \quad \text{and} \quad \frac{\partial f}{\partial y}.$$

By definition, we therefore have

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

and

$$\frac{\partial f}{\partial y} = \lim_{k \to 0} \frac{f(x, y + k) - f(x, y)}{k}.$$

A partial derivative is therefore obtained by keeping all but one variable fixed, and taking the ordinary derivative with respect to this one variable.

Example 3. Let $f(x, y) = x^2y^3$. Then

$$\frac{\partial f}{\partial x} = 2xy^3 \quad \text{and} \quad \frac{\partial f}{\partial y} = 3x^2y^2.$$

We observe that the partial derivatives are themselves functions. This is the reason why the notation D_if is sometimes more useful than the notation $\partial f/\partial x_i$. It allows us to write $D_if(P)$ for any point P in the set where the partial is defined. There cannot be any ambiguity or confusion with a (meaningless) symbol $D_i(f(P))$, since $f(P)$ is a number. Thus $D_if(P)$ means $(D_if)(P)$. It is the value of the function D_if at P.

Example 4. Let $f(x, y) = \sin xy$. To find $D_2f(1, \pi)$, we first find $\partial f/\partial y$, or $D_2f(x, y)$, which is simply

$$D_2f(x, y) = (\cos xy)x.$$

Hence

$$D_2f(1, \pi) = (\cos \pi) \cdot 1 = -1.$$

Also,

$$D_2f\left(3, \frac{\pi}{4}\right) = \left(\cos \frac{3\pi}{4}\right) \cdot 3 = -\frac{1}{\sqrt{2}} \cdot 3 = -\frac{3}{\sqrt{2}} \cdot$$

Let f be defined in an open set U and assume that the partial derivatives of f exist at each point X of U. The **vector**

$$\left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right) = (D_1f(X), \ldots, D_nf(X)),$$

whose components are the partial derivatives, will be called the **gradient** of f at X and will be denoted by $\operatorname{grad} f(x)$. One must read this

$$(\operatorname{grad} f)(X),$$

but we shall usually omit the parentheses around grad f. Sometimes one also writes ∇f instead of grad f.

If f is a function of two variables (x, y), then we have

$$\nabla f(x, y) = \operatorname{grad} f(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right).$$

Example 5. Let $f(x, y) = x^2y^3$. Then

$$\operatorname{grad} f(x, y) = (2xy^3, 3x^2y^2),$$

so that in this case,

$$\operatorname{grad} f(1, 2) = (16, 12).$$

Thus the gradient of a function f associates a **vector** to a point X.

If f is a function of three variables (x, y, z), then

$$\operatorname{grad} f\cancel{(x, y, z)} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right).$$

Using the formula for the derivative of a sum of two functions, and the derivative of a constant times a function, we conclude at once that the gradient satisfies the following properties:

Theorem 1. *Let f, g be two functions defined on an open set U, and assume that their partial derivatives exist at every point of U. Let c be a number. Then*

$$\operatorname{grad}(f + g) = \operatorname{grad} f + \operatorname{grad} g$$
$$\operatorname{grad}(cf) = c \operatorname{grad} f.$$

You should carry out the details of the proof as an exercise.

We shall give later several geometric and physical interpretations for the gradient.

Exercises

Find the partial derivatives

$$\frac{\partial f}{\partial x}, \quad \frac{\partial f}{\partial y}, \quad \text{and} \quad \frac{\partial f}{\partial z},$$

for the following functions $f(x, y)$ or $f(x, y, z)$.

1. $xy + z$
2. $x^2y^5 + 1$
3. $\sin(xy) + \cos z$
4. $\cos(xy)$
5. $\sin(xyz)$
6. e^{xyz}
7. $x^2 \sin(yz)$
8. xyz
9. $xz + yz + xy$
10. $x \cos(y - 3z) + \arcsin(xy)$
11. Find $\operatorname{grad} f(P)$ if P is the point $(1, 2, 3)$ in Exercises 1, 2, 6, 8, and 9.
12. Find $\operatorname{grad} f(P)$ if P is the point $(1, \pi, \pi)$ in Exercises 4, 5, 7.

13. Find grad $f(P)$ if

$$f(x, y, z) = \log(z + \sin(y^2 - x))$$

and

$$P = (1, -1, 1).$$

14. Find the partial derivatives of x^y.

Find the gradient of the following functions at the given point.

15. $f(x, y, z) = e^{-2x}\cos(yz)$ at $(1, \pi, \pi)$
16. $f(x, y, z) = e^{3x+y}\sin(5z)$ at $(0, 0, \pi/6)$
17. Prove that an open ball of radius $a > 0$ centered at some point Q is in fact an open set.

§3. *Differentiability and gradient*

Let f be a function defined on an open set U. Let X be a point of U. For all vectors H such that $\|H\|$ is small (and $H \neq O$), the point $X + H$ also lies in the open set. However we *cannot* form a quotient

$$\frac{f(X + H) - f(X)}{H}$$

because it is meaningless to divide by a vector. In order to define what we mean for a function f to be differentiable, we must therefore find a way which does not involve dividing by H.

We reconsider the case of functions of one variable. Let us fix a number x. We had defined the derivative to be

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

Let

$$g(h) = \frac{f(x + h) - f(x)}{h} - f'(x).$$

Then $g(h)$ is not defined when $h = 0$, but

$$\lim_{h \to 0} g(h) = 0.$$

We can write

$$f(x + h) - f(x) = f'(x)h + hg(h).$$

This relation has meaning so far only when $h \neq 0$. However, we observe that if we define $g(0)$ to be 0, then the preceding relation is obviously true when $h = 0$ (because we just get $0 = 0$).

Furthermore, we can replace h by $-h$ if we replace g by $-g$. Thus we have shown that if f is differentiable, there exists a function g such that

$$(1) \qquad f(x + h) - f(x) = f'(x)h + |h|g(h),$$

$$\lim_{h \to 0} g(h) = 0.$$

Conversely, suppose that there exists a number a and a function $g(h)$ such that

(1a) $$f(x + h) - f(x) = ah + |h|g(h),$$
$$\lim_{h \to 0} g(h) = 0.$$

We find for $h \neq 0$,

$$\frac{f(x + h) - f(x)}{h} = a + \frac{|h|}{h} g(h).$$

Taking the limit as h approaches 0, we observe that

$$\lim_{h \to 0} \frac{|h|}{h} g(h) = 0.$$

Hence the limit of the Newton quotient exists and is equal to a. Hence f is differentiable, and its derivative $f'(x)$ is equal to a.

Therefore, the existence of a number a and a function g satisfying (1a) above could have been used as the definition of differentiability in the case of functions of one variable. The great advantage of (1) is that no h appears in the denominator. It is this relation which will suggest to us how to define differentiability for functions of several variables, and how to prove the chain rule for them.

We now consider a function of n variables.

Let f be a function defined on an open set U. Let X be a point of U. If $H = (h_1, \dots, h_n)$ is a vector such that $\|H\|$ is small enough, then $X + H$ will also be a point of U and so $f(X + H)$ is defined. Note that

$$X + H = (x_1 + h_1, \dots, x_n + h_n).$$

This is the generalization of the $x + h$ with which we dealt previously.

When f is a function of two variables, which we write (x, y), then we use the notation $H = (h, k)$ so that

$$X + H = (x + h, y + k).$$

The point $X + H$ is close to X and we are interested in the difference $f(X + H) - f(X)$, which is the difference of the value of the function at $X + H$ and the value of the function at X. If this difference approaches 0 when H approaches O, then we say that f is **continuous**. We say that f is differentiable at X if the partial derivatives $D_1f(X), \dots, D_nf(X)$ exist, and if there exists a function g (defined for small H) such that

$$\lim_{H \to O} g(H) = 0 \qquad \left(\text{also written } \lim_{\|H\| \to O} g(H) = 0\right)$$

and

$$f(X + H) - f(X) = D_1f(X)h_1 + \cdots + D_nf(X)h_n + \|H\|g(H).$$

With the other notation for partial derivatives, this last relation reads:

$$f(X + H) - f(X) = \frac{\partial f}{\partial x_1} h_1 + \cdots + \frac{\partial f}{\partial x_n} h_n + \|H\| g(H).$$

We say that f is **differentiable** in the open set U if it is differentiable at every point of U, so that the above relation holds for every point X in U.

In view of the definition of the gradient in §2, we can rewrite our fundamental relation in the form

$$\boxed{f(X + H) - f(X) = \big(\operatorname{grad} f(X)\big) \cdot H + \|H\| g(H).} \tag{2}$$

The term $\|H\|g(H)$ has an order of magnitude smaller than the previous term involving the dot product. This is one advantage of the present notation. We know how to handle the formalism of dot products and are accustomed to it, and its geometric interpretation. This will help us later in interpreting the gradient geometrically.

For the moment, we observe that the gradient is the only vector which will make formula (2) valid (cf. Exercise 5).

In two variables, the definition of differentiability reads

$$f(x + h, y + k) - f(x, y) = \frac{\partial f}{\partial x} h + \frac{\partial f}{\partial y} k + \|H\| g(H).$$

We view the term

$$\frac{\partial f}{\partial x} h + \frac{\partial f}{\partial y} k$$

as an approximation to $f(X + H) - f(X)$, depending in a particularly simple way on h and k.

If we use the abbreviation

$$\operatorname{grad} f = \nabla f,$$

then formula (2) can be written

$$\boxed{f(X + H) - f(X) = \nabla f(X) \cdot H + \|H\| g(H).}$$

As with $\operatorname{grad} f$, one must read $(\nabla f)(X)$ and not the meaningless $\nabla\big(f(X)\big)$ since $f(X)$ is a number for each value of X, and thus it makes no sense to apply ∇ to a number. The symbol ∇ is applied to the function f, and $(\nabla f)(X)$ is the value of ∇f at X.

Example. Suppose that we consider values for H pointing only in the direction of the standard unit vectors. In the case of two variables, consider for instance $H = (h, 0)$. Then for such H, the condition for differentiability reads:

$$f(X + H) = f(x + h, y) = f(x, y) + \frac{\partial f}{\partial x} h + |h|g(H).$$

In higher dimensional space, let $E_i = (\ldots, 0, 1, 0, \ldots)$ be the i-th unit vector. Let $H = hE_i$ for some number h, so that

$$H = (\ldots, 0, h, 0, \ldots).$$

Then for such H,

$$f(X + H) = f(X + hE_i) = f(X) + \frac{\partial f}{\partial x_i} h + |h|g(H).$$

Example. We can often estimate error terms with an expression $\|H\|g(H)$, where $g(H)$ approaches 0 as $\|H\|$ approaches 0, by using standard properties of the absolute value, namely

$$|a + b| \leqq |a| + |b|.$$

For instance, let $H = (h, k)$ where h, k are numbers. Then by definition,

$$\|H\| = \sqrt{h^2 + k^2} \qquad \text{and} \qquad \|H\|^2 = h^2 + k^2.$$

Observe that

$$h^2 \leqq h^2 + k^2 = \|H\|^2.$$

Hence

$$|h| \leqq \|H\|.$$

Similarly,

$$|h^2 + hk| \leqq |h^2| + |hk| = |h|^2 + |h|\,|k|.$$

Hence

$$|h^2 + hk| \leqq \|H\|^2 + \|H\|\,\|H\| \leqq 2\|H\|^2.$$

Example. You should read this example in connection with the last step of the proof of the next theorem. If you do not wish to put too much emphasis on theory, take the next theorem for granted and skip both this example and the proof. Let g_1, g_2 be functions defined for small values of H such that

$$\lim_{H \to O} g_1(H) = 0 \qquad \text{and} \qquad \lim_{H \to O} g_2(H) = 0.$$

We want to see that the expression

$$hg_1(H) + kg_2(H)$$

can be put in the form $\|H\|g(H)$ where $\lim_{H\to O} g(H) = 0$. We write

$$hg_1(H) + kg_2(H) = \|H\| \frac{h}{\|H\|} g_1(H) + \|H\| \frac{k}{\|H\|} g_2(H)$$

$$= \|H\| \left[\frac{h}{\|H\|} g_1(H) + \|H\| \frac{k}{\|H\|} g_2(H) \right].$$

Let $g(H)$ be the expression in brackets. Each factor $h/\|H\|$ and $k/\|H\|$ has absolute value $\leqq 1$. Hence each one of the terms inside the bracket approaches 0 as H approaches O. Thus we have written

$$hg_1(H) + kg_2(H) = \|H\|g(H),$$

as desired.

Theorem 2. *Let f be a function defined on some open set U. Assume that its partial derivatives exist for every point in this open set, and that they are continuous. Then f is differentiable.*

Proof. For simplicity of notation, we shall use two variables. Thus we deal with a function $f(x, y)$. We let $H = (h, k)$. Let (x, y) be a point in U, and take H small, $H \neq (0, 0)$. We have to consider the difference $f(X + H) - f(X)$, which is simply

$$f(x + h, y + k) - f(x, y).$$

This is equal to

$$f(x + h, y + k) - f(x, y + k) + f(x, y + k) - f(x, y).$$

Applying the mean value theorem for functions of *one* variable, and applying the definition of partial derivatives, we see that there is a number s between x and $x + h$ such that

$$f(x + h, y + k) - f(x, y + k) = D_1f(s, y + k)h. \tag{3}$$

Similarly, there is a number t between y and $y + k$ such that

$$f(x, y + k) - f(x, y) = D_2f(x, t)k. \tag{4}$$

We shall now analyze the expressions on the right-hand side of equations (3) and (4).

Let

$$g_1(H) = D_1f(s, y + k) - D_1f(x, y).$$

As H approaches O, $(s, y + k)$ approaches (x, y) because s is between x and $x + h$. Since $D_1 f$ is continuous, it follows that

$$\lim_{H \to O} g_1(H) = 0.$$

But

$$D_1 f(s, y + k) = D_1 f(x, y) + g_1(H).$$

Hence equation (3) can be rewritten as

$$(5) \qquad f(x + h, y + k) - f(x, y + k) = D_1 f(x, y)h + hg_1(H).$$

By a similar argument, we can rewrite equation (4) in the form

$$(6) \qquad f(x, y + k) - f(x, y) = D_2 f(x, y)k + kg_2(H)$$

with some function $g_2(H)$ such that

$$\lim_{H \to O} g_2(H) = 0.$$

If we add (5) and (6) we obtain

$$(7)\ f(X + H) - f(X) = D_1 f(X)h + D_2 f(X)k + hg_1(H) + kg_2(H).$$

In view of the example given before our theorem, we see that the last two terms on the right are of the form $\|H\|g(H)$. This proves the theorem.

Remark 1. If we dealt with n variables, then we would consider the expression for $f(X + H) - f(X)$ given by

$$\begin{aligned} &f(x_1 + h_1, \ldots, x_n + h_n) - f(x_1, x_2 + h_2, \ldots, x_n + h_n) \\ &+ f(x_1, x_2 + h_2, \ldots, x_n + h_n) - f(x_1, x_2, \ldots, x_n + h_n) \\ &\qquad \vdots \\ &+ f(x_1, \ldots, x_{n-1}, x_n + h_n) - f(x_1, \ldots, x_n). \end{aligned}$$

We would then apply the mean value theorem at each step, take the sum, and argue in essentially the same way as with two variables.

Remark 2. Some sort of smoothness assumption on the function besides the existence of the partial derivatives must be made in order to insure that it is differentiable at a point. For instance, consider the function f defined by

$$f(x, y) = \frac{xy}{x^2 + y^2} \qquad \text{if} \qquad (x, y) \neq (0, 0)$$

$$f(0, 0) = 0.$$

You should have worked out the level lines for this function, and found that they are given by straight lines through the origin. In particular, you see that the function is not continuous at the origin. However, its partial

derivatives exist and are easily computed by using the definitions, namely:

$$D_1 f(0, 0) = \lim_{h \to 0} \frac{f(0 + h, 0) - f(0, 0)}{h}$$

$$= \lim_{h \to 0} \frac{h \cdot 0 - 0}{h} = \lim_{h \to 0} 0 = 0.$$

Similarly, $D_2 f(0, 0) = 0$. Now do Exercise 8.

Exercises

1. Let $f(x, y) = 2x - 3y$. What is $\partial f/\partial x$ and $\partial f/\partial y$?
2. Let $A = (a, b)$ and let f be the function on $\mathbf{R}^2$ such that $f(X) = A \cdot X$. Let $X = (x, y)$. In terms of the coordinates of A, determine $\partial f/\partial x$ and $\partial f/\partial y$.
3. Let $A = (a, b, c)$ and let f be the function on $\mathbf{R}^3$ such that $f(X) = A \cdot X$. Let $X = (x, y, z)$. In terms of the coordinates of A, determine $\partial f/\partial x$, $\partial f/\partial y$, and $\partial f/\partial z$.
4. Generalize the above two exercises to n-space.
5. Let f be defined on an open set U. Let X be a point of U. Let A be a vector, and let g be a function defined for small H, such that

$$\lim_{H \to O} g(H) = 0.$$

Assume that

$$f(X + H) - f(X) = A \cdot H + \|H\| g(H).$$

Prove that $A = \operatorname{grad} f(X)$. You may do this exercise in 2 variables first and then in 3 variables, and let it go at that. Use coordinates, e.g. let $A = (a, b)$ and $X = (x, y)$. Use special values of H.

6. Let $H = (h, k)$. Prove:
(a) $|h^2 + 3hk| \leqq 4\|H\|^2$.
(b) $|h^3 + h^2 k + k^3| \leqq 3\|H\|^3$.
(c) $|3hk^2 + 2h^3| \leqq 5\|H\|^3$.
(d) $|(h + k)^4| \leqq 16\|H\|^4$.
(e) $|(h + k)|^3 < 8\|H\|^3$.
7. Let

$$g(h, k) = \frac{h^2 - k^2}{h^2 + k^2}$$

be defined for $(h, k) \neq (0, 0)$. Find

$$\lim_{h \to 0} g(h, k), \qquad \lim_{k \to 0} \left[\lim_{h \to 0} g(h, k) \right]$$

$$\lim_{k \to 0} g(h, k), \qquad \lim_{h \to 0} \left[\lim_{k \to 0} g(h, k) \right].$$

8. Compute the partial derivatives of the function $f(x, y)$ given at the end of the section at any point $(x, y) \neq (0, 0)$ by the usual formulas. You see that the partial derivatives exist everywhere, but the function is not continuous.

CHAPTER IV

The Chain Rule and the Gradient

In this chapter, we prove the chain rule for functions of several variables and give a number of applications. Among them will be several interpretations for the gradient. These form one of the central points of our theory. They show how powerful the tools we have accumulated turn out to be.

§1. *The chain rule*

Let f be a function defined on some open set U. Let $t \mapsto X(t)$ be a curve such that the values $X(t)$ are contained in U. Then we can form the composite function $f \circ X$, which is a function of t, given by

$$(f \circ X)(t) = f(X(t)).$$

As an example, take $f(x, y) = e^x \sin(xy)$. Let $X(t) = (t^2, t^3)$. Then

$$f(X(t)) = e^{t^2} \sin(t^5).$$

This is a function of t in the old sense of functions of one variable.

The chain rule tells us how to find the derivative of this function, provided we know the gradient of f and the derivative X'. Its statement is as follows.

Chain Rule. *Let f be a function which is defined and differentiable on an open set U. Let $X: I \to \mathbf{R}^n$ be a differentiable curve (defined for some interval of numbers t) such that the values $X(t)$ lie in the open set U. Then the function*

$$t \mapsto f(X(t))$$

is differentiable (as a function of t), and

$$\frac{df(X(t))}{dt} = (\text{grad}\, f(X(t))) \cdot X'(t).$$

In the notation dX/dt, this also reads

$$\frac{df(X(t))}{dt} = (\text{grad}\, f)(X(t)) \cdot \frac{dX}{dt}.$$

Before proving the chain rule, we restate it in terms of components. If $X = (x_1, \ldots, x_n)$ then

$$\boxed{\frac{d(f(X(t)))}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt}.}$$

If f is a function of two variables (x, y) then

$$\boxed{\frac{df(X(t))}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.}$$

This can be applied to the seemingly more general situation when x, y are functions of more than one variable t. Suppose for instance that

$$x = \varphi(t, u) \qquad \text{and} \qquad y = \psi(t, u)$$

are differentiable functions of two variables. Let

$$g(t, u) = f(\varphi(t, u), \psi(t, u)).$$

If we keep u fixed and take the partial derivative of g with respect to t, then we can apply our chain rule, and obtain

$$\frac{\partial g}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}.$$

The components are of course useful in computations, to determine partial derivatives explicitly, but they will not be used in the proof.

Proof of the chain rule. By definition, we must investigate the quotient

$$\frac{f(X(t+h)) - f(X(t))}{h}.$$

Let

$$K = K(t, h) = X(t+h) - X(t).$$

Then our quotient can be rewritten in the form

$$\frac{f(X(t) + K) - f(X(t))}{h}.$$

Using the definition of differentiability for f, we have

$$f(X + K) - f(X) = (\operatorname{grad} f)(X) \cdot K + \|K\| g(K)$$

and

$$\lim_{\|K\| \to 0} g(K) = 0.$$

Replacing K by what it stands for, namely $X(t+h) - X(t)$, and dividing by h, we obtain:

$$\frac{f(X(t+h)) - f(X(t))}{h} = (\operatorname{grad} f)(X(t)) \cdot \frac{X(t+h) - X(t)}{h}$$
$$\pm \left\| \frac{X(t+h) - X(t)}{h} \right\| g(K).$$

As h approaches 0, the first term of the sum approaches what we want, namely

$$(\operatorname{grad} f)(X(t)) \cdot X'(t).$$

The second term approaches

$$\pm \|X'(t)\| \lim_{h \to 0} g(K),$$

and when h approaches 0, so does $K = X(t+h) - X(t)$. Hence the second term of the sum approaches 0. This proves our chain rule.

Example 1. Let $f(x, y) = x^2 + 2xy$. Let $x = r\cos\theta$ and $y = r\sin\theta$. Let $g(r, \theta) = f(r\cos\theta, r\sin\theta)$ be the composite function. Find $\partial g/\partial\theta$.

We have

$$\frac{\partial x}{\partial \theta} = -r\sin\theta \quad \text{and} \quad \frac{\partial y}{\partial \theta} = r\cos\theta.$$

Hence

$$\frac{\partial g}{\partial \theta} = (2x + 2y)(-r\sin\theta) + 2x(r\cos\theta).$$

If you want the answer completely in terms of r, θ, you can substitute $r\cos\theta$ and $r\sin\theta$ for x and y respectively in this expression.

Example 2. Let $w = f(x, y, z) = e^{xy}\cos z$ and let

$$x = tu, \qquad y = \sin(tu), \qquad z = u^2.$$

Then

$$\frac{\partial w}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial u}$$

$$= ye^{xy}(\cos z)t + xe^{xy}(\cos z)(\cos tu)t - e^{xy}(\sin z)2u$$

$$= \sin(tu)e^{tu\sin(tu)}(\cos u^2)t + tue^{tu\sin(tu)}(\cos u^2)(\cos tu)t$$
$$- e^{tu\sin(tu)}(\sin u^2)2u.$$

In this last expression, we have substituted the values for x, y, z in terms of t and u, thus giving the partial derivative completely in terms of these variables.

Example 3. Sometimes the letters x and y are occupied to denote variables which are not the first and second variables of the function f. In this case, other letters must be used if we wish to replace D_1f and D_2f by partial derivatives with respect to these variables. For example, let

$$u = f(x^2 - y, xy).$$

To find $\partial u/\partial x$, we let

$$s = x^2 - y \qquad \text{and} \qquad t = xy.$$

Then

$$\begin{aligned}\frac{\partial u}{\partial x} &= \frac{\partial f}{\partial s}\frac{\partial s}{\partial x} + \frac{\partial f}{\partial t}\frac{\partial t}{\partial x}\\ &= \frac{\partial f}{\partial s}2x + \frac{\partial f}{\partial t}y = D_1f(s, t)2x + D_2f(s, t)y.\end{aligned}$$

The function u depends on x, y and we may write $u = g(x, y)$. Then

$$\frac{\partial g}{\partial x} = \frac{\partial f}{\partial s}2x + \frac{\partial f}{\partial t}y. \tag{1}$$

Similarly,

$$\frac{\partial g}{\partial y} = \frac{\partial f}{\partial s}(-1) + \frac{\partial f}{\partial t}x. \tag{2}$$

We may then solve the linear equations (1) and (2), and we find for instance

$$\frac{\partial f}{\partial t} = \frac{1}{y + 2x^2}\left[\frac{\partial g}{\partial x} + 2x\frac{\partial g}{\partial y}\right].$$

The advantage of the D_1f, D_2f notation is that it does not depend on a choice of letters, and makes it clear that we take the partial derivatives of f with respect to the first and second variables. On the other hand, it is slightly more clumsy to write $D_1f(s, t)$ rather than $\partial f/\partial s$. Thus the second notation, **when used with an appropriate choice of variables,** is shorter and a little more mechanical. We emphasize, however, that it can only be used when the letters denoting the variables have been fixed properly.

Example 4. Let f be a function on $\mathbf{R}^3$. Let us interpret f as giving the temperature, so that at any point X in $\mathbf{R}^3$, the value of the function $f(X)$ is the temperature at X. Suppose that a bug moves in space along a differentiable curve, which we may denote in parametric form by

$$t \mapsto B(t).$$

Thus $B(t) = \big(x(t), y(t), z(t)\big)$ is the position of the bug at time t. Let us assume that the bug starts from a point where he feels that the temperature

is comfortable, and therefore that the temperature is constant along the path on which he moves. In other words, f is constant along the curve $B(t)$. This means that for all values of t, we have

$$f(B(t)) = c,$$

where c is constant. Differentiating with respect to t, and using the chain rule, we find that

$$\operatorname{grad} f\,(B(t)) \cdot B'(t) = 0.$$

This means that the gradient of f is perpendicular to the velocity vector at every point of the curve.

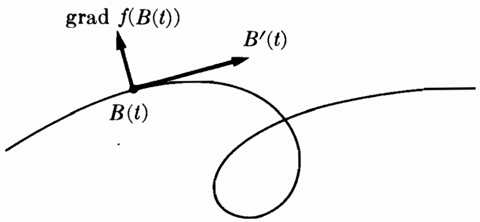


Figure 1

Example 5. Let $f(x, y, z) = g(x^2 - 3zy + xz)$, where g is a differentiable function of one variable. Then the chain rule becomes much simpler, and we find

$$\frac{\partial f}{\partial x} = g'(x^2 - 3zy + xz)(2x + z).$$

We denote the derivative of g by g' as usual. We do not write it as dg/dx, because x is a letter which is already occupied for other purposes. We could let

$$u = x^2 - 3zy + xz,$$

in which case it would be all right to write

$$\frac{\partial f}{\partial x} = \frac{dg}{du}\frac{\partial u}{\partial x},$$

and we would get the same answer as above. In general, if $h(x, y, z)$ is a function of x, y, z, and g is a function of one variable, then we may form the composite function

$$f(x, y, z) = g(h(x, y, z)).$$

We then have

$$\frac{\partial f}{\partial x} = g'(h(x, y, z))\frac{\partial h}{\partial x}.$$

Written in terms of the first partial, we have the longer (but more accurate) expression

$$D_1 f(x, y, z) = g'(h(x, y, z))D_1 h(x, y, z).$$

Through practice, you will recognize which notation to use most efficiently, depending on the cases to be considered.

Example 6. Let $g(t, x, y) = f(t^2x, ty)$. Then

$$\frac{\partial g}{\partial t} = D_1 f(t^2x, ty)2tx + D_2 f(t^2x, ty)y.$$

Here again, since the letter x is occupied, we cannot write $\partial f/\partial x$ for $D_1 f$.

Exercises

(All functions are assumed to be differentiable as needed.)

1. If $x = u(r, s, t)$ and $y = v(r, s, t)$ and $z = f(x, y)$, write out the formula for

$$\frac{\partial z}{\partial r} \quad \text{and} \quad \frac{\partial z}{\partial t}.$$

2. Find the partial derivatives with respect to x, y, s, and t for the following functions.

   (a) $f(x, y, z) = x^3 + 3xyz - y^2z$, $x = 2t + s$, $y = -t - s$, $z = t^2 + s^2$
   (b) $f(x, y) = (x + y)/(1 - xy)$, $x = \sin 2t$, $y = \cos(3t - s)$

3. Let $f(x, y, z) = (x^2 + y^2 + z^2)^{1/2}$. Find $\partial f/\partial x$ and $\partial f/\partial y$.
4. Let $r = (x_1^2 + \cdots + x_n^2)^{1/2}$. What is $\partial r/\partial x_i$?
5. If $u = f(x - y, y - x)$, show that

$$\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} = 0.$$

6. If $u = x^3 f(y/x, z/x)$, show that

$$x\frac{\partial u}{\partial x} + y\frac{\partial u}{\partial y} + z\frac{\partial u}{\partial z} = 3u.$$

7. (a) Let $x = r\cos\theta$ and $y = r\sin\theta$. Let $z = f(x, y)$. Show that

$$\frac{\partial z}{\partial r} = \frac{\partial f}{\partial x}\cos\theta + \frac{\partial f}{\partial y}\sin\theta, \qquad \frac{1}{r}\frac{\partial z}{\partial \theta} = -\frac{\partial f}{\partial x}\sin\theta + \frac{\partial f}{\partial y}\cos\theta.$$

   (b) If we let $z = g(r, \theta) = f(r\cos\theta, r\sin\theta)$, show that

$$\left(\frac{\partial g}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial g}{\partial \theta}\right)^2 = \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2.$$

8. (a) Let g be a function of r, let $r = \|X\|$, and $X = (x, y, z)$. Let $f(X) = g(r)$. Show that

$$\left(\frac{dg}{dr}\right)^2 = \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + \left(\frac{\partial f}{\partial z}\right)^2.$$

(b) Let $g(x, y) = f(x + y, x - y)$, where f is a differentiable function of two variables, say $f = f(u, v)$. Show that

$$\frac{\partial g}{\partial x}\frac{\partial g}{\partial y} = \left(\frac{\partial f}{\partial u}\right)^2 - \left(\frac{\partial f}{\partial v}\right)^2.$$

(c) Let $g(x, y) = f(2x + 7y)$, where f is a differentiable function of one variable. Show that

$$2\frac{\partial g}{\partial y} = 7\frac{\partial g}{\partial x}.$$

9. Let g be a function of r, and $r = \|X\|$. Let $f(X) = g(r)$. Find $\operatorname{grad} f(X)$ for the following functions.

(a) $g(r) = 1/r$ (b) $g(r) = r^2$
(c) $g(r) = 1/r^3$ (d) $g(r) = e^{-r^2}$
(e) $g(r) = \log\dfrac{1}{r}$ (f) $g(r) = 4/r^m$

10. Let $x = u\cos\theta - v\sin\theta$, and $y = u\sin\theta + v\cos\theta$, with θ equal to a constant. Let $f(x, y) = g(u, v)$. Show that

$$\left(\frac{\partial g}{\partial u}\right)^2 + \left(\frac{\partial g}{\partial v}\right)^2 = \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2.$$

The next five exercises concern certain parametizations, and some of the results from them will be used in Exercises 16 and 17.

11. Let A, B be two unit vectors such that $A \cdot B = 0$. Let

$$F(t) = (\cos t)A + (\sin t)B.$$

Show that $F(t)$ lies on the sphere of radius 1 centered at the origin, for each value of t.

12. Let P, Q be two points on the sphere of radius 1, centered at the origin. Let $L(t) = P + t(Q - P)$, with $0 \leqq t \leqq 1$. If there exists a value of t in $[0, 1]$ such that $L(t) = O$, show that $t = \frac{1}{2}$, and that $P = -Q$.

13. Let P, Q be two points on the sphere of radius 1. Assume that $P \neq -Q$. Show that there exists a differentiable curve joining P and Q on the sphere of radius 1, centered at the origin. [*Hint:* Divide $L(t)$ in Exercise 12 by its length.]

14. If P, Q are two unit vectors such that $P = -Q$, show that there exists a differentiable curve joining P and Q on the sphere of radius 1, centered at

the origin. You may assume that there exists a unit vector A which is perpendicular to P. Then use Exercise 11.

15. Parametrize the ellipse
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$
by a differentiable curve.

16. Let f be a differentiable function (in two variables) such that $\operatorname{grad} f(X) = cX$ for some constant c and all X in 2-space. Show that f is constant on any circle of radius $a > 0$, centered at the origin. [*Hint:* Put $x = a \cos t$ and $y = a \sin t$ and find df/dt.]

17. (a) Generalize the preceding exercise to the case of n variables. You may assume that any two points on the sphere of radius a centered at the origin are connected by a differentiable curve.
(b) Let f be a differentiable function in n variables, and assume that there exists a function g such that $\operatorname{grad} f(X) = g(X)X$. Show that f is constant on the sphere of radius $a > 0$ centered at the origin. (In other words, in Exercise 16, the hypothesis about the constant c can be weakened to an arbitrary function.)

18. Let $r = \|X\|$. Let g be a differentiable function of one variable whose derivative is never equal to 0. Let $f(X) = g(r)$. Show that $\operatorname{grad} f(X)$ is parallel to X for $X \neq O$.

19. Let f be a differentiable function of two variables and assume that there is an integer $m \geqq 1$ such that
$$f(tx, ty) = t^m f(x, y)$$
for all numbers t and all x, y. Prove **Euler's relation**
$$x\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y} = mf(x, y).$$

20. Generalize Exercise 19 to n variables, namely let f be a differentiable function of n variables and assume that there exists an integer $m \geqq 1$ such that $f(tX) = t^m f(X)$ for all numbers t and all points X in $\mathbf{R}^n$. Show that
$$x_1\frac{\partial f}{\partial x_1} + \cdots + x_n\frac{\partial f}{\partial x_n} = mf(X),$$
which can also be written $X \cdot \operatorname{grad} f(X) = mf(X)$. How does this exercise apply to Exercise 6?

21. Let f be a differentiable function defined on all of $\mathbf{R}^n$. Assume that $f(tP) = tf(P)$ for all numbers t and all points P in $\mathbf{R}^n$. Show that for all P we have
$$f(P) = \operatorname{grad} f(O) \cdot P.$$

§2. *Tangent plane*

Let f be a differentiable function and c a number. The set of points X such that $f(X) = c$ and $\operatorname{grad} f(X) \neq O$ is called a **surface.**

Let $X(t)$ be a differentiable curve. We shall say that the curve **lies on** the surface if, for all t, we have

$$f(X(t)) = c.$$

This simply means that all the points of the curve satisfy the equation of the surface. If we differentiate this relation, we get from the chain rule:

$$\operatorname{grad} f\,(X(t)) \cdot X'(t) = 0.$$

Let P be a point of the surface, and let $X(t)$ be a curve on the surface passing through P. This means that there is a number t_0 such that $X(t_0) = P$. For this value t_0, we obtain

$$\operatorname{grad} f(P) \cdot X'(t_0) = 0.$$

Thus the gradient of f at P is perpendicular to the tangent vector of the curve at P. [We assume that $X'(t_0) \neq O$.] This is true for *any* differentiable curve passing through P. It is therefore very reasonable to **define** the **plane** (or hyperplane) **tangent** to the surface at P to be the plane passing through P and perpendicular to the vector $\operatorname{grad} f(P)$. (We know from Chapter XVII how to find such planes.) This definition applies only when $\operatorname{grad} f(P) \neq O$. If $\operatorname{grad} f(P) = O$, then we do not define the notion of tangent plane.

The fact that $\operatorname{grad} f(P)$ is perpendicular to every curve passing through P on the surface also gives us an interpretation of the gradient as being perpendicular to the surface

$$f(X) = c,$$

which is one of the level surfaces for the function f (Fig. 2).

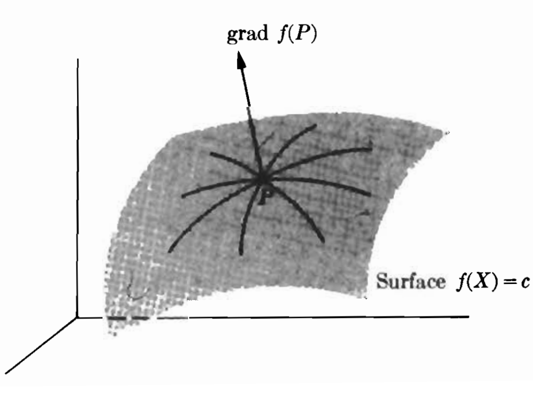


Figure 2

Example 1. Find the tangent plane to the surface

$$x^2 + y^2 + z^2 = 3$$

at the point (1, 1, 1).

Let $f(X) = x^2 + y^2 + z^2$. Then at the point $P = (1, 1, 1)$,

$$\operatorname{grad} f(P) = (2, 2, 2).$$

The equation of a plane passing through P and perpendicular to a vector N is

$$X \cdot N = P \cdot N.$$

In the present case, this yields

$$2x + 2y + 2z = 2 + 2 + 2 = 6.$$

Observe that our arguments also give us a means of finding a vector perpendicular to a curve in 2-space at a given point, simply by applying the preceding discussion to the plane instead of 3-space.

Example 2. Find the tangent line to the curve

$$x^2y + y^3 = 10$$

at the point (1, 2), and find a vector perpendicular to the curve at that point.

Let $f(x, y) = x^2y + y^3$. The gradient at the given point P is easily computed, and we find

$$\operatorname{grad} f(P) = (4, 13).$$

This is a vector N perpendicular to the curve at the given point. The tangent line is also given by $X \cdot N = P \cdot N$, and thus is

$$4x + 13y = 4 + 26 = 30.$$

Example 3. A surface may also be given in the form $z = g(x, y)$ where g is some function of two variables. In this case, the tangent plane is determined by viewing the surface as expressed by the equation

$$g(x, y) - z = 0.$$

For instance, suppose the surface is given by $z = x^2 + y^2$. We wish to determine the tangent plane at (1, 2, 5). Let $f(x, y, z) = x^2 + y^2 - z$. Then $\operatorname{grad} f(x, y, z) = (2x, 2y, -1)$ and

$$\operatorname{grad} f(1, 2, 5) = (1, 4, -1).$$

The equation of the tangent plane at $P = (1, 2, 5)$ perpendicular to $N = (1, 4, -1)$ is

$$x + 4y - z = P \cdot N = 4.$$

This is the desired equation.

Exercises

1. Find the equation of the tangent plane and normal line to each of the following surfaces at the specific point.
   (a) $x^2 + y^2 + z^2 = 49$ at $(6, 2, 3)$
   (b) $xy + yz + zx - 1 = 0$ at $(1, 1, 0)$
   (c) $x^2 + xy^2 + y^3 + z + 1 = 0$ at $(2, -3, 4)$
   (d) $2y - z^3 - 3xz = 0$ at $(1, 7, 2)$
   (e) $x^2y^2 + xz - 2y^3 = 10$ at $(2, 1, 4)$
   (f) $\sin xy + \sin yz + \sin xz = 1$ at $(1, \pi/2, 0)$
2. Let $f(x, y, z) = z - e^x \sin y$, and $P = (\log 3, 3\pi/2, -3)$. Find:
   (a) $\operatorname{grad} f(P)$,
   (b) the normal line at P to the level surface for f which passes through P,
   (c) the tangent plane to this surface at P.
3. Find the parametric equation of the tangent line to the curve of intersection of the following surfaces at the indicated point.
   (a) $x^2 + y^2 + z^2 = 49$ and $x^2 + y^2 = 13$ at $(3, 2, -6)$
   (b) $xy + z = 0$ and $x^2 + y^2 + z^2 = 9$ at $(2, 1, -2)$
   (c) $x^2 - y^2 - z^2 = 1$ and $x^2 - y^2 + z^2 = 9$ at $(3, 2, 2)$

   [*Note.* The tangent line above may be defined to be the line of intersection of the tangent planes of the given point.]
4. Let $f(X) = 0$ be a differentiable surface. Let Q be a point which does not lie on the surface. Given a differentiable curve $X(t)$ on the surface, defined on an open interval, give the formula for the distance between Q and a point $X(t)$. Assume that this distance reaches a minimum for $t = t_0$. Let $P = X(t_0)$. Show that the line joining Q to P is perpendicular to the curve at P.
5. Find the equation of the tangent plane to the surface $z = f(x, y)$ at the given point P when f is the following function:
   (a) $f(x, y) = x^2 + y^2$, $P = (3, 4, 25)$
   (b) $f(x, y) = x/(x^2 + y^2)^{1/2}$, $P = (3, -4, \frac{3}{5})$
   (c) $f(x, y) = \sin(xy)$ at $P = (1, \pi, 0)$
6. Find the equation of the tangent plane to the surface $x = e^{2y-z}$ at $(1, 1, 2)$.

§3. *Directional derivative*

Let f be defined on an open set and assume that f is differentiable. Let P be a point of the open set, and let A be a unit vector (i.e. $\|A\| = 1$). Then $P + tA$ is the parametric equation of a straight line in the direction of A and passing through P. We observe that

$$\frac{d(P + tA)}{dt} = A.$$

For instance, if $n = 2$ and $P = (p, q)$, $A = (a, b)$, then

$$P + tA = (p + ta, q + tb),$$

or in terms of coordinates,

$$x = p + ta, \qquad y = q + tb.$$

Hence

$$\frac{dx}{dt} = a \quad \text{and} \quad \frac{dy}{dt} = b$$

so that

$$\frac{d(P + tA)}{dt} = (a, b) = A.$$

The same argument works in higher dimensions.

Hence by the chain rule, if we take the derivative of the function $t \mapsto f(P + tA)$, which is defined for small values of t, we obtain

$$\frac{df(P + tA)}{dt} = \operatorname{grad} f(P + tA) \cdot A.$$

When t is equal to 0, this derivative is equal to

$$\operatorname{grad} f(P) \cdot A.$$

For obvious geometrical reasons, we call it the **directional derivative** of f in the direction of A. We interpret it as the rate of change of f along the straight line in the direction of A, at the point P. Thus if we agree on the notation $D_A f(P)$ for the directional derivative of f at P in the direction of the unit vector A, then we have

$$\boxed{D_A f(P) = \left.\frac{df(P + tA)}{dt}\right|_{t=0} = \operatorname{grad} f(P) \cdot A.}$$

In using this formula, the reader should remember that A is taken to be a **unit vector.** When a direction is given in terms of a vector whose length is not 1, then one must first divide this vector by its length before applying the formula.

Example. Let $f(x, y) = x^2 + y^3$ and let $B = (1, 2)$. Find the directional derivative of f in the direction of B, at the point $(-1, 3)$.

We note that B is not a unit vector. Its length is $\sqrt{5}$. Let

$$A = \frac{1}{\sqrt{5}} B.$$

Then A is a unit vector having the same direction as B. Let $P = (-1, 3)$.

Then $\operatorname{grad} f(P) = (-2, 27)$. Hence by our formula, the directional derivative is equal to:

$$\operatorname{grad} f(P) \cdot A = \frac{1}{\sqrt{5}}(-2 + 54) = \frac{52}{\sqrt{5}}.$$

Consider again a differentiable function f on an open set U.

Let P be a point of U. *Let us assume that* $\operatorname{grad} f(P) \neq O$, and let A be a unit vector. We know that

$$\operatorname{grad} f(P) \cdot A = \|\operatorname{grad} f(P)\| \, \|A\| \cos \theta,$$

where θ is the angle between $\operatorname{grad} f(P)$ and A. Since $\|A\| = 1$, we see that the directional derivative is equal to $\|\operatorname{grad} f(P)\| \cos \theta$. The value of $\cos \theta$ varies between -1 and $+1$ when we select all possible unit vectors A.

The maximal value of $\cos \theta$ is obtained when we select A such that $\theta = 0$, i.e. when we select A to have the same direction as $\operatorname{grad} f(P)$. In that case, the directional derivative is equal to the length of the gradient.

Thus we have obtained another interpretation for the gradient:

Its direction is that of maximal increase of the function, and its length is the rate of increase of the function in that direction.

The directional derivative in the direction of A is at a minimum when $\cos \theta = -1$. This is the case when we select A to have opposite direction to $\operatorname{grad} f(P)$. That direction is therefore the direction of maximal decrease of the function.

For example, f might represent a temperature distribution in space. At any point P, a particle which feels cold and wants to become warmer fastest should move in the direction of $\operatorname{grad} f(P)$. Another particle which is warm and wants to cool down fastest should move in the direction of $-\operatorname{grad} f(P)$.

Exercises

1. Let $f(x, y, z) = z - e^x \sin y$, and $P = (\log 3, 3\pi/2, -3)$. Find:
   (a) the directional derivative of f at P in the direction of $(1, 2, 2)$,
   (b) the maximum and minimum values for the directional derivatives of f at P.
2. Find the directional derivatives of the following functions at the specified points in the specified directions.
   (a) $\log(x^2 + y^2)^{1/2}$ at $(1, 1)$, direction $(2, 1)$
   (b) $xy + yz + zx$ at $(-1, 1, 7)$, direction $(3, 4, -12)$
   (c) $4x^2 + 9y^2$ at $(2, 1)$ in the direction of maximum directional derivative
3. A temperature distribution in space is given by the function

$$f(x, y) = 10 + 6 \cos x \cos y + 3 \cos 2x + 4 \cos 3y.$$

At the point $(\pi/3, \pi/3)$, find the direction of greatest increase of temperature, and the direction of greatest decrease of temperature.

4. In what direction are the following functions of X increasing most rapidly at the given point?

   (a) $x/\|X\|^{3/2}$ at $(1, -1, 2)$ $\quad (X = (x, y, z))$
   (b) $\|X\|^5$ at $(1, 2, -1, 1)$ $\quad (X = (x, y, z, w))$

5. Find the tangent plane to the surface $x^2 + y^2 - z^2 = 18$ at the point $(3, 5, -4)$.

6. Let $f(x, y, z) = (x + y)^2 + (y + z)^2 + (z + x)^2$. What is the direction of greatest increase of the function at the point $(2, -1, 2)$. What is the directional derivative of f in this direction at that point?

7. Let $f(x, y) = x^2 + xy + y^2$. What is the direction in which f is increasing most rapidly at the point $(-1, 1)$? Find the directional derivative of f in this direction.

§4. *Conservation law*

As a final application of the chain rule, we derive the conservation law of physics.

Let U be an open set. By a **vector field** on U we mean an association which to every point of U associates a vector of the same dimension.

If f is a differentiable function on U, then we observe that $\operatorname{grad} f$ is a vector field, which associates the vector $\operatorname{grad} f(P)$ to the point P of U.

A vector field in physics is often interpreted as a field of forces.

If F is a vector field on U, and X a point of U, then we denote by $F(X)$ the vector associated to X by F and call it the value of F at X, as usual.

If F is a vector field, and if there exists a differentiable function f such that $F = \operatorname{grad} f$, then the vector field is called **conservative.** Since $-\operatorname{grad} f = \operatorname{grad}(-f)$, it does not matter whether we use f or $-f$ in the definition of conservative.

Let us assume that F is a conservative field on U, and let Φ be a differentiable function such that for all points X in U we have

$$F(X) = -\operatorname{grad} \Phi.$$

In physics, one interprets Φ as a potential function. Suppose that a particle of mass m moves along a differentiable curve $X(t)$ in U, and let us assume that this particle obeys Newton's law:

$$F(X) = mX'', \qquad \text{that is} \qquad F(X(t)) = mX''(t)$$

for all t where $X(t)$ is defined. Then according to our hypotheses,

$$mX'' + \operatorname{grad} \Phi(X) = O.$$

Take the dot product of both sides with X'. We obtain

$$mX' \cdot X'' + \operatorname{grad} \Phi(X) \cdot X' = 0.$$

But the derivative (with respect to t) of X'^2 is $2X' \cdot X''$. The derivative with respect to t of $\Phi(X(t))$ is equal to

$$\operatorname{grad} \Phi\,(X) \cdot X'$$

by the chain rule. Hence the expression on the left of our last equation is the derivative of the **function**

$$\tfrac{1}{2}mX'^2 + \Phi(X),$$

and that derivative is 0. Hence this function is equal to a constant. This is what one means by the conservation law.

The function $\frac{1}{2}mX'^2$ is called the **kinetic energy**, and the conservation law states that the sum of the kinetic and potential energies is constant.

It is not true that all vector fields are conservative. We shall discuss the problem of determining which ones are conservative in the next chapter.

The fields of classical physics are for the most part conservative. For instance, consider a force which is inversely proportional to the square of the distance from the point to the origin, and in the direction of the position vector. Then there is a constant C such that for $X \neq O$ we have

$$F(X) = C \frac{1}{\|X\|^2} \frac{X}{\|X\|},$$

because $\dfrac{X}{\|X\|}$ is a unit vector in the direction of X. Thus

$$F(X) = C \frac{1}{r^3} X,$$

where $r = \|X\|$. A potential function for F is given by

$$-\frac{C}{r}.$$

This is immediately verified by taking the partial derivatives of this function.

Exercises

1. Find a potential function for a force field which is inversely proportional to the distance from the point to the origin, and is in the direction of the position vector.
2. Same question, replacing "distance" with "cube of the distance".
3. Let k be an integer $\geqq 1$. Find a potential function for the vector field F given by

$$F(X) = \frac{1}{r^k} X, \qquad \text{where } r = \|X\|.$$

[*Hint:* Cf. Exercise 9(f) of §1.]

CHAPTER V

Potential Functions and Curve Integrals

We are going to deal systematically with the possibility of finding a potential function for a vector field. The discussion of the existence of such a function will be limited to the case of two variables. Actually, there is no essential difficulty in extending the results to arbitrary n-space, but we leave this to the reader. (Cf. the answer section.)

The problem is one of integration, and the line integrals are a natural continuation of the integrals at the end of §1 (taken on vertical and horizontal lines).

§1. *Potential functions*

Let F be a vector field on an open set U. If φ is a differentiable function on U such that $F = \text{grad}\,\varphi$, then we say that φ is a **potential function** for F. (Or, in hip terminology, a **pot** function.)

One can raise two questions about potential functions. Are they unique, and do they exist?

We consider the first question, and we shall be able to give a satisfactory answer to it. The problem is analogous to determining an integral for a function of one variable, up to a constant, and we shall formulate and prove the analogous statement in the present situation.

We recall that even in the case of functions of one variable, it is *not* true that whenever two functions f, g are such that

$$\frac{df}{dx} = \frac{dg}{dx},$$

then f and g differ by a constant, unless we assume that f, g are defined on some interval. As we emphasized in the *First Course*, we could for instance take

$$f(x) = \begin{cases} \dfrac{1}{x} + 5 & \text{if} \quad x < 0, \\[2mm] \dfrac{1}{x} - \pi & \text{if} \quad x > 0, \end{cases}$$

$$g(x) = \frac{1}{x} \quad \text{if} \quad x \neq 0.$$

Then f, g have the same derivative, but there is no constant C such that for all $x \neq 0$ we have $f(x) = g(x) + C$.

In the case of functions of several variables, we shall have to make a similar restriction on the domain of definition of the functions.

Let U be an open set and let P, Q be two points of U. We shall say that P, Q can be joined by a **differentiable curve** if there exists a differentiable curve $X(t)$ (with t ranging over some interval of numbers) which is contained in U, and two values of t, say t_1 and t_2 in that interval, such that

$$X(t_1) = P \qquad \text{and} \qquad X(t_2) = Q.$$

For example, if U is the entire plane, then any two points can be joined by a straight line. In fact, if P, Q are two points, then we take

$$X(t) = P + t(Q - P).$$

When $t = 0$, then $X(0) = P$. When $t = 1$, then $X(1) = Q$.

It is not always the case that two points of an open set can be joined by a straight line. We have drawn a picture of two points P, Q in an open set U which cannot be so joined.

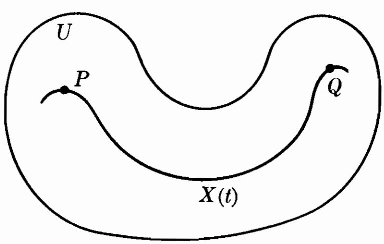


Figure 1

We are now in position to state the theorem we had in mind.

Theorem 1. *Let U be an open set, and assume that any two points in U can be joined by a differentiable curve. Let f, g be two differentiable functions on U. If* $\operatorname{grad} f(X) = \operatorname{grad} g(X)$ *for every point X of U, then there exists a constant C such that*

$$f(X) = g(X) + C$$

for all points X of U.

Proof. We note that $\operatorname{grad}(f - g) = \operatorname{grad} f - \operatorname{grad} g = O$, and we must prove that $f - g$ is constant. Letting $\varphi = f - g$, we see that it suffices to prove: If $\operatorname{grad} \varphi(X) = O$ for every point X of U, then φ is constant.

Let P be a fixed point of U and let Q be any other point. Let $X(t)$ be a differentiable curve joining P to Q, which is contained in U, and defined

over an interval. The derivative of the function $\varphi(X(t))$ is, by the chain rule,

$$\operatorname{grad}\varphi\,(X(t))\cdot X'(t).$$

But $X(t)$ is a point of U for all values of t in the interval. Hence by our assumption, the derivative of $\varphi(X(t))$ is 0 for all t in the interval. Hence there is a constant C such that

$$\varphi(X(t)) = C$$

for all t in the interval. In other words, the function φ is constant on the curve. Hence $\varphi(P) = \varphi(Q)$.

This result is true for any point Q of U. Hence φ is constant on U, as was to be shown.

Our theorem proves the uniqueness of potential functions (within the restrictions placed by our extra hypothesis on the open set U).

We still have the problem of determining when a vector field F admits a potential function.

We first make some remarks in the case of functions of two variables. Let F be a vector field (in 2-space), so that we can write

$$F(x, y) = (f(x, y), g(x, y))$$

with functions f and g, defined over a suitable open set. We want to know when there exists a function $\varphi(x, y)$ such that

$$\frac{\partial\varphi}{\partial x} = f \quad \text{and} \quad \frac{\partial\varphi}{\partial y} = g.$$

Such a function would be a potential function for F, by definition. (We assume throughout that all hypotheses of differentiability are satisfied as needed.)

Suppose that such a function φ exists. Then

$$\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}\left(\frac{\partial\varphi}{\partial x}\right) \quad \text{and} \quad \frac{\partial g}{\partial x} = \frac{\partial}{\partial x}\left(\frac{\partial\varphi}{\partial y}\right).$$

We shall show in the next chapter that under suitable hypotheses, the two partial derivatives on the right are equal. This means that if there exists a potential function for F, then

$$\frac{\partial f}{\partial y} = \frac{\partial g}{\partial x}.$$

This gives us a simple test in practice to tell whether a potential function may exist.

Theorem 2. *Let f, g be differentiable functions having continuous partial derivatives on an open set U in 2-space. If*

$$\frac{\partial f}{\partial y} \neq \frac{\partial g}{\partial x}$$

then the vector field given by $F(x, y) = (f(x, y), g(x, y))$ does not have a potential function.

Example. Consider the vector field given by

$$F(x, y) = (x^2y, \sin xy).$$

Then we let $f(x, y) = x^2y$ and $g(x, y) = \sin xy$. We have:

$$\frac{\partial f}{\partial y} = x^2 \quad \text{and} \quad \frac{\partial g}{\partial x} = y \cos xy.$$

Since $\partial f/\partial y \neq \partial g/\partial x$, it follows that the vector field does not have a potential function.

We shall prove in §3 that the converse of Theorem 2 is true in some very important cases. Before stating and proving the pertinent theorem, we first discuss an auxiliary situation.

Exercises

Determine which of the following vector fields have potential functions. The vector fields are described by the functions $(f(x, y), g(x, y))$.

1. $(1/x, xe^{xy})$
2. $(\sin(xy), \cos(xy))$
3. (e^{xy}, e^{x+y})
4. $(3x^4y^2, x^3y)$
5. $(5x^4y, x\cos(xy))$
6. $\left(\dfrac{x}{\sqrt{x^2 + y^2}}, 3xy^2\right)$

§2. *Differentiating under the integral*

Let f be a continuous function on a rectangle $a \leqq x \leqq b$ and $c \leqq y \leqq d$. We can then form a function of y by taking

$$\psi(y) = \int_a^b f(x, y)\, dx.$$

Example 1. We can determine explicitly the function ψ if we let $f(x, y) = \sin(xy)$, namely:

$$\psi(y) = \int_0^\pi \sin(xy)\, dx = -\left.\frac{\cos(xy)}{y}\right|_{x=0}^{x=\pi} = -\frac{\cos(\pi y) - 1}{y}.$$

We are interested in finding the derivative of ψ. The next theorem allows us to do this in certain cases, by differentiating with respect to y under the integral sign.

Theorem 3. *Assume that f is continuous on the preceding rectangle, and that $D_2 f$ exists and is continuous. Let*

$$\psi(y) = \int_a^b f(x, y)\, dx.$$

Then ψ is differentiable, and

$$\frac{d\psi}{dy} = D\psi(y) = \int_a^b D_2 f(x, y)\, dx = \int_a^b \frac{\partial f(x, y)}{\partial y}\, dx.$$

Proof. By definition, we have to investigate the Newton quotient for ψ. We have

$$\frac{\psi(y + h) - \psi(y)}{h} = \int_a^b \left[\frac{f(x, y + h) - f(x, y)}{h}\right] dx.$$

We then have to find

$$\lim_{h \to 0} \int_a^b \frac{f(x, y + h) - f(x, y)}{h}\, dx.$$

If we knew that we can put the limit sign inside the integral, we would then conclude that the preceding limit is equal to

$$\int_a^b \lim_{h \to 0} \frac{f(x, y + h) - f(x, y)}{h}\, dx = \int_a^b D_2 f(x, y)\, dx,$$

thus proving our theorem. We shall not give the argument which justifies moving the limit sign inside the integral, because it depends on (ϵ, δ) considerations which are mostly omitted in this book.

Example 2. Letting $f(x, y) = \sin(xy)$ as before, we find that

$$D_2 f(x, y) = x \cos(xy).$$

If we let

$$\psi(y) = \int_0^\pi f(x, y)\, dx,$$

then

$$D\psi(y) = \int_0^\pi D_2 f(x, y)\, dx = \int_0^\pi x \cos(xy)\, dx.$$

By evaluating this last integral, or by differentiating the expression found for ψ at the beginning of the section, the reader will find the same value, namely

$$D\psi(y) = -\left[\frac{\pi y \sin(\pi y) - \cos(\pi y)}{y^2} + \frac{1}{y^2}\right].$$

We can apply the previous theorem using any x as upper limit of the integration. Thus we may let

$$\psi(x, y) = \int_a^x f(t, y)\,dt,$$

in which case the theorem reads

$$\frac{\partial \psi}{\partial y} = D_2\psi(x, y) = \int_a^x D_2 f(t, y)\,dt = \int_a^x \frac{\partial f(t, y)}{\partial y}\,dt.$$

We use t as a variable of integration to distinguish it from the x which is now used as an end point of the interval $[a, x]$ instead of $[a, b]$.

The preceding way of determining the derivative of ψ with respect to y is called **differentiating under the integral sign.** *Note that it is completely different from the differentiation in the fundamental theorem of calculus.* In this case, we have an integral

$$g(x) = \int_a^x f(t)\,dt,$$

and

$$\frac{dg}{dx} = Dg(x) = f(x).$$

Thus when f is a function of two variables, and ψ is defined as above, the fundamental theorem of calculus states that

$$\frac{\partial \psi}{\partial x} = D_1\psi(x, y) = f(x, y).$$

For example, if we let

$$\psi(x, y) = \int_0^x \sin(ty)\,dt,$$

then

$$D_1\psi(x, y) = \sin(xy),$$

but by Theorem 3,

$$D_2\psi(x, y) = \int_0^x \cos(ty)t\,dt.$$

Exercises

In each of the following cases, find $D_1\psi(x, y)$ and $D_2\psi(x, y)$, by evaluating the integrals.

1. $\psi(x, y) = \displaystyle\int_1^x e^{ty}\,dt$

2. $\psi(x, y) = \displaystyle\int_0^x \cos(ty)\,dt$

3. $\psi(x, y) = \displaystyle\int_1^x (y + t)^2\,dt$

4. $\psi(x, y) = \displaystyle\int_1^x e^{y+t}\,dt$

5. $\psi(x, y) = \int_1^x e^{y-t}\,dt$

6. $\psi(x, y) = \int_0^x t^2y^3\,dt$

7. $\psi(x, y) = \int_1^x \frac{\log(ty)}{t}\,dt$

8. $\psi(x, y) = \int_1^x \sin(3ty)\,dt$

§3. Local existence of potential functions

We shall state a theorem which will give us conditions under which the converse of Theorem 2 is true.

Theorem 4. *Let f, g be differentiable functions on an open set of the plane. If this open set is the entire plane, or if it is an open disc, or the inside of a rectangle, if the partial derivatives of f, g exist and are continuous, and if*

$$\frac{\partial f}{\partial y} = \frac{\partial g}{\partial x},$$

then the vector field $F(x, y) = (f(x, y), g(x, y))$ has a potential function.

We shall indicate how a proof of Theorem 4 might go for a rectangle after we have discussed some examples.

Example 1. Determine whether the vector field F given by

$$F(x, y) = (e^{xy}, e^{x+y})$$

has a potential function.

Here, $f(x, y) = e^{xy}$ and $g(x, y) = e^{x+y}$. We have:

$$\frac{\partial f}{\partial y} = xe^{xy} \quad \text{and} \quad \frac{\partial g}{\partial x} = e^{x+y}.$$

Since these are not equal, we know that there cannot be a potential function.

If the partial derivatives $\partial f/\partial y$ and $\partial g/\partial x$ turn out to be equal, then one can try to find a potential function by integrating with respect to one of the variables. Thus we try to find

$$\int f(x, y)\,dx,$$

keeping y constant, and taking the ordinary integral of functions of one variable. If we can find such an integral, it will be a function $\psi(x, y)$, whose partial with respect to x will be equal to $f(x, y)$ (by definition). Adding a function of y, we can then adjust it so that its partial with respect to y is equal to $g(x, y)$.

Example 2. Let $F(x, y) = (2xy, x^2 + 3y^2)$. Determine whether this vector field has a potential function, and if it does, find it.

Applying the test which we mentioned above, we find that a potential function may exist. To find it, we consider first the integral

$$\int 2xy\,dx,$$

viewing y as constant. We obtain x^2y for the indefinite integral. We must now find a function $u(y)$ such that

$$\frac{\partial}{\partial y}(x^2y + u(y)) = x^2 + 3y^2.$$

This means that we must find a function $u(y)$ such that

$$x^2 + \frac{du}{dy} = x^2 + 3y^2,$$

or in other words,

$$\frac{du}{dy} = 3y^2.$$

This is a simple integration problem in one variable, and we find $u(y) = y^3$. Thus finally, if we let

$$\varphi(x, y) = x^2y + y^3,$$

we see that φ is a potential function for F.

Proof of Theorem 4. We let the rectangle be defined by

$$a \leqq x \leqq b \quad \text{and} \quad c \leqq y \leqq d.$$

We let

$$\varphi(x, y) = \int_a^x f(t, y)\,dt + \int_c^y g(a, u)\,du.$$

Then the second integral on the right does not depend on x, and by the fundamental theorem of calculus,

$$D_1\varphi(x, y) = f(x, y)$$

as wanted. On the other hand, using Theorem 3, and differentiating with respect to y, we get:

$$\begin{aligned}
D_2\varphi(x, y) &= \int_a^x D_2f(t, y)\,dt + g(a, y)\\
&= \int_a^x D_1g(t, y)\,dt + g(a, y)\\
&= g(t, y)\Big|_{t=a}^{t=x} + g(a, y)\\
&= g(x, y) - g(a, y) + g(a, y)\\
&= g(x, y)
\end{aligned}$$

thus yielding the desired expression for the second partial of φ. This proves Theorem 4.

Note that the proof is entirely similar to that of the example preceding it. The first integral with respect to x solves the requirements or the first partial of φ, and we correct it by an integral involving only y in order to adjust the answer to give the desired partial with respect to y.

Exercises

Determine which of the following vector fields admit potential functions.

1. $(e^x, \sin xy)$
2. $(2x^2y, y^3)$
3. $(2xy, y^2)$
4. $(y^2x^2, x + y^4)$

Find potential functions for the following vector fields.

5. (a) $F(X) = \dfrac{1}{r} X$ (b) $F(X) = \dfrac{1}{r^2} X$
   (c) $F(X) = r^n X$ (if n is an integer $\neq -2$). In this Exercise,
   $$r = \|X\|, \quad \text{and} \quad X \neq O.$$
6. $(4xy, 2x^2)$
7. $(xy \cos xy + \sin xy, x^2 \cos xy)$
8. $(3x^2y^2, 2x^3y)$
9. $(2x, 4y^3)$
10. (a) (ye^{xy}, xe^{xy}) (b) $(y \cos xy, x \cos xy)$
11. Let $r = \|X\|$. Let g be a differentiable function of one variable. Show that the vector field defined by
   $$F(X) = \frac{g'(r)}{r} X$$
   in the domain $X \neq O$ always admits a potential function. What is this potential function?
12. Generalize Theorem 4 to functions of three variables. Assume that we have a vector field $F = (f_1, f_2, f_3)$ defined on a 3-dimensional rectangle
   $$[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3],$$
   satisfying the conditions
   $$\frac{\partial f_i}{\partial x_j} = \frac{\partial f_j}{\partial x_i} \qquad (\text{or } D_j f_i = D_i f_j)$$
   for all indices i, j. Prove that this vector field has a potential function. Extend this to more than 3 variables.
13. Find a potential function f for the following vector fields F given as $F(x, y, z)$.
   (a) $(2x, 3y, 4z)$ (b) $(y + z, x + z, x + y)$
   (c) $(e^{y+2z}, xe^{y+2z}, 2xe^{y+2z})$ (d) $(y \sin z, x \sin z, xy \cos z)$
   (e) $(yz, xz + z^3, xy + 3yz^2)$ (f) $(e^{yz}, xze^{yz}, xye^{yz})$
   (g) $(z^2, 2y, 2xz)$ (h) $(yz \cos xy, xz \cos xy, \sin xy)$

§4. *Curve integrals*

Let U be an open set (of n-space), and let F be a vector field on U. We can represent F by components:

$$F(X) = (f_1(X), \ldots, f_n(X)),$$

each f_i being a function. When $n = 2$,

$$F(X) = (f(x, y), g(x, y)).$$

If each function $f_1(X), \ldots, f_n(X)$ is continuous, then we shall say that F is a **continuous** vector field. If each function $f_1(X), \ldots, f_n(X)$ is differentiable, then we shall say that F is a **differentiable** vector field.

We shall also deal with curves. Rather than use the letter X to denote a curve, we shall use another letter, for instance C, to avoid certain confusions which might arise in the present context. Furthermore, it is now convenient to assume that our curve C is defined on a **closed** interval $I = [a, b]$, with $a < b$. For each number t in I, the value $C(t)$ is a point in n-space. We shall say that the curve C lies in U if $C(t)$ is a point of U for all t in I. We say that C is **continuously differentiable** if its derivative $C'(t) = dC/dt$ exists and is continuous. We abbreviate the expression "continuously differentiable" by saying that the curve is a C^1-curve, or of class C^1.

Let F be a continuous vector field on U, and let C be a continuously differentiable curve in U. The dot product

$$F(C(t)) \cdot \frac{dC}{dt}$$

is a **function** of t, and it can be shown easily that this function is continuous (by ϵ and δ techniques which we always omit).

Example 1. Let $F(x, y) = (e^{xy}, y^2)$, and $C(t) = (t, \sin t)$. Then

$$C'(t) = (1, \cos t)$$

and

$$F(C(t)) = (e^{t \sin t}, \sin^2 t).$$

Hence

$$F(C(t)) \cdot C'(t) = e^{t \sin t} + (\cos t)(\sin^2 t).$$

Suppose that C is defined on the interval $[a, b]$. We define the **integral of F along C** to be

$$\int_C F = \int_a^b F(C(t)) \cdot \frac{dC}{dt}\, dt.$$

This integral is a direct generalization of the familiar notion of the integral of functions of one variable. If we are given a function $f(u)$, and u is a function of t, then

$$\int_{u(a)}^{u(b)} f(u)\,du = \int_a^b f\big(u(t)\big)\frac{du}{dt}\,dt.$$

(This is the formula describing the substitution method for evaluating integrals.)

In n-space, $C(a)$ and $C(b)$ are points, and our curve passes through these two points. Thus the integral we have written down can be interpreted as an integral of the vector field, along the curve, between the two points. It will also be convenient to write the integral in the form

$$\int_{P,C}^{Q} F = \int_{C(a)}^{C(b)} F(C)\cdot dC$$

to denote the integral along the curve C, from P to Q.

Example 2. Let $F(x, y) = (x^2y, y^3)$. Find the integral of F along the straight line from the origin to the point $(1, 1)$.

We can parametrize the line in the form

$$C(t) = (t, t).$$

Thus

$$F\big(C(t)\big) = (t^3, t^3).$$

Furthermore,

$$\frac{dC}{dt} = (1, 1).$$

Hence

$$F\big(C(t)\big)\cdot\frac{dC}{dt} = 2t^3.$$

The integral we must find is therefore equal to:

$$\int_C F = \int_0^1 2t^3\,dt = \frac{2t^4}{4}\bigg|_0^1 = \frac{1}{2}\cdot$$

Remark 1. Our integral of a vector field along a curve is defined for **parametrized curves.** In practice, a curve is sometimes given in a non-parametrized way. For instance, we may want to integrate over the curve defined by $y = x^2$. Then we select some parametrization which is usually the most natural, in this case

$$x = t, \qquad y = t^2.$$

In general, if a curve is defined by a function $y = g(x)$, we select the

parametrization

$$x = t, \qquad y = g(t).$$

For a circle of radius R centered at the origin, we select the parametrization

$$x = R\cos t, \qquad y = R\sin t, \qquad 0 \leqq t \leqq 2\pi.$$

whenever we wish to integrate counterclockwise.

For a straight line segment between two points P and Q, we take the parametrization C given by

$$C(t) = P + t(Q - P), \qquad 0 \leqq t \leqq 1.$$

The context should always make it clear which parametrization is intended.

Remark 2. We may be given a finite number of C^1-curves forming a path as indicated in the following figure:

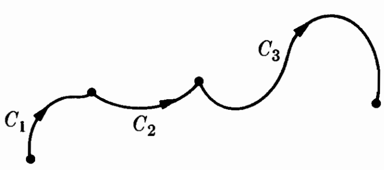


Figure 2

Thus formally, we define a (piecewise C^1) **path** C to be a finite sequence $\{C_1, \ldots, C_m\}$, where each C_i is a C^1-curve, defined on an interval $[a_i, b_i]$, such that the end point of C_i is the beginning point of C_{i+1}. Thus if $P_i = C_i(a_i)$ and $Q_i = C_i(b_i)$, then

$$Q_i = P_{i+1}.$$

We define the integral of F along such a path C to be the sum

$$\int_C F = \int_{C_1} F + \int_{C_2} F + \cdots + \int_{C_m} F.$$

We say that the path C is a **closed path** if the end point of C_m is the beginning point of C_1.

In the following picture, we have drawn a closed path such that the beginning point of C_1, namely P_1, is the end point of the path C_4, which joins P_4 to P_1.

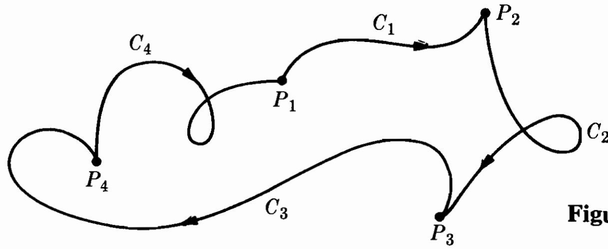


Figure 3

Example 3. Let $F(x, y) = (x^2, xy)$ and let the path consist of the segment of the parabola $y = x^2$ between $(0, 0)$ and $(1, 1)$, and the line segment from $(1, 1)$ and $(0, 0)$. (Cf. Fig. 4.)

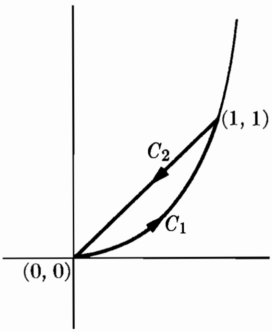


Figure 4

Then we let $C_1(t) = (t, t^2)$ and $C_2(t) = (1 - t, 1 - t)$. We let $C = \{C_1, C_2\}$. To find the integral of F along C we find the integral along C_1 and C_2, and add these integrals. We get:

$$\int_{C_1} F = \int_0^1 F(C_1(t)) \cdot (1, 2t)\, dt = \int_0^1 (t^2 + 2t^4)\, dt = \tfrac{1}{3} + \tfrac{2}{5}.$$

$$\int_{C_2} F = \int_0^1 F(C_2(t)) \cdot (-1, -1)\, dt = \int_0^1 -2(1 - 2t + t^2)\, dt = -\tfrac{2}{3}.$$

Hence

$$\int_C F = -\tfrac{1}{3} + \tfrac{2}{5}.$$

When the vector field F admits a potential function φ, then the integral of F along a curve has a simple expression in terms of φ.

Theorem 5. *Let F be a continuous vector field on the open set U and assume that $F =$ grad φ for some differentiable function φ on U. Let*

C be a C^1-curve in U, joining the points P and Q. Then

$$\int_{P,C}^{Q} F = \varphi(Q) - \varphi(P).$$

In particular, the integral of F is independent of the curve C joining P and Q.

Proof. Let C be defined on the interval $[a, b]$, so that $C(a) = P$ and $C(b) = Q$. By definition, we have

$$\int_{P,C}^{Q} F = \int_a^b F(C(t)) \cdot C'(t)\, dt = \int_a^b \operatorname{grad} \varphi\, (C(t)) \cdot C'(t)\, dt.$$

But the expression inside the integral is nothing but the derivative with respect to t of the function g given by $g(t) = \varphi(C(t))$, because of the chain rule. Thus our integral is equal to

$$\int_a^b g'(t)\, dt = g(b) - g(a) = \varphi(C(b)) - \varphi(C(a)).$$

This proves our theorem.

This theorem is easily extended to paths. We leave this to the reader.

We observe that in physics, one may interpret a vector field F as describing a force. Then the integral of this vector field along a path C describes the **work** done by the force along this path. In particular, when the vector field is conservative, as in Theorem 5, the work is expressed in terms of the potential function for F, and the end points of the path.

Example 4. Let $F(X) = kX/r^3$, where $r = \|X\|$, and k is a constant. This is the vector field inversely proportional to the square of the distance from the origin, used so often in physics. Then F has a potential function, namely the function φ such that $\varphi(X) = -k/r$. Thus the integral of F from $P = (1, 1, 1)$ to $Q = (1, 2, -1)$ is simply equal to

$$\varphi(Q) - \varphi(P) = -k\left(\frac{1}{\|Q\|} - \frac{1}{\|P\|}\right) = -k\left(\frac{1}{\sqrt{6}} - \frac{1}{\sqrt{3}}\right).$$

On the other hand, if P_1, Q_1 are two points at the same distance from the origin (i.e. lying on the same circle, centered at the origin), then the integral of F from P_1 to Q_1 along any curve is equal to 0.

Example 5. Let C be a closed curve, whose end point is equal to the beginning point P. In Theorem 5, when the vector field F admits a potential function φ, it follows that the integral of F over the closed curve is then equal to 0, because it is equal to

$$\varphi(P) - \varphi(P) = 0.$$

This allows us to give an example for a situation when a vector field $F = (f, g)$ satisfies the condition

$$\frac{\partial f}{\partial y} = \frac{\partial g}{\partial x}$$

but F does *not* have a potential function. Let

$$F(x, y) = \left(\frac{-y}{x^2 + y^2}, \frac{x}{x^2 + y^2}\right)$$

be the vector field of Exercise 11. A simple computation, left as an exercise, shows that it satisfies the above condition. Compute the integral of F over the closed circle of radius 1, centered at the origin. You will find a value $\neq 0$. This does *not* contradict Theorem 4, because the vector field is defined on the open set obtained from the plane by deleting the origin, i.e. the vector field is not defined at $(0, 0)$. The open set has a "hole" in it (a pinhole, in fact).

Exercises

Compute the curve integrals of the vector field over the indicated curves.

1. $F(x, y) = (x^2 - 2xy, y^2 - 2xy)$ along the parabola $y = x^2$ from $(-2, 4)$ to $(1, 1)$.
2. $(x, y, xz - y)$ over the line segment from $(0, 0, 0)$ to $(1, 2, 4)$.
3. Let $r = (x^2 + y^2)^{1/2}$. Let $F(X) = r^{-1}X$. Find the integral of F over the circle of radius 2, taken in counterclockwise direction.
4. Let C be a circle of radius 20 with center at the origin. Let F be a vector field such that $F(X)$ has the same direction as X. What is the integral of F around C?
5. What is the work done by the force $F(x, y) = (x^2 - y^2, 2xy)$ moving a particle of mass m along the square bounded by the coordinate axes and the lines $x = 3$, $y = 3$ in counterclockwise direction?
6. Let $F(x, y) = (cxy, x^6y^2)$, where c is a positive constant. Let a, b be numbers > 0. Find a value of a in terms of c such that the line integral of F along the curve $y = ax^b$ from $(0, 0)$ to the line $x = 1$ is independent of b.

Find the values of the indicated integrals of vector fields along the given curves in Exercises 7 through 13.

7. $(y^2, -x)$ along the parabola $x = y^2/4$ from $(0, 0)$ to $(1, 2)$.
8. $(x^2 - y^2, x)$ along the arc in the first quadrant of the circle $x^2 + y^2 = 4$ from $(0, 2)$ to $(2, 0)$.
9. (x^2y^2, xy^2) along the closed path formed by parts of the line $x = 1$ and the parabola $y^2 = x$, counterclockwise.

10. $(x^2 - y^2, x)$ counterclockwise around the circle $x^2 + y^2 = 4$.

11. (a) The vector field
$$\left(\frac{-y}{x^2 + y^2}, \frac{x}{x^2 + y^2}\right)$$
counterclockwise along the circle $x^2 + y^2 = 2$ from $(1, 1)$ to $(-\sqrt{2}, 0)$.
(b) The same vector field counterclockwise around the whole circle.
(c) Around the circle $x^2 + y^2 = 1$.
(d) Around the circle $x^2 + y^2 = r^2$.
(e) Verify that for this vector field, we have $\partial f/\partial y = \partial g/\partial x$. For a continuation of this train of thought, see Green's theorem.

12. The same vector field along the line $x + y = 1$ from $(0, 1)$ to $(1, 0)$.

13. $(2xy, -3xy)$ clockwise around the square bounded by the lines $x = 3$, $x = 5$, $y = 1$, $y = 3$.

14. Let $C = (C_1, \ldots, C_m)$ be a piecewise C^1-path in an open set U. Let F be a continuous vector field on U, admitting a differentiable potential function φ. Let P be the beginning point of the path and Q its end point. Show that
$$\int_{P,C}^{Q} F = \varphi(Q) - \varphi(P).$$
[*Hint:* Apply Theorem 5 to the beginning point P_i and end point Q_i for each curve C_i.]

15. Find the integral of the vector field $F(x, y, z) = (2x, 3y, 4z)$ along the straight line $C(t) = (t, t, t)$ between the points $(0, 0, 0)$ and $(1, 1, 1)$.

16. Find the integral of the vector field $F(x, y, z) = (y + z, x + z, x + y)$ along the straight line $C(t) = (t, t, t)$ between $(0, 0, 0)$ and $(1, 1, 1)$.

17. Find the integral of the vector field given in Exercises 15 and 16 between the given points along the curve $C(t) = (t, t^2, t^4)$. Compare your answers with those previously found.

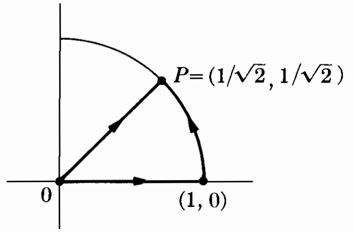
18. Let $F(x, y, z) = (y, x, 0)$. Find the integral of F along the straight line from $(1, 1, 1)$ to $(3, 3, 3)$.

19. Let P, Q be points of 3-space. Show that the integral of the vector field given by
$$F(x, y, z) = (z^2, 2y, 2xz)$$
from P to Q is independent of the curve selected between P and Q.

20. Let $F(x, y) = (x/r^3, y/r^3)$ where $r = (x^2 + y^2)^{1/2}$. Find the integral of F along the curve $C(t) = (e^t \cos t, e^t \sin t)$ from the point $(1, 0)$ to the point $(e^{2\pi}, 0)$.

21. Let $F(x, y, z) = (z^3y, z^3x, 3z^2xy)$. Show that the integral of F between two points is independent of the curve between the points.

22. Let $F(x, y) = (x^2y, xy^2)$.
(a) Does this vector field admit a potential function?
(b) Compute the integral of this vector field from O to the point P indicated on the figure, along the line segment from $(0, 0)$ to $(1/\sqrt{2}, 1/\sqrt{2})$.



(c) Compute the integral of this vector field from O to P along the path which consists of the segment from $(0, 0)$ to $(1, 0)$, and the arc of circle from $(1, 0)$ to P.

§5. Dependence of the integral on the path

By a path from now on, we mean a piecewise C^1-path, and all vector fields are assumed continuous.

Given two points P, Q in some open set U, and a vector field F on U, it may be that the integral of F along two paths from P to Q depends on the path. The main theorem of this section gives three equivalent conditions that this integral should be independent of the path. Before discussing this theorem, we describe what we mean by integrating along a curve in opposite direction.

Let $C: [a, b] \to \mathbf{R}^n$ be a curve. We define the **opposite curve** C^- (or the negative curve) by letting

$$C^-(t) = C(a + b - t).$$

Thus when $t = b$ we find that $C^-(b) = C(a)$, and when $t = a$ we find that $C^-(a) = C(b)$. As t increases from a to b, we see that $a + b - t$ decreases from b to a and thus we visualize C^- as going from $C(b)$ to $C(a)$ in reverse direction from C (Fig. 5).

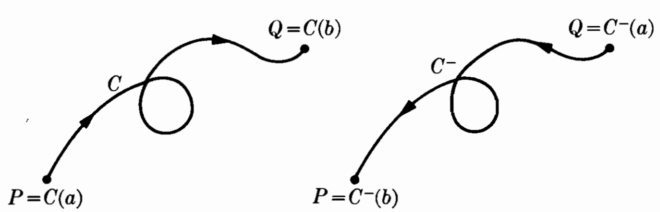


Figure 5

Lemma. *Let F be a vector field on the open set U, and let C be a curve in U, of class C^1, defined on the interval $[a, b]$. Then*

$$\int_{C^-} F = -\int_C F.$$

Proof. This is a simple application of the change of variables formula. Let $u = a + b - t$. Then $du/dt = -1$. By definition and the chain rule, we get:

$$\int_{C^-} F = \int_a^b F(C^-(t)) \cdot \frac{dC^-}{dt}\, dt$$
$$= \int_a^b F(C(a+b-t)) \cdot C'(a+b-t)(-1)\, dt.$$

We now change variables, with $du = -dt$. When $t = a$ then $u = b$, and when $t = b$ then $u = a$. Thus our integral is equal to

$$\int_b^a F(C(u)) \cdot C'(u)\, du = -\int_a^b F(C(u)) \cdot C'(u)\, du,$$

thereby proving the lemma.

The lemma expresses the expected result, that if we integrate the vector field along the opposite direction, then the value of the integral is the negative of the value obtained by integrating F along the curve itself.

Theorem 6. *Let U be an open set in $\mathbf{R}^n$ and let F be a vector field on U. Assume that any two points of U can be connected by a path in U. Then the following conditions are equivalent:*

(i) *The vector field F has a potential function.*
(ii) *The integral of F along any closed path in U is equal to* 0.
(iii) *If P, Q are two points in U then the integral of F from P to Q is independent of the path.*

Proof. Assume condition (i). Let $C = (C_1, \ldots, C_m)$ be a path in U where each C_i is a C^1-curve. Let P_i be the beginning point of C_i and let Q_i be its end point, so that $Q_i = P_{i+1}$. By Theorem 5, we find that

$$\int_C F = \varphi(Q_m) - \varphi(P_m) + \varphi(Q_{m-1}) - \varphi(P_{m-1}) + \cdots + \varphi(Q_1) - \varphi(P_1).$$

All intermediate terms cancel, leaving the first and the last terms, and our integral is equal to

$$\varphi(Q_m) - \varphi(P_1).$$

If the path is a closed path, then $Q_m = P_1$ and thus the integral is equal to 0. If $P_1 = P$ and $Q_m = Q$, then we see that the value of the integral is independent of the path; it depends only on P, Q and the potential function, namely $\varphi(Q) - \varphi(P)$. *Thus both conditions* (ii) *and* (iii) *follow from* (i).

Furthermore, condition (ii) *implies* (iii). Indeed, let C and D be paths from P to Q in U. Let $D = (D_1, \ldots, D_k)$ where each D_j is a C^1-curve.

Then we may form the opposite path

$$D^- = (D_k^-, \ldots, D_1^-),$$

and by the lemma,

$$\int_{D^-} F = -\int_D F.$$

If $C = (C_1, \ldots, C_m)$, then the path $(C_1, \ldots, C_m, D_k^-, \ldots, D_1^-)$ is a closed path from P to P (Fig. 6), and assuming (ii), we conclude that the integral of F along this closed path is equal to 0. Thus

$$\int_C F + \int_{D^-} F = 0.$$

From this it follows that

$$\int_C F = \int_D F,$$

whence condition (iii) holds.

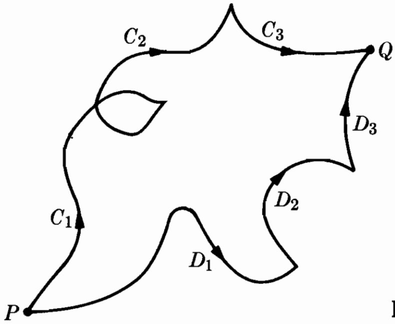


Figure 6

We shall now prove that condition (iii) implies (i). Let P_0 be a fixed point of U and define a function φ on U by the rule

$$\varphi(P) = \int_{P_0}^{P} F,$$

where the integral is taken along any path from P_0 to P. By assumption, this integral does not depend on the path, so we don't need to specify the path in the notation. We must show that the partial derivatives $D_i\varphi(P)$ exist for all P in U, and if the vector field F has coordinate functions

$$F = (f_1, \ldots, f_n),$$

then $D_i\varphi(P) = f_i(P)$.

To do this, let E_i be the unit vector with 1 in the i-th component and 0 in the other components. Then for any vector $X = (x_1, \ldots, x_n)$ we have $X \cdot E_i = x_i$. To determine $D_i\varphi(P)$ we must consider the Newton

quotient

$$\frac{\varphi(P + hE_i) - \varphi(P)}{h} = \frac{1}{h}\left[\int_{P_0}^{P+hE_i} F - \int_{P_0}^{P} F\right]$$

and show that its limit as $h \to 0$ is $f_i(P)$. The integral from P_0 to $P + hE_i$ can be taken along a path going first from P_0 to P and then from P to $P + hE_i$.

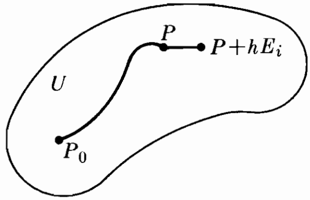


Figure 7

We can then cancel the integrals from P_0 to P and obtain

$$\frac{\varphi(P + hE_i) - \varphi(P)}{h} = \frac{\int_P^{P+hE_i} F(C) \cdot dC}{h},$$

taking the integral along any curve C between P and $P + hE_i$. In fact, we take C to be the parametrized straight line segment given by

$$C(t) = P + tE_i$$

with $0 \leqq t \leqq h$ in case h is positive. (The case of h negative is handled similarly. Cf. Exercise 1.) Then $C'(t) = E_i$ and

$$F(C(t)) \cdot C'(t) = f_i(C(t)).$$

The Newton quotient is therefore equal to

$$\frac{\int_0^h f_i(C(t))\, dt}{h}.$$

By the fundamental theorem of calculus, for any continuous function g we have (cf. Remark after the proof):

$$\lim_{h \to 0} \frac{1}{h} \int_0^h g(t)\, dt = g(0).$$

We apply this to the function given by $g(t) = f_i(C(t))$. Then

$$g(0) = f_i(C(0)) = f_i(P + 0E_i) = f_i(P).$$

Therefore we obtain the limit

$$\lim_{h \to 0} \frac{\varphi(P + hE_i) - \varphi(P)}{h} = f_i(P).$$

This proves what we wanted.

Remark. The use of the fundamental theorem of calculus in the preceding proof should be recognized as absolutely straightforward. If G is an indefinite integral for g, then

$$\int_0^h g(t)\,dt = G(h) - G(0),$$

and hence

$$\frac{1}{h}\int_0^h g(t)\,dt = \frac{G(h) - G(0)}{h}$$

is the ordinary Newton quotient for G. The fundamental theorem of calculus asserts precisely that the limit as $h \to 0$ is equal to $G'(0) = g(0)$.

Exercise

1. To take care of the case when h is negative in the proof of Theorem 6, use the parametrization $C(t) = P + thE_i$ with $0 \leqq t \leqq 1$. Making a change of variables, $u = th$, $du = h\,dt$, show that the proof follows exactly the same pattern as that given in the text.

CHAPTER VI

Higher Derivatives

In this chapter, we discuss two things which are of independent interest. First, we define partial differential operators (with constant coefficients). It is very useful to have facility in working with these formally.

Secondly, we apply them to the derivation of Taylor's formula for functions of several variables, which will be very similar to the formula for one variable. The formula, as before, tells us how to approximate a function by means of polynomials. In the present theory, these polynomials involve several variables, of course. We shall see that they are hardly more difficult to handle than polynomials in one variable in the matters under consideration.

The proof that the partial derivatives commute is tricky. It can be omitted without harm in a class allergic to theory, because the technique involved never reappears in the rest of this book.

§1. *Repeated partial derivatives*

Let f be a function of two variables, defined on an open set U in 2-space. Assume that its first partial derivative exists. Then $D_1 f$ (which we also write $\partial f/\partial x$ if x is the first variable) is a function defined on U. We may then ask for its first or second partial derivatives, i.e. we may form $D_2 D_1 f$ or $D_1 D_1 f$ if these exist. Similarly, if $D_2 f$ exists, and if the first partial derivative of $D_2 f$ exists, we may form $D_1 D_2 f$.

Suppose that we write f in terms of the two variables (x, y). Then we can write

$$D_1 D_2 f(x, y) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = (D_1(D_2 f))(x, y),$$

and

$$D_2 D_1 f(x, y) = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = (D_2(D_1 f))(x, y).$$

For example, let $f(x, y) = \sin(xy)$. Then

$$\frac{\partial f}{\partial x} = y\cos(xy) \qquad \text{and} \qquad \frac{\partial f}{\partial y} = x\cos(xy).$$

Hence

$$D_2D_1f(x, y) = -xy\sin(xy) + \cos(xy).$$

But differentiating $\partial f/\partial y$ with respect to x, we see that

$$D_1D_2f(x, y) = -xy\sin(xy) + \cos(xy).$$

These two repeated partial derivatives are equal!

The next theorem tells us that in practice, this will always happen.

Theorem 1. *Let f be a function of two variables, defined on an open set U of 2-space. Assume that the partial derivatives D_1f, D_2f, D_1D_2f, and D_2D_1f exist and are continuous. Then*

$$D_1D_2f = D_2D_1f.$$

Proof. A direct use of the definition of these partial and repeated partial derivatives would lead to a blind alley. Hence we shall have to use a special trick to pull through.

Let (x, y) be a point in U, and let $H = (h, k)$ be small, $h \neq 0$, $k \neq 0$. We consider the expression

$$g(x) = f(x, y + k) - f(x, y).$$

If we apply the mean value theorem to g, then we conclude that there exists a number s_1 between x and $x + h$ such that

$$g(x + h) - g(x) = g'(s_1)h,$$

or in other words, using the definitions of partial derivative:

$$g(x + h) - g(x) = [D_1f(s_1, y + k) - D_1f(s_1, y)]h. \tag{1}$$

But the difference on the left of this equation is

$$f(x + h, y + k) - f(x + h, y) - f(x, y + k) + f(x, y). \tag{2}$$

On the other hand, we can now apply the mean value theorem to the expression in brackets in (1) *with respect to the second variable.* If we do this, we see that the long expression in (2) is equal to

$$D_2D_1f(s_1, s_2)kh \tag{3}$$

for some number s_2 lying between y and $y + k$.

We now start all over again, and consider the expression

$$g_2(y) = f(x + h, y) - f(x, y).$$

We apply the mean value theorem to g_2, and conclude that there is a number t_2 between y and $y + k$ such that

$$g_2(y + k) - g_2(y) = g_2'(t_2)k,$$

or in other words, is equal to

(4) $$[D_2 f(x + h, t_2) - D_2 f(x, t_2)]k.$$

If you work out $g_2(y + k) - g_2(y)$, you will see that it is equal to the long expression of (2). Furthermore, proceeding as before, and applying the mean value theorem to the first variable in (4), we see that (4) becomes

(5) $$D_1 D_2 f(t_1, t_2)hk$$

for some number t_1 between x and $x + h$. Since (5) and (3) are both equal to the long expression in (2), they are equal to each other. Thus finally we obtain

$$D_2 D_1 f(s_1, s_2)kh = D_1 D_2 f(t_1, t_2)hk.$$

Since we assume from the beginning that $h \neq 0$ and $k \neq 0$, we can cancel hk, and get

$$D_2 D_1 f(s_1, s_2) = D_1 D_2 f(t_1, t_2).$$

Now as h, k approach 0, the left side of this equation approaches $D_2 D_1 f(x, y)$ because $D_2 D_1 f$ is assumed to be continuous. Similarly, the right-hand side approaches $D_1 D_2 f(x, y)$. We can therefore conclude that

$$D_1 D_2 f(x, y) = D_2 D_1 f(x, y),$$

as desired.

Consider now a function of three variables $f(x, y, z)$. We can then take three kinds of partial derivatives: D_1, D_2, or D_3 (in other notation, $\partial/\partial x$, $\partial/\partial y$, and $\partial/\partial z$). Let us assume throughout that all the partial derivatives which we shall consider exist and are continuous, so that we may form as many repeated partial derivatives as we please. Then using Theorem 1, we can show that it does not matter in which order we take these partials.

For instance, we see that

$$D_3 D_1 f = D_1 D_3 f.$$

This is simply an application of Theorem 1, keeping the second variable fixed. We may take a further partial derivative, for instance

$$D_1 D_3 D_1 f.$$

Here D_1 occurs twice and D_3 once. Then this expression will be equal to any other repeated partial derivative of f in which D_1 occurs twice and D_3 once. For example, we apply the theorem to the function $(D_1 f)$. Then the theorem allows us to interchange D_1 and D_3 in front of $(D_1 f)$ (always assuming that all partials we want to take exist and are con-

tinuous). We obtain

$$D_1D_3(D_1f) = D_3D_1(D_1f).$$

As another example, consider

$$D_2D_1D_3D_2f. \tag{6}$$

We wish to show that it is equal to $D_1D_2D_2D_3f$. By Theorem 1, we have $D_3D_2f = D_2D_3f$. Hence:

$$D_2D_1(D_3D_2f) = D_2D_1(D_2D_3f). \tag{7}$$

We then apply Theorem 1 again, and interchange D_2 and D_1 to obtain the desired expression.

In general, suppose that we are given three positive integers m_1, m_2, and m_3. We wish to take the repeated partial derivatives of f by using m_1 times the first partial D_1, using m_2 times the second partial D_2, and using m_3 times the third partial D_3. Then it does not matter in which order we take these partial derivatives, we shall always get the same answer.

To see this, note that by repeated application of Theorem 1, we can always interchange any occurrence of D_3 with D_2 or D_1 so as to push D_3 towards the right. We can perform such interchanges until all occurrences of D_3 occur furthest to the right, in the same way as we pushed D_3 towards the right going from expression (6) to expression (7). Once this is done, we start interchanging D_2 with D_1 until all occurrences of D_2 pile up just behind D_3. Once this is done, we are left with D_1 repeated a certain number of times on the left.

No matter with what arrangement of D_1, D_2, D_3 we started, we end up with the *same* arrangement, namely

$$\underbrace{D_1 \cdots D_1}_{m_1} \underbrace{D_2 \cdots D_2}_{m_2} \underbrace{D_3 \cdots D_3}_{m_3} f,$$

with D_1 occurring m_1 times, D_2 occurring m_2 times, and D_3 occurring m_3 times.

Exactly the same argument works for functions of more variables.

Exercises

In all problems, functions are assumed to be differentiable as needed.

Find the partial derivatives of order 2 for the following functions and verify explicitly in each case that $D_1D_2f = D_2D_1f$.

1. e^{xy}
2. $\sin(xy)$

3. $x^2y^3 + 3xy$
4. $2xy + y^2$
5. $e^{x^2+y^2}$
6. $\sin(x^2 + y)$
7. $\cos(x^3 + xy)$
8. $\arctan(x^2 - 2xy)$
9. e^{x+y}
10. $\sin(x + y)$.

Find $D_1D_2D_3f$ and $D_3D_2D_1f$ in the following cases.

11. xyz
12. x^2yz
13. e^{xyz}
14. $\sin(xyz)$
15. $\cos(x + y + z)$
16. $\sin(x + y + z)$
17. $(x^2 + y^2 + z^2)^{-1}$
18. $x^3y^2z + 2(x + y + z)$.
19. Let $x = r\cos\theta$ and $y = r\sin\theta$. Let $f(x, y) = g(r, \theta)$. Show that

$$\frac{\partial f}{\partial x} = \cos\theta\frac{\partial g}{\partial r} - \frac{\sin\theta}{r}\frac{\partial g}{\partial\theta}$$

$$\frac{\partial f}{\partial y} = \sin\theta\frac{\partial g}{\partial r} + \frac{\cos\theta}{r}\frac{\partial g}{\partial\theta}.$$

[*Hint:* Using the chain rule, find first $\frac{\partial g}{\partial r}$ and $\frac{\partial g}{\partial\theta}$ in terms of $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$. Then solve back a system of two equations in two unknowns.]

20. Let $x = r\cos\theta$ and $y = r\sin\theta$. Let $f(x, y) = g(r, \theta)$. Show that

$$\frac{\partial^2 g}{\partial r^2} + \frac{1}{r}\frac{\partial g}{\partial r} + \frac{1}{r^2}\frac{\partial^2 g}{\partial\theta^2} = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}.$$

21. Let $f(X) = g(r)$ (with $r = \|X\|$), and assume $X = (x, y, z)$. Show that

$$\frac{d^2g}{dr^2} + \frac{2}{r}\frac{dg}{dr} = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}.$$

22. Let $f(x, y)$ satisfy $f(tx, ty) = t^n f(x, y)$ for all t (n being some integer $\geqq 1$). Show that

$$x\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y} = nf(x, y).$$

23. Let f be as in Exercise 22. Show that

$$x^2\frac{\partial^2 f}{\partial x^2} + 2xy\frac{\partial^2 f}{\partial x\,\partial y} + y^2\frac{\partial^2 f}{\partial y^2} = n(n - 1)f(x, y).$$

(It is understood throughout that all functions are as many times differentiable as is necessary.)

24. A function of three variables $f(x, y, z)$ is said to satisfy **Laplace's equation** if

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = 0.$$

Verify that the following functions satisfy Laplace's equation.

(a) $x^2 + y^2 - 2z^2$ (b) $\log \sqrt{x^2 + y^2}$

(c) $\dfrac{1}{\sqrt{x^2 + y^2 + z^2}}$ (d) $e^{3x+4y} \cos(5z)$

25. Let $z = f(u, v)$ and $u = x + y$, $v = x - y$. Show that

$$\frac{\partial^2 z}{\partial x\, \partial y} = \frac{\partial^2 z}{\partial u^2} - \frac{\partial^2 z}{\partial v^2}.$$

26. Let $z = f(x + y) - g(x - y)$, where f, g are functions of one variable. Let $u = x + y$ and $v = x - y$. Show that

$$\frac{\partial^2 z}{\partial x^2} = \frac{\partial^2 z}{\partial y^2} = f''(u) - g''(v).$$

27. Let c be a constant, and let $z = \sin(x + ct) + \cos(2x + 2ct)$. Show that

$$\frac{\partial^2 z}{\partial t^2} = c^2 \frac{\partial^2 z}{\partial x^2}.$$

28. Let $z = f\left(\dfrac{x - y}{y}\right)$. Show that $x(\partial z/\partial x) + y(\partial z/\partial y) = 0$.

29. Let c be a constant, and let $z = f(x + ct) + g(x - ct)$. Let $u = x + ct$ and $v = x - ct$. Show that

$$\frac{\partial^2 z}{\partial t^2} = c^2 \frac{\partial^2 z}{\partial x^2} = c^2(f''(u) + g''(v)).$$

30. Let F be a vector field on an open set in 3-space, so that F is given by three coordinate functions, say $F = (f_1, f_2, f_3)$. Define the **curl** of F to be the vector field given by

$$(\text{curl } F)(x_1, x_2, x_3) = \left(\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3}, \frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1}, \frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2}\right).$$

Define the **divergence** of F to be the function $g = \text{div } F$ given by

$$g(x, y, z) = \frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial y} + \frac{\partial f_3}{\partial z}.$$

(a) Prove that

$$\text{div curl } F = 0.$$

(b) Prove that curl grad $\varphi = O$, for any function φ.

§2. *Partial differential operators*

We shall continue the discussion at the end of the last section, but we shall build up a convenient system to talk about iterated partial derivatives.

For simplicity, let us begin with functions of one variable x. We can then take only one type of derivative,

$$D = \frac{d}{dx}.$$

Let f be a function of one variable, and let us assume that all the iterated derivatives of f exist. Let m be a positive integer. Then we can take the m-th derivative of f, which we once denoted by $f^{(m)}$. We now write it

$$DD\cdots Df \qquad \text{or} \qquad \frac{d}{dx}\left(\frac{d}{dx}\cdots\left(\frac{df}{dx}\right)\cdots\right),$$

the derivative D (or d/dx) being iterated m times. What matters here is the number of times D occurs. We shall use the notation D^m or $(d/dx)^m$ to mean the iteration of D, m times. Thus we write

$$D^m f \qquad \text{or} \qquad \left(\frac{d}{dx}\right)^m f$$

instead of the above expressions. This is shorter. But even better, we have the rule

$$D^m D^n f = D^{m+n} f$$

for any positive integers m, n. So this iteration of derivatives begins to look like a multiplication. Furthermore, if we define $D^0 f$ to be simply f, then the rule above also holds if m, n are $\geqq 0$.

The expression D^m will be called a **simple differential operator of order** m (in one variable, so far).

Let us now look at the case of two variables, say (x, y). We can then take two partials D_1 and D_2 (or $\partial/\partial x$ and $\partial/\partial y$). Let m_1, m_2 be two integers $\geqq 0$. Instead of writing

$$\underbrace{D_1\cdots D_1}_{m_1}\underbrace{D_2\cdots D_2}_{m_2} f \qquad \text{or} \qquad \underbrace{\frac{\partial}{\partial x}\cdots\left(\frac{\partial}{\partial x}\right.}_{m_1}\underbrace{\left(\frac{\partial}{\partial y}\cdots\left(\frac{\partial f}{\partial y}\right)\right.}_{m_2}\left.\left.\cdots\right)\right),$$

we shall write

$$D_1^{m_1} D_2^{m_2} f \qquad \text{or} \qquad \left(\frac{\partial}{\partial x}\right)^{m_1}\left(\frac{\partial}{\partial y}\right)^{m_2} f.$$

For instance, taking $m_1 = 2$ and $m_2 = 5$ we would write

$$D_1^2 D_2^5 f.$$

This means: take the first partial twice and the second partial five times (in any order). (We assume throughout that all repeated partials exist and are continuous.)

An expression of type

$$D_1^{m_1} D_2^{m_2}$$

will be called a simple differential operator, and we shall say that its **order** is $m_1 + m_2$. In the example we just gave, the order is $5 + 2 = 7$.

It is now clear how to proceed with three or more variables, and it is no harder to express our thoughts in terms of n variables than in terms of three. Consequently, if we deal with functions of n variables, all of whose repeated partial derivatives exist and are continuous in some open set U, and if $D_1, \ldots, D_n$ denote the partial derivatives with respect to these variables, then we call an expression

$$D_1^{m_1} \cdots D_n^{m_n} \quad \text{or} \quad \left(\frac{\partial}{\partial x_1}\right)^{m_1} \cdots \left(\frac{\partial}{\partial x_n}\right)^{m_n}$$

a **simple differential operator,** $m_1, \ldots, m_n$ being integers $\geqq 0$. We say that its **order** is $m_1 + \cdots + m_n$.

Given a function f (satisfying the above stated conditions), and a simple differential operator D, we write Df to mean the function obtained from f by applying repeatedly the partial derivatives $D_1, \ldots, D_n$, the number of times being the number of times each D_i occurs in D.

Example 1. Consider functions of three variables (x, y, z). Then

$$D = \left(\frac{\partial}{\partial x}\right)^3 \left(\frac{\partial}{\partial y}\right)^5 \left(\frac{\partial}{\partial z}\right)^2$$

is a simple differential operator of order $3 + 5 + 2 = 10$. Let f be a function of three variables satisfying the usual hypotheses. To take Df means that we take the partial derivative with respect to z twice, the partial with respect to y five times, and the partial with respect to x three times.

We observe that a simple differential operator gives us a rule which to each function f associates another function Df.

As a matter of notation, referring to Example 1, one would also write the differential operator D in the form

$$D = \frac{\partial^{10}}{\partial x^3 \partial y^5 \partial z^2}.$$

In this notation, one would thus have

$$\left(\frac{\partial}{\partial x}\right)^2 f = \frac{\partial^2 f}{\partial x^2}$$

and

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x\, \partial y}.$$

All the above notations are used in the scientific literature, and this is the reason for including them here.

Warning. Do not confuse the two expressions

$$\left(\frac{\partial}{\partial x}\right)^2 f = \frac{\partial^2 f}{\partial x^2} \quad \text{and} \quad \left(\frac{\partial f}{\partial x}\right)^2,$$

which are usually *not* equal. For instance, if $f(x, y) = x^2y$, then

$$\frac{\partial^2 f}{\partial x^2} = 2y \quad \text{and} \quad \left(\frac{\partial f}{\partial x}\right)^2 = 4x^2y^2.$$

We shall now show how one can add simple differential operators and multiply them by constants.

Let D, D' be two simple differential operators. For any function f we define $(D + D')f$ to be $Df + D'f$. If c is a number, then we define $(cD)f$ to be $c(Df)$. In this manner, taking iterated sums, and products with constants, we obtain what we shall call **differential operators.** Thus a **differential operator** D is a sum of terms of type

$$cD_1^{m_1} \cdots D_n^{m_n},$$

where c is a number and $m_1, \ldots, m_n$ are integers $\geqq 0$.

Example 2. Dealing with two variables, we see that

$$D = 3\frac{\partial}{\partial x} + 5\left(\frac{\partial}{\partial x}\right)^2 - \pi\frac{\partial}{\partial x}\frac{\partial}{\partial y}$$

is a differential operator. Let $f(x, y) = \sin(xy)$. We wish to find Df. By definition,

$$\begin{aligned} Df(x, y) &= 3\frac{\partial f}{\partial x} + 5\left(\frac{\partial}{\partial x}\right)^2 f - \pi\frac{\partial}{\partial x}\frac{\partial f}{\partial y} \\ &= 3y\cos(xy) + 5(-y^2\sin(xy)) \\ &\quad - \pi[y(-\sin(xy))x + \cos(xy)]. \end{aligned}$$

We see that a differential operator associates with each function f (satisfying the usual conditions) another function Df.

Let c be a number and f a function. Let D_i be any partial derivative. Then

$$D_i(cf) = cD_if.$$

This is simply the old property that the derivative of a constant times a function is equal to the constant times the derivative of the function. Iterating partial derivatives, we see that this same property applies to differential operators. For any differential operator D, and any number c, we have

$$D(cf) = cDf.$$

Furthermore, if f, g are two functions (defined on the same open set, and having continuous partial derivatives of all orders), then for any partial derivative D_i, we have

$$D_i(f + g) = D_if + D_ig.$$

Iterating the partial derivatives, we find that for any differential operator D, we have

$$D(f + g) = Df + Dg.$$

Having learned how to add differential operators, we now learn how to multiply them.

Let D, D' be two differential operators. Then we define the differential operator DD' to be the one obtained by taking first D' and then D. In other words, if f is a function, then

$$(DD')f = D(D'f).$$

Example 3. Let

$$D = 3\frac{\partial}{\partial x} + 2\frac{\partial}{\partial y} \qquad \text{and} \qquad D' = \frac{\partial}{\partial x} + 4\frac{\partial}{\partial y}\cdot$$

Then

$$\begin{aligned} DD' &= \left(3\frac{\partial}{\partial x} + 2\frac{\partial}{\partial y}\right)\left(\frac{\partial}{\partial x} + 4\frac{\partial}{\partial y}\right) \\ &= 3\left(\frac{\partial}{\partial x}\right)^2 + 14\frac{\partial}{\partial x}\frac{\partial}{\partial y} + 8\left(\frac{\partial}{\partial y}\right)^2\cdot \end{aligned}$$

Differential operators multiply just like polynomials and numbers, and their addition and multiplication satisfy all the rules of addition and multiplication of polynomials. For instance:

If D, D' are two differential operators, then

$$DD' = D'D.$$

If D, D', D'' are three differential operators, then

$$D(D' + D'') = DD' + DD''.$$

It would be tedious to list all the properties here and to give in detail all the proofs (even though they are quite simple). We shall therefore omit these proofs. The main purpose of this section is to insure that you develop as great a facility in adding and multiplying differential operators as you have in adding and multiplying numbers or polynomials.

When a differential operator is written as a sum of terms of type

$$cD_1^{m_1} \cdots D_n^{m_n},$$

then we shall say that it is in **standard form.**

For example,

$$3\left(\frac{\partial}{\partial x}\right)^2 + 14\frac{\partial}{\partial x}\frac{\partial}{\partial y} + 8\left(\frac{\partial}{\partial y}\right)^2$$

is in standard form, but

$$\left(3\frac{\partial}{\partial x} + 2\frac{\partial}{\partial y}\right)\left(\frac{\partial}{\partial x} + 4\frac{\partial}{\partial y}\right)$$

is not.

Each term

$$cD_1^{m_1} \cdots D_n^{m_n}$$

is said to have degree $m_1 + \cdots + m_n$. If a differential operator is expressed as a sum of simple differential operators which all have the same degree, say m, then we say that it is **homogeneous** of degree m.

The differential operator of Example 2 is not homogeneous. The differential operator DD' of Example 3 is homogeneous of degree 2.

An important case of differential operators being applied to functions is that of monomials.

Example. Let $f(x, y) = x^3y^2$. Then

$$D_1f(x, y) = 3x^2y^2, \qquad D_1^2f(x, y) = 2 \cdot 3xy^2$$
$$D_1^3f(x, y) = 6y^2, \qquad D_1^4f(x, y) = 0.$$

Also observe that

$$D_1^3D_2^2f(x, y) = 3!2!$$

Example. The generalization of the above example is as follows, and will be important for Taylor's formula. Let

$$f(x, y) = x^iy^j$$

be a monomial, with exponents $i, j \geqq 0$. Then

$$D_1^iD_2^jf(x, y) = i!j!.$$

This is immediately verified, by differentiating x^i with respect to x, i times, thus getting rid of all powers of x: and differentiating y^j with respect

to y, j times, thus getting rid of all powers of y.

On the other hand, let r, s be integers $\geqq 0$ such that $i \neq r$ or $j \neq s$. Then

$$D_1^r D_2^s f(0, 0) = 0.$$

To see this, suppose that $r \neq i$. If $r > i$, then differentiating r times the power x^i yields 0. If $r < i$, then differentiating r times the power x^i yields

$$i(i-1)\cdots(i-r+1)x^{i-r},$$

and $i - r > 0$. Substituting $x = 0$ yields 0. The same argument works if $j \neq s$.

Exercises

Put the following differential operators in standard form.

1. $(3D_1 + 2D_2)^2$
2. $(D_1 + D_2 + D_3)^2$
3. $(D_1 - D_2)(D_1 + D_2)$
4. $(D_1 + D_2)^2$
5. $(D_1 + D_2)^3$
6. $(D_1 + D_2)^4$
7. $(2D_1 - 3D_2)(D_1 + D_2)$
8. $(D_1 - D_3)(D_2 + 5D_3)$
9. $\left(\frac{\partial}{\partial x} + 4\frac{\partial}{\partial y}\right)^3$
10. $\left(2\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)^2$
11. $\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^2$
12. $\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^3$

Find the values of the differential operator of Exercise 10 applied to the following functions at the given point.

13. x^2y at $(0, 1)$
14. xy at $(1, 1)$
15. $\sin(xy)$ at $(0, \pi)$
16. e^{xy} at $(0, 0)$
17. Compute $D_1^4 D_2^3 f(x, y)$ if $f(x, y)$ is
   (a) x^5y^4 (b) x^4y^2
   (c) x^4y^3 (d) $10x^4y^3$
18. Compute $D_1^7 D_2^9 f(0, 0)$ if $f(x, y)$ is
   (a) x^8y^7 (b) $3x^7y^9$
   (c) $11x^7y^9$ (d) $25x^6y^{11}$
19. Let $f(x, y) = 3x^2y + 4x^3y^4 - 7x^9y^4$. Find
   (a) $D_1^3 D_2^4 f(0, 0)$ (b) $D_1^9 D_2^4 f(0, 0)$
   (c) $D_1^2 D_2 f(0, 0)$ (d) $D_1^3 D_2 f(0, 0)$
20. Let $f(x, y, z) = 4x^2yz^3 - 5x^3y^4z + 7x^6y^{10}z^7$. Find
   (a) $D_1^2 D_2 D_3^2 f(0, 0, 0)$ (b) $D_1^2 D_2 D_3^3 f(0, 0, 0)$
   (c) $D_1^6 D_2^{10} D_3^7 f(0, 0, 0)$ (d) $D_1^5 D_2 D_3 f(0, 0, 0)$

21. Let f, g be two functions (of two variables) with continuous partial derivatives of order $\leqq 2$ in an open set U. Assume that

$$\frac{\partial f}{\partial x} = -\frac{\partial g}{\partial y} \quad \text{and} \quad \frac{\partial f}{\partial y} = \frac{\partial g}{\partial x}.$$

Show that

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0.$$

22. Let f be a function of three variables, defined for $X \neq O$ by $f(X) = 1/\|X\|$. Show that

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = 0.$$

23. In Exercise 20 of the preceding section, compute

$$\left(\frac{\partial}{\partial x}\right)^2 + \left(\frac{\partial}{\partial y}\right)^2$$

in terms of $\partial/\partial r$ and $\partial/\partial\theta$. Watch out! The coefficients are not constant.

§3. *Taylor's formula*

In the theory of functions of one variable, we derived an expression for the values of a function f near a point a by means of the derivatives, namely

$$f(a + h) = f(a) + f'(a)h + \frac{f^{(2)}(a)}{2!}h^2 + \cdots + \frac{f^{(r-1)}(a)}{(r-1)!}h^{r-1} + R_r,$$

where

$$R_r = \frac{f^{(r)}(c)}{r!}h^r,$$

for some point c between a and $a + h$. We shall now derive a similar formula for functions of several variables. We begin with the case of two variables.

We let $P = (a, b)$ and $H = (h, k)$. We assume that P is in an open set U and that f is a function on U all of whose partial derivatives up to order n exist and are continuous. We are interested in finding an expression

$$f(P + H) = f(P) + \ ?\ ?\ ?$$

The idea is to reduce the problem to the one variable case. Thus we define the function

$$g(t) = f(P + tH) = f(a + th, b + tk)$$

for $0 \leqq t \leqq 1$. We assume that U contains all points $P + tH$ for

$0 \leqq t \leqq 1$. Then

$$g(1) = f(P + H) \quad \text{and} \quad g(0) = f(P).$$

We can use Taylor's formula in one variable applied to the function g and we know that

$$g(1) = g(0) + \frac{g'(0)}{1!} + \cdots + \frac{g^{(r-1)}(0)}{(r-1)!} + \frac{g^{(r)}(\tau)}{r!}$$

for some number τ between 0 and 1, provided that g has r continuous derivatives. We shall now prove that the derivatives of g can be expressed in terms of the partial derivatives of f, and thus obtain the desired Taylor formula for f. We shall first do it for $n = 2$.

We let $x = a + th$ and $y = b + tk$. By the chain rule:

$$g'(t) = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} = \frac{\partial f}{\partial x} h + \frac{\partial f}{\partial y} k. \tag{1}$$

For the second derivative, we must find the derivative with respect to t of each one of the functions $\partial f/\partial x$ and $\partial f/\partial y$. By the chain rule applied to each such function, we have:

$$\begin{aligned}
\frac{d}{dt}\left(\frac{\partial f}{\partial x}\right) &= \frac{\partial^2 f}{\partial x^2}\frac{dx}{dt} + \frac{\partial^2 f}{\partial y\,\partial x}\frac{dy}{dt} = \frac{\partial^2 f}{\partial x^2} h + \frac{\partial^2 f}{\partial y\,\partial x} k, \\
\frac{d}{dt}\left(\frac{\partial f}{\partial y}\right) &= \frac{\partial^2 f}{\partial x\,\partial y} h + \frac{\partial^2 f}{\partial y^2} k.
\end{aligned} \tag{2}$$

Hence using (2) to take the derivative of (1), we find:

$$\begin{aligned}
g''(t) &= h\left[\frac{\partial^2 f}{\partial x^2} h + \frac{\partial^2 f}{\partial y\,\partial x} k\right] + k\left[\frac{\partial^2 f}{\partial x\,\partial y} h + \frac{\partial^2 f}{\partial y^2} k\right] \\
&= h^2 \frac{\partial^2 f}{\partial x^2} + 2hk \frac{\partial^2 f}{\partial x\,\partial y} + k^2 \frac{\partial^2 f}{\partial y^2}.
\end{aligned}$$

This expression can be rewritten more easily in terms of differential operators, namely we see that the expression for $g''(t)$ is equal to

$$\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^2 f.$$

If we wish to free the notation from the x and y, then we can use the notation

$$\begin{aligned}
g''(t) &= (hD_1 + kD_2)^2 f(P + tH) \\
&= (hD_1 + kD_2)^2 f(a + th, b + tk).
\end{aligned}$$

As usual, this means that we apply $(hD_1 + kD_2)^2$ to f, and then evaluate this function at the point $(a + th, b + tk)$.

The expression $hD_1 + kD_2$ looks like a dot product, and thus it is useful to abbreviate the notation and write

$$hD_1 + kD_2 = H \cdot \nabla.$$

With this abbreviation, our first derivative for g can then be written [from (1)]:

$$g'(t) = (H \cdot \nabla)f(P + tH),$$

and the second derivative can be written

$$g''(t) = (H \cdot \nabla)^2 f(P + tH).$$

Here again, we emphasize that $(H \cdot \nabla)$ and $(H \cdot \nabla)^2$ are first applied to f, so that strictly speaking we should write an extra set of parentheses, e.g.

$$g'(t) = \big((H \cdot \nabla)f\big)(a + th, b + tk)$$

and similarly for $g''(t)$.

The higher derivatives of g are determined similarly by induction.

Theorem 2. *Let r be a positive integer. Let f be a function defined on an open set U, and having continuous partial derivatives of orders $\leqq r$. Let P be a point of U, and H a vector. Let $g(t) = f(P + tH)$. Then*

$$g^{(r)}(t) = \big((H \cdot \nabla)^r f\big)(P + tH)$$

for all values of t such that $P + tH$ lies in U.

Proof. The case $r = 1$ (even $r = 2$) has already been verified. Suppose our formula proved for some integer r. Let $\psi = (H \cdot \nabla)^r f$. Then

$$g^{(r)}(t) = \psi(P + tH).$$

Hence by the case for $r = 1$ we get

$$g^{(r+1)}(t) = \big((H \cdot \nabla)\psi\big)(P + tH).$$

Substituting the value for ψ yields

$$g^{(r+1)}(t) = \big((H \cdot \nabla)^{r+1} f\big)(P + tH),$$

thus proving our theorem by induction.

In terms of the $\partial/\partial x$ and $\partial/\partial y$ notation, we see that

$$\boxed{g^{(r)}(t) = \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^r f(P + tH).}$$

Taylor's Formula. *Let f be a function defined on an open set U, and having continuous partial derivatives up to order r. Let P be a point of U, and H a vector. Assume that the line segment*

$$P + tH, \qquad 0 \leqq t \leqq 1,$$

is contained in U. Then there exists a number τ between 0 and 1 such that

$$f(P+H) = f(P) + \frac{(H\cdot\nabla)f(P)}{1!} + \cdots + \frac{(H\cdot\nabla)^{r-1}f(P)}{(r-1)!} + \frac{(H\cdot\nabla)^r f(P+\tau H)}{r!}.$$

Proof. This is obtained by plugging the expression for the derivatives of the function $g(t) = f(P + tH)$ into the Taylor formula for one variable. We see that

$$g^{(s)}(0) = (H\cdot\nabla)^s f(P)$$

and

$$g^{(r)}(\tau) = (H\cdot\nabla)^r f(P+\tau H).$$

This proves Taylor's formula as stated.

Rewritten in terms of the $\partial/\partial x$ and $\partial/\partial y$ notation, we have

$$\begin{aligned} f(a+h, b+k) = f(a,b) &+ \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right) f(a,b) + \cdots \\ &+ \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^{r-1} f(a,b) \\ &+ \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^{r} f(a+\tau h, b+\tau k). \end{aligned}$$

The powers of the differential operators

$$\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^s$$

are found by the usual binomial expansion. For instance:

$$\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^2 = h^2\frac{\partial^2}{\partial x^2} + 2hk\frac{\partial^2}{\partial x\,\partial y} + k^2\frac{\partial^2}{\partial y^2},$$

$$\begin{aligned} \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^3 &= h^3\left(\frac{\partial}{\partial x}\right)^3 + 3h^2k\left(\frac{\partial}{\partial x}\right)^2\left(\frac{\partial}{\partial y}\right) \\ &\quad + 3hk^2\left(\frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial y}\right)^2 + k^3\left(\frac{\partial}{\partial y}\right)^3. \end{aligned}$$

Example 1. Find the terms of degree $\leqq 2$ in the Taylor formula for the function $f(x, y) = \log(1 + x + 2y)$ at the point $(2, 1)$.

We compute the partial derivatives. They are:

$$f(2, 1) = \log 5,$$

$$D_1 f(x, y) = \frac{1}{1 + x + 2y}, \qquad D_1 f(2, 1) = \frac{1}{5} = \frac{\partial f}{\partial x}(2, 1),$$

$$D_2 f(x, y) = \frac{2}{1 + x + 2y}, \qquad D_2 f(2, 1) = \frac{2}{5} = \frac{\partial f}{\partial y}(2, 1),$$

$$D_1^2 f(x, y) = -\frac{1}{(1 + x + 2y)^2}, \qquad D_1^2 f(2, 1) = -\frac{1}{25} = \frac{\partial^2 f}{\partial x^2}(2, 1),$$

$$D_2^2 f(x, y) = -\frac{4}{(1 + x + 2y)^2}, \qquad D_2^2 f(2, 1) = -\frac{4}{25} = \frac{\partial^2 f}{\partial y^2}(2, 1),$$

$$D_1 D_2 f(x, y) = -\frac{2}{(1 + x + 2y)^2}, \; D_1 D_2 f(2, 1) = -\frac{2}{25} = \frac{\partial^2 f}{\partial x\, \partial y}(2, 1).$$

Hence

$$f(2 + h, 1 + k) = \log 5 + \left(\frac{1}{5} h + \frac{2}{5} k\right) + \frac{1}{2!}\left[-\frac{1}{25} h^2 - \frac{4}{25} hk - \frac{4}{25} k^2\right] + \cdots.$$

In many cases, we take $P = O$ and we wish to approximate $f(x, y)$ by a polynomial in x, y. Thus we let $H = (x, y)$. In that case, the notation $\partial/\partial x$ and $\partial/\partial y$ becomes even worse than usual since it is not entirely clear in taking the square

$$\left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y}\right)^2$$

what is to be treated as a constant and what is not. Thus it is better to write

$$(xD_1 + yD_2)^2,$$

and similarly for higher powers. We then obtain a polynomial expression for f, with a remainder term. The terms of degree $\leqq 3$ are as follows:

$$\begin{aligned}
&f(x, y)\\
&= f(0, 0) + D_1 f(0, 0)x + D_2 f(0, 0)y\\
&+ \frac{1}{2!}[D_1^2 f(0, 0)x^2 + 2D_1 D_2 f(0, 0)xy + D_2^2 f(0, 0)y^2]\\
&+ \frac{1}{3!}[D_1^3 f(0, 0)x^3 + 3D_1^2 D_2 f(0, 0)x^2 y + 3D_1 D_2^2 f(0, 0)xy^2 + D_2^3 f(0, 0)y^3]\\
&\qquad + R_4.
\end{aligned}$$

In general, the Taylor formula gives us an expression

$$f(x, y) = f(0, 0) + G_1(x, y) + \cdots + G_{r-1}(x, y) + R_r,$$

where $G_d(x, y)$ is a homogeneous polynomial in x, y of degree d, and R_r is the remainder term. We call

$$f(0, 0) + G_1(x, y) + \cdots + G_s(x, y)$$

the **polynomial approximation of** f, of degree $\leqq s$.

We write polynomials in one variable as sums

$$\sum_{i=0}^{n} c_i x^i = c_0 + c_1 x + \cdots + c_n x^n.$$

In a similar way, we can write polynomials in several variables,

$$G(x, y) = \sum_{i=0}^{n} \sum_{j=0}^{m} c_{ij} x^i y^j.$$

Let r, s be a pair of integers $\geqq 0$. Then

$$D_1^r D_2^s G(0, 0) = r!s!c_{rs},$$

by the example at the end of §2. Hence we have a simple expression for the coefficients of the polynomial,

$$\boxed{c_{ij} = \frac{D_1^i D_2^j G(0, 0)}{i!j!}.}$$

On the other hand, from the binomial expansion

$$(xD_1 + yD_2)^m = \sum_{i=0}^{m} \binom{m}{i} x^i y^{m-i} D_1^i D_2^{m-i},$$

and the value of the binomial coefficient,

$$\binom{m}{i} = \frac{m!}{i!(m - i)!},$$

we find that

$$\frac{(xD_1 + yD_2)^m}{m!} = \sum_{i=0}^{m} \frac{x^i y^{m-i}}{i!(m - i)!} D_1^i D_2^{m-i}.$$

Consequently,

$$\frac{(xD_1 + yD_2)^m f(0, 0)}{m!} = \sum_{i=0}^{m} c_{ij} x^i y^j = G_m(x, y)$$

is a polynomial in x, y, such that $i + j = m$, so all its monomials have

the same degree, and the coefficients are given by

$$(*) \qquad \boxed{c_{ij} = \frac{D_1^i D_2^j f(0, 0)}{i!\, j!}.}$$

The general Taylor polynomial of degree $\leqq s$ is therefore of the form

$$G(x, y) = \sum_{i+j \leqq s} c_{ij} x^i y^j,$$

where the coefficients c_{ij} are given by the above formula (*). Again, the example at the end of §2 shows that the partial derivatives up to total order s of this polynomial coincide with the derivatives of f, when evaluated at $(0, 0)$. Thus we may say:

The Taylor polynomial of a function f up to order s is that polynomial having the same partial derivatives as the function up to order s, when evaluated at $(0, 0)$.

Example 2. Find the polynomial approximation of the function

$$f(x, y) = \log (1 + x + 2y)$$

up to degree 2.

We computed the partial derivatives in Example 1. For the present application, we have

$$f(0, 0) = 0,$$

$$D_1 f(0, 0) = 1, \qquad D_2 f(0, 0) = 2,$$

$$D_1^2 f(0, 0) = -1, \qquad D_2^2 f(0, 0) = -4,$$

$$D_1 D_2 f(0, 0) = -2.$$

Hence the polynomial approximation of f up to degree 2 is

$$G(x, y) = x + 2y - \tfrac{1}{2}(x^2 + 4xy + 4y^2).$$

Example 3. In some special cases, there is a way of getting the polynomial approximation to the function more simply by using more directly the Taylor formula for one variable. Consider for instance the function $f(x, y) = \sin(xy)$. For any number u we know that

$$\sin u = u + R_3(u)$$

where $R_3(u)$ is the remainder of the Taylor formula for the sine function of one variable. From the *First Course in Calculus*, you should know

that this remainder term satisfies the estimate

$$|R_3(u)| \leqq \frac{|u|^3}{3!}.$$

Thus if we let $u = xy$, then we find that

$$\sin(xy) = xy + R_3(xy),$$

and hence

$$|\sin(xy) - xy| \leqq \frac{|xy|^3}{3!}.$$

Also we see that, for example, for $y \neq 0$, we have

$$\frac{\sin(xy) - xy}{y} = \frac{R_3(xy)}{y}.$$

In particular, we get the estimate

$$\left|\frac{\sin(xy) - xy}{y}\right| \leqq \frac{|x|^3|y|^2}{3!}.$$

From this we see that the limit of

$$\frac{\sin(xy) - xy}{y}$$

as (x, y) approaches $(0, 0)$ is equal to 0.

Example 4. We can find the polynomial approximation of the function in Example 2 by this method. We know from the theory of the logarithm in one variable that

$$\log(1 + u) = u - \tfrac{1}{2}u^2 + \text{terms of higher degree}.$$

Hence putting $u = x + 2y$, we find that

$$\log(1 + x + 2y) = x + 2y - \tfrac{1}{2}(x + 2y)^2 + \text{terms of higher degree}$$

and therefore the polynomial approximation of our function up to degree 2 is given by

$$G(x, y) = x + 2y - \tfrac{1}{2}(x^2 + 4xy + 4y^2),$$

which is an easier way, rather than finding all the partials, because we could just use the theory of functions of one variable.

Finally, we observe that the treatment of functions of several variables follows exactly the same pattern. In this case, we let

$$H = (h_1, \ldots, h_n)$$

and

$$H \cdot \nabla = h_1 D_1 + \cdots + h_n D_n = h_1 \frac{\partial}{\partial x_1} + \cdots + h_n \frac{\partial}{\partial x_n} \cdot$$

Not a single word need be changed in Theorems 2 and 3 to get Taylor's formula for several variables.

Exercises

Find the terms up to order 2 in the Taylor formula of the following functions (taking $P = O$).

1. $\sin(xy)$
2. $\cos(xy)$
3. $\log(1 + xy)$
4. $\sin(x^2 + y^2)$
5. e^{x+y}
6. $\cos(x^2 + y)$
7. $(\sin x)(\cos y)$
8. $e^x \sin y$
9. $x + xy + 2y^2$
10. Does $\dfrac{\sin(xy)}{x}$ approach a limit as (x, y) approaches $(0, 0)$? If so, what limit?
11. Same questions for

$$\frac{e^{xy} - 1}{x} \quad \text{and} \quad \frac{\log(1 + x^2 + y^2)}{x^2 + y^2}.$$

12. Same questions for

$$\frac{\cos(xy) - 1}{x}.$$

13. Same questions for

$$\frac{\sin(xy) - xy}{x^2 y}.$$

14. Find the terms up to order 3 in Taylor's formula for the function $e^x \cos y$.
15. What is the term of degree 7 in Taylor's formula for the function

$$x^3 - 2xy^4 + (x - 1)^2 y^{10}?$$

16. Show that if $f(x, y, z)$ is a polynomial in x, y, z, then it is equal to its own Taylor series, i.e. there exists an integer n such that $R_n = 0$.
17. Find the polynomial approximation of the function

$$f(x, y) = \log(1 + x + 2y)$$

up to degree 3.

18. In each one of Exercises 1 through 9, find the terms of degree $\leqq 2$ in the Taylor expansion of the function at the indicated point.

1. $P = (1, \pi)$	2. $P = (1, \pi)$	3. $P = (2, 3)$
4. $P = (\sqrt{\pi}, \sqrt{\pi})$	5. $P = (1, 2)$	6. $P = (0, \pi)$
7. $P = (\pi/2, \pi)$	8. $P = (2, \pi/4)$	9. $P = (1, 1)$

19. Let f be a function of two variables with continuous partial derivatives of order $\leqq 2$. Assume that $f(O) = 0$ and also that $f(ta, tb) = t^2 f(a, b)$ for all numbers t and all vectors (a, b). Show that for all points $P = (a, b)$ we have
$$f(P) = \frac{(P \cdot \nabla)^2 f(O)}{2!}.$$

20. Let U be an open set having the following property. Given two points X, Y in U, the line segment joining X and Y is contained in the open set.
(a) What is the parametric equation for this line segment?
(b) Let f have continuous partial derivatives in U. Assume that
$$\|\operatorname{grad} f(P)\| \leqq M$$
for some number M, and all points P in U. Show that for any two points X, Y in U we have
$$|f(X) - f(Y)| \leqq M\|X - Y\|.$$

§4. *Integral expressions*

Quite often, instead of the mean value type of remainder obtained previously in Taylor's formula, it is useful to deal with an integral form of the remainder. For instance, we have

$$f(x, y) = f(0, 0) + \int_0^1 \frac{d}{dt}(f(tx, ty))\, dt. \tag{1}$$

This is a direct application of the definition of the integral, since we can put $\psi(t) = f(tx, ty)$, and since
$$\int_0^1 \frac{d\psi}{dt}\, dt = \psi(1) - \psi(0).$$

If we now evaluate the derivative with respect to t under the integral, using the chain rule, we obtain
$$f(x, y) = f(0, 0) + \int_0^1 [D_1 f(tx, ty)x + D_2 f(tx, ty)y]\, dt$$
or
$$\boxed{f(x, y) = f(0, 0) + x g_1(x, y) + y g_2(x, y),}$$

where

$$g_1(x, y) = \int_0^1 D_1 f(tx, ty)\, dt \qquad \text{and} \qquad g_2(x, y) = \int_0^2 D_2 f(tx, ty)\, dt.$$

The advantage of such an expression is that the dependence of g_1 and g_2 on (x, y) is quite smooth—just as smooth as that of $D_1 f$ and $D_2 f$. From Chapter V, §2 we know that we can differentiate under the integral sign with respect to x and y. Thus this type of expression is often better than the remainder of Taylor's formula, with an undetermined τ which usually cannot be given explicitly, and depends on the specific choice of (x, y).

Exercise

1. Let f be a function of two variables, with continuous partials of order $\leqq 2$. Assume that $f(0, 0) = 0$ and $D_i f(0, 0) = 0$ for $i = 1, 2$. Show that there exist continuous functions h_i such that

$$f(x, y) = h_1(x, y)x^2 + h_2(x, y)xy + h_3(x, y)y^2.$$

[*Hint:* Apply the arguments of the text to the functions

$$D_1 f(tx, ty) \qquad \text{and} \qquad D_2 f(tx, ty)$$

using an integral with respect to some new variable, say s.]

CHAPTER VII

Maximum and Minimum

When we studied functions of one variable, we found maxima and minima by first finding critical points, i.e. points where the derivative is equal to 0, and then determining by inspection which of these are maxima or minima. We can carry out a similar investigation for functions of several variables. The condition that the derivative is equal to 0 must be replaced by the vanishing of all partial derivatives.

§1. *Critical points*

Let f be a differentiable function defined on an open set U. Let P be a point of U. If all partial derivatives of f are equal to 0 at P, then we say that P is a **critical point** of the function. In other words, for P to be a critical point, we must have

$$D_1 f(P) = 0, \quad \dots, \quad D_n f(P) = 0.$$

Example. Find the critical points of the function $f(x, y) = e^{-(x^2+y^2)}$. Taking the partials, we see that

$$\frac{\partial f}{\partial x} = -2xe^{-(x^2+y^2)} \quad \text{and} \quad \frac{\partial f}{\partial y} = -2ye^{-(x^2+y^2)}.$$

The only value of (x, y) for which both these quantities are equal to 0 is $x = 0$ and $y = 0$. Hence the only critical point is $(0, 0)$.

A critical point of a function of one variable is a point where the derivative is equal to 0. We have seen examples where such a point need not be a local maximum or a local minimum, for instance as in the following picture:

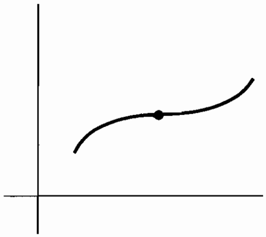


Figure 1

A fortiori, a similar thing may occur for functions of several variables. However, once we have found critical points, it is usually not too difficult to tell by inspection whether they are of this type or not.

Let f be any function (differentiable or not), defined on an open set U. We shall say that a point P of U is a **local maximum** for the function if there exists an open ball (of positive radius) B, centered at P, such that for all points X of B, we have

$$f(X) \leqq f(P).$$

As an exercise, define **local minimum** in an analogous manner.

In the case of functions of one variable, we took an open interval instead of an open ball around the point P. Thus our notion of local maximum in n-space is the natural generalization of the notion in 1-space.

Theorem 1. *Let f be a function which is defined and differentiable on an open set U. Let P be a local maximum for f in U. Then P is a critical point of f.*

Proof. The proof is exactly the same as for functions of one variable. In fact, we shall prove that the directional derivative of f at P in any direction is 0. Let H be a non-zero vector. For small values of t, $P + tH$ lies in the open set U, and $f(P + tH)$ is defined. Furthermore, for small values of t, tH is small, and hence $P + tH$ lies in our open ball such that

$$f(P + tH) \leqq f(P).$$

Hence the function of one variable $g(t) = f(P + tH)$ has a local maximum at $t = 0$. Hence its derivative $g'(0)$ is equal to 0. By the chain rule, we obtain as usual:

$$\operatorname{grad} f(P) \cdot H = 0.$$

This equation is true for every non-zero vector H, and hence

$$\operatorname{grad} f(P) = O.$$

This proves what we wanted.

Exercises

Find the critical points of the following functions.

1. $x^2 + 4xy - y^2 - 8x - 6y$
2. $x + y \sin x$
3. $x^2 + y^2 + z^2$
4. $(x + y)e^{-xy}$
5. $xy + xz$
6. $\cos(x^2 + y^2 + z^2)$
7. x^2y^2
8. $x^4 + y^2$
9. $(x - y)^4$
10. $x \sin y$

11. $x^2 + 2y^2 - x$

12. $e^{-(x^2+y^2+z^2)}$

13. $e^{(x^2+y^2+z^2)}$

14. In each of the preceding exercises, find the minimum value of the given function, and give all points where the value of the function is equal to this minimum. [*Do this exercise after you have read* §3.]

§2. *The quadratic form*

Let f be a differentiable function on an open set U, and assume that all partial derivatives up to order 3 exist and are continuous. Let P be a point of U, and assume that P is a critical point of f. We assume that we work in 2-space, so that we can express f near the point $P = (a, b)$ in the form

$$\begin{aligned} f(a+h, b+k) = {} & f(a, b) \\ & + \frac{1}{2}\left[h^2 \frac{\partial^2 f}{\partial x^2}(a, b) + 2hk \frac{\partial^2 f}{\partial x\, \partial y}(a, b) + k^2 \frac{\partial^2 f}{\partial y^2}(a, b)\right] \\ & + R_3, \end{aligned}$$

where R_3 is a remainder term. Actually, we prefer to write in the other notation:

$$\begin{aligned} f(a+h, b+k) = {} & f(a, b) \\ & + \tfrac{1}{2}[h^2 D_1^2 f(a, b) + 2hk D_1 D_2 f(a, b) + k^2 D_2^2 f(a, b)] \\ & + R_3 \end{aligned}$$

because we want to use x, y for other purposes.

The **function** $q(x, y)$ of x, y given by

$$q(x, y) = \tfrac{1}{2}[x^2 D_1^2 f(P) + 2xy D_1 D_2 f(P) + y^2 D_2^2 f(P)]$$

is called the **quadratic form** associated with f at P, whenever P is a critical point of f. This quadratic form approximates the values of f so that one gets some general idea of the behavior of f near P when the terms of degree 1 vanish.

Example 1. Let $f(x, y) = e^{-(x^2+y^2)}$. Then it is a simple matter to verify that

$$\operatorname{grad} f(0, 0) = 0.$$

We let $P = (0, 0)$ be the origin. Standard computations show that

$$D_1^2 f(O) = -2, \qquad D_1 D_2 f(O) = 0, \qquad D_2^2 f(O) = -2.$$

Substituting these values in the general formula gives the expression for

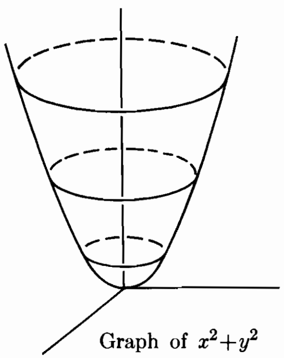
the quadratic form, namely

$$q(x, y) = -(x^2 + y^2).$$

We see that the quadratic form is nothing but the term of degree 2 in the Taylor expansion of the function at the given point.

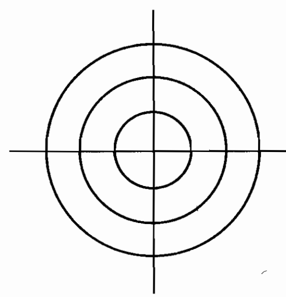
We shall now describe the level curves for some quadratic forms to get an idea of their behavior near the origin.

Example 2. $q(x, y) = x^2 + y^2$. Then a graph of the function q and the level curves look like those in Figs. 2 and 3.



Graph of x^2+y^2

Figure 2



Level curves

Figure 3

In this example, we see that the origin (0, 0) is a local minimum point for the form.

Example 3. $q(x, y) = -(x^2 + y^2)$. The graph and level curves look like these:

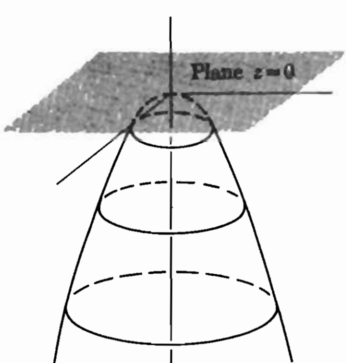


Figure 4

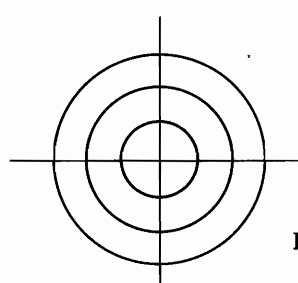


Figure 5

The origin is a local maximum for the form.

Example 4. $q(x, y) = x^2 - y^2$. The level curves are then hyperbolas, determined for each number c by the equation $x^2 - y^2 = c$:

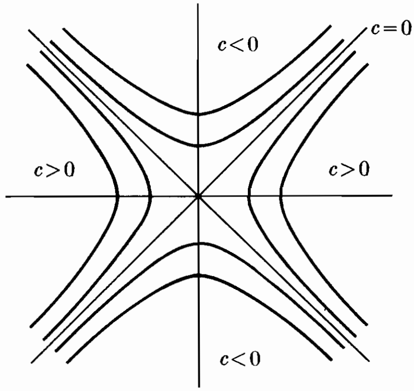


Figure 6

Of course, when $c = 0$, we get the two straight lines as shown.

Example 5. $q(x, y) = xy$. The level curves look like the following (similar to the preceding example, but turned around):

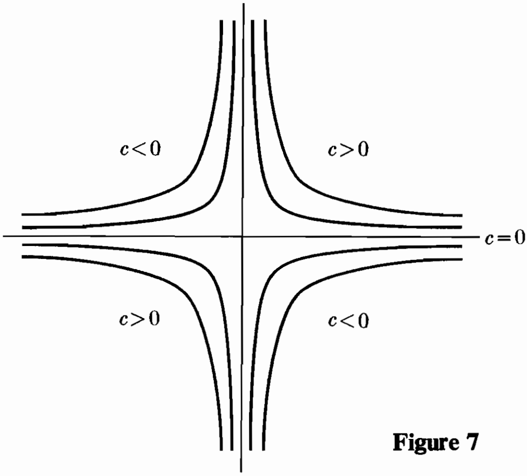


Figure 7

In Examples 4 and 5, we see that the origin, which is a critical point, is neither a local maximum nor local minimum.

Exercises

1. Let $f(x, y) = 3x^2 - 4xy + y^2$. Show that the origin is a critical point of f.
2. More generally, let a, b, c be numbers. Show that the function f given by $f(x, y) = ax^2 + bxy + cy^2$ has the origin as a critical point.

3. Find the quadratic form associated to the function f at the critical points P in the Exercises of §1.
4. Sketch the level curves for the following quadratic forms. Determine whether the origin is a local maximum, minimum, or neither.

(a) $q(x, y) = 2x^2 - y^2$ (b) $q(x, y) = 3x^2 + 4y^2$
(c) $q(x, y) = -(4x^2 + 5y^2)$ (d) $q(x, y) = y^2 - x^2$
(e) $q(x, y) = 2y^2 - x^2$ (f) $q(x, y) = y^2 - 4x^2$
(g) $q(x, y) = -(3x^2 + 2y^2)$ (h) $q(x, y) = 2xy$

§3. *Boundary points*

In considering intervals, we had to distinguish between closed and open intervals. We must make an analogous distinction when considering sets of points in space.

Let S be a set of points, in some n-space. Let P be a point of S. We shall say that P is an **interior point** of S if there exists an open ball B of positive radius, centered at P, and such that B is contained in S. The next picture illustrates an interior point (for the set consisting of the region enclosed by the curve).

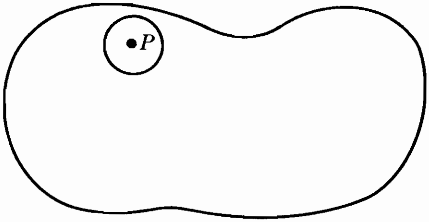


Figure 8

We have also drawn an open ball around P.

From the very definition, we conclude that the set consisting of all interior points of S is an open set.

A point P (not necessarily in S) is called a **boundary point** of S if every open ball B centered at P includes a point of S, and also a point which is not in S. We illustrate a boundary point in the following picture:

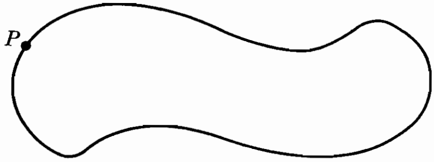


Figure 9

For example, the set of boundary points of the closed ball of radius $a > 0$ is the sphere of radius a. In 2-space, the plane, the region consisting of all points with $y > 0$ is open. Its boundary points are the points lying on the x-axis.

If a set contains all of its boundary points, then we shall say that the set is **closed.**

Finally, a set is said to be **bounded** if there exists a number $b > 0$ such that, for every point X of the set, we have

$$\|X\| \leqq b.$$

We are now in a position to state the existence of maxima and minima for continuous functions.

Theorem 2. *Let S be a closed and bounded set. Let f be a continuous function defined on S. Then f has a maximum and a minimum in S. In other words, there exists a point P in S such that*

$$f(P) \geqq f(X)$$

for all X in S, and there exists a point Q in S such that

$$f(Q) \leqq f(X)$$

for all X in S.

We shall not prove this theorem. It depends on an analysis which is beyond the level of this course.

When trying to find a maximum (say) for a function f, one should first determine the critical points of f in the interior of the region under consideration. If a maximum lies in the interior, it must be among these critical points.

Next, one should investigate the function on the boundary of the region. By parametrizing the boundary, one frequently reduces the problem of finding a maximum on the boundary to a lower-dimensional problem, to which the technique of critical points can also be applied.

Finally, one has to compare the possible maximum of f on the boundary and in the interior to determine which points are maximum points.

Example. In the Example in §1, we observe that the function

$$f(x, y) = e^{-(x^2+y^2)}$$

becomes very small as x or y becomes large. Consider some big closed disc centered at the origin. We know by Theorem 2 that the function has a maximum in this disc. Since the value of the function is small on the boundary, it follows that this maximum must be an interior point, and hence that the maximum is a critical point. But we found in the Example in §1 that the only critical point is at the origin. Hence we conclude that the origin is *the* only maximum of the function $f(x, y)$. The value of f at the origin is $f(0, 0) = 1$. Furthermore, the function has no minimum, because $f(x, y)$ is always positive and approaches 0 as x and y become large.

Exercises

Find the maximum and minimum points of the following functions in the indicated region.

1. $x + y$ in the square with corners at $(\pm 1, \pm 1)$
2. (a) $x + y + z$ in the region $x^2 + y^2 + z^2 < 1$
   (b) $x + y$ in the region $x^2 + y^2 < 1$
3. $xy - (1 - x^2 - y^2)^{1/2}$ in the region $x^2 + y^2 \leqq 1$
4. $144x^3y^2(1 - x - y)$ in the region $x \geqq 0$ and $y \geqq 0$ (the first quadrant together with its boundary)
5. $(x^2 + 2y^2)e^{-(x^2+y^2)}$ in the plane
6. (a) $(x^2 + y^2)^{-1}$ in the region $(x - 2)^2 + y^2 \leqq 1$
   (b) $(x^2 + y^2)^{-1}$ in the region $x^2 + (y - 2)^2 \leqq 1$
7. Which of the following functions have a maximum and which have a minimum in the whole plane?

   (a) $(x + 2y)e^{-x^2-y^4}$ (b) e^{x-y}

   (c) $e^{x^2-y^2}$ (d) $e^{x^2+y^{10}}$

   (e) $(3x^2 + 2y^2)e^{-(4x^2+y^2)}$ (f) $-x^2e^{x^4+y^{10}}$

   (g) $\begin{cases} \dfrac{x^2 + y^2}{|x| + |y|} & \text{if} \quad (x, y) \neq (0, 0) \\ 0 & \text{if} \quad (x, y) = (0, 0) \end{cases}$
8. Which is the point on the curve $(\cos t, \sin t, \sin(t/2))$ farthest from the origin?

§4. *Lagrange multipliers*

In this section, we shall investigate another method for finding the maximum or minimum of a function on some set of points. This method is particularly well adapted to the case when the set of points is described by means of an equation.

We shall work in 3-space. Let g be a differentiable function of three variables x, y, z. We consider the surface

$$g(X) = 0.$$

Let U be an open set containing this surface, and let f be a differentiable function defined for all points of U. We wish to find those points P on the surface $g(X) = 0$ such that $f(P)$ is a maximum or a minimum on the surface. In other words, we wish to find all points P such that $g(P) = 0$, and either

$$f(P) \geqq f(X) \quad \text{for all } X \text{ such that } g(X) = 0,$$

or

$$f(P) \leqq f(X) \quad \text{for all } X \text{ such that } g(X) = 0.$$

Any such point will be called an **extremum for f subject to the constraint g.**

In what follows, we consider only points P such that $g(P) = 0$ but $\operatorname{grad} g(P) \neq O$

Theorem 3. *Let g be a continuously differentiable function on an open set U. Let S be the set of points X in U such that $g(X) = 0$ but*

$$\operatorname{grad} g\,(X) \neq O.$$

Let f be a continuously differentiable function on U and assume that P is a point of S such that P is an extremum for f on S. (In other words, P is an extremum for f, subject to the constraint g.) Then there exists a number λ such that

$$\operatorname{grad} f\,(P) = \lambda \operatorname{grad} g\,(P).$$

Proof. Let $X: J \to S$ be a differentiable curve on the surface S passing through P, say $X(t_0) = P$. Then the function $t \mapsto f(X(t))$ has a maximum or a minimum at t_0. Its derivative

$$\frac{d}{dt} f(X(t))$$

is therefore equal to 0 at t_0. But this derivative is equal to

$$\operatorname{grad} f\,(P) \cdot X'(t_0) = 0.$$

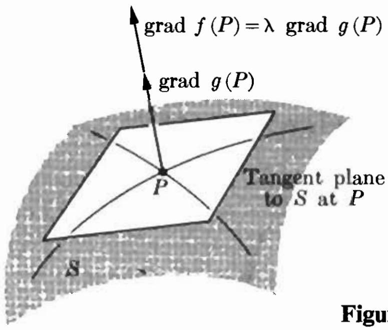


Figure 10

Hence $\operatorname{grad} f\,(P)$ is perpendicular to every curve on the surface passing through P (Fig. 10). Under these circumstances, and the hypothesis that $\operatorname{grad} g(P) \neq O$, there exists a number λ such that

$$(1) \qquad \operatorname{grad} f\,(P) = \lambda \operatorname{grad} g\,(P),$$

or in other words, $\operatorname{grad} f(P)$ has the same, or opposite direction, as $\operatorname{grad} g(P)$, provided it is not O. This is rather clear, since the direction of $\operatorname{grad} g(P)$ is the direction perpendicular to the surface, and we have seen that $\operatorname{grad} f(P)$ is also perpendicular to the surface.

Conversely, when we want to find an extremum point for f subject to the constraint g, we find all points P such that $g(P) = 0$, and such that relation (1) is satisfied. We can then find our extremum points among these by inspection.

(Note that this procedure is analogous to the procedure used to find maxima or minima for functions of one variable. We first determined all points at which the derivative is equal to 0, and then determined maxima or minima by inspection.)

Example 1. Find the maximum of the function $f(x, y) = x + y$ subject to the constraint $x^2 + y^2 = 1$.

We let $g(x, y) = x^2 + y^2 - 1$, so that S consists of all points (x, y) such that $g(x, y) = 0$. We have

$$\begin{aligned} \operatorname{grad} f(x, y) &= (1, 1), \\ \operatorname{grad} g(x, y) &= (2x, 2y). \end{aligned}$$

Let (x_0, y_0) be a point for which there exists a number λ satisfying

$$\operatorname{grad} f(x_0, y_0) = \lambda \operatorname{grad} g(x_0, y_0),$$

or in other words

$$1 = 2x_0\lambda \quad \text{and} \quad 1 = 2y_0\lambda.$$

Then $x_0 \neq 0$ and $y_0 \neq 0$. Hence $\lambda = 1/2x_0 = 1/2y_0$, and consequently $x_0 = y_0$. Since the point (x_0, y_0) must satisfy the equation $g(x_0, y_0) = 0$, we get the possibilities:

$$x_0 = \pm \frac{1}{\sqrt{2}} \quad \text{and} \quad y_0 = \pm \frac{1}{\sqrt{2}}\cdot$$

It is then clear that $(1/\sqrt{2}, 1/\sqrt{2})$ is a maximum for f since the only other possibility $(-1/\sqrt{2}, -1/\sqrt{2})$ is a point at which f takes on a negative value, and $f(1/\sqrt{2}, 1/\sqrt{2}) = 2/\sqrt{2} > 0$.

Example 2. Find the extrema for the function $x^2 + y^2 + z^2$ subject to the constraint $x^2 + 2y^2 - z^2 - 1 = 0$.

Computing the partial derivatives of the functions f and g, we find that we must solve the system of equations

(a) $2x = \lambda \cdot 2x,$ (b) $2y = \lambda \cdot 4y,$
(c) $2z = \lambda \cdot (-2z),$ (d) $g(X) = x^2 + 2y^2 - z^2 - 1 = 0.$

Let (x_0, y_0, z_0) be a solution. If $z_0 \neq 0$, then from (c) we conclude that $\lambda = -1$. The only way to solve (a) and (b) with $\lambda = -1$ is that $x = y = 0$. In that case, from (d), we would get

$$z_0^2 = -1,$$

which is impossible. Hence any solution must have $z_0 = 0$.

If $x_0 \neq 0$, then from (a) we conclude that $\lambda = 1$. From (b) and (c) we then conclude that $y_0 = z_0 = 0$. From (d), we must have $x_0 = \pm 1$. In this manner, we have obtained two solutions satisfying our conditions, namely

$$(1, 0, 0) \quad \text{and} \quad (-1, 0, 0).$$

Similarly, if $y_0 \neq 0$, we find two more solutions, namely

$$(0, \sqrt{\tfrac{1}{2}}, 0) \quad \text{and} \quad (0, -\sqrt{\tfrac{1}{2}}, 0).$$

These four points are therefore the extrema of the function f subject to the constraint g.

If we ask for the minimum of f, then a direct computation shows that the last two points

$$(0, \pm\sqrt{\tfrac{1}{2}}, 0)$$

are the only possible solutions (because $1 > \frac{1}{2}$).

Exercises

1. (a) Find the minimum of the function $x + y^2$ subject to the constraint $2x^2 + y^2 = 1$.
   (b) Find its maximum.
2. Find the maximum value of $x^2 + xy + y^2 + yz + z^2$ on the sphere of radius 1.
3. Let $A = (1, 1, -1)$, $B = (2, 1, 3)$, $C = (2, 0, -1)$. Find the point at which the function

$$f(X) = (X - A)^2 + (X - B)^2 + (X - C)^2$$

reaches its minimum, and find the minimum value.

4. Do Exercise 3 in general, for any three distinct vectors

$$A = (a_1, a_2, a_3), \qquad B = (b_1, b_2, b_3), \qquad C = (c_1, c_2, c_3).$$

5. Find the maximum of the function $3x^2 + 2\sqrt{2}\,xy + 4y^2$ on the circle of radius 3 in the plane.

6. Find the maximum of the functions xyz subject to the constraints $x \geqq 0$, $y \geqq 0$, $z \geqq 0$, and $xy + yz + xz = 2$.

7. Find the maximum and minimum distance from points on the curve

$$5x^2 + 6xy + 5y^2 = 0$$

to the origin in the plane.

8. Find the extreme values of the function $\cos^2 x + \cos^2 y$ subject to the constraint $x - y = \pi/4$ and $0 \leqq x \leqq \pi$.

9. Find the points on the surface $z^2 - xy = 1$ nearest to the origin.

10. Find the extreme values of the function xy subject to the condition $x + y = 1$.

11. Find the shortest distance between the point $(1, 0)$ and the curve $y^2 = 4x$.

12. Find the maximum and minimum points of the function

$$f(x, y, z) = x + y + z$$

in the region $x^2 + y^2 + z^2 \leqq 1$.

13. Find the extremum values of the function $f(x, y, z) = x - 2y + 2z$ on the sphere $x^2 + y^2 + z^2 = 1$.

14. Find the maximum of the function $f(x, y, z) = x + y + z$ on the sphere $x^2 + y^2 + z^2 = 4$.

15. Find the extreme values of the function f given by $f(x, y, z) = xyz$ subject to the condition $x + y + z = 1$.

16. Find the extreme values of the function given by $f(x, y, z) = (x + y + z)^2$ subject to the condition $x^2 + 2y^2 + 3z^2 = 1$.

17. Find the minimum of the function $f(x, y, z) = x^2 + y^2 + z^2$ subject to the condition $3x + 2y - 7z = 5$.

18. In general, if a, b, c, d are numbers with not all of a, b, c equal to 0, find the minimum of the function $x^2 + y^2 + z^2$ subject to the condition

$$ax + by + cz = d.$$

19. Find the maximum and minimum value of the function

$$f(x, y) = x^2 + 2y^2 - x$$

on the disc of radius 1 centered at the origin.

20. Find the shortest distance from a point on the ellipse $x^2 + 4y^2 = 4$ to the line $x + y = 4$.

PART TWO

MATRICES, LINEAR MAPS, AND DETERMINANTS

The present book is organized by topics. This means that we may go deeper in one part than is necessary to read another part. Certain portions may therefore be omitted without impairing the understanding of later portions.

Specifically, the reader may omit this entire part, and go immediately to the chapter on multiple integrals, which can be read after Chapter I. In a one-term course, you can cover Green's theorem also without any knowledge of matrices or determinants.

The amount of "linear algebra" discussed here is kept at a minimum, and is intended for applications to the Jacobian matrix in Chapter XI, and to the surface or volume integrals occurring in Chapter XV. Thus you may postpone this part until you read those sections. These applications describe in various contexts how an arbitrary mapping can be approximated by a linear mapping.

CHAPTER VIII

Matrices

§1. Matrices

We consider a new kind of object, matrices.

Let n, m be two integers $\geqq 1$. An array of numbers

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{pmatrix}$$

is called a **matrix.** We can abbreviate the notation for this matrix by writing it (a_{ij}), $i = 1, \ldots, m$ and $j = 1, \ldots, n$. We say that it is an m by n matrix, or an $m \times n$ matrix. The matrix has m **rows** and n **columns.** For instance, the first column is

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}$$

and the second row is $(a_{21}, a_{22}, \ldots, a_{2n})$. We call a_{ij} the ij-**entry** or ij-**component** of the matrix.

If you look back at Chapter I, §1, the example of 7-space taken from economics gives rise to a 7×7 matrix (a_{ij}) $(i, j = 1, \ldots, 7)$, where a_{ij} is the amount spent by the i-th industry on the j-th industry. Thus keeping the notation of that example, if $a_{25} = 50$, this means that the auto industry bought 50 million dollars' worth of stuff from the chemical industry during the given year.

Example 1. The following is a 2×3 matrix:

$$\begin{pmatrix} 1 & 1 & -2 \\ -1 & 4 & -5 \end{pmatrix}.$$

It has two rows and three columns.

The rows are $(1, 1, -2)$ and $(-1, 4, -5)$. The columns are

$$\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \quad \begin{pmatrix} -2 \\ -5 \end{pmatrix}.$$

Thus the rows of a matrix may be viewed as n-tuples, and the columns may be viewed as vertical m-tuples. A vertical m-tuple is also called a **column vector.**

A vector $(x_1, \ldots, x_n)$ is a $1 \times n$ matrix. A column vector

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

is an $n \times 1$ matrix.

When we write a matrix in the form (a_{ij}), then i denotes the row and j denotes the column. In Example 1, we have for instance $a_{11} = 1$, $a_{23} = -5$.

A single number (a) may be viewed as a 1×1 matrix.

Let (a_{ij}), $i = 1, \ldots, m$ and $j = 1, \ldots, n$ be a matrix. If $m = n$, then we say that it is a **square** matrix. Thus

$$\begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & -1 & 5 \\ 2 & 1 & -1 \\ 3 & 1 & -1 \end{pmatrix}$$

are both square matrices.

We have a **zero matrix,** in which $a_{ij} = 0$ for all i, j. It looks like this:

$$\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

We shall write it O. We note that we have met so far with the zero number, zero vector, and zero matrix.

We shall now define addition of matrices and multiplication of matrices by numbers.

We define addition of matrices only when they have the same size. Thus let m, n be fixed integers $\geqq 1$. Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $m \times n$ matrices. We define $A + B$ to be the matrix whose entry in the i-th row and j-th column is $a_{ij} + b_{ij}$. In other words, we add matrices of the same size componentwise.

Example 2. Let

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 3 & 4 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 5 & 1 & -1 \\ 2 & 1 & -1 \end{pmatrix}$$

Then

$$A + B = \begin{pmatrix} 6 & 0 & -1 \\ 4 & 4 & 3 \end{pmatrix}.$$

If A, B are both $1 \times n$ matrices, i.e. n-tuples, then we note that our addition of matrices coincides with the addition which we defined in Chapter I for n-tuples.

If O is the zero matrix, then for any matrix A (of the same size, of course), we have $O + A = A + O = A$. This is trivially verified.

We shall now define the multiplication of a matrix by a number. Let c be a number, and $A = (a_{ij})$ be a matrix. We define cA to be the matrix whose ij-component is ca_{ij}. We write $cA = (ca_{ij})$. Thus we multiply each component of A by c.

Example 3. Let A, B be as in Example 2. Let $c = 2$. Then

$$2A = \begin{pmatrix} 2 & -2 & 0 \\ 4 & 6 & 8 \end{pmatrix} \quad \text{and} \quad 2B = \begin{pmatrix} 10 & 2 & -2 \\ 4 & 2 & -2 \end{pmatrix}.$$

For any matrix A we let $-A$ be the matrix obtained by multiplying each component of A with -1. If $A = (a_{ij})$, then

$$-A = (-1)A = (-a_{ij}).$$

For instance, if

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 3 & 4 \end{pmatrix}$$

is the matrix of Example 2, then

$$-A = (-1)A = \begin{pmatrix} -1 & 1 & 0 \\ -2 & -3 & -4 \end{pmatrix}.$$

Observe that for any matrix A we have

$$A + (-A) = A - A = 0.$$

We define one more notion related to a matrix. Let $A = (a_{ij})$ be an $m \times n$ matrix. The $n \times m$ matrix $B = (b_{ji})$ such that $b_{ji} = a_{ij}$ is called the **transpose** of A, and is also denoted by ${}^tA$. Taking the transpose of a matrix amounts to changing rows into columns and vice versa. If A is the matrix which we wrote down at the beginning of this section, then ${}^tA$ is the matrix

$$\begin{pmatrix} a_{11} & a_{21} & a_{31} & \cdots & a_{m1} \\ a_{12} & a_{22} & a_{32} & \cdots & a_{m2} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & a_{3n} & \cdots & a_{mn} \end{pmatrix}.$$

To take a special case:

$$\text{If } A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 3 & 5 \end{pmatrix}, \qquad \text{then} \qquad {}^tA = \begin{pmatrix} 2 & 1 \\ 1 & 3 \\ 0 & 5 \end{pmatrix}.$$

If $A = (2, 1, -4)$ is a **row vector,** then

$${}^tA = \begin{pmatrix} 2 \\ 1 \\ -4 \end{pmatrix}$$

is a **column vector.**

Exercises

1. Let

$$A = \begin{pmatrix} 1 & 2 & 3 \\ -1 & 0 & 2 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} -1 & 5 & -2 \\ 1 & 1 & -1 \end{pmatrix}.$$

Find $A + B$, $3B$, $-2B$, $A + 2B$, $2A + B$, $A - B$, $A - 2B$, $B - A$.

2. Let

$$A = \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} -1 & 1 \\ 0 & -3 \end{pmatrix}.$$

Find $A + B$, $3B$, $-2B$, $A + 2B$, $A - B$, $B - A$.

3. In Exercise 1, find ${}^tA$ and ${}^tB$.
4. In Exercise 2, find ${}^tA$ and ${}^tB$.
5. If A, B are arbitrary $m \times n$ matrices, show that

$${}^t(A + B) = {}^tA + {}^tB.$$

6. If c is a number, show that ${}^t(cA) = c\,{}^tA$.

7. If $A = (a_{ij})$ is a square matrix, then the elements a_{ii} are called the **diagonal** elements. How do the diagonal elements of A and ${}^tA$ differ?
8. Find ${}^t(A + B)$ and ${}^tA + {}^tB$ in Exercise 2.
9. Find $A + {}^tA$ and $B + {}^tB$ in Exercise 2.
10. A matrix A is said to be **symmetric** if $A = {}^tA$. Show that for any square matrix A, the matrix $A + {}^tA$ is symmetric.
11. Write down the row vectors and column vectors of the matrices A, B in Exercise 1.
12. Write down the row vectors and column vectors of the matrices A, B in Exercise 2.

§2. *Multiplication of matrices*

We shall now define the product of matrices. Let $A = (a_{ij})$, $i = 1, \dots, m$ and $j = 1, \dots, n$ be an $m \times n$ matrix. Let $B = (b_{jk})$, $j = 1, \dots, n$ and $k = 1, \dots, s$ be an $n \times s$ matrix.

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \qquad B = \begin{pmatrix} b_{11} & \cdots & b_{1s} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{ns} \end{pmatrix}.$$

We define the product AB to be the $m \times s$ matrix whose ik-coordinate is

$$\sum_{j=1}^{n} a_{ij}b_{jk} = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk}.$$

If $A_1, \dots, A_m$ are the row vectors of the matrix A, and if $B^1, \dots, B^s$ are the column vectors of the matrix B, then the ik-coordinate of the product AB is equal to $A_i \cdot B^k$. Thus

$$AB = \begin{pmatrix} A_1 \cdot B^1 & \cdots & A_1 \cdot B^s \\ \vdots & & \vdots \\ A_m \cdot B^1 & \cdots & A_m \cdot B^s \end{pmatrix}.$$

Multiplication of matrices is therefore a generalization of the dot product.

Example 1. Let

$$A = \begin{pmatrix} 2 & 1 & 5 \\ 1 & 3 & 2 \end{pmatrix}, \qquad B = \begin{pmatrix} 3 & 4 \\ -1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Then AB is a 2×2 matrix, and computations show that

$$AB = \begin{pmatrix} 2 & 1 & 5 \\ 1 & 3 & 2 \end{pmatrix} \begin{pmatrix} 3 & 4 \\ -1 & 2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 15 & 15 \\ 4 & 12 \end{pmatrix}.$$

Example 2. Let

$$C = \begin{pmatrix} 1 & 3 \\ -1 & -1 \end{pmatrix}.$$

Let A, B be as in Example 1. Then

$$BC = \begin{pmatrix} 3 & 4 \\ -1 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 3 \\ -1 & -1 \end{pmatrix} = \begin{pmatrix} -1 & 5 \\ -3 & -5 \\ 1 & 5 \end{pmatrix}$$

and

$$A(BC) = \begin{pmatrix} 2 & 1 & 5 \\ 1 & 3 & 2 \end{pmatrix}\begin{pmatrix} -1 & 5 \\ -3 & -5 \\ 1 & 5 \end{pmatrix} = \begin{pmatrix} 0 & 30 \\ -8 & 0 \end{pmatrix}.$$

Compute $(AB)C$. What do you find?

Let A be an $m \times n$ matrix and let B be an $n \times 1$ matrix, i.e. a column vector. Then AB is again a column vector. The product looks like this:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}$$

where

$$c_i = \sum_{j=1}^{n} a_{ij}b_j = a_{i1}b_1 + \cdots + a_{in}b_n.$$

If $X = (x_1, \ldots, x_m)$ is a row vector, i.e. a $1 \times m$ matrix, then we can form the product XA, which looks like this:

$$(x_1, \ldots, x_m)\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = (y_1, \ldots, y_n),$$

where

$$y_k = x_1a_{1k} + \cdots + x_ma_{mk}.$$

In this case, XA is a $1 \times n$ matrix, i.e. a row vector.

If A is a square matrix, then we can form the product AA, which will be a square matrix of the same size as A. It is denoted by A^2. Similarly, we can form A^3, A^4, and in general, A^n for any positive integer n.

We define the unit $n \times n$ matrix to be the matrix having diagonal components all equal to 1, and all other components equal to 0. Thus

the unit $n \times n$ matrix, denoted by I_n, looks like this:

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

We can then define $A^0 = I$ (the unit matrix of the same size as A).

Theorem 1. *Let A, B, C be matrices. Assume that A, B can be multiplied, and A, C can be multiplied, and B, C can be added. Then A, $B + C$ can be multiplied, and we have*

$$A(B + C) = AB + AC.$$

If x is a number, then

$$A(xB) = x(AB).$$

Proof. Let A_i be the i-th row of A, and let B^k, C^k be the k-th column of B and C respectively. Then $B^k + C^k$ is the k-th column of $B + C$. By definition, the ik-component of AB is $A_i \cdot B^k$, the ik-component of AC is $A_i \cdot C^k$, and the ik-component of $A(B + C)$ is $A_i \cdot (B^k + C^k)$. Since

$$A_i \cdot (B^k + C^k) = A_i \cdot B^k + A_i \cdot C^k,$$

our first assertion follows. As for the second, observe that the k-th column of xB is xB^k. Since

$$A_i \cdot xB^k = x(A_i \cdot B^k),$$

our second assertion follows.

Theorem 2. *Let A, B, C be matrices such that A, B can be multiplied and B, C can be multiplied. Then A, BC can be multiplied, so can AB, C, and we have*

$$(AB)C = A(BC).$$

Proof. Let $A = (a_{ij})$ be an $m \times n$ matrix, let $B = (b_{jk})$ be an $n \times r$ matrix, and let $C = (c_{kl})$ be an $r \times s$ matrix. The product AB is an $m \times r$ matrix, whose ik-component is equal to the sum

$$a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk}.$$

We shall abbreviate this sum using our $\sum$ notation by writing

$$\sum_{j=1}^{n} a_{ij}b_{jk}.$$

By definition, the il-component of $(AB)C$ is equal to

$$\sum_{k=1}^{r}\left[\sum_{j=1}^{n} a_{ij}b_{jk}\right]c_{kl} = \sum_{k=1}^{r}\left[\sum_{j=1}^{n} a_{ij}b_{jk}c_{kl}\right].$$

The sum on the right can also be described as the sum of all terms

$$a_{ij}b_{jk}c_{kl},$$

where j, k range over all integers $1 \leqq j \leqq n$ and $1 \leqq k \leqq r$ respectively.

If we had started with the jl-component of BC and then computed the il-component of $A(BC)$ we would have found exactly the same sum, thereby proving the theorem.

Exercises

1. Let I be the unit $n \times n$ matrix. Let A be an $n \times r$ matrix. What is IA? If A is an $m \times n$ matrix, what is AI?
2. Let O be the matrix all of whose coordinates are 0. Let A be a matrix of a size such that the product AO is defined. What is AO?
3. In each one of the following cases, find $(AB)C$ and $A(BC)$.

(a) $A = \begin{pmatrix} 2 & 1 \\ 3 & 1 \end{pmatrix}$, $B = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix}$, $C = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}$

(b) $A = \begin{pmatrix} 2 & 1 & -1 \\ 3 & 1 & 2 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 3 & -1 \end{pmatrix}$, $C = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$

(c) $A = \begin{pmatrix} 2 & 4 & 1 \\ 3 & 0 & -1 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 1 & -1 \\ 3 & 1 & 5 \end{pmatrix}$, $C = \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ -1 & 4 \end{pmatrix}$

4. Let A, B be square matrices of the same size, and assume that $AB = BA$. Show that $(A + B)^2 = A^2 + 2AB + B^2$, and

$$(A + B)(A - B) = A^2 - B^2,$$

using the properties of matrices stated in Theorem 1.

5. Let

$$A = \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}.$$

Find AB and BA.

6. Let

$$C = \begin{pmatrix} 7 & 0 \\ 0 & 7 \end{pmatrix}.$$

Let A, B be as in Exercise 5. Find CA, AC, CB, and BC. State the general rule including this exercise as a special case.

7. Let $X = (1, 0, 0)$ and let

$$A = \begin{pmatrix} 3 & 1 & 5 \\ 2 & 0 & 1 \\ 1 & 1 & 7 \end{pmatrix}.$$

What is XA?

8. Let $X = (0, 1, 0)$, and let A be an arbitrary 3×3 matrix. How would you describe XA? What if $X = (0, 0, 1)$? Generalize to similar statements concerning $n \times n$ matrices, and their products with unit vectors.

9. Let $A = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}$. What is A^2?

10. Let $A = \begin{pmatrix} 0 & 0 \\ -5 & 0 \end{pmatrix}$. What is A^2?

11. (a) Let A be the matrix

$$\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Find A^2, A^3. Generalize to 4×4 matrices.

(b) Let A be the matrix

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Compute A^2, A^3, and A^4.

12. Let X be the indicated column vector, and A the indicated matrix. Find AX as a column vector.

(a) $X = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$, $A = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 2 & 0 & -1 \end{pmatrix}$ (b) $X = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, $A = \begin{pmatrix} 2 & 1 & 5 \\ 0 & 1 & 1 \end{pmatrix}$

(c) $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ (d) $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, $A = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$

13. Let
$$A = \begin{pmatrix} 2 & 1 & 3 \\ 4 & 1 & 5 \end{pmatrix}.$$
Find AX for each of the following values of X.

(a) $X = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ (b) $X = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ (c) $X = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ (d) $X = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$.

14. Let
$$A = \begin{pmatrix} 3 & 7 & 5 \\ 1 & -1 & 4 \\ 2 & 1 & 8 \end{pmatrix}.$$
Find AX for each of the values of X given in Exercise 13.

15. Let
$$X = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad A \quad \begin{pmatrix} a_{11} & \cdots & a_{14} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{m4} \end{pmatrix}.$$
What is AX?

16. Let X be a column vector having all its components equal to 0 except the i-th component which is equal to 1. Let A be an arbitrary matrix, whose size is such that we can form the product AX. What is AX?

17. Let X be a row vector having all its components equal to 0 except the j-th component which is equal to 1. Let A be an arbitrary matrix, whose size is such that we can form the product XA. What is XA? Work out special cases when X has 2 components, then when X has three components.

18. Let a, b be numbers, and let
$$A = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}.$$
What is AB? What is A^n where n is a positive integer?

19. If A is a square $n \times n$ matrix, we call a square matrix B an **inverse** for A if $AB = BA = I_n$. Show that if B, C are inverses for A, then $B = C$.

20. Show that the matrix A in Exercise 18 has an inverse. What is this inverse?

21. Show that if A, B are $n \times n$ matrices which have inverses, then AB has an inverse.

22. Determine all 2×2 matrices A such that $A^2 = O$.

23. Let $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. Show that $A^2 = \begin{pmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix}$.

    Determine A^n by induction for any positive integer n.

24. Find a 2×2 matrix A such that $A^2 = -I = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$.

25. Let

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

    Find A^2, A^3, A^4.

26. Let A be a diagonal matrix, with diagonal elements $a_1, \ldots, a_n$. What is A^2, A^3, A^k for any positive integer k?

27. Let

$$A = \begin{pmatrix} 0 & 1 & 6 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{pmatrix}.$$

    Find A^3.

CHAPTER IX

Linear Mappings

We shall first define the general notion of a mapping, which generalizes the notion of a function. Among mappings, the linear mappings are the most important. A good deal of mathematics is devoted to reducing questions concerning arbitrary mappings to linear mappings. For one thing, they are interesting in themselves, and many mappings are linear. On the other hand, it is often possible to approximate an arbitrary mapping by a linear one, whose study is much easier than the study of the original mapping. This is done in the calculus of several variables.

§1. *Mappings*

As usual, a collection of objects will be called a **set.** A member of the collection is also called an **element** of the set. It is useful in practice to use short symbols to denote certain sets. For instance we denote by $\mathbf{R}$ the set of all numbers. To say that "x is a number" or that "x is an element of $\mathbf{R}$" amounts to the same thing. The set of n-tuples of numbers will be denoted by $\mathbf{R}^n$. Thus "X is an element of $\mathbf{R}^n$" and "X is an n-tuple" mean the same thing. Instead of saying that u is an element of a set S, we shall also frequently say that u **lies in** S and we sometimes write $u \in S$. If S and S' are two sets, and if every element of S' is an element of S, then we say that S' is a **subset** of S. Thus the set of rational numbers is a subset of the set of (real) numbers. To say that S is a subset of S' is to say that S is part of S'. To denote the fact that S is a subset of S', we write $S \subset S'$.

If S_1, S_2 are sets, then the **intersection** of S_1 and S_2, denoted by $S_1 \cap S_2$, is the set of elements which lie in both S_1 and S_2. The **union** of S_1 and S_2, denoted by $S_1 \cup S_2$, is the set of elements which lie in S_1 or S_2.

Let S, S' be two sets. A **mapping** from S to S' is an association which to every element of S associates an element of S'. Instead of saying that F is a mapping from S into S', we shall often write the symbols $F: S \to S'$. A mapping will also be called a **map,** for the sake of brevity.

A function is a special type of mapping, namely it is a mapping from a set into the set of numbers, i.e. into $\mathbf{R}$.

We extend to mappings some of the terminology we have used for functions. For instance, if $T: S \to S'$ is a mapping, and if u is an element of S, then we denote by $T(u)$, or Tu, the element of S' associated to u by T. We call $T(u)$ the **value** of T at u, or also the **image** of u under T. The symbols $T(u)$ are read "T of u". The set of all elements $T(u)$, when u ranges over all elements of S, is called the **image** of T. If W is a subset of S, then the set of elements $T(w)$, when w ranges over all elements of W, is called the **image** of W under T, and is denoted by $T(W)$.

Let $F: S \to S'$ be a map from a set S into a set S'. If x is an element of S, we often write

$$x \mapsto F(x)$$

with a special arrow $\mapsto$ to denote the image of x under F. Thus, for instance, we would speak of the map F such that $F(x) = x^2$ as the map $x \mapsto x^2$.

Example. Let S and S' be both equal to $\mathbf{R}$. Let $f: \mathbf{R} \to \mathbf{R}$ be the function $f(x) = x^2$ (i.e. the function whose value at a number x is x^2). Then f is a mapping from $\mathbf{R}$ into $\mathbf{R}$. Its image is the set of numbers $\geqq 0$.

Example. Let S be the set of numbers $\geqq 0$, and let $S' = \mathbf{R}$. Let $g: S \to S'$ be the function such that $g(x) = x^{1/2}$. Then g is a mapping from S into $\mathbf{R}$.

Example. Let S be the set $\mathbf{R}^3$, i.e. the set of 3-tuples. Let $A = (2, 3, -1)$. Let $L: \mathbf{R}^3 \to \mathbf{R}$ be the mapping whose value at a vector $X = (x, y, z)$ is $A \cdot X$. Then $L(X) = A \cdot X$. If $X = (1, 1, -1)$, then the value of L at X is 6.

Just as we did with functions, we describe a mapping by giving its values. Thus, instead of making the statement in Example 5 describing the mapping L, we would also say: Let $L: \mathbf{R}^3 \to \mathbf{R}$ be the mapping $L(X) = A \cdot X$. This is somewhat incorrect, but is briefer, and does not usually give rise to confusion. More correctly, we can write $X \mapsto L(X)$ or $X \mapsto A \cdot X$ with the special arrow $\mapsto$ to denote the effect of the map L on the element X.

Example. Let $F: \mathbf{R}^2 \to \mathbf{R}^2$ be the mapping given by

$$F(x, y) = (2x, 2y).$$

Describe the image under F of the points lying on the circle $x^2 + y^2 = 1$.

Let (x, y) be a point on the circle of radius 1.

Let $u = 2x$ and $v = 2y$. Then u, v satisfy the relation

$$(u/2)^2 + (v/2)^2 = 1$$

or in other words,

$$\frac{u^2}{4} + \frac{v^2}{4} = 1.$$

Hence (u, v) is a point on the circle of radius 2. Therefore the image under F of the circle of radius 1 is a subset of the circle of radius 2. Conversely, given a point (u, v) such that

$$u^2 + v^2 = 4,$$

let $x = u/2$ and $y = v/2$. Then the point (x, y) satisfies the equation $x^2 + y^2 = 1$, and hence is a point on the circle of radius 1. Furthermore, $F(x, y) = (u, v)$. Hence every point on the circle of radius 2 is the image of some point on the circle of radius 1. We conclude finally that the image of the circle of radius 1 under F is precisely the circle of radius 2.

Note. In general, let S, S' be two sets. To prove that $S = S'$, one frequently proves that S is a subset of S' and that S' is a subset of S. This is what we did in the preceding argument.

Observe that the association

$$(x, y) \mapsto (2x, 2y)$$

is a dilation, i.e. a stretching by a factor of 2. Each point (x, y) is set on the point $(2x, 2y)$ which lies on the same ray from the origin, at twice the distance from the origin, as illustrated on Fig. 1.

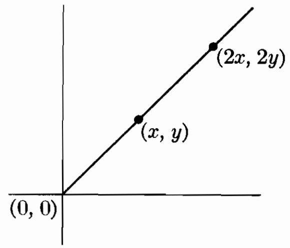


Figure 1

Example. In general, let r be a positive number. The association

$$(x, y) \mapsto (rx, ry)$$

is called **dilation** by the factor of r. We can also define it in 3-space, by

$$(x, y, z) \mapsto (rx, ry, rz).$$

We shall study such dilations later when we take up area and volume, and we shall see how these change under dilations.

Example. A curve in space as we studied in Chapter II was a mapping. For instance, we can define a map

$$F: \mathbf{R} \mapsto \mathbf{R}^3$$

by the association

$$t \mapsto (2t, 10^t, t^3).$$

Thus $F(t) = (2t, 10^t, t^3)$, and the value of F at 2 is

$$F(2) = (4, 100, 8).$$

In such a mapping we call

$$f_1(t) = 2t, \qquad f_2(t) = 10^t, \qquad f_3(t) = t^3$$

the **coordinate functions** of the mapping.

In general, a mapping $F: \mathbf{R} \to \mathbf{R}^3$ can always be expressed in terms of such functions, and we write

$$F(t) = \big(f_1(t), f_2(t), f_3(t)\big).$$

Example. Polar Coordinate Mapping. Let $F: \mathbf{R}^2 \to \mathbf{R}^2$ be the mapping defined by

$$F(r, \theta) = (r\cos\theta, r\sin\theta).$$

Thus we may put

$$x = r\cos\theta$$
$$y = r\sin\theta.$$

Then F is a mapping, which is called the **polar coordinate mapping.** We see that x and y depend on r, θ, and x, y are the coordinate functions of the mapping. Again, we shall study this mapping later when we change coordinates in a double integral. You should get well acquainted with this mapping, and we work out one example of what it does. Let S be the rectangle consisting of all points (r, θ) such that

$$0 \leqq r \leqq 2 \quad \text{and} \quad 0 \leqq \theta \leqq \pi/2.$$

We want to describe the image of S under the polar coordinate mapping.

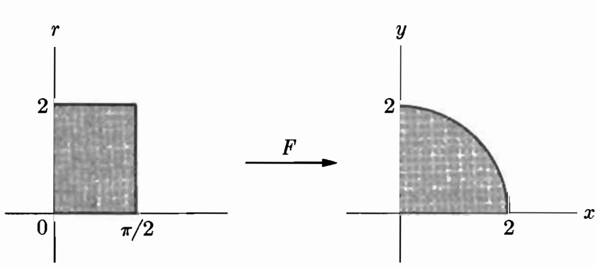


Figure 2

The image of S under the polar coordinate map F consists of all points (x, y) whose polar coordinates (r, θ) satisfy the above inequalities. We see that the image is just the sector of radius 2 in the first quadrant as shown on Fig. 2.

Example. Translations. Let A be a vector, say in the plane. We let

$$T_A\colon \mathbf{R}^2 \to \mathbf{R}^2$$

be the mapping such that

$$T_A(X) = X + A.$$

We call T_A the translation by A. On Fig. 3 we have drawn the translations of various points P, Q, M under translation by A. We may describe the image of a point P under translation by A as the point obtained from P by moving P in the direction of A, for a distance equal to the distance between O and A. Of course, the same notion also works in higher dimensional space. If A is an n-tuple, then

$$T_A\colon \mathbf{R}^n \to \mathbf{R}^n$$

is the mapping defined by the same equation as above, namely

$$T_A(X) = X + A.$$

You can visualize the picture (at least in $\mathbf{R}^3$) similarly.

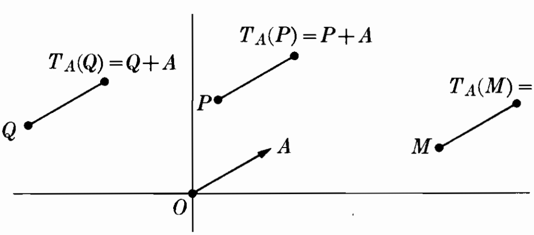


Figure 3

Example. You should not forget the identity mapping I, defined on any set S, and such that $I(x) = x$ for all x in S.

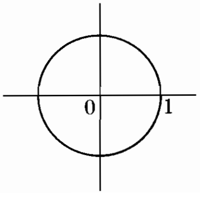
Exercises

1. Let $L(X) = A \cdot X$, where $A = (2, 3, -1)$. Give $L(X)$ when X is the vector: (a) $(1, 2, -3)$ (b) $(-1, 5, 0)$ (c) $(2, 1, 1)$
2. Let $F\colon \mathbf{R} \to \mathbf{R}^2$ be the mapping such that $F(t) = (e^t, t)$. What is $F(1)$, $F(0)$, $F(-1)$?
3. Let $A = (1, 1, -1, 3)$. Let $F\colon \mathbf{R}^4 \to \mathbf{R}$ be the mapping such that for any vector $X = (x_1, x_2, x_3, x_4)$ we have $F(X) = X \cdot A + 2$. What is the value of $F(X)$ when (a) $X = (1, 1, 0, -1)$ and (b) $X = (2, 3, -1, 1)$?

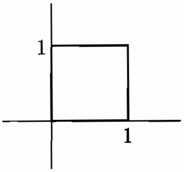
In each case, to prove that the image is equal to a certain set S, you must prove that the image is contained in S, and also that every element of S is in the image.

4. Let $F: \mathbf{R}^2 \to \mathbf{R}^2$ be the mapping defined by $F(x, y) = (2x, 3y)$. Describe the image of the points lying on the circle $x^2 + y^2 = 1$.
5. Let $F: \mathbf{R}^2 \to \mathbf{R}^2$ be the mapping defined by $F(x, y) = (xy, y)$. Describe the image under F of the straight line $x = 2$.
6. Let F be the mapping defined by $F(x, y) = (e^x \cos y, e^x \sin y)$. Describe the image under F of the line $x = 1$. Describe more generally the image under F of a line $x = c$, where c is a constant.
7. Let F be the mapping defined by $F(t, u) = (\cos t, \sin t, u)$. Describe geometrically the image of the (t, u)-plane under F.
8. Let F be the mapping defined by $F(x, y) = (x/3, y/4)$. What is the image under F of the ellipse
$$\frac{x^2}{9} + \frac{y^2}{16} = 1?$$
9. Draw the images of the following sets S under the polar coordinate mapping. In each case, the set S consists of all points (r, θ) satisfying the stated inequalities.
(a) $0 \leqq r \leqq 1$ and $0 \leqq \theta \leqq \pi/3$
(b) $0 \leqq r \leqq 3$ and $0 \leqq \theta \leqq 3\pi/4$
(c) $1 \leqq r \leqq 2$ and $\pi/4 \leqq \theta \leqq 3\pi/4$
(d) $1 \leqq r \leqq 2$ and $\pi/3 \leqq \theta \leqq 2\pi/3$
(e) $2 \leqq r \leqq 3$ and $\pi/6 \leqq \theta \leqq \pi/4$
(f) $2 \leqq r \leqq 3$ and $\pi/6 \leqq \theta \leqq \pi/3$
(g) $3 \leqq r \leqq 4$ and $\pi/2 \leqq \theta \leqq 2\pi/3$
10. In general, let S be the rectangle defined by the inequalities
$$0 < r_1 \leqq r \leqq r_2 \quad \text{and} \quad 0 \leqq \theta_1 \leqq \theta \leqq \theta_2.$$
Describe the image of S under the polar coordinate mapping.
11. Let $A = (-1, 2)$. Draw the image of the point X under translation by A when
(a) $X = (2, 3)$ (b) $X = (-5, 2)$ (c) $X = (1, 1)$
12. The identity mapping of $\mathbf{R}^n$ is equal to a translation T_A for some vector A. True or false? If true, which vector A?
13. Draw the image of the following figures under translation T_A, where $A = (-1, 2)$.

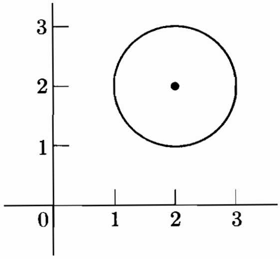
(a) The circle as shown:



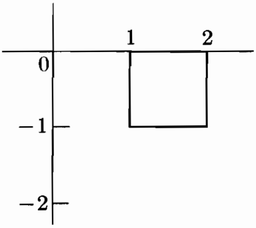
(b) The square, as shown:



(c) The circle as shown:



(d) The square as shown:



§2. *Linear mappings*

Consider two Euclidean spaces $\mathbf{R}^n$ and $\mathbf{R}^m$. In the applications, the values for m and n are 1, 2, or 3, but they can all occur, so it is just as easy to leave them indeterminate for what we are about to say.

A mapping

$$L\colon \mathbf{R}^n \to \mathbf{R}^m$$

is called a **linear mapping** if it satisfies the following properties:

LM 1. *For any elements X, Y in $\mathbf{R}^n$ we have*

$$L(X + Y) = L(X) + L(Y).$$

LM 2. *If c is a number, then*

$$L(cX) = cL(X).$$

These properties should remind you of properties of multiplication of matrices, and also of the dot product of n-tuples. These in fact provide us with the examples which interest us for this course.

Example. Let $A = (3, 1, -2)$. Then we have a linear map

$$L_A\colon \mathbf{R}^3 \to \mathbf{R}$$

defined by the dot product,

$$L_A(X) = A \cdot X.$$

If $X = (x, y, z)$, then

$$L_A(X) = 3x + y - 2z.$$

In general, let

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

be an $m \times n$ matrix. We can then associate with A a map

$$L_A\colon \mathbf{R}^n \to \mathbf{R}^m$$

by letting

$$L_A(X) = AX$$

for every **column** vector X in $\mathbf{R}^n$. Thus L_A is defined by the association $X \mapsto AX$, the product being the product of matrices. That L_A is linear is simply a special case of Theorem 1, Chapter VIII, §2, namely the theorem concerning properties of multiplication of matrices. Indeed, we have

$$A(X + Y) = AX + AY \qquad \text{and} \qquad A(cX) = cAX$$

for all vectors X, Y in $\mathbf{R}^n$ and all numbers c. We call L_A the linear map **associated** with the matrix A. We also say that A is the **matrix representing** the linear map L_A.

Example. If

$$A = \begin{pmatrix} 2 & 1 \\ -1 & 5 \end{pmatrix} \qquad \text{and} \qquad X = \begin{pmatrix} 3 \\ 7 \end{pmatrix},$$

then

$$L_A(X) = \begin{pmatrix} 2 & 1 \\ -1 & 5 \end{pmatrix}\begin{pmatrix} 3 \\ 7 \end{pmatrix} = \begin{pmatrix} 6 + 7 \\ -3 + 35 \end{pmatrix} = \begin{pmatrix} 13 \\ 32 \end{pmatrix}.$$

Theorem 1. *If A, B are $m \times n$ matrices and if $L_A = L_B$, then $A = B$. In other words, if matrices A, B give rise to the same linear map, then they are equal.*

Proof. By definition, we have $A_i \cdot X = B_i \cdot X$ for all i, if A_i is the

i-th row of A and B_i is the i-th row of B. Hence $(A_i - B_i) \cdot X = 0$ for all i and all X. Hence $A_i - B_i = O$, and $A_i = B_i$ for all i. Hence $A = B$.

It can easily be shown that every linear map from $\mathbf{R}^n$ into $\mathbf{R}^m$ is of the form L_A for some matrix A, in other words, the above example is the most general type of linear map from $\mathbf{R}^n$ into $\mathbf{R}^m$. The matrix A is called the **matrix associated with the linear map.** We shall give the proof when $n = 2$.

Let $E^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $E^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ be the standard unit vectors. Let $L: \mathbf{R}^2 \to \mathbf{R}^2$ be a linear map such that

$$L(E^1) = \begin{pmatrix} a \\ c \end{pmatrix} \quad \text{and} \quad L(E^2) = \begin{pmatrix} b \\ d \end{pmatrix}.$$

We shall prove that the matrix associated with L is precisely

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

First note that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix} = L(E^1)$$

and

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix} = L(E^2).$$

Let $X = \begin{pmatrix} x \\ y \end{pmatrix}$, so that $X = xE^1 + yE^2$. Then

$$\begin{aligned} L(X) = L(xE^1) + L(yE^2) &= xL(E^1) + yL(E^2) \\ &= xAE^1 + yAE^2 \\ &= A(xE^1 + yE^2) \\ &= AX. \end{aligned}$$

This proves that $L(X) = AX$, and therefore that A is the matrix representing L. A similar proof can be given for $\mathbf{R}^3$, or $\mathbf{R}^n$.

Example. Let $L: \mathbf{R}^2 \to \mathbf{R}^2$ be a linear map such that

$$L(E^1) = \begin{pmatrix} 3 \\ 5 \end{pmatrix} \quad \text{and} \quad L(E^2) = \begin{pmatrix} -2 \\ 9 \end{pmatrix}.$$

Then the matrix associated with L is the matrix

$$A = \begin{pmatrix} 3 & -2 \\ 5 & 9 \end{pmatrix}.$$

You can check that it has the desired effect on the unit vectors, namely:

$$\begin{pmatrix} 3 & -2 \\ 5 & 9 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$$

and

$$\begin{pmatrix} 3 & -2 \\ 5 & 9 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ 9 \end{pmatrix}.$$

Exercises

1. In each case, find the vector $L_A(X)$.

(a) $A = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$, $X = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$ (b) $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $X = \begin{pmatrix} 5 \\ 1 \end{pmatrix}$

(c) $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $X = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$ (d) $A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, $X = \begin{pmatrix} 7 \\ -3 \end{pmatrix}$

2. Let r be a number. Let $F_r\colon \mathbf{R}^n \to \mathbf{R}^n$ be the dilation mapping, defined by the formula

$$F_r(x) = rX.$$

Exhibit a matrix A such that $F_r(X) = AX$.

3. Let a, b be numbers. Let $F_{a,b}\colon \mathbf{R}^2 \to \mathbf{R}^2$ be the mapping such that

$$F_{a,b}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax \\ by \end{pmatrix}.$$

Exhibit a matrix A such that $F_{a,b}(X) = AX$.

4. Let a_1, a_2, a_3 be numbers. If $X = (x, y, z)$ let $F(X) = (a_1x, a_2y, a_3z)$. Writing X as a column vector, exhibit a matrix A such that $F(X) = AX$.

5. Let $F(x, y, z) = (x, y)$. Writing X as a column vector, exhibit a matrix A such that $F(X) = AX$.

6. Let $F(x, y, z) = x$. Writing X as a column vector, exhibit a matrix A such that $F(X) = AX$.

7. Let $F(x, y, z) = (x, z)$. Writing X as a column vector, exhibit a matrix A such that $F(X) = AX$.

8. Same question if $F(x, y, z) = (y, z)$.

9. Let $F: \mathbf{R}^4 \to \mathbf{R}^2$ be the mapping such that

$$F(x_1, x_2, x_3, x_4) = (x_1, x_2).$$

Writing X as a column vector, exhibit a matrix A such that $F(X) = AX$.

10. Let $F: \mathbf{R}^4 \to \mathbf{R}^3$ be the mapping such that

$$F(x_1, x_2, x_3, x_4) = (x_1, x_2, x_3).$$

Writing X as a column vector, exhibit a matrix A such that $F(X) = AX$.

11. Let A be an element of $\mathbf{R}^3$. Suppose that the translation by A is a linear map. What is the only possibility for A? If $A \neq O$, can T_A be a linear map? Proof?

12. Let $L: \mathbf{R}^2 \to \mathbf{R}^2$ be the linear map such that

$$L(E^1) = \begin{pmatrix} -5 \\ 7 \end{pmatrix} \quad \text{and} \quad L(E^2) = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

What is the matrix associated with L?

13. Same question if

$$L(E^1) = \begin{pmatrix} -1 \\ 4 \end{pmatrix} \quad \text{and} \quad L(E^2) = \begin{pmatrix} 2 \\ 6 \end{pmatrix}.$$

14. Let $L: \mathbf{R}^3 \to \mathbf{R}^3$ be a linear map such that

$$L(E^1) = \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix}, \quad L(E^2) = \begin{pmatrix} -2 \\ 7 \\ 9 \end{pmatrix}, \quad L(E^3) = \begin{pmatrix} 8 \\ -5 \\ 2 \end{pmatrix}.$$

Here,

$$E^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad E^2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad E^3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

What is the matrix associated with L? Verify that it has the desired effect on the unit vectors.

15. Write out the proof that if E^1, E^2, E^3 are the standard unit vectors in $\mathbf{R}^3$, and if $L: \mathbf{R}^3 \to \mathbf{R}^3$ is the linear map such that

$$L(E^1) = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix}, \quad L(E^2) = \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}, \quad L(E^3) = \begin{pmatrix} a_{13} \\ a_{23} \\ a_{33} \end{pmatrix},$$

then the matrix A associated with L is the matrix (a_{ij}), that is

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

16. Let $L\colon \mathbf{R}^3 \to \mathbf{R}^3$ be the linear map such that

$$L(E^1) = \begin{pmatrix} -3 \\ 5 \\ 0 \end{pmatrix}, \quad L(E^2) = \begin{pmatrix} 4 \\ 1 \\ -7 \end{pmatrix}, \quad L(E^3) = \begin{pmatrix} 5 \\ -2 \\ 8 \end{pmatrix}.$$

What is the matrix associated with L? Verify directly that it has the desired effect on the unit vectors.

17. Let $L\colon \mathbf{R} \to \mathbf{R}^n$ be a linear map. Prove that there exists a vector A in $\mathbf{R}^n$ such that for all t in $\mathbf{R}$ we have

$$L(t) = tA.$$

18. Let $L\colon \mathbf{R}^2 \to \mathbf{R}^3$ be a linear map. Let

$$E^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad E^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

be the unit vectors in $\mathbf{R}^2$. Suppose that

$$L(E^1) = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix}, \quad L(E^2) = \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}.$$

In terms of the a_{ij}, what is the matrix A associated with L?

19. Let $L\colon \mathbf{R}^2 \to \mathbf{R}^3$ be a linear map, and suppose that E^1, E^2 are the unit vectors in $\mathbf{R}^2$. Let

$$L(E^1) = \begin{pmatrix} 3 \\ 1 \\ -4 \end{pmatrix} \quad \text{and} \quad L(E^2) = \begin{pmatrix} -5 \\ 7 \\ -8 \end{pmatrix}.$$

What is the matrix A associated with L?

§3. *Geometric applications*

Let P, A be elements of $\mathbf{R}^n$. We define the **line segment** between P and $P + A$ to be the set of all points

$$P + tA, \qquad 0 \leqq t \leqq 1.$$

This line segment is illustrated in the following picture.

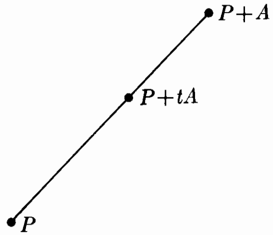


Figure 4

For instance, if $t = \frac{1}{2}$, then $P + \frac{1}{2}A$ is the point midway between P and $P + A$. Similarly, if $t = \frac{1}{3}$, then $P + \frac{1}{3}A$ is the point one-third of the way between P and $P + A$ (Fig. 5).

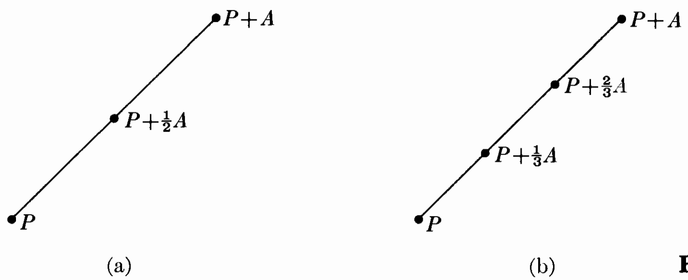


Figure 5

If P, Q are elements of $\mathbf{R}^n$, let $A = Q - P$. Then the line segment between P and Q is the set of all points $P + tA$, or

$$P + t(Q - P), \qquad 0 \leqq t \leqq 1.$$

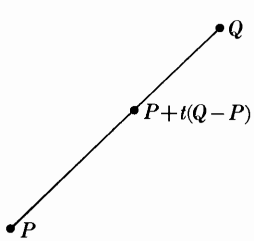


Figure 6

Observe that we can rewrite the expression for these points in the form

$$(1 - t)P + tQ, \qquad 0 \leqq t \leqq 1, \tag{1}$$

and letting $s = 1 - t$, $t = 1 - s$, we can also write it as

$$sP + (1 - s)Q, \qquad 0 \leqq s \leqq 1.$$

Finally, we can write the points of our line segment in the form

$$t_1P + t_2Q \tag{2}$$

with $t_1, t_2 \geqq 0$ and $t_1 + t_2 = 1$. Indeed, letting $t = t_2$, we see that every point which can be written in the form (2) satisfies (1). Conversely, we let $t_1 = 1 - t$ and $t_2 = t$ and see that every point of the form (1) can be written in the form (2).

Let $L: \mathbf{R}^n \to \mathbf{R}^m$ be a linear map. Let S be the line segment in $\mathbf{R}^n$ between two points P, Q. Then the image $L(S)$ of this line segment is the line segment in $\mathbf{R}^m$ between the points $L(P)$ and $L(Q)$. This is obvious from (2), because

$$L(t_1P + t_2Q) = t_1L(P) + t_2L(Q).$$

We shall now generalize this discussion to higher dimensional figures. Let P, Q be elements of $\mathbf{R}^n$, and assume that they are $\neq 0$, and Q is not a scalar multiple of P. We define the **parallelogram spanned** by P and Q to be the set of all points

$$t_1P + t_2Q, \qquad 0 \leqq t_i \leqq 1 \quad \text{for} \quad i = 1, 2.$$

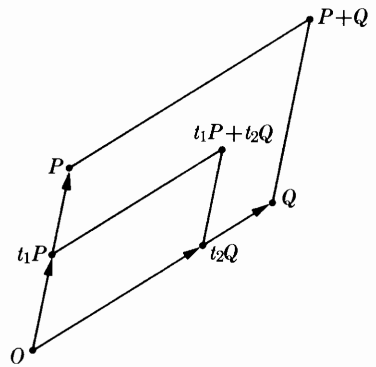


Figure 7

This definition is clearly justified since t_1P is a point of the segment between O and P (Fig. 7), and t_2Q is a point of the segment between O and Q. For all values of t_1, t_2 ranging independently between 0 and 1, we see geometrically that $t_1v + t_2w$ describes all points of the parallelogram.

At the end of §1 we defined **translations.** We obtain the most general parallelogram (Fig. 8) by taking the translation of the parallelogram just described. Thus if A is an element of $\mathbf{R}^n$, the translation by A of the parallelogram spanned by P and Q consists of all points

$$A + t_1P + t_2Q, \quad 0 \leqq t_i \leqq 1 \quad \text{for} \quad i = 1, 2.$$

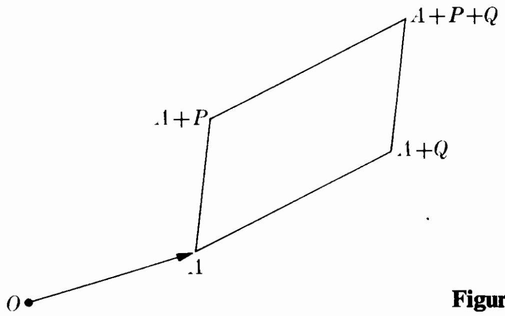


Figure 8

As with line segments, we see that if

$$L\colon \mathbf{R}^n \to \mathbf{R}^m$$

is a linear map, and if S is a parallelogram as described above, then the image of S is again a parallelogram, provided that $L(P)$ and $L(Q)$ do not lie on the same line through the origin (i.e. $L(P)$ is not a scalar multiple of $L(Q)$). This is immediately seen, because the image of S under L consists of all points

$$L(A + t_1P + t_2Q) = L(A) + t_1L(P) + t_2L(Q),$$

with

$$0 \leqq t_i \leqq 1 \qquad \text{for} \qquad i = 1, 2.$$

We see again the usefulness of the conditions for linearity **LM 1** and **LM 2.**

Example. Let S be the parallelogram spanned by the vectors $P = (1, 2)$ and $(-1, 5)$. Let $L\colon \mathbf{R}^2 \to \mathbf{R}^2$ be the linear map L_A, where A is the matrix

$$\begin{pmatrix} 3 & 1 \\ -1 & 5 \end{pmatrix}.$$

Then, writing P, Q as vertical vectors, we obtain

$$L(P) = AP = \begin{pmatrix} 3 & 1 \\ -1 & 5 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 9 \end{pmatrix}$$

$$L(Q) = AQ = \begin{pmatrix} 3 & 1 \\ -1 & 5 \end{pmatrix}\begin{pmatrix} -1 \\ 5 \end{pmatrix} = \begin{pmatrix} 2 \\ 26 \end{pmatrix}.$$

Hence the image of S under L is the parallelogram spanned by the vectors $(5, 9)$ and $(2, 26)$.

On the next figure, we have drawn a typical situation of the image of a parallelogram under a linear map.

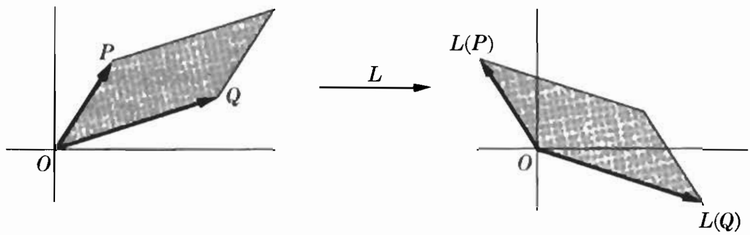


Figure 9

A similar discussion can be carried out in 3-space. It is good practice for you to write it up yourself. Do Exercise 5.

Exercises

1. Let L be the linear map represented by the matrix

$$\begin{pmatrix} 1 & -1 \\ 2 & 3 \end{pmatrix}.$$

Let S be the line segment between P and Q. Draw the image of S under L, indicating $L(P)$ and $L(Q)$ in each of the following cases.

(a) $P = (2, 1)$ and $Q = (-1, 1)$
(b) $P = (3, -1)$ and $Q = (1, 2)$
(c) $P = (1, 1)$ and $Q = (1, -1)$
(d) $P = (2, -1)$ and $Q = (1, 2)$

2. In cases (a), (b), (c), and (d) of Exercise 1, let T be the parallelogram spanned by P and Q. Draw the image of T by the linear map L of Exercise 1, indicating in each case $L(P)$ and $L(Q)$.

3. Let $E^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $E^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ be the standard unit vectors. Write down their images under the linear map L represented by the matrix

$$\begin{pmatrix} 3 & -1 \\ 5 & 2 \end{pmatrix}.$$

Let S be the square spanned by E^1 and E^2. Draw the image of this square under L, indicating $L(E^1)$ and $L(E^2)$.

4. Let E^1, E^2 again be the standard unit vectors, drawn vertically. Let L be the linear map represented by the matrix

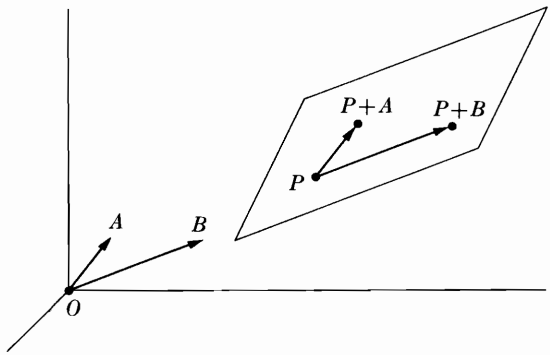
$$\begin{pmatrix} -2 & 3 \\ 1 & 5 \end{pmatrix}.$$

Let S be the square spanned by E^1, E^2. Draw the image $L(S)$, again indicating $L(E^1)$ and $L(E^2)$.

5. (a) Give a definition of the **box (parallelepiped)** spanned by three vectors A, B, C in $\mathbf{R}^3$.
   (b) Let $L\colon \mathbf{R}^3 \to \mathbf{R}^3$ be a linear map. Prove that the image of such a box under L is again a box, spanned by $L(A)$, $L(B)$, $L(C)$ (provided that the segments from O to $L(A)$, $L(B)$, $L(C)$, respectively, do not all lie in a plane, otherwise you get a "degenerate" box).
   (c) Draw a picture for this in 3-dimensional space.
6. Let L be the linear map of $\mathbf{R}^3$ into itself represented by the matrix
$$\begin{pmatrix} -3 & 1 & 4 \\ 2 & 2 & 1 \\ 1 & -2 & 5 \end{pmatrix}.$$
   Let S be the cube spanned by the three unit vectors E^1, E^2, E^3. Give explicitly three vectors spanning $L(S)$.
7. Same questions as in Exercise 6, if L is represented by the matrix
$$\begin{pmatrix} 2 & 4 & -6 \\ 3 & 7 & 5 \\ -1 & 2 & -8 \end{pmatrix}.$$
8. Let $X(t) = P + tA$, with t in $\mathbf{R}$, be the parametrization of a straight line in $\mathbf{R}^n$. Let $L\colon \mathbf{R}^n \to \mathbf{R}^m$ be a linear map. Suppose that $L(A) \neq O$. Prove that the image of the straight line is a straight line.
9. Let S be a line passing through two distinct points P and Q, in $\mathbf{R}^n$. Let $L\colon \mathbf{R}^n \to \mathbf{R}^m$ be a linear map, such that $L(P) \neq L(Q)$.
   (a) Give a parametric representation of the line S.
   (b) Give a parametric representation of the line $L(S)$.
10. Let A, B be non-zero vectors in $\mathbf{R}^n$ and assume that neither is a scalar multiple of the other. Such vectors are called **independent.** We define the **plane spanned by A and B** to be the set of all points
$$tA + sB,$$
    for all real numbers t, s. Observe that this is the 2-dimensional analogue of the parametrization of a line. Let $L\colon \mathbf{R}^n \to \mathbf{R}^m$ be a linear map. Assume that $L(A)$ and $L(B)$ are independent. Prove that the image of the plane spanned by A and B is a plane (spanned by which vectors?).
11. Let A, B be independent vectors in $\mathbf{R}^n$, and let P be a point. We define the **plane through P parallel to A, B** to be the set of all points
$$P + tA + sB,$$

where t, s range over all real numbers. Let $L: \mathbf{R}^n \to \mathbf{R}^m$ be a linear map such that $L(A)$ and $L(B)$ are independent. Prove that the image of the preceding plane is also a plane.

The plane of Exercise 10 looks like this.



It is the translation by P of the plane in Exercise 9.

§4. *Composition and inverse of mappings*

This section will be useful for Chapter XI, §3, §4 *and Chapter XIII. You can omit these without impairing your understanding of the rest of the book.*

Before we discuss linear mappings, we have to make some more remarks on mappings in general. You recall that in studying functions of one variable, you met composite functions and the chain rule for differentiation. We shall meet a similar situation in several variables.

In one variable, let

$$f: \mathbf{R} \to \mathbf{R} \quad \text{and} \quad g: \mathbf{R} \to \mathbf{R}$$

be functions. Then we can form the **composite function** $g \circ f$, defined by

$$(g \circ f)(x) = g(f(x)).$$

Let U, V, W be sets. Let

$$F: U \to V \quad \text{and} \quad G: V \to W$$

be mappings. Then we can form the **composite mapping** from U into W,

denoted by $G \circ F$. It is by definition the mapping defined by

$$(G \circ F)(u) = G(F(u))$$

for all u in U.

Example. Let $G: \mathbf{R}^2 \to \mathbf{R}^2$ be the mapping such that

$$G(Y) = 3Y.$$

Let $F: \mathbf{R}^2 \to \mathbf{R}^2$ be the mapping such that $F(X) = X + A$, where

$$A = (1, -2).$$

Then

$$G(F(X)) = G(X + A) = (X + A) = 3X + 3A.$$

Our mapping $G \circ F$ is the composite of a translation and a dilation.

Example. Let $G: \mathbf{R}^2 \to \mathbf{R}^3$ be the mapping such that

$$G(x, y) = (x^2, xy, \sin y).$$

If (u, v, w) are the coordinates of $\mathbf{R}^3$, we may set

$$u = x^2, \qquad v = xy, \qquad w = \sin y.$$

Let $F: \mathbf{R}^3 \to \mathbf{R}^3$ be the mapping such that

$$F(u, v, w) = (u^3, uv, vw).$$

Then

$$F(\mathrm{G}(x, y)) = (x^6, x^3y, xy \sin y).$$

The composition of mappings is associative. More precisely, let U, V, W, S be sets. Let

$$F: U \to V, \qquad G: V \to W, \qquad \textit{and} \qquad H: W \to S$$

be mappings. Then

$$H \circ (G \circ F) = (H \circ G) \circ F.$$

Proof. Here again, the proof is very simple. By definition, we have, for any element u of U:

$$(H \circ (G \circ F))(u) = H((G \circ F)(u)) = H(G(F(u))).$$

On the other hand,

$$((H \circ G) \circ F)(u) = (H \circ G)(F(u)) = H(G(F(u))).$$

By definition, this means that $(H \circ G) \circ F = H \circ (G \circ F)$.

If S is any set, the **identity mapping** I_S is defined to be the map such that $I_S(x) = x$ for all $x \in S$. We note that the identity map is both injective and surjective. If we do not need to specify the reference to S (because it is made clear by the context), then we write I instead of I_S. Thus we have $I(x) = x$ for all $x \in S$.

Finally, we define inverse mappings. Let $F: S \to S'$ be a mapping from one set into another set. We say that F has an **inverse** if there exists a mapping $G: S' \to S$ such that

$$G \circ F = Id_S \quad \text{and} \quad F \circ G = Id_{S'}.$$

By this we mean that the composite maps $G \circ F$ and $F \circ G$ are the identity mappings of S and S' respectively.

Example. Let $S = S'$ be the set of all numbers $\geqq 0$. Let

$$f\colon S \to S'$$

be the map such that $f(x) = x^2$. Then f has an inverse mapping, namely the map $g: S \to S$ such that $g(x) = \sqrt{x}$.

Example. Let $\mathbf{R}^+$ be the set of numbers > 0 and let $f: \mathbf{R} \to \mathbf{R}^+$ be the map such that $f(x) = e^x$. Then f has an inverse mapping which is nothing but the logarithm.

Example. Let A be a vector in $\mathbf{R}^3$ and let

$$T_A\colon \mathbf{R}^3 \to R^3$$

be the translation by A. By definition, we recall that this means

$$T_A(X) = X + A.$$

If B is another vector in $\mathbf{R}^3$, then the composite mapping $T_B \circ T_A$ has the value

$$\begin{aligned}(T_B \circ T_A)(X) &= T_B(T_A(X)) \\ &= T_B(X + A) \\ &= X + A + B.\end{aligned}$$

If $B = -A$, we see that

$$T_{-A}(T_A(X)) = X + A - A = X,$$

and similarly that $T_A(T_{-A}(X)) = X$. Hence T_{-A} is the inverse mapping of T_A. In words, we may say that the inverse mapping of translation by A is translation by $-A$. Of course, the same holds in $\mathbf{R}^n$.

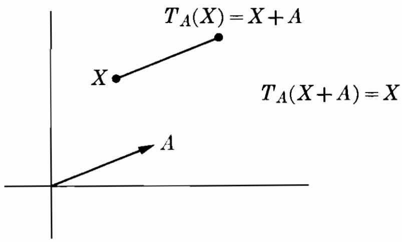


Figure 10

Let

$$f: S \to S'$$

be a map. We say that f is **injective** if whenever $x, y \in S$ and $x \neq y$, then $f(x) \neq f(y)$. In other words, f is injective means that f takes on distinct values at distinct elements of S. For example, the map

$$f: \mathbf{R} \to \mathbf{R}$$

such that $f(x) = x^2$, is not injective, because $f(1) = f(-1) = 1$. Also the function $x \mapsto \sin x$ is not injective, because $\sin x = \sin (x + 2\pi)$. However, the map $f: \mathbf{R} \to \mathbf{R}$ such that $f(x) = x + 1$ is injective, because if $x + 1 = y + 1$, then $x = y$.

Again, let $f: S \to S'$ be a mapping. We shall say that f is **surjective** if the image of f is all of S'. Again, the map

$$f: \mathbf{R} \to \mathbf{R}$$

such that $f(x) = x^2$, is not surjective, because its image consists of all numbers $\geqq 0$, and this image is not equal to all of R. On the other hand, the map of $\mathbf{R}$ into $\mathbf{R}$ given by $x \mapsto x^3$ is surjective, because given a number y there exists a number x such that $y = x^3$ (the cube root of y). Thus every number is in the image of our map.

Let $\mathbf{R}^+$ be the set of real numbers $\geqq 0$. As a matter of convention, we agree to distinguish between the maps

$$\mathbf{R} \to \mathbf{R} \quad \text{and} \quad \mathbf{R}^+ \to \mathbf{R}^+$$

given by the same formula $x \mapsto x^2$. The point is that when we view the association $x \mapsto x^2$ as a map of $\mathbf{R}$ into $\mathbf{R}$, then it is not surjective, and it is not injective. But when we view this formula as defining a map from $\mathbf{R}^+$ into $\mathbf{R}^+$, then it gives both an injective and surjective map of $\mathbf{R}^+$

into itself, because every positive number has a positive square root, and such a positive square root is uniquely determined.

In general, when dealing with a map $f: S \to S'$, we must therefore always specify the sets S and S', to be able to say that f is injective, or surjective, or neither. To have a completely accurate notation, we should write

$$f_{S,S'}$$

or some such symbol which specifies S and S' into the notation, but this becomes too clumsy, and we prefer to use the context to make our meaning clear.

Let

$$f: S \to S'$$

be a map which has an inverse mapping g. Then f is both injective and surjective.

Proof. Let $x, y \in S$ and $x \neq y$. Let $g: S' \to S$ be the inverse mapping of f. If $f(x) = f(y)$, then we must have

$$x = g(f(x)) = g(f(y)) = y,$$

which is impossible. Hence $f(x) \neq f(y)$, and therefore f is injective. To prove that f is surjective, let $z \in S'$. Then

$$f(g(z)) = z$$

by definition of the inverse mapping, and hence $z = f(x)$, where $x = g(z)$. This proves that f is surjective.

The converse of the statement we just proved is also true, namely:

Let $f: S \to S'$ be a map which is both injective and surjective. Then f has an inverse mapping.

Proof. Given $z \in S'$, since f is surjective, there exists $x \in S$ such that $f(x) = z$. Since f is injective, this element x is uniquely determined by z, and we can therefore define

$$g(z) = x.$$

By definition of g, we find that $f(g(z)) = z$, and $g(f(x)) = x$, so that g is an inverse mapping for f.

Thus we can say that a map $f: S \to S'$ has an inverse mapping if and only if f is both injective and surjective.

Using another terminology, we can also say that a map

$$f\colon S \to S'$$

which has an inverse mapping establishes a one-one correspondence between the elements of S and the elements of S'.

We shall, of course, be mostly concerned with linear mappings.

Let $F\colon \mathbf{R}^n \to \mathbf{R}^m$ and $G\colon \mathbf{R}^m \to R^s$ be linear maps. Then the composite map $G \circ F$ is also a linear map.

Proof. This is very easy to prove. Let u, v be elements of U. Since F is linear, we have $F(u + v) = F(u) + F(v)$. Hence

$$(G\circ F)(u + v) = G\big(F(u + v)\big) = G\big(F(u) + F(v)\big).$$

Since G is linear, we obtain

$$G\big(F(u) + F(v)\big) = G\big(F(u)\big) + G\big(F(v)\big).$$

Hence

$$(G\circ F)(u + v) = (G\circ F)(u) + (G\circ F)(v).$$

Next, let c be a number. Then

$$\begin{aligned} (G\circ F)(cu) &= G\big(F(cu)\big) \\ &= G\big(cF(u)\big) \qquad \text{(because } F \text{ is linear)} \\ &= cG\big(F(u)\big) \qquad \text{(because } G \text{ is linear).} \end{aligned}$$

This proves that $G\circ F$ is a linear mapping.

We can also see this with matrices. Suppose that A is the matrix associated with F, and B is the matrix associated with G. Then by definition, we have

$$F(X) = AX, \qquad \text{for } X \text{ in } \mathbf{R}^n,$$

and

$$G(Y) = BY, \qquad \text{for } Y \text{ in } \mathbf{R}^m.$$

Hence

$$G\big(F(X)\big) = B(AX) = (BA)X,$$

and we see that the product BA is the matrix associated with the linear map $G \circ F$. In other words, the product of the matrices associated with G and F, respectively, is the matrix associated with $G \circ F$.

Let $F\colon \mathbf{R}^n \to \mathbf{R}^n$ be a linear mapping. We shall say that F is **invertible** if there exists a linear mapping

$$G\colon \mathbf{R}^n \to \mathbf{R}^n$$

such that $G \circ F = I$ and $F \circ G = I$. [It can be shown that if an inverse for F exists as a mapping, then this inverse is necessarily linear, but we don't give the proof. It is an easy exercise.] Similarly, let A be an $n \times n$ matrix. We say that A is **invertible** if there exists an $n \times n$ matrix B such that $AB = BA = I_n$ is the unit $n \times n$ matrix. We denote B by A^{-1}.

If F is a linear mapping as above, then we know that it has an associated matrix A, such that

$$F(X) = AX, \qquad \text{all } X \text{ in } \mathbf{R}^n.$$

Suppose that F is invertible, and that G is its inverse linear mapping. Then G also has an associated matrix B, and since $G(F(X)) = X$, we must have

$$BAX = X,$$

for all X in $\mathbf{R}^n$. Similarly, we must also have $ABX = X$ for all X in $\mathbf{R}^n$. In particular, this must be true if X is any one of the standard unit vectors, and from this we see that $AB = BA = I_n$ is the unit $n \times n$ matrix. Thus $B = A^{-1}$. In other words:

If A is the matrix associated with an invertible linear mapping $L: \mathbf{R}^n \to \mathbf{R}^n$, then A^{-1} is the matrix associated with the inverse of L.

It is usually a tedious process to find the inverse of a matrix, and this process involves linear equations. For 2×2 matrices, however, the process is short. We shall discuss it in connection with determinants.

Exercises

1. Let $F: \mathbf{R}^3 \to \mathbf{R}^3$ be the map such that $F(X) = 7X$. Prove that F has an inverse mapping, and that this inverse is linear. Do the same if $F: \mathbf{R}^n \to \mathbf{R}^n$ is defined by the same formula.
2. Let $F: \mathbf{R}^n \to \mathbf{R}^n$ be the map such that $F(X) = -8X$. Prove that F is invertible, and write down its inverse explicitly.
3. Let c be a number $\neq 0$ and let $L: \mathbf{R}^n \to \mathbf{R}^n$ be the map such that $F(X) = cX$. Prove that L has an inverse linear map, and write it down explicitly.
4. Let A, B, C be square matrices of the same size and assume that they are invertible. Prove that AB is invertible, and express its inverse in terms of A^{-1} and B^{-1}. Also show that ABC is invertible.
5. Let A be a square matrix such that $A^2 = 0$. Show that I-A is invertible. (I is the unit matrix of the same size as A.)
6. Let A be a square matrix such that $A^2 + 2A + I = 0$. Show that A is invertible.
7. Let A be a square matrix such that $A^3 = 0$. Show that I-A is invertible.

CHAPTER X

Determinants

In this chapter we carry out the theory of determinants for the case of 2×2 and 3×3 matrices.

§1. Determinants of order 2

Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

be a 2×2 matrix. We define its determinant to be $ad - cb$. Thus the determinant is a number. We denote it by

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

For example, the determinant of the matrix

$$\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$$

is equal to $2 \cdot 4 - 1 \cdot 1 = 7$. The determinant of

$$\begin{pmatrix} -2 & -3 \\ 4 & 5 \end{pmatrix}$$

is equal to $(-2) \cdot 5 - (-3) \cdot 4 = -10 + 12 = 2$.

Theorem 1. *If A is a 2×2 matrix, then the determinant of A is equal to the determinant of the transpose of A. In other words,*

$$D(A) = D({}^tA).$$

Proof. This is immediate from the definition of the determinant. We

have

$$|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \quad \text{and} \quad |{}^tA| = \begin{vmatrix} a & c \\ b & d \end{vmatrix},$$

and

$$ad - bc = ad - cb.$$

Of course, the property expressed in Theorem 2 is very simple. We give it here because it is satisfied by 3×3 determinants which will be studied later.

Consider a 2×2 matrix A with columns A^1, A^2. The determinant $D(A)$ has interesting properties with respect to these columns, which we shall describe. Thus it is useful to use the notation

$$D(A) = D(A^1, A^2)$$

to emphasize the dependence of the determinant on its columns. If the two columns are denoted by

$$B = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix},$$

then we would write

$$D(B, C) = \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix} = b_1c_2 - c_1b_2.$$

We may view the determinant as a certain type of "product" between the columns B and C. To what extent does this product satisfy the same rules as the product of numbers. Answer: To some extent, which we now determine precisely.

To begin with, this "product" satisfies distributivity. In the determinant notation, this means:

D1. *If* $B = B' + B''$, *i.e.*

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} b_1' \\ b_2' \end{pmatrix} + \begin{pmatrix} b_1'' \\ b_2'' \end{pmatrix},$$

then

$$D(B' + B'', C) = D(B', C) + D(B'', C).$$

Similarly, if $C = C' + C''$, *then*

$$D(B, C' + C'') = D(B, C') + D(B, C'').$$

Proof. Of course, the proof is quite simple using the definition of the determinant. We have

$$\begin{aligned} D(B' + B'', C) &= \begin{vmatrix} b_1' + b_1'' & c_1 \\ b_2' + b_2'' & c_2 \end{vmatrix} \\ &= (b_1' + b_1'')c_2 - (b_2' + b_2'')c_1 \\ &= b_1'c_2 + b_1''c_2 - b_2'c_1 - b_2''c_1 \\ &= D(B', C) + D(B'', C). \end{aligned}$$

Distributivity on the other side is proved similarly.

D2. *If x is a number, then*

$$D(xB, C) = x \cdot D(B, C) = D(B, xC).$$

Proof. We have

$$\begin{aligned} D(xB, C) = \begin{vmatrix} xb_1 & c_1 \\ xb_2 & c_2 \end{vmatrix} = xb_1c_2 - xb_2c_1 &= x(b_1c_2 - b_2c_1) \\ &= xD(B, C). \end{aligned}$$

Again, the other equality is proved similarly.

Properties **D1** and **D2** may be expressed by saying that *the determinant is linear as a function of each column.*

D3. *If the two columns of the matrix are equal, then the determinant is equal to* 0. *In other words,*

$$D(B, B) = 0.$$

Proof. This is obvious, because

$$\begin{vmatrix} b_1 & b_1 \\ b_2 & b_2 \end{vmatrix} = b_1b_2 - b_2b_1 = 0.$$

The two vectors

$$E^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad E^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

are the standard **unit vectors.** The matrix formed by them, namely

$$E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

is the **unit matrix.** We have:

D4. *If E is the matrix, then $D(E) = D(E^1, E^2) = 1$.*

This is obvious.

These four basic properties are fundamental, and other properties can be deduced from them, without going back to the definition of the determinant in terms of the components of the matrix.

D5. *If we add a multiple of one column to the other, then the value of the determinant does not change. In other words, let x be a number. Then*

$$D(B + xC, C) = D(B, C) \quad \textit{and} \quad D(B, C + xB) = D(B, C).$$

Written out in terms of components, the first relation reads

$$\begin{vmatrix} b_1 + xc_1 & c_1 \\ b_2 + xc_2 & c_2 \end{vmatrix} = \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}.$$

Proof. Using **D1, D2, D3** in succession, we find that

$$\begin{aligned} D(B + xC, C) &= D(B, C) + D(xC, C) \\ &= D(B, C) + xD(C, C) = D(B, C). \end{aligned}$$

A similar proof applies to $D(B, C + xB)$.

D6. *If the two columns are interchanged, then the value of the determinant changes by a sign. In other words, we have*

$$D(B, C) = -D(C, B).$$

Proof. Again, we use **D1, D2, D4** successively, and get

$$\begin{aligned} 0 = D(B + C, B + C) &= D(B, B + C) + D(C, B + C) \\ &= D(B, B) + D(B, C) + D(C, B) + D(C, C) \\ &= D(B, C) + D(C, B). \end{aligned}$$

This proves that $D(B, C) = -D(C, B)$, as desired.

Of course, you can also give a proof using the components of the matrix. Do this as an exercise. However, there is some point in doing it as above, because in the study of determinants in the higher-dimensional case later, a proof with components becomes much messier, while the proof following the same pattern as the one we have given remains neat.

Exercises

1. Compute the following determinants.

(a) $\begin{vmatrix} 3 & -5 \\ 4 & 2 \end{vmatrix}$ (b) $\begin{vmatrix} 2 & -1 \\ -3 & 4 \end{vmatrix}$ (c) $\begin{vmatrix} -3 & 4 \\ 2 & -1 \end{vmatrix}$

(d) $\begin{vmatrix} -5 & 3 \\ 4 & 6 \end{vmatrix}$ (e) $\begin{vmatrix} 3 & 3 \\ -7 & -8 \end{vmatrix}$ (f) $\begin{vmatrix} -5 & -4 \\ 6 & 3 \end{vmatrix}$

2. Compute the determinant

$$\begin{vmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{vmatrix}$$

for any real number θ.

3. Compute the determinant

$$\begin{vmatrix} \cos\theta & \sin\theta \\ \sin\theta & \cos\theta \end{vmatrix}$$

when

(a) $\theta = \pi$, (b) $\theta = \pi/2$, (c) $\theta = \pi/3$, (d) $\theta = \pi/4$.

4. Prove:

(a) The other half of **D1.**
(b) The other half of **D2.**
(c) The other half of **D5.**

5. Let c be a number, and let A be a 2×2 matrix. Define cA to be the matrix obtained by multiplying all components of A by c. How does $D(cA)$ differ from $D(A)$?

§2. *Determinants of order 3*

We shall define the determinant for 3×3 matrices, and we shall see that it satisfies properties analogous to those of the 2×2 case.

Let

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

be a 3×3 matrix. We define its **determinant** according to the formula

known as the **expansion by a row,** say the first row. That is, we define

$$(1) \qquad \boxed{D(A) = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33}\end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33}\end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32}\end{vmatrix}.}$$

and we denote $D(A)$ also with the two vertical bars

$$D(A) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33}\end{vmatrix}.$$

We may describe the sum in (1) as follows. Let A_{ij} be the matrix obtained from A by deleting the i-th row and the j-th column. Then the sum for $D(A)$ can be written as

$$a_{11}D(A_{11}) - a_{12}D(A_{12}) + a_{13}D(A_{13}).$$

In other words, each term consists of the product of an element of the first row and the determinant of the 2×2 matrix obtained by deleting the first row and the j-th column, and putting the appropriate sign to this term as shown.

Example. Let

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 4 \\ -3 & 2 & 5\end{pmatrix}.$$

Then

$$A_{11} = \begin{pmatrix} 1 & 4 \\ 2 & 5\end{pmatrix}, \quad A_{12} = \begin{pmatrix} 1 & 4 \\ -3 & 5\end{pmatrix}, \quad A_{13} = \begin{pmatrix} 1 & 1 \\ -3 & 2\end{pmatrix}$$

and our formula for the determinant of A yields

$$\begin{aligned} D(A) &= 2\begin{vmatrix} 1 & 4 \\ 2 & 5\end{vmatrix} - 1\begin{vmatrix} 1 & 4 \\ -3 & 5\end{vmatrix} + 0\begin{vmatrix} 1 & 1 \\ -3 & 2\end{vmatrix} \\ &= 2(5 - 8) - 1(5 + 12) + 0 \\ &= -23.\end{aligned}$$

Thus the determinant is a number. To compute this number in the above example, we computed the determinants of the 2×2 matrices explicitly. We can also expand these in the general definition, and thus we find a six-term expression for the determinant of a general 3×3 matrix $A = (a_{ij})$, namely:

$$\boxed{\begin{aligned} D(A) = {} & a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{12}a_{21}a_{33} \\ & + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}. \end{aligned}} \tag{2}$$

Do not memorize (2). Remember only (1), and write down (2) only when needed for specific purposes.

We could have used the other rows to expand the determinant, instead of the first row. For instance, the expansion according to the second row is given by

$$\begin{aligned} & -a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{22}\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} - a_{23}\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ & = -a_{21}D(A_{21}) + a_{22}D(A_{22}) - a_{23}D(A_{23}). \end{aligned}$$

Again, each term is the product of a_{2j} with the determinant of the 2×2 matrix obtained by deleting the second row and j-th column, together with the appropriate sign in front of each term. This sign is determined according to the pattern:

$$\begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}.$$

If you write down the two terms for each one of the 2×2 determinants in the expansion according to the second row, you will obtain six terms, and you will find immediately that they give you the same value which we wrote down in formula (2). Thus expanding according to the second row gives the same value for the determinant as expanding according to the first row.

Furthermore, we can also expand according to any one of the columns. For instance, expanding according to the first column, we find that

$$a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$$

yields precisely the same six terms as in (2), if you write down each one of the two terms corresponding to each one of the 2×2 determinants in the above expression.

Example. We compute the determinant

$$\begin{vmatrix} 3 & 0 & 1 \\ 1 & 2 & 5 \\ -1 & 4 & 2 \end{vmatrix}$$

by expanding according to the second column. The determinant is equal to

$$2\begin{vmatrix} 3 & 1 \\ -1 & 2 \end{vmatrix} - 4\begin{vmatrix} 3 & 1 \\ 1 & 5 \end{vmatrix} = 2(6 - (-1)) - 4(15 - 1) = -42.$$

Note that the presence of 0 in the first row and second column eliminates one term in the expansion, since this term is equal to 0.

If we expand the above determinant according to the third column, we find the same value, namely

$$+1\begin{vmatrix} 1 & 2 \\ -1 & 4 \end{vmatrix} - 5\begin{vmatrix} 3 & 0 \\ -1 & 4 \end{vmatrix} + 2\begin{vmatrix} 3 & 0 \\ 1 & 2 \end{vmatrix} = -42.$$

Theorem 2. *If A is a 3×3 matrix, then $D(A) = D({}^tA)$. In other words, the determinant of A is equal to the determinant of the transpose of A.*

Proof. This is true because expanding $D(A)$ according to rows or columns gives the same value, namely the expression in (2).

Exercises

1. Write down the expansion of a 3×3 determinant according to the third row, the second column, and the third column, and verify in each case that you get the same six terms as in (2).
2. Compute the following determinants by expanding according to the second row, and also according to the third column, as a check for your computation. Of course, you should find the same value.

(a) $\begin{vmatrix} 2 & 1 & 2 \\ 0 & 3 & -1 \\ 4 & 1 & 1 \end{vmatrix}$ (b) $\begin{vmatrix} 3 & -1 & 5 \\ -1 & 2 & 1 \\ -2 & 4 & 3 \end{vmatrix}$ (c) $\begin{vmatrix} 2 & 4 & 3 \\ -1 & 3 & 0 \\ 0 & 2 & 1 \end{vmatrix}$

(d) $\begin{vmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \\ 0 & 2 & 7 \end{vmatrix}$ (e) $\begin{vmatrix} -1 & 5 & 3 \\ 4 & 0 & 0 \\ 2 & 7 & 8 \end{vmatrix}$ (f) $\begin{vmatrix} 3 & 1 & 2 \\ 4 & 5 & 1 \\ -1 & 2 & -3 \end{vmatrix}$

3. Compute the following determinants.

(a) $\begin{vmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 7 \end{vmatrix}$ (b) $\begin{vmatrix} -3 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -8 \end{vmatrix}$ (c) $\begin{vmatrix} 6 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -2 \end{vmatrix}$

4. Let a, b, c be numbers. In terms of a, b, c, what is the value of the determinant

$$\begin{vmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{vmatrix} ?$$

5. Find the determinants of the following matrices.

(a) $\begin{pmatrix} 1 & 2 & 5 \\ 0 & 1 & 7 \\ 0 & 0 & 3 \end{pmatrix}$ (b) $\begin{pmatrix} -1 & 5 & 20 \\ 0 & 4 & 8 \\ 0 & 0 & 6 \end{pmatrix}$

(c) $\begin{pmatrix} 2 & -6 & 9 \\ 0 & 1 & 4 \\ 0 & 0 & 8 \end{pmatrix}$ (d) $\begin{pmatrix} -7 & 98 & 54 \\ 0 & 2 & 46 \\ 0 & 0 & -1 \end{pmatrix}$

(e) $\begin{pmatrix} 1 & 4 & 6 \\ 0 & 0 & 1 \\ 0 & 0 & 8 \end{pmatrix}$ (f) $\begin{pmatrix} 4 & 0 & 0 \\ -5 & 2 & 0 \\ 79 & 54 & 1 \end{pmatrix}$

(g) $\begin{pmatrix} 1 & 5 & 2 \\ 0 & 2 & 7 \\ 0 & 0 & 4 \end{pmatrix}$ (h) $\begin{pmatrix} -5 & 0 & 0 \\ 7 & 2 & 0 \\ -9 & 4 & 1 \end{pmatrix}$

6. In terms of the components of the matrix, what is the value of the determinant:

(a) $\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{vmatrix}$? (b) $\begin{vmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$?

§3. *Additional properties of determinants*

We shall now see that 3×3 determinants satisfy the properties **D1** through **D6**, listed previously for 2×2 determinants. These properties are concerned with the columns of the matrix, and hence it is useful to use the same notation which we used before. If A^1, A^2, A^3 are the columns of the 3×3 matrix A, then we write

$$D(A) = D(A^1, A^2, A^3).$$

For the rest of this section, we assume that our column and row vectors have dimension 3; that is, that they have three components. Thus any column vector B in this section can be written in the form

$$B = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

D1. *Suppose that the first column can be written as a sum,*

$$A^1 = B + C,$$

that is,

$$\begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}.$$

Then

$$D(B + C, A^2, A^3) = D(B, A^2, A^3) + D(C, A^2, A^3).$$

and the analogous rule holds with respect to the second and third columns.

Proof. We use the definition of the determinant, namely the expansion according to the first row. We see that each term splits into a sum of two terms corresponding to B and C. For instance,

$$a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} = b_1\begin{vmatrix} a_{22} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + c_1\begin{vmatrix} a_{22} & a_{23} \\ a_{31} & a_{33} \end{vmatrix},$$

$$a_{12}\begin{vmatrix} b_2 + c_2 & a_{23} \\ b_3 + c_3 & a_{33} \end{vmatrix} = a_{12}\begin{vmatrix} b_2 & a_{23} \\ b_3 & a_{33} \end{vmatrix} + a_{12}\begin{vmatrix} c_2 & a_{23} \\ c_3 & a_{33} \end{vmatrix},$$

$$a_{13}\begin{vmatrix} b_2 + c_2 & a_{22} \\ b_3 + c_3 & a_{32} \end{vmatrix} = a_{13}\begin{vmatrix} b_2 & a_{22} \\ b_3 & a_{32} \end{vmatrix} + a_{13}\begin{vmatrix} c_2 & a_{22} \\ c_3 & a_{32} \end{vmatrix}.$$

Summing with the appropriate sign yields the desired relation.

D2. *If x is a number, then*

$$D(xA^1, A^2, A^3) = x \cdot D(A^1, A^2, A^3),$$

and similarly for the other columns.

Proof. We have

$$D(xA^1, A^2, A^3) = xa_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} xa_{21} & a_{23} \\ xa_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} xa_{21} & a_{22} \\ xa_{31} & a_{32} \end{vmatrix}$$

$$= x \cdot D(A^1, A^2, A^3).$$

The proof is similar for the other columns.

D3. *If two columns of the matrix are equal, then the determinant is equal to* 0.

Proof. Suppose that $A^1 = A^2$, and look at the expansion of the determinant according to the first row. Then $a_{11} = a_{12}$, and the first two terms cancel. The third term is equal to 0 because it involves a 2×2 determinant whose two columns are equal. The proof for the other cases is similar. (Other cases: $A^2 = A^3$ and $A^1 = A^3$.)

In the 3×3 case, we also have the **unit vectors,** namely

$$E^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad E^2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad E^3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

and the unit **3×3 matrix,** namely

$$E = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

D4. *If E is the unit matrix, then $D(E) = D(E^1, E^2, E^3) = 1$.*

Proof. This is obvious from the expansion according to the first row.

Observe that to prove our basic four properties, we needed to use the definition of the determinant, i.e. its expansion according to the first row. For the remaining properties, we can give a proof which is not based directly on this expansion, but only on the formalism of **D1** through **D4.** This has the advantage of making the arguments easier, and in fact of making them completely analogous to those used in the 2×2 case. We carry them out.

D5. *If we add a multiple of one column to another, then the value of the determinant does not change. In other words, let x be a number. Then for instance,*

$$D(A^1, A^2 + xA^1, A^3) = D(A^1, A^2, A^3),$$

and similarly in all other cases.

Proof. We have

$$\begin{aligned} D(A^1, A^2, + xA^1, A^3) &= D(A^1, A^2, A^3) + D(A^1, xA^1, A^3) && \text{(by } \mathbf{D1}) \\ &= D(A^1, A^2, A^3) + x \cdot D(A^1, A^1, A^3) && \text{(by } \mathbf{D2}) \\ &= D(A^1, A^2, A^3) && \text{(by } \mathbf{D4}). \end{aligned}$$

This proves what we wanted. The proofs of the other cases are similar.

D6. *If two adjacent columns are interchanged, then the determinant changes by a sign. In other words, we have*

$$D(A^1, A^3, A^2) = -D(A^1, A^2, A^3),$$

and similarly in the other case.

Proof. We use the same method as before. We find

$$\begin{aligned} 0 &= D(A^1, A^2 + A^3, A^2 + A^3) \\ &= D(A^1, A^2, A^2 + A^3) + D(A^1, A^3, A^2 + A^3) \\ &= D(A^1, A^2, A^2) + D(A^1, A^2, A^3) + D(A^1, A^3, A^2) + D(A^1, A^3, A^3) \\ &= D(A^1, A^2, A^3) + D(A^1, A^3, A^2), \end{aligned}$$

using **D1** and **D3.** This proves **D6** in this case, and the other cases are proved similarly.

Using these rules, especially **D5,** we can compute determinants a little more efficiently. For instance, we have already noticed that when a 0 occurs in the given matrix, we can expand according to the row (or column) in which this 0 occurs, and it eliminates one term. Using **D5** repeatedly,

we can change the matrix so as to get as many zeros as possible, and then reduce the computation to one term.

Furthermore, knowing that the determinant of A is equal to the determinant of its tranpose, we can also conclude that properties **D1** through **D6** hold for rows instead of columns. For instance, we can state **D6** for rows:

If two adjacent rows are interchanged, then the determinant changes by a sign.

As an exercise, state all the other properties for rows.

Example. Compute the determinant

$$\begin{vmatrix} 3 & 0 & 1 \\ 1 & 2 & 5 \\ -1 & 4 & 2 \end{vmatrix}.$$

We already have 0 in the first row. We subtract two times the second row from the third row. Our determinant is then equal to

$$\begin{vmatrix} 3 & 0 & 1 \\ 1 & 2 & 5 \\ -3 & 0 & -8 \end{vmatrix}.$$

We expand according to the second column. The expansion has only one term $\neq 0$, with a $+$ sign, and that is:

$$2\begin{vmatrix} 3 & 1 \\ -3 & -8 \end{vmatrix}.$$

The 2×2 determinant can be evaluated by our definition of $ad - bc$, and we find the value

$$2(-24 - (-3)) = -42.$$

Example. We compute the determinant

$$\begin{vmatrix} 4 & 7 & 10 \\ 3 & 7 & 5 \\ 5 & -1 & 10 \end{vmatrix}.$$

We subtract two times the second row from the first row, and then from the third row, yielding

$$\begin{vmatrix} -2 & -7 & 0 \\ 3 & 7 & 5 \\ -1 & -15 & 0 \end{vmatrix},$$

which we expand according to the third column, and get

$$\begin{aligned} -5(30 - 7) &= -5(23) \\ &= -115. \end{aligned}$$

Note that the term has a minus sign, determined by our usual pattern of signs.

Determinants can also be defined for $n \times n$ matrices, satisfying analogous properties to **D1** through **D6.** The proofs are similar, but involve sometimes more complicated notation, so we shall not go into them.

Exercises

1. (a) Write out in full and prove property **D1** with respect to the second column and the third column.
   (b) Same thing for property **D2.**
2. Prove the two cases not treated in the text for property **D3.**
3. Prove **D5** in the case
   (a) you add a multiple of the third column to the first;
   (b) you add a multiple of the second column to the first;
   (c) you add a multiple of the third column to the second.
4. If you interchange the first and third columns of the given matrix, how does its determinant change? What about interchanging the first and third row?
5. Compute the following determinants.

(a) $\begin{vmatrix} 2 & 1 & 2 \\ 0 & 3 & -1 \\ 4 & 1 & 1 \end{vmatrix}$ (b) $\begin{vmatrix} 3 & -1 & 5 \\ -1 & 2 & 1 \\ -2 & 4 & 3 \end{vmatrix}$ (c) $\begin{vmatrix} 2 & 4 & 3 \\ -1 & 3 & 0 \\ 0 & 2 & 1 \end{vmatrix}$

(d) $\begin{vmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \\ 0 & 2 & 7 \end{vmatrix}$ (e) $\begin{vmatrix} -1 & 5 & 3 \\ 4 & 0 & 0 \\ 2 & 7 & 8 \end{vmatrix}$ (f) $\begin{vmatrix} 3 & 1 & 2 \\ 4 & 5 & 1 \\ -1 & 2 & -3 \end{vmatrix}$

6. Compute the following determinants.

(a) $\begin{vmatrix} 1 & 1 & 3 \\ -1 & 1 & 0 \\ 1 & 2 & 5 \end{vmatrix}$ (b) $\begin{vmatrix} 3 & 2 & 1 \\ 4 & 1 & 2 \\ 1 & 5 & 7 \end{vmatrix}$ (c) $\begin{vmatrix} 3 & 1 & 1 \\ 2 & 5 & 5 \\ 8 & 7 & 7 \end{vmatrix}$

(d) $\begin{vmatrix} 4 & -9 & 2 \\ 4 & -9 & 2 \\ 3 & 1 & 0 \end{vmatrix}$ (e) $\begin{vmatrix} 4 & -1 & 1 \\ 2 & 0 & 0 \\ 1 & 5 & 7 \end{vmatrix}$ (f) $\begin{vmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 8 & 5 & 7 \end{vmatrix}$

(g) $\begin{vmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 27 \end{vmatrix}$ (h) $\begin{vmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 9 \end{vmatrix}$ (i) $\begin{vmatrix} 2 & -1 & 4 \\ 3 & 1 & 5 \\ 1 & 2 & 3 \end{vmatrix}$

7. In general, what is the determinant of a diagonal matrix

$$\begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{vmatrix} ?$$

8. Compute the following determinants, making the computation as easy as you can.

(a) $\begin{vmatrix} 4 & -9 & 2 \\ 4 & -9 & 2 \\ 3 & 1 & 5 \end{vmatrix}$ (b) $\begin{vmatrix} 4 & -1 & 1 \\ 2 & 0 & 0 \\ 1 & 5 & 7 \end{vmatrix}$ (c) $\begin{vmatrix} 2 & -1 & 4 \\ 1 & 1 & 5 \\ 1 & 2 & 3 \end{vmatrix}$

(d) $\begin{vmatrix} 3 & 1 & 1 \\ 2 & 5 & 5 \\ 8 & 7 & 7 \end{vmatrix}$ (e) $\begin{vmatrix} 2 & 1 & 1 \\ 3 & 1 & 5 \\ 4 & -2 & 3 \end{vmatrix}$ (f) $\begin{vmatrix} -4 & 4 & 2 \\ 5 & 1 & 3 \\ 2 & 1 & 4 \end{vmatrix}$

(g) $\begin{vmatrix} 7 & 3 & 2 \\ 1 & -1 & 1 \\ 2 & 1 & 3 \end{vmatrix}$ (h) $\begin{vmatrix} 3 & 2 & 1 \\ 1 & 1 & 1 \\ -1 & 3 & 4 \end{vmatrix}$ (i) $\begin{vmatrix} -2 & -1 & 1 \\ 3 & 1 & -1 \\ -1 & 2 & 3 \end{vmatrix}$

(j) $\begin{vmatrix} 2 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 2 & 2 \end{vmatrix}$ (k) $\begin{vmatrix} -4 & 1 & 2 \\ 3 & 2 & 1 \\ -1 & -1 & 1 \end{vmatrix}$ (l) $\begin{vmatrix} -1 & 3 & 2 \\ 3 & -1 & 1 \\ 6 & -2 & 2 \end{vmatrix}$

9. Let c be a number and multiply each component a_{ij} of a 3×3 matrix A by c, thus obtaining a new matrix which we denote by cA. How does $D(A)$ differ from $D(cA)$?

10. Let x_1, x_2, x_3 be numbers. Show that

$$\begin{vmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{vmatrix} = (x_2 - x_1)(x_3 - x_2)(x_3 - x_1).$$

11. Suppose that A^1 is a sum of three columns, say

$$A^1 = B^1 + B^2 + B^3.$$

Using **D1** twice, prove that

$$D(B^1 + B^2 + B^3, A^2, A^3)$$
$$= D(B^1, A^2, A^3) + D(B^2, A^2, A^3) + D(B^3, A^2, A^3).$$

Using summation notation, we can write this in the form

$$D(B^1 + B^2 + B^3, A^2, A^3) = \sum_{j=1}^{3} D(B^j, A^2, A^3),$$

which is shorter. In general, suppose that

$$A^1 = \sum_{j=1}^{n} B^j$$

is a sum of n columns. Using the summation notation, express similarly

$$D(A^1, A^2, A^3)$$

as a sum of (how many?) terms.

12. Let x_j $(j = 1, 2, 3)$ be numbers. Let

$$A^1 = x_1C^1 + x_2C^2 + x_3C^3.$$

Prove that

$$D(A^1, A^2, A^3) = \sum_{j=1}^{3} x_jD(C^j, A^2, A^3).$$

State and prove the analogous statement when

$$A^1 = \sum_{j=1}^{n} x_jC^j.$$

13. State the analogous property to that of Exercise 12 with respect to the second column. Then with respect to the third column.

14. If $a(t)$, $b(t)$, $c(t)$, $d(t)$ are functions of t, one can form the determinant

$$\begin{vmatrix} a(t) & b(t) \\ c(t) & d(t) \end{vmatrix},$$

just as with numbers. Write out in full the determinant

$$\begin{vmatrix} \sin t & \cos t \\ -\cos t & \sin t \end{vmatrix}.$$

15. Write out in full the determinant

$$\begin{vmatrix} t+1 & t-1 \\ t & 2t+5 \end{vmatrix}.$$

16. Let $f(t)$, $g(t)$ be two functions having derivatives of all orders. Let $\varphi(t)$ be the function obtained by taking the determinant

$$\varphi(t) = \begin{vmatrix} f(t) & g(t) \\ f'(t) & g'(t) \end{vmatrix}.$$

Show that

$$\varphi'(t) = \begin{vmatrix} f(t) & g(t) \\ f''(t) & g''(t) \end{vmatrix},$$

i.e. the derivative is obtained by taking the derivative of the bottom row.

17. Let

$$A(t) = \begin{pmatrix} b_1(t) & c_1(t) \\ b_2(t) & c_2(t) \end{pmatrix}$$

be a 2×2 matrix of differentiable functions. Let $B(t)$ and $C(t)$ be its column vectors. Let

$$\varphi(t) = \operatorname{Det}(A(t)).$$

Show that

$$\varphi'(t) = D(B'(t), C(t)) + D(B(t), C'(t)).$$

§4. *Independence of vectors*

In the geometric applications of Chapter IX, we studied parallelograms and parallelotopes spanned by vectors. Let us look at the situation in 3-space. Let A, B, C be vectors in $\mathbf{R}^3$, and suppose that A, B are independent. We define the **plane spanned by A and B** to be the set of all points

$$xA + yB,$$

with all real numbers x, y. When $x = y = 0$ we obtain the origin, so the plane passes through the origin and looks like this.

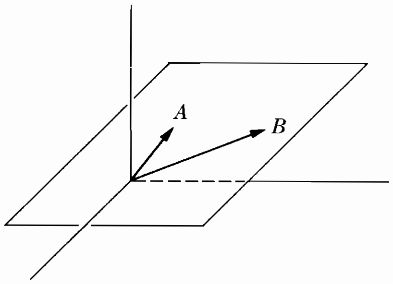


Figure 1

We say that C is **independent** of A and B if C does not lie in the above plane, i.e. if C cannot be written in the form

$$C = xA + yB$$

with some numbers x and y. Geometrically, this means that C points in a direction outside the plane, as shown on Fig. 2.

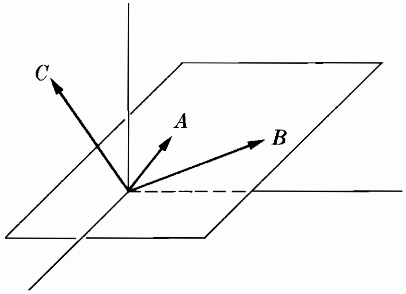


Figure 2

We shall now see that the determinant gives us a criterion when C is independent of A and B.

Theorem 3. *Let A, B, C be in $\mathbf{R}^3$. If $D(A, B, C) \neq 0$, then C is independent of A and B.*

Proof. Suppose that $C = xA + yB$ with some numbers x, y. Then

$$\begin{aligned} D(A, B, C) &= D(A, B, xA + yB) \\ &= D(A, B, xA) + D(A, B, yB) \\ &= xD(A, B, A) + yD(A, B, B) \\ &= 0 \qquad \text{(why?).} \end{aligned}$$

This is against our hypothesis, and thus proves our theorem.

Exercises

In the following exercises, let A, B, C be in $\mathbf{R}^3$ and assume that the determinant $D(A, B, C)$ is $\neq 0$. Prove

1. There is no number x such that $B = xA$.
2. There is no number x such that $B = xC$.
3. A is independent of B and C.
4. B is independent of A and C.
5. Let x, y, z be numbers such that $xA + yB + zC = 0$. Then $x = y = z = 0$.
6. Draw a picture of the set of all points

$$xA + yB + zC,$$

with $0 \leqq x \leqq 1$, $0 \leqq y \leqq 1$, and $0 \leqq z \leqq 1$, in 3-space. This set is called the **box** (or **parallelotope**) spanned by A, B, C.

§5. Determinant of a product

Theorem 4. *Let A, B be 3×3 matrices. Then*

$$D(AB) = D(A)D(B).$$

In other words, the determinant of a product is the product of the determinants.

Proof. Let $AB = C$ and let C^m be the m-th column of C. From the definition of the product of matrices, one sees that if X is a column vector, then

$$AX = x_1A^1 + x_2A^2 + x_3A^3.$$

Apply this remark to each one of the columns of B successively, that is, $X = B^1$, $X = B^2$, and $X = B^3$ to find the respective columns of C. We conclude that

$$C^m = b_{1m}A^1 + b_{2m}A^2 + b_{3m}A^3.$$

Therefore

$$\begin{aligned} D(AB) = D(C) &= D\left(\sum_{i=1}^{3} b_{i1}A^i, \sum_{j=1}^{3} b_{j2}A^j, \sum_{k=1}^{3} b_{k3}A^k\right) \\ &= \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} b_{i1}b_{j2}b_{k3}D(A^i, A^j, A^k). \end{aligned}$$

Here we have used repeatedly linearity with respect to each column. Any term on the right in the sum will be 0 if $i = j$, or $i = k$, or $j = k$. The other terms will correspond to a permutation of A^1, A^2, A^3, and there will be six such terms. If you write them out, and interchange columns making the

appropriate sign change, you will find that the sum is equal to the six-term expansion for the determinant of B times the determinant of A, in other words

$$D(AB) = D(B)D(A).$$

This proves our theorem.

Observe that if A is invertible and $AB = I$, then we necessarily have $D(A) \neq 0$, because according to Theorem 4,

$$1 = D(I) = D(A)D(B).$$

The converse is also true, that is: If $D(A) \neq 0$, then A is invertible. We shall discuss it in the next section.

§6. *Inverse of a matrix*

Theorem 5. *Let A be a square matrix such that $D(A) \neq 0$. Then A is invertible.*

Let us consider the 2×2 case. Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

be a 2×2 matrix, and assume that its determinant $ad - bc \neq 0$. We wish to find an inverse for A, that is a 2×2 matrix

$$X = \begin{pmatrix} x & y \\ z & w \end{pmatrix}$$

such that

$$AX = XA = I.$$

Let us look at the first requirement, $AX = I$, which, written out in full, looks like this:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x & y \\ z & w \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Let us look at the first column of AX. We must solve the equations

$$(*) \qquad \begin{aligned} ax + bz &= 1, \\ cx + dz &= 0. \end{aligned}$$

This is a system of two equations in two unknowns, x and z, which we know how to solve. Similarly, looking at the second column, we see that we must solve a system of two equations in the unknowns y, w, namely

$$(**)\qquad \begin{aligned} ay + bw &= 0, \\ cy + dw &= 1. \end{aligned}$$

Example. Let

$$A = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix}.$$

We seek a matrix X such that $AX = I$. We must therefore solve the systems of linear equations

$$\begin{aligned} 2x + z &= 1, \\ 4x + 3z &= 0, \end{aligned} \qquad \text{and} \qquad \begin{aligned} 2y + w &= 0, \\ 4y + 3w &= 1. \end{aligned}$$

By the ordinary method of solving two equations in two unknowns, we find

$$x = 1, \qquad z = -1 \qquad \text{and} \qquad y = -\tfrac{1}{2}, \qquad w = 1.$$

Thus the matrix

$$X = \begin{pmatrix} 1 & -\frac{1}{2} \\ -1 & 1 \end{pmatrix}$$

is such that $AX = I$. The reader will also verify by direct multiplication that $XA = I$. This solves for the desired inverse.

The same procedure, of course, works for the general systems (*) and (**). Consider (*). Multiply the first equation by d, multiply the second equation by b, and subtract. We get

$$(ad - bc)x = d,$$

whence

$$x = \frac{d}{ad - bc}.$$

We see that the determinant of A occurs in the denominator. You can solve similarly for y, z, w and you will find similar expressions with only $D(A)$ in the denominator. This proves Theorem 5 in the 2×2 case.

The proof in the 3×3 case is also done by solving linear equations, but we shall omit it.

Exercises

1. Find the inverses of the following matrices.

(a) $\begin{pmatrix} 2 & -1 \\ 5 & 2 \end{pmatrix}$ (b) $\begin{pmatrix} 3 & 4 \\ -2 & 1 \end{pmatrix}$ (c) $\begin{pmatrix} 5 & 1 \\ 1 & 2 \end{pmatrix}$ (d) $\begin{pmatrix} -2 & -1 \\ -3 & -4 \end{pmatrix}$

2. Write down the general formula for the inverse of a 2×2 matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

PART THREE

MAPPINGS FROM VECTORS TO VECTORS

One way in which we use linear algebra is to approximate an arbitrary mapping by a linear mapping. The derivative of an arbitrary mapping is in a sense the "best" linear approximation to the mapping. This notion is discussed in the chapter of this part. When studying parametrized surfaces in Chapter XV, it should be useful to have understood the idea of linear approximation, since it allows us to define the tangent plane to such a surface in a natural way, as the image of the approximating linear map.

If you are not interested in the idea of the proof behind the implicit function theorem, you can work mechanically without knowing anything about inverse mappings in order to compute the derivative of an implicit function.

CHAPTER XI

Applications to Functions of Several Variables

Having acquired the language of linear maps and matrices, we shall be able to define the derivative of a mapping, or rather, of a differentiable mapping. The theoretical considerations involved in the proof of the general chain rule of §3 become of course a little abstract. But you should note that it is precisely the availability of the notion of linear mapping which allows us to give a statement of the chain rule, and a proof, which runs exactly parallel to the proof for functions of one variable, as given in the *First Course*. The analysis profits from algebra, and conversely, the algebra of linear mappings finds a neat application which enhances its attractiveness.

§1. The derivative as a linear map

We shall interpret our notion of differentiability given in Chapter III in terms of linear mappings.

Let U be an open set in $\mathbf{R}^n$. Let f be a function defined on U. Let P be a point of U, and assume that f is differentiable at P. Then there is a vector A, and a function g such that for all small vectors H we can write

$$f(P + H) = f(P) + A \cdot H + \|H\|g(H), \tag{1}$$

and

$$\lim_{\|H\| \to 0} g(H) = 0. \tag{2}$$

The vector A, expressed in terms of coordinates, is none other than the vector of partial derivatives:

$$A = \operatorname{grad} f(P) = \big(D_1 f(P), \ldots, D_n f(P)\big).$$

We have seen that there is a linear map $L = L_A$ such that

$$L(H) = A \cdot H.$$

Our condition that f is differentiable may therefore be expressed by saying that there is a linear map $L\colon \mathbf{R}^n \to \mathbf{R}$ and a function g defined for

sufficiently small H, such that

$$f(P + H) = f(P) + L(H) + \|H\|g(H) \tag{3}$$

and

$$\lim_{\|H\| \to 0} g(H) = 0.$$

In the case of functions of one variable, we have of course the same kind of formula, namely

$$f(a + h) = f(a) + ch + |h|g(h)$$

where

$$\lim_{h \to 0} g(h) = 0.$$

Here, a, h are numbers, and so is the ordinary derivative c. But the map $L_c\colon \mathbf{R} \to \mathbf{R}$ such that $L_c(h) = ch$ (multiplication by the number c) is a linear map, so that also in this case, we can write

$$f(a + h) = f(a) + L_c(h) + |h|g(h).$$

Up to now, we did not define the notion of derivative for functions of several variables. We now define the **derivative** of f at P to be this linear map, which we shall denote by $Df(P)$ or also $f'(P)$. This notation is therefore entirely similar to the notation used for functions of one variable. We could not make the definition before we knew what a linear map was. All the theory developed in Chapters II through VII could be carried out knowing only dot products, and this is the reason we postponed making the general definition of derivative until now.

If L is a linear map, then it will be useful to omit some parentheses in order to simplify the notation. Thus we shall sometimes write Lv instead of $L(v)$. With this convention, we can write (3) in the form

$$f(P + H) = f(P) + Df(P)H + \|H\|g(H), \tag{4}$$

or also

$$f(P + H) = f(P) + f'(P)H + \|H\|g(H). \tag{5}$$

These ways of expressing differentiability are those which generalize to arbitrary mappings.

Let U be an open set in $\mathbf{R}^n$. Let $F\colon U \to \mathbf{R}^m$ be a mapping. Let P be a point of U. We shall say that F is **differentiable** at P if there exists a linear map

$$L\colon \mathbf{R}^n \to \mathbf{R}^m$$

and a mapping G defined for all vectors H sufficiently small, such that we have

$$F(P + H) = F(P) + LH + \|H\|G(H) \tag{6}$$

and

$$\lim_{\|H\| \to 0} G(H) = O. \tag{7}$$

If such a linear mapping L exists, then we interpret (6) as saying that L approximates F up to an error term whose magnitude is small, near the point P.

A linear map L satisfying conditions (6) and (7) will be said to be **tangent** to F at P. It is also said to be the best linear approximation to F at P.

Just as before, we define a map ψ defined for small H to be $o(H)$ ("little oh of H") if

$$\lim_{\|H\| \to 0} \frac{\psi(H)}{\|H\|} = O.$$

Then we can write our definition of differentiability in the form

$$F(P + H) = F(P) + L(H) + o(H),$$

where L is a linear map.

Theorem 1. *Suppose that there exist linear maps L, M which are tangent to F at P. Then $L = M$. In other words, if there exists one linear map which is tangent to F at P, then there is only one.*

Proof. Suppose that there are two mappings G_1, G_2 such that for all sufficiently small H, we have

$$F(P + H) = F(P) + LH + \|H\|G_1(H),$$
$$F(P + H) = F(P) + MH + \|H\|G_2(H),$$

and

$$\lim_{\|H\| \to 0} G_1(H) = O, \qquad \lim_{\|H\| \to 0} G_2(H) = O.$$

We must show that for any vector Y we have $LY = MY$. Let t range over small positive numbers. Then tY is small, and $P + tY$ lies in U. Thus $F(P + tY)$ is defined. By hypothesis, we have

$$F(P + tY) = F(P) + L(tY) + \|tY\|G_1(tY),$$
$$F(P + tY) = F(P) + M(tY) + \|tY\|G_2(tY).$$

Subtracting, we obtain

$$O = L(tY) - M(tY) + \|tY\|[G_1(tY) - G_2(tY)].$$

Let $G = G_1 - G_2$. Since L, M are linear, we can write $L(tY) = tL(Y)$ and $M(tY) = tM(Y)$. Consequently, we obtain

$$tM(Y) - tL(Y) = t\|Y\|G(tY).$$

Take $t \neq 0$. Dividing by t yields

$$M(Y) - L(Y) = \|Y\|G(tY).$$

As t approaches 0, $G(tY)$ approaches O also. Hence the right-hand side of this last equation approaches O. But $M(Y) - L(Y)$ is a fixed vector. The only way this is possible is that $M(Y) - L(Y) = O$, in other words, $M(Y) = L(Y)$, as was to be shown.

If there exists a linear map tangent to F at P, we shall denote this linear map by $F'(P)$, or $DF(P)$ and call it the **derivative** of F at P. We may therefore write

$$F(P + H) = F(P) + F'(P)H + \|H\|G(H)$$

instead of (6).

In the next section, we shall see how the linear map $F'(P)$ can be computed, or rather how its matrix can be computed when we deal with vectors as n-tuples.

Exercises

1. Let $f: \mathbf{R} \to \mathbf{R}$ be a function, and let a be a number. Assume that there exists a linear map L tangent to f at a. Show that

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

2. Conversely, assume that the limit

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

exists and is equal to a number b. Let L_b be the linear map such that $L_b(x) = bx$ for all numbers x. Show that L_b is tangent to f at a. It is customary to identify the number b and the linear map L_b, and to call either one the derivative of f at a.

3. Let $L: \mathbf{R} \to \mathbf{R}^n$ be a linear map from the reals into $\mathbf{R}^n$. Show that there is some element v in $\mathbf{R}^n$ such that $L(x) = xv$ for all numbers x.

4. Going back to Chapter II, let $X(t)$ be a curve, defined for all numbers t, say. Discuss in a manner analogous to Exercises 1 and 2 the derivative dX/dt, and the linear map $L_t\colon \mathbf{R} \to \mathbf{R}^n$ which is tangent to X at t.

§2. *The Jacobian matrix*

Throughout this section, all our vectors will be vertical vectors. We let $D_1, \ldots, D_n$ be the usual partial derivatives. Thus $D_i = \partial/\partial x_i$.

Let $F\colon \mathbf{R}^n \to \mathbf{R}^m$ be a mapping. We can represent F by coordinate functions. In other words, there exist functions $f_1, \ldots, f_m$ such that

$$F(X) = \begin{pmatrix} f_1(X) \\ f_2(X) \\ \vdots \\ f_m(X) \end{pmatrix} = {}^t(f_1(X), \ldots, f_m(X)).$$

To simplify the typography, we shall sometimes write a vertical vector as the transpose of a horizontal vector, as we have just done.

We view X as a column vector, $X = {}^t(x_1, \ldots, x_n)$.

Let us assume that the partial derivatives of each function f_i $(i = 1, \ldots, m)$ exist. We can then form the matrix of partial derivatives:

$$\left(\frac{\partial f_i}{\partial x_j}\right) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{pmatrix} = \begin{pmatrix} D_1 f_1(X) & \cdots & D_n f_1(X) \\ \vdots & & \vdots \\ D_1 f_m(X) & \cdots & D_n f_m(X) \end{pmatrix}$$

$i = 1, \ldots, m$ and $j = 1, \ldots, n$. This matrix is called the **Jacobian** matrix of F, and is denoted by $J_F(X)$.

In the case of two variables (x, y), say F is given by functions (f, g), so that

$$F(x, y) = (f(x, y), g(x, y)),$$

then Jacobian matrix is

$$J_F(x, y) = \begin{pmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \\ \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \end{pmatrix}.$$

(As we have done just now, we sometimes write the vectors horizontally, although to be strictly correct, they should be written vertically.)

Example 1. Let $F: \mathbf{R}^2 \to \mathbf{R}^2$ be the mapping defined by

$$F(x, y) = \begin{pmatrix} x^2 + y^2 \\ e^{xy} \end{pmatrix} = \begin{pmatrix} f(x, y) \\ g(x, y) \end{pmatrix}.$$

Find the Jacobian matrix $J_F(P)$ for $P = (1, 1)$.

The Jacobian matrix at an arbitrary point (x, y) is

$$\begin{pmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \\[2ex] \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \end{pmatrix} = \begin{pmatrix} 2x & 2y \\ ye^x & xe^y \end{pmatrix}.$$

Hence when $x = 1$, $y = 1$, we find:

$$J_F(1, 1) = \begin{pmatrix} 2 & 2 \\ e & e \end{pmatrix}.$$

Example 2. Let $F: \mathbf{R}^2 \to \mathbf{R}^3$ be the mapping defined by

$$F(x, y) = \begin{pmatrix} xy \\ \sin x \\ x^2y \end{pmatrix}.$$

Find $J_F(P)$ at the point $P = (\pi, \pi/2)$.

The Jacobian matrix at an arbitrary point (x, y) is

$$\begin{pmatrix} y & x \\ \cos x & 0 \\ 2xy & x^2 \end{pmatrix}.$$

Hence

$$J_F\left(\pi, \frac{\pi}{2}\right) = \begin{pmatrix} \pi/2 & \pi \\ -1 & 0 \\ \pi^2 & \pi^2 \end{pmatrix}.$$

Theorem 2. *Let U be an open set in $\mathbf{R}^n$. Let $F: U \to \mathbf{R}^m$ be a mapping, having coordinate functions $f_1, \ldots, f_m$. Assume that each function f_i is differentiable at a point X of U. Then F is differentiable at X, and the matrix representing the linear map $DF(X) = F'(X)$ is the Jacobian matrix $J_F(X)$.*

Proof. For each integer i between 1 and m, there is a function g_i such

that

$$\lim_{\|H\|\to 0} g_i(H) = 0,$$

and such that we can write

$$f_i(X + H) = f_i(X) + \operatorname{grad} f_i\,(X) \cdot H + \|H\| g_i(H).$$

We view X and $F(X)$ as vertical vectors. By definition, we can then write

$$F(X + H) = {}^t\big(f_1(X + H), \ldots, f_m(X + H)\big).$$

Hence

$$F(X + H) = \begin{pmatrix} f_1(X) \\ \vdots \\ f_m(X) \end{pmatrix} + \begin{pmatrix} \operatorname{grad} f_1\,(X) \cdot H \\ \vdots \\ \operatorname{grad} f_m\,(X) \cdot H \end{pmatrix} + \|H\| \begin{pmatrix} g_1(H) \\ \vdots \\ g_m(H) \end{pmatrix}.$$

The term in the middle, involving the gradients, is precisely equal to the product of the Jacobian matrix, times H, i.e. to

$$J_F(X)H.$$

Let $G(H) = {}^t\big(g_1(H), \ldots, g_m(H)\big)$ be the vector on the right. Then

$$F(X + H) = F(X) + J_F(X)H + \|H\|G(H).$$

As $\|H\|$ approaches 0, each coordinate of $G(H)$ approaches 0. Hence $G(H)$ approaches O; in other words,

$$\lim_{\|H\|\to 0} G(H) = O.$$

Hence the linear map represented by the matrix $J_F(X)$ is tangent to F at X. Since such a linear map is unique, we have proved our theorem.

Let U be open in $\mathbf{R}^n$ and $F\colon U \to \mathbf{R}^n$ be a differentiable map into the same dimensional space. Then the Jacobian matrix $J_F(X)$ is a square matrix, and its determinant is called the **Jacobian determinant** of F at X. We denote it by

$$\Delta_F(X).$$

Example 3. Let F be as in Example 2, $F(x, y) = (x^2 + y^2, e^{xy})$. Then the Jacobian determinant is equal to

$$\Delta_F(x, y) = \begin{vmatrix} 2x & 2y \\ ye^x & xe^y \end{vmatrix} = 2x^2e^y - 2y^2e^x.$$

In particular,

$$\Delta_F(1, 1) = 2e - 2e = 0,$$
$$\Delta_F(1, 2) = 2e^2 - 8e.$$

Example 4. An important map is given by the polar coordinates, $F\colon \mathbf{R}^2 \to \mathbf{R}^2$ such that $F(r, \theta) = (r\cos\theta, r\sin\theta)$. We can view the map as defined on all of $\mathbf{R}^2$, although when selecting polar coordinates, we take $r \geqq 0$. We see that F maps a rectangle into a circular sector (Fig. 1).

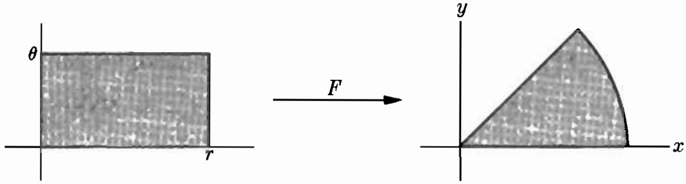


Figure 1

It is easy to compute the Jacobian matrix and determinant. Do this as Exercise 7.

Exercises

1. In each of the following cases, compute the Jacobian matrix of F.
   (a) $F(x, y) = (x + y, x^2y)$ (b) $F(x, y) = (\sin x, \cos xy)$
   (c) $F(x, y) = (e^{xy}, \log x)$ (d) $F(x, y, z) = (xz, xy, yz)$
   (e) $F(x, y, z) = (xyz, x^2z)$ (f) $F(x, y, z) = (\sin xyz, xz)$
2. Find the Jacobian matrix of the mappings in Exercise 1 evaluated at the following points.
   (a) $(1, 2)$ (b) $(\pi, \pi/2)$ (c) $(1, 4)$
   (d) $(1, 1, -1)$ (e) $(2, -1, -1)$ (f) $(\pi, 2, 4)$
3. Let $L\colon \mathbf{R}^n \to \mathbf{R}^m$ be a linear map. Show that for each point X of $\mathbf{R}^n$ we have $L'(X) = L$.
4. Find the Jacobian matrix of the following maps.
   (a) $F(x, y) = (xy, x^2)$ (b) $F(x, y, z) = (\cos xy, \sin xy, xz)$
5. Find the Jacobian determinant of the map in Exercise 1(a). Determine all points where the Jacobian determinant is equal to 0.
6. Find the Jacobian determinant of the map in Exercise 1(b).
7. Let $F\colon \mathbf{R}^2 \to \mathbf{R}^2$ be the map defined by
$$F(r, \theta) = (r\cos\theta, r\sin\theta),$$
   in other words the polar coordinates map
$$x = r\cos\theta, \qquad y = r\sin\theta.$$
   Find the Jacobian matrix and Jacobian determinant of this mapping. Determine all points (r, θ) where the Jacobian determinant vanishes.
8. Let $F\colon \mathbf{R}^3 \to \mathbf{R}^3$ be the mapping defined by
$$F(r, \theta, \varphi) = (r\sin\varphi\cos\theta, r\sin\varphi\sin\theta, r\cos\varphi)$$

or in other words

$$x = r \sin \varphi \cos \theta, \qquad y = r \sin \varphi \sin \theta, \qquad z = r \cos \varphi.$$

Find the Jacobian matrix and Jacobian determinant of this mapping.

9. Find the Jacobian matrix and determinant of the map

$$F(r, \theta) = (e^r \cos \theta, e^r \sin \theta).$$

Show that the Jacobian determinant is never 0. Show that there exist two distinct points (r_1, θ_1) and (r_2, θ_2) such that

$$F(r_1, \theta_1) = F(r_2, \theta_2).$$

§3. *The chain rule*

In the *First Course*, we proved a chain rule for composite functions. Earlier in this book, a chain rule was given for a composite of a function and a map defined for real numbers, but having values in $\mathbf{R}^n$. In this section, we give a general formulation of the chain rule for arbitrary compositions of mappings.

Let U be an open set in $\mathbf{R}^n$, and let V be an open set in $\mathbf{R}^m$. Let $F: U \to \mathbf{R}^m$ be a mapping, and assume that all values of F are contained in V. Let $G: V \to \mathbf{R}^s$ be a mapping. Then we can form the composite mapping $G \circ F$ from U into $\mathbf{R}^s$.

Let X be a point of U. Then $F(X)$ is a point of V by assumption. Let us assume that F is differentiable at X, and that G is differentiable at $F(X)$. We know that $F'(X)$ is a linear map from $\mathbf{R}^n$ into $\mathbf{R}^m$, and $G'(F(X))$ is a linear map from $\mathbf{R}^m$ into $\mathbf{R}^s$. Thus we may compose these two linear maps to give a linear map $G'(F(X)) \circ F'(X)$ from $\mathbf{R}^n$ into $\mathbf{R}^s$.

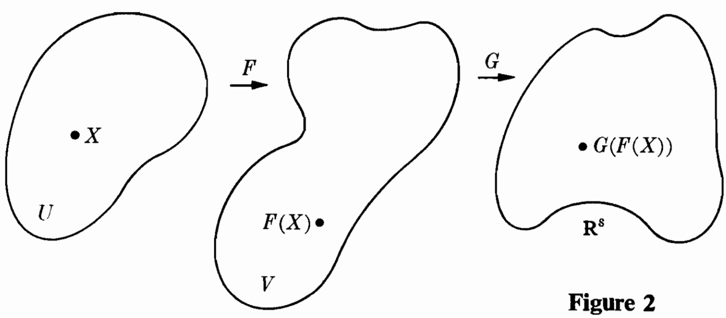


Figure 2

The next theorem tells us what the derivative of $G \circ F$ is in terms of the derivative of F at X, and the derivative of G at $F(X)$. Please observe how the statement and proof of the theorem will be entirely parallel to the statement and proof of the theorem for the chain rule in the *First Course*.

Theorem 3. *Let U be an open set in $\mathbf{R}^n$, let V be an open set in $\mathbf{R}^m$. Let $F: U \to \mathbf{R}^m$ be a mapping such that all values of F are contained in V. Let $G: V \to \mathbf{R}^s$ be a mapping. Let X be a point of U such that F is differentiable at X. Assume that G is differentiable at $F(X)$. Then the composite mapping $G \circ F$ is differentiable at X, and its derivative is given by*

$$(G \circ F)'(X) = G'\big(F(X)\big) \circ F'(X).$$

Proof. By definition of differentiability, there exists a mapping Φ_1 such that

$$\lim_{\|H\| \to 0} \Phi_1(H) = O$$

and

$$F(X + H) = F(X) + F'(X)H + \|H\|\Phi_1(H).$$

Similarly, there exists a mapping Φ_2 such that

$$\lim_{\|K\| \to 0} \Phi_2(K) = O,$$

and

$$G(Y + K) = G(Y) + G'(Y)K + \|K\|\Phi_2(K).$$

We let $K = K(H)$ be

$$K = F(X + H) - F(X) = F'(X)H + \|H\|\Phi_1(H).$$

Then

$$\begin{aligned} G\big(F(X + H)\big) &= G\big(F(X) + K\big) \\ &= G\big(F(X)\big) + G'\big(F(X)\big)K + o(K). \end{aligned}$$

Using the fact that $G'\big(F(X)\big)$ is linear, and

$$K = F(X + H) - F(X) = F'(X)H + \|H\|\Phi_1(H),$$

we can write

$$\begin{aligned} (G \circ F)(X + H) = (G \circ F)(X) &+ G'\big(F(X)\big)F'(X)H \\ &+ \|H\|G'\big(F(X)\big)\Phi_1(H) + o(K). \end{aligned}$$

Using simple estimates which we do not give in detail, we conclude that

$$(G \circ F)(X + H) = (G \circ F)(X) + G'\big(F(X)\big)F'(X)H + o(H).$$

This proves that the linear map

$$G'\big(F(X)\big)F'(X)$$

is tangent to $G \circ F$ at X. It must therefore be equal to $(G \circ F)'(X)$, as was to be shown.

§4. *Inverse mappings and implicit functions*

Let U be open in $\mathbf{R}^n$ and let $F: U \to \mathbf{R}^n$ be a map, given by coordinate functions:

$$F(X) = (f_1(X), \ldots, f_n(X)).$$

If all the partial derivatives of all functions f_i exist and are continuous, we say that F is a C^1-**map.** We say that F is C^1-**invertible** on U if the image $F(U)$ is an open set V, and if there exists a C^1-map $G: V \to U$ such that $G \circ F$ and $F \circ G$ are the respective identity mappings on U and V.

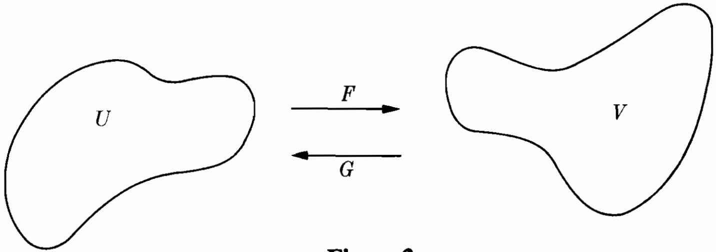


Figure 3

Example 1. Let A be a fixed vector, and let $F: \mathbf{R}^n \to \mathbf{R}^n$ be the translation by A, namely $F(X) = X + A$. Then F is C^1-invertible, its inverse being translation by $-A$.

Example 2. Let U be the subset of $\mathbf{R}^2$ consisting of all pairs (r, θ) with $r > 0$ and $0 < \theta < \pi$. Let

$$F(r, \theta) = (r \cos \theta, r \sin \theta).$$

Let $x = r \cos \theta$ and $y = r \sin \theta$. Then the image of U is the upper half-plane consisting of all (x, y) such that $y > 0$, and arbitrary x (Fig. 4).

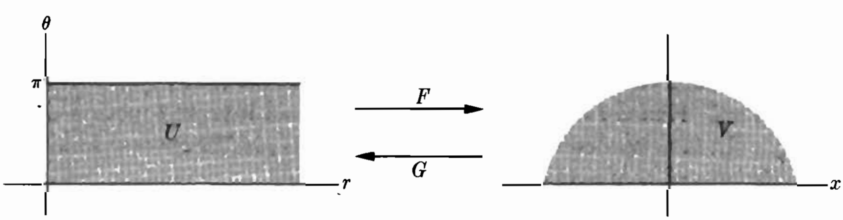


Figure 4

We can solve for the inverse map G, namely:

$$r = \sqrt{x^2 + y^2} \qquad \text{and} \qquad \theta = \arccos \frac{x}{r}$$

so that

$$G(x, y) = \left(\sqrt{x^2 + y^2}, \arccos \frac{x}{r}\right) \cdot$$

In many applications, a map is not necessarily invertible, but has still a useful property locally. Let P be a point of U. We say that F is **locally C^1-invertible at P** if there exists an open set U_1 contained in U and containing P such that F is C^1-invertible on U_1.

Example 3. If we view $F(r, \theta) = (r \cos \theta, r \sin \theta)$ as defined on all of $\mathbf{R}^2$, then F is not C^1-invertible on all of $\mathbf{R}^2$, but given any point, it is locally invertible at that point. One could see this by giving an explicit inverse map as we did in Example 2. At any rate, from Example 2, we see that F is C^1-invertible on the set $r > 0$ and $0 < \theta < \pi$.

In most cases, it is not possible to define an inverse map by explicit formulas. However, there is a very important theorem which allows us to conclude that a map is locally invertible at a point.

Inverse Mapping Theorem. *Let $F\colon U \to \mathbf{R}^n$ be a C^1-map. Let P be a point of U. If the Jacobian determinant $\Delta_F(P)$ is not equal to* 0, *then F is locally C^1-invertible at P.*

A proof of this theorem is too involved to be given in this book. However, we make the following comment. The fact that the determinant $\Delta_F(P)$ is not 0 implies (and in fact is equivalent with) the fact that the Jacobian matrix is invertible, and the Jacobian matrix represents the linear map $F'(P)$. Thus the inverse mapping theorem asserts that if the derivative $F'(P)$ is invertible, then the map F itself is locally invertible at P. Since it is usually very easy to determine whether the Jacobian determinant vanishes or not, we see that the inverse mapping theorem gives us a simple criterion for local invertibility.

Example 4. Consider the case of one variable, $y = f(x)$. In the *First Course*, we proved that if $f'(x_0) \neq 0$ at a point x_0, then there is an inverse function defined near $y_0 = f(x_0)$. Indeed, say $f'(x_0) > 0$. By continuity, assuming that f' is continuous (i.e. f is C^1), we know that $f'(x) > 0$ for x close to x_0. Hence f is strictly increasing, and an inverse function exists near x_0. In fact, we determined the derivative. If g is the inverse function, then we proved that

$$g'(y_0) = f'(x_0)^{-1}.$$

Example 5. The formula for the derivative of the inverse function in the case of one variable can be generalized to the case of the inverse mapping theorem. Suppose that the map $F\colon U \to V$ has a C^1-inverse $G\colon V \to U$. Let X be a point of U. Then $G \circ F = I$ is the identity, and

since I is linear, we see directly from the definition of the derivative that $I'(X) = I$. Using the chain rule, we find that

$$I = (G \circ F)'(X) = G'\big(F(X)\big) \circ F'(X)$$

for all X in U. In particular, this means that if $Y = F(X)$, then

$$\boxed{G'(Y) = F'(X)^{-1}}$$

where the inverse in this last expression is to be understood as the inverse of the linear map $F'(X)$. Thus we have generalized the formula for the derivative of an inverse function.

Example 6. Let $F(x, y) = (e^x \cos y, e^x \sin y)$. Show that F is locally invertible at every point.

We find that

$$J_F(x, y) = \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix}, \quad \text{whence} \quad \Delta_F(x, y) = e^x \neq 0.$$

Since the Jacobian determinant is not 0, it follows that F is locally invertible at (x, y) for all x, y.

Example 7. *Let U be open in $\mathbf{R}^2$ and let $f\colon U \to \mathbf{R}$ be a C^1-function. Let (a, b) be a point of U. Assume that $D_2 f(a, b) \neq 0$. Then the map F given by*

$$(x, y) \mapsto F(x, y) = \big(x, f(x, y)\big)$$

is locally invertible at (a, b).

Proof. All we have to do is compute the Jacobian matrix and determinant. We have

$$J_F(x, y) = \begin{pmatrix} 1 & 0 \\ \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{pmatrix}$$

so that

$$J_F(a, b) = \begin{pmatrix} 1 & 0 \\ D_1 f(a, b) & D_2 f(a, b) \end{pmatrix}$$

and hence

$$\Delta_F(a, b) = D_2 f(a, b).$$

By assumption, this is not 0, and the inverse mapping theorem implies what we want.

The result of Example 7 can be used to discuss implicit functions.

Again let $f: U \to \mathbf{R}$ be as in Example 7, and assume that $f(a, b) = c$. We ask whether there is some differentiable function $y = \varphi(x)$ defined near $x = a$ such that $\varphi(a) = b$ and

$$f(x, \varphi(x)) = c$$

for all x near a. If such a function φ exists, then we say that $y = \varphi(x)$ is the **function determined implicitly by** f.

Theorem 4 (Implicit Function Theorem). *Let U be open in $\mathbf{R}^2$ and let $f: U \to \mathbf{R}$ be a C^1-function. Let (a, b) be a point of U, and let $f(a, b) = c$. Assume that $D_2 f(a, b) \neq 0$. Then there exists an implicit function $y = \varphi(x)$ which is C^1 in some interval containing a, and such that $\varphi(a) = b$.*

Proof. We apply Example 7 and use the notation of that example. Thus we let

$$F(x, y) = (x, f(x, y)).$$

We know that $F(a, b) = (a, c)$ and that there exists a C^1-inverse G defined locally near (a, c). The inverse map G has two coordinate functions, and we can write $G(x, z) = (x, g(x, z))$ for some function g. Thus we put $y = g(x, z)$, and $z = f(x, y)$. We define

$$\varphi(x) = g(x, c).$$

Then on the one hand,

$$F(x, \varphi(x)) = F(x, g(x, c)) = F(G(x, c)) = (x, c),$$

and on the other hand,

$$F(x, \varphi(x)) = (x, f(x, \varphi(x))).$$

This proves that $f(x, \varphi(x)) = c$. Furthermore, by definition of an inverse map, $G(a, c) = (a, b)$ so that $\varphi(a) = b$. This proves the implicit function theorem.

Example 8. Let $f(x, y) = x^2 + y^2$ and let $(a, b) = (1, 1)$. Then $c = f(1, 1) = 2$. We have $D_2 f(x, y) = 2y$ so that

$$D_2 f(1, 1) = 2 \neq 0,$$

so the implicit function $y = \varphi(x)$ near $x = 1$ exists. In this case, we can of course solve explicitly for y, namely

$$y = \sqrt{2 - x^2}.$$

Example 9. We take $f(x, y) = x^2 + y^2$ as in Example 8, and $(a, b) = (-1, -1)$. Then again $c = f(-1, -1) = 2$, and

$$D_2 f(-1, -1) = -2 \neq 0.$$

In this case we can still solve for y in terms of x, namely

$$y = -\sqrt{2 - x^2}.$$

In general, the equation $f(x, y) = c$ defines some curve as in the following picture.

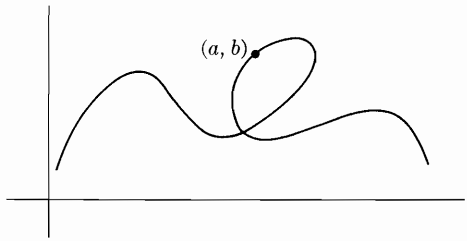


Figure 5

Near the point (a, b) as indicated in the picture, we see that there is an implicit function:

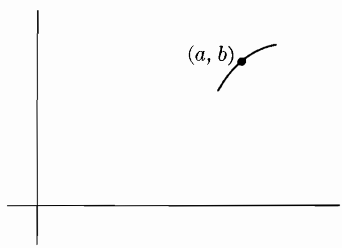


Figure 6

but that one could not define the implicit function for all x, only for those x near a.

Example 10. Let $f(x, y) = x^2y + 3y^3x^4 - 4$. Take $(a, b) = (1, 1)$ so that $f(a, b) = 0$. Then $D_2f(x, y) = x^2 + 9y^2x^4$ and

$$D_2f(1, 1) = 10 \neq 0.$$

Hence the implicit function $y = \varphi(x)$ exists, but there is no simple way to solve for it. We can also determine the derivative $\varphi'(1)$. Indeed, differentiating the equation $f(x, y) = 0$, knowing that $y = \varphi(x)$ is a differentiable function, we find

$$2xy + x^2y' + 12y^3x^3 + 9y^2y'x^4 = 0,$$

whence we can solve for $y' = \varphi'(x)$, namely

$$\varphi'(x) = y' = -\frac{2xy + 12y^3x^3}{x^2 + 9y^2x^4}.$$

Hence

$$\varphi'(1) = -\frac{2 + 12}{1 + 9} = -\frac{7}{5}.$$

In Exercise 4 we give the general formula for an arbitrary function f.

Example 11. In general, given any function $f(x, y) = 0$ and $y = \varphi(x)$ we can find $\varphi'(x)$ by differentiating in the usual way. For instance, suppose

$$x^3 + 4y \sin(xy) = 0.$$

Then taking the derivative with respect to x, we find

$$3x^2 + 4y' \sin(xy) + 4y \cos(xy)(y + xy').$$

We then solve for y' as

$$y' = -\frac{4y^2 \cos(xy) + 3x^2}{4 \sin(xy) + 4xy \cos(xy)}$$

whenever $4 \sin(xy) + 4xy \cos(xy) \neq 0$. Similarly, we can solve for y'' by differentiating either of the last two expressions. In the present case, this gets complicated.

Exercises

1. Determine whether the following mappings are locally C^1-invertible at the given point.
   (a) $F(x, y) = (x^2 - y^2, 2xy)$ at $(x, y) \neq (0, 0)$
   (b) $F(x, y) = (x^3y + 1, x^2 + y^2)$ at $(1, 2)$
   (c) $F(x, y) = (x + y, y^{1/4})$ at $(1, 16)$
   (d) $F(x, y) = \left(\dfrac{x}{x^2 + y^2}, \dfrac{y}{x^2 + y^2}\right)$ at $(x, y) \neq (0, 0)$
   (e) $F(x, y) = (x + x^2 + y, x^2 + y^2)$ at $(x, y) = (5, 8)$
2. Determine whether the following mappings are locally C^1—invertible at the indicated point.
   (a) $F(x, y) = (x + y, x^2y)$ at $(1, 2)$
   (b) $F(x, y) = (\sin x, \cos xy)$ at $(\pi, \pi/2)$
   (c) $F(x, y) = (e^{xy}, \log x)$ at $(1, 4)$
   (d) $F(x, y, z) = (xz, xy, yz)$ at $(1, 1, -1)$
3. Show that the map defined by $F(x, y) = (e^x \cos y, e^x \sin y)$ is not invertible on all of $\mathbf{R}^2$, even though it is locally invertible everywhere.
4. Let $y = \varphi(x)$ be an implicit function satisfying $f(x, \varphi(x)) = 0$, both f, φ being C^1. Show that
$$\varphi'(x) = -\frac{D_1 f(x, \varphi(x))}{D_2 f(x, \varphi(x))}$$
   wherever $D_2 f(x, \varphi(x)) \neq 0$.
5. Find an expression for $\varphi''(x)$ by differentiating the preceding expression for $\varphi'(x)$.
6. Let $f(x, y) = (x - 2)^3 y + xe^{y-1}$. Is $D_2 f(a, b) \neq 0$ at the following points (a, b)?
   (a) $(1, 1)$ (b) $(0, 0)$ (c) $(2, 1)$

7. Let f be a C^1-function of 3 variables (x, y, z) defined on an open set U of $\mathbf{R}^3$. Let (a, b, c) be a point of U, and assume $f(a, b, c) = 0$, $D_3f(a, b, c) \neq 0$. Show that there exists a C^1-function $\varphi(x, y)$ defined near (a, b) such that

$$f(x, y, \varphi(x, y)) = 0 \quad \text{and} \quad \varphi(a, b) = c.$$

We call φ the implicit function $z = \varphi(x, y)$ determined by f at (a, b).

8. In Exercise 7, show that

$$D_1\varphi(a, b) = -\frac{D_1f(a, b, c)}{D_3f(a, b, c)}.$$

9. For each of the following functions f, show that $f(x, y) = 0$ defines an implicit function $y = \varphi(x)$ at the given point (a, b), and find $\varphi'(a)$.
   (a) $f(x, y) = x^2 - xy + y^2 - 3$ at $(1, 2)$
   (b) $f(x, y) = x \cos xy$ at $(1, \pi/2)$
   (c) $f(x, y) = 2e^{x+y} - x + y$ at $(1, -1)$
   (d) $f(x, y) = xe^y - y + 1$ at $(-1, 0)$
   (e) $f(x, y) = x + y + x \sin y$ at $(0, 0)$
   (f) $f(x, y) = x^5 + y^5 + xy + 4$ at $(2, -2)$

10. For each of the following functions $f(x, y, z)$, show that $f(x, y, z) = 0$ defines an implicit function $z = \varphi(x, y)$ at the given point (a, b, c) and find $D_1\varphi(a, b)$ and $D_2\varphi(a, b)$.
   (a) $f(x, y, z) = x + y + z + \cos xyz$ at $(0, 0, -1)$
   (b) $f(x, y, z) = z^3 - z - xy \sin z$ at $(0, 0, 0)$
   (c) $f(x, y, z) = x^3 + y^3 + z^3 - 3xyz - 4$ at $(1, 1, 2)$
   (d) $f(x, y, z) = x + y + z - e^{xyz}$ at $(0, \frac{1}{2}, \frac{1}{2})$

11. Let $f(x, y, z) = x^3 - 2y^2 + z^2$. Show that $f(x, y, z) = 0$ defines an implicit function $x = \varphi(y, z)$ at the point $(1, 1, 1)$. Find $D_1\varphi$ and $D_2\varphi$ at the point $(1, 1)$.

12. In Exercise 10, show that $f(x, y, z) = 0$ also determines y as an implicit function of (x, z) and z as an implicit function of (x, y) at the given point. Find the partial derivatives of these implicit functions at the given point.

PART FOUR

MULTIPLE INTEGRATION

The theory of integration has been separated logically into two parts, for the convenience of those using the text in different circumstances.

For those who use the text only one term, or who wish to deal with multiple integration early, Chapter XII gives the basic techniques of multiple integrals, and is independent of the linear algebra or determinants. It can be read immediately after the chapter on vectors, i.e. after Chapter I.

Similarly, the first section on Green's theorem can be read after knowing about curve integrals and double integrals. It is independent of Chapter XIII and of the linear algebra. It provides a good application at a quite elementary level for both curve integrals and double integrals, by showing the relation between the two.

Chapter XV, the last, is independent of the change of variables formula. It requires essentially only multiple integration, and the algebra of vectors. The section on Jacobian matrices is used incidentally, to give more geometric motivation to tangent planes and area.

Chapter XIII on the change of variables formula is the most expandable for a class pressed for time. It uses determinants. If there is time, it can be used to give a more direct proof for the value of the volume of an elementary spherical region, computed ad hoc in Chapter XII.

CHAPTER XII

Multiple Integrals

When studying functions of one variable, it was possible to give essentially complete proofs for the existence of an integral of a continuous function over an interval. The investigation of the integral involved lower sums and upper sums.

In order to develop a theory of integration for functions of several variables, it becomes necessary to have techniques whose degree of sophistication is somewhat greater than that which is available to us. Hence we shall only state results, and omit most of the proofs, except in special cases. These results will allow us to compute multiple integrals.

Even in these special cases, the proofs should be omitted in any class which is not very hip on theory.

We shall also list various formulas giving double and triple integrals in terms of polar coordinates, and we give a geometric argument to make them plausible. Here again, the general formula for changing variables in a multiple integral can be handled theoretically (and elegantly) only when much more machinery is available than we have at present. The proofs properly belong to an advanced calculus course. [Cf. *Introduction to Analysis.*]

§1. Double integrals

We begin by discussing the analogue of upper and lower sums associated with partitions.

Let R be a region of the plane (Fig. 1), and let f be a function defined on R. We shall say that f is **bounded** if there exists a number M such that $|f(X)| \leqq M$ for all X in R.

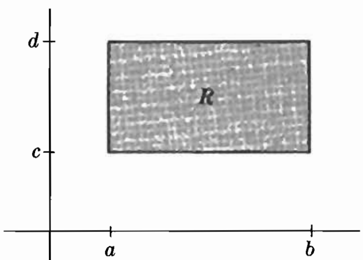


Figure 1

Let a, b be two numbers with $a \leqq b$, and let c, d be two numbers with $c \leqq d$. We consider the closed interval $[a, b]$ on the x-axis and the closed interval $[c, d]$ on the y-axis. These determine a rectangle R in the plane, consisting of all pairs of points (x, y) with $a \leqq x \leqq b$ and $c \leqq y \leqq d$.

The rectangle R above will be denoted by $[a, b] \times [c, d]$.

Let I denote the interval $[a, b]$. By a partition P_I of I we mean a sequence of numbers

$$x_1 = a \leqq x_2 \leqq \cdots \leqq x_m = b$$

which we also write as $P_I = (x_1, \ldots, x_m)$. Similarly, by a partition P_J of the interval $J = [c, d]$ we mean a sequence of numbers

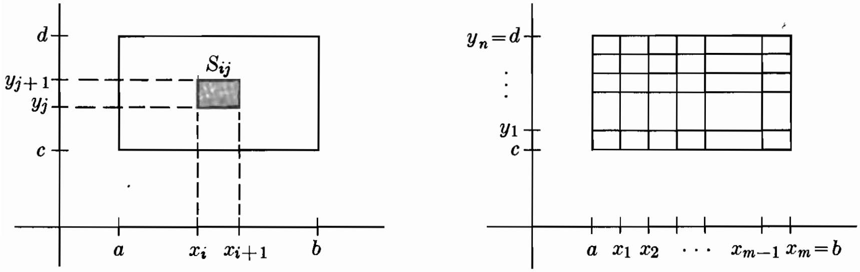
$$y_1 = c \leqq y_2 \leqq \cdots \leqq y_n = d$$

which we write as $P_J = (y_1, \ldots, y_n)$.

Each pair of small intervals $[x_i, x_{i+1}]$ and $[y_j, y_{j+1}]$ determines a rectangle

$$S_{ij} = [x_i, x_{i+1}] \times [y_j, y_{j+1}].$$

(Cf. Fig. 2(a).) We denote symbolically by $P = P_I \times P_J$ the partition of R into rectangles S_{ij} and we call such S_{ij} a **subrectangle** of the **partition** (Fig. 2(b)).



(a) **Figure 2** (b)

If R is a rectangle as above, we define its **2-dimensional volume** (that is, its area) to be the obvious thing, namely

$$\text{Area}(R) = (d - c)(b - a).$$

Thus the area of each subrectangle S_{ij} is $(y_{j+1} - y_j)(x_{i+1} - x_i)$.

Let A be a region in the plane, and let f be a function defined on A. As usual, we say that f is **continuous** at a point P of A if

$$\lim_{X \to P} f(X) = f(P).$$

We say that f is continuous on A if it is continuous at every point of A.

If S is a set and f a function on S which reaches a maximum on S, we let

$$\max_S f$$

denote this maximum value. It is a value $f(v)$ for some point v in S such that $f(v) \geqq f(w)$ for all w in S. Similarly, we let

$$\min_S f$$

be the minimum value of the function on S, if it exists. We recall a fact which we do not prove, that a continuous function on a closed and bounded set always takes on a maximum and minimum value. For instance, a continuous function on a closed interval $[a, b]$ always has a maximum. A continuous function on a rectangle as above also has a maximum, and a minimum.

We then form sums which are analogous to the lower and upper sums used to define the integral of functions of one variable. If P denotes the partition as above, and f is a continuous function on R, we define

$$L(P, f) = \sum_S (\min_S f)\,\mathrm{Area}(S),$$
$$U(P, f) = \sum_S (\max_S f)\,\mathrm{Area}(S).$$

The symbol $\sum_S$ means that we must take the sum over all subrectangles of the partition. In terms of the indices i, j, we can rewrite say the lower sum as

$$\begin{aligned} L(P, f) &= \sum_{i=1}^{m} \sum_{j=1}^{n} (\min_{S_{ij}} f)(y_{j+1} - y_j)(x_{i+1} - x_i) \\ &= \sum_i \sum_j (\min_{S_{ij}} f)\,\mathrm{Area}(S_{ij}), \end{aligned}$$

and similarly for the upper sum.

Let v_{ij} be a point in the small rectangle S_{ij} such that $f(v_{ij})$ is a maximum of f on this rectangle. Then the upper sum $U(P, f)$ can be written also in the form

$$\begin{aligned} U(P, f) &= \sum_{i=1}^{m} \sum_{j=1}^{n} f(v_{ij})(y_{j+1} - y_j)(x_{i+1} - x_i). \\ &= \sum_i \sum_j f(v_{ij})\,\mathrm{Area}(S_{ij}) \end{aligned}$$

If $f(v_{ij})$ is neither a maximum nor a minimum for f on S_{ij}, then the above sum lies between the upper and lower sum, and is called a **Riemann sum** for f.

Just as in the case of functions of one variable, we can then take refinements of partitions. If P_I' is a partition of I, we say that P_I' is a refinement of P_I if every number of P_I is among the numbers of P_I'. If P_J' is a refinement of P_J, then we call $P_I' \times P_J' = P'$ a refinement of P.

We omit the proof of the following lemma, which is entirely similar to the one variable case.

Lemma. *If P' is a refinement of P, then*

$$L(P,f) \leqq L(P',f) \leqq U(P',f) \leqq U(P,f).$$

In other words, the lower sums increase under refinements of the partition, while the upper sums decrease.

We define f to be **integrable** on R if there exists a unique number which is greater than or equal to every lower sum, and less than or equal to every upper sum. Formulated in another way, we can say that f is integrable on R if and only if the least upper bound of all lower sums is equal to the greatest lower bound of all upper sums. If this number exists, we call it the **integral** of f, and denote it by

$$\int_R f \qquad \text{or} \qquad \iint_R f(x, y)\, dy\, dx.$$

We can interpret the integral as a volume under certain conditions. Namely, suppose that $f(x, y) \geqq 0$ for all (x, y) in R. The value $f(x, y)$ may be viewed as a height above the point (x, y), and we may consider the integral of f as the volume of the 3-dimensional region lying above the rectangle R and bounded from above by the graph of f (Fig. 3).

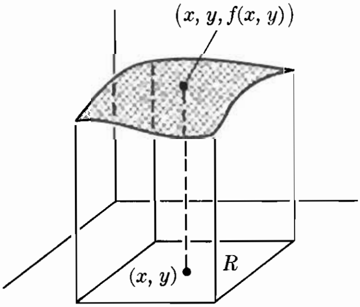


Figure 3

Each term

$$(\min_S f)\ \text{Area}(S)$$

is the volume of a rectangular box whose base is the rectangle S in the (x, y)-plane, and whose height is $\min_S f$. The volume of such a box is precisely $(\min_S f)$ Area(S), where, as we said above, Area(S) is the 2-dimensional volume of S, that is its area. This box lies below the 3-dimen-

sional region bounded from above by the graph of f. Similarly, the term

$$(\max_S f)\ \mathrm{Area}(S)$$

is the volume of a box whose base is S and whose height is $\max_S f$. This box lies above the above region. This makes our interpretation of the integral as volume clear.

Also, as in one variable, a positive function on a region may be viewed as a density, and thus if $f \geqq 0$ on R, then we also interpret

$$\iint_R f(x, y)\, dy\, dx$$

as the **mass** of R.

Theorem 1. *Assume that f, g are functions on the rectangle R, and are integrable. Then $f + g$ is integrable. If k is a number, then kf is integrable. We have:*

$$\int_R (f + g) = \int_R f + \int_R g \qquad \text{and} \qquad \int_R (kf) = k\int_R f.$$

In other words, integrable functions form a vector space, and the integral is a linear map on this vector space.

Proof. Let P be a partition of R and let S be a subrectangle of the partition. For any point v in S we have

$$\min_S f \leqq f(v) \qquad \text{and} \qquad \min_S g \leqq g(v),$$

whence

$$\min_S f + \min_S g \leqq f(v) + g(v).$$

Thus $\min_S f + \min_S g$ is a lower bound for all values $f(v) + g(v)$. Hence by definition of a greatest lower bound, we obtain the inequality

$$\min_S f + \min_S g \leqq \min_{v\in S} \big(f(v) + g(v)\big) = \min_S (f + g).$$

Consequently we get

$$\begin{aligned} L(P, f) + L(P, g) &= \sum_S (\min_S f)\ \mathrm{Area}(S) + \sum_S (\min_S g)\ \mathrm{Area}(S) \\ &= \sum_S (\min_S f + \min_S g)\ \mathrm{Area}(S) \\ &\leqq \sum_S \min_S (f + g)\ \mathrm{Area}(S) \\ &= L(P, f + g). \end{aligned}$$

By a similar argument, we find that

$$L(P,f) + L(P,g) \leqq L(P,f+g) \leqq U(P,f+g) \leqq U(P,f) + U(P,g).$$

Since $L(P,f)$ and $U(P,f)$ are arbitrarily close together for suitable partitions, and $L(P,g)$, $U(P,g)$ are arbitrarily close together for suitable partitions, we see by the usual squeezing process of limits that $L(P,f+g)$ and $U(P,f+g)$ are arbitrarily close together for suitable partitions. This proves that $f+g$ is integrable.

As for the constant k, we note that

$$\min_S (kf) = \min_{v\in S} \big(kf(v)\big) = k \cdot \min_{v\in S} f(v).$$

Hence k comes out as a factor in each term of the lower sum, and similarly for the upper sum, so that

$$kL(P,f) = L(P,kf) \leqq U(P,kf) = kU(P,f).$$

From this our second assertion follows.

Theorem 2. *If f, g are integrable on R, and $f \leqq g$, then*

$$\int_R f \leqq \int_R g.$$

Proof. We have for each subrectangle S of a partition P:

$$\min_S f \leqq f(v) \leqq g(v)$$

for all v in S. Hence $\min_S f$ is a lower bound for the values of g on S, and hence

$$\min_S f \leqq \min_S g.$$

Consequently

$$L(P,f) = \sum_S (\min_S f)\,\mathrm{Area}(S) \leqq \sum_S (\min_S g)\,\mathrm{Area}(S) = L(P,g) \leqq \int_R g.$$

Since $\int_R g$ is an upper bound for $L(P,f)$ it follows that the least upper bound of all lower sums for f is $\leqq \int_R g$, in other words

$$\int_R f \leqq \int_R g,$$

as was to be shown.

Theorem 3c. *Let R be a rectangle, and let f be a function defined and continuous on R. Then f is integrable on R.*

We shall not give the proof of Theorem 3c.

We need a somewhat more general discussion to deal with applications which arise naturally in practice. A function f is usually not given on a

rectangle but on some region A in the plane. We say that A is **bounded** if there exists a number M such that $\|X\| \leqq M$ for all points X in A. Any bounded region is contained in a rectangle, as shown on Fig. 4.

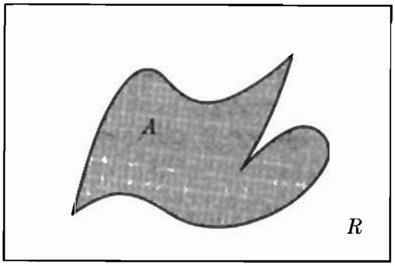


Figure 4

The set of boundary points of the region A will be called the **boundary** of A. We shall say that the boundary is **smooth** if it consists of a finite number of smooth curves. A smooth curve means a C^1 curve, i.e. a curve parametrized such that the coordinate functions have continuous derivatives, as studied in Chapter II. The boundary of A in Fig. 4 consists of three such curves. We draw a finite number of C^1 curves in the next picture.

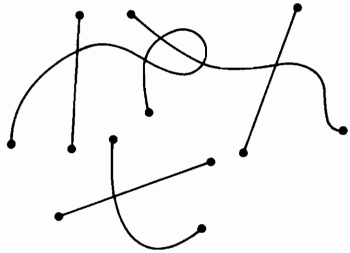


Figure 5

Suppose the function f is defined on a region A as in Fig. 4, so that A is bounded and has a boundary which is smooth. If we want to integrate f over the region A, then it is natural to extend the definition of f to the whole rectangle R, by letting

$$f(v) = 0$$

for every point v in R such that v does not lie in A. Then even if we assume that f is continuous on A, we see that f is not continuous on R. The points of discontinuity are precisely the points of the boundary of A. Therefore we cannot apply Theorem 3c directly, and we need a minor adjustment of our definitions to deal with this case, which we now discuss.

Suppose that instead of being continuous on R the function f is merely bounded, and so has a least upper bound and a greatest lower bound.

Let P be a partition of R, and let S be a subrectangle of the partition. By

$$\text{lub}_S f = \underset{v \in S}{\text{lub}}\, f(v)$$

we mean the least upper bound of all values $f(v)$ for v in S. We take it as a known property of the real numbers that any bounded set of numbers has a least upper bound, and also a greatest lower bound. Similarly, we denote by

$$\text{glb}_S f = \underset{v \in S}{\text{glb}}\, f(v)$$

the greatest lower bound of all values of f on S. We may then form upper and lower sums with the least upper bound and greatest lower bound, respectively, that is:

$$U(P, f) = \sum_S (\text{lub}_S\, f)\, \text{Area}(S)$$

and

$$L(P, f) = \sum_S (\text{glb}_S\, f)\, \text{Area}(S).$$

Then Theorems 1 and 2 hold for this more general type of function, and the proofs are the same, replacing max by lub and min by glb. We also have an extension of Theorem 3c which applies to all practical cases which we shall meet.

Theorem 3. *Let R be a rectangle and let f be a function defined on R, bounded, and continuous except possibly at the points lying on a finite number of smooth curves. Then f is integrable on R.*

Again, we shall not prove Theorem 3. However, we make some comments to indicate the main idea in the proof.

Suppose we are interested in the area of the region A, contained in the rectangle R as in Fig. 4. Let us partition the rectangle into small rectangles S_{ij} as before. Let f be the function which takes on the value 1 in A and has the value 0 at any point not in A. Let v_{ij} be a point in S_{ij}. Let us consider an approximating sum

$$\sum_{ij} f(v_{ij}) \cdot \text{Area}(S_{ij}).$$

If the rectangle S_{ij} lies entirely within the region A, then $f(v_{ij}) = 1$ and the above sum has a term contributing the area of S_{ij}. If the rectangle S_{ij} lies entirely outside the region A, then $f(v_{ij}) = 0$, and the corresponding term in the sum is equal to 0. Therefore the terms in the sum which may or may not give a positive contribution are those such that S_{ij} touches the boundary of A. Suppose that the diameter of each rectangle S_{ij} is small,

say at most ϵ. We can achieve this by taking a very fine partition of the rectangle. Let L be the length of the boundary. Then the contribution to the above sum arising from those terms meeting the boundary of A will be approximately equal to ϵL, and therefore will tend to 0 as ϵ tends to 0. This means that if we make the partition very fine, we get a good approximation to the area of A by means of the above sum. A similar argument applies when we deal with a more general function f.

Let A be a region in the plane, contained in a rectangle R (Fig. 4). Let f be a function defined on A. We denote by f_A the function which has the same values as f at points of A, and such that $f_A(Q) = 0$ if Q is a point not in A. Then f_A is defined on the rectangle R, and we define

$$\int_A f = \int_R f_A$$

provided that f_A is integrable. By Theorem 3, we note that if the boundary of A is smooth, and if f is continuous on A, then f_A is continuous except at all points lying on the boundary of A, and hence f_A is integrable.

We now have one more property of the integral which is convenient to integrate a function over several regions.

Theorem 4. *Let A be a bounded region in the plane, expressed as a union of two regions A_1 and A_2 having no points in common except possibly boundary points. Assume that the boundaries of A, A_1, A_2 are smooth. If f is a function defined on A and continuous except at a finite number of smooth curves, then*

$$\int_A f = \int_{A_1} f + \int_{A_2} f.$$

Furthermore, if A is itself some smooth curve, contained in a rectangle R, and if f is a bounded function on R which has the value 0 except possibly for points of A, then

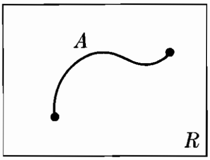
$$\int_A f = 0.$$

We shall not give the proof of Theorem 4, which anyhow is intuitively clear. In Fig. 6(a) we have drawn a smooth curve in R where f may not be 0, and such that $f(v) = 0$ if v lies in R but v is not a point of A. Then

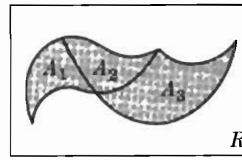
$$\int_A f = 0.$$

This is reasonable because the 2-dimensional area of a curve is 0. In Fig. 6(b) we have drawn three regions A_1, A_2, A_3 which have only smooth

curves in common. The integral of a function f over the three regions is then the sum of the integrals of f over each region separately.



(a)



(b)

Figure 6

Exercises

1. Let f be a continuous bounded function on a rectangle R. Let M be a number such that $|f(v)| \leqq M$ for all v in R. Show that

$$\left|\int_R f\right| \leqq M\,\text{Area}(R).$$

This estimate is also true for an integrable function f. Does your proof apply to this more general case? It should.

§2. *Repeated integrals*

To compute the integral we shall investigate double integrals.

Let f be a function defined on our rectangle. For each x in the interval $[a, b]$ we have a function f_x of y given by $f_x(y) = f(x, y)$, and this function f_x is defined on the interval $[c, d]$. Assume that for each x the function f_x is integrable over this interval (in the old sense of the word, for functions of one variable). We may then form the integral

$$\int_c^d f_x(y)\,dy = \int_c^d f(x, y)\,dy.$$

The expression we obtain depends on the particular value of x chosen in the interval $[a, b]$, and is thus a function of x. Assume that this function is integrable over the interval $[a, b]$. We can then take the integral

$$\int_a^b \left[\int_c^d f(x, y)\,dy\right] dx, \qquad \text{also written} \qquad \int_a^b \int_c^d f(x, y)\,dy\,dx,$$

which is called the **repeated integral** of f.

Example 1. Let $f(x, y) = x^2y$. Find the repeated integral of f over the rectangle determined by the intervals $[1, 2]$ on the x-axis and $[-3, 4]$ on the y-axis.

We must find the repeated integral

$$\int_1^2 \int_{-3}^4 f(x, y)\, dy\, dx.$$

To do this, we first compute the integral with respect to y, namely

$$\int_{-3}^4 x^2 y\, dy.$$

For a fixed value of x, we can take x^2 out of the integral, and hence this inner integral is equal to

$$x^2 \int_{-3}^4 y\, dy = x^2 \frac{y^2}{2}\bigg|_{-3}^4$$

$$= \frac{7x^2}{2}.$$

We then integrate with respect to x, namely

$$\int_1^2 \frac{7x^2}{2}\, dx = \frac{49}{6}.$$

Thus the integral of f over the rectangle is equal to $\frac{49}{6}$.

The repeated integral is useful in computing a double integral because of the following theorem, which will be proved after discussing some examples.

Theorem 5. *Let R be a rectangle $[a, b] \times [c, d]$, and let f be integrable on R. Assume that for each x in $[a, b]$ the function f_x given by*

$$f_x(y) = f(x, y)$$

is integrable on $[c, d]$. Then the function

$$x \mapsto \int_c^d f(x, y)\, dy$$

is integrable on $[a, b]$, and

$$\int_R f = \int_a^b \left[\int_c^d f(x, y)\, dy\right] dx.$$

Geometrically speaking, the inner integral for a fixed value of x gives the area of a cross section as indicated in the following figure. Then integrating such areas yields the volume of the 3-dimensional figure bounded below by the rectangle R, and above by the graph of f.

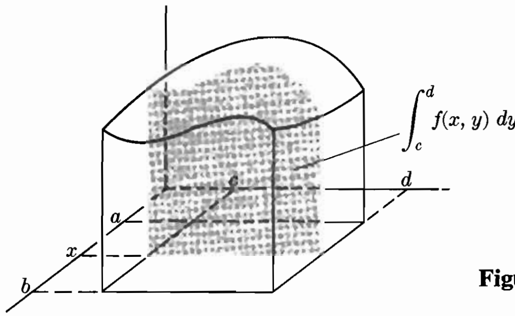


Figure 7

The following situation will arise frequently in practice.

Let g_1, g_2 be two smooth functions on a closed interval $[a, b]$ ($a \leqq b$) such that $g_1(x) \leqq g_2(x)$ for all x in that interval. Let c, d be numbers such that

$$c < g_1(x) \leqq g_2(x) < d$$

for all x in the interval $[a, b]$. Then g_1, g_2 determine a region A lying between $x = a$, $x = b$, and the two curves $y = g_1(x)$ and $y = g_2(x)$. (Cf. Fig. 8.)

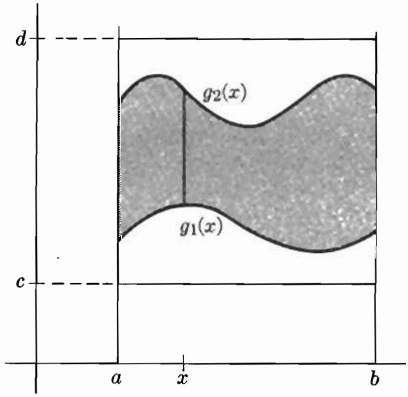


Figure 8

Let f be a function which is continuous on the region A, and define f on the rectangle $[a, b] \times [c, d]$ to be equal to 0 at any point of the rectangle not lying in the region A. For any value x in the interval $[a, b]$ the integral

$$\int_c^d f(x, y)\, dy$$

can be written as a sum:

$$\int_c^{g_1(x)} f(x, y)\, dy + \int_{g_1(x)}^{g_2(x)} f(x, y)\, dy + \int_{g_2(x)}^d f(x, y)\, dy.$$

Since $f(x, y) = 0$ whenever $c \leqq y < g_1(x)$ and $g_2(x) < y \leqq d$, it follows that the two extreme integrals are equal to 0. Thus the repeated integral of f over the rectangle is in fact equal to the repeated integral

$$\boxed{\int_a^b \left[\int_{g_1(x)}^{g_2(x)} f(x, y)\, dy\right] dx.}$$

Regions of the type described by two functions g_1, g_2 as above are the most common type of regions with which we deal.

From Theorem 5 and the preceding discussion, we obtain:

Corollary. *Let g_1, g_2 be two smooth functions defined on a closed interval $[a, b]$ $(a \leqq b)$ such that $g_1(x) \leqq g_2(x)$ for all x in that interval. Let f be a continuous function on the region A lying between $x = a$, $x = b$, and the two curves $y = g_1(x)$ and $y = g_2(x)$. Then*

$$\int_A f = \int_a^b \left[\int_{g_1(x)}^{g_2(x)} f(x, y)\, dy\right] dx;$$

in other words, the double integral is equal to the repeated integral.

We shall give the proof of Theorem 5 below. Before doing that, we first give examples showing how to apply Theorem 5, or rather its corollary.

Example 2. Let $f(x, y) = x^2 + y^2$. Find the integral of f over the region A bounded by the straight line $y = x$ and the parabola $y = x^2$ (Fig. 9).

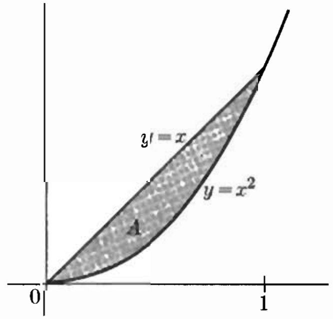


Figure 9

In this case, we have $g_1(x) = x^2$ and $g_2(x) = x$. Thus our integral is

equal to

$$\int_0^1 \left[\int_{x^2}^x (x^2 + y^2)\, dy \right] dx.$$

Now the inner integral is given by

$$\int_{x^2}^x (x^2 + y^2)\, dy = x^2 y + \frac{y^3}{3}\bigg|_{x^2}^{x}$$
$$= x^3 + \frac{x^3}{3} - x^4 - \frac{x^6}{3}.$$

Hence the repeated integral is equal to

$$\int_A f = \int_0^1 \left(x^3 + \frac{x^3}{3} - x^4 - \frac{x^6}{3} \right) dx = \frac{x^4}{4} + \frac{x^4}{12} - \frac{x^5}{5} - \frac{x^7}{21}\bigg|_0^1$$
$$= \frac{1}{4} + \frac{1}{12} - \frac{1}{5} - \frac{1}{21}.$$

(We don't need to simplify the number on the right.)

Given a region A, it is frequently possible to break it up into smaller regions having only boundary points in common, and such that each smaller region is of the type we have just described. In that case, to compute the integral of a function over A, we can apply Theorem 4.

Example 3. Let $f(x, y) = 2xy$. Find the integral of f over the triangle bounded by the lines $y = 0$, $y = x$, and the line $x + y = 2$.

The region is as shown in Fig. 10.

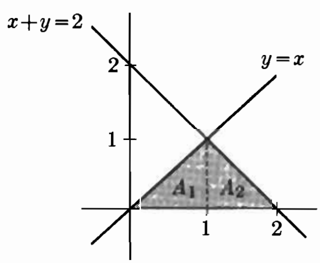


Figure 10

We break up our region into the portion from 0 to 1 and the portion from 1 to 2. These correspond to the small triangles A_1, A_2, as indicated in the picture. Then

$$\int_{A_1} f = \int_0^1 \left[\int_0^x 2xy\, dy \right] dx \qquad \text{and} \qquad \int_{A_2} f = \int_1^2 \left[\int_0^{2-x} 2xy\, dy \right] dx.$$

There is no difficulty in evaluating these integrals, and we leave them to you.

Finally, we define the **area** of a region A to be integral of the function 1 over A, i.e.

$$\text{Area}(A) = \iint_A 1\, dy\, dx.$$

Example 4. Find the area of the region bounded by the straight line $y = x$ and the curve $y = x^2$.

The region has been sketched in Example 2. By definition,

$$\text{Area}(A) = \int_0^1 \int_{x^2}^{x} dy\, dx = \int_0^1 (x - x^2)\, dx$$

$$= \frac{x^2}{2} - \frac{x^3}{3}\bigg|_0^1 = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}.$$

We also observe that the same arguments as before apply if we interchange the role of x and y. Thus for the rectangle R we also have

$$\int_R f(x, y)\, dy\, dx = \int_R f(x, y)\, dx\, dy = \int_c^d \left[\int_a^b f(x, y)\, dx\right] dy.$$

The same goes for a region described by functions

$$x = g_1(y)$$

and

$$x = g_2(y)$$

with $g_1 \leqq g_2$ between $y = c$ and $y = d$.

If A is a region in the plane bounded by a finite number of smooth curves, and f is a function on A such that $f(x) \geqq 0$ for $x \in A$, then we can interpret f as a density function, and we also call the integral $\int_A f$ the **mass** of A.

Example 5. Find the integral of the function $f(x, y) = x^2y^2$ over the region bounded by the lines $y = 1$, $y = 2$ and $x = y$ (Fig. 11).

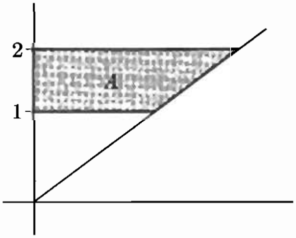


Figure 11

We have to compute the integral as prescribed, namely:

$$\int_1^2 \left[\int_0^y x^2y^2\,dx\right] dy = \int_1^2 y^2 \frac{x^3}{3}\bigg|_0^y dy = \int_1^2 \frac{y^5}{3}\,dy = \frac{7}{2}.$$

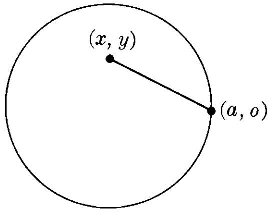
We can also say that the preceding integral, namely $7/2$, is the mass of A corresponding to the density given by the function f. Of course the units of mass are those determined by the units of density.

Example. Find the mass of a disc of radius a if the density is proportional to the square of the distance from a point on the circumference.

We take the circle surrounding the disc to have equation

$$x^2 + y^2 = a^2,$$

and select the point on the circumference to be $(a, 0)$, as shown on the figure.



Then the density function is

$$f(x, y) = k[(x - a)^2 + y^2].$$

The mass is therefore given by the integral

$$2\int_{-a}^{a}\int_0^{\sqrt{a^2-x^2}} k[(x - a)^2 + y^2]\,dy\,dx = \tfrac{3}{2}\,k\pi a^4.$$

We now give the proof of Theorem 5. We let R be the product of the intervals I, J so $R = I \times J$. We consider a partition $P = P_I \times P_J$ of R, where P_I, P_J are partitions of the intervals I, J respectively. Each subrectangle of P can then be written as

$$S = S_I \times S_J,$$

where S_I is a subinterval of I and S_J a subinterval of J. Then

$$\text{Area}(S) = l(S_J)l(S_I),$$

where l denotes the length of an interval.

We denote the function

$$x \mapsto \int_c^d f(x, y)\, dy$$

by $\int_J f$, so that the value of this function at x is

$$\int_J f_x = \int_c^d f(x, y)\, dy.$$

We have:

$$\begin{aligned} L(P_I \times P_J, f) &= \sum_S (\mathrm{glb}_S f)\, \mathrm{Area}(S) \\ &= \sum_{S_I} \sum_{S_J} (\mathrm{glb}_{S_I \times S_J} f)\, \mathrm{Area}(S_I \times S_J) \\ (*) \qquad &= \sum_{S_I} \Big[\sum_{S_J} \Big(\operatorname*{glb}_{(x,y) \in S_I \times S_J} f(x, y) \Big) l(S_J) \Big] l(S_I). \end{aligned}$$

For any x in I we have

$$\begin{aligned} \sum_{S_J} \Big(\operatorname*{glb}_{(x,y) \in S_I \times S_J} f(x, y) \Big) l(S_J) &\leqq \sum_{S_J} \operatorname*{glb}_{y \in S_J} f(x, y) l(S_J) \\ &= L(P_J, f_x) \\ &\leqq \int_J f_x, \end{aligned}$$

because each term in the expression on the left involves a glb over all (x, y) rather than only over y, and thus contributes less to the sum. Thus the expression on the left is a lower bound for the expression on the right. From this we conclude that the expression on the left is $\leqq \operatorname*{glb}_{x \in S_I} \int_J f_x$, and hence by (*)

$$\begin{aligned} L(P, f) = L(P_I \times P_J, f) &\leqq \sum_{S_I} \Big(\operatorname*{glb}_{x \in S_I} \int_J f_x \Big)\, l(S_I) \\ &= L\left(P_I, \int_J f\right) \\ &\leqq U\left(P_I, \int_J f\right). \end{aligned}$$

By similar arguments applied to upper sums instead of lower sums, we conclude that

$$L(P, f) \leqq L\left(P_I, \int_J f\right) \leqq U\left(P_I, \int_J f\right) \leqq U(P_I \times P_J, f) = U(P, f).$$

Since f is assumed to be integrable, it follows that for suitable partitions, $L(P, f)$ and $U(P, f)$ are arbitrarily close together. Thus the lower sums for the function $\int_J f$ and the upper sums for this function are arbitrarily

close for suitable partitions P. This implies that the function $\int_J f$ is integrable, and the fact that these lower sums and upper sums are squeezed between $L(P,f)$ and $U(P,f)$ shows that the double integral is equal to the repeated integral, as desired.

Exercises

1. Find the value of the following repeated integrals.

(a) $\int_0^2 \int_1^3 (x + y)\,dx\,dy$ (b) $\int_0^2 \int_1^{x^2} y\,dy\,dx$ (c) $\int_0^1 \int_{y^2}^{y} \sqrt{x}\,dx\,dy$

(d) $\int_0^{\pi} \int_0^{x} x \sin y\,dy\,dx$ (e) $\int_1^2 \int_y^{y^2} dx\,dy$ (f) $\int_0^{\pi} \int_0^{\sin x} y\,dy\,dx$

(g) $\int_0^{\pi/2} \int_0^2 r^2 \cos \theta\,dr\,d\theta$ (h) $\int_0^{2\pi} \int_0^{1-\cos\theta} r^3 \cos^2 \theta\,dr\,d\theta$

(i) $\int_0^{\arctan 3/2} \int_0^{2 \sec \theta} r\,dr\,d\theta$

2. Sketch the regions described by the following inequalities.

(a) $|x| \leqq 1,\ -1 \leqq y \leqq 2$ (b) $|x| \leqq 3,\ |y| \leqq 4$

(c) $x + y \leqq 1,\ x \geqq 0,\ y \geqq 0$ (d) $0 \leqq y \leqq |x|,\ 0 \leqq x \leqq 5$

(e) $0 \leqq x \leqq y,\ 0 \leqq y \leqq 5$ (f) $|x| + |y| \leqq 1$

3. Find the integral of the following functions.

(a) $x \cos(x + y)$ over the triangle whose vertices are $(0, 0)$, $(\pi, 0)$, and (π, π).

(b) e^{x+y} over the region defined by $|x| + |y| \leqq 1$.

(c) $x^2 - y^2$ over the region bounded by the curve $y = \sin x$ between 0 and π.

(d) $x^2 + y$ over the triangle whose vertices are $(-\frac{1}{2}, \frac{1}{2})$, $(1, 2)$, $(1, -1)$.

4. Find the integrals of the following functions over the indicated region.

(a) $f(x, y) = x$ over the region bounded by $y = x^2$ and $y = x^3$.

(b) $f(x, y) = y$ over the same region as in (a).

(c) $f(x, y) = x^2$ over the region bounded by $y = x$, $y = 2x$, and $x = 2$.

5. Let a be a number > 0. Show that the area of the region consisting of all points (x, y) such that $|x| + |y| \leqq a$, is $(2a)^2/2!$.

6. Find the following integrals and sketch the region of integration in each case.

(a) $\int_1^2 \int_{x^2}^{x^3} x\,dy\,dx$ (b) $\int_1^{-1} \int_x^{2x} e^{x+y}\,dy\,dx$

(c) $\int_0^2 \int_1^3 |x - 2| \sin y\,dx\,dy$ (d) $\int_0^{\pi/2} \int_{-y}^{y} \sin x\,dx\,dy$

(e) $\int_{-1}^{1}\int_{0}^{|x|} dy\,dx$ (f) $\int_{0}^{\pi/2}\int_{0}^{\cos y} x \sin y\,dx\,dy$

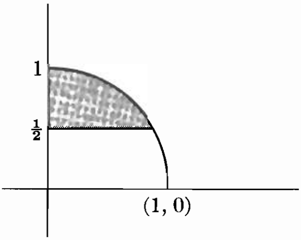
7. Sketch the region defined by $x \geqq 0$, $x^2 + y^2 \leqq 2$, and $x^2 + y^2 \geqq 1$. Determine the integral of $f(x, y) = x^2$ over this region. If you wait till you study polar coordinates in the next section, you will do this exercise more easily.

8. Integrate the function f over the indicated region.
   (a) $f(x, y) = 1/(x + y)$ over the region bounded by the lines $y = x$, $x = 1$, $x = 2$, $y = 0$.
   (b) $f(x, y) = x^2 - y^2$ over the region defined by the inequalities

$$0 \leqq x \leqq 1 \qquad \text{and} \qquad x^2 - y^2 \geqq 0.$$

   (c) $f(x, y) = x \sin xy$ over the rectangle $0 \leqq x \leqq \pi$ and $0 \leqq y \leqq 1$.
   (d) $f(x, y) = x^2 - y^2$ over the triangle whose vertices are $(-1, 1)$, $(0, 0)$, $(1, 1)$.
   (e) $f(x, y) = 1/(x + y + 1)$ over the square $0 \leqq x \leqq 1$, $0 \leqq y \leqq 1$.

9. Compute the integral of the function $f(x, y) = xy$ over the region defined by the inequalities $0 \leqq x^2 + y^2 \leqq 1$, $0 \leqq x$, $\frac{1}{2} \leqq y$, sketched below.



§3. *Polar coordinates*

It is frequently more convenient to describe a region by means of polar coordinates than with the "rectangular" coordinates of the preceding section. Such a region can then be described as the image of a simpler region as follows.

Let a, b be numbers with $0 \leqq a \leqq b \leqq 2\pi$. Let c, d be two numbers with $0 \leqq c \leqq d$. Then the inequalities

$$a \leqq \theta \leqq b \qquad \text{and} \qquad c \leqq r \leqq d$$

describe a rectangle in the (r, θ)-plane. Under the map

$$G\colon \mathbf{R}^2 \to \mathbf{R}^2$$

given by $G(r, \theta) = (r\cos\theta, r\sin\theta)$, that is

$$x = r\cos\theta, \qquad y = r\sin\theta,$$

this rectangle goes into a circular region as shown in Fig. 12.

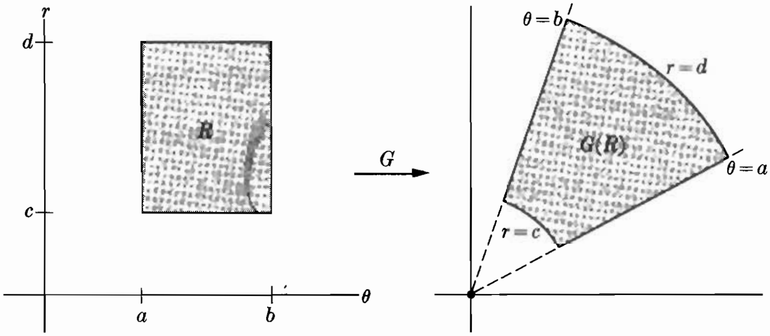


Figure 12

The preceding map G is called the **polar coordinate** map.
Consider partitions

$$a = \theta_1 \leqq \theta_2 \leqq \cdots \leqq \theta_n = b, \qquad c = r_1 \leqq r_2 \leqq \cdots \leqq r_m = d$$

of the two intervals $[a, b]$ and $[c, d]$. Each pair of intervals $[\theta_i, \theta_{i+1}]$ and $[r_j, r_{j+1}]$ determines a small region as shown in the following figure.

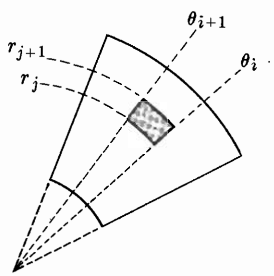


Figure 13

The area of such a region is equal to the difference between the area of the sector having angle $\theta_{i+1} - \theta_i$ and radius r_{j+1}, and the area of the sector having the same angle but radius r_j. The area of a sector having angle θ and radius r is equal to

$$\frac{\theta}{2\pi}\pi r^2 = \frac{\theta r^2}{2}.$$

Consequently the difference mentioned above is equal to

$$\frac{(\theta_{i+1} - \theta_i)r_{j+1}^2}{2} - \frac{(\theta_{i+1} - \theta_i)r_j^2}{2} = (\theta_{i+1} - \theta_i)\frac{(r_{j+1} + r_j)}{2}(r_{j+1} - r_j).$$

We note that

$$\frac{r_{j+1} + r_j}{2} = \bar{r}_j$$

and the area of the small region is therefore equal to

$$\bar{r}_j(r_{j+1} - r_j)(\theta_{i+1} - \theta_i).$$

If f is a function on the (x, y)-plane, it determines a function of (r, θ) by the formula

$$f^*(r, \theta) = f(r\cos\theta, r\sin\theta).$$

Then

$$\sum_{j=1}^{m}\sum_{i=1}^{n} f^*(r_j, \theta_i)r_j(r_{j+1} - r_j)(\theta_{i+1} - \theta_i)$$

is a Riemann sum on the product $[a, b] \times [c, d]$. Consequently the following theorem is now very plausible.

Theorem 6. *Let $R = [a, b] \times [c, d]$ be as above, and let G be the polar coordinate map. Let f be bounded and continuous on $G(R)$, except possibly at a finite number of smooth curves. Let f^* be the corresponding function of (r, θ). Then*

$$\iint_R f^*(r, \theta)r\,dr\,d\theta = \iint_{G(R)} f(x, y)\,dy\,dx.$$

In the next chapter, we shall state another theorem which gives another justification for this change of variables formula. We do not prove any of these statements in this course, since the rigorous proofs depend on more developed techniques.

As with rectangular coordinates, we can deal with more general regions. Let g_1, g_2 be two smooth functions defined on the interval $[a, b]$ and assume

$$0 \leqq g_1(\theta) \leqq g_2(\theta)$$

for all θ in that interval. Let A be the region consisting of all points (θ, r) such that $a \leqq \theta \leqq b$ and $g_1(\theta) \leqq r \leqq g_2(\theta)$. We can select two num-

bers $c, d \geqq 0$ such that

$$c \leqq g_1(\theta) \leqq g_2(\theta) \leqq d$$

for all θ in the interval $[c, d]$. Let f be continuous on $G(A)$ and extend f to the circular sector of radius d between $\theta = a$ and $\theta = b$ by giving it the value 0 outside $G(A)$. *Then the integral of Theorem* 6 *taken over this sector is equal to the repeated integral*

$$\int_a^b \int_{g_1(\theta)}^{g_2(\theta)} f^*(\theta, r) r \, dr \, d\theta.$$

The following picture shows a typical region $G(A)$ under consideration. The important thing to remember about the formula of Theorem 6 is the appearance of an extra r inside the integral.

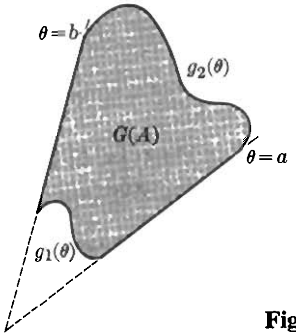


Figure 14

We also remark that a region could be described by taking θ as a function of r, and letting r vary between two constant values. In view of Theorem 5, we can evaluate the double integral of Theorem 6 by repeated integration first with respect to θ and then with respect to r.

In dealing with polar coordinates, it is useful to remember the equation of a circle. Let $a > 0$. Then

$$r = a \cos \theta, \qquad -\pi/2 \leqq \theta \leqq \pi/2,$$

is the equation of a circle of radius $a/2$ and center $(a/2, 0)$. Similarly,

$$r = a \sin \theta, \qquad 0 \leqq \theta \leqq \pi,$$

is the equation of a circle of radius $a/2$ and center $(0, a/2)$. You can

easily show this, as an exercise, using the relations

$$r = \sqrt{x^2 + y^2}, \qquad x = r\cos\theta, \qquad y = r\sin\theta.$$

(*Note.* The coordinates of the center above are given in rectangular coordinates.)

Example. Find the integral of the function $f(x, y) = x^2$ over the region enclosed by the curve given in polar coordinates by the equation

$$r = (1 - \cos\theta).$$

The function of the polar coordinates (r, θ) corresponding to f is given by

$$f^*(r, \theta) = r^2\cos^2\theta.$$

The region in the polar coordinate space is described by the inequalities

$$0 \leqq r \leqq 1 - \cos\theta \qquad \text{and} \qquad 0 \leqq \theta \leqq 2\pi.$$

This region in the (x, y)-plane looks like this:

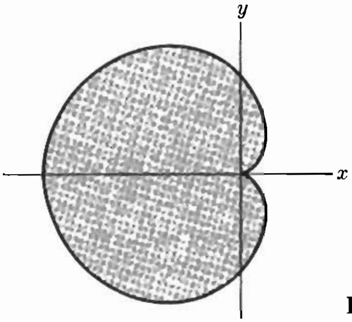


Figure 15

The desired integral is therefore the integral

$$\int_0^{2\pi}\int_0^{1-\cos\theta} r^3\cos^2\theta\,dr\,d\theta.$$

We integrate first with respect to r, which is easy, and see that our integral is equal to

$$\int_0^{2\pi} \tfrac{1}{4}(1 - \cos\theta)^4\cos^2\theta\,d\theta.$$

The evaluation of this integral is done by techniques of the first course in

calculus. We expand out the expression of the fourth power, and get a sum of terms involving $\cos^k \theta$ for $k = 0, \ldots, 6$. The reader should know how to integrate powers of the cosine, using repeatedly the formula

$$\cos^2 \theta = \frac{1 + \cos 2\theta}{2},$$

or using the recursion formula in terms of lower powers. No matter what method the reader uses, he will find the final answer to be

$$\frac{49\pi}{32}.$$

Exercises

1. By changing to polar coordinates, find the integral of $e^{x^2+y^2}$ over the region consisting of the points (x, y) such that $x^2 + y^2 \leqq 1$.
2. Find the volume of the region lying over the disc $x^2 + (y - 1)^2 \leqq 1$ and bounded from above by the function $z = x^2 + y^2$.
3. Find the integral of $e^{-(x^2+y^2)}$ over the circular disc bounded by
$$x^2 + y^2 = a^2, \qquad a > 0.$$
4. What is
$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)}\, dx\, dy?$$
5. Find the mass of a square plate of side a if the density is proportional to the square of the distance from a vertex.
6. Find the mass of a circular disk of radius a if the density is proportional to the square of the distance from a point on the circumference.
7. Find the mass of a plate bounded by one arch of the curve $y = \sin x$, and the x-axis, if the density is proportional to the distance from the x-axis.

Evaluate the following integrals. Take $a > 0$.

8. $\displaystyle\int_{-a}^{a} \int_{-\sqrt{a^2-x^2}}^{\sqrt{a^2-x^2}} dy\, dx$
9. $\displaystyle\int_0^a \int_0^{\sqrt{a^2-y^2}} (x^2 + y^2)\, dx\, dy$
10. $\displaystyle\int_0^{a/\sqrt{2}} \int_y^{\sqrt{a^2-y^2}} x\, dx\, dy$
11. Find the area inside the curve $r = a(1 + \cos \theta)$ and outside the circle $r = a$.

12. The base of a solid is the region of Exercise 11 and the top is given by the function $f(x, y) = x$. Find the volume.

13. Find the area enclosed by the curve $r^2 = 2a^2 \cos 2\theta$.

14. The base of a solid is the area of Exercise 13, and the top is bounded by the function (in terms of polar coordinates) $f(r, \theta) = \sqrt{2a^2 - r^2}$. Find the volume.

15. Find the integral of the function

$$f(x, y) = \frac{1}{(x^2 + y^2 + 1)^{3/2}}$$

over the disc of radius a centered at the origin. Letting a tend to infinity, show that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dy\, dx = 2\pi.$$

16. Answer the same question for the function

$$f(x, y) = \frac{1}{(x^2 + y^2 + 2)^2}.$$

17. Find the integral of the function

$$f(x, y) = \frac{1}{(x^2 + y^2)^3}$$

over the region between the two circles of radius 2 and radius 3, centered at the origin.

18. (a) Find the integral of the function $f(x, y) = x$ over the region bounded in polar coordinates by $r = 1 - \cos \theta$.
(b) Let a be a number > 0. Find the integral of the function $f(x, y) = x^2$ over the region bounded in polar coordinates by $r = a(1 - \cos \theta)$.

19. Sketch the region defined by $x \geqq 0$, $x^2 + y^2 \leqq 2$ and $x^2 + y^2 \geqq 1$. Determine the integral of the following functions over this region.
(a) $f(x, y) = x^2$ (b) $f(x, y) = x$ (c) $f(x, y) = y$.

20. Let n be a positive integer, and let $f(x, y) = 1/r^n$, where $r = \sqrt{x^2 + y^2}$.
(a) Find the integral of this function over the region contained between two circles of radii a and b respectively, with $0 < a < b$.
(b) For which values of n does this integral approach a limit as $a \to 0$?

§4. *Triple integrals*

The entire discussion concerning 2-dimensional integrals generalizes to higher dimensions. We discuss briefly the 3-dimensional case.

A 3-dimensional rectangle (rectangular parallelepiped) can be written

as a product of three intervals:

$$R = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3].$$

It looks like this.

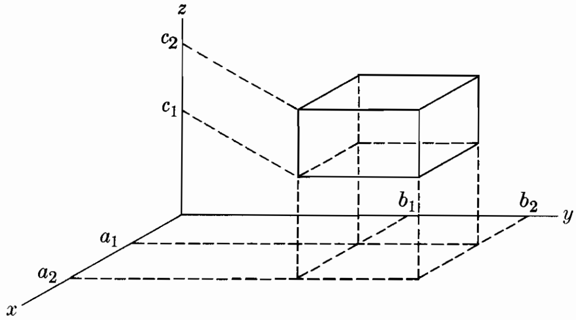


Figure 16

A partition P of R is then determined by partitions P_1, P_2, P_3 of the three intervals respectively, and partitions R into 3-dimensional subrectangles, which we denote again by S.

If f is a bounded function on R, we may then form upper and lower sums. Indeed, we define the volume of the rectangle R above to be the 3-dimensional volume

$$\mathrm{Vol}(R) = (b_3 - a_3)(b_2 - a_2)(b_1 - a_1)$$

and similarly for the subrectangles of the partition. Then we have

$$L(P, f) = \sum_S (\mathrm{glb}_S f)\,\mathrm{Vol}(S),$$

$$U(P, f) = \sum_S (\mathrm{lub}_S f)\,\mathrm{Vol}(S).$$

A refinement P' of P is determined by refinements P_1', P_2', P_3' of P_1, P_2, P_3 respectively, and the Lemma of §1 extends to this case.

Again, we say that f is integrable if the least upper bound of the lower sums is equal to the greatest lower bound of the upper sums, and if this

is the case, we define it to be the integral of f over R, written

$$\int_R f = \iiint\limits_R f(x, y, z)\, dz\, dy\, dx$$

if the variables are x, y, z.

If $f \geqq 0$, then we interpret this integral as the 4-dimensional volume of the 4-dimensional region lying in 4-space, bounded from below by R, and from above by the graph of f. Of course, we cannot draw this figure because it is in 4-space, but the terminology goes right over.

Theorem 1 and Theorem 2 are again valid, that is the integral is linear, and satisfies the usual inequality.

The criterion of Theorem 3 for a function to be integrable also has an analogue. In this case, however, we have to parametrize the boundary of a 3-dimensional region by 2-dimensional smooth pieces of surfaces. Thus let T be a 2-dimensional rectangle, and let

$$F\colon T \to \mathbf{R}^3$$

be a map. If F is of class C^1 we shall say that F is smooth, and we call the image of F a smooth surface (Fig. 17).

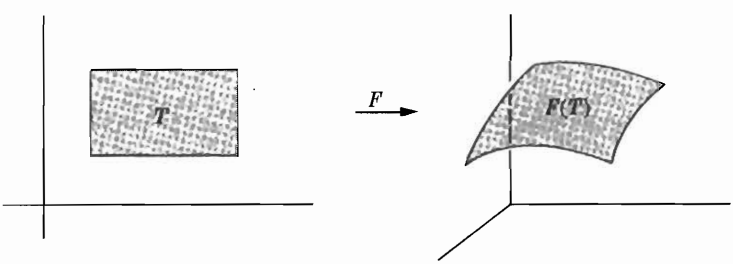


Figure 17

The analogue of Theorem 3 is then:

Let R be a 3-dimensional rectangle, and let f be a function defined on R, bounded and continuous except possibly at the points lying on a finite number of smooth surfaces. Then f is integrable on R.

Again we can integrate over a more general region than a rectangle, provided such a region A has a boundary which is contained in a finite number of smooth surfaces. Then Theorem 4 holds. If A denotes a 3-dimensional region and f is a function on A, we denote the integral of

f over A by

$$\int_A f \qquad \text{or} \qquad \iiint_A f(x, y, z)\, dz\, dy\, dx.$$

If we view A as a solid piece of material, and f is interpreted as a density distribution over A, then the integral of f over A may be interpreted as the **mass** of A.

To compute multiple integrals in the 3-dimensional case, we have the same situation as in the 2-dimensional case.

The theorem concerning the relation with repeated integrals holds, so that if R is the rectangle given by

$$R = [a_1, b_1] \times [a, b_{22}] \times [a_3, b_3],$$

then

$$\int_R f = \int_{a_1}^{b_1}\left[\int_{a_2}^{b_2}\left(\int_{a_3}^{b_3} f(x, y, z)\, dz\right) dy\right]dx.$$

Of course, the repeated integral can be evaluated in any order.

Example 1. Find the integral of the function $f(x, y, z) = \sin x$ over the rectangle $0 \leqq x \leqq \pi$, $2 \leqq y \leqq 3$, and $-1 \leqq z \leqq 1$.

The integral is equal to

$$\int_0^\pi \int_2^3 \int_{-1}^1 \sin x\, dz\, dy\, dx.$$

If we first integrate with respect to z, we get $z|_{-1}^{1} = 2$. Next with respect to y, we get $y|_2^3 = 1$. We are then reduced to the integral

$$\int_0^\pi 2 \sin x\, dx = -2\cos x\Big|_0^\pi = -2(\cos \pi - \cos 0) = 4.$$

We also have the integral over regions determined by inequalities.

Case 1. Rectangular coordinates. *Let a, b be numbers, $a \leqq b$. Let g_1, g_2 be two smooth functions defined on the interval $[a, b]$ such that*

$$g_1(x) \leqq g_2(x),$$

and let $h_1(x, y) \leqq h_2(x, y)$ be two smooth functions defined on the region consisting of all points (x, y) such that

$$a \leqq x \leqq b \qquad \text{and} \qquad g_1(x) \leqq y \leqq g_2(x).$$

Let A be the set of points (x, y, z) such that

$$a \leqq x \leqq b, \quad g_1(x) \leqq y \leqq g_2(x), \quad \text{and} \quad h_1(x, y) \leqq z \leqq h_2(x, y).$$

Let f be continuous on A. Then

$$\int_A f = \int_a^b \left[\int_{g_1(x)}^{g_2(x)} \left(\int_{h_1(x,y)}^{h_2(x,y)} f(x, y, z)\, dz \right) dy \right] dx.$$

For simplicity, the integral on the right will also be written without the brackets.

Example. Consider the tetrahedron T spanned by 0 and the three unit vectors (Fig. 18).

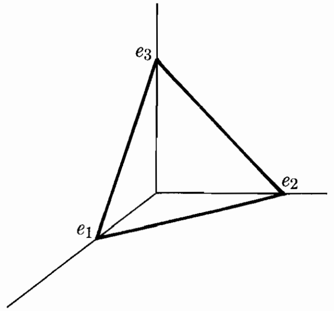


Figure 18

This tetrahedron is described by the inequalities:

$$0 \leqq x \leqq 1, \qquad 0 \leqq y \leqq 1 - x, \qquad 0 \leqq z \leqq 1 - x - y.$$

Hence if f is a function on the tetrahedron, its integral over T is given by

$$\int_T f = \int_0^1 \int_0^{1-x} \int_0^{1-x-y} f(x, y, z)\, dz\, dy\, dx.$$

For the constant function 1, the integral gives you the volume of the tetrahedron, and you should have no difficulty in evaluating it, finding the value $\frac{1}{6}$.

Case 2. Cylindrical coordinates. Analogously to the polar coordinate map in 2-space, we consider the cylindrical coordinate map in three space, given by

$$G(r, \theta, z) = (r \cos \theta, r \sin \theta, z).$$

In other words,

$$x = r \cos \theta$$
$$y = r \sin \theta$$
$$z = z.$$

The image of a box B defined by the inequalities:

$$0 \leqq \theta_1 \leqq \theta \leqq \theta_2 \leqq 2\pi, \quad 0 \leqq r_1 \leqq r \leqq r_2, \quad \text{and} \quad z_1 \leqq z \leqq z_2,$$

under the map G is shown in the following picture.

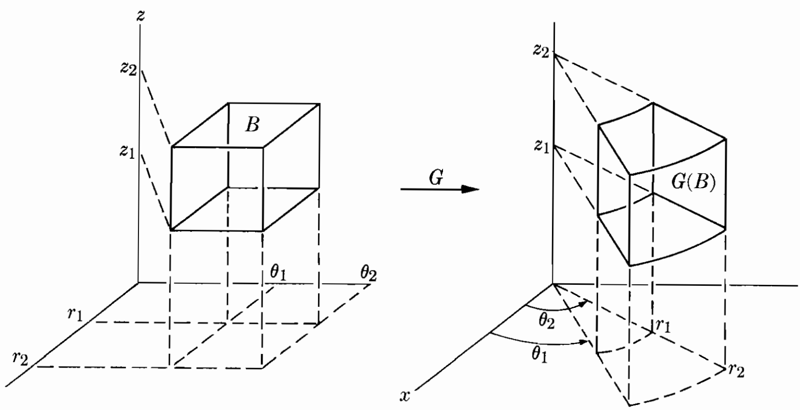


Figure 19

The volume of the elementary region on the right, which is the image of the box under the cylindrical coordinate map, is equal to the area of the base times the altitude, and is therefore equal to

$$(z_2 - z_1)\left(\frac{r_2^2 - r_1^2}{2}\right)(\theta_2 - \theta_1).$$

This expression can be rewritten in the form

$$\bar{r}(z_2 - z_1)(r_2 - r_1)(\theta_2 - \theta_1),$$

where

$$\bar{r} = \frac{r_2 + r_1}{2}.$$

Therefore if a function fs given in terms of the rectangular coordinates over some region, which is the image $G(A)$ of some region in the (r, θ, z)-space, then its integral is given in terms of the cylindrical coordinates by

$$\boxed{\iiint_{G(A)} f(x, y, z)\, dz\, dy\, dx = \iiint_{A} f(r\cos\theta, r\sin\theta, z) r\, dz\, dr\, d\theta.}$$

Indeed, the same kind of argument applies as with polar coordinates.

A region may be described by inequalities given by functions. For instance, *let A be the region in the (θ, r, z)-space consisting of all points (θ, r, z) satisfying conditions:*

$$a \leqq \theta \leqq b, \qquad (b \leqq a + 2\pi),$$
$$0 \leqq g_1(\theta) \leqq r \leqq g_2(\theta),$$

with smooth functions g_1, g_2 defined on the interval $[a, b]$, and

$$h_1(\theta, r) \leqq z \leqq h_2(\theta, r),$$

with smooth functions h_1, h_2 defined on the 2-dimensional region bounded by $\theta = a$, $\theta = b$, and g_1, g_2, i.e. the region consisting of all points (θ, r) satisfying the above inequalities. Let G be the map of cylindrical coordinates given above. Let f be a continuous function on the region $G(A)$ in the (x, y, z)-space. Let

$$f^*(\theta, r, z) = f(r\cos\theta, r\sin\theta, z).$$

Then

$$\boxed{\int_{G(A)} f = \int_a^b \int_{g_1(\theta)}^{g_2(\theta)} \int_{h_1(\theta,r)}^{h_2(\theta,r)} f^*(\theta, r, z) r\,dz\,dr\,d\theta.}$$

The function which we denote by f^* may be viewed as the function f in terms of the cylindrical coordinates.

Example. Find the mass of a solid bounded by the polar coordinates $\pi/3 \leqq \theta \leqq 2\pi/3$ and $r = \cos\theta$ and by $z = 0$, $z = r$, if the density is given by the function

$$f^*(r, \theta, z) = 3r.$$

The mass is given by the integral

$$\int_{\pi/3}^{2\pi/3} \int_0^{\cos\theta} \int_0^r 3r \cdot r\,dz\,dr\,d\theta.$$

Integrating the inner integral with respect to z yields $3r^2 r = 3r^3$. Integrating with respect to r between 0 and $\cos\theta$ yields

$$\left.\frac{3r^4}{4}\right|_0^{\cos\theta} = \frac{3\cos^4\theta}{4}.$$

Finally we integrate with respect to θ, using elementary techniques of integration: $\cos^2\theta = (1 + \cos 2\theta)/2$ so that

$$\begin{aligned}\cos^4\theta &= \tfrac{1}{4}(1 + 2\cos 2\theta + \cos^2 2\theta)\\ &= \frac{1}{4}\left(1 + 2\cos 2\theta + \frac{1 + \cos 4\theta}{2}\right).\end{aligned}$$

Case 3. Spherical coordinates. We consider the region in coordinates (ρ, θ, φ) described by

$$0 \leqq \rho, \qquad 0 \leqq \varphi \leqq \pi, \qquad 0 \leqq \theta \leqq 2\pi.$$

These coordinates can be used to describe a point in 3-space as shown on the following picture.

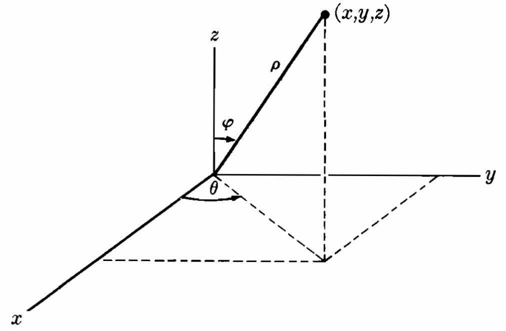


Figure 20

In fact, we let

$$\rho = \sqrt{x^2 + y^2 + z^2}.$$

We denote this by ρ to distinguish it from the polar coordinate r in the (x, y)-plane. We see that

$$x^2 + y^2 = \rho^2 - z^2 = \rho^2 \sin^2 \varphi$$

so that the polar r is given by

$$r = \sqrt{x^2 + y^2} = \rho \sin \varphi.$$

In taking the square root, we do not need to use the absolute value $|\sin \varphi|$ because we take $0 \leqq \varphi \leqq \pi$ so that $\sin \varphi \geqq 0$ for our values of φ.
We can now integrate this between the given limits, and we find

$$\frac{3}{4}\int_{\pi/3}^{2\pi/3} \cos^4 \theta \, d\theta = \frac{3}{16}\left(\frac{\pi}{3} - \sqrt{3} + \frac{\pi}{6} + \frac{\sqrt{3}}{8}\right).$$

Note. In the above example, the function is already given in terms of (r, θ, z). It corresponds to the function $f(x, y, z) = 3\sqrt{x^2 + y^2}$. Indeed, taking $f(r \cos \theta, r \sin \theta, z)$ yields $3r$.

Example. Let us find the volume of the region inside the cylinder $r = 4 \cos \theta$, bounded above by the sphere $r^2 + z^2 = 16$, and below by the plane $z = 0$. In the (x, y)-plane, the equation $r = 4 \cos \theta$ is that of a circle, with $0 \leqq \theta \leqq \pi$. The region is then defined by means of the other two inequalities

$$0 \leqq z \leqq 16 - r^2 \qquad \text{and} \qquad 0 \leqq r \leqq 4 \cos \theta.$$

Therefore the desired volume V is the integral

$$V = \int_0^{\pi} \int_0^{4\cos\theta} \int_0^{\sqrt{16-r^2}} r\,dz\,dr\,d\theta$$

$$= \int_0^{\pi} \int_0^{4\cos\theta} r\sqrt{16-r^2}\,dr\,d\theta$$

$$= -\tfrac{64}{3}\int_0^{\pi} (\sin^3\theta - 1)\,d\theta = \tfrac{64}{9}(3\pi - 4).$$

From the formulas $x = r\cos\theta$ and $y = r\sin\theta$, we then obtain the relationship between (x, y, z) and (ρ, θ, φ), namely:

$$x = \rho\sin\varphi\cos\theta,$$
$$y = \rho\sin\varphi\sin\theta,$$
$$z = \rho\cos\varphi.$$

We can also say that we have a mapping $G\colon \mathbf{R}^3 \to \mathbf{R}^3$ given by

$$G(\rho, \theta, \varphi) = (\rho\sin\varphi\cos\theta, \rho\sin\varphi\sin\theta, \cos\varphi),$$

Let R be the 3-dimensional rectangle in the (ρ, θ, φ)-space described by the inequalities:

$$\boxed{\begin{gathered} \theta_1 \leqq \theta \leqq \theta_2, \qquad (\theta_2 \leqq \theta_1 + 2\pi), \\ 0 \leqq \rho_1 \leqq \rho \leqq \rho_2, \\ 0 \leqq \varphi_1 \leqq \varphi \leqq \varphi_2 \leqq \pi. \end{gathered}}$$

The image of R under the map G is then an elementary spherical region as shown in the next picture.

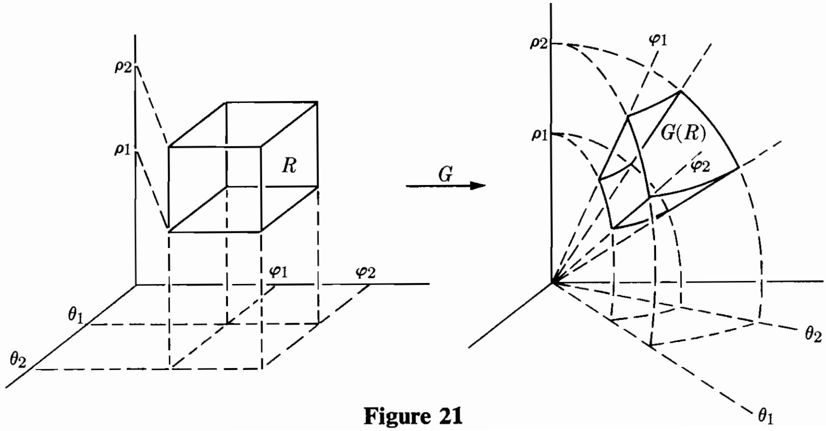


Figure 21

The volume of the elementary spherical region $G(R)$ just described is equal to

$$\boxed{\left(\frac{\rho_2^3}{3} - \frac{\rho_1^3}{3}\right)(\cos\varphi_1 - \cos\varphi_2)(\theta_2 - \theta_1).}$$

In order to see this, we shall find the volume of a slightly simpler region, namely that lying above a cone and inside a sphere as shown on the next figure.

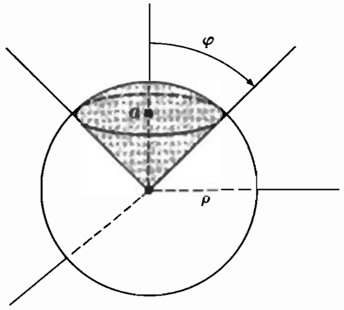


Figure 22

The radius of the sphere is ρ, and the angle of the cone is φ, as shown on the figure. We let a be the height at which the cone meets the sphere. The volume of this region consists of two pieces. The first is the volume of a cone of height a, and whose base is

$$b = \rho \sin\varphi.$$

Observe that $a = \rho\cos\varphi$. The volume of this cone is therefore equal to

$$\frac{\pi}{3}\rho^3 \sin^2\varphi\cos\varphi.$$

The other piece lies below the spherical dome, and can be obtained as a volume of revolution of the curve $x^2 + y^2 = \rho^2$, letting x range between a and ρ. You should know how to do this, and you will find the answer

$$\pi(\tfrac{2}{3}\rho^3 - \rho^3\cos\varphi + \tfrac{1}{3}\rho^3\cos^3\varphi).$$

Adding our two volumes together, and noting that

$$\cos^3\varphi = \cos^2\varphi\cos\varphi,$$

we find that the volume of the region lying above the cone and inside the sphere is equal to

$$\tfrac{2}{3}\pi\rho^3 - \tfrac{2}{3}\pi\rho^3\cos\varphi.$$

The volume of this region lying between angles φ_1 and φ_2 is obtained by subtracting, and is equal to

$$\tfrac{2}{3}\pi\rho^3 (\cos\varphi_1 - \cos\varphi_2).$$

Considering only the part lying between the spheres of radii ρ_1 and ρ_2, we obtain its volume again by subtraction, and get

$$\tfrac{2}{3}\pi(\rho_2^3 - \rho_1^3)(\cos\varphi_1 - \cos\varphi_2).$$

Finally, we have to take that part lying between angles θ_1 and θ_2, that is, take the fraction

$$\frac{\theta_2 - \theta_1}{2\pi}$$

of this last volume. In this way, we obtain precisely the desired volume of the elementary spherical region of Fig. 21.

Using the mean value theorem, we find that

$$\frac{\rho_2^3}{3} - \frac{\rho_1^3}{3} = \bar{\rho}^2(\rho_2 - \rho_1),$$

for some number $\bar{\rho}$ between ρ_1 and ρ_2. Again by the mean value theorem, we find that

$$\cos\varphi_1 - \cos\varphi_2 = \sin\bar{\varphi}\,(\varphi_2 - \varphi_1).$$

Hence the volume of the elementary spherical region $G(R)$ is equal to

$$\bar{\rho}^2 \sin\bar{\varphi}\,(\rho_2 - \rho_1)(\varphi_2 - \varphi_1)(\theta_2 - \theta_1).$$

By forming Riemann sums we already had in polar coordinates, it is therefore reasonable that the triple integral of a function f over a region $G(A)$, which is obtained as the image under the spherical coordinate map, can be expressed in terms of spherical coordinates by the formula:

$$\boxed{\iiint_A f(G(\rho,\theta,\varphi))\rho^2 \sin\varphi\, d\rho\, d\varphi\, d\theta = \iiint_{G(A)} f(x, y, z)\, dz\, dy\, dx.}$$

As usual, $f(G(\rho, \theta, \varphi)) = f^*(\rho, \theta, \varphi)$ is the value of the function at the given point (x, y, z) in terms of the spherical coordinates of the point (ρ, θ, φ).

Example. As a check, let us apply the general formula directly to see if it gives us the same answer for the volume of the elementary spherical

region $G(R)$. We are supposed to evaluate the integral

$$\iiint_{G(R)} dz\,dy\,dx = \int_{\theta_1}^{\theta_2}\int_{\varphi_1}^{\varphi_2}\int_{\rho_1}^{\rho_2} \rho^2 \sin\varphi\, d\rho\, d\varphi\, d\theta$$

In this case, the repeated 3-fold integral splits into separate integrals with respect to ρ, φ, θ independently. These integrals are of course very simple to evaluate. In this case, the limits of integration are constant. Integrating with respect to ρ yields the factor $\frac{1}{3}(\rho_2^3 - \rho_1^3)$. Integrating with respect to φ yields the factor $(\cos\varphi_1 - \cos\varphi_2)$. Integrating with respect to θ yields the factor $(\theta_2 - \theta_1)$. Thus the evaluation of the integral checks with the arguments given previously.

Example. Find the volume above the cone $z^2 = x^2 + y^2$ and inside the sphere $x^2 + y^2 + z^2 = 1$ (Fig. 23).

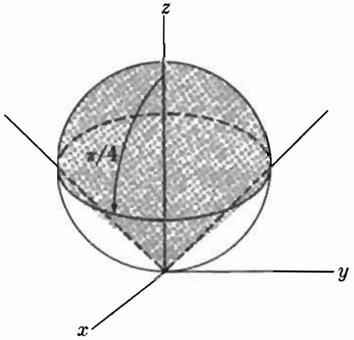


Figure 23

The equation for the sphere in spherical coordinates is obtained by the values

$$\rho^2 = x^2 + y^2 + z^2$$

and

$$z = \rho\cos\varphi.$$

Thus the sphere is given by the equation

$$\rho = \cos\varphi.$$

The cone is given by $\cos^2\varphi = \sin^2\varphi$, and since $0 \leqq \varphi \leqq \pi$ this is the same as $\varphi = \pi/4$. Thus the region of integration is the image under the spherical coordinate map of the region A described by the inequalities:

$$0 \leqq \theta \leqq 2\pi, \qquad 0 \leqq \varphi \leqq \pi/4, \qquad 0 \leqq \rho \leqq \cos\varphi.$$

Hence our volume is equal to the integral

$$\int_{G(A)} 1 = \int_0^{2\pi}\int_0^{\pi/4}\int_0^{\cos\varphi} \rho^2 \sin\varphi\, d\rho\, d\varphi\, d\theta.$$

The inside integral with respect to ρ is equal to

$$(\sin\varphi)\frac{\rho^3}{3}\Big|_0^{\cos\varphi} = \frac{1}{3}\cos^3\varphi\sin\varphi.$$

This is now easily integrated with respect to φ, and yields

$$\frac{1}{3}\,\frac{-\cos^4\varphi}{4}\Big|_0^{\pi/4} = \frac{1}{12}\left(-\frac{1}{4}+1\right) = \frac{3}{48}\cdot$$

Finally, we integrate with respect to θ, and the final answer is therefore equal to

$$\tfrac{3}{48}\cdot 2\pi = \tfrac{1}{8}\pi.$$

Example. Find the mass of a solid body S determined by the inequalities of spherical coordinates:

$$0 \leqq \theta \leqq \frac{\pi}{2}, \qquad \frac{\pi}{4} \leqq \varphi \leqq \arctan 2, \qquad 0 \leqq \rho \leqq \sqrt{6},$$

if the density, given as a function of the spherical coordinates (ρ, θ, φ), is equal to $1/\rho$.

To find the mass, we have to integrate the given function over the region. The integral is given by

$$\int_0^{\pi/2}\int_{\pi/4}^{\arctan 2}\int_0^{\sqrt{6}} \frac{1}{\rho}\rho^2\sin\varphi\, d\rho\, d\varphi\, d\theta.$$

Performing the repeated integral, we obtain

$$\frac{3\pi}{2}\left(\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{5}}\right).$$

We note that in the present example, the limits of integration are constants, and hence the repeated integral is equal to a product of the integrals

$$\int_0^{\pi/2} d\theta \cdot \int_{\pi/4}^{\arctan 2}\sin\varphi\, d\varphi \cdot \int_0^{\sqrt{6}} \rho\, d\rho.$$

Each integration can be performed separately. Of course, this does not hold when the limits of integration are non-constant functions.

As before, we have a similar integral when the boundaries of integration are not constant. We state the result:

Let a, b be numbers such that $0 \leqq b - a \leqq 2\pi$. Let $g_1(\theta)$, $g_2(\theta)$ be smooth functions of θ, defined on the interval $a \leqq \theta \leqq b$ such that

$$0 \leqq g_1(\theta) \leqq g_2(\theta) \leqq \pi.$$

Let h_1, h_2 be functions of two variables, defined and smooth on the region

consisting of all points (θ, φ) *such that*

$$a \leqq \theta \leqq b,$$
$$g_1(\theta) \leqq \varphi \leqq g_2(\theta)$$

and such that $0 \leqq h_1(\theta, \varphi) \leqq h_2(\theta, \varphi)$ *for all* (θ, φ) *in this region. Let* A *be the 3-dimensional region in the* (θ, φ, ρ)*-space consisting of all points such that*

$$a \leqq \theta \leqq b,$$
$$g_1(\theta) \leqq \varphi \leqq g_2(\theta),$$
$$h_1(\theta, \varphi) \leqq \rho \leqq h_2(\theta, \varphi).$$

Let G *be the spherical coordinate map, and let* f *be a continuous function on* $G(A)$. *Let* $f^*(\theta, \varphi, \rho) = f(G(\theta, \varphi, \rho))$. *Then*

$$\boxed{\int_{G(A)} f = \int_a^b \int_{g_1(\theta)}^{g_2(\theta)} \int_{h_1(\theta,\varphi)}^{h_2(\theta,\varphi)} f^*(\theta, \varphi, \rho)\rho^2 \sin\varphi \, d\rho \, d\varphi \, d\theta.}$$

Exercises

1. Find the volume inside the sphere

$$x^2 + y^2 + z^2 = a^2,$$

by using spherical coordinates.

2. Find the integral

$$\int_0^\pi \int_0^{\sin\theta} \int_0^{\rho\cos\theta} \rho^2 \, dz \, d\rho \, d\theta.$$

3. (a) Find the mass of a spherical ball of radius $a > 0$ if the density at any point is equal to a constant k times the distance of that point to the center.
(b) Find the integral of the function

$$f(x, y, z) = \frac{1}{\sqrt{x^2 + y^2 + z^2}}$$

over the spherical shell of inside radius a and outside radius 1. Assume $0 < a < 1$. What is the limit of this integral as $a \to 0$?

4. Find the mass of a spherical shell of inside radius a and outside radius b if the density at any point is inversely proportional to the distance from the center.

5. Find the integral of the function

$$f(x, y, z) = x^2$$

over that portion of the cylinder

$$x^2 + y^2 = a^2$$

lying between the planes

$$z = 0 \quad \text{and} \quad z = b > 0.$$

6. Find the mass of a sphere of radius a if the density at any point is proportional to the distance from a fixed plane passing through a diameter.
7. Find the volume of the region bounded by the cylinder $y = \cos x$, and the planes

$$z = y, \quad x = 0, \quad x = \pi/2, \quad \text{and} \quad z = 0.$$

8. Find the volume of the region bounded above by the sphere

$$x^2 + y^2 + z^2 = 1$$

and below by the surface

$$z = x^2 + y^2.$$

9. Find the volume of that portion of the sphere $x^2 + y^2 + z^2 = a^2$, which is inside the cylinder $r = a \sin \theta$, using cylindrical coordinates.
10. Find the volume above the cone $z^2 = x^2 + y^2$ and inside the sphere $\rho = 2a \cos \varphi$ (spherical coordinates). [Draw a picture. What is the center of the sphere? What is the equation of the cone in spherical coordinates?]
11. Find the volumes of the following regions, in 3-space.
    (a) Bounded above by the plane $z = 1$, and below by the top half of $z^2 = x^2 + y^2$.
    (b) Bounded above and below by $z^2 = x^2 + y^2$, and on the sides by $x^2 + y^2 + z^2 = 1$.
    (c) Bounded above by $z = x^2 + y^2$, below by $z = 0$, and on the sides by $x^2 + y^2 = 1$.
    (d) Bounded above by $z = x$, and below by $z = x^2 + y^2$.
12. Find the integral of the following functions over the indicated region, in 3-space.
    (a) $f(x, y, z) = x^2$ over the tetrahedron bounded by the plane

$$12x + 20y + 15z = 60,$$

    and the coordinate planes.
    (b) $f(x, y, z) = y$ over the tetrahedron as in (a).
    (c) $f(x, y, z) = 7yz$ over the region on the positive side of the (x, z)-plane, bounded by the planes $y = 0$, $z = 0$, and $z = a$ (for some positive number a), and the cylinder $x^2 + y^2 = b^2$ $(b > 0)$.
13. Find the volume of the region bounded by the cylinder $r^2 = 16$, by the plane $z = 0$, and below the plane $y = 2z$.
14. Let n be a positive integer, and let $f(x, y, z) = 1/\rho^n$, where

$$\rho = \sqrt{x^2 + y^2 + z^2}.$$

(a) Find the integral of the function

$$f(x, y, z) = 1/\rho^n$$

over the region contained between two spheres of radii a and b respectively, with $0 < a < b$.

(b) For which values of n does this integral approach a limit as $a \to 0$? Compare with the similar result which you may have worked out in the preceding section for a function of two variables.

§5. *Center of mass*

Double and triple integrals have an application to finding the center of mass of a body in the plane or in 3-space. Let A be such a body, say in the plane, and let f be its density function, giving the density at every point. Let m be the total mass. Let $(\bar{x}, \bar{y})$ be the coordinates of the center of mass. Then they are given by the integrals:

$$\bar{x} = \frac{1}{m}\iint_A xf(x, y)\,dy\,dx = \frac{\displaystyle\iint_A xf(x, y)\,dy\,dx}{\displaystyle\iint_A f(x, y)\,dy\,dx}.$$

$$\bar{y} = \frac{1}{m}\iint_A yf(x, y)\,dy\,dx = \frac{\displaystyle\iint_A yf(x, y)\,dy\,dx}{\displaystyle\iint_A f(x, y)\,dy\,dx}.$$

In 3-space, we would of course use the triple integral of $xf(x, y, z)$ and $yf(x, y, z)$ over the body. For instance, the third coordinate of the center of mass of a body of total mass m in 3-space is given by

$$\bar{z} = \frac{1}{m}\iiint_A zf(x, y, z)\,dx\,dy\,dz.$$

Example. Let us find the center of mass of the part of the first quadrant lying in the disc of radius 1, as shown on Fig. 24. We assume in this case that the density is uniform, say equal to 1.

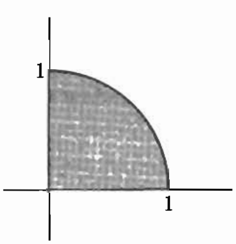


Figure 24

The total mass m is equal to $\pi/4$, and

$$\bar{x} = \frac{1}{m} \iint_A x\, dy\, dx.$$

The integral is best evaluated by changing variables, i.e. using polar coordinates. Thus we find:

$$\iint_A x\, dy\, dx = \int_0^{\pi/2} \int_0^1 r \cos\theta\, r\, dr\, d\theta = \tfrac{1}{3}.$$

Hence

$$\bar{x} = \frac{4}{3\pi}.$$

Similarly, or by symmetry, we have $\bar{y} = \dfrac{4}{3\pi}$ also.

Example. Let us find the z-coordinate of the center of mass of the part of the unit ball consisting of all points (x, y, z) whose coordinates are $\geqq 0$. If A denotes this part of the ball, then we have

$$\bar{z} = \frac{1}{m} \iiint_A z\, dx\, dy\, dz.$$

By using spherical coordinates, the integral is equal to

$$\int_0^{\pi/2} \int_0^{\pi/2} \int_0^1 \rho \cos\varphi\, \rho^2 \sin\varphi\, d\rho\, d\varphi\, d\theta.$$

Again we easily find the value $\pi/16$. We also know that the mass of the total ball is $\frac{4}{3}\pi$. Hence the mass of our part of the ball is $\dfrac{1}{8} \cdot \dfrac{4\pi}{3} = \dfrac{\pi}{6}$, so that

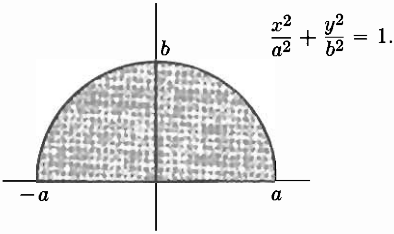
$$\bar{z} = \frac{\pi}{16} \cdot \frac{6}{\pi} = \frac{3}{8}.$$

Exercises

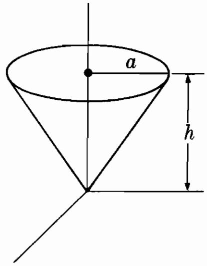
In each of the following cases, find the center of mass of the given body, assuming that the density is equal to 1.

1. The triangle whose vertices are $(0, 0)$, $(3, 0)$, and $(0, 5)$.
2. The region enclosed by the parabola $y = 6x - x^2$ and the line $y = x$.

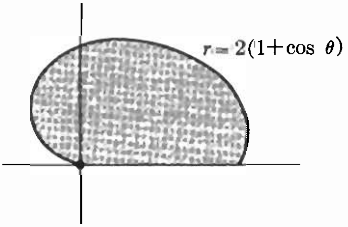
3. The upper half of the region enclosed by the ellipse as shown on the figure.



4. The region enclosed by the parabolas $y = 2x - x^2$ and $y = 3x^2 - 6x$.
5. The region enclosed by one arch of the curve $y = \sin x$.
6. The region bounded by the curves $y = \sin x$ and $y = \cos x$, for $0 \leqq x \leqq \pi/4$.
7. The region bounded by $y = \log x$ and $y = 0$, $1 \leqq x \leqq a$.
8. The inside of a cone of height h and base radius r, as shown on the figure.



9. Find (a) mass and (b) the center of mass of a plate bounced by the upper half of the curve $r = 2(1 + \cos \theta)$ (in polar coordinates) if the density is proportional to the distance from the origin. The plate is drawn on the next figure.



10. Find (a) the mass and (b) the center of mass of a right circular cylinder of radius a and height h if the density is proportional to the distance from the base.

11. (a) Find the mass of a circular plate of radius a whose density is proportional to the distance from the center.
    (b) Find the center of mass of this plate.
    (c) Find the center of mass of one quadrant of this plate.
12. Find the mass of a circular cylinder of radius a and height h if the density is proportional to the square of the distance from the axis.
13. Find the center of mass of a cone of height h and radius of the base equal to a, if the density is proportional to the distance from the base.

CHAPTER XIII

The Change of Variables Formula

§1. Determinants as area and volume

We shall study the manner in which area changes under an arbitrary mapping by approximating this mapping with a linear map. Therefore, first we study how area and volume change under a linear map, and this leads us to interpret the determinant as area and volume according as we are in $\mathbf{R}^2$ or $\mathbf{R}^3$.

Let us first consider $\mathbf{R}^2$. Let

$$A = \begin{pmatrix} a \\ c \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b \\ d \end{pmatrix}$$

be two non-zero vectors in the plane, and suppose that they are not scalar multiples of each other. We have already seen that they span a parallelogram, as shown on Fig. 1.

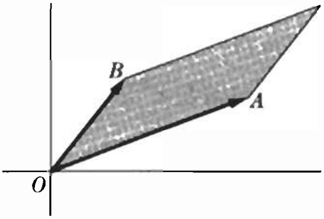


Figure 1

***Theorem 1 in* $\mathbf{R}^2$.** *Let A, B be non-zero elements of $\mathbf{R}^2$, which are not scalar multiples of each other. Then the area of the parallelogram spanned by A and B is equal to the absolute value of the determinant $|D(A, B)|$.*

Proof. We assume known that this area is equal to the product of the lengths of the base times the altitude, and this is equal to

$$|A|\,|B|\,|\sin\theta|,$$

where θ is the angle between A and B (i.e. between $\overrightarrow{OA}$ and $\overrightarrow{OB}$). This is

illustrated on Fig. 2.

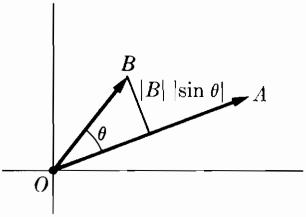


Figure 2

Note that

$$|\sin \theta| = \sqrt{1 - \cos^2 \theta},$$

and recall from the theory of the dot product that

$$\cos \theta = \frac{A \cdot B}{|A|\,|B|}.$$

We have

$$\begin{aligned}\text{Area of parallelogram} &= |A|\,|B|\sqrt{1 - \frac{(A \cdot B)^2}{|A|^2\,|B|^2}} \\ &= \sqrt{|A|^2\,|B|^2 - (A \cdot B)^2}.\end{aligned}$$

All that remains to be done is to plug in the coordinates of A and B to see what we want come out. Indeed, the above expression is equal to the square root of

$$(a^2 + c^2)(b^2 + d^2) - (ab + cd)^2.$$

If you expand this out, you will find that this last expression is equal to

$$(ad - bc)^2.$$

Consequently, the area of the parallelogram is equal to

$$\sqrt{(ad - bc)^2} = |ad - bc| = |D(A, B)|.$$

This proves our assertion.

***Theorem 1 in* $\mathbf{R}^3$.** *Let A, B, C be vectors in $\mathbf{R}^3$, and assume that the segments $\overrightarrow{OA}$, $\overrightarrow{OB}$, $\overrightarrow{OC}$ do not all lie in a plane. Then the volume of the box spanned by A, B, C is equal to the absolute value of the determinant.*

$$|D(A, B, C)|.$$

Proof: Similar arguments to those which applied in $\mathbf{R}^2$ show us that the area of the base of the box, spanned by A and B, is equal to

$$(*) \qquad \sqrt{|A|^2\,|B|^2 - (A \cdot B)^2}.$$

Look at Fig. 3.

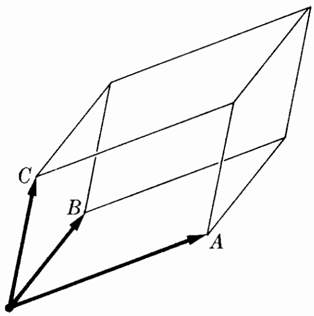


Figure 3

The volume of the box is equal to the area of this base times the altitude, and this altitude is equal to the length of the projection of C along a vector perpendicular to A and B. You should now have read the section on cross products, because the simplest way to handle the present situation is to use the cross product. We know that $A \times B$ is such a vector, perpendicular to A and B. The projection of C on $A \times B$ is equal to

$$\frac{C \cdot (A \times B)}{(A \times B) \cdot (A \times B)} A \times B,$$

where the number in front of $A \times B$ is the component of C along $A \times B$ as studied in Chapter I. Therefore the length of this projection is equal to

$$(**) \qquad \frac{|C \cdot (A \times B)|}{|A \times B|}.$$

On the other hand, if you look at property **CP 6** of the cross product in Chapter I, §6, you will find that (*) is equal to $|A \times B|$. Therefore, the volume of the box spanned by A, B, C, which is equal to the product of (*) and (**), is seen to be equal to

$$|C \cdot (A \times B)|.$$

All that remains to be done is for you to plug in the coordinates, to see that

this is equal to the absolute value of the determinant. You let

$$A = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \quad B = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}, \quad C = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix},$$

use the definition of the cross product of $A \times B$, and then dot with C. You will find precisely the six terms which give the determinant $D(A, B, C)$, up to a sign, which is killed by the absolute value. This proves Theorem 1 in $\mathbf{R}^3$.

Example. Let $A = (3, 1)$ and $B = (2, -5)$. Then the area of the parallelogram spanned by A and B is equal to the absolute value of the determinant

$$\begin{vmatrix} 3 & 1 \\ 2 & -5 \end{vmatrix} = -15 - 2 = -17.$$

Hence this area is equal to 17. *Note:* We wrote our vectors horizontally. We get the same determinant as if we write them vertically, namely

$$\begin{vmatrix} 3 & 2 \\ 1 & -5 \end{vmatrix} = -17,$$

because we know that the determinant of the transpose of a matrix is equal to the determinant of the matrix.

We interpret Theorem 1 in terms of linear maps. Given vectors A, B in the plane, we know that there exists a unique linear map

$$L\colon \mathbf{R}^2 \to \mathbf{R}^2$$

such that $L(E^1) = A$ and $L(E^2) = B$. In fact, if

$$A = aE^1 + cE^2, \qquad B = bE^1 + dE^2,$$

then the matrix associated with the linear map is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Furthermore, if we denote by S the unit cube spanned by E^1, E^2, and by P the parallelogram spanned by A, B, then P is the image under L of S, that is $L(S) = P$. Indeed, as we have seen, for $0 \leqq t_i \leqq 1$ we have

$$L(t_1E^1 + t_2E^2) = t_1L(E^1) + t_2L(E^2) = t_1A + t_2B.$$

Let us define the **determinant of a linear map** to be the determinant of its associated matrix. We conclude that

$$(\text{Area of } P) = |\operatorname{Det}(L)|.$$

Example. The area of the parallelogram spanned by the vectors (2, 1) and (3, −1) (Fig. 4) is equal to the absolute value of

$$\begin{vmatrix} 2 & 1 \\ 3 & -1 \end{vmatrix} = -5,$$

and hence is equal to 5.

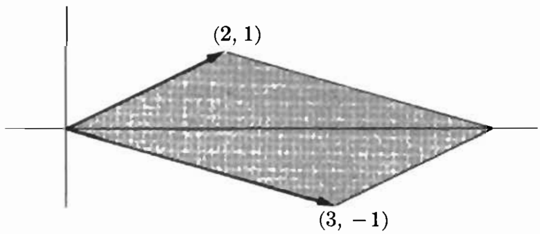


Figure 4

Theorem 2. *Let P be a parallelogram spanned by two vectors in $\mathbf{R}^2$. Let $L\colon \mathbf{R}^2 \to \mathbf{R}^2$ be a linear map. Then*

$$\text{Area of } L(P) = |\operatorname{Det} L| \,(\text{Area of } P).$$

Proof. Suppose that P is spanned by two vectors A, B. Then $L(P)$ is spanned by $L(A)$ and $L(B)$. (Cf. Fig. 5). There is a linear map $L_1\colon \mathbf{R}^2 \to \mathbf{R}^2$ such that

$$L_1(E^1) = A \quad \text{and} \quad L_1(E^2) = B.$$

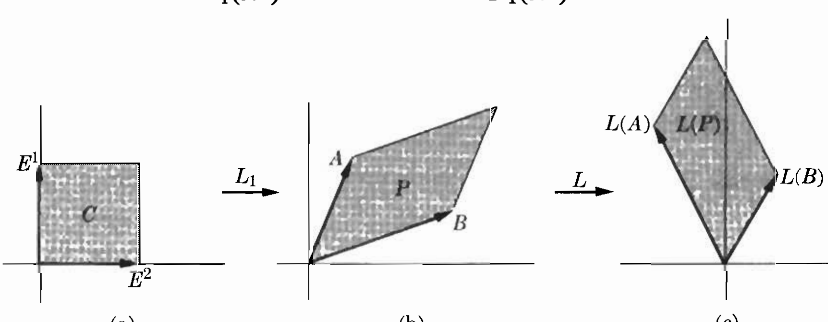


Figure 5

Then $P = L_1(S)$, where S is the unit square, and

$$L(P) = L\big(L_1(S)\big) = (L \circ L_1)(S).$$

By what we proved above in (*), we obtain

$$\text{Area } L(P) = |\text{Det}(L \circ L_1)| = |\text{Det}(L)\,\text{Det}(L_1)| = |\text{Det}(L)|\,\text{Area}(P),$$

thus proving our assertion.

Corollary. *For any rectangle R with sides parallel to the axes, and any linear map*

$$L\colon \mathbf{R}^2 \to \mathbf{R}^2,$$

we have

$$\text{Area } L(R) = |\text{Det}(L)|\,\text{Area}(R).$$

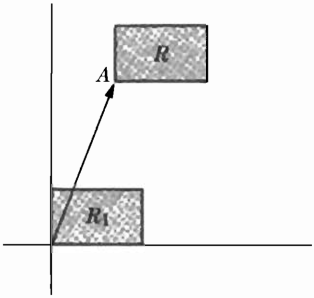


Figure 6

Proof. The rectangle R is equal to the translation of a rectangle R_1 as shown on Fig. 6, with one corner at the origin, that is

$$R = R_1 + A.$$

Then

$$L(R) = L(R_1) + L(A).$$

The area of $L(R_1)$ is the same as the area of $L(R_1) + L(A)$ (i.e. the translation of $L(R_1)$ by $L(A)$). All we have to do is apply Theorem 2 to complete the proof.

Example. The area of the parallelogram spanned by the vectors

$$(3, -2) \qquad \text{and} \qquad (4, 1)$$

is equal to the absolute value of the determinant

$$\begin{vmatrix} 3 & -2 \\ 4 & 1 \end{vmatrix}.$$

The determinant is equal to 11, so this is also the area.

Example. The area of the parallelogram spanned by the vectors

$$(3, 2) \quad \text{and} \quad (4, 1)$$

is equal to the absolute value of the determinant

$$\begin{vmatrix} 3 & 2 \\ 4 & 1 \end{vmatrix}.$$

The determinant is equal to -5, so the area is equal to 5.

Example. The volume of the box spanned by the vectors

$$(3, 0, 1), \quad (1, 2, 5), \quad (-1, 4, 2)$$

is equal to 42, because the determinant

$$\begin{vmatrix} 3 & 0 & 1 \\ 1 & 2 & 5 \\ -1 & 4 & 2 \end{vmatrix}$$

has the value -42.

We can also formulate Theorem 2 and its corollary in 3-space.

Theorem 2 in 3 space. *Let P be a parallelotope (box) in 3-space, spanned by three vectors. Let $L\colon \mathbf{R}^3 \to \mathbf{R}^3$ be a linear map. Then*

$$\text{Volume of } L(P) = |\text{Det } L| \text{ (Volume of } P).$$

Corollary. *For any rectangular box R in 3-space and any linear map $L\colon \mathbf{R}^3 \to \mathbf{R}^3$, we have*

$$\text{Vol } L(R) = |\text{Det } (L)| \text{ Vol } (R).$$

The proofs are exactly like those in 2-space, drawing 3-dimensional boxes instead of 2-dimensional rectangles.

Exercises

1. Find the area of the parallelogram spanned by
   (a) $(-3, 5)$ and $(2, -1)$. (b) $(2, 3)$ and $(4, -1)$.
2. Find the area of the parallelogram spanned by the following vectors.
   (a) $(2, 1)$ and $(-4, 5)$ (b) $(3, 4)$ and $(-2, -3)$

3. Find the area of the parallelogram such that three corners of the parallelogram are given by the following points
   (a) $(1, 1)$, $(2, -1)$, $(4, 6)$ (b) $(-3, 2)$, $(1, 4)$, $(-2, -7)$
   (c) $(2, 5)$, $(-1, 4)$, $(1, 2)$ (d) $(1, 1)$, $(1, 0)$, $(2, 3)$
4. Find the volume of the parallelepiped spanned by the following vectors in 3-space.
   (a) $(1, 1, 3)$, $(1, 2, -1)$, $(1, 4, 1)$ (b) $(1, -1, 4)$, $(1, 1, 0)$, $(-1, 2, 5)$
   (c) $(-1, 2, 1)$, $(2, 0, 1)$, $(1, 3, 0)$ (d) $(-2, 2, 1)$, $(0, 1, 0)$, $(-4, 3, 2)$

§2. *Dilations*

This section will serve as an introduction to the general change of variables formula, and the interpretation of determinants as area and volume.

Let r be a positive number. If A is a vector in $\mathbf{R}^n$ (in practice, $\mathbf{R}^2$ or $\mathbf{R}^3$) we call rA the dilation of A by r. Thus dilation by r is a linear mapping,

$$A \mapsto rA.$$

We wish to analyze what happens to area in $\mathbf{R}^2$, and volume in $\mathbf{R}^3$, under a dilation. We start with the simplest case, that of a rectangle. Consider a rectangle whose sides have lengths a, b, as on Fig. 7(a). If we multiply the sides of the rectangle by r, we obtain a rectangle with sides ra, rb as on Fig. 7(b). The area of the dilated rectangle is equal to

$$rarb = r^2ab.$$

Thus dilation by r changes the area of the rectangle by r^2.

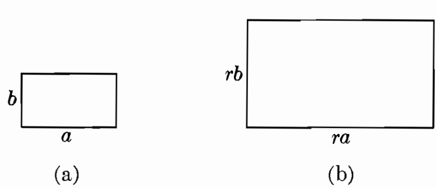


Figure 7

In general, let S be an arbitrary region in the plane $\mathbf{R}^2$, whose area can be approximated by the area of a finite number of rectangles. Then the area of S itself changes by r^2 under dilation by r, in other words,

$$\text{Area of } rS = r^2(\text{area of } S).$$

For instance, let D be the disc of radius r, so that D_1 is the disc of radius 1, centered at the origin (Fig. 8). Then $D_r = rD_1$.

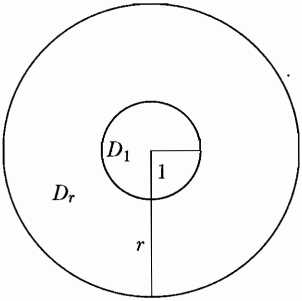


Figure 8

If π is the area of the disc of radius 1, then πr^2 must be the area of the disc of radius r. Of course, we knew this already, but we find this result here again from another point of view. More generally, consider a region S inside a curve as in Fig. 9(a), and let us draw the dilation of S by r in Fig. 9(b). To justify that the area changes by r^2, we draw a grid, approximating the areas by squares.

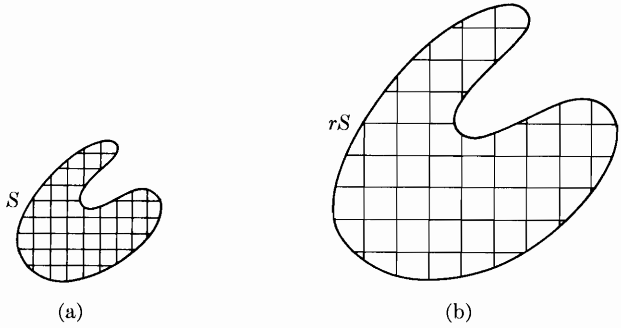
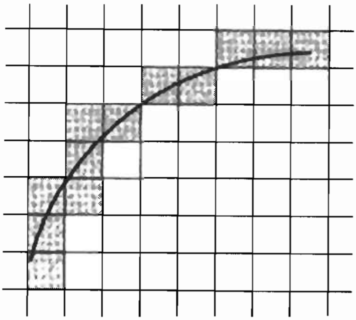


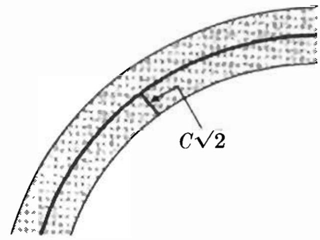
Figure 9

Under dilation by r, the area of each square gets multiplied by r^2, and so the sum of the areas of these squares, which approximates the area of S, also gets multiplied by r^2.

The question, of course, arises as to whether the squares lying inside S, and formed by a sufficiently fine grid, actually approximate S. We can see that they do, as follows. Let the sides of the squares in the grid have length c. (Fig. 10a.) Suppose that a square intersects the curve which



(a)



(b) **Figure 10**

bounds S. Let Z be this curve. Then any point in the square is at distance at most $c\sqrt{2}$ from the curve Z. This is because the distance between any two points of the square is at most $c\sqrt{2}$ (the length of the diagonal of the square). Let us draw a band of width $c\sqrt{2}$ on each side of the curve, as shown on Fig. 10(b). Then all the squares which intersect the curve must lie within that band. It is very plausible that the area of the band is at most equal to

$$2c\sqrt{2} \text{ times the length of the curve.}$$

Thus if we take c to be very small, i.e. if we take the grid to be a very fine grid, we see that the area of the region S is approximated by the area covered by the squares lying entirely inside the region. This explains why the area of S will get multiplied by r^2 under dilation by r.

We can also make a mixed dilation. Let r, s be two positive numbers. Consider the mapping of $\mathbf{R}^2$ given by

$$(x, y) \mapsto (rx, sy).$$

Thus we dilate the first coordinate by r and the second by s. If a rectangle R has sides of lengths a, b respectively, then the image of the rectangle under this mapping will be a rectangle with sides of lengths ra, sb. Hence the area of the image will be

$$rasb = rsab.$$

Thus the area changes by a factor of rs under our mapping.

An argument as before shows that if we submit a region S to such a mapping $F_{r,s}$ such that

$$F_{r,s}(x, y) = (rx, sy),$$

then its area will change by a factor of rs.

Example. We now have a very easy way of finding the area of an ellipse defined by an equation

$$\frac{x^2}{9} + \frac{y^2}{16} = 1.$$

Indeed, let $u = x/3$ and $v = y/4$. Then

$$u^2 + v^2 = 1,$$

and the ellipse is equal to the image of the circle under the mapping

$$(u, v) \mapsto (3u, 4v).$$

Hence the area of the ellipse is equal to $3 \cdot 4\pi = 12\pi$. Note how we did this without integration! However, the technique of the small grid is of course exactly the same technique which was used in the theory of the integral.

We can also develop the same ideas in 3-space. Consider dilation by r in 3-space, namely consider the mapping

$$(x, y, z) \mapsto (rx, ry, rz).$$

If P is a rectangular box with sides a, b, c, then its dilation by r will be a box with sides ra, rb, rc, and the dilated box will have volume

$$rarbrc = r^3abc.$$

Thus the volume of a box changes by r^3 under dilation by r.

Similarly, let r, s, t be positive numbers, and consider the linear map

$$F_{r,s,t}\colon \mathbf{R}^3 \mapsto \mathbf{R}^3$$

such that

$$F_{r,s,t}(x, y, z) = (rx, sy, tz).$$

We view this as a mixed dilation. If a rectangular box has sides of lengths a, b, c, then under $F_{r,s,t}$ it gets transformed into a box with sides ra, sb, tc whose volume is

$$rasbtc = rstabc.$$

Thus the volume gets multiplied by rst.

If we approximate an arbitrary region in 3-spaces by cubes, then we see in a manner analogous to that of 2-space that the volume of the region changes by a factor of r^3 under dilation by r, and changes by a factor of rst under the mixed dilation $F_{r,s,t}$.

Example. Find the volume of the region bounded by the equation

$$\frac{x^2}{9} + \frac{y^2}{16} + \frac{z^2}{25} = 1.$$

To do this, let

$$u = \frac{x}{3}, \qquad v = \frac{y}{4}, \qquad w = \frac{z}{5}.$$

The inequality

$$u^2 + v^2 + w^2 \leqq 1$$

defines the unit ball in $\mathbf{R}^3$, and our given region is obtained from this unit ball by the mixed dilation

$$F_{3,4,5}.$$

Assuming that the volume of the unit ball in $\mathbf{R}^3$ is equal to $\frac{4}{3}\pi$, we conclude that the volume of our region is equal to

$$3 \cdot 4 \cdot 5 \cdot \tfrac{4}{3}\pi = 80\pi.$$

In the next section, we investigate how area and volume change under general linear maps, not just dilations and mixed dilations.

Exercises

1. Find the area of the region bounded by the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

2. Find the volume of the region bounded by the surface

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

In both exercises, a, b, c are positive numbers. Use the ideas of this section.

3. Let A be the region in 3-space defined by the inequalities

$$0 \leqq x_i \qquad \text{and} \qquad x_1^4 + x_2^4 + x_3^4 \leqq 1.$$

Let C be the volume of this region.

(a) In terms of C, what is the volume of the region defined by the inequalities

$$0 \leqq x_i \qquad \text{and} \qquad x_1^4 + x_2^4 + x_3^4 \leqq 29?$$

(b) Same question if instead of 29 on the right you have a positive number r.

4. Let A be the region in 3-space defined by the inequalities

$$0 \leqq x_i \quad \text{and} \quad \sum_{i=1}^{3} x_i^5 \leqq 1.$$

Let C be the volume of this region.

(a) In terms of C, what is the volume of the region defined by the inequalities

$$0 \leqq x_i \quad \text{and} \quad \sum_{i=1}^{3} x_i^5 \leqq 33?$$

(b) Same question if instead of 33 you have an arbitrary positive number r on the right.

§3. Change of variables formula in two dimensions

Let R be a rectangle in $\mathbf{R}^2$ and suppose that R is contained in some open set U. Let

$$G\colon U \to \mathbf{R}^2$$

be a C^1-map. If G has two coordinate functions,

$$G(u, v) = \big(g_1(u, v), g_2(u, v)\big),$$

this means that the partial derivatives of g_1, g_2 exist and are continuous. We let $G(u, v) = (x, y)$, so that

$$x = g_1(u, v) \quad \text{and} \quad y = g_2(u, v).$$

Then the Jacobian determinant of the map G is by definition

$$\Delta_G(u, v) = \begin{vmatrix} \dfrac{\partial g_1}{\partial u} & \dfrac{\partial g_1}{\partial v} \\[2ex] \dfrac{\partial g_2}{\partial u} & \dfrac{\partial g_2}{\partial v} \end{vmatrix}.$$

This determinant is nothing but the determinant of the linear map $G'(u, v)$, which is the tangent linear map to G at (u, v).

Theorem 3. *Assume that G is C^1-invertible on the interior of the rectangle R. Let f be a function on $G(R)$ which is continuous except on a finite number of smooth curves. Then*

$$\int_R (f \circ G)|\Delta_G| = \int_{G(R)} f$$

or in terms of coordinates,

$$\iint\limits_R f\big(G(u, v)\big)|\Delta_G(u, v)|\, du\, dv = \iint\limits_{G(R)} f(x, y)\, dy\, dx.$$

The proof of Theorem 3 is not easy to establish rigorously, depending on ϵ and δ arguments. However, we can make it plausible in view of Theorem 2.

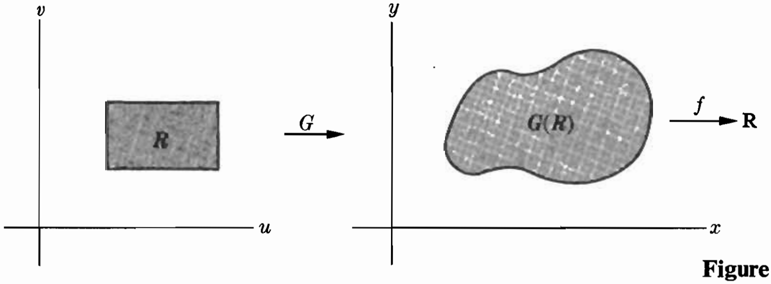


Figure 11

Indeed, suppose first that f is a constant function, say $f(x, y) = 1$ for all (x, y). Then the integral on the right, over $G(R)$, is simply the area of $G(R)$, and our formula reduces to

$$\int_R |\Delta_G| = \int_{G(R)} 1.$$

As we pointed out before, Δ_G is the determinant of the approximating linear map to G. If G is itself linear, then $G'(u, v) = G$ for all u, v and in this case, our formula reduces to Theorem 2, or rather its corollary. In the general case, one has to show that when one approximates G by its tangent linear map, which depends on (u, v), and then integrates $|\Delta_G|$ one still obtains the same result. Cf., for instance, my *Introduction to Analysis* for a complete proof. A special case will be proved in the next chapter.

When f is not a constant function, one still has the problem of reducing this case to the case of constant functions. This is done by taking a partition of R into small rectangles S, and then approximating f on each $G(S)$ by a constant function. Again, the details are out of the bounds of this book.

We shall now see how we recover the integral in terms of polar coordinates from the general Theorem 3.

Example 1. Let $x = r\cos\theta$ and $y = r\sin\theta$, $r \geqq 0$. Then in this case, we have computed previously the determinant, which is

$$\Delta_G(r, \theta) = r.$$

Thus we find again the formula

$$\iint_R f(r\cos\theta, r\sin\theta) r\, dr\, d\theta = \iint_{G(R)} f(x, y)\, dy\, dx.$$

Of course, we have to take a rectangle for which the map

$$G(r, \theta) = (r\cos\theta, r\sin\theta)$$

is invertible on the interior of the rectangle. For instance, we can take

$$0 \leqq r_1 \leqq r \leqq r_2 \quad \text{and} \quad 0 \leqq \theta_1 \leqq \theta \leqq \theta_2 \leqq 2\pi.$$

The image of the rectangle R is the portion $G(R)$ of the sector as shown in Fig. 12.

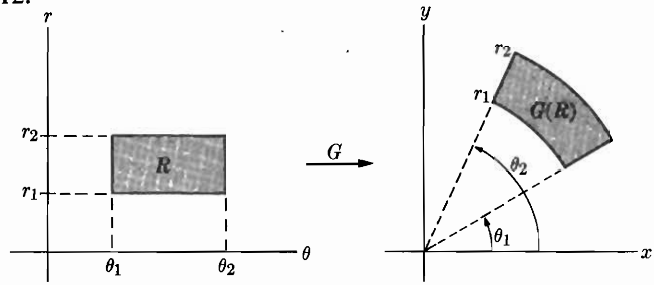


Figure 12

For the next example, we observe that if G is a linear map L, represented by a matrix M, then a Jacobian matrix of G is equal to this matrix M, and hence its Jacobian determinant is the determinant of M.

Example 2. Let T be the triangle whose vertices are $(1, 2)$, $(3, -1)$, and $(0, 0)$. Find the area of this triangle

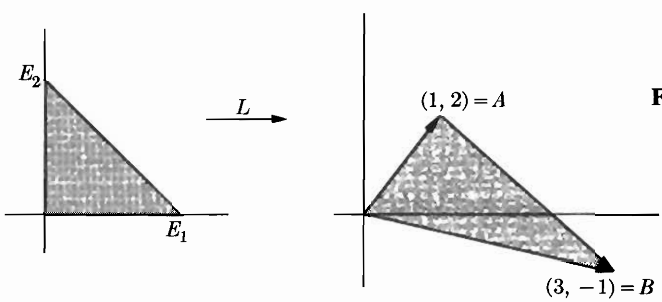


Figure 13

The triangle T is the image of the triangle spanned by 0, E_1, E_2 under a linear map, namely the linear map L such that

$$L(E_1) = (1, 2)$$

and

$$L(E_2) = (3, -1).$$

It is verified at once that $|\mathrm{Det}\,(L)| = 7$. Since the area of the triangle spanned by 0, E_1, E_2 is $\frac{1}{2}$, it follows that the desired area is equal to $\frac{7}{2}$.

Example 3. Let $(x, y) = G(u, v) = (e^u \cos v, e^u \sin v)$. Let R be the rectangle in the (u, v)-space defined by the inequalities $0 \leqq u \leqq 1$ and $0 \leqq v \leqq \pi$. It is not difficult to show that G satisfies the hypotheses of

Theorem 3, but we shall assume this. The Jacobian matrix of G is given by

$$\begin{pmatrix} e^u \cos v & -e^u \sin v \\ e^u \sin v & e^u \cos v \end{pmatrix}$$

so that its Jacobian determinant is equal to

$$\Delta_G(u, v) = e^{2u}.$$

Let $f(x, y) = x^2$. Then $f^*(u, v) = e^{2u} \cos^2 v$. According to Theorem 3, the integral of f over $G(R)$ is given by the integral

$$\int_0^1 \int_0^\pi e^{4u} \cos^2 v \, du \, dv$$

which can be evaluated very simply by integrating e^{4u} with respect to u and $\cos^2 v$ with respect to v, and taking the product. The final answer is then equal to

$$\frac{e^4 - 1}{4} + \tfrac{1}{2}\pi.$$

Example 4. Let S be the region enclosed by ellipse defined by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \qquad a, b > 0.$$

Its area is πab. (Why?) Let L be the linear map represented by the matrix

$$\begin{pmatrix} 1 & 2 \\ 3 & -5 \end{pmatrix}.$$

Its determinant is equal to -11. Hence the area of the image of S under L is $11\pi ab$.

Exercises

In the following exercises, you may assume that the map G satisfies the hypotheses of Theorem 3.

1. Let $(x, y) = G(u, v) = (u^2 - v^2, 2uv)$. Let A be the region defined by $u^2 + v^2 \leqq 1$ and $0 \leqq u$, $0 \leqq v$. Find the integral of the function

$$f(x, y) = 1/(x^2 + y^2)^{1/2}$$

over $G(A)$.

2. (a) Let $(x, y) = G(u, v)$ be the same map as in Exercise 1. Let A be the square $0 \leqq u \leqq 2$ and $0 \leqq v \leqq 2$. Find the area of $G(A)$.
(b) Find the integral of $f(x, y) = x$ over $G(A)$.

3. (a) Let R be the rectangle whose corners are $(1, 2)$, $(1, 5)$, $(3, 2)$, and $(3, 5)$. Let G be the linear map represented by the matrix

$$\begin{pmatrix} 2 & 1 \\ -1 & 3 \end{pmatrix}.$$

Find the area of $G(R)$.

(b) Same question if G is represented by the matrix $\begin{pmatrix} 3 & 2 \\ 1 & -6 \end{pmatrix}$.

4. Let $(x, y) = G(u, v) = (u + v, u^2 - v)$. Let A be the region in the first quadrant bounded by the axes and the line $u + v = 2$. Find the integral of the function $f(x, y) = 1/\sqrt{1 + 4x + 4y}$ over $G(A)$.

5. Let R be the unit square in the (u, v)-plane, defined by the inequalities

$$0 \leqq u \leqq 1 \quad \text{and} \quad 0 \leqq v \leqq 1.$$

(a) Sketch the image $F(R)$ of R under the mapping F such that

$$F(u, v) = (u, u + v^2).$$

In other words, $x = u$ and $y = u + v^2$.

(b) Compute the integral of the function $f(x, y) = x$ over the region $F(R)$ by using the change of variables formula.

6. Compute the area enclosed by the ellipse, defined by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} \leqq 1.$$

Take $a, b > 0$.

7. Let $(x, y) = G(u, v) = (u, v(1 + u^2))$. Let R be the rectangle $0 \leqq u \leqq 3$ and $0 \leqq v \leqq 2$. Find the integral of $f(x, y) = x$ over $G(R)$.

8. Let G be the linear map represented by the matrix

$$\begin{pmatrix} 3 & 0 \\ 1 & 5 \end{pmatrix}.$$

If A is the interior of a circle of radius 10, what is the area of $G(A)$?

9. Let G be the linear map of Exercise 8, and let A be the ellipse defined as in Exercise 6. What is the area of $G(A)$?

10. Let T be the triangle bounded by the x-axis, the y-axis, and the line $x + y = 1$. Let φ be a continuous function of one variable on the interval $[0, 1]$. Let m, n be positive integers. Show that

$$\iint_T \varphi(x + y)x^m y^n \, dy \, dx = c_{m,n} \int_0^1 \varphi(t)t^{m+n+1} \, dt,$$

where $c_{m,n}$ is the constant given by the integral $\int_0^1 (1 - t)^m t^n \, dt$. [*Hint:* Let $x = u - v$ and $y = v$.]

11. Let B be the region bounded by the ellipse $x^2/a^2 + y^2/b^2 = 1$. Find the integral
$$\iint_B y\,dy\,dx.$$

§4. Change of variables formula in three dimensions

The formula has the same shape as in two dimensions, namely:

Change of Variables Formula. *Let A be a bounded region in $\mathbf{R}^3$ whose boundary consists of finite number of smooth surfaces. Let A be contained in some open set U, and let*
$$G\colon U \to \mathbf{R}^3$$
be a C^1-map, which we assume to be C^1-invertible on the interior of A. Let f be a function on $G(A)$, bounded and continuous except on a finite number of smooth surfaces. Then

$$\boxed{\iiint_A f\big(G(u, v, w)\big)\,|\Delta_G(u, v, w)|\,du\,dv\,dw = \iiint_{G(A)} f(x, y, z)\,dz\,dy\,dx.}$$

In the 3-dimensional case, the Jacobian matrix of G at every point is then a 3×3 matrix.

Example. Let R be the 3-dimensional rectangle spanned by the three unit vectors E_1, E_2, E_3. Let A_1, A_2, A_3 be three vectors in 3-space, and let
$$G\colon \mathbf{R}^3 \to \mathbf{R}^3$$
be the linear map such that $G(E_1) = A_i$. Then $G(R)$ is a parallelotope (not necessarily rectangular). (Cf. Fig. 14.)

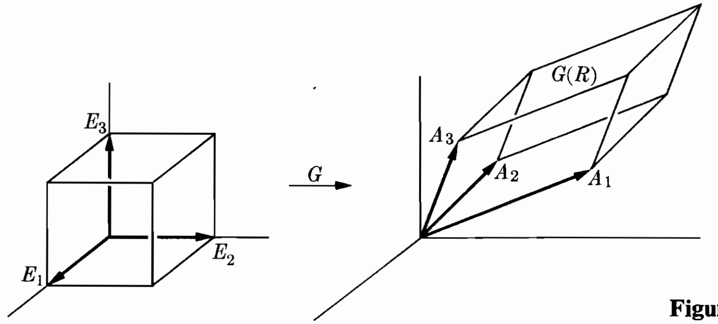


Figure 14

The Jacobian matrix of the map is constant, and is equal to the determinant of the matrix representing the linear map.

The volume of the unit cube is equal to 1. Hence the volume of $G(R)$ is equal to $|\operatorname{Det}(G)|$.

For instance, if

$$\begin{aligned} A_1 &= (3, 1, 2), \\ A_2 &= (1, -1, 4), \\ A_3 &= (2, 1, 0), \end{aligned}$$

then

$$\operatorname{Det}(G) = \begin{vmatrix} 3 & 1 & 2 \\ 1 & -1 & 4 \\ 2 & 1 & 0 \end{vmatrix} = -4$$

so the volume of $G(R)$ is equal to 4.

Similarly, we find the volume of the tetrahedron spanned by the origin and the three vectors

$$A_1 = (3, 1, 4), \qquad A_2 = (-1, 2, 1), \qquad A_3 = (5, -2, 1).$$

We assume that you have computed the volume of the tetrahedron spanned by the unit vectors, and found $\frac{1}{6}$. There is a unique linear map L which carries E_i on A_i. Hence the volume of our tetrahedron is equal to $\frac{1}{6}$ times the absolute value of the determinant of this linear map, that is to $\frac{1}{6}$ times the absolute value of the determinant

$$\begin{vmatrix} 3 & 1 & 4 \\ -1 & 2 & 1 \\ 5 & -2 & 1 \end{vmatrix} = -14.$$

The answer is 14/6.

If we are given the four vertices of a tetrahedron and want to find its volume, then we subtract one vertex from the others. This gives us a tetrahedron with one vertex at the origin, whose volume can be found by the above procedure.

Example. Consider the cylindrical coordinates map, given by

$$G(r, \theta, z) = (r\cos\theta, r\sin\theta, z).$$

Compute its Jacobian matrix, and its Jacobian determinant. You will easily find

$$\Delta_G(r, \theta, z) = r,$$

so that the general formula for changing variables gives you the same

result that was found in Chapter XII by looking at the volume of an elementary region, image of a box under the map G.

Example. Let G be the map of spherical coordinates, given by

$$G(\rho, \theta, \varphi) = (\rho \sin \varphi \cos \theta, \rho \sin \varphi \sin \theta, \cos \varphi).$$

Again you should compute the Jacobian matrix and the Jacobian determinant. You will find:

$$\Delta_G(\rho, \theta, \varphi) = \rho^2 \sin \varphi.$$

This gives a justification for the formula of Chapter XII in terms of the change of variables formula, which in the present case reads just like the result of Chapter XII, namely:

$$\boxed{\iiint_A f(G(\rho, \theta, \varphi))\rho^2 \sin \varphi \, d\rho \, d\varphi \, d\theta = \iiint_{G(A)} f(x, y, z) \, dz \, dy \, dx.}$$

Exercise. Carry out in detail the computation of the preceding two examples.

Exercises

1. (a) Let $G: \mathbf{R}^3 \to \mathbf{R}^3$ be the map which sends spherical coordinates (θ, φ, ρ) into cylindrical coordinates (θ, r, z). Write down the Jacobian matrix for this map, and its Jacobian determinant.
   (b) Write down the change of variables formula for this case.
2. Let A be a region in $\mathbf{R}^3$ and assume that its volume is equal to k. Let $G: \mathbf{R}^3 \to \mathbf{R}^3$ be the map such that $G(x, y, z) = (ax, by, cz)$, where a, b, c are positive numbers. What is the volume of $G(A)$?
3. Find the volume of the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leqq 1.$$

   by the change of variables formula, and by the method of dilations.
4. Find the volume of the solid which is the image of a ball of radius a under the linear map represented by the matrix

$$\begin{pmatrix} 1 & -1 & 1 \\ 0 & 2 & 5 \\ 0 & 0 & 7 \end{pmatrix}.$$

5. (a) Find the volume of the tetrahedron T determined by the inequalities

$$0 \leqq x, \quad 0 \leqq y, \quad 0 \leqq z \quad \text{and} \quad x + y + z \leqq 1.$$

(b) This tetrahedron can also be written in the form

$$t_1E_1 + t_2E_2 + t_3E_3 \quad \text{with} \quad t_1 + t_2 + t_3 \leqq 1, \quad 0 \leqq t_i.$$

If L is the linear map such that $L(E_i) = A_i$, show that $L(T)$ is described by similar inequalities. We call it the tetrahedron spanned by 0, A_1, A_2, A_3.

(c) Determine the volume of the tetrahedron spanned by the origin and the three vectors (1, 1, 2), (2, 0, −1), (3, 1, 2).

(d) Using the fact that the volume of a region does not change under translation, determine the volume of the tetrahedron spanned by the four points (1, 1, 1), (2, 2, 3), (3, 1, 0), and (4, 2, 3).

6. (a) Determine the volume of the tetrahedron spanned by the four points (2, 1, 0), (3, −1, 1), (−1, 1, 2), (0, 0, 1).

(b) Same question for the four points (3, 1, 2), (2, 0, 0), (4, 1, 5), (5, −1, 1).

CHAPTER XIV

Green's Theorem

§1. Statement of the theorem

In this chapter, we shall change slightly our notation concerning curve integrals.

Suppose we are given a vector field on some open set U in the plane. Then this vector field has two components, i.e. we can write

$$F(x, y) = \big(P(x, y), Q(x, y)\big),$$

where P, Q are functions of two variables (x, y). In everything that follows, we assume that all functions we deal with are C^1, i.e. that these functions have continuous partial derivatives.

Let $C\colon [a, b] \to U$ be a curve. We shall use a new notation for the integral of F over C, namely we write

$$\int_C F = \int_a^b F\big(C(t)\big) \cdot C'(t)\,dt = \int_C P(x, y)\,dx + Q(x, y)\,dy$$

or abbreviate this as

$$\int_C P\,dx + Q\,dy.$$

This is reasonable since the curve gives

$$x = x(t)$$

and

$$y = y(t)$$

as functions of t, and

$$F\big(C(t)\big) \cdot \frac{dC}{dt} = P(x, y)\frac{dx}{dt} + Q(x, y)\frac{dy}{dt}.$$

Green's Theorem. *Let P, Q be C^1-functions on a region A, which is the interior of a closed piecewise C^1-path C, parametrized counterclockwise. Then*

$$\int_C P\,dx + Q\,dy = \iint_A \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dy\,dx.$$

The region and its boundary may look as follows:

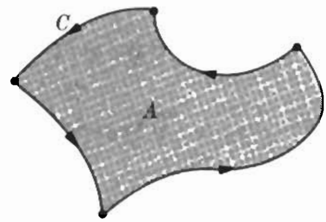


Figure 1

It is difficult to prove Green's theorem in general, partly because it is difficult to make rigorous the notion of "interior" of a path, and also the notion of counterclockwise. In practice, for any specifically given region, it is always easy, however. That it may be difficult in general is already suggested by drawing a somewhat less simple region as follows:

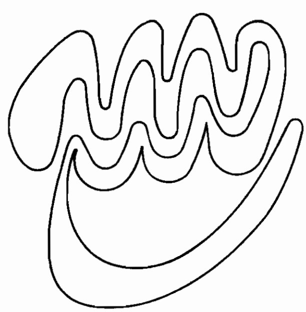


Figure 2

We shall therefore prove Green's theorem only in special cases, where we can give the region and the parametrization of its boundary explicitly.

Case 1. Suppose that the region A is given by the inequalities

$$a \leqq x \leqq b \qquad \text{and} \qquad g_1(x) \leqq y \leqq g_2(x)$$

in the same manner as we studied before in Chapter XII, §2.

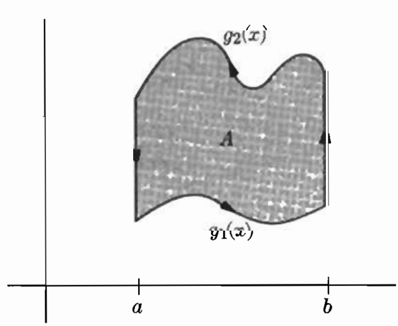


Figure 3

The boundary of A then consists of four pieces, the two vertical segments, and the pieces parametrized by the maps:

$$\gamma_1 : t \mapsto (t, g_1(t)), \qquad a \leqq t \leqq b,$$
$$\gamma_2 : t \mapsto (t, g_2(t)), \qquad a \leqq t \leqq b.$$

Then we can prove one-half of Green's theorem, namely

$$\int_C P\,dx = \iint_A -\frac{\partial P}{\partial y}\,dy\,dx.$$

Proof. We have

$$\begin{aligned}\iint_A \frac{\partial P}{\partial y}\,dy\,dx &= \int_a^b \int_{g_1(x)}^{g_2(x)} D_2P(x, y)\,dy\,dx \\ &= \int_a^b \left(P(x, y)\Big|_{g_1(x)}^{g_2(x)}\right) dx \\ &= \int_a^b [P(x, g_2(x)) - P(x, g_1(x))]\,dx \\ &= \int_{\gamma_2} P\,dx - \int_{\gamma_1} P\,dx.\end{aligned}$$

However, the boundary of A, oriented counterclockwise, consists of four pieces,

$$\gamma_1, \gamma_2^-, \gamma_3, \gamma_4,$$

where γ_2^- is the opposite curve to γ_2, and γ_3, γ_4 are the vertical segments. One sees at once that the integrals

$$\int_{\gamma_3} P\,dx \qquad \text{and} \qquad \int_{\gamma_4} P\,dx$$

are equal to 0, and thus we obtain the formula in this case.

Case 2. Suppose that the region is given by similar inequalities as in Case 1, but with respect to the y-axis. In other words, the region A is defined by inequalities

$$c \leqq y \leqq d \qquad \text{and} \qquad g_1(y) \leqq x \leqq g_2(y).$$

Then we prove the other half of Green's theorem, namely

$$\iint_A \frac{\partial Q}{\partial x}\,dy\,dx = \int_C Q\,dy.$$

Proof. We take the integral with respect to x first:

$$\iint_A \frac{\partial Q}{\partial x}\,dx\,dy = \int_c^d \left[\int_{g_1(y)}^{g_2(y)} D_1Q(x,y)\,dx\right]dy$$

$$= \int_c^d [Q(g_2(y), y) - Q(g_1(y), y)]\,dy.$$

In this case, the integral of $Q\,dy$ over the horizontal segments is equal to 0, and hence our formula is proved.

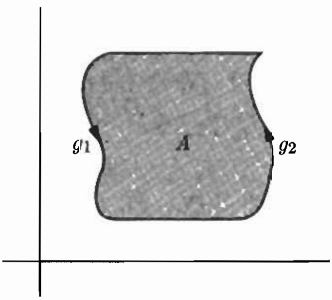


Figure 4

In particular, if a region is of a type satisfying both the preceding conditions, then the full theorem follows. Examples of such regions are rectangles and triangles and interiors of circles:

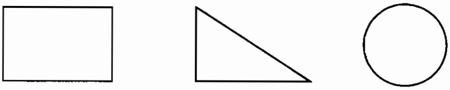


Figure 5

We have therefore proved Green's theorem in these cases.

Frequently, a region can be decomposed into regions of the preceding types. We draw a picture to illustrate this, namely the annulus lying between two circles.

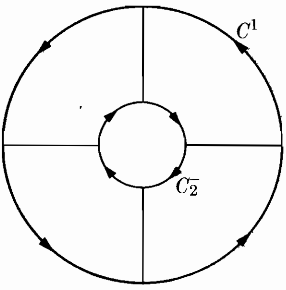


Figure 6

By drawing four line segments as shown, we decompose this annulus

into four regions, and it would thus suffice to prove Green's theorem for each one of these four regions. None of them yet satisfies the desired hypotheses, but one more decomposition will do for each region, as shown in the next picture.

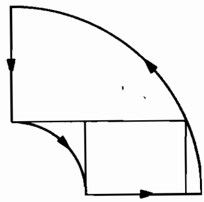


Figure 7

Consequently if we denote by C_1 the outside circle taken counterclockwise, and by C_2 the inside circle taken counterclockwise, if we let $C = \{C_1, C_2^-\}$, and if A is the region between C_1 and C_2, then

$$\iint_A \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dy\, dx = \int_C P\,dx + Q\,dy$$

$$= \int_{C_1} P\,dx + Q\,dy - \int_{C_2} P\,dx + Q\,dy.$$

Example 1. Let A be the region between two concentric circles C_1, C_2 as shown, both with counterclockwise orientation (Fig. 8).

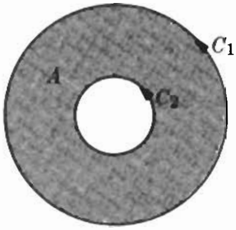


Figure 8

Let $F = (P, Q)$ be a vector field on A and suppose that

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}.$$

Then the left-hand side in the above relation is equal to 0, and consequently we see that the integral of F over C_1 is equal to the integral of F over C_2, in other words

$$\int_{C_1} P\,dx + Q\,dy = \int_{C_2} P\,dx + Q\,dy.$$

Of course, if F is the gradient of a function, then both these integrals are 0. However, we saw previously that there exist vector fields satis-

fying the condition $\partial P/\partial y = \partial Q/\partial x$, but not having potential functions, e.g.

$$F(x, y) = \left(\frac{-y}{x^2 + y^2}, \frac{x}{x^2 + y^2}\right).$$

Example 2. Find the integral of the vector field

$$F(x, y) = (y + 3x, 2y - x)$$

counterclockwise around the ellipse $4x^2 + y^2 = 4$.

Let $P(x, y) = y + 3x$ and $Q(x, y) = 2y - x$. Then $\partial Q/\partial x = -1$ and $\partial P/\partial y = -1$. By Green's theorem, we get

$$\int_C P\,dx + Q\,dy = \iint_A (-2)\,dy\,dx = -2 \text{ Area } (A),$$

where Area (A) is the area of the ellipse, which is known to be 2π ($= \pi ab$ when the ellipse is in the form $x^2/a^2 + y^2/b^2 = 1$).

Exercises

1. Use Green's theorem to find the integral $\int_C y^2\,dx + x\,dy$ when C is the following curve (taken counterclockwise).
   (a) The square with vertices (0, 0), (2, 0), (2, 2), (0, 2).
   (b) The square with vertices $(\pm 1, \pm 1)$.
   (c) The circle of radius 2 centered at the origin.
   (d) The circle of radius 1 centered at the origin.
   (e) The square with vertices $(\pm 2, 0)$, $(0, \pm 2)$.
   (f) The ellipse $x^2/a^2 + y^2/b^2 = 1$.
2. Let A be a region, which is the interior of a closed curve C oriented counterclockwise. Show that the area of A is given by

$$\text{Area } (A) = \tfrac{1}{2}\int_C -y\,dx + x\,dy = \int_C x\,dy.$$

3. Let C_1 be the closed path consisting of the vertical segment on the line $x = 2$, and the piece of the parabola

$$y^2 = 2(x + 2)$$

lying to the left of this segment, as shown on Fig. 9. We assume that C_1 is oriented counterclockwise. Find the integral

$$\int_{C_1} \frac{-y}{x^2 + y^2}\,dx + \frac{x}{x^2 + y^2}\,dy.$$

[*Hint:* Reduce this to an integral over the circle of radius 1.]

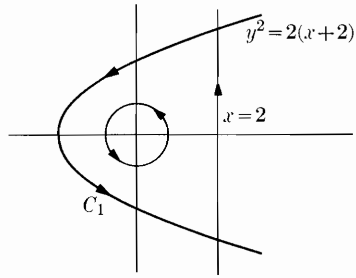


Figure 9

4. Assume that the function f satisfies Laplace's equation,

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0,$$

on a region A which is the interior of a curve C, oriented counterclockwise. Show that

$$\int_C \frac{\partial f}{\partial y} dx - \frac{\partial f}{\partial x} dy = 0.$$

5. If $F = (P, Q)$ is a vector field, we recall that its divergence is defined to be $\operatorname{div} F = \partial P/\partial x + \partial Q/\partial y$. Let

$$C(t) = (g_1(t), g_2(t)), \qquad a \leqq t \leqq b$$

be a closed curve. Define the right normal vector at t to be the vector

$$N(t) = (g_2'(t), -g_1'(t)).$$

Verify that this is a vector perpendicular to the curve. Show that if F is a vector field on a region A, which is the interior of the closed curve C, oriented counterclockwise, then

$$\iint_A (\operatorname{div} F)\, dy\, dx = \int_a^b F \cdot N\, dt$$

Note that $\|N(t)\| = \|C'(t)\| = v(t)$. Since $s(t) = \int v(t)\, dt$, the integral on the right is often expressed in terms of s. Let $\mathbf{n} = N/\|N\|$ be the unit vector in the direction of N. Then our formula reads:

$$\iint_A (\operatorname{div} F)\, dy\, dx = \int_C F \cdot \mathbf{n}\, ds.$$

6. Let $C: [a, b] \to U$ be a C^1-curve in an open set U of the plane. If f is a function on U (assumed to be differentiable as needed), we define

$$\int_C f = \int_a^b f(C(t))\|C'(t)\|\, dt$$

$$= \int_a^b f(C(t)) \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt.$$

For $r > 0$, let $x = r\cos\theta$ and $y = r\sin\theta$. Let φ be the function of r

defined by

$$\varphi(r) = \frac{1}{2\pi r}\int_{C_r} f = \frac{1}{2\pi r}\int_0^{2\pi} f(r\cos\theta, r\sin\theta)r\,d\theta.$$

where C_r is the circle of radius r, parametrized as above. Assume that f satisfies Laplace's equation

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0.$$

Show that $\varphi(r)$ does not depend on r and in fact

$$f(0, 0) = \frac{1}{2\pi r}\int_{C_r} f.$$

[*Hint:* First take $\varphi'(r)$ and differentiate under the integral, with respect to r. Let D_r be the disc of radius r which is the interior of C_r. Using Exercise 5, you will find that

$$\begin{aligned}\varphi'(r) &= \frac{1}{2\pi r}\iint_{D_r} \operatorname{div}\operatorname{grad} f(x, y)\,dy\,dx\\ &= \frac{1}{2\pi r}\iint_{D_r}\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right)dy\,dx\\ &= 0.\end{aligned}$$

Taking the limit as $r \to 0$, prove the desired assertion.]

§2. *Application to the change of variables formula*

When a region A is the interior of a closed path, then we can use Green's theorem to prove the change of variables formula in special cases. Indeed, Green's theorem reduces a double integral to an integral over a curve, and change of variables formulas for curves are easier to establish than for 2-dimensional areas. Thus we begin by looking at a special case of a change of variables formula for curves.

Let $C: [a, b] \to U$ be a C^1-curve in an open set of $\mathbf{R}^2$. Let $G: U \to \mathbf{R}^2$ be a C^2-map, given by coordinate functions,

$$G(u, v) = (x, y) = \big(f(u, v), g(u, v)\big).$$

Then the composite $G \circ C$ is a curve. If $C(t) = \big(\alpha(t), \beta(t)\big)$, then

$$G \circ C(t) = G\big(C(t)\big) = \big(f(\alpha(t), \beta(t)), g(\alpha(t), \beta(t))\big).$$

Example 1. Let $G(u, v) = (u, -v)$ be the reflection along the horizontal axis. If $C(t) = (\cos t, \sin t)$, then

$$G \circ C(t) = (\cos t, -\sin t).$$

Thus $G \circ C$ again parametrizes the circle, but observe that the orientation of $G \circ C$ is opposite to that of C, i.e. it is clockwise!

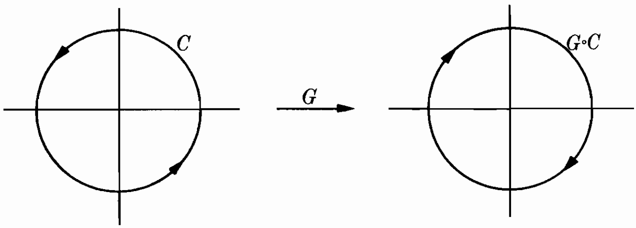


Figure 10

The reason for this reversal of orientation is that the Jacobian determinant of G is negative, namely it is the determinant of

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Thus a map G is said to **preserve orientation** if $\Delta_G(u, v) > 0$ for all (u, v) in the domain of definition of G. For simplicity, we only consider such maps G.

Green's theorem leads us to consider the integral

$$\int_{G \circ C} x \, dy.$$

By definition and the chain rule, we have

$$\begin{aligned} \int_{G \circ C} x \, dy &= \int_a^b f(C(t)) \left(\frac{\partial y}{\partial u} \frac{du}{dt} + \frac{\partial y}{\partial v} \frac{dv}{dt} \right) dt \\ &= \int_C f(u, v) \frac{\partial y}{\partial u} du + f(u, v) \frac{\partial y}{\partial v} dv. \end{aligned}$$

This is true for any C^1-curve as above. Hence it remains true for any piecewise C^1-path, consisting of a finite number of curves.

We are now ready to state and prove the change of variables formula in the case to which Green's theorem applies.

Let U be open in $\mathbf{R}^2$, and let A be a region which is the interior of a closed path C (piecewise C^1 as usual) contained in U. Let

$$G: U \to \mathbf{R}^2$$

be a C^2-map, which is C^1-invertible on U and such that $\Delta_G > 0$. Then

$G(A)$ is a region which is the interior of the path $G \circ C$. We then have

$$\boxed{\iint_{G(A)} dy\,dx = \iint_{A} \Delta_G(u, v)\,du\,dv.}$$

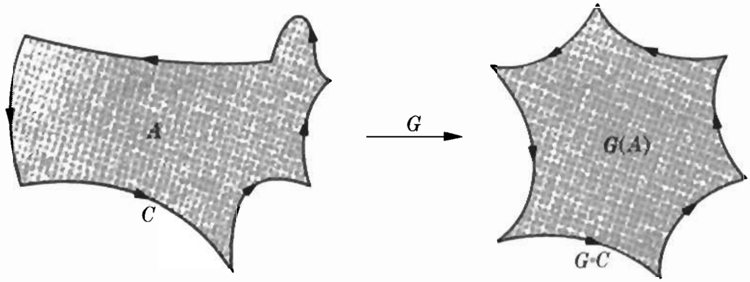


Figure 11

Proof. Let $G(u, v) = \big(f(u, v), g(u, v)\big)$ be expressed by its coordinates. We have, using Green's theorem:

$$\begin{aligned}
\iint_{G(A)} dy\,dx &= \int_{G\circ C} x\,dy = \int_C f\frac{\partial g}{\partial u}\,du + f\frac{\partial g}{\partial v}\,dv \\
&= \iint_A \left[\frac{\partial}{\partial u}\left(f\frac{\partial g}{\partial v}\right) - \frac{\partial}{\partial v}\left(f\frac{\partial g}{\partial u}\right)\right] du\,dv \\
&= \iint_A \left[\frac{\partial f}{\partial u}\frac{\partial g}{\partial v} + f\frac{\partial g}{\partial u\,\partial v} - f\frac{\partial g}{\partial u\,\partial v} - \frac{\partial f}{\partial v}\frac{\partial g}{\partial u}\right] du\,dv \\
&= \iint_A \left[\frac{\partial f}{\partial u}\frac{\partial g}{\partial v} - \frac{\partial g}{\partial u}\frac{\partial f}{\partial v}\right] du\,dv \\
&= \iint_A \Delta_G(u, v)\,du\,dv,
\end{aligned}$$

thus proving what we wanted.

Exercises

1. Under the same assumptions as the theorem in this section, assume that $\varphi = \varphi(x, y)$ is a continuous function on $G(A)$, and that we can write $\varphi(x, y) = \partial Q/\partial x$ for some continuous function Q. Prove the more general

formula

$$\iint_{G(A)} \varphi(x, y)\, dy\, dx = \iint_{A} \varphi\big(G(u, v)\big)\, \Delta_G(u, v)\, du\, dv.$$

[*Hint:* Let $P = 0$ and follow the same pattern of proof as in the text.]

2. Let $(x, y) = G(u, v)$ as in the text. We suppose that $G\colon U \to \mathbf{R}^2$, and that F is a vector field on $G(U)$. Then $F \circ G$ is a vector field on U. Let C be a C^1-curve in U. Show that

$$\int_{G \circ C} F = \int_C (F \circ G) \cdot \frac{\partial G}{\partial u}\, du + (F \circ G) \cdot \frac{\partial G}{\partial v}\, dv.$$

[Let $F(x, y) = (P(x, y), Q(x, y))$ and apply the definitions.]

CHAPTER XV

Surface Integrals

§1. *Parametrization, tangent plane, and normal vector*

Let us first recall that a curve can be described by an equation, like

$$x^2 + y^2 = 1,$$

or it can be given parametrically, as when we set

$$x = \cos\theta,$$
$$y = \sin\theta,$$

with $0 \leqq \theta \leqq 2\pi$. A similar situation will occur for surfaces, and we consider first the parametric representation.

Let R be a region in the plane, whose variables are denoted by (t, u). Let

$$X: R \to \mathbf{R}^3$$

be a mapping, which can be written in terms of its coordinate functions

$$X(t, u) = \big(x_1(t, u), x_2(t, u), x_3(t, u)\big),$$

where x_1, x_2, x_3 are functions from R into the real numbers. We say that such a mapping is C^1 if each coordinate function is differentiable, and if its partial derivatives are continuous. In this case, we may view X as parametrizing a surface in $\mathbf{R}^3$, as shown on Fig. 1.

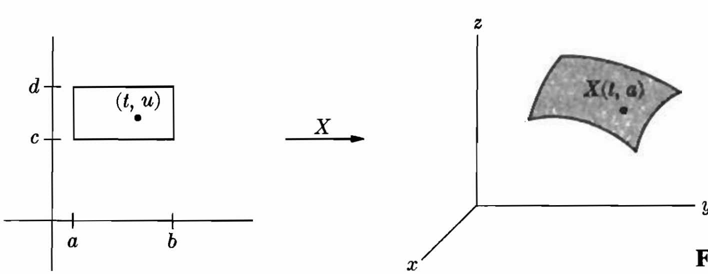


Figure 1

If x, y, z are the three coordinates of $\mathbf{R}^3$, then we also write the parametrization of our surface in the form

$$\begin{aligned} x &= f_1(u, v) \quad &\text{or}\quad & x(u, v), \\ y &= f_2(u, v) \quad &\text{or}\quad & y(u, v), \\ z &= f_3(u, v) \quad &\text{or}\quad & z(u, v). \end{aligned}$$

Example. We parametrize the sphere of radius ρ by means of spherical coordinates, as studied in Chapter XII, namely

$$\begin{aligned} x &= \rho \sin \varphi \cos \theta, \\ y &= \rho \sin \varphi \sin \theta, \\ z &= \rho \cos \varphi. \end{aligned}$$

The region R in $\mathbf{R}^2$ is the rectangle described by the inequalities

$$0 \leqq \varphi \leqq \pi$$

and

$$0 \leqq \theta < 2\pi.$$

Our mapping "wraps" this rectangle around the sphere. If we evaluate

$$x^2 + y^2 + z^2,$$

and use relations like $\sin^2 \theta + \cos^2 \theta = 1$, we get the value ρ^2. This kind of technique shows us how to get back the equation in rectangular coordinates from the parametrization.

Example. A torus (i.e. a doughnut-shaped surface) can be given parametrically by the functions:

$$\begin{aligned} x &= (a + b \cos \varphi) \cos \theta, \\ y &= (a + b \cos \varphi) \sin \theta, \\ z &= b \sin \varphi. \end{aligned}$$

The torus is centered at the origin, and $a > 0$ is the distance from the origin to the center of a cross section, as shown on Fig. 2. The variables φ, θ satisfy the inequalities

$$0 \leqq \varphi < 2\pi$$

and

$$0 \leqq \theta < 2\pi.$$

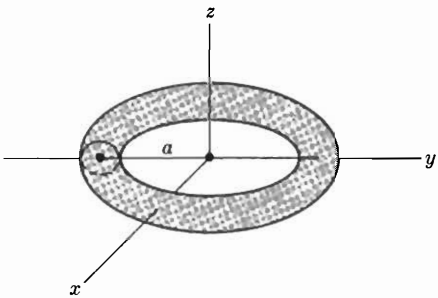


Figure 2

The number $b > 0$ is the radius of a cross section. The angle φ determines the rotation of a point in this cross section, as shown in Fig. 3.

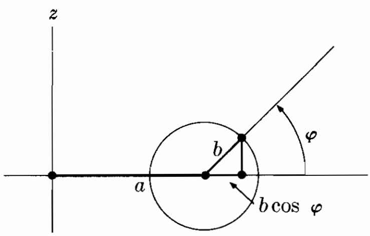


Figure 3

It is clear from this picture that the elevation z of a point is given by $b \sin \varphi$. If we project the point on the (x, y)-plane, then the distance of this projection from the origin is exactly

$$a + b \cos \varphi.$$

To get the x-coordinate of this projection, we have to multiply the projection with $\cos \theta$, and to get the y-coordinate of this projection, we have to multiply the projection with $\sin \theta$, as shown on Fig. 4.

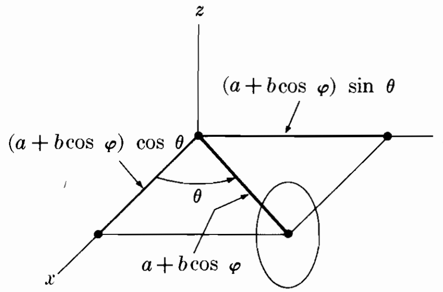


Figure 4

Let R be a region in $\mathbf{R}^2$, and let

$$X\colon R \to \mathbf{R}^3$$

be the parametrization of a surface. We assume that X is of class C^1. We have already studied the derivative of X at a point, and in Chapter XI, we defined it as the tangent linear map. If

$$X(t, u) = \begin{pmatrix} x_1(t, u) \\ x_2(t, u) \\ x_3(t, u) \end{pmatrix}$$

is represented by coordinates, then the derivative

$$X'(t, u)\colon \mathbf{R}^2 \to \mathbf{R}^3$$

is a linear map, represented by the Jacobian matrix

$$J_X(t, u) = \begin{pmatrix} \dfrac{\partial x_1}{\partial t} & \dfrac{\partial x_1}{\partial u} \\[2ex] \dfrac{\partial x_2}{\partial t} & \dfrac{\partial x_2}{\partial u} \\[2ex] \dfrac{\partial x_3}{\partial t} & \dfrac{\partial x_3}{\partial u} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial x}{\partial t} & \dfrac{\partial x}{\partial u} \\[2ex] \dfrac{\partial y}{\partial t} & \dfrac{\partial y}{\partial u} \\[2ex] \dfrac{\partial z}{\partial t} & \dfrac{\partial z}{\partial u} \end{pmatrix}.$$

For simplicity, we shall express ourselves as if this Jacobian matrix were actually a linear map. If we apply it to the two unit (vertical) vectors

$$E^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and

$$E^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

then we obtain two vectors A^1, A^2 which are nothing but

$$A^1 = J_X(t, u)E^1 = \frac{\partial X}{\partial t} \qquad \text{and} \qquad A^2 = J_X(t, u)E^2 = \frac{\partial X}{\partial u},$$

viewing $X(t, u)$ as a vertical vector. The picture is as follows.

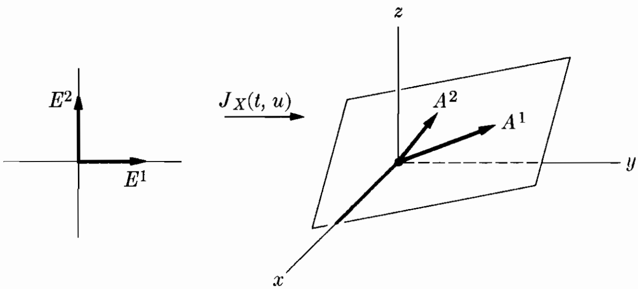


Figure 5

We shall say that (t, u) is a **regular** point for X if the two vectors A^1, A^2 span a plane in $\mathbf{R}^3$. The translation of this plane to the point $X(t, u)$ is called the **tangent plane** of the surface at the given point. This is illustrated on Fig. 6. It is the plane passing through the point $X(t, u)$, parallel to the vectors $A^1 = \partial X/\partial t$ and $A^2 = \partial X/\partial u$.

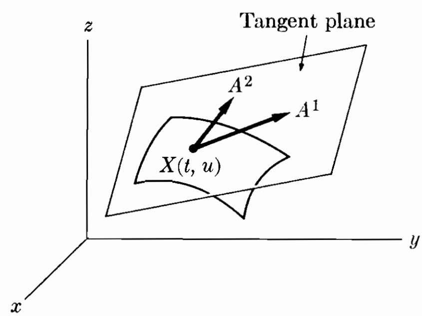


Figure 6

We now assume that you have read the section on the cross product in Chapter I. Then you realize that if A, B are non-zero vectors in $\mathbf{R}^3$, and are not parallel, their cross product

$$A \times B = (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1)$$

is perpendicular to both of them, as illustrated on Fig. 7.

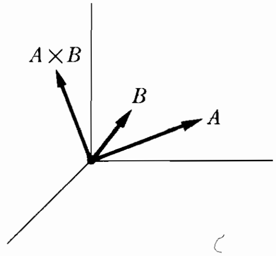


Figure 7

If we want a vector of norm 1 perpendicular to both A and B, all we have to do is divide $A \times B$ by its norm.

In the case of a parametrized surface, we can do this with the two vectors A^1 and A^2 as above. Of course, $B \times A = -A \times B$ is also perpendicular to both A and B. We use the notation

$$N = \frac{\partial X}{\partial t} \times \frac{\partial X}{\partial u}$$

whenever the surface is given parametrically by $(t, u) \mapsto X(t, u)$. Then $N = N(t, u)$ is a vector perpendicular to the surface, as shown on Fig. 8.

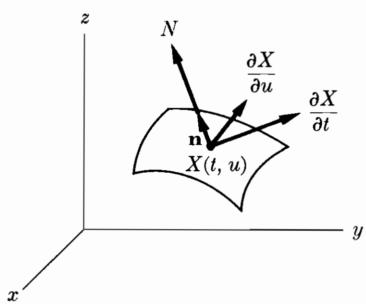


Figure 8

If we have chosen the orientation, i.e. the order of t, u, such that N points outwards from the surface, and if we denote by $\mathbf{n}$ the **outward unit normal vector** to the surface, then we have

$$\mathbf{n} = \frac{N}{\|N\|} = \frac{\dfrac{\partial X}{\partial t} \times \dfrac{\partial X}{\partial u}}{\left\|\dfrac{\partial X}{\partial t} \times \dfrac{\partial X}{\partial u}\right\|}.$$

Example. We compute the above quantities in the case of the parametrization of the sphere given above. We get very easily

$$\frac{\partial X}{\partial \varphi} = \begin{pmatrix} \rho \cos \varphi \cos \theta \\ \rho \cos \varphi \sin \theta \\ -\rho \sin \varphi \end{pmatrix} \quad \text{and} \quad \frac{\partial X}{\partial \theta} = \begin{pmatrix} -\rho \sin \varphi \sin \theta \\ \rho \sin \varphi \cos \theta \\ 0 \end{pmatrix}.$$

Hence

$$N(\varphi, \theta) = \frac{\partial X}{\partial \varphi} \times \frac{\partial X}{\partial \theta} = \begin{pmatrix} \rho^2 \sin^2 \varphi \cos \theta \\ \rho^2 \sin^2 \varphi \sin \theta \\ \rho^2 \sin \varphi \cos \varphi \end{pmatrix} = \rho \sin \varphi \; X(\varphi, \theta).$$

Since $\sin\varphi$ and ρ are $\geqq 0$, we see that N has the same direction as the position vector $X(\varphi, \theta)$, and therefore points outward. Taking the square root of the sum of the squares of the coordinates, we find

$$\left\| \frac{\partial X}{\partial \varphi} \times \frac{\partial X}{\partial \theta} \right\| = \rho^2 |\sin\varphi| = \rho^2 \sin\varphi.$$

Hence

$$\mathbf{n} = \frac{1}{\rho} X(\varphi, \theta).$$

Exercises

1. Compute the coordinates of the vectors $\partial X/\partial\theta$ and $\partial X/\partial\varphi$, when X is the mapping parametrizing the torus as in Example 2. Compute the norms of these vectors.

In each one of the following exercises, where you are given a parametrization $X(t, u)$, compute the tangent vectors $\dfrac{\partial X}{\partial t}, \dfrac{\partial X}{\partial u}$, their cross product, and the norm of this cross product. In each case, get an equation in cartesian coordinates for the surface parametrized by X. Draw the picture of the surface.

2. **The cone.** Let α be a fixed number, $0 < \alpha < \pi/2$. Let

$$X(\theta, z) = (z\sin\alpha\cos\theta, z\sin\alpha\sin\theta, z\cos\alpha),$$

$0 \leqq \theta < 2\pi$ and $0 \leqq z \leqq h$. Describe how you get a cone of height h.

3. **Paraboloid.** Let $X(t, \theta) = (at\cos\theta, at\sin\theta, t^2)$, with

$$0 \leqq \theta < 2\pi \quad \text{and} \quad 0 \leqq t \leqq h.$$

4. **Ellipsoid.** Let $a, b, c > 0$. Let

$$X(\varphi, \theta) = (a\sin\varphi\cos\theta, b\sin\varphi\sin\theta, c\cos\varphi).$$

5. **Cylinder.** Let $a > 0$. Let

$$X(\theta, z) = (a\cos\theta, a\sin\theta, z),$$

with $0 \leqq \varphi < 2\pi$, and $h_1 \leqq z \leqq h_2$.

6. **Surface of revolution.** (around the z-axis). Let f be a function of one variable r, defined for $r_1 \leqq r \leqq r_2$. Let $0 \leqq \varphi < 2\pi$, and let

$$X(r, \theta) = (r\cos\theta, r\sin\theta, f(r)).$$

§2. *Surface area*

Let A, B be non-zero vectors in $\mathbf{R}^3$, and assume that they are not parallel. Then they span a parallelogram, as shown on Fig. 9, and this parallelogram is contained in a plane.

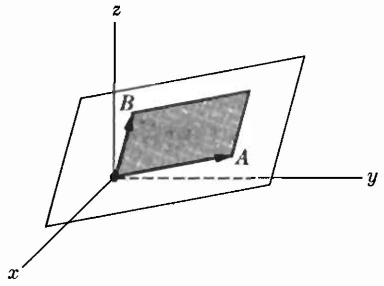


Figure 9

If θ is the angle between A and B, then the area of this parallelogram is precisely equal to

$$\|A\|\ \|B\|\ |\sin \theta|,$$

as one sees at once from Fig. 10, and as we already mentioned in Chapter I.

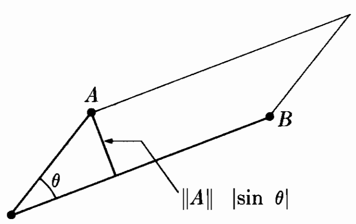


Figure 10

We observe that $\|A\|\ \|B\|\ |\sin \theta|$ is precisely the norm of $A \times B$. Thus in 3-space, we may say that the area of the parallelogram spanned by A and B is equal to

$$\|A \times B\|.$$

We apply this to surfaces. At each point, the tangent linear map $X'(t, u)$ of a parametrizing map $X(t, u)$ transforms the unit square spanned by E^1, E^2 into a parallelogram spanned by

$$\frac{\partial X}{\partial t} \quad \text{and} \quad \frac{\partial X}{\partial u}.$$

We can view this transformation as the local stretching effect on the area

of the square, and by the preceding remark, the area of this parallelogram is equal to

$$\left\| \frac{\partial X}{\partial t} \times \frac{\partial X}{\partial u} \right\| .$$

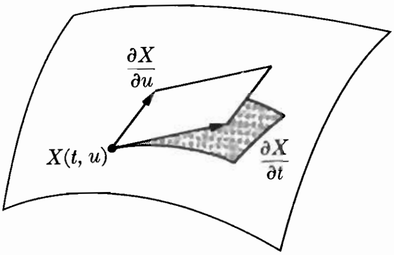


Figure 11

Assume that X is defined on a region R, and that the mapping

$$(t, u) \mapsto X(t, u)$$

is injective, except for a finite number of smooth curves in R. Also assume that the coordinate functions of f are C^1, and that all the points of R are regular, except for a finite number of smooth curves. It is then reasonable to define the **area of the parametrized surface** to be the integral

$$\boxed{\iint_S d\sigma = \iint_R \left\| \frac{\partial X}{\partial t} \times \frac{\partial X}{\partial u} \right\| dt\, du.}$$

Example. Let us compute the area of a sphere, whose parametrization was given in §1. We had already computed that

$$\left\| \frac{\partial X}{\partial \varphi} \times \frac{\partial X}{\partial \theta} \right\| = \rho^2 \sin \varphi.$$

Hence the area of the sphere is equal to

$$\int_0^{2\pi} \int_0^{\pi} \rho^2 \sin \varphi \, d\varphi \, d\theta.$$

Since ρ^2 is constant, we take it out of the integral. It is a trivial matter to carry out the integration, and we find that the desired area is equal to $4\pi\rho^2$.

Example. Sometimes a surface is given by the graph of a function

$$z = f(x, y),$$

defined over some region R of the (x, y)-plane. In this case, we use $t = x$ and $u = y$ as the parameters, so that

$$X(x, y) = \big(x, y, f(x, y)\big).$$

Thus the case when a surface is so defined is a special case of the general parametrization. In this special case, we find

$$\frac{\partial X}{\partial x} = \begin{pmatrix} 1 \\ 0 \\ \dfrac{\partial f}{\partial x} \end{pmatrix} \quad \text{and} \quad \frac{\partial X}{\partial y} = \begin{pmatrix} 0 \\ 1 \\ \dfrac{\partial f}{\partial y} \end{pmatrix}.$$

Consequently

$$\frac{\partial X}{\partial x} \times \frac{\partial X}{\partial y} = \left(-\frac{\partial f}{\partial x}, -\frac{\partial f}{\partial y}, 1\right),$$

and

$$\left\|\frac{\partial X}{\partial x} \times \frac{\partial X}{\partial y}\right\| = \sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}.$$

The **area of the surface** is given by the integral

$$\boxed{\iint_R \sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}\, dx\, dy.}$$

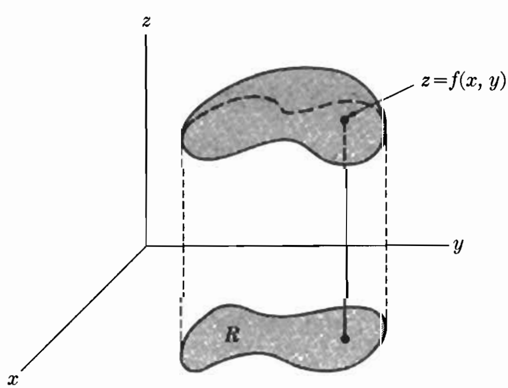


Figure 12

Symbolically we may write in this case

$$\boxed{d\sigma = \sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}\, dx\, dy.}$$

Example. It may also happens that a surface is defined implicitly by an equation

$$g(x, y, z) = 0,$$

and that over a certain region R of the (x, y)-plane, we can then solve for z by a function

$$z = f(x, y),$$

satisfying this equation, that is

$$g(x, y, f(x, y)) = 0.$$

Taking the partials with respect to x and y, we find the relations:

$$\frac{\partial f}{\partial x} = \frac{\partial g/\partial x}{\partial g/\partial z} \quad \text{and} \quad \frac{\partial f}{\partial y} = \frac{\partial g/\partial y}{\partial g/\partial z}.$$

We can now use the formula for the area obtained in the preceding example, and thus obtain a formula for the area just in terms of the given g, namely:

$$\boxed{\iint_R \frac{\sqrt{(\partial g/\partial x)^2 + (\partial g/\partial y)^2 + (\partial g/\partial z)^2}}{|\partial g/\partial z|}\, dx\, dy.}$$

Example. Take the special case of this formula arising from the equation of a sphere

$$x^2 + y^2 + z^2 - a^2 = 0,$$

where $a > 0$ is the radius. Then $g(x, y, z)$ is the expression on the left, and the partials are trivially computed:

$$\frac{\partial g}{\partial x} = 2x, \quad \frac{\partial g}{\partial y} = 2y, \quad \frac{\partial g}{\partial z} = 2z.$$

We can solve for z explicitly in terms of x, y by letting

$$z = \sqrt{a^2 - x^2 - y^2} = f(x, y),$$

where (x, y) ranges over the points in the disc of radius a in the plane. The surface is then the upper hemisphere.

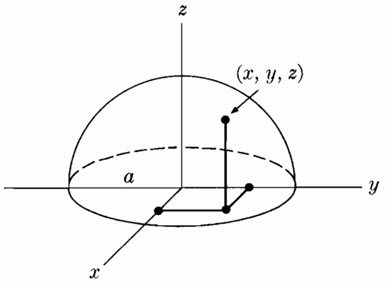


Figure 13

We can again compute the area of this hemisphere by the integral

$$\iint_R \frac{\sqrt{4x^2 + 4y^2 + 4z^2}}{|2z^2|}\,dx\,dy = \iint_R \frac{a}{z}\,dx\,dy,$$

using the fact that $x^2 + y^2 + z^2 = a^2$. Using polar coordinates, we know how to evaluate this last integral. We get

$$\text{Area of hemisphere} = a\int_0^{2\pi}\int_0^a \frac{1}{\sqrt{a^2 - r^2}}\,r\,dr\,d\theta.$$

Integrating 1 with respect to θ between 0 and 2π yields 2π. The integral with respect to r is reducible to the form

$$\int \frac{1}{\sqrt{u}}\,du,$$

and is therefore easily found. Thus, finally, we obtain the value

$$2\pi a^2$$

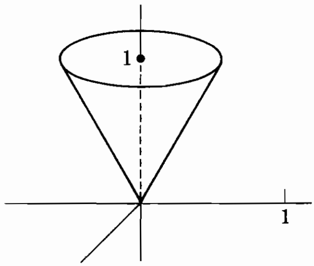
for the area of the hemisphere. Naturally, this jibes with the answer found from the parametrization by means of spherical coordinates.

Remark. Just as in the case of curves, it can be shown that the area of a surface is independent of the parametrization selected. This amounts to a change of variables in a 2-dimensional integral, but we shall omit the proof.

Exercises

Compute the following areas.

1. (a) A cone as shown on the following figure.



(b) The cone of height h obtained by rotating the line $z = 3x$ around the z-axis.

2. The surface $z = x^2 + y^2$ lying above the disc of radius 1 in the (x, y)-plane.
3. The surface $2z = 4 - x^2 - y^2$ over the disc of radius $\sqrt{2}$ in the (x, y)-plane.
4. $z = xy$ over the disc of radius 1.
5. The surface given parametrically by

$$X(t, \theta) = (t\cos\theta, t\sin\theta, \theta),$$

with $0 \leqq t \leqq 1$ and $0 \leqq \theta \leqq 2\pi$.

6. The surface given parametrically by

$$X(t, u) = (t + u, t - u, t),$$

with $0 \leqq t \leqq 1$ and $0 \leqq \theta \leqq 2\pi$. [*Hint:* Use $t = \sinh u = (e^u - e^{-u})/2$.]

7. The part of the sphere $x^2 + y^2 + z^2 = 1$ between the planes $z = 1/\sqrt{2}$ and $z = -1/\sqrt{2}$.
8. The part of the sphere $x^2 + y^2 + z^2 = 1$ inside the cone $x^2 + y^2 = z^2$.
9. The torus, using the parametrization in §1, assuming that the cross section has radius 1.

§3. *Surface integrals*

Let R be a region in the plane, and let

$$X\colon R \to \mathbf{R}^3$$

be the parametrization of a surface by a smooth mapping X. Let S be the image of X, i.e. the surface, and let ψ be a function on S. Then when ψ is sufficiently smooth, we define the **integral of ψ over** S by the formula

$$\iint_S \psi\, d\sigma = \iint_R \psi(X(t, u)) \left\| \frac{\partial X}{\partial t} \times \frac{\partial X}{\partial u} \right\| dt\, du.$$

When ψ is the constant 1, then our formula expresses simply the area of the parametrized surface.

Example. Suppose that ψ is the function representing a positive density of the surface. Then the integral above is interpreted as the mass m of the surface, corresponding to this density.

Example. Let ψ be a density as above, and m the mass. The integrals

$$\bar{x} = \frac{1}{m} \iint_S x\,\psi(x, y, z)\, d\sigma,$$

$$\bar{y} = \frac{1}{m} \iint_S y\,\psi(x, y, z)\, d\sigma,$$

$$\bar{z} = \frac{1}{m} \iint_S z\,\psi(x, y, z)\, d\sigma$$

give the coordinates $(\bar{x}, \bar{y}, \bar{z})$ of the center of mass of the surface. For instance, suppose that we want to find the center of mass of a hemisphere of radius a, having constant density c. We use the spherical coordinate parametrization of §1. The hemisphere is the one lying above the (x, y)-plane as in Fig. 14.

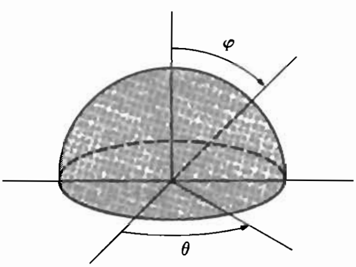


Figure 14

By symmetry, it is easy to see that $\bar{x} = \bar{y} = 0$. We have $z = a\cos\varphi$. The third coordinate $\bar{z}$ is given by the integral

$$\bar{z} = \frac{c}{m} \iint_S z\,d\sigma = \frac{c}{m} \int_0^{2\pi} \int_0^{\pi/2} a\cos\varphi \cdot a^2 \sin\varphi \, d\varphi \, d\theta,$$

which is easily evaluated to be $ca^3\pi/m$. The total mass is equal to the density times the area, since the density is constant, and we know that the area of the hemisphere is $2\pi a^2$. Hence we find

$$\bar{z} = a/2.$$

Example. Let $X: R \to \mathbf{R}^3$ parametrize a surface, and suppose that the image of X, that is the surface, is contained in some open set U in $\mathbf{R}^3$. Let F be a vector field on U, that is a mapping

$$F: U \to \mathbf{R}^3.$$

We assume that F is as smooth as needed. We define the **integral of the vector field along the surface** in a manner similar to the integral of a vector field along a curve in the lower dimensional case. Namely, let $\mathbf{n}$ be the outward normal unit vector to the surface, it being assumed that we have agreed on an orientation of the surface which determines its outside and inside. Then

$$F \cdot \mathbf{n}$$

is the projection of the vector field along the normal to the surface, and we define the above integral by the formula

$$\iint_S F \cdot \mathbf{n}\, d\sigma = \iint_R F \cdot \mathbf{n} \left\| \frac{\partial X}{\partial t} \times \frac{\partial X}{\partial u} \right\| dt\, du.$$

By definition, we have

$$\mathbf{n} \left\| \frac{\partial X}{\partial t} \times \frac{\partial X}{\partial u} \right\| = \frac{\partial X}{\partial t} \times \frac{\partial X}{\partial u} \cdot$$

Hence our integral for F over the surface can be rewritten

$$\boxed{\iint_S F \cdot \mathbf{n}\, d\sigma = \iint_R F\big(X(t, u)\big) \cdot \left(\frac{\partial X}{\partial t} \times \frac{\partial X}{\partial u}\right) dt\, du.}$$

An important physical example is given by a fluid flow, subject to a force field G, so that we may interpret G as a vector field,

$$G: U \to \mathbf{R}^3.$$

Let ψ be the function representing the density of the fluid, so that $\psi(x, y, z)$

is the density at a given point (x, y, z), and is a number. We call

$$F(x, y, z) = \psi(x, y, z)G(x, y, z)$$

the flux of the flow, and visualize it as in Fig. 15.

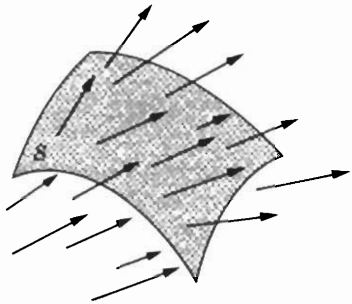


Figure 15

The amount of fluid passing through the surface per unit time is then given by the integral of the flux over the surface, namely

$$\iint_S F \cdot \mathbf{n}\, d\sigma,$$

where F is the flux.

It is not true that all surfaces can be oriented so that we can define an outside and an inside. The well-known Moebius strip gives an example when this cannot be done. In all the applications that we deal with, however, it is geometrically clear what is meant by the inside and outside. It is fairly difficult to give a definition in general, and so we don't go into this.

Observe that when we give a parametrization

$$(t, u) \mapsto X(t, u),$$

we could interchange the role of t, u as the first and second variable, respectively. Thus, for instance, if

$$X(t, u) = (t, u, t^2 + u^2),$$

we could let

$$Y(u, t) = (t, u, t^2 + u^2).$$

Then

$$\frac{\partial Y}{\partial u} \times \frac{\partial Y}{\partial t} = -\frac{\partial X}{\partial t} \times \frac{\partial X}{\partial u}.$$

Interchanging the variables amounts to changing the orientation. The two normal vectors corresponding to these two parametrizations have opposite direction. In finding the integral of a vector field with respect to a given parametrization, one must therefore agree on what is the "inside" and what is the "outside" of the surface, and check that the normal vector obtained from the cross product of the two partial derivatives points to the outside.

Example. Compute the integral of the vector field

$$F(x, y) = (x, y)$$

over the sphere $x^2 + y^2 + z^2 = a^2 \quad (a > 0)$. We use the parametrization of §1. Then

$$N(\varphi, \theta) = \frac{\partial X}{\partial \varphi} \times \frac{\partial X}{\partial \theta} = a \sin \varphi\, X(\varphi, \theta).$$

Thus $N(\varphi, \theta)$ is a positive multiple of the position vector X, and hence points outwards. So we get

$$F\big(X(\varphi, \theta)\big) \cdot N(\varphi, \theta) = (a \sin \varphi)[(a \sin \varphi \cos \theta)^2 + (a \sin \varphi \sin \theta)^2],$$

and

$$\iint_R F \cdot N \, d\varphi \, d\theta = a^3 \int_0^{2\pi} \int_0^{\pi} \sin^3 \varphi \, d\varphi \, d\theta = \frac{8\pi a^3}{3}.$$

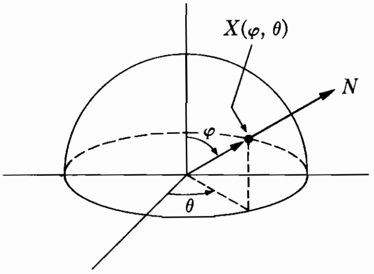


Figure 16

Example. Let S be the paraboloid defined by the equation

$$z = x^2 + y^2.$$

We can use x, y as parameters, and represent S parametrically by

$$X(x, y) = (x, y, x^2 + y^2).$$

Then

$$N(x, y) = (1, 0, 2x) \times (0, 1, 2y) = (-2x, -2y, 1).$$

Thus with the parametrization as given, we see from the picture that N points inside the paraboloid.

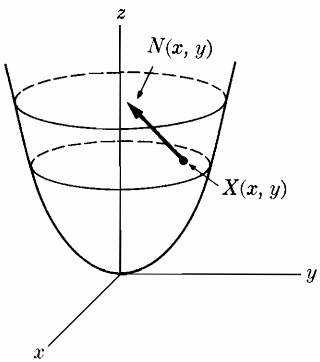


Figure 17

For instance, when x, y are positive, say equal to 1, then

$$N(1, 1) = (-2, -2, 1),$$

which points inward. Consequently, if we want the integral of a vector field F with respect to the outward orientation, then we have to take minus the integral $\iint F \cdot N \, dx \, dy$. To handle a concrete case, let

$$F(x, y, z) = (y, -x, z^2).$$

We want to compute the integral of F over the part of the paraboloid determined by the inequality

$$0 \leqq z \leqq 1.$$

We have

$$F(X(x, y)) \cdot N(x, y) = -2xy + 2xy + z^2 = z^2 = (x^2 + y^2)^2.$$

Hence

$$\iint_S F \cdot \mathbf{n} \, d\sigma = -\iint_R (x^2 + y^2)^2 \, dx \, dy,$$

where R is the unit disc in the (x, y)-plane. Changing to polar coordinates, it is easy to evaluate this integral, which is equal to

$$\int_0^{2\pi}\int_0^1 r^4\, r\, dr\, d\theta = \frac{\pi}{3}\cdot$$

The desired integral is therefore equal to $-\pi/3$. Note that in the present case, we have

$$\mathbf{n} = -\frac{N}{\|N\|}\cdot$$

Exercises

Integrate the following function over the indicated surface.

1. The function $x^2 + y^2$ over the same upper hemisphere as in the example in the text.
2. The function $(x^2 + y^2)z$ over this same hemisphere.
3. The function $(x^2 + y^2)z^2$ over this same hemisphere.
4. The function $z(x^2 + y^2)^2$ over this same hemisphere.
5. The function z over the surface

$$z = 1 - x^2 - y^2, \qquad z \geqq 0.$$

   (Use polar coordinates and sketch the surface.)
6. The function x over the cone $x^2 + y^2 = z^2$, $0 \leqq z \leqq a$.
7. The function x over the part of the sphere $x^2 + y^2 + z^2 = a^2$ contained inside the cone of Exercise 6.
8. The function x^2 over the cylinder defined by $x^2 + y^2 = a^2$, and $0 \leqq z \leqq 1$, excluding its top and bottom.
9. The same function x^2 over the top and bottom of the cylinder.
10. ***Theorem of Pappus.*** Let $C\colon [a, b] \to \mathbf{R}^2$ be the parametrization of a smooth curve, say

$$C(t) = (f(t), z(t)),$$

   which we view as lying in the (x, z)-plane, as shown on Fig. 18. We assume that $f(t) \geqq 0$. Let $\bar{x}$ be the x-coordinate of the center of mass of this curve in the (x, z)-plane. Prove that the area of the surface of revolution of this curve is equal to

$$2\pi\bar{x}L,$$

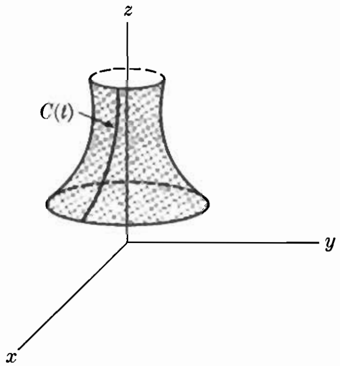


Figure 18

where L is the length of the curve. *Hint*: Parametrize the surface of revolution by the mapping

$$X(t, \theta) = (f(t)\cos\theta,\ f(t)\sin\theta,\ z(t)).$$

What is θ in Fig. 18? Recall that $\bar{x}$ is given by

$$\bar{x} = \frac{1}{L}\int_a^b f(t)\|C'(t)\|\,dt.$$

How does this apply to get the area of torus in a simple way?

11. Let S be the center of a sphere of radius a and centered at O. Let P be a fixed point, either inside or outside the sphere, but not on S. Let

$$r(X) = \|X - P\|.$$

Show that

$$\iint_S \frac{1}{r}\,d\sigma = \begin{cases} 4\pi a & \text{if } P \text{ is inside the sphere} \\ \dfrac{4\pi a^2}{\|P\|} & \text{if } P \text{ is outside the sphere.} \end{cases}$$

Find the integrals of the following vector fields over the given surfaces.

12. $F(x, y, z) = \dfrac{1}{\sqrt{x^2 + y^2}}(y, -y, 1)$ over the paraboloid

$$z = 1 - x^2 - y^2, \qquad 0 \leqq z \leqq 1.$$

(Draw the picture.)

13. The same vector field as in Exercise 12, over the lower hemisphere of a sphere centered at the origin, of radius 1.

14. The vector field $F(x, y, z) = (y, -x, 1)$ over the surface

$$X(t, \theta) = (t \cos \theta, t \sin \theta, \theta),$$

$0 \leqq t \leqq 1$ and $0 \leqq \theta \leqq 2\pi$.

15. The vector field $F(x, y, z) = (x^2, y^2, z^2)$ over the surface

$$X(t, u) = (t + u, t - u, t),$$

$0 \leqq t \leqq 2$ and $1 \leqq u \leqq 3$.

16. The vector field $F(X) = X$, over the part of the sphere $x^2 + y^2 + z^2 = 1$ between the planes $z = 1/\sqrt{2}$ and $z = -1/\sqrt{2}$.
17. The vector field $F(x, y, z) = (x, 0, 0)$ over the part of the unit sphere inside the cone $x^2 + y^2 = z^2$.
18. The vector field $F(x, y, z) = (x, y^2, z)$ over the triangle determined by the plane $x + y + z = 1$, and the coordinate planes.
19. The vector field $F(x, y, z) = (x, y, z^2)$ over the cylinder defined by $x^2 + y^2 = a^2, 0 \leqq z \leqq 1$,
    (a) excluding the top and bottom,
    (b) including the top and bottom.
20. The vector field $F(x, y, z) = (xy, y^2, y^3)$ over the boundary of the unit cube

$$0 \leqq x \leqq 1, \qquad 0 \leqq y \leqq 1, \qquad 0 \leqq z \leqq 1.$$

21. The vector field $F(x, y, z) = (xz, 0, 1)$ over the upper hemisphere of radius 1.

§4. *Curl and divergence of a vector field*

Let U be an open set in $\mathbf{R}^3$, and let

$$F: U \to \mathbf{R}^3$$

be a vector field. Thus F associates a vector to each point of U, and F is given by three coordinate functions,

$$F(x, y, z) = \big(f_1(X), f_2(X), f_3(X)\big).$$

We assume that F is as differentiable as needed, usually of class C^1 suffices, i.e. each coordinate function is differentiable and has continuous partial derivatives.

We define the **divergence** of F to be the *function*

$$\operatorname{div} F = \frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial y} + \frac{\partial f_3}{\partial z}.$$

Thus the divergence is the sum of the partial derivatives of the coordinate functions, taken with respect to the corresponding variables.

Example. Let $F(x, y, z) = (\sin xy, e^{xz}, 2x + yz^4)$. Then

$$\begin{aligned}(\operatorname{div} F)(x, y, z) &= y\cos xy + 0 + 4yz^3\\ &= y\cos xy + 4yz^3.\end{aligned}$$

As a matter of notation, one sometimes writes symbolically

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) = (D_1, D_2, D_3),$$

where D_1, D_2, D_3 are the partial derivative operators with respect to the corresponding variables. Then one also writes

$$\operatorname{div} F = \nabla \cdot F = D_1 f_1 + D_2 f_2 + D_3 f_3.$$

We shall interpret the divergence geometrically later. Similarly, we now define the **curl** of F, and interpret it geometrically later. We define

$$\begin{aligned}\operatorname{curl} F &= \left(\frac{\partial f_3}{\partial y} - \frac{\partial f_2}{\partial z}, \frac{\partial f_1}{\partial z} - \frac{\partial f_3}{\partial x}, \frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}\right)\\ &= (D_2 f_3 - D_3 f_2, D_3 f_1 - D_1 f_3, D_1 f_2 - D_2 f_1).\end{aligned}$$

The curl of F is therefore also a *vector* field.

Again, we use the symbolic notation

$$\operatorname{curl} F = \nabla \times F = \begin{vmatrix} E_1 & E_2 & E_3 \\ D_1 & D_2 & D_3 \\ f_1 & f_2 & f_3 \end{vmatrix},$$

where E_1, E_2, E_3 are the standard unit vectors. The "determinant" on the right is to be interpreted symbolically, using an expansion according to the first row. For instance, the first term in such an expansion is obtained by taking E_1 and "multiplying" it by the "determinant"

$$\begin{vmatrix} D_2 & D_3 \\ f_2 & f_3 \end{vmatrix} = D_2 f_3 - D_3 f_2.$$

This means that the first component of curl F is $D_2 f_3 - D_3 f_2$. The other components are obtained by a similar formal operation on the "determinant", with respect to the second and third components of the first row.

Example. Let F be the same vector field as in the preceding example. Then

$$\operatorname{curl} F = \begin{vmatrix} E_1 & E_2 & E_3 \\ D_1 & D_2 & D_3 \\ \sin xy & e^{xy} & 2x + yz^4 \end{vmatrix}$$

$$= (z^4 - 0, 2 - 0, ye^{xy} - x \sin xy)$$

$$= (z^4, 2, ye^{xy} - x \sin xy).$$

Exercises

Compute the divergence and the curl of the following vector fields.

1. $F(x, y, z) = (x^2, xyz, yz^2)$
2. $F(x, y, z) = (y \log x, x \log y, xy \log z)$
3. $F(x, y, z) = (x^2, \sin xy, e^x yz)$
4. $F(x, y, z) = (e^{xy} \sin z, e^{xz} \sin y, e^{yz} \cos x)$
5. Let φ be a smooth function. Prove that curl grad $\varphi = 0$.
6. Prove that div curl $F = 0$.
7. Let $\nabla^2 = \nabla \cdot \nabla = D_1^2 + D_2^2 + D_3^2 = \left(\frac{\partial}{\partial x}\right)^2 + \left(\frac{\partial}{\partial y}\right)^2 + \left(\frac{\partial}{\partial z}\right)^2$. A function f is said to be **harmonic** if $\nabla^2 f = 0$. Prove that the following functions are harmonic.

   (a) $\dfrac{1}{\sqrt{x^2 + y^2 + z^2}}$ (b) $x^2 - y^2 + 2z$

   (c) If f is harmonic, prove that div grad $f = 0$.
8. Let $F(X) = c\dfrac{X}{\|X\|^3}$, where c is constant. Prove that div $F = 0$ and that curl $F = 0$.
9. Prove that div $(F \times G) = G \cdot \operatorname{curl} F - F \cdot \operatorname{curl} G$, if F, G are vector fields.
10. Prove that div (grad $f \times$ grad g) $= 0$, if f, g are functions.

§5. *Divergence theorem*

In this section, we let U be a 3-dimensional region in $\mathbf{R}^3$, whose boundary is a closed surface which is smooth, except for a finite number of smooth curves. For instance, a 3-dimensional rectangular box is such a region. The inside of a sphere, or of an ellipsoid is such a region. The region

bounded by the plane $z = 2$, and inside the paraboloid $z = x^2 + y^2$ is such a region, illustrated in Fig. 19.

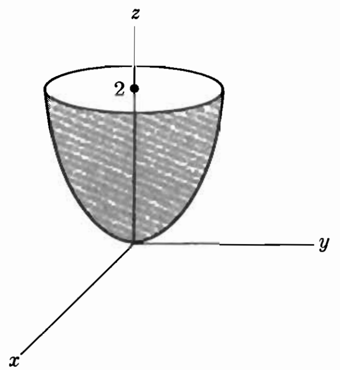


Figure 19

Note that the boundary consists of two pieces, the surface of the paraboloid and the disc on top, each of which can be easily parametrized.

Divergence Theorem. *Let U be a region in 3-space, forming the inside of a surface S which is smooth, except for a finite number of smooth curves. Let F be a C^1 vector field on an open set containing U and S. Let* $\mathbf{n}$ *be the unit outward normal vector to S. Then*

$$\iint_S F \cdot \mathbf{n}\, d\sigma = \iiint_U \operatorname{div} F\, dV,$$

where the expression on the right is simply the triple integral of the function div F *over the region U.*

It is not easy to give a proof of the divergence theorem in general, but we shall give it in a special case of a rectangular box. This makes the general case very plausible, because we could reduce the general case to the special case by the following steps:

(i) Analyze how surface integrals change (or rather do not change) when we change the variables.

(ii) Reduce the theorem to a "local one" where the region admits one parametrization from a rectangular box. This can be done by various chopping-up processes, some of which are messy, some of which are neat, but all of which take up a fair amount of space to establish fully.

(iii) Combine the first and second steps, reducing the local theorem concerning the region to the theorem concerning a box, by means of the change of variables formula.

We now prove the theorem for a box, expressed as a product of intervals:

$$[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3],$$

and illustrated in Fig. 20.

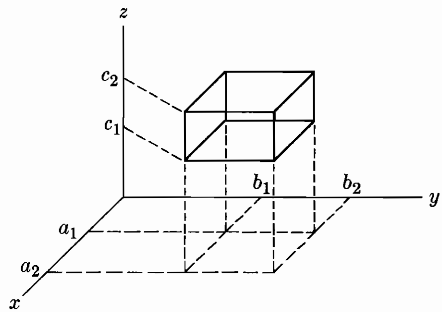


Figure 20

The surface surrounding the box consists of six sides, so that the integral over S will be a sum of six integrals, each one taken over one of the sides.

Let S_1 be the front face. We can parametrize S_1 by

$$X(y, z) = (a_2, y, z),$$

with y, z satisfying the inequalities

$$b_1 \leqq y \leqq b_2 \qquad \text{and} \qquad c_1 \leqq z \leqq c_2.$$

Let $\mathbf{n}_1$ be the unit outward normal vector on S_1. Then

$$\mathbf{n}_1 = (1, 0, 0).$$

If $F = (f_1, f_2, f_3)$, then $F \cdot \mathbf{n}_1 = f_1$, and hence

$$\iint_{S_1} F \cdot \mathbf{n}\, d\sigma = \int_{c_1}^{c_2} \int_{b_1}^{b_2} f_1(a_2, y, z)\, dy\, dz.$$

Similarly, let S_2 be the back face, parametrized by

$$X(y, z) = (a_1, y, z),$$

with y, z satisfying the same inequalities as above. Then

$$\mathbf{n}_2 = -(1, 0, 0),$$

the geometric interpretation being that the outward unit normal vector

points to the back of the box drawn on Fig. 20. Hence

$$\iint_{S_2} F \cdot \mathbf{n}\, d\sigma = \int_{c_1}^{c_2} \int_{b_1}^{b_1} -f_1(a_1, y, z)\, dy\, dz.$$

Adding the integrals over S_1 and S_2 yields

$$\iint_{S_1} + \iint_{S_2} F \cdot \mathbf{n}\, d\sigma = \int_{c_1}^{c_2} \int_{b_1}^{b_2} [f_1(a_2, y, z) - f_1(a_1, y, z)]\, dy\, dz$$

$$= \int_{c_1}^{c_2} \int_{b_1}^{b_2} \int_{a_1}^{a_2} D_1 f_1(x, y, z)\, dx\, dy\, dz$$

$$= \iiint_U D_1 f_1\, dV.$$

We now carry out a similar argument for the right side and the left side, as well as the top side and the bottom side. We find that the sums of the surface integral taken over these pairs of sides are equal to

$$\iiint_U D_2 f_2\, dV$$

and

$$\iiint_U D_3 f_3\, dV,$$

respectively. Adding all three volume integrals yields

$$\iint_S F \cdot \mathbf{n}\, d\sigma = \iiint_U (D_1 f_1 + D_2 f_2 + D_3 f_3)\, dV,$$

which is precisely the integral of the divergence, thus proving what we wanted.

Example. Let us compute the integral of the vector field

$$F(x, y, z) = (x^2, y^2, z^2)$$

over the unit cube by using the divergence theorem. The divergence of F is equal to $2x + 2y + 2z$, and hence the integral is equal to

$$\int_0^1 \int_0^1 \int_0^1 (2x + 2y + 2z)\, dx\, dy\, dz,$$

which is easily evaluated to give the value 3.

Example. Let us compute the integral of the vector field

$$F(x, y, z) = (x, y, z),$$

that is $F(X) = X$ over the sphere of radius a. The divergence of F is equal to

$$\frac{\partial x}{\partial x} + \frac{\partial y}{\partial y} + \frac{\partial z}{\partial z} = 3.$$

The ball B is the inside of the sphere. By the divergence theorem, we get

$$\iint_S F \cdot \mathbf{n}\, d\sigma = \iiint_B 3\, dV = 3 \cdot \tfrac{4}{3}\pi a^3 = 4\pi a^3.$$

Note that the volume integral over the ball B of radius a is the integral of the constant 3, and hence is equal to 3 times the volume of the ball.

The divergence theorem has an interesting application, which can be used to interpret the divergence geometrically.

Corollary. *Let $B(t)$ be the solid ball of radius $t > 0$, centered at a point P in $\mathbf{R}^3$. Let $S(t)$ denote the boundary of the ball,* i.e. *the sphere of radius t, centered at P. Let F be a C^1 vector field, and let $V(t)$ denote the volume of $B(t)$. Let $\mathbf{n}$ denote the unit normal vector pointing out from the spheres. Then*

$$(\operatorname{div} F)(P) = \lim_{t \to 0} \frac{1}{V(t)} \iint_{S(t)} F \cdot \mathbf{n}\, d\sigma.$$

Proof. Let $g = \operatorname{div} F$. Since g is continuous by assumption, we can write

$$g(X) = g(P) + h(X),$$

where

$$\lim_{X \to P} h(X) = 0.$$

Using the divergence theorem, we get

$$\begin{aligned} \frac{1}{V(t)} \iint_{S(t)} F \cdot \mathbf{n}\, d\sigma &= \frac{1}{V(t)} \iiint_{B(t)} \operatorname{div} F\, dV \\ &= \frac{1}{V(t)} \iiint_{B(t)} g(P)\, dV + \frac{1}{V(t)} \cdot \iiint_{B(t)} h\, dV. \end{aligned}$$

Observe that $g(P) = (\operatorname{div} F)(P)$ is constant, and hence can be taken out of the first integral. The simple integral of dV over $B(t)$ yields the volume $V(t)$, which cancels, so that the first term is equal to $(\operatorname{div} F)(P)$, which is the desired answer.

There remains to show that the second term approaches 0 as t approaches 0. But this is clear: The function h approaches 0, and the integral on the right can be estimated as follows:

$$\left|\frac{1}{V(t)} \iiint_{B(t)} h\, dV\right| \leqq \operatorname*{Max}_{\|X-P\| \leqq t} |h(X)| \frac{1}{V(t)} \iiint_{B(t)} dV$$
$$\leqq \operatorname*{Max}_{\|X-P\| \leqq t} |h(X)|.$$

As $t \to 0$, the maximum of $h(X)$ for $\|X - P\| \leqq t$ approaches 0, thus proving what we wanted.

The integral expression under the limit sign in the corollary can be interpreted as the flow going outside the sphere per unit time, in the direction of the unit outward normal vector. Dividing by the volume of the ball $B(t)$, we obtain the mass per unit volume flowing out of the sphere. Thus we get an interpretation for the divergence of F at P as the rate of change of mass per unit volume per unit time at P.

Exercises

1. Compute explicitly the integrals over the top, bottom, right, and left sides of the box to check in detail the remaining steps of the proof of the divergence theorem, left to the reader in the text, as "similar arguments".

2. Let S be the boundary of the unit cube,

$$0 \leqq x \leqq 1, \qquad 0 \leqq y \leqq 1, \qquad 0 \leqq z \leqq 1.$$

   Compute the integral of the vector field $F(x, y, z) = (xy, y^2, y^2)$ over the surface of this cube.

3. Calculate the integral

$$\iint_S (\operatorname{curl} F) \cdot \mathbf{n}\, d\sigma$$

   where F is the vector field

$$F(x, y, z) = (-y, x^2, z^3),$$

   and S is the surface

$$x^2 + y^2 + z^2 = 1, \qquad -\tfrac{1}{2} \leqq z \leqq 1.$$

   Don't make things more complicated than they need be.

4. Find the integral of the vector field

$$F(X) = \frac{X}{\|X\|}$$

   over the sphere of radius 4.

Find the integral of the following vector fields over the indicated surface.

5. $F(x, y, z) = (yz, xz, xy)$ over the cube centered at the origin and with sides of length 2.
6. $F(x, y, z) = (x^2, y^2, z^2)$ over the same cube.
7. $F(x, y, z) = (x - y, y - z, x - y)$ over the same cube.
8. $F(X) = X$ over the same cube.
9. $F(x, y, z) = (x + y, y + z, x + z)$ over the surface bounded by the paraboloid
$$z = 4 - x^2 - y^2,$$
and the disc of radius 2 centered at the origin in the (x, y)-plane.
10. $F(x, y, z) = (2x, 3y, z)$ over the surface bounding the region enclosed by the cylinder
$$x^2 + y^2 = 4$$
and the planes $z = 1$ and $z = 3$.
11. $F(x, y, z) = (x, y, z)$, over the surface bounding the region enclosed by the paraboloid $z = x^2 + y^2$, the cylinder $x^2 + y^2 = 9$, and the plane $z = 0$.
12. $F(x, y, z) = (x + y, y + z, x + z)$ over the surface bounding the region defined by the inequalities
$$0 \leqq x^2 + y^2 \leqq 9 \quad \text{and} \quad 0 \leqq z \leqq 5.$$
13. $F(x, y, z) = (3x^2, xy, z)$ over the tetrahedron bounded by the coordinate planes and the plane $x + y + z = 1$.
14. Let f be a harmonic function, that is a function satisfying
$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = 0.$$
Let S be a closed smooth surface bounding a region U in 3-space. Let f be a harmonic function on an open set containing the region and its boundary. If $\mathbf{n}$ is the unit normal vector to the surface pointing outward, let $D_{\mathbf{n}}f$ be the directional derivative of f in the direction of $\mathbf{n}$. Prove that
$$\iint_S D_{\mathbf{n}} f \, d\sigma = 0.$$
[*Hint:* Let $F = \operatorname{grad} f$.]
15. Assumptions being as in Exercise 14, prove that
$$\iint_S f D_{\mathbf{n}} f \, d\sigma = \iiint_U \|\operatorname{grad} f\|^2 \, dV.$$
[*Hint:* Let $F = f \operatorname{grad} f$.]

§6. *Stokes' theorem*

We recall Green's theorem in the plane. It stated that if S is a plane region bounded by a closed path C, oriented counterclockwise, and F is a vector field on some open set containing the region, $F = (f_1, f_2)$, then

$$\iint_S (D_1 f_2 - D_2 f_1)\, d\sigma = \int_C F \cdot dC.$$

Of course in the plane with variables (x, y), $d\sigma = dx\, dy$.

We can now ask for a similar theorem in 3-space, when the surface lies in 3-space, and the surface is bounded by a curve in 3-space. The analogous statement is true, and is called Stokes' theorem:

Stokes' Theorem. *Let S be a smooth surface in $\mathbf{R}^3$, bounded by a closed curve C. Assume that the surface is orientable, and that the boundary curve is oriented so that the surface lies to the left of the curve. Let F be a C^1 vector field in an open set containing the surface S and its boundary. Then*

$$\iint_S (\operatorname{curl} F) \cdot \mathbf{n}\, d\sigma = \int_C F \cdot dC.$$

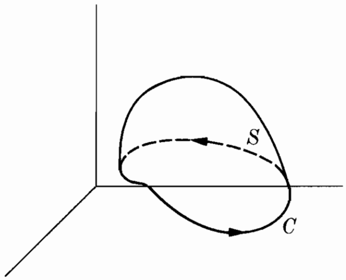


Figure 21

When the surface consists of a finite number of smooth pieces, and the boundary also consists of a finite number of smooth curves, then the analogous statement holds, by taking a sum over these pieces.

We shall not prove Stokes' theorem. The proof can be reduced to that of Green's theorem in the plane by making an analysis of the way both sides of the formula behave under changes of variables, i.e. changes of parametrization. Note that Green's theorem in the plane is a special case, because then the unit normal vector is simply $(0, 0, 1)$, and the curl of F dotted with the unit normal vector is simply the third component of the curl, namely

$$D_1 f_2 - D_2 f_1.$$

Thus Green's theorem in the plane makes the 3-dimensional analogue quite plausible.

Stokes' theorem has an interesting consequence as follows. Suppose that two surfaces S_1 and S_2 are bounded by a curve C, and lie on opposite sides of the curve, as on Fig. 22. Then

$$\iint_{S_1} (\operatorname{curl} F) \cdot \mathbf{n}\, d\sigma = -\iint_{S_2} (\operatorname{curl} F) \cdot \mathbf{n}\, dr$$

because these integrals are equal to the integrals of F over the boundary curve with opposite orientations. We have also drawn separately the surfaces S_1 and S_2 having C as boundary. Observe that taken together, S_1 and S_2 bound the inside of a 3-dimensional region.

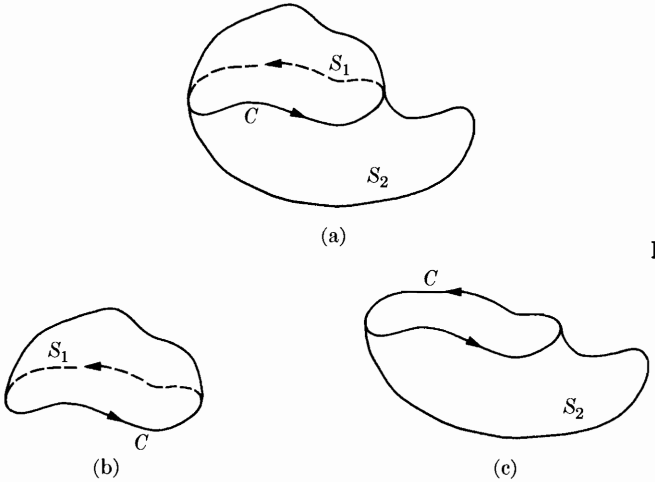


Figure 22

Similarly, we can consider a ball, bounded by a sphere. The two hemispheres have a common boundary, namely the circle in the plane as on Fig. 23. Note that C is oriented so that S_1 lies to the left of C, but S_2 lies to the right of C.

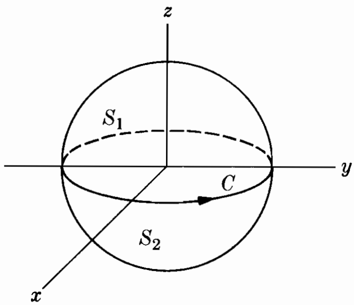


Figure 23

By the divergence theorem, we know that if S denotes the union of S_1 and S_2, then

$$\iint_S (\operatorname{curl} F) \cdot \mathbf{n}\, d\sigma = \iiint_U \operatorname{div} \operatorname{curl} F\, dV.$$

However, div curl $F = 0$. Hence the integral above is equal to 0. This corresponds to the fact that

$$\iint_{S_1} (\operatorname{curl} F) \cdot \mathbf{n}\, d\sigma = -\iint_{S_2} (\operatorname{curl} F) \cdot \mathbf{n}\, d\sigma$$

because the integral over S_1 is equal to the integral of F over C, whereas the integral over S_2 is equal to the integral of F over C^-, which is the same as C but oriented in the opposite direction.

Example. We shall verify Stokes' theorem for the vector field

$$F(x, y, z) = \big(z - y,\ x + z,\ -(x + y)\big),$$

and the surface bounded by the paraboloid

$$z = 4 - x^2 - y^2$$

and the plane $z = 0$, as on Fig. 24.

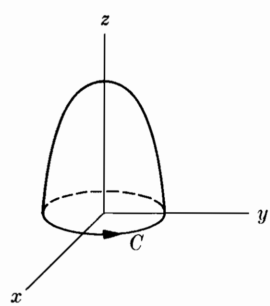


Figure 24

First we compute the integral over the boundary curve, which is just the circle

$$x^2 + y^2 = 4.$$

We parametrize the circle by $x = 2\cos\theta$ and $y = 2\sin\theta$ as usual. Then

$$\begin{aligned} F \cdot dC &= (z - y)\,dx + (x + z)\,dy - (x + y)\,dz \\ &= -2\sin\theta(-2\sin\theta\, d\theta) + 2\cos\theta(2\cos\theta)\, d\theta \\ &= 4\, d\theta. \end{aligned}$$

Consequently,

$$\int_C F \cdot dC = \int_0^{2\pi} 4\, d\theta = 8\pi.$$

Now we evaluate the surface integral. First we get the curl, namely

$$\operatorname{curl} F = \begin{vmatrix} E_1 & E_2 & E_3 \\ D_1 & D_2 & D_3 \\ z - y & x + z & -x - y \end{vmatrix} = (-2, 2, 2).$$

We can compute the normal vector as in §1, or by observing that the surface is defined by the equation

$$f(x, y, z) = z - 4 + x^2 + y^2 = 0,$$

and then finding

$$\operatorname{grad} f(x, y, z) = (2x, 2y, 1),$$

so that

$$\mathbf{n} = \frac{1}{\sqrt{4x^2 + 4y^2 + 1}}(2x, 2y, 1).$$

Then

$$\iint_S \operatorname{curl} F \cdot \mathbf{n}\, d\sigma = \iint_D (-4x + 4y + 2)\, dx\, dy,$$

where D is the disc defined by $x^2 + y^2 \leqq 4$. This last integral is easily found to be equal to 8π, which is, of course, the same value as the integral of F over the curve in the first part of the example.

Remark. Green's and Stokes' theorems are special cases of higher dimensional theorems expressing a relation between an integral over a region in space, and another integral over the boundary of the region. To give a systematic treatment requires somewhat more elaborate foundations, and lies beyond the bounds of this course.

Exercises

Verify Stokes' theorem in each one of the following cases.

1. $F(x, y, z) = (z, x, y)$, S defined by $z = 4 - x^2 - y^2$, $z \geqq 0$.
2. $F(x, y, z) = (x^2 + y, yz, x - z^2)$ and S is the triangle defined by the plane $2x + y + 2z = 2$ and $x, y, z \geqq 0$.

3. $F(x, y, z) = (x, z, -y)$ and the surface is the portion of the sphere of radius 2 centered at the origin, such that $y \geqq 0$.
4. $F(x, y, z) = (x, y, 0)$ and the surface is the part of the paraboloid $z = x^2 + y^2$ inside the cylinder $x^2 + y^2 = 4$.
5. $F(x, y, z) = (y + x, x + z, z^2)$, and the surface is that part of the cone $z^2 = x^2 + y^2$ between the planes $z = 0$ and $z = 1$.

Compute the integral $\iint_S \operatorname{curl} F \cdot \mathbf{n}\, d\sigma$ by means of Stokes' theorem.

6. $F(x, y, z) = (y, z, x)$ over the triangle with vertices at the unit points $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$.
7. $F(x, y, z) = (x + y, y - z, x + y + z)$ over the hemisphere

$$x^2 + y^2 + z^2 = a^2, \qquad z \geqq 0.$$

APPENDIX

Fourier Series

In this appendix, we discuss a little more systematically the scalar product in the context of spaces of functions. This may be covered at the same time that Chapter I is discussed, but I place the material as an appendix in order not to interrupt the discussion of ordinary vectors after Chapter I.

§1. General scalar products

Let V be the set (also called the space) of continuous functions on some interval, say the interval $[-\pi, \pi]$ which is of interest in Fourier series. We define the **scalar product** of functions f, g in V to be the number

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)g(x)\,dx.$$

This scalar product satisfies conditions analogous to those of Chapter I, namely:

SP 1. *We have* $\langle v, w\rangle = \langle w, v\rangle$ *for all* v, w *in* V.

SP 2. *If* u, v, w *are elements of* V, *then*

$$\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle.$$

SP 3. *If* x *is a number, then*

$$\langle xu, v\rangle = x\langle u, v\rangle = \langle u, xv\rangle.$$

SP 4. *For all* v *in* V *we have* $\langle v, v\rangle \geqq 0$, *and* $\langle v, v\rangle > 0$ *if* $v \neq 0$.

The verification of these properties amounts to recalling simple properties of the integral. For instance, for **SP 1,** we have

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)g(x)\,dx = \int_{-\pi}^{\pi} g(x)f(x)\,dx = \langle g, f\rangle.$$

We leave the verification of **SP 2** and **SP 3** as exercises. To prove **SP 4,** suppose that f is a non-zero function, This means that there exists some

point c in the interval $[-\pi, \pi]$ such that $f(c) \neq 0$. Then

$$\langle f, f \rangle = \int_{-\pi}^{\pi} f(x)^2 \, dx,$$

and $f(x)^2$ is a function which is always $\geqq 0$, and such that

$$f(c)^2 > 0.$$

Thus the graph of $f(x)^2$ may look like this.

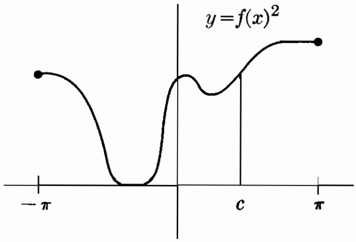


Figure 1

Let $p(x) = f(x)^2$. Geometrically, the integral of $p(x)$ from $-\pi$ to π is the area under the curve $y = p(x)$ between $-\pi$ and π, and this area cannot be 0 since $p(c) > 0$, so the area is > 0. We can give a more formal argument by observing that by continuity, there is an interval of radius r around c and a number $s > 0$ such that

$$p(x) \geqq s$$

for all x in this interval. Then by the definition of the integral according to lower sums,

$$\int_{-\pi}^{\pi} p(x) \, dx \geqq rs > 0.$$

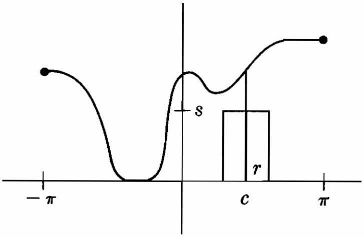


Figure 2

All the discussion of Chapter I which was carried out using only the four properties **SP 1** through **SP 4** is not seen to be valid in the present

context. For instance, we define elements v, w in V to be **orthogonal,** or **perpendicular**, and write $v \perp w$, if and only if $\langle v, w\rangle = 0$. We define the **norm** of v to be

$$\|v\| = \sqrt{\langle v, v\rangle}.$$

Remark. In analogy with ordinary Euclidean space, elements of V are also sometimes called **vectors.** More generally, one can define the general notion of a vector space, which is simply a set whose elements can be added and multiplied by numbers in such a way as to satisfy the basic properties of addition and multiplication (e.g. associativity and commutativity). Continuous functions on an interval form such a space. In an arbitrary vector space, one can then define the notion of a scalar product satisfying the above four conditions. For our purposes, which is to concentrate on the calculus part of the subject, we work right away in this function space. However, you should observe throughout that all the arguments of this section use only the basic axioms. Of course, when we want to find the norm of a specific function, like $\sin 3x$, then we use specifically the fact that we are working with the scalar product defined by the integral.

We shall now summarize a few properties of the norm.

If c is any number, then we immediately get

$$\|cv\| = |c|\,\|v\|,$$

because

$$\|cv\| = \sqrt{\langle cv, cv\rangle} = \sqrt{c^2\langle v, v\rangle} = |c|\,\|v\|.$$

Thus we see the same type of arguments as in Chapter I apply here. In fact, any argument given in Chapter I which does not use coordinates applies to our more general situation. We shall see further examples as we go along.

As before, we say that an element $v \in V$ is a **unit vector** if $\|v\| = 1$. If $v \in V$ and $v \neq 0$, then $v/\|v\|$ is a unit vector.

The following two identities follow directly from the definition of the length.

The Pythagoras theorem. *If v, w are perpendicular, then*

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2.$$

The parallelogram law. *For any v, w we have*

$$\|v + w\|^2 + \|v - w\|^2 = 2\|v\|^2 + 2\|w\|^2.$$

The proofs are trivial. We give the first, and leave the second as an exer-

cise. For the first, we have

$$\begin{aligned}\|v + w\|^2 = \langle v + w, v + w\rangle &= \langle v, v\rangle + 2\langle v, w\rangle + \langle w, w\rangle \\ &= \|v\|^2 + \|w\|^2.\end{aligned}$$

Let w be an element of V such that $\|w\| \neq 0$. For any v there exists a unique number c such that $v - cw$ is perpendicular to w. Indeed, for $v - cw$ to be perpendicular to w we must have

$$\langle v - cw, w\rangle = 0,$$

whence $\langle v, w\rangle - \langle cw, w\rangle = 0$ and $\langle v, w\rangle = c\langle w, w\rangle$. Thus

$$c = \frac{\langle v, w\rangle}{\langle w, w\rangle}\cdot$$

Conversely, letting c have this value shows that $v - cw$ is perpendicular to w. We call c the **component of v along w**. This component is also called the **Fourier coefficient** of v with respect to w, to fit the applications in the theory of Fourier Series.

In particular, if w is a unit vector, then the component of v along w is simply

$$c = \langle c, w\rangle.$$

Example. Let V be the space of continuous functions on $[-\pi, \pi]$. Let f be the function given by $f(x) = \sin kx$, where k is some integer > 0. Then

$$\begin{aligned}\|f\| = \sqrt{\langle f, f\rangle} &= \left(\int_{-\pi}^{\pi} \sin^2 kx\, dx\right)^{1/2} \\ &= \sqrt{\pi}.\end{aligned}$$

If g is any continuous function on $[-\pi, \pi]$, then the Fourier coefficient of g with respect to f is

$$\frac{\langle g, f\rangle}{\langle f, f\rangle} = \frac{1}{\pi}\int_{-\pi}^{\pi} g(x)\sin kx\, dx.$$

Let c be the component of v along w. As with the case of n-space, we define the **projection** of v along w to be the vector cw, because of our usual picture:

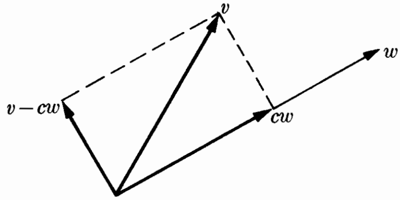


Figure 3

Exactly the same arguments which we gave in Chapter I can now be used to get the **Schwarz inequality,** namely:

Theorem 1. *For all $v, w \in V$ we have*

$$|\langle v, w\rangle| \leqq \|v\|\,\|w\|.$$

Proof. If $w = 0$, then both sides are equal to 0 and our inequality is obvious. Next, assume that $w = e$ is a unit vector, that is $e \in V$ and $\|e\| = 1$. If c is the component of v along e, then $v - ce$ is perpendicular to e, and also perpendicular to ce. Hence by the Pythagoras theorem, we find

$$\begin{aligned}\|v\|^2 &= \|v - ce\|^2 + \|ce\|^2\\ &= \|v - ce\|^2 + c^2.\end{aligned}$$

Hence $c^2 \leqq \|v\|^2$, so that $|c| \leqq \|v\|$. Finally, if w is arbitrary $\neq 0$, then $e = w/\|w\|$ is a unit vector, so that by what we just saw,

$$\left|\left\langle v, \frac{w}{\|w\|}\right\rangle\right| \leqq \|v\|.$$

This yields

$$|\langle v, w\rangle| \leqq \|v\|\,\|w\|,$$

as desired.

Theorem 2. *If $v, w \in V$, then*

$$\|v + w\| \leqq \|v\| + \|w\|.$$

Proof. Exactly the same as that of the analogous theorem in Chapter I, §4.

Let $v_1, \ldots, v_n$ be non-zero elements of V which are mutually perpendicular, that is $\langle v_i, v_j\rangle = 0$ if $i \neq j$. Let c_j be the component of v along v_i. Then

$$v - c_1v_1 - \cdots - c_nv_n$$

is perpendicular to $v_1, \ldots, v_n$. To see this, all we have to do is to take the product with v_j for any j. All the terms involving $\langle v_i, v_j\rangle$ will give 0 if $i \neq j$, and we shall have two remaining terms

$$\langle v, v_j\rangle - c_j\langle v_j, v_j\rangle$$

which cancel. Thus subtracting linear combinations as above orthogonalizes v with respect to $v_1, \ldots, v_n$. The next theorem shows that $c_1v_1 + \cdots + c_nv_n$ gives the closest approximation to v as a linear combination of $v_1, \ldots, v_n$.

Theorem 3. *Let $v_1, \ldots, v_n$ be vectors which are mutually perpendicular, and such that $\|v_i\| \neq 0$ for all i. Let v be an element of V, and let c_i be the component of v along v_i. Let $a_1, \ldots, a_n$ be numbers. Then*

$$\left\| v - \sum_{k=1}^{n} c_k v_k \right\| \leqq \left\| v - \sum_{k=1}^{n} a_k v_k \right\| .$$

Proof. We know that

$$v - \sum_{k=1}^{n} c_k v_k$$

is perpendicular to each v_i, $i = 1, \ldots, n$. Hence it is perpendicular to any linear combination of $v_1, \ldots, v_n$. Now we have:

$$\begin{aligned} \|v - \textstyle\sum a_k v_k\|^2 &= \|v - \textstyle\sum c_k v_k + \textstyle\sum (c_k - a_k) v_k\|^2 \\ &= \|v - \textstyle\sum c_k v_k\|^2 + \|\textstyle\sum (c_k - a_k) v_k\|^2 \end{aligned}$$

by the Pythagoras theorem. This proves that

$$\|v - \textstyle\sum c_k v_k\|^2 \leqq \|v - \textstyle\sum a_k v_k\|^2,$$

and thus our theorem is proved.

The next theorem is known as the **Bessel inequality.**

Theorem 4. *If $v_1, \ldots, v_n$ are mutually perpendicular unit vectors, and if c_i is the Fourier coefficient of v with respect to v_i, then*

$$\sum_{i=1}^{n} c_i^2 \leqq \|v\|^2.$$

Proof. We have

$$\begin{aligned} 0 &\leqq \langle v - \textstyle\sum c_i v_i, v - \textstyle\sum c_i v_i \rangle \\ &= \langle v, v \rangle - \textstyle\sum 2c_i \langle v, v_i \rangle + \textstyle\sum c_i^2 \\ &= \langle v, v \rangle - \textstyle\sum c_i^2. \end{aligned}$$

From this our inequality follows.

Exercises

1. Prove **SP 2** and **SP 3**, using simple properties of the integral.
2. Let $f_1, \ldots, f_n$ be functions in V which are mutually perpendicular, that is

$$\langle f_i, f_j \rangle = 0 \quad \text{if} \quad i \neq j,$$

and assume that none of the functions f_i is 0. Let $c_1, \ldots, c_n$ be numbers

such that

$$c_1 f_1 + \cdots + c_n f_n = 0$$

(the zero function). Prove that all c_i are equal to 0.

3. Let f be a fixed element of V. Let W be the subset of elements h in V such that h is perpendicular to f. Prove that if h_1, h_2 lie in W, then $h_1 + h_2$ lies in W. If c is a number and h is perpendicular to f, prove that ch is also perpendicular to f.
4. Write out the inequalities of Theorem 1 and Theorem 2 explicitly in terms of the integrals. Appreciate the fact that the notation of the text, following that of Chapter I, gives a much neater way, and a more geometric way, of expressing these inequalities.
5. Let m, n be positive integers. Prove that the functions

$$1, \sin nx, \cos mx$$

are mutually orthogonal. Use formulas like

$$\sin A \cos B = \tfrac{1}{2}[\sin (A + B) + \sin (A - B)],$$
$$\cos A \cos B = \tfrac{1}{2}[\cos (A + B) + \cos (A - B)].$$

6. Let $\varphi_n(x) = \cos nx$ and $\psi_n(x) = \sin nx$, for a positive integer n. Let φ_0 be the function such that $\varphi_0(x) = 1$, i.e. the constant function 1. Verify by performing the integrals that

$$\|\varphi_n\| = \|\psi_n\| = \sqrt{\pi} \qquad \text{and} \qquad \|\varphi_0\| = \sqrt{2\pi}.$$

7. Let V be the set of continuous functions on the interval [0, 1]. Define the scalar product in V by the integral

$$\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx.$$

(a) Prove that this satisfies conditions **SP 1** through **SP 4.** How would you define $\|f\|$ in the present context?
(b) Let $f(x) = x$ and $g(x) = x^2$. Find $\langle f, g\rangle$.
(c) With f, g as in (b), find $\|f\|$ and $\|g\|$.
(d) Let $h(x) = 1$, the constant function 1. Find $\langle f, h\rangle$, $\langle g, h\rangle$, and $\|h\|$.

§2. *Computation of Fourier series*

In the previous section we used continuous functions on the interval $[-\pi, \pi]$. For many applications one has to deal with somewhat more general functions. A convenient class of functions is that of piecewise

continuous functions. We say that f is **piecewise continuous** if it is continuous except at a finite number of points, and if at each such point c the limits

$$\lim_{\substack{h\to 0\\ h>0}} f(c-h) \quad \text{and} \quad \lim_{\substack{h\to 0\\ h>0}} f(c+h)$$

both exist. The graph of a piecewise continuous function then looks like this:

Figure 4

Let V be the set of functions on the interval $[-\pi, \pi]$ which are piecewise continuous. If f, g are in V, so is the sum $f + g$.

If c is a number, the function cf is also in V, so functions in V can be added and multiplied by numbers, to yield again functions in V. Furthermore, if f, g are piecewise continuous then the ordinary product fg is also piecewise continuous. We can then form the scalar product $\langle f, g\rangle$ since the integral is defined for piecewise continuous functions, and the three properties **SP 1** through **SP 3** are satisfied. However, the scalar product is not positive definite. A function f which is such that $f(x) = 0$ except at a finite number of points has norm 0.

Thus it is convenient, instead of **SP 4,** to formulate a slightly weaker condition:

Weak SP 4. *For all v in V we have* $\langle v, v\rangle \geqq 0$.

We then call the scalar product **positive** (not necessarily definite).

We define the **norm** of an element as before, and we ask: For which elements of V is the norm equal to 0? The answer is simple.

Theorem 5. *Let V be the space of functions which are piecewise continuous on the interval $[-\pi, \pi]$. Let f be in V. Then $\|f\| = 0$ if and only if $f(x) = 0$ for all but a finite number of points x in the interval.*

Proof. First, it is clear that if $f(x) = 0$ except for a finite number of x, then

$$\|f\|^2 = \int_{-\pi}^{\pi} f(x)^2\,dx = 0.$$

(Draw the picture of $f(x)^2$.) Conversely, suppose f is piecewise continuous on $[-\pi, \pi]$ and suppose we have a partition of $[-\pi, \pi]$ into intervals such that f is continuous on each subinterval $[a_i, a_{i+1}]$ except possibly at the end points a_i, $i = 0, \dots, r-1$. Suppose that $\|f\| = 0$, so that also

$\|f\|^2 = 0 = \langle f, f\rangle$. This means that

$$\int_{-\pi}^{\pi} f(x)^2\,dx = 0,$$

and the integral is the sum of the integrals over the smaller intervals, so that

$$\sum_{i=0}^{r-1} \int_{a_i}^{a_{i+1}} f(x)^2\,dx = 0.$$

Each integral satisfies

$$\int_{a_i}^{a_{i+1}} f(x)^2\,dx \geqq 0$$

and hence each such integral is equal to 0. However, since f is continuous on an interval $[a_i, a_{i+1}]$ except possibly at the end points, we must have $f(x)^2 = 0$ for $a_i < x < a_{i+1}$, whence $f(x) = 0$ for $a_i < x < a_{i+1}$. Hence $f(x) = 0$ except at a finite number of points.

The space V of piecewise continuous functions on $[-\pi, \pi]$ is not finite dimensional. Instead of dealing with a finite number of orthogonal vectors, we must now deal with an infinite number.

For each positive integer n we consider the functions

$$\varphi_n(x) = \cos nx, \qquad \psi_n(x) = \sin nx,$$

and we also consider the function

$$\varphi_0(x) = 1.$$

It is verified by easy direct integrations that

$$\|\varphi_n\| = \|\psi_n\| = \sqrt{\pi} \quad \text{if} \quad n \neq 0,$$
$$\|\varphi_0\| = \sqrt{2\pi}.$$

Hence the Fourier coefficients of a function f with respect to our functions 1, $\cos nx$, $\sin nx$ are equal to:

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,dx, \qquad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx,$$

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

Furthermore, the functions 1, $\cos nx$, $\sin mx$ are easily verified to be mutually orthogonal. In other words, for any pair of distinct functions

f, g among 1, $\cos nx$, $\sin mx$ we have $\langle f, g\rangle = 0$. This means:

If $m \neq n$ and $n \geqq 0$, then

$$\int_{-\pi}^{\pi} \cos nx \cos mx\, dx = 0, \qquad \int_{-\pi}^{\pi} \sin nx \sin mx = 0;$$

and for any m, n:

$$\int_{-\pi}^{\pi} \cos nx \sin mx\, dx = 0.$$

The verifications of these orthogonalities are mere exercises in elementary calculus, which you should have already done in §1.

The **Fourier series** of a function f (piecewise continuous) is defined to be the series

$$a_0 + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx).$$

The partial sum

$$s_n(x) = a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx)$$

is simply the projection of the function f on the space generated by the functions 1, $\cos kx$, $\sin kx$ for $k = 1, \ldots, n$. In the present infinite dimensional case, we write

$$f \sim a_0 + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx).$$

The sense in which one can replace the sign $\sim$ by an equality depends on various theorems whose proofs go beyond this course. One of these theorems is the following:

Theorem 6. *Assume that the piecewise continuous function f on $[-\pi, \pi]$ is orthogonal to every one of the functions 1, $\cos nx$, $\sin nx$. Then $f(x) = 0$ except at a finite number of x. If f is continuous, then $f = 0$.*

Theorem 6 shows at least that a continuous function is entirely determined by its Fourier series. There is another sense, however, in which we would like f to be equal to its Fourier series, namely we would like the values $f(x)$ to be given by

$$\begin{aligned} f(x) &= a_0 + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx) \\ &= a_0 + \lim_{n\to\infty} \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx). \end{aligned}$$

It is false in general that if f is merely continuous then $f(x)$ is given by

the series. However, it is true under some reasonable conditions, for instance:

Theorem 7. *Let $-\pi < x < \pi$ and assume that f is differentiable in some open interval containing x, and has a continuous derivative in this interval. Then $f(x)$ is equal to the value of the Fourier series.*

Example 1. Find the Fourier series of the function f such that

$$\begin{aligned} f(x) &= 0 \quad \text{if} \quad -\pi < x < 0, \\ f(x) &= 1 \quad \text{if} \quad 0 < x < \pi. \end{aligned}$$

The graph of f is as follows.

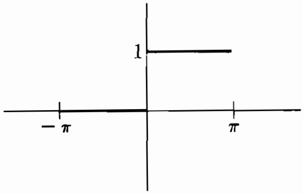


Figure 5

Since the Fourier coefficients are determined by an integral, it does not matter how we define f at $-\pi$, 0, or π. We have

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,dx = \frac{1}{2\pi}\int_0^{\pi} dx = \tfrac{1}{2},$$

$$a_n = \frac{1}{\pi}\int_0^{\pi} \cos nx\,dx = 0,$$

$$\begin{aligned} b_n &= \frac{1}{\pi}\int_0^{\pi} \sin nx\,dx = \frac{1}{\pi n}(-\cos nx)\Big|_0^{\pi} \\ &= \begin{cases} 0 & \text{if } n \text{ is even,} \\ \dfrac{2}{\pi n} & \text{if } n \text{ is odd.} \end{cases} \end{aligned}$$

Hence the Fourier series of f is:

$$f(x) \sim \tfrac{1}{2} + \sum_{m=0}^{\infty} \frac{2}{(2m+1)\pi} \sin (2m+1)x.$$

By Theorem 7, we know that $f(x)$ is actually given by the series except at the points $-\pi$, 0, and π.

Example 2. Find the Fourier series of the function f such that $f(x) = -1$ if $-\pi < x < 0$ and $f(x) = x$ if $0 < x < \pi$.

The graph of f is as follows.

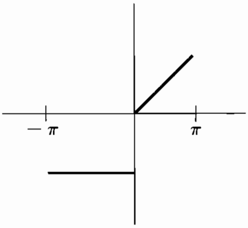


Figure 6

Again we compute the Fourier coefficients. We evaluate the integral over each of the intervals $[-\pi, 0]$ and $[0, \pi]$ since the function is given by different formulas over these intervals. We have

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{0}(-1)\,dx + \frac{1}{2\pi}\int_0^{\pi} x\,dx = \frac{1}{2} + \frac{\pi}{4},$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{0}(-1)\cos nx\,dx + \frac{1}{\pi}\int_0^{\pi} x\cos nx\,dx$$

$$= \begin{cases} 0 & \text{if } n \text{ is even,} \\ -\dfrac{2}{\pi n^2} & \text{if } n \text{ is odd,} \end{cases}$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{0}(-1)\sin nx\,dx + \frac{1}{\pi}\int_0^{\pi} x\sin nx\,dx$$

$$= \begin{cases} -\dfrac{1}{n} & \text{if } n \text{ is even,} \\ \dfrac{2}{\pi n} + \dfrac{1}{n} & \text{if } n \text{ is odd.} \end{cases}$$

Thus we obtain:

$$f(x) = \frac{1}{2} + \frac{\pi}{4} + \sum_{k=1}^{n}(a_k\cos kx + b_k\sin kx).$$

The equality is valid for $-\pi < x < 0$ and $0 < x < \pi$ by Theorem 7.

Example 3. Find the Fourier series of the function $\sin^2 x$.

We have

$$\sin^2 x = \frac{1-\cos 2x}{2} = \frac{1}{2} - \frac{1}{2}\cos 2x.$$

This is already written as a Fourier series, so the expression on the right is the desired Fourier series.

A function f is said to be **periodic** of period 2π if we have

$$f(x + 2\pi) = f(x)$$

for all x. For such a function, we then have by induction $f(x + 2\pi n) = f(x)$ for all positive integers n. Furthermore, letting $t = x + 2\pi$, we see also that

$$f(t - 2\pi) = f(t)$$

for all t, and hence $f(x - 2\pi n) = f(x)$ for all x and all positive integers n.

Given a piecewise continuous function on the interval $-\pi \leqq x < \pi$, we can extend it to a piecewise continuous function which is periodic of period 2π over all of **R**, simply by periodicity.

Example 4. Let $f(x) = x$ on $-\pi \leqq x < \pi$. If we extend f by periodicity, then the graph of the extended function looks like this:

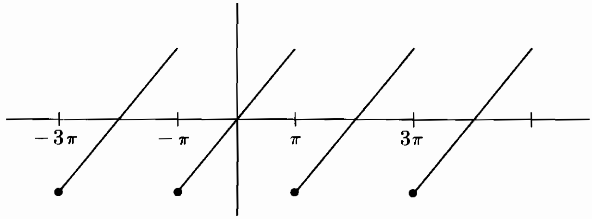


Figure 7

Example 5. Let f be the function on the interval $-\pi \leqq x < \pi$ given by:

$$\begin{aligned} f(x) &= 0 \quad \text{if} \quad -\pi \leqq x \leqq 0, \\ f(x) &= 1 \quad \text{if} \quad 0 < x < \pi. \end{aligned}$$

Then the graph of the function extended by periodicity looks like this:

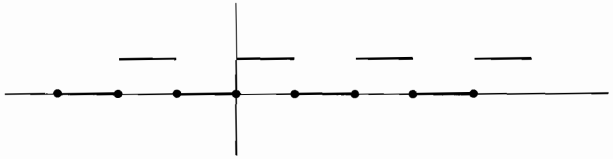


Figure 8

Example 6. Let f be the function on the interval $-\pi \leqq x < \pi$ given by $f(x) = e^x$. Then the graph of the extended function looks like this:

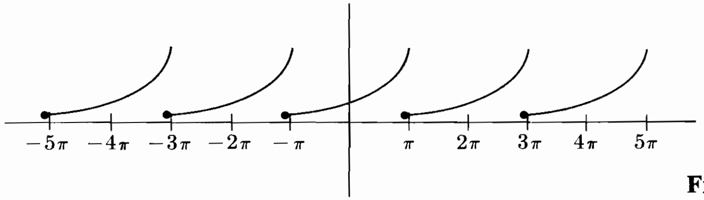


Figure 9

On the other hand, we may also be given a function over the interval $[0, 2\pi]$ and then extend this function by periodicity.

Example 7. Let $f(x) = x$ on the interval $0 \leqq x < 2\pi$. The graph of the function extended by periodicity to all of $\mathbf{R}$ looks like this:

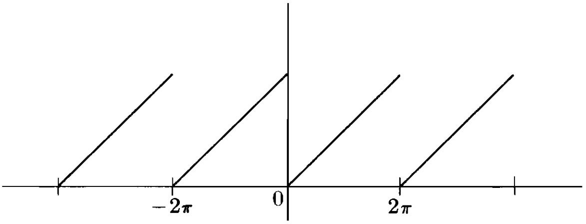


Figure 10

This is different from the function in Example 4, since in the present case, the extended function is never negative. When the function is given on an interval $[0, 2\pi]$, we compute the Fourier coefficients by taking the integral from 0 to 2π. In the present case, we therefore have:

$$a_0 = \frac{1}{2\pi}\int_0^{2\pi} x\,dx = \pi,$$

$$a_n = \frac{1}{\pi}\int_0^{2\pi} x\cos nx\,dx = 0 \qquad \text{for all } n,$$

$$b_n = \frac{1}{\pi}\int_0^{2\pi} x\sin nx\,dx = -\frac{2}{n}\cdot$$

Hence we have, for $0 < x < 2\pi$:

$$x = \pi - 2\left(\sin x + \frac{\sin 2x}{2} + \frac{\sin 3x}{3} + \cdots\right)\cdot$$

Exercises

1. (a) Let $f(x)$ be the function such that $f(x) = 2$ if $0 \leqq x < \pi$ and $f(x) = -1$ if $-\pi \leqq x < 0$. Compute $\|f\|$.
   (b) Same question, if $f(x) = x$ for $0 \leqq x < \pi$ and $f(x) = -1$ for $-\pi \leqq x < 0$.
2. If f is periodic of period 2π and a, b are numbers, show that

$$\int_a^b f(x)\,dx = \int_{a+2\pi}^{b+2\pi} f(x)\,dx = \int_{a-2\pi}^{b-2\pi} f(x)\,dx.$$

[*Hint:* Change variables, letting $u = x - 2\pi$, $du = dx$.] Also, prove:

$$\int_{-\pi}^{\pi} f(x + a)\, dx = \int_{-\pi}^{\pi} f(x)\, dx = \int_{-\pi+a}^{\pi+a} f(x)\, dx.$$

[*Hint:* Split the integral over the bounds $-\pi + a$, $-\pi$, π, $\pi + a$.]

3. Let f be an even function, that is $f(x) = f(-x)$, for all x. Assume that f is periodic of period 2π. Show that all its Fourier coefficients with respect to $\sin nx$ are 0. Let g be an odd function (that is $g(-x) = -g(x)$). Show that all its Fourier coefficients with respect to $\cos nx$ are 0.
4. Compute the Fourier series of the functions, given on the interval $-\pi < x < \pi$ by the following $f(x)$:

   (a) x (b) x^2 (c) $|x|$ (d) $\sin^2 x$
   (e) $|\sin x|$ (f) $|\cos x|$ (g) $\sin^3 x$ (h) $\cos^3 x$
5. Show that the following relations hold:

   (a) For $0 < x < 2\pi$ and $a \neq 0$,

   $$\pi e^{ax} = (e^{2a\pi} - 1)\left(\frac{1}{2a} + \sum_{k=1}^{\infty} \frac{a \cos kx - k \sin kx}{k^2 + a^2}\right).$$

   (b) For $0 < x < 2\pi$ and a not an integer,

   $$\pi \cos ax = \frac{\sin 2a\pi}{2a} + \sum_{k=1}^{\infty} \frac{a \sin 2a\pi \cos kx + k(\cos 2a\pi - 1) \sin kx}{a^2 - k^2}.$$

   (c) Letting $x = \pi$ in part (b), conclude that

   $$\frac{a\pi}{\sin a\pi} = 1 + 2a^2 \sum_{k=1}^{\infty} \frac{(-1)^k}{a^2 - k^2}.$$

   (d) For $0 < x < 2\pi$,

   $$\frac{(\pi - x)^2}{4} = \frac{\pi^2}{12} + \sum_{k=1}^{\infty} \frac{\cos kx}{k^2}.$$

Answers to Exercises

I am much indebted to Mr. Mitchell Luskin for the answers to the exercises.

Chapter I, §1

	$A + B$	$A - B$	$3A$	$-2B$
1.	$(1, 0)$	$(3, -2)$	$(6, -3)$	$(2, -2)$
2.	$(-1, 7)$	$(-1, -1)$	$(-3, 9)$	$(0, -8)$
3.	$(1, 0, 6)$	$(3, -2, 4)$	$(6, -3, 15)$	$(2, -2, -2)$
4.	$(-2, 1, -1)$	$(0, -5, 7)$	$(-3, -6, 9)$	$(2, -6, 8)$
5.	$(3\pi, 0, 6)$	$(-\pi, 6, -8)$	$(3\pi, 9, -3)$	$(-4\pi, 6, -14)$
6.	$(15 + \pi, 1, 3)$	$(15 - \pi, -5, 5)$	$(45, -6, 12)$	$(-2\pi, -6, 2)$

Chapter I, §2

1. No **2.** Yes **3.** No **4.** Yes **5.** No **6.** Yes **7.** Yes **8.** Yes

Chapter I, §3

1. (a) 5 (b) 10 (c) 30 (d) 14 (e) $\pi^2 + 10$ (f) 245
2. (a) -3 (b) 12 (c) 2 (d) -17 (e) $2\pi^2 - 16$ (f) $15\pi - 10$
4. (b) and (d) **7.** $\langle ff\rangle = \frac{2}{3}$, $\langle gg\rangle = \frac{2}{5}$, $\langle fg\rangle = 0$

Chapter I, §4

1. (a) $\sqrt{5}$ (b) $\sqrt{10}$ (c) $\sqrt{30}$
(d) $\sqrt{14}$ (e) $\sqrt{10 + \pi^2}$ (f) $\sqrt{245}$
2. (a) $\sqrt{2}$ (b) 4 (c) $\sqrt{3}$
(d) $\sqrt{26}$ (e) $\sqrt{58 + 4\pi^2}$ (f) $\sqrt{10 + \pi^2}$
3. (a) $(\frac{3}{2}, -\frac{3}{2})$ (b) $(0, 3)$ (c) $(-\frac{2}{3}, \frac{2}{3}, \frac{2}{3})$
(d) $(\frac{17}{26}, -\frac{51}{26}, \frac{34}{13})$ (e) $\dfrac{\pi^2 - 8}{2\pi^2 + 29}(2\pi, -3, 7)$ (f) $\dfrac{15\pi - 10}{10 + \pi^2}(\pi, 3, -1)$
4. (a) $(-\frac{6}{5}, \frac{3}{5})$ (b) $(-\frac{6}{5}, \frac{18}{5})$ (c) $(\frac{2}{15}, -\frac{1}{15}, \frac{1}{3})$
(d) $-\frac{17}{14}(-1, -2, 3)$ (e) $\dfrac{2\pi^2 - 16}{\pi^2 + 10}(\pi, 3, -1)$ (f) $\dfrac{3\pi - 2}{49}(15, -2, 4)$
5. (a) $\dfrac{35}{\sqrt{41 \cdot 35}}, \dfrac{6}{\sqrt{41 \cdot 6}}, 0$ (b) $\dfrac{1}{\sqrt{17 \cdot 26}}, \dfrac{6}{\sqrt{41 \cdot 17}}, \dfrac{25}{\sqrt{26 \cdot 41}}$
13. 0, 0 **14.** $\sqrt{2}$ **15.** $\sqrt{\dfrac{\pi}{3}}, \sqrt{\pi}$ **16.** $\sqrt{2\pi}$ **17.** $\sqrt{\dfrac{\pi}{n}}, \sqrt{\dfrac{\pi}{m}}$

Chapter I, §5

1. $X = (1, 1, -1) + t(3, 0, -4)$ **2.** $X = (-1, 5, 2) + t(-4, 9, 1)$
3. $y = x + 8$ **4.** $4y = 5x - 7$ **6.** (c) and (d)
7. (a) $x - y + 3z = -1$ (b) $3x + 2y - 4z = 2\pi + 26$
(c) $x - 5z = -33$
8. (a) $2x + y + 2z = 7$ (b) $7x - 8y - 9z = -29$
(c) $y + z = 1$
9. $(3, -9, -5)$, $(1, 5, -7)$ (Others would be constant multiples of these.)
10. (a) $2(t^2 + 5)^{1/2}$ **11.** $(15t^2 + 26t + 21)^{1/2}$, $\sqrt{146/15}$
12. $(-2, 1, 5)$ **13.** $(11, 13, -7)$
14. (a) $X = (1, 0, -1) + t(-2, 1, 5)$
(b) $x = (-10, -13, 7) + t(11, 13, -7)$
15. (a) $-\frac{1}{3}$ (b) $\dfrac{2}{\sqrt{42}}$ (c) $\dfrac{4}{\sqrt{66}}$ (d) $-\dfrac{2}{\sqrt{18}}$
16. (a) $(-4, \frac{11}{2}, \frac{15}{2})$ (b) $(\frac{25}{13}, \frac{10}{13}, -\frac{9}{13})$ **17.** $(1, 3, -2)$ **18.** $2/\sqrt{3}$
20. (a) $\dfrac{8}{\sqrt{35}}$ (b) $\dfrac{13}{\sqrt{21}}$
21. (a) $(-\frac{3}{2}, 4, \frac{1}{2})$ (b) $(-\frac{2}{3}, \frac{11}{3}, 0)$, $(-\frac{7}{3}, \frac{13}{3}, 1)$
(c) $(0, \frac{17}{5}, -\frac{2}{5})$ (d) $(-1, \frac{19}{5}, \frac{1}{5})$ **22.** $\dfrac{P + Q}{2}$

Chapter I, §6

1. $(-4, -3, 1)$ **2.** $(-1, 1, -1)$ **3.** $(-9, 6, -1)$ **4.** 0
5. E_3, E_1, E_2 in that order
7. $(0, -1, 0)$ and $(0, 0, 0)$; no
9. (a) $2\sqrt{131}$ (b) $\sqrt{245}$ (c) $\sqrt{470}$ (d) $\sqrt{381}$

Chapter II, §1

1. $(e^t, -\sin t, \cos t)$ **2.** $\left(2\cos 2t, \dfrac{1}{1+t}, 1\right)$ **3.** $(-\sin t, \cos t)$
4. $(-3\sin 3t, 3\cos 3t)$ **7.** B
8. $\left(\dfrac{1}{2}, \dfrac{\sqrt{3}}{2}\right) + t\left(\dfrac{1}{2}, \dfrac{\sqrt{3}}{2}\right)$, $(-1, 0) + t(-1, 0)$, or $y = \sqrt{3}x$, $y = 0$
9. $ex + y + 2z = e^2 + 3$ **10.** $x + y = 1$
11. $\sqrt{(X(T) - Q) \cdot (X(T) - Q)}$
16. (a) $(0, 1, \pi/8) + t(-4, 0, 1)$ (b) $(1, 2, 1) + t(1, 2, 2)$
(c) $(e^3, e^{-3}, 3\sqrt{2}) + t(3e^3, -3e^{-3}, 3\sqrt{2})$ (d) $(1, 1, 1) + t(1, 3, 4)$
19. $(2, 0, 4)$ and $(18, 4, 12)$ or when $t = -3$ or 1
27. $$Y'(t) = \frac{dY}{dt} = \frac{d}{dt}[X(t) \times X'(t)]$$
$$= X'(t) \times X'(t) + X(t) \times X''(t) = X(t) \times X''(t)$$

28. $$\begin{aligned} Y'(t) = \frac{dY}{dt} &= \frac{d}{dt}[X(t)\cdot(X'(t)\times X''(t))] \\ &= X'(t)\cdot(X'(t)\times X''(t)) + X(t)\cdot\frac{d}{dt}[X'(t)\times X''(t)] \\ &= X(t)\cdot(X'(t)\times X'''(t)) \end{aligned}$$

Chapter II, §2

1. $\sqrt{2}$ **2.** $2\sqrt{13}$

3. (a) $\frac{\pi}{8}\sqrt{17}$ (b) $\frac{3}{2}(\sqrt{41}-1)+\frac{5}{4}\left(\log\frac{6+\sqrt{41}}{5}\right)$ (c) $e-\frac{1}{e}$

4. (a) 8 (b) $4-2\sqrt{2}$

5. (a) $\sqrt{5}-\sqrt{2}+\log\dfrac{2+2\sqrt{2}}{1+\sqrt{5}}$ (b) $\sqrt{26}-\sqrt{10}+\log\frac{5}{3}\left(\dfrac{1+\sqrt{10}}{1+\sqrt{26}}\right)$

6. $\operatorname{Log}(\sqrt{2}+1)$

Chapter II, §3

3. $\dfrac{r}{r^2+c^2}$

4. (a) The norm of $-\frac{11}{98}(1, 2, 3)+\frac{1}{7}(0, 1, 3)$ (b) 2
(c) The norm of $\frac{11}{98}(1, -2, 3)+\frac{1}{7}(0, 1, -3)$.

6. $k(t)=\dfrac{|f''(t)|}{(1+(f')^2)^{3/2}}$ **7.** $\dfrac{(1+t^2)^{3/2}}{t}$, $\min=\dfrac{\sqrt{2}}{2}$

8. (a) $\dfrac{|-\sin t|}{(1+\cos^2 t)^{3/2}}$ (b) 1 (c) $\frac{3}{10}$

9. $\dfrac{(1+4t^2)^{3/2}}{2}$ **10.** $\dfrac{(a^2\sin^2 t+b^2\cos^2 t)^{3/2}}{ab}$

11. $|t\pi|$ **12.** $\dfrac{s}{2}$ **14.** $\dfrac{a}{v^2}$

Chapter III, §2

	$\partial f/\partial x$	$\partial f/\partial y$	$\partial f/\partial z$
1.	y	x	1
2.	$2xy^5$	$5x^2y^4$	0
3.	$y\cos(xy)$	$x\cos(xy)$	$-\sin(z)$
4.	$-y\sin(xy)$	$-x\sin(xy)$	0
5.	$yz\cos(xyz)$	$xz\cos(xyz)$	$xy\cos(xyz)$
6.	yze^{xyz}	xze^{xyz}	xye^{xyz}
7.	$2x\sin(yz)$	$x^2z\cos(yz)$	$x^2y\cos(yz)$
8.	yz	xz	xz
9.	$z+y$	$z+x$	$x+y$
10.	$\cos(y-3z)+\dfrac{y}{\sqrt{1-x^2y^2}}$	$-x\sin(y-3z)+\dfrac{x}{\sqrt{1-x^2y^2}}$	$3x\sin(y-3z)$

11. (1) $(2, 1, 1)$ (2) $(64, 80, 0)$ (6) $(6e^6, 3e^6, 2e^6)$
(8) $(6, 3, 2)$ (9) $(5, 4, 3)$

12. (4) $(0, 0, 0)$ (5) $(\pi^2 \cos \pi^2, \pi \cos \pi^2, \pi \cos \pi^2)$
(7) $(2 \sin \pi^2, \pi \cos \pi^2, \pi \cos \pi^2)$

13. $(-1, -2, 1)$

14. $\dfrac{\partial x^y}{\partial x} = yx^{y-1}$ $\dfrac{\partial x^y}{\partial y} = x^y \ln x$

15. $(-2e^{-2} \cos \pi^2, -\pi e^{-2} \sin \pi^2, -\pi e^{-2} \sin \pi^2)$

16. $(\frac{3}{2}, \frac{1}{2}, \frac{5}{2})$

Chapter III, §3

1. $2, -3$ **2.** a, b **3.** a, b, c

7. $\lim_{h\to 0} g(h, k) = -1$ $\lim_{k\to 0}\left[\lim_{h\to 0} g(h, k)\right] = -1$

$\lim_{k\to 0} g(h, k) = 1$ $\lim_{h\to 0}\left[\lim_{k\to 0} g(h, k)\right] = 1$

Chapter IV, §1

1. $\dfrac{\partial z}{\partial r} = \dfrac{\partial f}{\partial x}\dfrac{\partial u}{\partial r} + \dfrac{\partial f}{\partial y}\dfrac{\partial v}{\partial r}$ and $\dfrac{\partial z}{\partial t} = \dfrac{\partial f}{\partial x}\dfrac{\partial u}{\partial t} + \dfrac{\partial f}{\partial y}\dfrac{\partial v}{\partial t}$

2. (a) $\dfrac{\partial f}{\partial x} = 3x^2 + 3yz,$ $\dfrac{\partial f}{\partial y} = 3xz - 2yz$

$$\frac{\partial f}{\partial s} = (3x^2 + 3yz) + (3xz - 2yz)(-1) + (3xy - y^2)2s$$

$$\frac{\partial f}{\partial t} = (3x^2 + 3yz)2 + (3xz - 2yz)(-1) + (3xy - y^2)2t$$

(b) $\dfrac{\partial f}{\partial x} = \dfrac{y^2 + 1}{(1 - xy)^2},$ $\dfrac{\partial f}{\partial y} = \dfrac{x^2 + 1}{(1 - xy)^2}$

$$\frac{\partial f}{\partial s} = \frac{(x^2 + 1)\sin(3t - s)}{(1 - xy)^2}$$

$$\frac{\partial f}{\partial t} = \frac{2(y^2 + 1)\cos 2t - 3(x^2 + 1)\sin(3t - s)}{(1 - xy)^2}$$

3. $\dfrac{\partial f}{\partial x} = \dfrac{x}{(x^2 + y^2 + z^2)^{1/2}},$ $\dfrac{\partial f}{\partial y} = \dfrac{y}{(x^2 + y^2 + z^2)^{1/2}}$

4. $\dfrac{\partial r}{\partial x_i} = \dfrac{x_i}{r}$

9. (a) $-X/r^3$ (b) $2X$ (c) $-3X/r^5$ (d) $2e^{-r^2}X$ (e) $-X/r^2$
(f) $-4mX/r^{m+2}$

Chapter IV, §2

1.	Plane	Line
(a)	$6x + 2y + 3z = 49$	$X = (6, 2, 3) + t(12, 4, 6)$
(b)	$x + y + 2z = 2$	$X = (1, 1, 0) + t(1, 1, 2)$
(c)	$13x + 15y + z = -15$	$X = (2, -3, 4) + t(13, 15, 1)$
(d)	$6x - 2y + 15z = 22$	$X = (1, 7, 2) + t(-6, 2, -15)$
(e)	$4x + y + z = 13$	$X = (2, 1, 4) + t(8, 2, 2)$
(f)	$z = 0$	$X = (1, \pi/2, 0) + t(0, 0, \pi/2 + 1)$

2. (a) $(3, 0, 1)$ (b) $X = \left(\log 3, \dfrac{3\pi}{2}, -3\right) + t(3, 0, 1)$

(c) $3x + z = 3\log 3 - 3$

3. (a) $X = (3, 2, -6) + t(2, -3, 0)$ (b) $X = (2, 1, -2) + t(-5, 4, -3)$

(c) $X = (3, 2, 2) + t(2, 3, 0)$ **4.** $\sqrt{(X(t) - Q)^2}$

5. (a) $6x + 8y - z = 25$ (b) $16x + 12y - 125z = -75$

(c) $\pi x + y + z = 2\pi$ **6.** $x - 2y + z = 1$

Chapter IV, §3

1. (a) $\frac{5}{3}$ (b) $\max = \sqrt{10}$ $\min = -\sqrt{10}$

2. (a) $\dfrac{3}{2\sqrt{5}}$ (b) $\frac{48}{13}$ (c) $2\sqrt{145}$

3. Increasing $\left(-\dfrac{9\sqrt{3}}{2}, -\dfrac{3\sqrt{3}}{2}\right)$, decreasing $\left(\dfrac{9\sqrt{3}}{2}, \dfrac{3\sqrt{3}}{2}\right)$

4. (a) $\left(\dfrac{9}{2\cdot 6^{7/4}}, \dfrac{3}{2\cdot 6^{7/4}}, -\dfrac{6}{2\cdot 6^{7/4}}\right)$ (b) $(1, 2, -1, 1)$

5. $3x + 5y + 4z = 18$ **6.** $6\sqrt{6}$ **7.** $\sqrt{2}$

Chapter IV, §4

1. $\log \|X\|$ **2.** $-\dfrac{1}{2r^2}$ **3.** $\begin{cases} \log r, & k = 2 \\ \dfrac{1}{(2-k)r^{k+2}}, & k \neq 2 \end{cases}$

Chapter V, §1

1. No **2.** No **3.** No **4.** No **5.** No **6.** No

Chapter V, §2

1. $D_1\psi(x, y) = e^{xy}$, $D_2\psi(x, y) = \dfrac{xe^{xy} - e^{y}}{y} - \dfrac{e^{xy} - e^{y}}{y^2}$

2. $D_1\psi(x, y) = \cos(xy)$, $D_2\psi(x, y) = \dfrac{x}{y}\cos(xy) - \dfrac{\cos(xy)}{y^2}$

3. $D_1\psi(x, y) = (y + x)^2$
$D_2\psi(x, y) = 2yx - 2y + x^2 - 1$

4. $D_1\psi(x, y) = e^{y+x}$
$D_2\psi(x, y) = e^{y+x} - e^{y+1}$

5. $D_1\psi(x, y) = e^{y-x}$
$D_2\psi(x, y) = -e^{y-x} + e^{y-1}$

6. $D_1\psi(x, y) = x^2y^3$
$D_2\psi(x, y) = y^2x^3$

7. $D_1\psi(x, y) = \dfrac{\log(xy)}{x}$
$D_2\psi(x, y) = x\log(xy) - 1$

8. $D_1\psi(x, y) = \sin(3xy)$

$$D_2\psi(x, y) = \frac{\cos 3xy - \cos 3y}{3y^2} + \frac{3x \sin 3xy - 3\cos 3y}{y}$$

Chapter V, §3

1. No **2.** No **3.** No **4.** No

5. (a) r (b) $\log r$ (c) $\dfrac{r^{n+2}}{n+2}$ if $n \neq -2$

6. $2x^2y$ **7.** $x \sin xy$ **8.** x^3y^2

9. $x^2 + y^4$ **10.** (a) e^{xy} (b) $\sin xy$

11. $g(r)$

12. Given the vector field $F = (f_1, \ldots, f_n)$ in n-space, defined on a rectangle, $[a_1, b_1] \times \cdots \times [a_n, b_n]$. Assume that

$$\frac{\partial f_i}{\partial x_j} = \frac{\partial f_j}{\partial x_i} \qquad (\text{or} \quad D_j f_i = D_i f_j)$$

for all indices i, j. For $n = 3$, define (x, y, z) to be

$$\int_{a_1}^{x} f_1(t, y, z)\,dt + \int_{a_2}^{y} f_2(a_1, t, z)\,dt + \int_{a_3}^{z} f_3(a_1, a_2, t)\,dt,$$

and similarly for n variables. Using the hypothesis and the fact that a partial derivative of parameters can be taken in and out of an integral, you will find easily that φ is a potential function for F.

Conversely, given a vector field $F = (f_1, \ldots, f_n)$ on an open set U, if there exists a potential function, and if the partial derivatives of the functions f_i exist and are continuous, then the relations

$$\frac{\partial f_i}{\partial x_j} = \frac{\partial f_j}{\partial x_i}$$

must be satisfied for all i, j, for the same reason as that given in the text for two variables. This generalizes Theorem 4.

13. (a) $x^2 + \frac{3}{2}y^2 + 2z^2$
(b) $x + y + z$
(c) xe^{y+2z}
(d) $xy \sin z$
(e) $xyz + z^3y$
(f) xe^{yz}
(g) $xz^2 + y^2$
(h) $z \sin xy$

Chapter V, §4

1. $-\frac{369}{10}$ **2.** $\frac{23}{6}$ **3.** 0 **4.** 0 **5.** 54 **6.** $\sqrt{3C/2}$ **7.** $\frac{4}{3}$ **8.** $-\pi - \frac{8}{3}$

9. $\frac{4}{15}$ **10.** 4π

11. (a) $3\pi/4$ (b) 0 (c) 0 (d) 0

12. $-\pi/2$ **13.** 56 **15.** $\frac{9}{2}$ **16.** 3 **17.** (a) 4 (b) 3 **18.** 8 **20.** $1 - e^{-2\pi}$

22. (a) No (b) $\frac{1}{8}$ (c) 0

Chapter VI, §1

	$\partial^2 f/\partial x^2$	$\partial^2 f/\partial y^2$	$\partial^2 f/\partial x\,\partial y$
1.	$y^2 e^{xy}$	$x^2 e^{xy}$	$yxe^{xy} + e^{xy}$
2.	$-y^2 \sin xy$	$-x^2 \sin xy$	$-xy \sin xy + \cos xy$
3.	$2y^3$	$6x^2y$	$6xy^2 + 3$
4.	0	2	2
5.	$2e^{x^2+y^2} + 4x^2e^{x^2+y^2}$	$e^{x^2+y^2}(2 + 4y^2)$	$4xye^{x^2+y^2}$
6.	$2\cos(x^2 + y)$ $-4x^2 \sin(x^2 + y)$	$-\sin(x^2 + y)$	$-2x \sin(x^2 + y)$
7.	$-(3x^2 + y)^2 \cos(x^3 + xy)$ $-6x \sin(x^3 + xy)$	$-x^2 \cos(x^3 + xy)$	$-(3x^2 + y)x \cos(x^3 + xy)$ $-\sin(x^3 + xy)$

8. $$\frac{\partial^2 f}{\partial x^2} = \frac{2(1 + (x^2 - 2xy)^2) - 2(2x - 2y)^2(x^2 - 2xy)}{(1 + (x^2 - 2xy)^2)^2}$$

$$\frac{\partial^2 f}{y^2} = \frac{[1 + (x^2 - 2xy)^2](-2) + 2x[2(x^2 - 2xy)(-2x)]}{[1 + (x^2 - 2xy)^2]^2}$$

$$\frac{\partial^2 f}{\partial x\,\partial y} = \frac{-2[1 + (x^2 - 2xy)^2] - (2x - 2y)(x^2 - 2xy)(-2x)2}{[1 + (x^2 - 2xy)^2]^2}$$

9. All three $= e^{x+y}$ **10.** All three $= -\sin(x + y)$

11. 1 **12.** $2x$ **13.** $e^{xyz}(1 + 3xyz + x^2y^2z^2)$

14. $(1 - x^2y^2z^2)\cos xyz - 3xyz \sin xyz$

15. $\sin(x + y + z)$ **16.** $-\cos(x + y + z)$

17. $-\dfrac{48xyz}{(x^2 + y^2 + z^2)^4}$ **18.** $6x^2y$

Chapter VI, §2

1. $9D_1^2 + 12D_1D_2 + 4D_2^2$

2. $D_1^2 + D_2^2 + D_3^2 + 2D_1D_2 + 2D_2D_3 + 2D_1D_3$

3. $D_1^2 - D_2^2$ **4.** $D_1^2 + 2D_1D_2 + D_2^2$

5. $D_1^3 + 3D_1^2D_2 + 3D_1D_2^2 + D_2^3$

6. $D_1^4 + 4D_1^3D_2 + 6D_1^2D_2^2 + 4D_1D_2^3 + D_2^4$

7. $2D_1^2 - D_1D_2 - 3D_2^2$ **8.** $D_1D_2 - D_3D_2 + 5D_1D_3 - 5D_3^2$

9. $\left(\dfrac{\partial}{\partial x}\right)^3 + 12\left(\dfrac{\partial}{\partial x}\right)^2 \dfrac{\partial}{\partial y} + 48\dfrac{\partial}{\partial x}\left(\dfrac{\partial}{\partial y}\right)^2 + 64\left(\dfrac{\partial}{\partial y}\right)^3$

10. $4\left(\dfrac{\partial}{\partial x}\right)^2 + 4\dfrac{\partial}{\partial x}\dfrac{\partial}{\partial y} + \left(\dfrac{\partial}{\partial y}\right)^2$ **11.** $h^2\left(\dfrac{\partial}{\partial x}\right)^2 + 2hk\dfrac{\partial}{\partial x}\dfrac{\partial}{\partial y} + k^2\left(\dfrac{\partial}{\partial y}\right)^2$

12. $h^3\left(\dfrac{\partial}{\partial x}\right)^3 + 3h^2k\left(\dfrac{\partial}{\partial x}\right)^2\dfrac{\partial}{\partial y} + 3hk^2\dfrac{\partial}{\partial x}\left(\dfrac{\partial}{\partial y}\right)^2 + k^3\left(\dfrac{\partial}{\partial y}\right)^3$

13. 8 **14.** 4 **15.** 4 **16.** 1 **17.** (a) $2880xy$ (b) 0 (c) 144 (d) 1440

18. (a) 0 (b) $3 \cdot 7!9!$ (c) $11 \cdot 7!9!$ (d) 0

19. (a) 576 (b) $-7 \cdot 9!4!$ (c) 6 (d) 0

20. (a) 0 (b) 48 (c) $7 \cdot 6!10!7!$ (d) 0

23. $\left(\dfrac{\partial}{\partial r}\right)^2 + \dfrac{1}{r}\dfrac{\partial^2}{\partial r\,\partial \theta}$

Chapter VI, §3

1. xy **2.** 1 **3.** xy **4.** $x^2 + y^2$

5. $1 + x + y + \frac{x^2}{2} + xy + \frac{y^2}{2}$ **6.** $1 - \frac{y^2}{2}$ **7.** x

8. $y + xy$ **9.** $x + xy + 2y^2$ **10.** Yes, 0 **11.** (a) Yes, 0 (b) Yes, 1

12. Yes, 0 **13.** Yes, 0 **14.** $1 + x + \frac{x^2}{2} - \frac{y^2}{2} + \frac{x^3}{6} - \frac{xy^2}{2}$ **15.** 0

17. Terms up to degree 2 given in text. Term of degree 3 is $\frac{1}{3}(x + 2y)^3$.

18. (1) $-\pi(x - 1) + (y - \pi)$

(2) $-1 + \frac{\pi^2}{2}(x - 1)^2 + \pi(x - 1)(y - \pi)$

(3) $\log 7 + \frac{3}{7}(x - 2) + \frac{2}{7}(y - 3) - \frac{9}{18}(x - 2)^2$
$+ \frac{1}{49}(x - 2)(y - 3) - \frac{4}{98}(y - 3)^2$

(4) $2\sqrt{\pi}(x - \sqrt{\pi}) + 2\sqrt{\pi}(y - \sqrt{\pi}) + (x - \sqrt{\pi})^2 + (y - \sqrt{\pi})^2$

(5) $e^3 + e^3(x - 1) + e^3(y - 2) + \frac{e^3}{2}(x - 1)^2$

$$+ e^3(x - 1)(y - 2) + \frac{e^3}{2}(y - 2)^2$$

(6) $-1 + \frac{1}{2}(y - \pi)^2$

(7) $1 - \frac{1}{2}(x - \pi/2)^2 - \frac{1}{2}(y - \pi)^2$

(8) $\frac{e^2\sqrt{2}}{2} + \frac{e^2\sqrt{2}}{2}(x - 2) + \frac{e^2\sqrt{2}}{2}(y - \pi/2)$

$$+ \frac{e^2\sqrt{2}}{4}(x - 2)^2 + \frac{e^2\sqrt{2}}{2}(x - 2)(y - \pi/4)$$

$$+ \frac{e^2\sqrt{2}}{4}(y - \pi/4)^2$$

(9) $4 + 2(x - 1) + 5(y - 1) + (x - 1)(y - 1) + 2(y - 1)^2$

20. (a) $X + t(Y - X)$, $0 \leqq t \leqq 1$

(b) By the mean value theorem applied to the function

$$g(t) = f\big(X + t(Y - X)\big),$$

we get

$$f(Y) - f(X) = \big(\operatorname{grad} f(Z)\big) \cdot (Y - X)$$

for some Z on the line segment. Now use the Schwarz inequality.

Chapter VI, §4

1. First observe that for each point X we have

$$f(X) - f(O) = \int_0^1 Df(tX)\,dt,$$

where $D = x_1 D_1 + \cdots + x_n D_n$. Assuming that $f(O) = 0$, and repeating

the argument, assuming that $\nabla f(O) = 0$, we obtain

$$f(X) = \int_0^1 \int_0^1 tD^2 f(stX)\, ds\, dt.$$

Thus we find

$$f(X) = \sum_{i,j=1}^{n} h_{ij}(X) x_i x_j,$$

where

$$h_{ij}(X) = \int_0^1 \int_0^1 tD_iD_j f(stX)\, ds\, dt, \qquad \text{if} \quad i \neq j,$$

$$h_{ji}(X) = \int_0^1 \int_0^1 \tfrac{1}{2} tD_iD_j f(stX)\, ds\, dt, \qquad \text{if} \quad i = j.$$

We have $h_{ij} = h_{ji}$ because $D_iD_j = D_jD_i$.

Chapter VII, §1

1. $(2, 1)$, neither max nor min
2. $\big((2n + 1)\pi, 1\big)$ and $(2n\pi, -1)$, neither max nor min
3. $(0, 0, 0)$, min, value 0
4. $\pm(\sqrt{2}/2, \sqrt{2}/2)$, neither local max nor min. [*Hint*: Change variables, letting $u = x + y$ and $v = x - y$. Then the critical points are at $\pm(\sqrt{2}, 0)$, and in the (u, v)-plane, near these points, the function increases in one direction and decreases in the other.]
5. All points of form $(0, t, -t)$, neither max nor min.
6. All (x, y, z) with $x^2 + y^2 + z^2 = 2n\pi$ are local max, value 1.
All (x, y, z) with $x^2 + y^2 + z^2 = (2n + 1)\pi$ are local min, value -1.
7. All points $(x, 0)$ and $(0, y)$ are mins, value 0.
8. $(0, 0)$, min, value 0 **9.** (t, t), min, value 0
10. $(0, n\pi)$, neither max nor min **11.** $(1/2, 0)$, min, value $-1/4$
12. $(0, 0, 0)$, max, value 1 **13.** $(0, 0, 0)$, min, value 1

Chapter VII, §2

3. (1) $x^2 + 4xy - y^2$
(2) At $\big((2n + 1)\pi, 1\big)$, $-xy$. At $(2n\pi, -1)$, $+xy$.
(3) $x^2 + y^2 + z^2$
(4) $-\dfrac{1}{\sqrt{2}} e^{-1/2}\left(\dfrac{x^2}{2} + 3xy + \dfrac{y^2}{2}\right)$ at $(\sqrt{2}/2, \sqrt{2}/2)$

$\dfrac{1}{\sqrt{2}} e^{1/2}\left(\dfrac{x^2}{2} + 3xy + \dfrac{y^2}{2}\right)$ at $(-\sqrt{2}/2, -\sqrt{2}/2)$

(5) $xy + xz$
(6) At (a, b, c) such that $a^2 + b^2 + c^2 = 2n\pi$, the form is

$$-2(a^2x^2 + b^2y^2 + c^2z^2) - 4(abxy + acxz + bcyz).$$

At the point (a, b, c) such that $a^2 + b^2 + c^2 = (2n + 1)\pi$, the form is

$$2(a^2x^2 + b^2y^2 + c^2z^2) + 4(abxy + acxz + bcyz).$$

(7) At points $(a, 0)$ we get a^2y^2. At points $(0, b)$, we get b^2x^2.
(8) y^2 (9) 0 (10) $\pm xy$ (11) $x^2 + 2y^2$
(12) $-x^2 - y^2 - z^2$ (13) $x^2 + y^2 + z^2$

4. (a) Neither (b) Min (c) Max (d) Neither (e) Neither (f) Neither (g) Max (h) Neither

Chapter VII, §3

1. Min $= -2$ at $(-1, -1)$, max $= 2$ at $(1, 1)$ **2.** None
3. Max $\frac{1}{2}$ at $(\sqrt{2}/2, \sqrt{2}/2)$ and $(-\sqrt{2}/2, -\sqrt{2}/2)$
4. Max at $(\frac{1}{2}, \frac{1}{3})$, no min
5. Min 0 at $(0, 0)$, max $2/e$ at $(0, \pm 1)$, rel. max at $(\pm 1, 0)$
6. Max $= 1$ at $(0, 1)$, min $= 1/9$ at $(0, 3)$
7. (a) Both (b) Neither (c) Neither (d) Min (e) Both (f) Max (g) Min
8. $t = (2n + 1)\pi$, so $(-1, 0, 1)$ and $(-1, 0, -1)$

Chapter VII, §4

1. (a) $-1/\sqrt{2}$ (b) $9/8$ **2.** $1 + 1/\sqrt{2}$ **3.** At $(\frac{5}{3}, \frac{2}{3}, \frac{1}{3})$ min $= 12$
4. $X = \frac{1}{3}(A + B + C)$, min value is $\frac{2}{3}(A^2 + B^2 + C^2 - AB - AC - BC)$
5. 45 at $\pm(\sqrt{3}, \sqrt{6})$ **6.** $(\frac{2}{3})^{3/2}$ at $\sqrt{\frac{2}{3}}\,(1, 1, 1)$ **7.** Min 0, max 0
8. Max at $(\pi/8, -\pi/8)$, value $2\cos^2(\pi/8)$; min at $(5\pi/8, 3\pi/8)$ value $\cos^2(5\pi/8) + \cos^2(3\pi/8)$
9. $(0, 0, \pm 1)$ **10.** No min, max $= \frac{1}{4}$ at $(\frac{1}{2}, \frac{1}{2})$ **11.** 1
12. Max $= \sqrt{3}$ at $\sqrt{3}/3(1, 1, 1)$, min $= -\sqrt{3}$ at $-\sqrt{3}/3(1, 1, 1)$.
13. Values 3 and -3 at $(\frac{1}{3}, -\frac{2}{3}, \frac{2}{3})$ and $(-\frac{1}{3}, \frac{2}{3}, -\frac{2}{3})$
14. $2\sqrt{3}$ at (x, x, x) with $x = \sqrt{4/3}$ **15.** Max $= 1/27$, no min
16. Max $= 121/49$, min $= 0$ **17.** $25/62$ **18.** $d^2/(a^2 + b^2 + c^2)$
20. $\left(\dfrac{11 - 8\sqrt{5}}{2}\right)^{1/2}$

Chapter VIII, §1

1. $A + B = \begin{pmatrix} 0 & 7 & 1 \\ 0 & 1 & 1 \end{pmatrix}$, $3B = \begin{pmatrix} -3 & 15 & -6 \\ 3 & 3 & -3 \end{pmatrix}$

$-2B = \begin{pmatrix} 2 & -10 & 4 \\ -2 & -2 & 2 \end{pmatrix}$, $A + 2B = \begin{pmatrix} -1 & 12 & -1 \\ 1 & 2 & 0 \end{pmatrix}$

$2A + B = \begin{pmatrix} 1 & 9 & 4 \\ -1 & 1 & 3 \end{pmatrix}$, $A - B = \begin{pmatrix} 2 & -3 & 5 \\ -2 & -1 & 3 \end{pmatrix}$

$A - 2B = \begin{pmatrix} 3 & -8 & 7 \\ -3 & -2 & 4 \end{pmatrix}$, $B - A = \begin{pmatrix} -2 & 3 & -5 \\ 2 & 1 & -3 \end{pmatrix}$

2. $A + B = \begin{pmatrix} 0 & 0 \\ 2 & -2 \end{pmatrix}$, $3B = \begin{pmatrix} -3 & 3 \\ 0 & -9 \end{pmatrix}$ $-2B = \begin{pmatrix} 2 & -2 \\ 0 & 6 \end{pmatrix}$,

$A + 2B = \begin{pmatrix} -1 & 1 \\ 2 & -5 \end{pmatrix}$, $A - B = \begin{pmatrix} 2 & -2 \\ 2 & 4 \end{pmatrix}$, $B - A = \begin{pmatrix} -2 & 2 \\ -2 & -4 \end{pmatrix}$

3. ${}^tA = \begin{pmatrix} 1 & -1 \\ 2 & 0 \\ 3 & 2 \end{pmatrix}$, ${}^tB = \begin{pmatrix} -1 & 1 \\ 5 & 1 \\ -2 & -1 \end{pmatrix}$

4. ${}^tA = \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix}$, ${}^tB = \begin{pmatrix} -1 & 0 \\ 1 & -3 \end{pmatrix}$ **7.** Same

8. $\begin{pmatrix} 0 & 2 \\ 0 & -2 \end{pmatrix}$, same **9.** $A + {}^tA = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, $B + {}^tB = \begin{pmatrix} -2 & 1 \\ 1 & -6 \end{pmatrix}$

11. Rows of A: $(1, 2, 3)$, $(-1, 0, 2)$

Columns of A: $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$, $\begin{pmatrix} 2 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$

Rows of B: $(-1, 5, -2)$, $(1, 1, -1)$

Columns of B: $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$, $\begin{pmatrix} 5 \\ 1 \end{pmatrix}$, $\begin{pmatrix} -2 \\ -1 \end{pmatrix}$

12. Rows of A: $(1, -1)$, $(2, 1)$ Columns of A: $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

Rows of B: $(-1, 1)$, $(0, -3)$ Columns of B: $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 1 \\ -3 \end{pmatrix}$

Chapter VIII, §2

1. $IA = AI = A$ **2.** 0

3. (a) $\begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix}$ (b) $\begin{pmatrix} 10 \\ 14 \end{pmatrix}$ (c) $\begin{pmatrix} 33 & 37 \\ 11 & -18 \end{pmatrix}$

5. $AB = \begin{pmatrix} 4 & 2 \\ 5 & -1 \end{pmatrix}$, $BA = \begin{pmatrix} 2 & 4 \\ 4 & 1 \end{pmatrix}$

6. $AC = CA = \begin{pmatrix} 7 & 14 \\ 21 & -7 \end{pmatrix}$, $BC = CB = \begin{pmatrix} 14 & 0 \\ 7 & 7 \end{pmatrix}$.

If $C = xI$, where x is a number, then $AC = CA = xA$.

7. $(3, 1, 5)$, first row **8.** Second row, third row, i-th row **9.** O **10.** O

11. (a) $A^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $A^3 = O$ matrix. If $B = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ then

$$B^2 = \begin{pmatrix} 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B^3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \text{and } B^4 = O.$$

(b) $A^2 = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$, $A^3 = \begin{pmatrix} 1 & 3 & 6 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}$, $A^4 = \begin{pmatrix} 1 & 4 & 10 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{pmatrix}$

12. (a) $\begin{pmatrix} 4 \\ 9 \\ 5 \end{pmatrix}$ (b) $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ (c) $\begin{pmatrix} x_2 \\ 0 \end{pmatrix}$ (d) $\begin{pmatrix} 0 \\ x_1 \end{pmatrix}$

13. (a) $\begin{pmatrix} 2 \\ 4 \end{pmatrix}$ (b) $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ (c) $\begin{pmatrix} 3 \\ 5 \end{pmatrix}$ (d) $\begin{pmatrix} 4 \\ 6 \end{pmatrix}$

14. (a) $\begin{pmatrix}3\\1\\2\end{pmatrix}$ (b) $\begin{pmatrix}7\\-1\\1\end{pmatrix}$ (c) $\begin{pmatrix}5\\4\\8\end{pmatrix}$ (d) $\begin{pmatrix}12\\3\\9\end{pmatrix}$

15. Second column of A **16.** i-th column of A **17.** j-th row of A

18. $\begin{pmatrix}1 & a+b\\0 & 1\end{pmatrix}$, $\begin{pmatrix}1 & na\\0 & 1\end{pmatrix}$ **20.** $\begin{pmatrix}1 & -a\\0 & 1\end{pmatrix}$ **21.** $(AB)^{-1} = B^{-1}A^{-1}$

22. $\begin{pmatrix}a & b\\ \dfrac{-a^2}{b} & -a\end{pmatrix}$ for any $a, b \neq 0$; if $b = 0$, then $\begin{pmatrix}0 & 0\\0 & 0\end{pmatrix}$.

23. $A^n = \begin{pmatrix}\cos n\theta & -\sin n\theta\\ \sin n\theta & \cos n\theta\end{pmatrix}$ **24.** $\begin{pmatrix}0 & 1\\-1 & 0\end{pmatrix}$

25. $\begin{pmatrix}1 & 0 & 0\\0 & 4 & 0\\0 & 0 & 9\end{pmatrix}$, $\begin{pmatrix}1 & 0 & 0\\0 & 8 & 0\\0 & 0 & 27\end{pmatrix}$, $\begin{pmatrix}1 & 0 & 0\\0 & 16 & 0\\0 & 0 & 81\end{pmatrix}$

26. Diagonal matrix with diagonal $a_1{}^k, a_2{}^k, \ldots a_n{}^k$ **27.** $A^3 = 0$

Chapter IX, §1

1. (a) 11 (b) 13 (c) 6 **2.** (a) $(e, 1)$ b) $(1, 0)$ (c) $(1/e, -1)$
3. (a) 1 (b) 11 **4.** Ellipse $9x^2 + 4y^2 = 36$ **5.** Line $x = 2y$
6. Circle $x^2 + y^2 = e^2$, circle $x^2 + y^2 = e^{2c}$
7. Cylinder, radius 1, z-axis = axis of cylinder **8.** Circle $x^2 + y^2 = 1$
12. $A = O$

Chapter IX, §2

1. (a) $(5, 3)$ (b) $(5, 0)$ (c) $(5, 1)$ (d) $(0, -3)$

2. $\begin{pmatrix}r & 0 & \cdots & 0\\0 & r & & \vdots\\ \vdots & & \ddots & 0\\0 & & \cdots & r\end{pmatrix}$ **3.** $\begin{pmatrix}a & 0\\0 & b\end{pmatrix}$ **4.** $\begin{pmatrix}a_1 & 0 & 0\\0 & a_2 & 0\\0 & 0 & a_3\end{pmatrix}$ **5.** $\begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\end{pmatrix}$

6. $(1, 0, 0)$ **7.** $\begin{pmatrix}1 & 0 & 0\\0 & 0 & 1\end{pmatrix}$ **8.** $\begin{pmatrix}0 & 1 & 0\\0 & 0 & 1\end{pmatrix}$ **9.** $\begin{pmatrix}1 & 0 & 0 & 0\\0 & 1 & 0 & 0\end{pmatrix}$

10. $\begin{pmatrix}1 & 0 & 0 & 0\\0 & 1 & 0 & 0\\0 & 0 & 1 & 0\end{pmatrix}$ **11.** Only $A = 0$. **12.** $\begin{pmatrix}-5 & 3\\7 & 1\end{pmatrix}$

13. $\begin{pmatrix}-1 & 2\\4 & 6\end{pmatrix}$ **14.** $\begin{pmatrix}1 & -2 & 8\\3 & 7 & -5\\4 & 9 & 2\end{pmatrix}$ **16.** $\begin{pmatrix}-3 & 4 & 5\\5 & 1 & -2\\0 & -7 & 8\end{pmatrix}$

17. Let $A = L(1)$. For any number t, we have by linearity, $L(t) = L(t \cdot 1)$ $tL(1) = tA$.

18. $\begin{pmatrix}a_{11} & a_{12}\\a_{21} & a_{22}\\a_{31} & a_{32}\end{pmatrix}$ **19.** $\begin{pmatrix}3 & -5\\1 & 7\\-4 & -8\end{pmatrix}$

Chapter IX, §3

3. $L(E^1) = \begin{pmatrix} 3 \\ 5 \end{pmatrix}, \quad L(E^2) = \begin{pmatrix} -1 \\ 2 \end{pmatrix}.$

5. It is the set of all points

$$t_1A + t_2B + t_3C,$$

with numbers t_i satisfying $0 \leqq t_i \leqq 1$ for $i = 1,2,3$. Let S be this parallelepiped. The image of S under L is the set $L(S)$ consisting of all points

$$t_1L(A) + t_2L(B) + t_3L(C),$$

with t_i satisfying the above inequality. Hence it is a parallelepiped if $L(A)$, $L(B)$, $L(C)$ do not all lie in a plane.

6. $\begin{pmatrix} -3 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix}, \begin{pmatrix} 4 \\ 1 \\ 5 \end{pmatrix}$

7. The three column vectors of the matrix
8. It is the set of points $L(P) + tL(A)$ with all t in $\mathbf{R}$.
9. (a) $P + t(Q - P)$ (b) $L(P) + tL(Q - P) = L(P) + t[L(Q) - L(P)]$
10. It is the set of points $tL(A) + sL(B)$, with t, s in $\mathbf{R}$.
11. It is the set of points $L(P) + tL(A) + sL(B)$ with t, s in $\mathbf{R}$.

Chapter IX, §4

1. Inverse of F is the map G such that $G(X) = (1/7)X$.

2. $G(X) = (-1/8)X$ **3.** $G(X) = c^{-1}X$.

4. $(AB)^{-1} = B^{-1}A^{-1}$; $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$. Just multiply out

$$ABB^{-1}A^{-1} = I \quad \text{and} \quad ABCC^{-1}B^{-1}A^{-1} = I.$$

The same also holds taking the multiplication on the other side.

5. $(I + A)(I - A) = (I - A)(I + A) = I^2 - A^2 = I$ so $I + A$ is an inverse for $I - A$.

6. $I = A(-2I - A)$, so $-(2I + A)$ is an inverse (it commutes with A).

7. We have $(I - A)(I + A + A^2) = (I + A + A^2)(I - A) = I - A^3 = I$, so $I + A + A^2$ is an inverse for $I - A$.

Chapter X, §1

1. (a) 26 (b) 5 (c) -5 (d) -42 (e) -3 (f) 9
2. 1 **3.** (a) 1 (b) -1 (c) $-\frac{1}{2}$ (d) 0 **5.** $D(cA) = c^2D(A)$.

Chapter X, §2

2. (a) -20 (b) 5 (c) 4 (d) 5 (e) -76 (f) -14
3. (a) 140 (b) 120 (c) -60
4. abc **5.** (a) 3 (b) -24 (c) 16 (d) 14 (e) 0 (f) 8 (g) 8 (h) -10
6. $a_{11}a_{22}a_{33}$ both (a) and (b)

Chapter X, §3

5. (a) -20 (b) 7 (c) 4 (d) 5 (e) -76 (f) -14

6. (a) 1 (b) -42 (c) 0 (d) 0 (e) 12 (f) 14 (g) 108 (h) 135 (i) 10

7. $a_{11}a_{22}a_{33}$

8. (a) 0 (b) 24 (c) -12 (d) 0 (e) 27 (f) -54 (g) -21 (h) -4
(i) 5 (j) 0 (k) -18 (l) 0

9. $D(cA) = c^3 D(A)$

14. 1 **15.** $t^2 + 8t + 5$

Chapter X, §4

1. If a number x is such that $B = xA$, then

$$D(A, B, C) = D(A, xA, C) = xD(A, A, C) = 0,$$

contrary to assumption.

5. Let x, y, z be numbers such that $xA + yB + zC = 0$. Then

$$\begin{aligned} 0 = D(O, B, C) &= D(xA + yB + zC, B, C) \\ &= xD(A, B, C) + yD(B, B, C) + zD(C, B, C) \\ &= xD(A, B, C). \end{aligned}$$

Since $D(A, B, C) \neq 0$ by assumption, it follows that $x = 0$. A similar argument computing $D(A, O, C)$ and $D(A, B, O)$ shows that $y = 0$ and $z = 0$.

Chapter X, §6

1. (a) $\begin{pmatrix} \frac{2}{9} & \frac{1}{9} \\ -\frac{5}{9} & \frac{2}{9} \end{pmatrix}$ (b) $\begin{pmatrix} \frac{1}{11} & -\frac{4}{11} \\ \frac{2}{11} & \frac{3}{11} \end{pmatrix}$ (c) $\begin{pmatrix} \frac{2}{9} & -\frac{1}{9} \\ -\frac{1}{9} & \frac{5}{9} \end{pmatrix}$ (d) $\begin{pmatrix} -\frac{4}{5} & \frac{1}{5} \\ \frac{3}{5} & -\frac{2}{5} \end{pmatrix}$

2. $\dfrac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

Chapter XI, §2

1. (a) $\begin{pmatrix} 1 & 1 \\ 2xy & x^2 \end{pmatrix}$ (b) $\begin{pmatrix} \cos x & 0 \\ -y \sin xy & -x \sin xy \end{pmatrix}$ (c) $\begin{pmatrix} ye^{xy} & xe^{xy} \\ 1/x & 0 \end{pmatrix}$

(d) $\begin{pmatrix} z & 0 & x \\ y & x & 0 \\ 0 & z & y \end{pmatrix}$ (e) $\begin{pmatrix} yz & xz & xy \\ 2xz & 0 & x^2 \end{pmatrix}$

(f) $\begin{pmatrix} yz \cos xyz & xz \cos xyz & yx \cos xyz \\ z & 0 & x \end{pmatrix}$

2. (a) $\begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}$ (b) $\begin{pmatrix} -1 & 0 \\ -\frac{\pi}{2} \sin \frac{\pi^2}{2} & -\pi \sin \frac{\pi^2}{2} \end{pmatrix}$ (c) $\begin{pmatrix} 4e^4 & e^4 \\ 1 & 0 \end{pmatrix}$

(d) $\begin{pmatrix} -1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$ (e) $\begin{pmatrix} 1 & -2 & -2 \\ -4 & 0 & 4 \end{pmatrix}$ (f) $\begin{pmatrix} 8 & 4\pi & 2\pi \\ 4 & 0 & \pi \end{pmatrix}$

4. (a) $\begin{pmatrix} y & x \\ 2x & 0 \end{pmatrix}$ (b) $\begin{pmatrix} -y \sin xy & -x \sin xy & 0 \\ y \cos xy & x \cos xy & 0 \\ z & 0 & x \end{pmatrix}$

5. $\Delta_F(X) = x^2 - 2xy$. $\Delta_F(X) = 0$ when $x = 0$, y arbitrary, and also at all points with $x = 2y$.
6. $\Delta_F(X) = -x \cos x \sin xy$
7. $\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$, r; determinant vanishes only for $r = 0$.
8. $\begin{pmatrix} \sin\varphi\cos\theta & -r\sin\varphi\sin\theta & r\cos\varphi\cos\theta \\ \cos\varphi & 0 & -r\sin\varphi \\ \sin\varphi\sin\theta & r\sin\varphi\cos\theta & r\cos\varphi\sin\theta \end{pmatrix}$

   Determinant $r^2 \sin\varphi$
9. $\begin{pmatrix} e^r\cos\theta & -e^r\sin\theta \\ e^r\sin\theta & e^r\cos\theta \end{pmatrix}$

   Determinant is e^{2r}. $F(r, \theta) = F(r, \theta + 2\pi)$.

Chapter XI, §4

1. Yes in all cases **2.** (a), (b), (c), (d) all locally C^1-invertible
3. $F(x, y) = F(x, y + 2\pi)$
5. Letting $y = \varphi(x)$, we have

$$\varphi''(x) = \frac{-1}{D_2f(x,y)^2}\left[\begin{matrix} D_2f(x,y)\big(D_1^2f(x,y) + D_2D_1f(x,y)\varphi'(x)\big) \\ -D_1f(x,y)\big(D_1D_2f(x,y) + D_2^2f(x,y)\varphi'(x)\big)\end{matrix}\right].$$

6. (a) No (b) Yes (c) Yes
9. (a) We have $2x - y - xy' + 2yy' = 0$. This yields $\varphi'(1) = 0$.

   (b) $\varphi'(1) = 1$ (c) $\varphi'(1) = -\frac{1}{3}$ (d) $\varphi'(-1) = -\frac{1}{2}$ (e) $\varphi'(0) = -1$
   (f) $\varphi'(2) = \frac{39}{41}$
10. (a) both -1
   (b) $D_1\varphi(0, 0) = 0$; $D_2\varphi(0, 0) = 0$
   (c) $D_1\varphi(1, 1) = \frac{1}{3}$; $D_2\varphi(1, 1) = \frac{1}{3}$
   (d) $D_1\varphi(0, \frac{1}{2}) = -\frac{3}{4}$; $D_2\varphi(0, \frac{1}{2}) = -1$
11. $D_1\varphi = \frac{4}{3}$, $D_2\varphi = -\frac{2}{3}$
12. (a) $D_1\varphi(0, -1) = -1$; $D_2\varphi(0, -1) = -1$
   (b) $D_1\varphi(0, 0) = 0$; $D_2\varphi(0, 0) = -1$
   (c) $D_1\varphi(1, 2) = -1$; $D_2\varphi(1, 2) = 3$
   (d) $D_1\varphi(\frac{1}{2}, \frac{1}{2}) = -\frac{3}{4}$; $D_2\varphi(\frac{1}{2}, \frac{1}{2}) = -1$

Chapter XII, §2

1. (a) 12 (b) $\frac{11}{5}$ (c) $\frac{1}{10}$ (d) $2 + \pi^2/2$ (e) $\frac{5}{6}$ (f) $\pi/4$ (g) $\frac{8}{3}$
   (h) $\frac{49}{32}\pi$ (i) 3
3. (a) $\frac{1}{4} - \pi$ (b) $e - 1/e$ (c) $\pi^2 - \frac{40}{9}$ (d) $\frac{63}{32}$
4. (a) $\frac{1}{20}$ (b) $\frac{1}{35}$ (c) 4
6. (a) $\frac{49}{20}$ (b) $e^{-3}/3 - e^{-2}/2 - e^3/3 + e^2/2$
   (c) $1 - \cos 2$ (d) 0 (e) 1 (f) $\frac{1}{6}$
7. $3\pi/8$
8. (a) $\log 2$ (b) $\frac{1}{3}$ (c) π (d) $-\frac{1}{3}$ (e) $\log\frac{27}{16}$ **9.** $\frac{9}{128}$

Chapter XII, §3

1. $(e-1)\pi$ **2.** $3\pi/2$ **3.** $\pi(1-e^{-a^2})$ **4.** π **5.** $2ka^4/3$ **6.** $3k\pi a^4/2$
7. $k\pi/4$ **8.** πa^2 **9.** $\pi a^4/8$ **10.** $a^3\sqrt{2}/6$ **11.** $a^2(\pi+8)/4$
12. $a^3(15\pi+32)/24$ **13.** $2a^2$

14. $\dfrac{8\sqrt{2}}{3}a^3\left(-\dfrac{1}{3}+\dfrac{\pi}{4}\right)$ **15.** $2\pi[-(a^2+1)^{-1/2}+1]$

16. $2\pi\left[-\dfrac{1}{2(a^2+2)}+\dfrac{1}{4}\right]\cdot$ Limit $=\pi/2$

17. $\pi\frac{65}{2592}$ **18.** (a) $-5\pi/4$ (b) $\frac{49}{32}\pi a^4$ **19.** (a) $3\pi/4$ (b) 0 (c) 0

20. (a) $2\pi\left[\dfrac{b^{-n+2}-a^{-n+2}}{-n+2}\right]$ if $n\neq 2$

$$2\pi[\log b-\log a] \quad \text{if} \quad n=2.$$

(b) The integral approaches a limit of $n=0, 1$

Chapter XII, §4

1. $\frac{4}{3}\pi a^3$ **2.** 0 **3.** (a) $ka^4\pi$ (b) $2\pi(1-a^2)$ **4.** $2\pi k(b^2-a^2)$ **5.** $\pi ba^4/4$

6. $k\pi a^4/2$ **7.** $\pi/8$ **8.** $2\pi\left[-\frac{1}{3}(1-r_0^2)^{3/2}+\dfrac{1}{3}-\dfrac{r_0^4}{4}\right]$

$$\text{where } r_0^2=\frac{-1+\sqrt{5}}{2}$$

9. $\frac{2}{9}a^3(3\pi-4)$ **10.** πa^3 **11.** (a) $\pi/3$ (b) $2\pi\sqrt{2}/3$ (c) $\pi/2$ (d) $\pi/32$
12. (a) 25 (b) 15/2 (c) $7a^2b^3/3$ **13.** 64/3

14. (a) $4\pi\left[\dfrac{b^{3-n}-a^{3-n}}{3-n}\right]$ if $n\neq 3$

$4\pi[\log b-\log a]$ if $n=3$.

(b) The integral approaches a limit of $n=0, 1, 2$.

Chapter XII, §5

1. $(1, 5/3)$ **2.** $(5/2, 2)$ **3.** $\left(0, \dfrac{3b}{2\pi}\right)$ **4.** $(1, -415)$ **5.** $(\pi/2, \pi/8)$

6. $\bar{x}=\dfrac{\pi}{2}+\dfrac{\pi\sqrt{2}}{4}-1-\sqrt{2}$, $\bar{y}=\dfrac{\sqrt{2}+1}{4}$

7. $\bar{x}=\dfrac{2a^2\log a-a^2+1}{4(a\log a-a+1)}$, $\bar{y}=\dfrac{a(\log a)^2}{2(a\log a-a+1)}-1$

8. $(0, 0, \frac{3}{4}h)$ **9.** (a) $\frac{20}{3}k\pi$ (b) $\left(\dfrac{21}{10}, \dfrac{96}{25\pi}\right)$

10. (a) $\frac{1}{2}k\pi h^2r^2$ (b) $\frac{2}{3}h$ **11.** (a) $\frac{2}{3}ka^3\pi$ (b) $(0, 0)$ (c) $\left(\dfrac{3a}{2\pi}, \dfrac{3a}{2\pi}\right)$

12. $\frac{1}{2}ka^4h\pi$ **13.** $\left(0, 0, \dfrac{2h}{5}\right)$

Chapter XIII, §1

1. (a) 7 (b) 14 **2.** (a) 14 (b) 1
3. (a) 11 (b) 38 (c) 8 (d) 1 **4.** (a) 10 (b) 22 (c) 11 (d) 0

Chapter XIII, §2

1. πab **2.** $\frac{4}{3}\pi abc$ **3.** (a) $29^{3/4}$ (b) $r^{3/4}$ **4.** (a) $33^{3/5}$ (b) $r^{3/5}$

Chapter XIII, §3

1. π **2.** (a) $\frac{128}{3}$ (b) 0 **3.** (a) 42 (b) 120 **4.** 2 **5.** $\frac{1}{2}$ **6.** πab **7.** $\frac{99}{2}$
8. 1500π **9.** $15\pi ab$ **11.** 0

Chapter XIII, §4

1. (a) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & \rho\cos\varphi & \sin\varphi \\ 0 & -\rho\sin\varphi & \cos\varphi \end{pmatrix}$ and determinant is ρ.

(b) $$\iiint_A f(G(\theta, \varphi, \rho))\rho \, d\rho \, d\varphi \, d\theta = \iiint_{G(A)} f(\theta, r, z) \, dz \, dr \, d\theta$$

2. $abck$ **3.** $\frac{4}{3}\pi abc$ **4.** $\frac{4}{3}\pi a^3 \cdot 14$
5. (a) $\frac{1}{6}$ (c) $\frac{1}{3}$ (d) $\frac{1}{3}$ **6.** (a) $\frac{7}{6}$ (b) $\frac{3}{2}$

Chapter XIV, §1

1. (a) -4 (b) 4 (c) 4π (d) π (e) 8 (f) πab **3.** 0

Chapter XV, §1

1. $$\frac{\partial X}{\partial \theta} = (-(a + b\cos\varphi)\sin\theta, (a + b\cos\varphi)\cos\theta, 0)$$

$$\left\|\frac{\partial X}{\partial \theta}\right\| = a + b\cos\varphi$$

$$\frac{\partial X}{\partial \varphi} = ((a - b\sin\varphi)\cos\theta, (a - b\sin\varphi)\sin\theta, b\cos\varphi)$$

$$\left\|\frac{\partial X}{\partial \varphi}\right\| = \sqrt{a^2 + b^2 - 2ab\sin\varphi}$$

2. $$\frac{\partial X}{\partial \theta} = (-z\sin\alpha\sin\theta, z\sin\alpha\cos\theta, 0)$$

$$\frac{\partial X}{\partial z} = (\sin\alpha\cos\theta, \sin\alpha\sin\theta, \cos\alpha)$$

$$\frac{\partial X}{\partial \theta} \times \frac{\partial X}{\partial z} = (z\sin\alpha\cos\theta\cos\alpha, z\sin\alpha\sin\theta\cos\alpha, -z\sin^2\alpha)$$

$$\left\|\frac{\partial X}{\partial \theta} \times \frac{\partial X}{\partial z}\right\| = z\sin\alpha$$

Equation of surface is $x^2 + y^2 = (\tan\alpha)^2 z^2$.

3. $\dfrac{\partial X}{\partial t} = (a\cos\theta,\ a\sin\theta,\ 2t)$

$$\frac{\partial X}{\partial \theta} = (-at\sin\theta,\ at\cos\theta,\ 0)$$

$$\frac{\partial X}{\partial t} \times \frac{\partial X}{\partial \theta} = (-2at^2\cos\theta,\ -2at^2\sin\theta,\ a^2t)$$

$$\left\|\frac{\partial X}{\partial t} \times \frac{\partial X}{\partial \theta}\right\| = \sqrt{4a^2t^4 + a^4t^2}$$

The equation is $x^2 + y^2 = a^2z$.

4. $\dfrac{\partial X}{\partial \varphi} = (a\cos\varphi\cos\theta,\ b\cos\varphi\sin\theta,\ -c\sin\varphi)$

$$\frac{\partial X}{\partial \theta} = (-a\sin\varphi\sin\theta,\ b\sin\varphi\cos\theta,\ 0)$$

$$\frac{\partial X}{\partial \varphi} \times \frac{\partial X}{\partial \theta} = (cb\sin^2\varphi\cos\theta,\ ac\sin^2\varphi\sin\theta,\ ab\sin\varphi\cos\varphi)$$

$$\left\|\frac{\partial X}{\partial \varphi} \times \frac{\partial X}{\partial \theta}\right\| = \sqrt{c^2b^2\sin^4\varphi^4\cos^2\theta + a^2c^2\sin^4\varphi\sin^2\theta + a^2b^2\sin^2\varphi\cos^2\varphi}$$

The equation is $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1.$

5. $\dfrac{\partial X}{\partial \theta} = (-a\sin\theta,\ a\cos\theta,\ 0)$

$$\frac{\partial X}{\partial z} = (a\cos\theta,\ a\sin\theta,\ 1)$$

$$\frac{\partial X}{\partial \theta} \times \frac{\partial X}{\partial z} = (a\cos\theta,\ a\sin\theta,\ a^2)$$

$$\left\|\frac{\partial X}{\partial \theta} \times \frac{\partial X}{\partial z}\right\| = \sqrt{a^2 + a^4}$$

The equation is $x^2 + y^2 = a^2$

6. $\dfrac{\partial X}{\partial r} = (\cos\theta,\ \sin\theta, f'(r))$

$$\frac{\partial X}{\partial \theta} = (-r\sin\theta,\ r\cos\theta,\ 0)$$

$$\frac{\partial X}{\partial r} \times \frac{\partial X}{\partial \theta} = (-f'(r)r\cos\theta,\ -f'(r)r\sin\theta,\ r)$$

$$\left\|\frac{\partial X}{\partial r} \times \frac{\partial X}{\partial \theta}\right\| = r\sqrt{f'(r)^2 + 1}$$

Equation is $z = f(\sqrt{x^2 + y^2})$.

Chapter XV, §*2*

1. (a) $\pi\sqrt{2}$ (b) $\frac{10}{9}\pi h^2$ **2.** $\dfrac{\pi}{6}(5\sqrt{5}-1)$ **3.** $2\pi(\sqrt{3}-\frac{1}{3})$

4. $\frac{2}{3}\pi(2\sqrt{2}-1)$ **5.** $\frac{1}{8}e^{2\,\text{arcsinh}\,1}-\frac{1}{8}e^{-2\,\text{arcsinh}\,1}+\frac{1}{2}\sinh 1$

6. $4\sqrt{6}$ **7.** $2\sqrt{2}\pi$ **8.** $2\pi(1-\sqrt{2}/2)$ **9.** $4\pi^2 a$

Chapter XV, §*3*

1. $4\pi a^4/3$ **2.** $\pi a^5/2$ **3.** $4\pi a^6/15$ **4.** $\pi a^7/3$ **5.** $\dfrac{\pi}{60}(25\sqrt{5}-11)$

6. 0 **7.** 0 **8.** πa^3 **9.** $\pi a^4/2$ **12.** $4\pi/3$ **13.** $\pi^2/4+2\pi$ **14.** 4π

15. $104/3$ **16.** $2\pi\sqrt{2}$ **17.** $\dfrac{\pi}{12}(8-5\sqrt{2})$ **18.** $5/12$

19. (a) $2\pi a^2$ (b) $3\pi a^2$ **20.** $3/2$ **21.** $5\pi/4$

Chapter XV, §*4*

1. $\nabla\cdot F = 2x + xz + 2yz$
$\nabla\times F = (z^2 - xy, 0, yz)$

2. $\nabla\cdot F = \dfrac{y}{x}+\dfrac{x}{y}+\dfrac{xy}{z}$
$\nabla\times F = (x\log z, -y\log z, \log y - \log x)$

3. $\nabla\cdot F = 2x + x\cos xy + e^x y$
$\nabla\times F = (e^x z, -e^x yz, y\cos xy)$

4. $\nabla\cdot F = ye^{xy}\sin z + e^{xz}\cos y + ye^{yz}\cos x$
$\nabla\times F = (ze^{yz}\cos x - xe^{xz}\sin y, e^{xy}\cos z + e^{yz}\sin x,$
$ze^{xz}\sin y - xe^{xy}\sin z)$

Chapter XV, §*5*

2. $3/2$ **3.** 0 **4.** 64π **5.** 0 **6.** 0 **7.** 16 **8.** 24 **9.** 24π **10.** 48π
11. $243\pi/2$ **12.** 135π **13.** $1/40$

Chapter XV, §*6*

1. 0 **2.** $-13/6$ **3.** 0 **4.** 0 **5.** 0 **6.** 0 **7.** $-\pi a^2$

Appendix, §*1*

1. $$\int_{-\pi}^{\pi} cf(x)\,dx = c\int_{-\pi}^{\pi} f(x)\,dx$$

and

$$\langle f, g+h\rangle = \int_{-\pi}^{\pi} f(x)[g(x)+h(x)]\,dx = \int_{-\pi}^{\pi}[f(x)g(x)+f(x)h(x)]\,dx$$
$$= \int_{-\pi}^{\pi} f(x)g(x)\,dx + \int_{-\pi}^{\pi} f(x)h(x)\,dx = \langle f, g\rangle + \langle f, h\rangle.$$

2. Take the scalar product with f_i. We obtain for each i,

$$0 = \langle c_1 f_1 + \cdots + c_n f_n, f_i \rangle = \sum_{k=1}^{n} c_k \langle f_k, f_i \rangle = c_i.$$

3. If $\langle h_1, f \rangle = 0$ and $\langle h_2, f \rangle = 0$, then

$$\langle h_1 + h_2, f \rangle = \langle h_1, f \rangle + \langle h_2, f \rangle = 0.$$

If c is a number and $\langle h, f \rangle = 0$, then $\langle ch, f \rangle = c\langle h, f \rangle = 0$.

4. $$\left| \int_{-\pi}^{\pi} f(x)g(x)\, dx \right| \leqq \left(\int_{-\pi}^{\pi} f(x)^2\, dx \right)^{1/2} \left(\int_{-\pi}^{\pi} g(x)^2\, dx \right)^{1/2}$$

$$\left(\int_{-\pi}^{\pi} [f(x) + g(x)]^2\, dx \right)^{1/2} \leqq \left(\int_{-\pi}^{\pi} f(x)^2\, dx \right)^{1/2} + \left(\int_{-\pi}^{\pi} g(x)^2\, dx \right)^{1/2}$$

7. (b) 1/4 (c) $\|f\| = \dfrac{\sqrt{3}}{3}$ and $\|g\| = \dfrac{\sqrt{5}}{5}$ (d) 1/2, 1/3, 1

Appendix, §2

1. (a) $\sqrt{5\pi}$ (b) $(\pi + \pi^3/3)^{1/2}$

4. (a) $\dfrac{x}{2} = \sin x - \dfrac{\sin 2x}{2} + \cdots + (-1)^{n+1} \dfrac{\sin nx}{n} + \cdots$

(b) $x^2 = \dfrac{\pi^2}{3} - 4\left(\cos x - \dfrac{\cos 2x}{2^2} + \cdots + (-1)^{n+1} \dfrac{\cos nx}{n^2} + \cdots \right)$

(c) $|x| = \dfrac{\pi}{2} - \dfrac{4}{\pi}\left(\cos x + \dfrac{\cos 3x}{3^2} + \cdots + \dfrac{\cos (2n+1)x}{(2n+1)^2} + \cdots \right)$

(f) $|\cos x| = \dfrac{4}{\pi}\left(\dfrac{1}{2} + \dfrac{\cos 2x}{3} + \cdots + (-1)^{n-1} \dfrac{\cos 2nx}{4n^2 - 1} + \cdots \right)$

(g) $\sin^3 x = \frac{3}{4} \sin x - \frac{1}{4} \sin 3x$

(d) $\dfrac{1}{2} - \dfrac{\cos 2x}{2}$

(e) $|\sin x| = \dfrac{4}{\pi}\left(\dfrac{1}{2} - \dfrac{\cos 2x}{3} - \cdots - \dfrac{\cos 2nx}{4n^2 - 1} - \cdots \right)$

(h) $\cos^3 x = \frac{3}{4} \cos x + \frac{1}{4} \cos 3x$

Index

ABCDEFGH79876543