LINEAR ALGEBRA

SECOND EDITION

SERGE LANG
Columbia University, New York, New York

ADDISON-WESLEY PUBLISHING COMPANY
Reading, Massachusetts · Menlo Park, California · London · Sydney · Manila

WORLD STUDENT SERIES EDITION

SECOND PRINTING 1972

A complete and unabridged reprint of the original American textbook, this World Student Series edition may be sold only in those countries to which it is consigned by Addison-Wesley or its authorized trade distributors. It may not be re-exported from the country to which it has been consigned, and it may not be sold in the United States of America or its possessions.

Library of Congress Catalog Card Number: 70-129-520

Foreword

The present book is meant as a text for a course in linear algebra, at the undergraduate level. Enough material has been included for a one-year course, but by suitable omissions, it will also be easy to use the book for one term.

During the past decade, the curriculum for algebra courses at the undergraduate level has shifted its emphasis towards linear algebra. The shift is partly due to the recognition that this part of algebra is easier to understand than some other parts (being less abstract, and in any case being directly motivated by spatial geometry), and partly because of the wide applications which exist for linear algebra. Consequently, I have started the book with the basic notion of vector in real Euclidean space, which sets the general pattern for much that follows. The chapters on groups and rings are included because of their important relation to the linear algebra, the group of invertible linear maps (or matrices) and the ring of linear maps of a vector space being perhaps the most striking examples of groups and rings. The fact that a vector space over a field can be viewed fruitfully as a module over its ring of endomorphisms is worth emphasizing as part of a linear algebra course. However, because of the general intent of the book, these chapters are not treated with quite the same degree of completeness which they might otherwise receive, and a short text on basic algebraic structures (groups, rings, fields, sets, etc.) will accompany this one to offer the opportunity of teaching a separate one-term course on these matters, principally intended for mathematics majors.

The tensor product, and especially the alternating product, are so important in courses in advanced calculus that it was imperative to insert a chapter on them, keeping the applications in mind. The limited purpose of the chapter here allows for concreteness and simplicity.

The appendix on convex sets pursues some of the geometric ideas of Chapter I, taking for granted some standard facts about continuous functions on compact sets, closures of sets, etc. It can essentially be read after Chapter I, and after knowing the definition of a linear map. Various odds and ends are given in a second appendix (including a proof of the algebraic closure of the complex numbers), which can be covered according to the judgement of the instructor.

The basic portion of this book, on vector spaces, matrices, linear maps, and determinants is now published separately as *Introduction to Linear Algebra*, with additional simplifications of language and text. For instance, we take vector spaces over the reals, we consider only the positive definite scalar product, we omit the dual space, etc., which are less worthy of emphasis for a first introduction, needed in immediate applications, e.g. in calculus. In the more complete text of a full course in linear algebra, these topics are of course included, as are many others, especially the structure theorems which form Part Two: spectral theorem, for symmetric, hermitian, unitary operators; triangulation theorems (including the Jordan normal form); primary decomposition; Schur's lemma; the Wedderburn-Rieffel theorem (with Rieffel's beautifully simple proof); etc. Of course, better students can handle the more complete book at once, but I hope that the separation will be pedagogically useful for others.

In this second edition, I have rewritten a few sections, and inserted a few new topics. I have also added many new exercises.

New York, 1970 SERGE LANG

Contents

Part One
Basic Theory

Chapter I

Vectors

Chapter II

Vector Spaces

Chapter III

Matrices

Chapter IV

Linear Mappings

Chapter V

Linear Maps and Matrices

Chapter VI

Scalar Products and Orthogonality

Chapter VII

Determinants

Part Two
Structure Theorems

Chapter VIII

Bilinear Forms and the Standard Operators

Chapter IX

Polynomials and Matrices

Chapter X

Triangulation of Matrices and Linear Maps

Chapter XI

Spectral Theorem

Chapter XII

Polynomials and Primary Decomposition

Part Three
Relations with Other Structures

Chapter XIII

Multilinear Products

Chapter XIV

Groups

Chapter XV

Rings

Appendix 1

Convex Sets

APPENDIX 2

Odds and Ends

APPENDIX 3

PART ONE

BASIC THEORY

This page intentionally left blank

CHAPTER I

Vectors

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

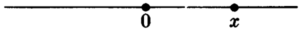
The cross product is included for the sake of completeness. It is almost never used in the rest of the book. It is the only aspect of the theory of vectors which is valid only in 3-dimensional space (not 2, nor 4, nor n-dimensional space). One significant feature of almost all the statements and proofs of this book (except for those concerning the cross product and determinants), 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*

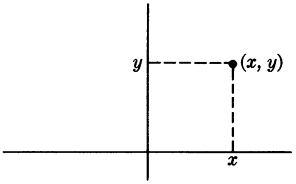
We know that a number can be used to represent a point on a line, once a unit length is selected.

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

These representations can be represented in a picture as follows.



(a) Point on a line



(b) Point in a plane

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 following picture illustrates this.

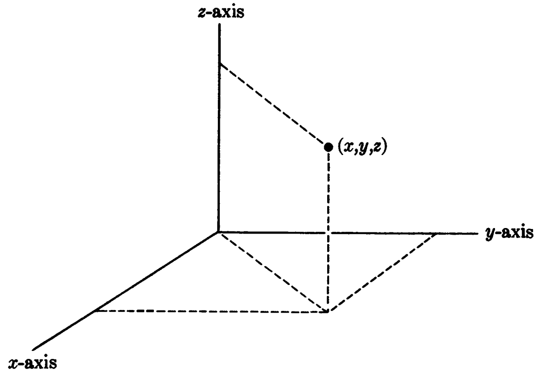


Figure 2

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

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

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

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

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

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

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 $O = (0, \dots, 0)$ 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 $A = (a_1, \dots, a_n)$ and $B = (b_1, \dots, b_n)$. Then

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

and

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

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 $(0, \dots, 0)$. 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).

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

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

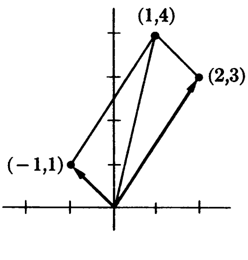


Figure 3

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

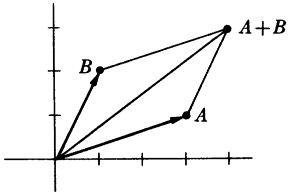


Figure 4

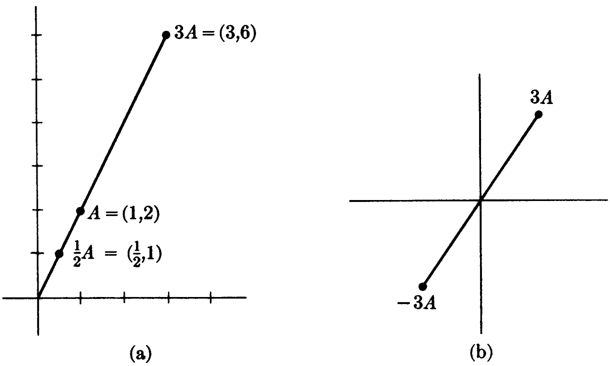


Figure 5

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).

Exercises

Find $A + B$, $A - B$, $3A$, $-2B$ in each of the following cases.

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. Draw the points of Exercises 1 through 4 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.

§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).

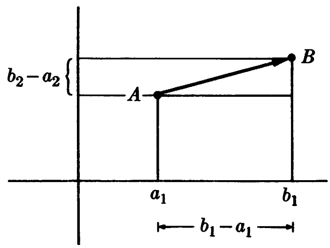


Figure 6

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 $\overrightarrow{O(A - B)}$, $\overrightarrow{BA}$.

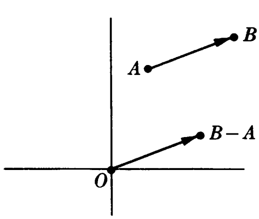


Figure 7

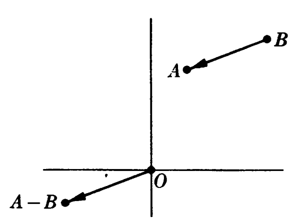
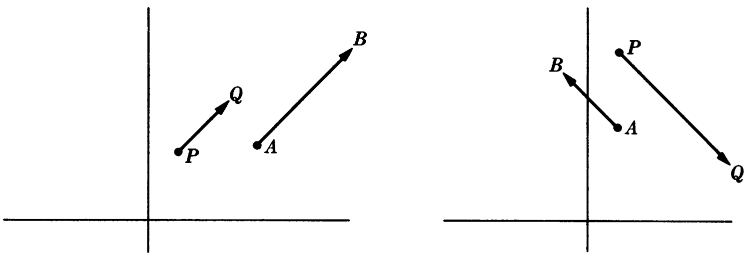


Figure 8

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.



(a) Same direction (b) Opposite direction

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.

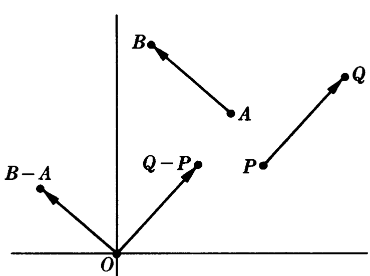


Figure 10

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*

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

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