DAVID M. BLOOM

Linear Algebra and Geometry

Linear algebra and geometry

DAVID M. BLOOM
Associate Professor of Mathematics
Brooklyn College of the City University of New York

CAMBRIDGE UNIVERSITY PRESS
CAMBRIDGE
LONDON · NEW YORK · MELBOURNE

Published by the Syndics of the Cambridge University Press
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
Bentley House, 200 Euston Road, London NW1 2DB
32 East 57th Street, New York, NY 10022, USA
296 Beaconsfield Parade, Middle Park, Melbourne 3206, Australia

First published 1979

Printed in the United States of America
Typeset by Offset Composition Services, Inc., Washington, D.C.
Printed and bound by R. R. Donnelley & Sons Company,
Crawfordsville, Ind.

Library of Congress Cataloging in Publication Data

Bloom, David M., 1936–

Linear algebra and geometry.

Includes index.

1. Algebras, Linear. 2. Geometry. I. Title.
QA184.B59 512'.5 77-26666

ISBN 0 521 21959 0 hard covers
ISBN 0 521 29324 3 paperback

Preface

In the spring semester of 1966, I taught two undergraduate courses at Brooklyn College: an "advanced elective" course in linear algebra, which I had taught before; and a course entitled "Higher Analytic Geometry," which I had not taught before. I was aware from the beginning of the semester, of course, that connections existed between the subject matter of the first course and that of the second, having first been exposed to such connections in a major way as a student in the beautiful course taught by O. Zariski at Harvard in 1957–58; but I was not fully aware of the extent of these connections. Thus, it came as something of a revelation to me to discover, as the term progressed, that *every* topic which we covered in the first course (literally without exception) was *significantly* applicable to the second course.

At this point I asked myself: if linear algebra and geometry can be so well integrated mathematically, why not integrate them pedagogically? Specifically, instead of two one-term courses, why not teach one full-year course in which the relations between linear algebra and geometry could be explored to their fullest extent? Here, it seemed, might lie an opportunity to illustrate the unity of mathematics and to counteract the prevalent tendency toward compartmentalization of knowledge.

It is in the spirit of the preceding remarks that I decided to write this book. There are other books available whose object is similar (Jaeger's *Introduction to Analytic Geometry and Linear Algebra*, for example), but they are still sufficiently few in number so that I feel no need for an excuse to contribute another one to the collection. To confess the truth, I am really writing this book for myself; every teacher eventually develops his or her own pet approaches to this or that topic, and the urge to self-expression is strong. However, because self-expression requires an audience to be fully satisfying, let me describe the probable audience for this book: students at the advanced undergraduate level who have already had a one-term introduction to abstract ("modern") algebra, and *have understood the latter reasonably well.* The necessary algebraic background is fully summarized in Chapter 1, but this chapter is intended mainly for reference and review; the student who is not already familiar with the concepts of group, ring, field, homomorphism, equivalence relation, and so on, is urged not only to read Chapter 1 carefully before going further, but also to consult one of the good books on abstract algebra (for example, [10] or [11]) for further details.

Some of the material in Chapter 1 will probably be new even to those readers (hopefully the majority of readers) who have already taken a course in abstract algebra. One may, if desired, postpone such material until it is needed. (For example, Theorem 1.3.6 is not needed until Section 4.7, and in that section a specific reference to Theorem 1.3.6 is given so that one will know where to look.) However, it is recommended that the reader spend at least some time with Chapter 1 to refresh his or her memory of abstract concepts, before going on to Chapter 2, which is the "real" beginning of the book.

Mathematically, this book is self-contained with four exceptions. *First exception*: Some of the results in Sections 1.12 (the integers), 1.15 (properties of rational, real, and complex numbers), and 1.16 (properties of polynomials) are stated without proof. (For example, I felt that this was not the place to construct the real number system; thus, the existence of such a system is assumed.) However, references are given in each of the above-mentioned sections to books in which the missing proofs (including constructions) may be found. *Second exception*: Certain concepts and results external to algebra, whose full development would be impractical in this book, are nevertheless needed for the treatment of some specific topics that I wished to include, namely: applications to area and volume (Sections 5.7 and 5.14), a treatment of rigid motions and orientation using the concept of continuity (Sections 6.6 and 6.7), and the use of determinants in analysis (Section 9.8 and Appendix E). These applications of algebra to other fields are ends in themselves; their deletion would result in no harm to the self-contained structure of the book (though the *spirit* of the book would suffer; this sort of thing is what makes mathematics interesting). *Third exception* (perhaps this is a subcase of the second exception): The proof of the main result of Section 7.6 depends on the fact that a real polynomial function of n variables is continuous. *Fourth exception*: Some of the proofs are left as exercises for the reader. Those exercises whose results are needed later are marked by asterisks.

With the exceptions just noted, any result that is used is proved. So as not to intimidate the reader, a few of the most formidable proofs are relegated to appendixes. There is both a pedagogical and a philosophical reason to include proofs of theorems in a textbook. The pedagogical reason is that after working one's way through a certain number of proofs, one begins to recognize patterns, methods, styles of proof and eventually develops skill in constructing proofs oneself. The philosophical reason is that in the presence of proof, one is freed from *dependence on authority*. It is not too good an idea to believe everything you're told, whether the teller be a politician, a newspaper, someone you love – or even a textbook on linear algebra and geometry. We, too, make mistakes; more than one statement made in a textbook (or even in a prestigious professional journal) has turned out to be false. The reader, coming across any given statement in this or another mathematics textbook, will perhaps assume that the statement has a high probability (say, 99.97%) of being true; but if he or she can verify the statement by following a printed proof, step-by-step, the probability increases. (One recalls the old advertisement for the multisection Sunday edition of the *New York Times*: "You don't have to read it all – but it's nice to know it's all there.")

A few words about my approach to geometry. In Chapter 2, all of the geometry is informal; we use it to illustrate and motivate the algebraic concepts of the chapter. Starting with Chapter 3, the approach changes: geometry is made formal, rigorous, algebraic. However, formal definitions are usually preceded by informal discussions for the purpose of motivation, and I have tried to formulate such definitions, whenever possible, so as to parallel the reader's previous geometric experience. For example, to define $\cos(\mathbf{u},\mathbf{v})$ as $(\mathbf{u}\cdot\mathbf{v})/|\mathbf{u}||\mathbf{v}|$ seems unnecessarily artificial when the right-triangle approach can be made rigorous. (Not *all* such artificialities are avoidable; e.g., distance is defined via the square-root formula. We have to start *somewhere*.)

The major geometric areas covered in this book are affine geometry, orthogonal (Euclidean) geometry, isometries, and quadric surfaces; other topics present but receiving less emphasis include area and volume and a very brief introduction to projective geometry (in an affine context). The n-dimensional case is considered throughout, both for the sake of generality and as a means of unifying the treatment.

For use in a two-semester undergraduate course, Chapters 2 through 7, excluding Sections 7.6 and 7.7, should suffice for all but the best students. Sections 5.14, 6.6, and 6.7 could be omitted without loss of continuity, but the material in these sections is likely to interest the student and should be covered if time permits. Sections 7.6 and 7.7 and Chapters 8 and 9 are more abstract than what precedes them and require a greater degree of mathematical maturity of the reader.

Even in mathematics there can be different paths to the same goal. Thus, I try now and then to present more than one proof of a theorem, more than one way of solving a problem, more than one way of defining something. Sometimes, these alternate approaches appear in the exercises.

Numerical examples appear throughout. Andrew Gleason, under whom I took two graduate courses, used to tell us, "You're no good if you can't compute." I agree.

Considerable use is made of the "equivalence relation" concept. For instance, parallelism is an equivalence relation among lines; an "angle" is an equivalence class of pairs of vectors (or pairs of rays); an "orientation" in real n-space is defined to be an equivalence class of orthonormal n-tuples.

Numbering (except in appendixes) is by chapter, section, and item, with the figures numbered separately from other items. Thus, for example, the first two numbered items in Chapter 3, Section 4 (other than figures) are Definition 3.4.1 and Theorem 3.4.2; the first two figures in this same section are Figure 3.4.1 and Figure 3.4.2. "1.11.1(d)" refers to the fourth part of item 1.11.1.

At the end of almost every section, exercises appear. The abbreviation SOL following an exercise means that a full or partial solution to the exercise appears in Appendix G at the end of the book. The abbreviation ANS means that an answer (but not the steps leading to the answer) appears in the appendix; the abbreviation SUG means that a mere hint or suggestion appears. As noted before, an asterisk beside an exercise means that the result of the exercise will be needed later.

The Halmos symbol ∎ is used throughout the book to indicate the end of a proof. The abbreviation R.A.A. (reductio ad absurdum) appears in several proofs and means "contradiction". (I do not understand why the latter abbreviation is so seldom used in mathematical writing; it seems no more abstruse than Q.E.D.)

Of the courses I have taken and the texts from which I have taught, so many have influenced this book that I could not possibly list them all. However, I wish to express special gratitude to Professor Robert Taylor, formerly of Columbia University, who first interested me in

algebra; to Oscar Zariski, whose course in geometry was a revelation to me; to Richard Brauer, whose advice and encouragement during my struggle with a doctoral thesis (and subsequently) was invaluable; to the late Moses Richardson of Brooklyn College, who encouraged me to start this book; to Melvin Hausner, whose own book served as a model in some respects; to my students, who have stimulated and challenged me and have provided more than one of the ideas that I have used in the book; and to my family, who patiently did without me for too many days and nights.

September, 1978 David M. Bloom

Contents

1 Background in abstract algebra

1.0 Introduction

This chapter is for the purpose of reference and review. It summarizes those fundamentals of abstract algebra that we will need later, and largely parallels a standard first course in abstract algebra. If you have already taken such a course, you probably should start right in with Chapter 2, though you will find it necessary to refer back to this chapter from time to time. The proofs are all here (except in Sections 1.15 and 1.16, where references to the literature replace proofs); examples are fewer in number than in an abstract algebra text per se, but I have tried to explain concepts clearly. If supplementary reading is desired, either of the references [10] or [11] listed in the References should prove helpful.

1.1 Sets, elements, basic notations

You probably have some idea of what is meant by a "set" and by an "element" of a set. For example, we can speak of "the set of all integers from 1 to 10"; the elements of this set are the integers from 1 to 10. Similarly, if we speak of "the set of all Presidents of the United States from 1955 to 1965", this set has three elements, namely Eisenhower, Kennedy, Johnson. These examples do *not* constitute a formal definition; in fact, we cannot really give a formal definition at all. A concept can be defined only in terms of simpler concepts, and there are no concepts logically simpler than those of "set" and "element". Thus we must formally treat the words "set" and "element" as *undefined terms*. However, this should not prevent us from believing that we know what the words mean.

Sets are often denoted by capital letters, and elements of sets by small letters. (There are exceptions.) The set of all integers is commonly denoted $\mathbf{Z}$; we shall do so throughout this book. The symbolic expression

$$x \in A$$

means "x is an element of (the set) A" (or some grammatical modification thereof; for instance, the sentence, "Let $x \in A$" is of course to be read, "Let x be an element of A", not "Let x is an element of A"). Phrases like "x belongs to A", "x is in A", "A contains x", etc., mean the same thing as "x is an element of A."

When we wish to specify the elements of a set, we enclose them in brackets which look like { }; thus, for example, if A is the set of all integers from 1 to 16, we may write

$$A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16\}.$$

If we do not feel like doing so much writing, an alternative is

$$A = \{x \in \mathbf{Z} : 1 \leqslant x \leqslant 16\},$$

which can be read, "A is the set consisting of *all* elements x of $\mathbf{Z}$ (that is, all integers x) such that $1 \leqslant x \leqslant 16$". (The colon means "such that".) In short, the elements of a set may be specified either by listing them or by describing them in terms of one or more properties. As a further illustration, if B is the set of all *even* integers from 1 to 10, then

$$B = \{2, 4, 6, 8, 10\} = \{x \in \mathbf{Z} : 1 \leqslant x \leqslant 10, x \text{ is even}\}.$$

The symbol $\varnothing$ denotes the *empty set*, a set having no elements. It is important not to confuse $\varnothing$ (which has no elements) with $\{\varnothing\}$ (which has one element, namely $\varnothing$).

Two sets are equal (i.e., the same) if they have the same elements; thus, to prove that the sets A and B are equal it suffices to show that (1) every element of A belongs to B, and (2) every element of B belongs to A. If (1) is true (whether (2) is true or not), A is called a *subset* of B (*notation*: $A \subseteq B$), and B is called an *overset*, or *superset*, of A. The notation $A \subset B$ means that A is a *proper* subset of B; that is, A is a subset of B but is not equal to B. (In this situation, we naturally call B a *proper overset* of A.) For example, $\{0, 2\}$ is a proper subset of $\{0, 1, 2, 3\}$; $\{3, 5, 7, 9\}$ is a proper overset of $\{3, 7\}$; $\{3\} \subset \{3, 5\}$; $\{1\} \subseteq \{1\}$ but not $\{1\} \subset \{1\}$.

A diagonal line through one of the symbols $=, \in, \subset, \subseteq, <, >$ (etc.) means "is not"; thus, for example, $3 \neq 5$; $7 \notin \{1, 2, 3, 4, 5, 6\}$; $\{3, 4, 5\} \not\subseteq \{3, 4\}$; $10 \not< 8$.

If A and B are sets, then $A \cup B$ (the *union* of A and B) is the set whose elements are the elements of A together with the elements of B. More generally, if $A_1, A_2, \ldots, A_n$ are sets, then the union of these n sets, denoted either $A_1 \cup \cdots \cup A_n$ or

$$\bigcup_{i=1}^{n} A_i,$$

is defined to be the set consisting of all elements which belong to at least one of the sets A_i. For example,

$$\{1, 2, 4, 6\} \cup \{2, 8\} \cup \{1, 7\} = \{1, 2, 4, 6, 7, 8\},$$

and if, say, $A_1 = \{1\}, A_2 = \{2\}, \ldots, A_n = \{n\}$, then

$$\bigcup_{i=1}^{5} A_i = \{1, 2, 3, 4, 5\}.$$

The *intersection* of two sets A and B (denoted $A \cap B$) is the set whose elements are the elements common to A and B. Similar definitions apply to more than two sets. For example,

$$\{1, 3, 5\} \cap \{3, 5, 7\} = \{3, 5\}$$
$$\{2, 3, 5\} \cap \{2, 5, 6\} \cap \{2, 3, 6\} = \{2\}$$
$$\{1, 2\} \cap \{3, 4\} = \varnothing.$$

Two sets are *disjoint* if they have empty intersection; that is, if they have no elements in common.

The sets $\{2, 3\}$, $\{3, 2\}$, and $\{2, 3, 2\}$ are equal, since they have the same elements; the order in which we write the elements does not matter, nor does it matter if some element is listed more than once. In some situations, however, such distinctions matter. If, for example, we assign coordinates (x, y) to points in the plane (as in elementary analytic geometry), the points (2, 3) and (3, 2) are different; similarly, the point (2, 3) in the plane is not the same as the point (2, 3, 2) in 3-space. Expressions like (2, 3) and (3, 2), written with ordinary parentheses instead of brackets, are called *ordered pairs*; expressions like (2, 3, 2) are called *ordered triples*. More generally, if n is any positive integer, then an expression of the form $(a_1, a_2, \ldots, a_n)$ is called an *ordered n-tuple*. (The word "ordered" is occasionally omitted to save space.) In the n-tuple $(a_1, a_2, \ldots, a_n)$, a_1 is called the *first coordinate*, a_2 is called the *second coordinate*, and so on. Two ordered n-tuples are equal only if they have the same first coordinate, the same second coordinate, ..., the same nth coordinate.

If A and B are sets, then $A \times B$ is the standard notation for the set of all ordered pairs (a, b) such that $a \in A$ and $b \in B$. For example, if $A = \{1, 2\}$ and $B = \{a, b, c\}$, then

$$A \times B = \{(1, a), (1, b), (1, c), (2, a), (2, b), (2, c)\}.$$

Observe that in this example A has two elements, B has three elements, $A \times B$ has six elements. Does this suggest to you a general result concerning the number of elements in $A \times B$? Does it suggest a reason for the notation "$A \times B$"?

1.2 Mappings

Let A, B be sets. A *mapping of A into B* is (to be informal) a rule that assigns to each element $x \in A$ exactly one element $x' \in B$. A mapping of A into B is normally denoted by a single letter, such as $\mathbf{T}$; the notation

$$\mathbf{T} : A \to B$$

is read, "$\mathbf{T}$ is a mapping of A into B". (The set A is then called the *domain* of $\mathbf{T}$.) If the mapping $\mathbf{T}$ assigns to the element $a \in A$ the element $b \in B$, then b is called the *image* of a under $\mathbf{T}$, or (more concisely) the $\mathbf{T}$-*image* of a; and we say that "a is mapped into b by $\mathbf{T}$". In this situation, we write

(1.2.1) $$a\mathbf{T} = b$$

or

$$\mathbf{T} : a \to b.$$

If b is the image of a under $\mathbf{T}$, then a is said to be a *pre-image* of b under $\mathbf{T}$. Note that although an element of A has exactly one image in B, an element $b \in B$ may have any number of pre-images in A; for that matter, it may have no pre-images in A.

The type of notation exhibited in 1.2.1 is called *right-hand notation*. Some authors prefer *left-hand notation*, in which one writes $\mathbf{T}(a) = b$ instead of $a\mathbf{T} = b$. (Left-hand notation is almost universal in elementary calculus textbooks, in which mappings are called "functions" and the sets A and B are usually intervals of real numbers.) Unfortunately, mathematicians as a body have never been able to agree on which hand to use. This is not entirely the fault of mathematicians; each notation is at times more convenient to use than the other, depending on the context. In this book, we shall usually use the right-hand notation 1.2.1, but exceptions will occur.

If A_0 is any subset of A and if $\mathbf{T} : A \to B$, then we use the notation $A_0\mathbf{T}$ for the set of all images (under $\mathbf{T}$) of elements of A_0. That is,

$$A_0\mathbf{T} = \{x\mathbf{T} : x \in A_0\}.$$

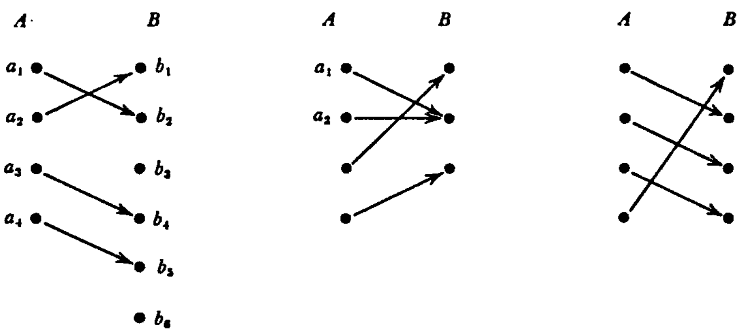
For example, if $A = B = \mathbf{Z}$ and the mapping $\mathbf{T} : \mathbf{Z} \to \mathbf{Z}$ is defined by

$n\mathbf{T} = n^2$, and if $A_0 = \{2, 5, 7\}$, then $A_0\mathbf{T} = \{4, 25, 49\}$. By analogy with our terminology regarding single elements, the set $A_0\mathbf{T}$ is called the *image* of the set A_0 under $\mathbf{T}$.

If $\mathbf{T} : A \to B$, then the set $A\mathbf{T}$ is called the *range* of $\mathbf{T}$; it consists of all elements of B which are images of elements of A. The range of $\mathbf{T}$ is a subset of B; it may or may not be the whole set B. Both possibilities are illustrated in Figure 1.2.1. In each part of Figure 1.2.1, dots denote distinct elements of A and B, and arrows indicate the correspondence between each element of A and its image in B. In part (a) of the figure, every element of A has exactly one image in B (as required), but two of the elements of B, namely, b_3 and b_6, have no pre-images in A; the range of $\mathbf{T}$ is, thus, $A\mathbf{T} = \{b_1, b_2, b_4, b_5\}$. If $\mathbf{T} : A \to B$ is a mapping such that *every* element of B has a pre-image in A, then $\mathbf{T}$ is said to be *onto* and we say that $\mathbf{T}$ maps A *onto* B. In other words, $\mathbf{T}$ is onto if and only if $A\mathbf{T} = B$. The mapping shown in Figure 1.2.1(b) is onto; the mapping shown in Figure 1.2.1(a) is not.

In Figure 1.2.1(b), every element of A has exactly one image in B, as required; but the images are not all distinct, since a_1 and a_2 have the same image. A mapping of $A \to B$ is called *one-to-one* (1–1) if no two elements of A have the same image (equivalently, if no element of B has more than one pre-image). Thus in Figure 1.2.1(a), the mapping is one-to-one, but not in Figure 1.2.1(b). To show that a given mapping $\mathbf{T} : A \to B$ is one-to-one, we must show that if $a_1 \neq a_2$ (in A) then $a_1\mathbf{T} \neq a_2\mathbf{T}$; or, equivalently, that if $a_1\mathbf{T} = a_2\mathbf{T}$ then $a_1 = a_2$. The latter is actually the more common method of proof; in a typical proof that a mapping $\mathbf{T}$ is one-to-one, the first sentence reads, "Assume $a_1\mathbf{T} = a_2\mathbf{T}$", and the last sentence reads, "Therefore $a_1 = a_2$".

Figure 1.2.1



one-to-one
not onto
(a)

onto
not one-to-one
(b)

one-to-one
onto
(c)