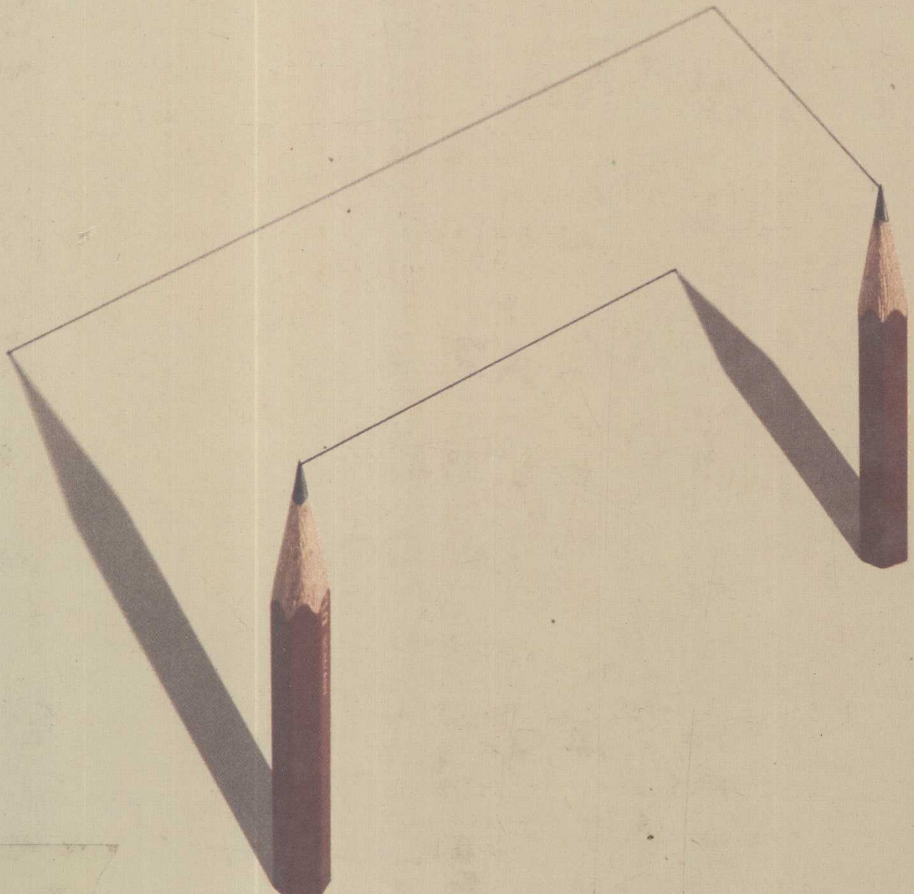
LINEAR ALGEBRA AND ITS APPLICATIONS

GILBERT STRANG



LINEAR ALGEBRA AND ITS APPLICATIONS

THIRD EDITION

GILBERT STRANG
Massachusetts Institute of Technology

HARCOURT BRACE JOVANOVICH, PUBLISHERS
San Diego

The lithographs on the cover were contributed by the distinguished artist Susumu Endo.
Back cover: Space and Space/Book and Pencil
Front cover: Space and Space/Pencil VI
Susumu Endo: 3-13-3 Jingumae, Shibuyaku Tokyo 150, Japan

ISBN: 0-15-551005-3
Library of Congress Catalog Card Number: 87-083019

Printed in the United States of America

PREFACE

Linear algebra is a fantastic subject. On the one hand it is clean and beautiful. If you have three vectors in 12-dimensional space, you can almost see them. A combination like the first plus the second minus twice the third is harder to visualize, but may still be possible. I don't know if anyone can see *all* such combinations, but somehow (in this course) you begin to do it. Certainly the combinations of those three vectors will not fill the whole 12-dimensional space. (I'm afraid the course has already begun; usually you get to read the preface for free!) What those combinations do fill out is something important—and not just to pure mathematicians.

That is the other side of linear algebra. It is *needed* and *used*. Ten years ago it was taught too abstractly, and the crucial importance of the subject was missed. Such a situation could not continue. Linear algebra has become as basic and as applicable as calculus, and fortunately it is easier. The curriculum had to change, and this is now widely accepted as an essential sophomore or junior course—a requirement for engineering and science, and a central part of mathematics.

The goal of this book is to show both sides—the beauty of linear algebra, and its value. The effort is not all concentrated on theorems and proofs, although the mathematics is there. The emphasis is less on rigor, and much more on understanding. *I try to explain rather than to deduce.* In the book, and also in class, ideas come with examples. Once you work with subspaces, you understand them. The ability to reason mathematically will develop, if it is given enough to do. And the essential ideas of linear algebra are *not too hard*.

I would like to say clearly that this is a book about mathematics. It is not so totally watered down that all the purpose is drained out. I do not believe that students or instructors want an empty course; three hours a week can achieve something worthwhile, provided the textbook helps. I hope and believe that you will see, behind the informal and personal style of this book, that it is written to teach real mathematics. There will be sections you omit, and there might be explanations you don't need,† but you cannot miss the underlying force of this subject. It moves simply and naturally from a line or a plane to the n-dimensional space $\mathbf{R}^n$. That step is mathematics at its best, and every student can take it.

One question is hard to postpone: How should the course start? Most students come to the first class already knowing something about linear equations. Still I believe that we must begin with n equations in n unknowns, $Ax = b$, and with the simplest and most useful method of solution—*Gaussian elimination* (not determinants!). It is a perfect introduction to matrix multiplication. And fortunately, even though the method is so straightforward, there are insights that are central to its understanding and new to almost every student. One is to recognize, as elimination goes from the original matrix A to an upper triangular U, that A is being factored into two triangular matrices: $A = LU$. That observation is not deep, and it is easy to verify, but it is tremendously important in practice. For me this is one indicator of a serious course, a dividing line from a presentation that deals only with row operations or A^{-1}.

Another question is to find the right speed. If matrix calculations are familiar, then *Chapter 1 must not be too slow*. It is Chapter 2 that demands more work, and that means work of a different kind—not the crunching of numbers which a computer can do, but an understanding of $Ax = b$ which starts with elimination and goes deeper. The class has to know that the gears have changed; ideas are coming. Instead of individual vectors, we need vector spaces. I am convinced that the four fundamental subspaces—the column space of A, its row space, and the nullspaces of A and A^{T}—are the most effective way to illustrate linear dependence and independence, and to understand "basis" and "dimension" and "rank." Those are developed gradually but steadily, and they generate examples in a completely natural way. They are also the key to $Ax = b$.

May I take one example, to show how an idea can be seen in different ways? It is the fundamental step of multiplying A times x, a matrix times a vector. At one level Ax is just numbers. At the next level it is a combination of the columns of A. At a third level it is a vector in the column space. (We are seeing a space of vectors, containing all combinations and not only this one.) To an algebraist, A represents a linear transformation and Ax is the result of applying it to x. All four are important, and the book must make the connections.

Chapters 1–5 are really the heart of a course in linear algebra. They contain a large number of applications to physics, engineering, probability and statistics, economics, and biology. Those are not tacked on at the end; they are part of the mathematics. Networks and graphs are a terrific source of rectangular matrices, essential in engineering and computer science and also perfect examples for teaching. What mathematics can do, and what linear algebra does so well, is to see patterns that are partly hidden in the applications. *This is a book that allows*

† My favorite proof comes in a book by Ring Lardner: " '*Shut up,' he explained.*"

pure mathematicians to teach applied mathematics. I believe the faculty can do the moving, and teach what students need. The effort is absolutely rewarding.

If you know earlier editions of the book, you will see changes. Section 1.1 is familiar but not 1.2. Certainly the spirit has not changed; this course is alive because its subject is. By teaching it constantly, I found a whole series of improvements—in the organization and the exercises (hundreds are new, over a very wide range), and also in the content. Most of these improvements will be visible only in the day-by-day effort of teaching and learning this subject—when the right explanation or the right exercise makes the difference. I mention two changes that are visible in the table of contents: *Linear transformations* are integrated into the text, and there is a new (and optional) section on the *Fast Fourier Transform.* That is perhaps the outstanding algorithm in modern mathematics, and it has revolutionized digital processing. It is nothing more than a fast way of multiplying by a certain matrix! You may only know (as I did) that the idea exists and is important. It is a pleasure to discover how it fits into linear algebra (and introduces complex numbers).

This is a first course in linear algebra. The theory is motivated, and reinforced, by genuine applications. At the same time, the goal is understanding—and the subject is well established. After Chapter 2 reaches beyond elimination and A^{-1} to the idea of a vector space, Chapter 3 concentrates on *orthogonality.* Geometrically, that is understood before the first lecture. Algebraically, the steps are familiar but crucial—to know when vectors are perpendicular, or which subspaces are orthogonal, or how to project onto a subspace, or how to construct an orthonormal basis. Do not underestimate that chapter. Then Chapter 4 presents *determinants*, the key link between $Ax = b$ and $Ax = \lambda x$. They give a test for invertibility which picks out the eigenvalues. That introduces the last big step in the course.

Chapter 5 puts diagonalization ahead of the Jordan form. The *eigenvalues* and *eigenvectors* take us directly from a matrix A to its powers A^k. They solve equations that evolve in time—dynamic problems, in contrast to the static problem $Ax = b$. They also carry information which is not obvious from the matrix itself—a Markov matrix has $\lambda_{\max} = 1$, an orthogonal matrix has all $|\lambda| = 1$, and a symmetric matrix has real eigenvalues. If your course reaches the beginning of Chapter 6, the connections between eigenvalues and pivots and determinants of symmetric matrices tie the whole subject together. (The last section of each chapter is optional.) Then Chapter 7 gives more concentrated attention to numerical linear algebra, which has become the foundation of scientific computing. And I believe that even a brief look at Chapter 8 allows a worthwhile but relaxed introduction to linear programming—my class is happy because it comes at the end, without examination.

I would like to mention the ***Manual for Instructors***, and another book. The manual contains solutions to all exercises (including the Review Exercises at the ends of Chapters 1–5), and also a collection of ideas and suggestions about applied linear algebra. I hope instructors will request a copy from the publisher (HBJ College Department, 7555 Caldwell Avenue, Chicago IL 60648). I also hope that readers of this book will go forward to the next one. That is called ***Introduction to Applied Mathematics***, and it combines linear algebra with differential equations into a text for modern applied mathematics and engineering mathematics. It includes Fourier analysis, complex variables, partial differential equations, numerical methods, and optimization—but the starting point is linear algebra. It is published by Wellesley-Cambridge Press (Box 157, Wellesley MA 02181) and the response has been tremendous—many departments have wanted a renewal of that course, to teach what is most needed.

This book, like that next one, aims to recognize what the computer can do (without being dominated by it). Solving a problem no longer means writing down an infinite series, or finding a formula like Cramer's rule, but constructing an effective algorithm. That needs good ideas: mathematics survives! The algebra stays clear and simple and stable. For elimination, the operation count in Chapter 1 has also a second purpose—to reinforce a detailed grasp of the n by n case, by actually counting the steps. But I do not do everything in class. The text should supplement as well as summarize the lectures.

In short, a book is needed that will permit the applications to be taught successfully, in combination with the underlying mathematics. That is the book I have tried to write.

In closing, this is a special opportunity for me to say thank you. I am extremely grateful to readers who have liked the book, and have seen what it stands for. Many have written with ideas and encouragement, and I mention only five names: Dan Drucker, Vince Giambalvo, Steve Kleiman, Beresford Parlett, and Jim Simmonds. Behind them is an army of friends and critics that I am proud to have. This third edition has been made better by what they have taught—to students and to the author. It was a very great pleasure to work with Sophia Koulouras, who typed the manuscript, and Michael Michaud, who designed the book and the cover. And above all, my gratitude goes to my wife and children and parents. The book is theirs too, and so is the spirit behind it—which in the end is everything. May I rededicate this book to my mother and father, who gave so much to it: Thank you both.

GILBERT STRANG

CONTENTS

3 ORTHOGONALITY

4 DETERMINANTS

5 EIGENVALUES AND EIGENVECTORS

6 POSITIVE DEFINITE MATRICES

7 COMPUTATIONS WITH MATRICES

8 LINEAR PROGRAMMING AND GAME THEORY

1

MATRICES AND GAUSSIAN ELIMINATION

INTRODUCTION ■ 1.1

The central problem of linear algebra is the solution of linear equations. The most important case, and the simplest, is when the number of unknowns equals the number of equations. Therefore we begin with this basic problem: ***n equations in n unknowns***.

There are two well-established ways to solve linear equations. One is the method of ***elimination***, in which multiples of the first equation are subtracted from the other equations—so as to remove the first unknown from those equations. This leaves a smaller system, of $n-1$ equations in $n-1$ unknowns. The process is repeated until there is only one equation and one unknown, which can be solved immediately. Then it is not hard to go backward, and find the other unknowns in reverse order; we shall work out an example in a moment. A second and more sophisticated way introduces the idea of ***determinants***. There is an exact formula called Cramer's rule, which gives the solution (the correct values of the unknowns) as a ratio of two n by n determinants. From the examples in a textbook it is not obvious which way is better ($n = 3$ or $n = 4$ is about the upper limit on the patience of a reasonable human being).

In fact, the more sophisticated formula involving determinants is a disaster in practice, and elimination is the algorithm that is constantly used to solve large systems of equations. Our first goal is to understand this algorithm. It is generally called ***Gaussian elimination***.

The idea is deceptively simple, and in some form it may already be familiar to the reader. But there are four aspects that lie deeper than the simple mechanics of elimination. Together with the algorithm itself, we want to explain them in this chapter. They are:

(1) The ***geometry*** of linear equations. It is not easy to visualize a 10-dimensional plane in 11-dimensional space. It is harder to see eleven of those planes intersecting at a single point in that space—but somehow it is almost possible. With three planes in three dimensions it can certainly be done. Then linear algebra moves the problem into four dimensions, or eleven dimensions, where the intuition has to imagine the geometry (and gets it right).

(2) The interpretation of elimination as a ***factorization*** of the coefficient matrix. We shall introduce *matrix notation* for the system of n equations, writing the unknowns as a vector x and the equations in the matrix shorthand $Ax = b$. Then *elimination amounts to factoring A into a product LU, of a lower triangular matrix L and an upper triangular matrix U.*

First we have to introduce matrices and vectors in a systematic way, as well as the rules for their multiplication. We also define the ***transpose*** A^{T} and the ***inverse*** A^{-1} of a matrix A.

(3) In most cases elimination goes forward without difficulties. In some exceptional cases it will ***break down***—either because the equations were written in the wrong order, which is easily fixed by exchanging them, or else because the equations fail to have a unique solution. In the latter case there may be ***no solution, or infinitely many***. We want to understand how, at the time of breakdown, the elimination process identifies each of these possibilities.

(4) It is essential to have a rough count of the ***number of operations*** required to solve a system by elimination. The expense in computing often determines the accuracy in the model. The computer can do millions of operations, but not very many trillions. And already after a million steps, roundoff error can be significant. (Some problems are sensitive; others are not.) Without trying for full detail, we want to see what systems arise in practice and how they are actually solved.

The final result of this chapter will be an elimination algorithm which is about as efficient as possible. It is essentially the algorithm that is in constant use in a tremendous variety of applications. And at the same time, understanding it in terms of matrices—the coefficient matrix, the matrices that carry out an elimination step or an exchange of rows, and the final triangular factors L and U—is an essential foundation for the theory.

THE GEOMETRY OF LINEAR EQUATIONS ■ 1.2

The way to understand this subject is by example. We begin with two extremely humble equations, recognizing that you could solve them without a course in linear algebra. Nevertheless I hope you will give Gauss a chance:

$$\begin{aligned} 2x - y &= 1 \\ x + y &= 5. \end{aligned}$$

There are two ways to look at that system, and our main point is to see them both.

The first approach concentrates on the separate equations, in other words on the *rows*. That is the most familiar, and in two dimensions we can do it quickly. The equation $2x - y = 1$ is represented by a *straight line* in the x-y plane. The line goes through the points $x = 1$, $y = 1$ and $x = \frac{1}{2}$, $y = 0$ (and also through $(0, -1)$ and $(2, 3)$ and all intermediate points). The second equation $x + y = 5$ produces a second line (Fig. 1.1a). Its slope is $dy/dx = -1$ and it crosses the first line at the solution. The point of intersection is the only point on both lines, and therefore it is the only solution to both equations. It has the coordinates $x = 2$ and $y = 3$—which will soon be found by a systematic elimination.

The second approach is not so familiar. It looks at the ***columns*** of the linear system. The two separate equations are really one ***vector equation***

$$x\begin{bmatrix} 2 \\ 1 \end{bmatrix} + y\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}.$$

The problem is ***to find the combination of the column vectors on the left side which produces the vector on the right side***. Those two-dimensional vectors are represented by the bold lines in Fig. 1.1b. The unknowns are the numbers x and y which

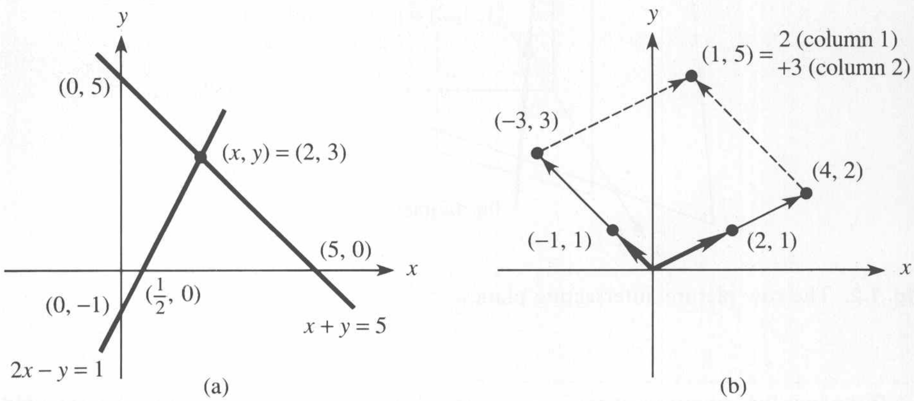


Fig. 1.1. The geometry by rows and by columns.

multiply the column vectors. The whole idea can be seen in that figure, where 2 times column 1 is added to 3 times column 2. Geometrically this produces a famous parallelogram. Algebraically it produces the correct vector $\begin{bmatrix}1\\5\end{bmatrix}$, on the right side of our equations. The column picture confirms that the solution is $x = 2$, $y = 3$.

More time could be spent on that example, but I would rather move forward to $n = 3$. Three equations are still manageable, and they have much more variety. As a specific example, consider

$$\begin{aligned} 2u + v + w &= 5 \\ 4u - 6v \quad &= -2 \\ -2u + 7v + 2w &= 9. \end{aligned} \tag{1}$$

Again we can study the rows or the columns, and we start with the rows. Each equation describes a ***plane*** in three dimensions. The first plane is $2u + v + w = 5$, and it is sketched in Fig. 1.2. It contains the points $(\frac{5}{2}, 0, 0)$ and $(0, 5, 0)$ and $(0, 0, 5)$. It is determined by those three points, or by any three of its points—provided they do not lie on a line. We mention in passing that *the plane* $2u + v + w = 10$ *is parallel to this one.* The corresponding points are $(5, 0, 0)$ and $(0, 10, 0)$ and $(0, 0, 10)$, twice as far away from the origin—which is the center point $u = 0, v = 0, w = 0$. Changing the right hand side moves the plane parallel to itself, and the plane $2u + v + w = 0$ goes through the origin.†

The second plane is $4u - 6v = -2$. It is drawn vertically, because w can take any value. The coefficient of w happens to be zero, but this remains a plane in

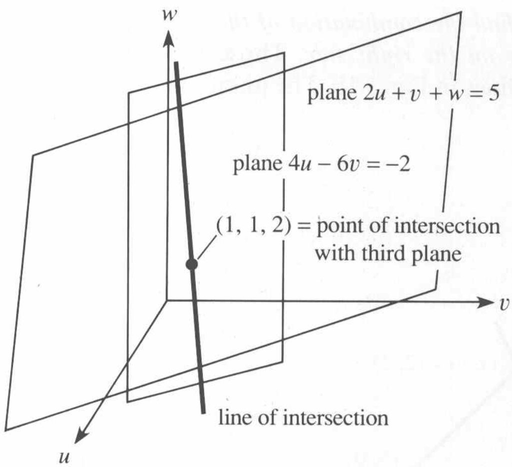


Fig. 1.2. The row picture: intersecting planes.

† If the first two equations were $2u + v + w = 5$ and $2u + v + w = 10$, the planes would not intersect and there would be no solution.

3-space. (If the equation were $4u = 3$, or even the extreme case $u = 0$, it would still describe a plane.) The figure shows the intersection of the second plane with the first. That intersection is a line. *In three dimensions a line requires two equations*; in n dimensions it will require $n - 1$.

Finally the third plane intersects this line in a point. The plane (not drawn) represents the third equation $-2u + 7v + 2w = 9$, and it crosses the line at $u = 1$, $v = 1$, $w = 2$. That point solves the linear system.

How does this picture extend into n dimensions? We will have n equations, and they contain n unknowns. The first equation still determines a "*plane*." It is no longer an ordinary two-dimensional plane in 3-space; somehow it has "*dimension* $n - 1$." It must be extremely thin and flat within n-dimensional space, although it would look solid to us. If time is the fourth dimension, then the plane $t = 0$ cuts through 4-dimensional space and produces the 3-dimensional universe we live in (or rather, the universe as it was at $t = 0$). Another plane is $z = 0$, which is also 3-dimensional; it is the ordinary x-y plane taken over all time. Those three-dimensional planes will intersect! What they have in common is the ordinary x-y plane at $t = 0$. We are down to two dimensions, and the next plane leaves a line. Finally a fourth plane leaves a single point. It is the point at the intersection of 4 planes in 4 dimensions, and it solves the 4 underlying equations.

I will be in trouble if that example from relativity goes any further. The point is that linear algebra can operate with any number of equations. The first one produces an $n - 1$-dimensional plane in n dimensions. The second equation determines another plane, and they intersect (we hope) in a smaller set of "dimension $n - 2$." Assuming all goes well, every new plane (every new equation) reduces the dimension by one. At the end, when all n planes are accounted for, the intersection has dimension zero. It is a *point*, it lies on all the planes, and its coordinates satisfy all n equations. It is the solution! That picture is intuitive—the geometry will need support from the algebra—but it is basically correct.

Column Vectors

We turn to the columns. This time the vector equation (the same equation as (1)) is

$$u \begin{bmatrix} 2 \\ 4 \\ -2 \end{bmatrix} + v \begin{bmatrix} 1 \\ -6 \\ 7 \end{bmatrix} + w \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix}. \tag{2}$$

Those are *three-dimensional column vectors*. The vector b on the right side has components $5, -2, 9$, and these components allow us to draw the vector. ***The vector b is identified with the point whose coordinates are*** $5, -2, 9$. Every point in three-dimensional space is matched to a vector, and vice versa. That was the idea of Descartes, who turned geometry into algebra by working with the coordinates of the point. We can write the vector in a column, or we can list its components

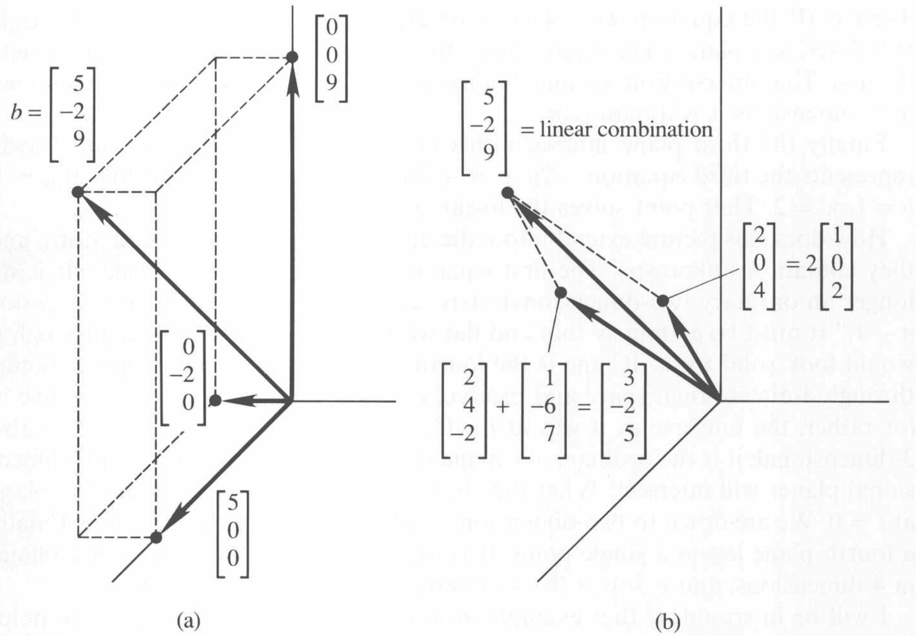


Fig. 1.3. The column picture: linear combination of columns equals b.

as $b = (5, -2, 9)$, or we can represent it geometrically by an arrow from the origin.† Throughout the book we use parentheses and commas when the components are listed horizontally, and square brackets (with no commas) when a column vector is printed vertically.

What really matters is ***addition of vectors*** and ***multiplication by a scalar*** (a number). In Fig. 1.3a you see a vector addition, which is carried out component by component:

$$\begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 9 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix}.$$

In the right figure there is a multiplication by 2 (and if it had been -2 the vector would have gone in the reverse direction):

$$2\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix}, \qquad -2\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ -4 \end{bmatrix}.$$

† Some authors prefer to say that the arrow is really the vector, but I think it doesn't matter; you can choose the arrow, or the point, or the three numbers. (They all start with the same origin (0, 0, 0).) In six dimensions it is probably easiest to choose the six numbers.

Also in the right figure is one of the central ideas of linear algebra. It uses *both* of the basic operations; vectors are *multiplied by numbers and then added*. The result is called a ***linear combination***, and in this case the linear combination is

$$\begin{bmatrix} 2 \\ 4 \\ -2 \end{bmatrix} + \begin{bmatrix} 1 \\ -6 \\ 7 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix}.$$

You recognize the significance of that special combination; it solves equation (2). The equation asked for multipliers u, v, w which produce the right side b. Those numbers are $u = 1$, $v = 1$, $w = 2$. They give the correct linear combination of the column vectors, and they also gave the point (1, 1, 2) in the row picture (where the three planes intersect).

Multiplication and addition are carried out on each component separately. Therefore linear combinations are possible provided the vectors have the same number of components. Note that all vectors in the figure were three-dimensional, even though some of their components were zero.

Do not forget the goal. It is to look beyond two or three dimensions into n dimensions. With n equations in n unknowns, there were n planes in the row picture. There are n vectors in the column picture, plus a vector b on the right side. The equations ask for a ***linear combination of the n vectors that equals b***. In this example we found one such combination (there are no others) but for certain equations that will be impossible. Paradoxically, the way to understand the good case is to study the bad one. Therefore we look at the geometry exactly when it breaks down, in what is called the *singular case*.

First we summarize:

Row picture: Intersection of n planes
Column picture: The right side b is a combination of the column vectors
Solution to equations: Intersection point of planes = coefficients in the combination of columns

The Singular Case

Suppose we are again in three dimensions, and the three planes in the row picture *do not intersect*. What can go wrong? One possibility, already noted, is that two planes may be parallel. Two of the equations, for example $2u + v + w = 5$ and $4u + 2v + 2w = 11$, may be inconsistent—and there is no solution (Fig. 1.4a shows an end view). In the two-dimensional problem, with lines instead of planes, this is the only possibility for breakdown. That problem is singular if the lines are parallel, and it is nonsingular if they meet. But three planes in three dimensions can be in trouble without being parallel.

The new difficulty is shown in Fig. 1.4b. All three planes are perpendicular to the page; from the end view they form a triangle. Every pair of planes intersects,