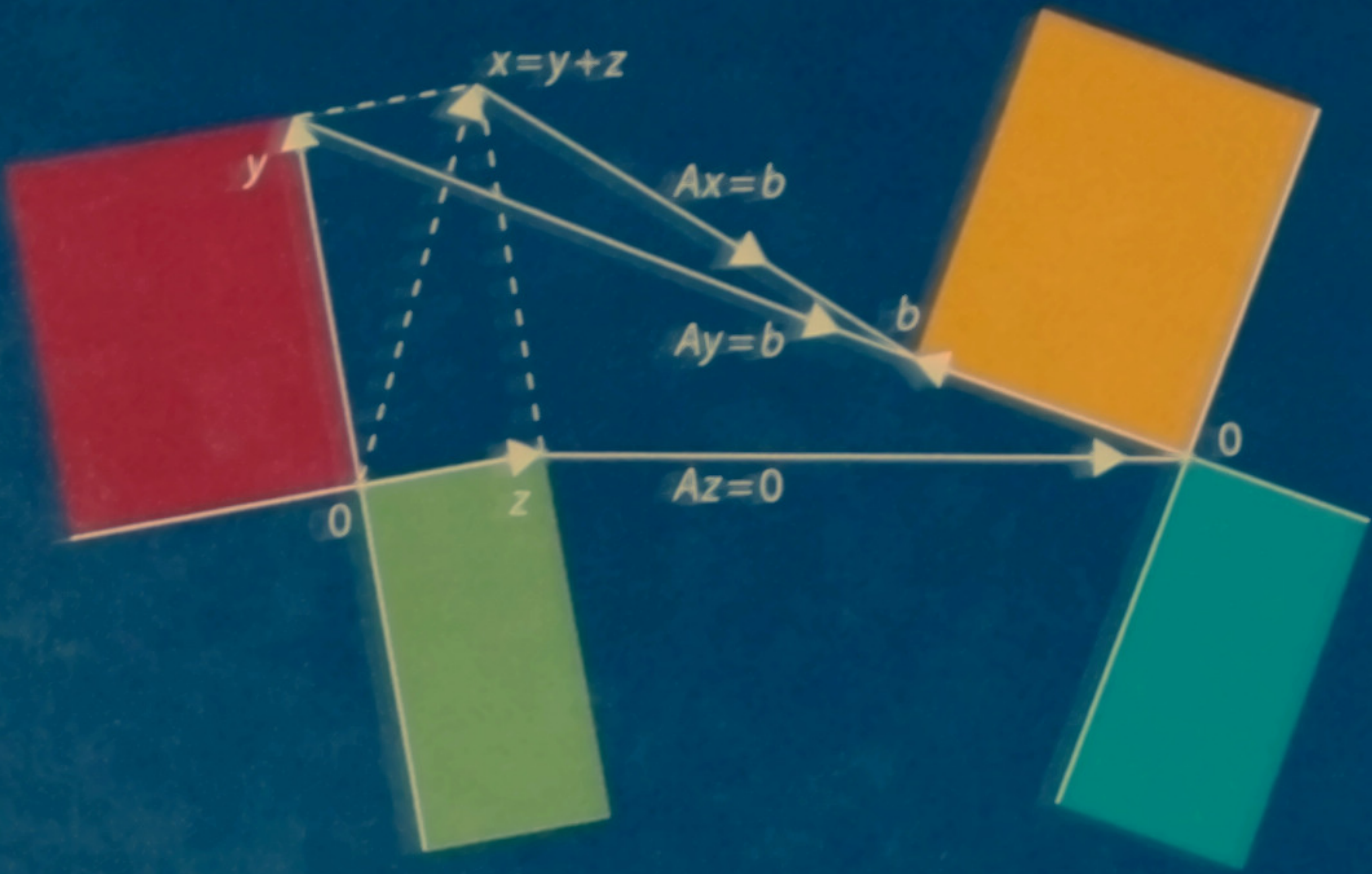
Introduction to

LINEAR ALGEBRA

FOURTH EDITION



GILBERT STRANG

INTRODUCTION TO LINEAR ALGEBRA

Fourth Edition

GILBERT STRANG
Massachusetts Institute of Technology

WELLESLEY - CAMBRIDGE PRESS
Box 812060 Wellesley MA 02482

Introduction to Linear Algebra, 4th Edition
Copyright ©2009 by Gilbert Strang
ISBN 978-0-9802327-1-4

Fourth International Edition
Copyright ©2009 by Gilbert Strang
ISBN 978-0-9802327-2-1

Typeset by www.valutone.co.in Printed in the United States of America 9 8 7 6 5 4

QA184.S78 2009 512'.5 93-14092

Other texts from Wellesley - Cambridge Press

Computational Science and Engineering, Gilbert Strang
ISBN 978-0-9614088-1-7 ISBN 0-9614088-1-2

Wavelets and Filter Banks, Gilbert Strang and Truong Nguyen
ISBN 978-0-9614088-7-9 ISBN 0-9614088-7-1

Introduction to Applied Mathematics, Gilbert Strang
ISBN 978-0-9614088-0-0 ISBN 0-9614088-0-4

An Analysis of the Finite Element Method,
2008 edition, Gilbert Strang and George Fix
ISBN 978-0-9802327-0-7 ISBN 0-9802327-0-8

Calculus Second edition (2010), Gilbert Strang
ISBN 978-0-9802327-4-5 ISBN 0-9802327-4-0

Wellesley - Cambridge Press
Box 812060
Wellesley MA 02482 USA
www.wellesleycambridge.com

gs@math.mit.edu
math.mit.edu/~gs
phone (781) 431-8488
fax (617) 253-4358

The website for this book is **math.mit.edu/linearalgebra**.
A Solutions Manual is available to instructors by email from the publisher.

Course material including syllabus and Teaching Codes and exams and also videotaped lectures are available on the teaching website: **web.mit.edu/18.06**

Linear Algebra is included in MIT's OpenCourseWare site **ocw.mit.edu**.
This provides video lectures of the full linear algebra course 18.06.

The front cover captures a central idea of linear algebra.
$Ax = b$ is solvable when b is in the (orange) column space of A.
One particular solution y is in the (red) row space: $Ay = b$.
Add any vector z from the (green) nullspace of A: $Az = 0$.
The complete solution is $x = y + z$. Then $Ax = Ay + Az = b$.

The cover design was the inspiration of a creative collaboration:
Lois Sellers (**birchdesignassociates.com**) and Gail Corbett.

Table of Contents

Preface

I will be happy with this preface if three important points come through clearly:

1. The beauty and variety of linear algebra, and its extreme usefulness
2. The goals of this book, and the new features in this Fourth Edition
3. The steady support from our linear algebra websites and the video lectures

May I begin with notes about two websites that are constantly used, and the new one.

ocw.mit.edu Messages come from thousands of students and faculty about linear algebra on this OpenCourseWare site. The 18.06 course includes video lectures of a complete semester of classes. Those lectures offer an independent review of the whole subject based on this textbook—the professor's time stays free and the student's time can be 3 a.m. (The reader doesn't have to be in a class at all.) A million viewers around the world have seen these videos (*amazing*). I hope you find them helpful.

web.mit.edu/18.06 This site has homeworks and exams (with solutions) for the current course as it is taught, and as far back as 1996. There are also review questions, Java demos, Teaching Codes, and short essays (*and the video lectures*). My goal is to make this book as useful as possible, with all the course material we can provide.

math.mit.edu/linearalgebra The newest website is devoted specifically to this Fourth Edition. It will be a permanent record of ideas and codes and good problems and solutions. Several sections of the book are directly available online, plus notes on teaching linear algebra. The content is growing quickly and contributions are welcome from everyone.

The Fourth Edition

Thousands of readers know earlier editions of *Introduction to Linear Algebra*. The new cover shows the **Four Fundamental Subspaces**—the row space and nullspace are on the left side, the column space and the nullspace of A^{T} are on the right. It is not usual to put the central ideas of the subject on display like this! You will meet those four spaces in Chapter 3, and you will understand why that picture is so central to linear algebra.

Those were named the Four Fundamental Subspaces in my first book, and they start from a matrix A. Each row of A is a vector in n-dimensional space. When the matrix has m rows, each column is a vector in m-dimensional space. The crucial operation in linear algebra is taking ***linear combinations of vectors***. (That idea starts on page 1 of the book and never stops.) *When we take all linear combinations of the column vectors, we get the column space.* If this space includes the vector b, we can solve the equation $Ax = b$.

I have to stop here or you won't read the book. May I call special attention to the new Section 1.3 in which these ideas come early—with two specific examples. You are not expected to catch every detail of vector spaces in one day! But you will see the first matrices in the book, and a picture of their column spaces, and even an *inverse matrix*. You will be learning the language of linear algebra in the best and most efficient way: by using it.

Every section of the basic course now ends with ***Challenge Problems***. They follow a large collection of review problems, which ask you to use the ideas in that section—the dimension of the column space, a basis for that space, the rank and inverse and determinant and eigenvalues of A. Many problems look for computations by hand on a small matrix, and they have been highly praised. The new Challenge Problems go a step further, and sometimes they go deeper. Let me give four examples:

Section 2.1: Which row exchanges of a Sudoku matrix produce another Sudoku matrix?

Section 2.4: From the shapes of A, B, C, is it faster to compute AB times C or A times BC?

Background: The great fact about multiplying matrices is that *AB times C gives the same answer as A times BC*. This simple statement is the reason behind the rule for matrix multiplication. If AB is square and C is a vector, it's faster to do BC first. Then multiply by A to produce ABC. The question asks about other shapes of A, B, and C.

Section 3.4: If $Ax = b$ and $Cx = b$ have the same solutions for every b, is $A = C$?

Section 4.1: What conditions on the four vectors r, n, c, ℓ allow them to be bases for the row space, the nullspace, the column space, and the left nullspace of a 2 by 2 matrix?

The Start of the Course

The equation $Ax = b$ uses the language of linear combinations right away. The vector Ax is *a combination of the columns of A*. The equation is asking for ***a combination that produces b***. The solution vector x comes at three levels and all are important:

1. ***Direct solution*** to find x by forward elimination and back substitution.

2. ***Matrix solution*** using the inverse of A: $x = A^{-1}b$ (if A has an inverse).

3. ***Vector space solution*** $x = y + z$ as shown on the cover of the book:

   Particular solution (to $Ay = b$) plus ***nullspace solution*** (to $Az = 0$)

Direct elimination is the most frequently used algorithm in scientific computing, and the idea is not hard. Simplify the matrix A so it becomes triangular—then all solutions come quickly. I don't spend forever on practicing elimination, it will get learned.

The speed of every new supercomputer is tested on $Ax = b$: it's pure linear algebra. IBM and Los Alamos announced a new world record of 10^{15} operations per second in 2008.

That *petaflop speed* was reached by solving many equations in parallel. High performance computers avoid operating on single numbers, they feed on whole submatrices.

The processors in the Roadrunner are based on the Cell Engine in PlayStation 3. What can I say, video games are now the largest market for the fastest computations.

Even a supercomputer doesn't want the inverse matrix: too slow. Inverses give the simplest formula $x = A^{-1}b$ but not the top speed. And everyone must know that determinants are even slower—there is no way a linear algebra course should begin with formulas for the determinant of an n by n matrix. Those formulas have a place, but not first place.

Structure of the Textbook

Already in this preface, you can see the style of the book and its goal. That goal is serious, to explain this beautiful and useful part of mathematics. You will see how the applications of linear algebra reinforce the key ideas. I hope every teacher will learn something new; familiar ideas can be seen in a new way. The book moves gradually and steadily from *numbers* to *vectors* to *subspaces*—each level comes naturally and everyone can get it.

Here are ten points about the organization of this book:

1. Chapter 1 starts with vectors and dot products. If the class has met them before, focus quickly on linear combinations. The new Section 1.3 provides three independent vectors whose combinations fill all of 3-dimensional space, and three dependent vectors in a plane. ***Those two examples are the beginning of linear algebra.***

2. Chapter 2 shows the row picture and the column picture of $Ax = b$. The heart of linear algebra is in that connection between the rows of A and the columns: the same numbers but very different pictures. Then begins the algebra of matrices: an elimination matrix E multiplies A to produce a zero. The goal here is to capture the whole process—start with A and end with an ***upper triangular*** U.

   Elimination is seen in the beautiful form $A = LU$. The ***lower triangular*** L holds all the forward elimination steps, and U is the matrix for back substitution.

3. Chapter 3 is linear algebra at the best level: ***subspaces***. The column space contains all linear combinations of the columns. The crucial question is: ***How many of those columns are needed***? The answer tells us the dimension of the column space, and the key information about A. We reach the Fundamental Theorem of Linear Algebra.

4. Chapter 4 has m equations and only n unknowns. It is almost sure that $Ax = b$ has no solution. We cannot throw out equations that are close but not perfectly exact. When we solve by ***least squares***, the key will be the matrix $A^{\mathrm{T}}A$. This wonderful matrix $A^{\mathrm{T}}A$ appears everywhere in applied mathematics, when A is rectangular.

5. ***Determinants*** in Chapter 5 give formulas for all that has come before—inverses, pivots, volumes in n-dimensional space, and more. We don't need those formulas to compute! They slow us down. But $\det A = 0$ tells when a matrix is singular, and that test is the key to eigenvalues.

6. ***Section 6.1 introduces eigenvalues for 2 by 2 matrices.*** Many courses want to see eigenvalues early. It is completely reasonable to come here directly from Chapter 3, because the determinant is easy for a 2 by 2 matrix. *The key equation is* $Ax = \lambda x$.

   Eigenvalues and eigenvectors are an astonishing way to understand a square matrix. They are not for $Ax = b$, they are for dynamic equations like $du/dt = Au$. The idea is always the same: *follow the eigenvectors.* In those special directions, A acts like a single number (the eigenvalue λ) and the problem is one-dimensional.

   Chapter 6 is full of applications. One highlight is ***diagonalizing a symmetric matrix***. Another highlight—not so well known but more important every day—is the diagonalization of *any matrix*. This needs two sets of eigenvectors, not one, and they come (of course!) from $A^{\mathrm{T}}A$ and AA^{T}. This Singular Value Decomposition often marks the end of the basic course and the start of a second course.

7. Chapter 7 explains the ***linear transformation*** approach—it is linear algebra without coordinates, the ideas without computations. Chapter 9 is the opposite—all about how $Ax = b$ and $Ax = \lambda x$ are really solved. Then Chapter 10 moves from real numbers and vectors to complex vectors and matrices. The Fourier matrix F is the most important complex matrix we will ever see. And the ***Fast Fourier Transform*** (multiplying quickly by F and F^{-1}) is a revolutionary algorithm.

8. Chapter 8 is full of applications, more than any single course could need:

   8.1 *Matrices in Engineering*—differential equations replaced by matrix equations

   8.2 *Graphs and Networks*—leading to the edge-node matrix for Kirchhoff's Laws

   8.3 *Markov Matrices*—as in Google's *PageRank* algorithm

   8.4 *Linear Programming*—a new requirement $x \geq 0$ and minimization of the cost

   8.5 *Fourier Series*—linear algebra for functions and digital signal processing

   8.6 *Matrices in Statistics and Probability*—$Ax = b$ is weighted by average errors

   8.7 *Computer Graphics*—matrices move and rotate and compress images.

9. Every section in the basic course ends with a ***Review of the Key Ideas***.

10. How should computing be included in a linear algebra course? It can open a new understanding of matrices—every class will find a balance. I chose the language of MATLAB as a direct way to describe linear algebra: eig(ones(4)) will produce the eigenvalues 4, 0, 0, 0 of the 4 by 4 all-ones matrix. *Go to* ***netlib.org*** *for codes.*

    You can freely choose a different system. More and more software is open source.

The new website ***math.mit.edu/linearalgebra*** provides further ideas about teaching and learning. Please contribute! Good problems are welcome by email: ***gs@math.mit.edu***. Send new applications too, linear algebra is an incredibly useful subject.

The Variety of Linear Algebra

Calculus is mostly about one special operation (the derivative) and its inverse (the integral). Of course I admit that calculus could be important .... But so many applications of mathematics are discrete rather than continuous, digital rather than analog. The century of data has begun! You will find a light-hearted essay called "Too Much Calculus" on my website. ***The truth is that vectors and matrices have become the language to know.***

Part of that language is the wonderful variety of matrices. Let me give three examples:

Symmetric matrix	***Orthogonal matrix***	***Triangular matrix***
$\begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}$	$\frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

A key goal is learning to "read" a matrix. You need to see the meaning in the numbers. This is really the essence of mathematics—patterns and their meaning.

May I end with this thought for professors. You might feel that the direction is right, and wonder if your students are ready. ***Just give them a chance***! Literally thousands of students have written to me, frequently with suggestions and surprisingly often with thanks. They know this course has a purpose, because the professor and the book are on their side. Linear algebra is a fantastic subject, enjoy it.

Help With This Book

I can't even name all the friends who helped me, beyond thanking Brett Coonley at MIT and Valutone in Mumbai and SIAM in Philadelphia for years of constant and dedicated support. The greatest encouragement of all is the feeling that you are doing something worthwhile with your life. Hundreds of generous readers have sent ideas and examples and corrections (and favorite matrices!) that appear in this book. *Thank you all.*

Background of the Author

This is my eighth textbook on linear algebra, and I have not written about myself before. I hesitate to do it now. It is the mathematics that is important, and the reader. The next paragraphs add something personal as a way to say that textbooks are written by people.

I was born in Chicago and went to school in Washington and Cincinnati and St. Louis. My college was MIT (and my linear algebra course was *extremely abstract*). After that came Oxford and UCLA, then back to MIT for a very long time. I don't know how many thousands of students have taken 18.06 (more than a million when you include the videos on ***ocw.mit.edu***). The time for a fresh approach was right, because this fantastic subject was only revealed to math majors—**we needed to open linear algebra to the world.**

Those years of teaching led to the Haimo Prize from the Mathematical Association of America. For encouraging education worldwide, the International Congress of Industrial and Applied Mathematics awarded me the first Su Buchin Prize. I am extremely grateful, more than I could possibly say. What I hope most is that you will like linear algebra.

Chapter 1

Introduction to Vectors

The heart of linear algebra is in two operations—both with vectors. We add vectors to get $v + w$. We multiply them by numbers c and d to get cv and dw. Combining those two operations (adding cv to dw) gives the ***linear combination*** $cv + dw$.

Linear combination $\quad cv + dw = c\begin{bmatrix} 1 \\ 1 \end{bmatrix} + d\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} c + 2d \\ c + 3d \end{bmatrix}$

Example $\quad v + w = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ is the combination with $c = d = 1$

Linear combinations are all-important in this subject! Sometimes we want one particular combination, the specific choice $c = 2$ and $d = 1$ that produces $cv + dw = (4, 5)$. Other times we want *all the combinations* of v and w (coming from all c and d).

The vectors cv lie along a line. When w is not on that line, **the combinations** $cv + dw$ **fill the whole two-dimensional plane**. (I have to say "two-dimensional" because linear algebra allows higher-dimensional planes.) Starting from four vectors u, v, w, z in four-dimensional space, their combinations $cu + dv + ew + fz$ are likely to fill the space—but not always. The vectors and their combinations could even lie on one line.

Chapter 1 explains these central ideas, on which everything builds. We start with two-dimensional vectors and three-dimensional vectors, which are reasonable to draw. Then we move into higher dimensions. The really impressive feature of linear algebra is how smoothly it takes that step into n-dimensional space. Your mental picture stays completely correct, even if drawing a ten-dimensional vector is impossible.

This is where the book is going (into n-dimensional space). The first steps are the operations in Sections 1.1 and 1.2. Then Section 1.3 outlines three fundamental ideas.

1.1 ***Vector addition $v + w$ and linear combinations $cv + dw$.***

1.2 ***The dot product $v \cdot w$ of two vectors and the length $\|v\| = \sqrt{v \cdot v}$.***

1.3 ***Matrices A, linear equations $Ax = b$, solutions $x = A^{-1}b$.***

1.1 Vectors and Linear Combinations

"You can't add apples and oranges." In a strange way, this is the reason for vectors. We have two separate numbers v_1 and v_2. That pair produces a ***two-dimensional vector*** $\boldsymbol{v}$:

$$\textbf{Column vector} \qquad \boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \qquad \begin{aligned} v_1 &= \text{first component} \\ v_2 &= \text{second component} \end{aligned}$$

We write $\boldsymbol{v}$ as a ***column***, not as a row. The main point so far is to have a single letter $\boldsymbol{v}$ (in ***boldface italic***) for this pair of numbers v_1 and v_2 (in *lightface italic*).

Even if we don't add v_1 to v_2, we do *add vectors*. The first components of $\boldsymbol{v}$ and $\boldsymbol{w}$ stay separate from the second components:

$$\begin{matrix}\textbf{VECTOR} \\ \textbf{ADDITION}\end{matrix} \qquad \boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \text{ and } \boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \text{ add to } \boldsymbol{v} + \boldsymbol{w} = \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \end{bmatrix}.$$

You see the reason. We want to add apples to apples. Subtraction of vectors follows the same idea: *The components of* $\boldsymbol{v} - \boldsymbol{w}$ *are* $v_1 - w_1$ *and* $v_2 - w_2$.

The other basic operation is *scalar multiplication*. Vectors can be multiplied by 2 or by -1 or by any number c. There are two ways to double a vector. One way is to add $\boldsymbol{v} + \boldsymbol{v}$. The other way (the usual way) is to multiply each component by 2:

$$\begin{matrix}\textbf{SCALAR} \\ \textbf{MULTIPLICATION}\end{matrix} \qquad 2\boldsymbol{v} = \begin{bmatrix} 2v_1 \\ 2v_2 \end{bmatrix} \text{ and } -\boldsymbol{v} = \begin{bmatrix} -v_1 \\ -v_2 \end{bmatrix}.$$

The components of $c\boldsymbol{v}$ are cv_1 and cv_2. The number c is called a "scalar".

Notice that the sum of $-\boldsymbol{v}$ and $\boldsymbol{v}$ is the zero vector. This is $\mathbf{0}$, which is not the same as the number zero! The vector $\mathbf{0}$ has components 0 and 0. Forgive me for hammering away at the difference between a vector and its components. Linear algebra is built on these operations $\boldsymbol{v} + \boldsymbol{w}$ and $c\boldsymbol{v}$—***adding vectors and multiplying by scalars***.

The order of addition makes no difference: $\boldsymbol{v} + \boldsymbol{w}$ equals $\boldsymbol{w} + \boldsymbol{v}$. Check that by algebra: The first component is $v_1 + w_1$ which equals $w_1 + v_1$. Check also by an example:

$$\boldsymbol{v} + \boldsymbol{w} = \begin{bmatrix} 1 \\ 5 \end{bmatrix} + \begin{bmatrix} 3 \\ 3 \end{bmatrix} = \begin{bmatrix} \mathbf{4} \\ \mathbf{8} \end{bmatrix} \qquad \boldsymbol{w} + \boldsymbol{v} = \begin{bmatrix} 3 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} \mathbf{4} \\ \mathbf{8} \end{bmatrix}.$$

Linear Combinations

Combining addition with scalar multiplication, we now form "*linear combinations*" of $\boldsymbol{v}$ and $\boldsymbol{w}$. Multiply $\boldsymbol{v}$ by c and multiply $\boldsymbol{w}$ by d; then add $c\boldsymbol{v} + d\boldsymbol{w}$.

DEFINITION ***The sum of*** $c\boldsymbol{v}$ ***and*** $d\boldsymbol{w}$ ***is a*** *linear combination* ***of*** $\boldsymbol{v}$ ***and*** $\boldsymbol{w}$.

Four special linear combinations are: sum, difference, zero, and a scalar multiple $c\boldsymbol{v}$:

$$\begin{aligned} 1\boldsymbol{v} + 1\boldsymbol{w} &= \text{sum of vectors in Figure 1.1a} \\ 1\boldsymbol{v} - 1\boldsymbol{w} &= \text{difference of vectors in Figure 1.1b} \\ 0\boldsymbol{v} + 0\boldsymbol{w} &= \textit{zero vector} \\ c\boldsymbol{v} + 0\boldsymbol{w} &= \text{vector } c\boldsymbol{v} \text{ in the direction of } \boldsymbol{v} \end{aligned}$$

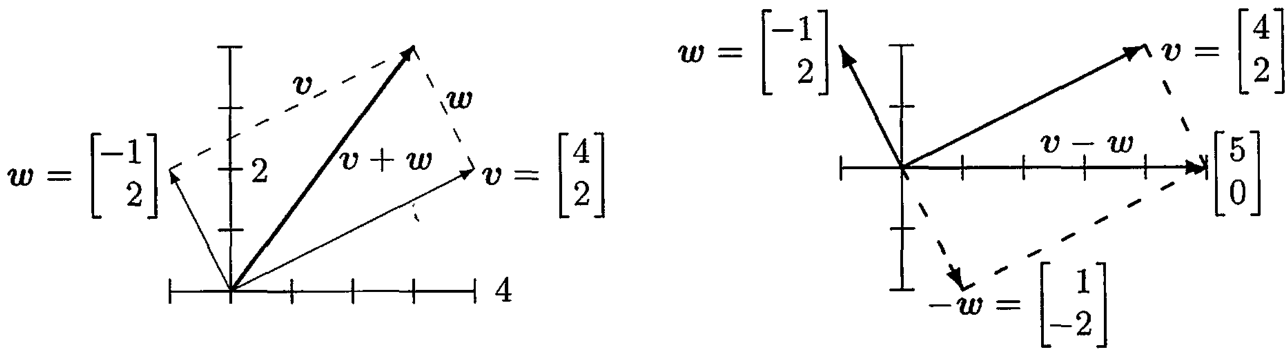
The zero vector is always a possible combination (its coefficients are zero). Every time we see a "space" of vectors, that zero vector will be included. This big view, taking *all* the combinations of $\boldsymbol{v}$ and $\boldsymbol{w}$, is linear algebra at work.

The figures show how you can visualize vectors. For algebra, we just need the components (like 4 and 2). That vector $\boldsymbol{v}$ is represented by an arrow. The arrow goes $v_1 = 4$ units to the right and $v_2 = 2$ units up. It ends at the point whose x, y coordinates are $4, 2$. This point is another representation of the vector—so we have three ways to describe $\boldsymbol{v}$:

Represent vector $\boldsymbol{v}$	Two numbers	Arrow from $(0, 0)$	Point in the plane

We add using the numbers. We visualize $\boldsymbol{v} + \boldsymbol{w}$ using arrows:

Vector addition (head to tail) ***At the end of*** $\boldsymbol{v}$, ***place the start of*** $\boldsymbol{w}$.



$$\boldsymbol{v} + \boldsymbol{w} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \qquad \boldsymbol{v} - \boldsymbol{w} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} - \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$$

Figure 1.1: Vector addition $\boldsymbol{v} + \boldsymbol{w} = (3,\ 4)$ produces the diagonal of a parallelogram. The linear combination on the right is $\boldsymbol{v} - \boldsymbol{w} = (5,\ 0)$.

We travel along $\boldsymbol{v}$ and then along $\boldsymbol{w}$. Or we take the diagonal shortcut along $\boldsymbol{v} + \boldsymbol{w}$. We could also go along $\boldsymbol{w}$ and then $\boldsymbol{v}$. In other words, $\boldsymbol{w} + \boldsymbol{v}$ gives the same answer as $\boldsymbol{v} + \boldsymbol{w}$.

These are different ways along the parallelogram (in this example it is a rectangle). The sum is the diagonal vector $v + w$.

The zero vector $\mathbf{0} = (0, 0)$ is too short to draw a decent arrow, but you know that $v + \mathbf{0} = v$. For $2v$ we double the length of the arrow. We reverse w to get $-w$. This reversing gives the subtraction on the right side of Figure 1.1.

Vectors in Three Dimensions

A vector with two components corresponds to a point in the xy plane. The components of v are the coordinates of the point: $x = v_1$ and $y = v_2$. The arrow ends at this point (v_1, v_2), when it starts from $(0, 0)$. Now we allow vectors to have three components (v_1, v_2, v_3).

The xy plane is replaced by three-dimensional space. Here are typical vectors (still column vectors but with three components):

$$v = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \quad \text{and} \quad w = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \quad \text{and} \quad v + w = \begin{bmatrix} 3 \\ 4 \\ 3 \end{bmatrix}.$$

The vector v corresponds to an arrow in 3-space. Usually the arrow starts at the "origin", where the xyz axes meet and the coordinates are $(0, 0, 0)$. The arrow ends at the point with coordinates v_1, v_2, v_3. There is a perfect match between the ***column vector*** and the ***arrow from the origin*** and the ***point where the arrow ends***.

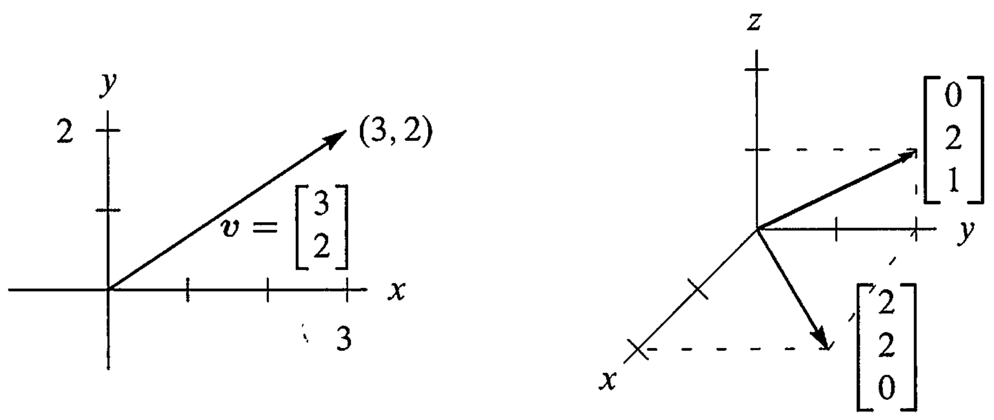


Figure 1.2: Vectors $\begin{bmatrix} x \\ y \end{bmatrix}$ and $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ correspond to points (x, y) and (x, y, z).

From now on $v = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$ ***is also written as*** $v = (1, 1, -1)$.

The reason for the row form (in parentheses) is to save space. But $\boldsymbol{v} = (1, 1, -1)$ is not a row vector! It is in actuality a column vector, just temporarily lying down. The row vector $[\,1 \quad 1 \quad -1\,]$ is absolutely different, even though it has the same three components. That row vector is the "transpose" of the column $\boldsymbol{v}$.

In three dimensions, $\boldsymbol{v} + \boldsymbol{w}$ is still found a component at a time. The sum has components $v_1 + w_1$ and $v_2 + w_2$ and $v_3 + w_3$. You see how to add vectors in 4 or 5 or n dimensions. When $\boldsymbol{w}$ starts at the end of $\boldsymbol{v}$, the third side is $\boldsymbol{v} + \boldsymbol{w}$. The other way around the parallelogram is $\boldsymbol{w} + \boldsymbol{v}$. Question: Do the four sides all lie in the same plane? *Yes.* And the sum $\boldsymbol{v} + \boldsymbol{w} - \boldsymbol{v} - \boldsymbol{w}$ goes completely around to produce the _____ vector.

A typical linear combination of three vectors in three dimensions is $\boldsymbol{u} + 4\boldsymbol{v} - 2\boldsymbol{w}$:

Linear combination
Multiply by 1, 4, −2
Then add

$$\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} + 4 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} - 2 \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix}.$$

The Important Questions

For one vector $\boldsymbol{u}$, the only linear combinations are the multiples $c\boldsymbol{u}$. For two vectors, the combinations are $c\boldsymbol{u} + d\boldsymbol{v}$. For three vectors, the combinations are $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$. Will you take the big step from *one* combination to *all* combinations? Every c and d and e are allowed. Suppose the vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ are in three-dimensional space:

1. What is the picture of *all* combinations $c\boldsymbol{u}$?
2. What is the picture of *all* combinations $c\boldsymbol{u} + d\boldsymbol{v}$?
3. What is the picture of *all* combinations $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$?

The answers depend on the particular vectors $\boldsymbol{u}, \boldsymbol{v}$, and $\boldsymbol{w}$. If they were zero vectors (a very extreme case), then every combination would be zero. If they are typical nonzero vectors (components chosen at random), here are the three answers. This is the key to our subject:

1. The combinations $c\boldsymbol{u}$ fill a ***line***.
2. The combinations $c\boldsymbol{u} + d\boldsymbol{v}$ fill a ***plane***.
3. The combinations $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$ fill ***three-dimensional space***.

The zero vector $(0, 0, 0)$ is on the line because c can be zero. It is on the plane because c and d can be zero. The line of vectors $c\boldsymbol{u}$ is infinitely long (forward and backward). It is the plane of all $c\boldsymbol{u} + d\boldsymbol{v}$ (combining two vectors in three-dimensional space) that I especially ask you to think about.

Adding all $c\boldsymbol{u}$ on one line to all $d\boldsymbol{v}$ on the other line fills in the plane in Figure 1.3.

When we include a third vector $\boldsymbol{w}$, the multiples $e\boldsymbol{w}$ give a third line. Suppose that third line is not in the plane of $\boldsymbol{u}$ and $\boldsymbol{v}$. Then combining all $e\boldsymbol{w}$ with all $c\boldsymbol{u} + d\boldsymbol{v}$ fills up the whole three-dimensional space.

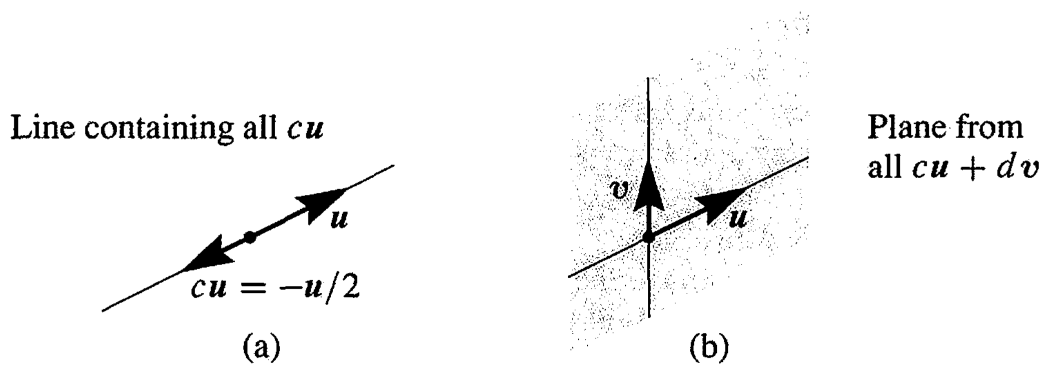


Figure 1.3: (a) Line through u. (b) The plane containing the lines through u and v.

This is the typical situation! **Line**, then **plane**, then **space**. But other possibilities exist. When w happens to be $cu + dv$, the third vector is in the plane of the first two. The combinations of u, v, w will not go outside that uv plane. We do not get the full three-dimensional space. Please think about the special cases in Problem 1.

■ REVIEW OF THE KEY IDEAS ■

1. A vector v in two-dimensional space has two components v_1 and v_2.
2. $v + w = (v_1 + w_1,\ v_2 + w_2)$ and $cv = (cv_1, cv_2)$ are found a component at a time.
3. A linear combination of three vectors u and v and w is $cu + dv + ew$.
4. Take *all* linear combinations of u, or u and v, or u, v, w. In three dimensions, those combinations typically fill a line, then a plane, and the whole space $\mathbf{R}^3$.

■ WORKED EXAMPLES ■

1.1 A The linear combinations of $v = (1, 1, 0)$ and $w = (0, 1, 1)$ fill a plane. *Describe that plane.* Find a vector that is *not* a combination of v and w.

Solution **The combinations $cv + dw$ fill a plane in $\mathbf{R}^3$.** The vectors in that plane allow any c and d. The plane of Figure 1.3 fills in between the "u-line" and the "v-line".

$$\text{Combinations} \quad cv + dw = c\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + d\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} c \\ c+d \\ d \end{bmatrix} \text{ fill a plane.}$$

Four particular vectors in that plane are $(0, 0, 0)$ and $(2, 3, 1)$ and $(5, 7, 2)$ and $(\pi, 2\pi, \pi)$. The second component $c + d$ is always the sum of the first and third components. *The vector* $(1, 2, 3)$ *is not in the plane, because* $2 \neq 1 + 3$.

Another description of this plane through $(0,0,0)$ is to know that $\boldsymbol{n} = (1,-1,1)$ is **perpendicular** to the plane. Section 1.2 will confirm that $90°$ angle by testing dot products: $\boldsymbol{v}\cdot\boldsymbol{n} = 0$ and $\boldsymbol{w}\cdot\boldsymbol{n} = 0$.

1.1 B For $\boldsymbol{v} = (1,0)$ and $\boldsymbol{w} = (0,1)$, describe all points $c\boldsymbol{v}$ with (1) *whole numbers* c (2) *nonnegative* $c \geq 0$. Then add all vectors $d\boldsymbol{w}$ and describe all $c\boldsymbol{v} + d\boldsymbol{w}$.

Solution

(1) The vectors $c\boldsymbol{v} = (c,0)$ with whole numbers c are **equally spaced points** along the x axis (the direction of $\boldsymbol{v}$). They include $(-2,0)$, $(-1,0)$, $(0,0)$, $(1,0)$, $(2,0)$.

(2) The vectors $c\boldsymbol{v}$ with $c \geq 0$ fill a ***half-line***. It is the *positive* x axis. This half-line starts at $(0,0)$ where $c = 0$. It includes $(\pi,0)$ but not $(-\pi,0)$.

(1′) Adding all vectors $d\boldsymbol{w} = (0,d)$ puts a vertical line through those points $c\boldsymbol{v}$. We have infinitely many ***parallel lines*** from (*whole number* c, *any number* d).

(2′) Adding all vectors $d\boldsymbol{w}$ puts a vertical line through every $c\boldsymbol{v}$ on the half-line. Now we have a ***half-plane***. It is the right half of the xy plane (any $x \geq 0$, any height y).

1.1 C Find two equations for the unknowns c and d so that the linear combination $c\boldsymbol{v} + d\boldsymbol{w}$ equals the vector $\boldsymbol{b}$:

$$\boldsymbol{v} = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \qquad \boldsymbol{w} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Solution In applying mathematics, many problems have two parts:

1 *Modeling part* Express the problem by a set of equations.

2 *Computational part* Solve those equations by a fast and accurate algorithm.

Here we are only asked for the first part (the equations). Chapter 2 is devoted to the second part (the algorithm). Our example fits into a fundamental model for linear algebra:

$$\text{Find } c_1,\ldots,c_n \quad \text{so that} \quad c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n = \boldsymbol{b}.$$

For $n = 2$ we could find a formula for the $\boldsymbol{c}$'s. The "elimination method" in Chapter 2 succeeds far beyond $n = 100$. For n greater than 1 million, see Chapter 9. Here $n = 2$:

Vector equation $$c\begin{bmatrix} 2 \\ -1 \end{bmatrix} + d\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

The required equations for c and d just come from the two components separately:

Two scalar equations $$\begin{aligned} 2c - d &= 1 \\ -c + 2d &= 0 \end{aligned}$$

You could think of those as two lines that cross at the solution $c = \frac{2}{3}$, $d = \frac{1}{3}$.

Problem Set 1.1

Problems 1–9 are about addition of vectors and linear combinations.

1 Describe geometrically (line, plane, or all of $\mathbf{R}^3$) all linear combinations of

$$\text{(a)}\ \begin{bmatrix}1\\2\\3\end{bmatrix} \text{ and } \begin{bmatrix}3\\6\\9\end{bmatrix} \qquad \text{(b)}\ \begin{bmatrix}1\\0\\0\end{bmatrix} \text{ and } \begin{bmatrix}0\\2\\3\end{bmatrix} \qquad \text{(c)}\ \begin{bmatrix}2\\0\\0\end{bmatrix} \text{ and } \begin{bmatrix}0\\2\\2\end{bmatrix} \text{ and } \begin{bmatrix}2\\2\\3\end{bmatrix}$$

2 Draw $\boldsymbol{v} = \begin{bmatrix}4\\1\end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix}-2\\2\end{bmatrix}$ and $\boldsymbol{v}+\boldsymbol{w}$ and $\boldsymbol{v}-\boldsymbol{w}$ in a single xy plane.

3 If $\boldsymbol{v} + \boldsymbol{w} = \begin{bmatrix}5\\1\end{bmatrix}$ and $\boldsymbol{v} - \boldsymbol{w} = \begin{bmatrix}1\\5\end{bmatrix}$, compute and draw $\boldsymbol{v}$ and $\boldsymbol{w}$.

4 From $\boldsymbol{v} = \begin{bmatrix}2\\1\end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix}1\\2\end{bmatrix}$, find the components of $3\boldsymbol{v} + \boldsymbol{w}$ and $c\boldsymbol{v} + d\boldsymbol{w}$.

5 Compute $\boldsymbol{u} + \boldsymbol{v} + \boldsymbol{w}$ and $2\boldsymbol{u} + 2\boldsymbol{v} + \boldsymbol{w}$. How do you know $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ lie in a plane?

$$\textbf{In a plane} \qquad \boldsymbol{u} = \begin{bmatrix}1\\2\\3\end{bmatrix}, \quad \boldsymbol{v} = \begin{bmatrix}-3\\1\\-2\end{bmatrix}, \quad \boldsymbol{w} = \begin{bmatrix}2\\-3\\-1\end{bmatrix}.$$

6 Every combination of $\boldsymbol{v} = (1, -2, 1)$ and $\boldsymbol{w} = (0, 1, -1)$ has components that add to ______. Find c and d so that $c\boldsymbol{v} + d\boldsymbol{w} = (3, 3, -6)$.

7 In the xy plane mark all nine of these linear combinations:

$$c\begin{bmatrix}2\\1\end{bmatrix} + d\begin{bmatrix}0\\1\end{bmatrix} \quad \text{with} \quad c = 0, 1, 2 \quad \text{and} \quad d = 0, 1, 2.$$

8 The parallelogram in Figure 1.1 has diagonal $\boldsymbol{v} + \boldsymbol{w}$. What is its other diagonal? What is the sum of the two diagonals? Draw that vector sum.

9 If three corners of a parallelogram are $(1, 1)$, $(4, 2)$, and $(1, 3)$, what are all three of the possible fourth corners? Draw two of them.

Problems 10–14 are about special vectors on cubes and clocks in Figure 1.4.

10 Which point of the cube is $\boldsymbol{i} + \boldsymbol{j}$? Which point is the vector sum of $\boldsymbol{i} = (1, 0, 0)$ and $\boldsymbol{j} = (0, 1, 0)$ and $\boldsymbol{k} = (0, 0, 1)$? Describe all points (x, y, z) in the cube.

11 Four corners of the cube are $(0, 0, 0)$, $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$. What are the other four corners? Find the coordinates of the center point of the cube. The center points of the six faces are ______.

12 How many corners does a cube have in 4 dimensions? How many 3D faces? How many edges? A typical corner is $(0, 0, 1, 0)$. A typical edge goes to $(0, 1, 0, 0)$.

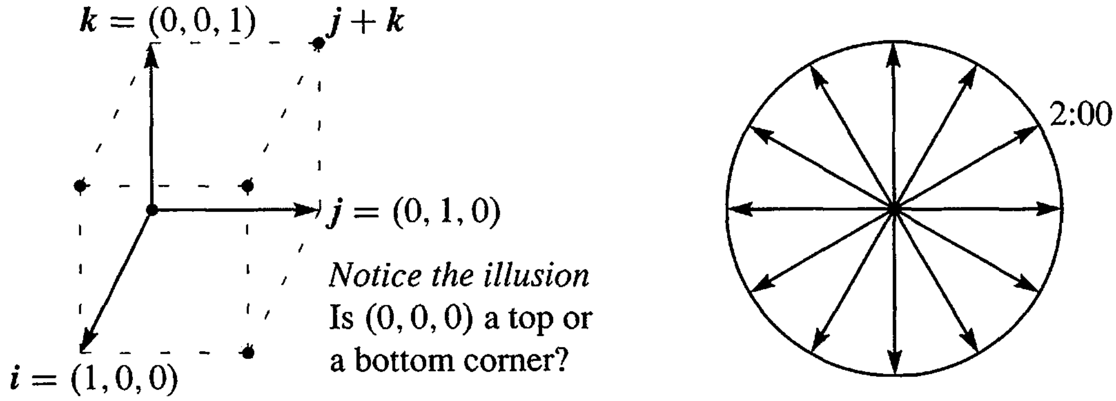


Figure 1.4: Unit cube from $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ and twelve clock vectors.

13 (a) What is the sum V of the twelve vectors that go from the center of a clock to the hours 1:00, 2:00, ..., 12:00?

(b) If the 2:00 vector is removed, why do the 11 remaining vectors add to 8:00?

(c) What are the components of that 2:00 vector $\boldsymbol{v} = (\cos\theta, \sin\theta)$?

14 Suppose the twelve vectors start from 6:00 at the bottom instead of $(0,0)$ at the center. The vector to 12:00 is doubled to $(0,2)$. Add the new twelve vectors.

Problems 15–19 go further with linear combinations of $\boldsymbol{v}$ and $\boldsymbol{w}$ (Figure 1.5a).

15 Figure 1.5a shows $\frac{1}{2}\boldsymbol{v} + \frac{1}{2}\boldsymbol{w}$. Mark the points $\frac{3}{4}\boldsymbol{v} + \frac{1}{4}\boldsymbol{w}$ and $\frac{1}{4}\boldsymbol{v} + \frac{1}{4}\boldsymbol{w}$ and $\boldsymbol{v} + \boldsymbol{w}$.

16 Mark the point $-\boldsymbol{v} + 2\boldsymbol{w}$ and any other combination $c\boldsymbol{v} + d\boldsymbol{w}$ with $c + d = 1$. Draw the line of all combinations that have $c + d = 1$.

17 Locate $\frac{1}{3}\boldsymbol{v} + \frac{1}{3}\boldsymbol{w}$ and $\frac{2}{3}\boldsymbol{v} + \frac{2}{3}\boldsymbol{w}$. The combinations $c\boldsymbol{v} + c\boldsymbol{w}$ fill out what line?

18 Restricted by $0 \le c \le 1$ and $0 \le d \le 1$, shade in all combinations $c\boldsymbol{v} + d\boldsymbol{w}$.

19 Restricted only by $c \ge 0$ and $d \ge 0$ draw the "cone" of all combinations $c\boldsymbol{v} + d\boldsymbol{w}$.

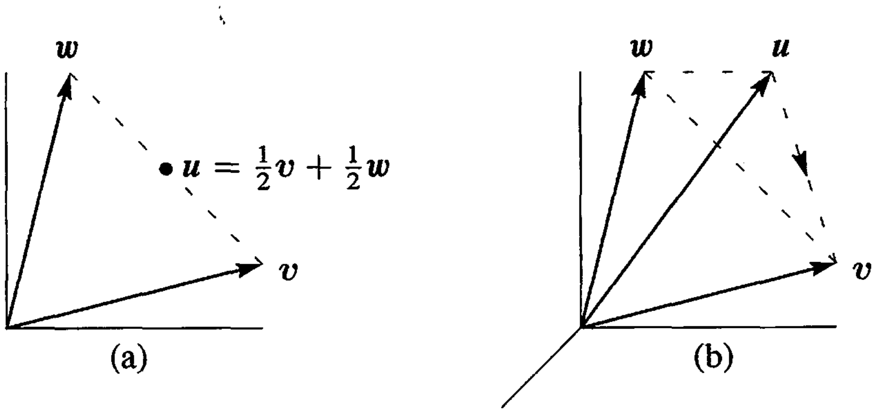


Figure 1.5: Problems **15–19** in a plane Problems **20–25** in 3-dimensional space

Problems 20–25 deal with $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ in three-dimensional space (see Figure 1.5b).

20 Locate $\frac{1}{3}\boldsymbol{u} + \frac{1}{3}\boldsymbol{v} + \frac{1}{3}\boldsymbol{w}$ and $\frac{1}{2}\boldsymbol{u} + \frac{1}{2}\boldsymbol{w}$ in Figure 1.5b. Challenge problem: Under what restrictions on c, d, e, will the combinations $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$ fill in the dashed triangle? To stay in the triangle, one requirement is $c \geq 0, d \geq 0, e \geq 0$.

21 The three sides of the dashed triangle are $\boldsymbol{v} - \boldsymbol{u}$ and $\boldsymbol{w} - \boldsymbol{v}$ and $\boldsymbol{u} - \boldsymbol{w}$. Their sum is _____. Draw the head-to-tail addition around a plane triangle of $(3, 1)$ plus $(-1, 1)$ plus $(-2, -2)$.

22 Shade in the pyramid of combinations $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$ with $c \geq 0, d \geq 0, e \geq 0$ and $c + d + e \leq 1$. Mark the vector $\frac{1}{2}(\boldsymbol{u} + \boldsymbol{v} + \boldsymbol{w})$ as inside or outside this pyramid.

23 If you look at *all* combinations of those $\boldsymbol{u}, \boldsymbol{v}$, and $\boldsymbol{w}$, is there any vector that can't be produced from $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$? Different answer if $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ are all in _____.

24 Which vectors are combinations of $\boldsymbol{u}$ and $\boldsymbol{v}$, and *also* combinations of $\boldsymbol{v}$ and $\boldsymbol{w}$?

25 Draw vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ so that their combinations $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$ fill only a line. Find vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ so that their combinations $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$ fill only a plane.

26 What combination $c\begin{bmatrix} 1 \\ 2 \end{bmatrix} + d\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ produces $\begin{bmatrix} 14 \\ 8 \end{bmatrix}$? Express this question as two equations for the coefficients c and d in the linear combination.

27 *Review Question.* In xyz space, where is the plane of all linear combinations of $\boldsymbol{i} = (1, 0, 0)$ and $\boldsymbol{i} + \boldsymbol{j} = (1, 1, 0)$?

Challenge Problems

28 Find vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ so that $\boldsymbol{v} + \boldsymbol{w} = (4, 5, 6)$ and $\boldsymbol{v} - \boldsymbol{w} = (2, 5, 8)$. This is a question with _____ unknown numbers, and an equal number of equations to find those numbers.

29 Find two different combinations of the three vectors $\boldsymbol{u} = (1, 3)$ and $\boldsymbol{v} = (2, 7)$ and $\boldsymbol{w} = (1, 5)$ that produce $\boldsymbol{b} = (0, 1)$. Slightly delicate question: If I take any three vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ in the plane, will there always be two different combinations that produce $\boldsymbol{b} = (0, 1)$?

30 The linear combinations of $\boldsymbol{v} = (a, b)$ and $\boldsymbol{w} = (c, d)$ fill the plane unless _____. Find four vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{z}$ with four components each so that their combinations $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w} + f\boldsymbol{z}$ produce all vectors (b_1, b_2, b_3, b_4) in four-dimensional space.

31 Write down three equations for c, d, e so that $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w} = \boldsymbol{b}$. Can you somehow find c, d, and e?

$$\boldsymbol{u} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} \quad \boldsymbol{v} = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} \quad \boldsymbol{w} = \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

1.2 Lengths and Dot Products

The first section backed off from multiplying vectors. Now we go forward to define the *"dot product"* of $\boldsymbol{v}$ and $\boldsymbol{w}$. This multiplication involves the separate products v_1w_1 and v_2w_2, but it doesn't stop there. Those two numbers are added to produce the single number $\boldsymbol{v} \cdot \boldsymbol{w}$. *This is the geometry section (lengths and angles).*

DEFINITION The ***dot product*** or ***inner product*** of $\boldsymbol{v} = (v_1, v_2)$ and $\boldsymbol{w} = (w_1, w_2)$ is the number $\boldsymbol{v} \cdot \boldsymbol{w}$:

$$\boldsymbol{v} \cdot \boldsymbol{w} = v_1w_1 + v_2w_2. \tag{1}$$

Example 1 The vectors $\boldsymbol{v} = (4, 2)$ and $\boldsymbol{w} = (-1, 2)$ have a *zero* dot product:

Dot product is zero
Perpendicular vectors
$$\begin{bmatrix} 4 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 2 \end{bmatrix} = -4 + 4 = 0.$$

In mathematics, zero is always a special number. For dot products, it means that *these two vectors are perpendicular.* The angle between them is 90°. When we drew them in Figure 1.1, we saw a rectangle (not just any parallelogram). The clearest example of perpendicular vectors is $\boldsymbol{i} = (1, 0)$ along the x axis and $\boldsymbol{j} = (0, 1)$ up the y axis. Again the dot product is $\boldsymbol{i} \cdot \boldsymbol{j} = 0 + 0 = 0$. Those vectors $\boldsymbol{i}$ and $\boldsymbol{j}$ form a right angle.

The dot product of $\boldsymbol{v} = (1, 2)$ and $\boldsymbol{w} = (3, 1)$ is 5. Soon $\boldsymbol{v} \cdot \boldsymbol{w}$ will reveal the angle between $\boldsymbol{v}$ and $\boldsymbol{w}$ (not 90°). Please check that $\boldsymbol{w} \cdot \boldsymbol{v}$ is also 5.

The dot product $\boldsymbol{w} \cdot \boldsymbol{v}$ ***equals*** $\boldsymbol{v} \cdot \boldsymbol{w}$. The order of $\boldsymbol{v}$ and $\boldsymbol{w}$ makes no difference.

Example 2 Put a weight of 4 at the point $x = -1$ (left of zero) and a weight of 2 at the point $x = 2$ (right of zero). The x axis will balance on the center point (like a see-saw). The weights balance because the dot product is $(4)(-1) + (2)(2) = 0$.

This example is typical of engineering and science. The vector of weights is $(w_1, w_2) = (4, 2)$. The vector of distances from the center is $(v_1, v_2) = (-1, 2)$. The weights times the distances, w_1v_1 and w_2v_2, give the "moments". The equation for the see-saw to balance is $w_1v_1 + w_2v_2 = 0$.

Example 3 Dot products enter in economics and business. We have three goods to buy and sell. Their prices are (p_1, p_2, p_3) for each unit—this is the "price vector" $\boldsymbol{p}$. The quantities we buy or sell are (q_1, q_2, q_3)—positive when we sell, negative when we buy. *Selling* q_1 *units at the price* p_1 *brings in* q_1p_1. The total income (quantities q times prices p) is ***the dot product*** $\boldsymbol{q} \cdot \boldsymbol{p}$ ***in three dimensions***:

$$\textbf{\textit{Income}} = (q_1, q_2, q_3) \cdot (p_1, p_2, p_3) = q_1p_1 + q_2p_2 + q_3p_3 = \textbf{\textit{dot product}}.$$

A zero dot product means that "the books balance". Total sales equal total purchases if $\boldsymbol{q} \cdot \boldsymbol{p} = 0$. Then $\boldsymbol{p}$ is perpendicular to $\boldsymbol{q}$ (in three-dimensional space). A supermarket with thousands of goods goes quickly into high dimensions.

Small note: Spreadsheets have become essential in management. They compute linear combinations and dot products. What you see on the screen is a matrix.

Main point To compute $\boldsymbol{v} \cdot \boldsymbol{w}$, multiply each v_i times w_i. Then add $\Sigma v_i w_i$.

Lengths and Unit Vectors

An important case is the dot product of a vector *with itself.* In this case $\boldsymbol{v}$ equals $\boldsymbol{w}$. When the vector is $\boldsymbol{v} = (1, 2, 3)$, the dot product with itself is $\boldsymbol{v} \cdot \boldsymbol{v} = \|\boldsymbol{v}\|^2 = 14$:

Dot product $\boldsymbol{v} \cdot \boldsymbol{v}$
Length squared

$$\|\boldsymbol{v}\|^2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 1 + 4 + 9 = \mathbf{14}.$$

Instead of a 90° angle between vectors we have 0°. The answer is not zero because $\boldsymbol{v}$ is not perpendicular to itself. The dot product $\boldsymbol{v} \cdot \boldsymbol{v}$ gives the *length of* $\boldsymbol{v}$ *squared.*

DEFINITION The ***length*** $\|\boldsymbol{v}\|$ of a vector $\boldsymbol{v}$ is the square root of $\boldsymbol{v} \cdot \boldsymbol{v}$:

Length = norm($\boldsymbol{v}$) **length** $= \|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}.$

In two dimensions the length is $\sqrt{v_1^2 + v_2^2}$. In three dimensions it is $\sqrt{v_1^2 + v_2^2 + v_3^2}$. By the calculation above, the length of $\boldsymbol{v} = (1, 2, 3)$ is $\|\boldsymbol{v}\| = \sqrt{14}$.

Here $\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}$ is just the ordinary length of the arrow that represents the vector. In two dimensions, the arrow is in a plane. If the components are 1 and 2, the arrow is the third side of a right triangle (Figure 1.6). The Pythagoras formula $a^2 + b^2 = c^2$, which connects the three sides, is $1^2 + 2^2 = \|\boldsymbol{v}\|^2$.

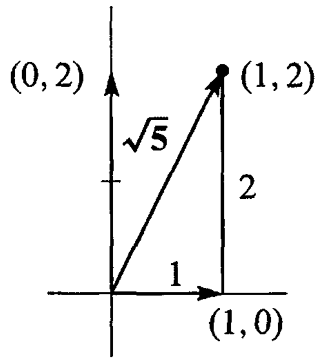
For the length of $\boldsymbol{v} = (1, 2, 3)$, we used the right triangle formula twice. The vector $(1, 2, 0)$ in the base has length $\sqrt{5}$. This base vector is perpendicular to $(0, 0, 3)$ that goes straight up. So the diagonal of the box has length $\|\boldsymbol{v}\| = \sqrt{5 + 9} = \sqrt{14}$.

The length of a four-dimensional vector would be $\sqrt{v_1^2 + v_2^2 + v_3^2 + v_4^2}$. Thus the vector $(1, 1, 1, 1)$ has length $\sqrt{1^2 + 1^2 + 1^2 + 1^2} = 2$. This is the diagonal through a unit cube in four-dimensional space. The diagonal in n dimensions has length $\sqrt{n}$.

The word "**unit**" is always indicating that some measurement equals "one". The unit price is the price for one item. A unit cube has sides of length one. A unit circle is a circle with radius one. Now we define the idea of a "unit vector".

DEFINITION ***A unit vector $\boldsymbol{u}$ is a vector whose length equals one.*** Then $\boldsymbol{u} \cdot \boldsymbol{u} = \mathbf{1}$.

An example in four dimensions is $\boldsymbol{u} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. Then $\boldsymbol{u} \cdot \boldsymbol{u}$ is $\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1$. We divided $\boldsymbol{v} = (1, 1, 1, 1)$ by its length $\|\boldsymbol{v}\| = 2$ to get this unit vector.



$$\begin{aligned} v \cdot v &= v_1^2 + v_2^2 + v_3^2 \\ 5 &= 1^2 + 2^2 \\ 14 &= 1^2 + 2^2 + 3^2 \end{aligned}$$

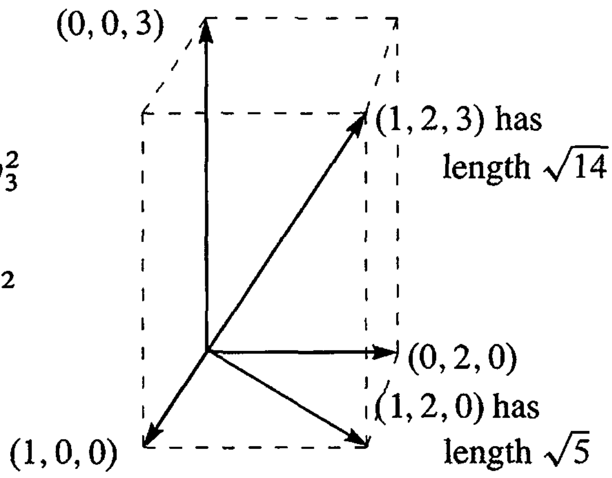


Figure 1.6: The length $\sqrt{v \cdot v}$ of two-dimensional and three-dimensional vectors.

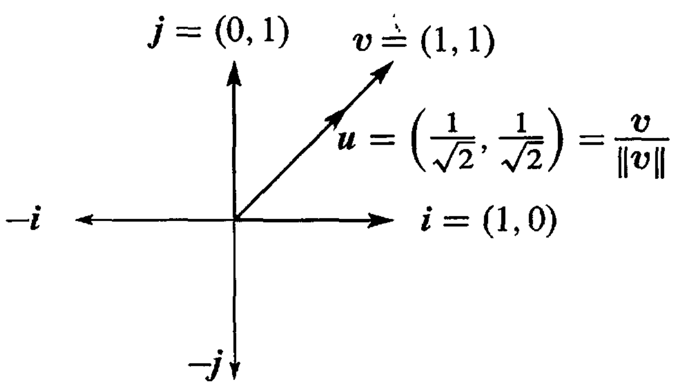
Example 4 The standard unit vectors along the x and y axes are written $\boldsymbol{i}$ and $\boldsymbol{j}$. In the xy plane, the unit vector that makes an angle "theta" with the x axis is $(\cos\theta, \sin\theta)$:

$$\textbf{Unit vectors} \quad \boldsymbol{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{u} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}.$$

When $\theta = 0$, the horizontal vector $\boldsymbol{u}$ is $\boldsymbol{i}$. When $\theta = 90^\circ$ (or $\frac{\pi}{2}$ radians), the vertical vector is $\boldsymbol{j}$. At any angle, the components $\cos\theta$ and $\sin\theta$ produce $\boldsymbol{u} \cdot \boldsymbol{u} = 1$ because $\cos^2\theta + \sin^2\theta = 1$. These vectors reach out to the unit circle in Figure 1.7. Thus $\cos\theta$ and $\sin\theta$ are simply the coordinates of that point at angle θ on the unit circle.

Since $(2, 2, 1)$ has length 3, the vector $(\frac{2}{3}, \frac{2}{3}, \frac{1}{3})$ has length 1. Check that $\boldsymbol{u} \cdot \boldsymbol{u} = \frac{4}{9} + \frac{4}{9} + \frac{1}{9} = 1$. For a unit vector, **divide any nonzero v by its length $\|v\|$**.

Unit vector $\quad \boldsymbol{u} = v/\|v\|$ **is a unit vector in the same direction as v.**



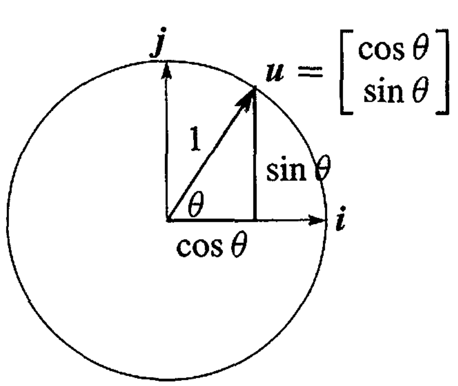


Figure 1.7: The coordinate vectors $\boldsymbol{i}$ and $\boldsymbol{j}$. The unit vector $\boldsymbol{u}$ at angle 45° (left) divides $v = (1, 1)$ by its length $\|v\| = \sqrt{2}$. The unit vector $\boldsymbol{u} = (\cos\theta, \sin\theta)$ is at angle θ.

The Angle Between Two Vectors

We stated that perpendicular vectors have $\boldsymbol{v} \cdot \boldsymbol{w} = 0$. The dot product is zero when the angle is 90°. To explain this, we have to connect angles to dot products. Then we show how $\boldsymbol{v} \cdot \boldsymbol{w}$ finds the angle between any two nonzero vectors $\boldsymbol{v}$ and $\boldsymbol{w}$.

Right angles *The dot product is* $\boldsymbol{v} \cdot \boldsymbol{w} = 0$ *when* $\boldsymbol{v}$ *is perpendicular to* $\boldsymbol{w}$.

Proof When $\boldsymbol{v}$ and $\boldsymbol{w}$ are perpendicular, they form two sides of a right triangle. The third side is $\boldsymbol{v} - \boldsymbol{w}$ (the hypotenuse going across in Figure 1.8). The *Pythagoras Law* for the sides of a right triangle is $a^2 + b^2 = c^2$:

$$\textbf{Perpendicular vectors} \quad \|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2 = \|\boldsymbol{v} - \boldsymbol{w}\|^2 \tag{2}$$

Writing out the formulas for those lengths in two dimensions, this equation is

$$\textbf{Pythagoras} \qquad (v_1^2 + v_2^2) + (w_1^2 + w_2^2) = (v_1 - w_1)^2 + (v_2 - w_2)^2. \tag{3}$$

The right side begins with $v_1^2 - 2v_1w_1 + w_1^2$. Then v_1^2 and w_1^2 are on both sides of the equation and they cancel, leaving $-2v_1w_1$. Also v_2^2 and w_2^2 cancel, leaving $-2v_2w_2$. (In three dimensions there would be $-2v_3w_3$.) Now divide by -2:

$$0 = -2v_1w_1 - 2v_2w_2 \quad \text{which leads to} \quad \boldsymbol{v_1w_1 + v_2w_2 = 0}. \tag{4}$$

Conclusion Right angles produce $\boldsymbol{v} \cdot \boldsymbol{w} = 0$. The dot product is zero when the angle is $\theta = 90°$. Then $\cos\theta = 0$. The zero vector $\boldsymbol{v} = \boldsymbol{0}$ is perpendicular to every vector $\boldsymbol{w}$ because $\boldsymbol{0} \cdot \boldsymbol{w}$ is always zero.

Now suppose $\boldsymbol{v} \cdot \boldsymbol{w}$ is **not zero.** It may be positive, it may be negative. The sign of $\boldsymbol{v} \cdot \boldsymbol{w}$ immediately tells whether we are below or above a right angle. The angle is less than 90° when $\boldsymbol{v} \cdot \boldsymbol{w}$ is positive. The angle is above 90° when $\boldsymbol{v} \cdot \boldsymbol{w}$ is negative. The right side of Figure 1.8 shows a typical vector $\boldsymbol{v} = (3, 1)$. The angle with $\boldsymbol{w} = (1, 3)$ is less than 90° because $\boldsymbol{v} \cdot \boldsymbol{w} = 6$ is positive.

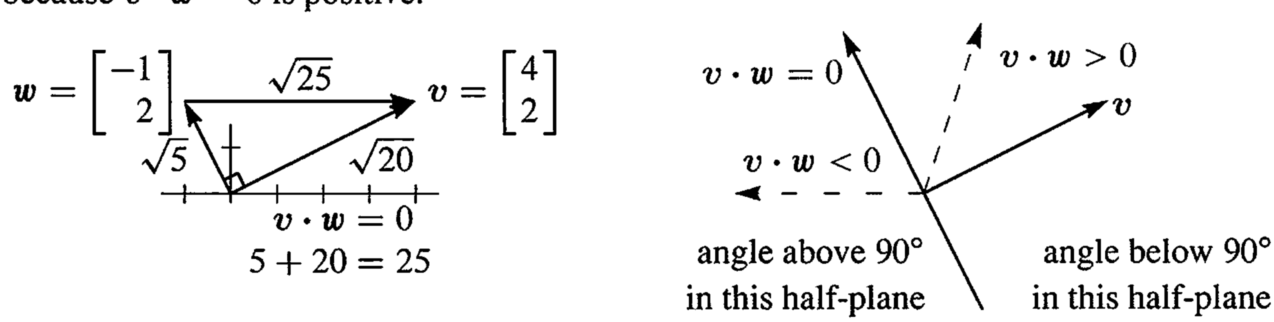


Figure 1.8: Perpendicular vectors have $\boldsymbol{v} \cdot \boldsymbol{w} = 0$. Then $\|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2 = \|\boldsymbol{v} - \boldsymbol{w}\|^2$.

The borderline is where vectors are perpendicular to $\boldsymbol{v}$. On that dividing line between plus and minus, $(1, -3)$ is perpendicular to $(3, 1)$. The dot product is zero.

The dot product reveals the exact angle θ. This is not necessary for linear algebra—you could stop here! Once we have matrices, we won't come back to θ. But while we are on the subject of angles, this is the place for the formula.

Start with **unit vectors $\boldsymbol{u}$ and $\boldsymbol{U}$**. The sign of $\boldsymbol{u} \cdot \boldsymbol{U}$ tells whether $\theta < 90°$ or $\theta > 90°$. Because the vectors have length 1, we learn more than that. ***The dot product $\boldsymbol{u} \cdot \boldsymbol{U}$ is the cosine of θ.*** This is true in any number of dimensions.

Unit vectors $\boldsymbol{u}$ and $\boldsymbol{U}$ at angle θ have $\boldsymbol{u} \cdot \boldsymbol{U} = \cos\theta$. **Certainly** $|\boldsymbol{u} \cdot \boldsymbol{U}| \leq 1$.

Remember that $\cos\theta$ is never greater than 1. It is never less than -1. ***The dot product of unit vectors is between -1 and 1.***

Figure 1.9 shows this clearly when the vectors are $\boldsymbol{u} = (\cos\theta, \sin\theta)$ and $\boldsymbol{i} = (1, 0)$. The dot product is $\boldsymbol{u} \cdot \boldsymbol{i} = \cos\theta$. That is the cosine of the angle between them.

After rotation through any angle α, these are still unit vectors. The vector $\boldsymbol{i} = (1, 0)$ rotates to $(\cos\alpha, \sin\alpha)$. The vector $\boldsymbol{u}$ rotates to $(\cos\beta, \sin\beta)$ with $\beta = \alpha + \theta$. Their dot product is $\cos\alpha\cos\beta + \sin\alpha\sin\beta$. From trigonometry this is the same as $\cos(\beta - \alpha)$. But $\beta - \alpha$ is the angle θ, so the dot product is $\cos\theta$.

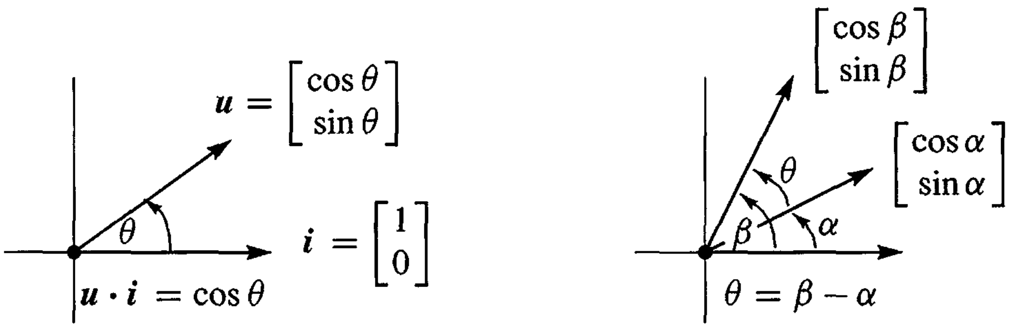


Figure 1.9: The dot product of unit vectors is the cosine of the angle θ.

Problem 24 proves $|\boldsymbol{u} \cdot \boldsymbol{U}| \leq 1$ directly, without mentioning angles. The inequality and the cosine formula $\boldsymbol{u} \cdot \boldsymbol{U} = \cos\theta$ are always true for unit vectors.

What if $\boldsymbol{v}$ and $\boldsymbol{w}$ are not unit vectors? Divide by their lengths to get $\boldsymbol{u} = \boldsymbol{v}/\|\boldsymbol{v}\|$ and $\boldsymbol{U} = \boldsymbol{w}/\|\boldsymbol{w}\|$. Then the dot product of those unit vectors $\boldsymbol{u}$ and $\boldsymbol{U}$ gives $\cos\theta$.

COSINE FORMULA If $\boldsymbol{v}$ and $\boldsymbol{w}$ are nonzero vectors then $\dfrac{\boldsymbol{v} \cdot \boldsymbol{w}}{\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|} = \cos\theta$.

Whatever the angle, this dot product of $v/\|v\|$ with $w/\|w\|$ never exceeds one. That is the ***"Schwarz inequality"*** $|v \cdot w| \leq \|v\|\,\|w\|$ for dot products—or more correctly the Cauchy-Schwarz-Buniakowsky inequality. It was found in France and Germany and Russia (and maybe elsewhere—it is the most important inequality in mathematics).

Since $|\cos\theta|$ never exceeds 1, the cosine formula gives two great inequalities:

SCHWARZ INEQUALITY	$\lvert v \cdot w\rvert \leq \|v\|\,\|w\|$
TRIANGLE INEQUALITY	$\|v + w\| \leq \|v\| + \|w\|$

Example 5 Find $\cos\theta$ for $v = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $w = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and check both inequalities.

Solution The dot product is $v \cdot w = 4$. Both v and w have length $\sqrt{5}$. The cosine is $4/5$.

$$\cos\theta = \frac{v \cdot w}{\|v\|\,\|w\|} = \frac{4}{\sqrt{5}\sqrt{5}} = \frac{4}{5}.$$

The angle is below 90° because $v \cdot w = 4$ is positive. By the Schwarz inequality, $v \cdot w = 4$ is less than $\|v\|\,\|w\| = 5$. Side $3 = \|v + w\|$ is less than side 1 + side 2, by the triangle inequality. For $v + w = (3, 3)$ that says $\sqrt{18} < \sqrt{5} + \sqrt{5}$. Square this to get $18 < 20$.

Example 6 The dot product of $v = (a, b)$ and $w = (b, a)$ is $2ab$. Both lengths are $\sqrt{a^2 + b^2}$. The Schwarz inequality in this case says that $2ab \leq a^2 + b^2$.

This is more famous if we write $x = a^2$ and $y = b^2$. The "geometric mean" $\sqrt{xy}$ is not larger than the "arithmetic mean" = average $\frac{1}{2}(x + y)$.

$$\begin{matrix}\textbf{Geometric} \\ \textbf{mean}\end{matrix} \leq \begin{matrix}\textbf{Arithmetic} \\ \textbf{mean}\end{matrix} \qquad ab \leq \frac{a^2 + b^2}{2} \quad \text{becomes} \quad \sqrt{xy} \leq \frac{x + y}{2}.$$

Example 5 had $a = 2$ and $b = 1$. So $x = 4$ and $y = 1$. The geometric mean $\sqrt{xy} = 2$ is below the arithmetic mean $\frac{1}{2}(1 + 4) = 2.5$.

Notes on Computing

Write the components of v as $v(1), \ldots, v(N)$ and similarly for w. In FORTRAN, the sum $v + w$ requires a loop to add components separately. The dot product also uses a loop to add the separate $v(j)w(j)$. Here are VPLUSW and VDOTW:

FORTRAN

```
   DO 10 J = 1,N                    DO 10 J = 1,N
10 VPLUSW(J) = v(J) + w(J)      10 VDOTW = VDOTW + V(J) * W(J)
```

MATLAB *and also* PYTHON *work directly with whole vectors, not their components.* No loop is needed. When v and w have been defined, $v + w$ is immediately understood.

Input v and w as rows—the prime $'$ transposes them to columns. $2v + 3w$ uses $*$ for multiplication by 2 and 3. The result will be printed unless the line ends in a semicolon.

MATLAB $\quad v = [2 \;\; 3 \;\; 4]' \;\; ; \quad w = [1 \;\; 1 \;\; 1]' \;\; ; \quad u = 2 * v + 3 * w$

The dot product $v \cdot w$ is usually seen as ***a row times a column*** (***with no dot***):

$$\text{Instead of} \quad \begin{bmatrix} 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \end{bmatrix} \quad \text{we more often see} \quad [1 \;\; 2] \begin{bmatrix} 3 \\ 4 \end{bmatrix} \quad \text{or} \quad v' * w$$

The length of v is known to MATLAB as norm (v). We could define it ourselves as sqrt $(v' * v)$, using the square root function—also known. The cosine we have to define ourselves! The angle (in radians) comes from the *arc cosine* (acos) function:

Cosine formula $\quad$ cosine $= v' * w/(\text{norm}\,(v) * \text{norm}\,(w))$

Angle formula $\quad$ angle $=$ acos (cosine)

An M-file would create a new function **cosine** (v, w) for future use. The M-files created especially for this book are listed at the end. R and PYTHON are open source software.

■ REVIEW OF THE KEY IDEAS ■

1. The dot product $v \cdot w$ multiplies each component v_i by w_i and adds all $v_i w_i$.
2. The length $\|v\|$ of a vector is the square root of $v \cdot v$.
3. $u = v/\|v\|$ is a ***unit vector***. Its length is 1.
4. The dot product is $v \cdot w = 0$ when vectors v and w are perpendicular.
5. The cosine of θ (the angle between any nonzero v and w) never exceeds 1:

$$\cos\theta = \frac{v \cdot w}{\|v\|\,\|w\|} \qquad \textit{Schwarz inequality} \quad |v \cdot w| \le \|v\|\,\|w\|.$$

Problem 21 will produce the ***triangle inequality*** $\|v + w\| \le \|v\| + \|w\|$.

■ WORKED EXAMPLES ■

1.2 A For the vectors $v = (3, 4)$ and $w = (4, 3)$ test the Schwarz inequality on $v \cdot w$ and the triangle inequality on $\|v + w\|$. Find $\cos\theta$ for the angle between v and w. When will we have *equality* $|v \cdot w| = \|v\|\,\|w\|$ and $\|v + w\| = \|v\| + \|w\|$?

Solution The dot product is $\boldsymbol{v} \cdot \boldsymbol{w} = (3)(4) + (4)(3) = 24$. The length of $\boldsymbol{v}$ is $\|\boldsymbol{v}\| = \sqrt{9+16} = 5$ and also $\|\boldsymbol{w}\| = 5$. The sum $\boldsymbol{v} + \boldsymbol{w} = (7,7)$ has length $7\sqrt{2} < 10$.

Schwarz inequality	$\lvert\boldsymbol{v} \cdot \boldsymbol{w}\rvert \le \|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$	is $24 < 25$.
Triangle inequality	$\|\boldsymbol{v} + \boldsymbol{w}\| \le \|\boldsymbol{v}\| + \|\boldsymbol{w}\|$	is $7\sqrt{2} < 5 + 5$.
Cosine of angle	$\cos\theta = \frac{24}{25}$	Thin angle from $\boldsymbol{v} = (3,4)$ to $\boldsymbol{w} = (4,3)$

Suppose one vector is a multiple of the other as in $\boldsymbol{w} = c\boldsymbol{v}$. Then the angle is $0°$ or $180°$. In this case $\lvert\cos\theta\rvert = 1$ and $\lvert\boldsymbol{v} \cdot \boldsymbol{w}\rvert$ *equals* $\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$. If the angle is $0°$, as in $\boldsymbol{w} = 2\boldsymbol{v}$, then $\|\boldsymbol{v} + \boldsymbol{w}\| = \|\boldsymbol{v}\| + \|\boldsymbol{w}\|$. The triangle is completely flat.

1.2 B Find a unit vector $\boldsymbol{u}$ in the direction of $\boldsymbol{v} = (3,4)$. Find a unit vector $\boldsymbol{U}$ that is perpendicular to $\boldsymbol{u}$. How many possibilities for $\boldsymbol{U}$?

Solution For a unit vector $\boldsymbol{u}$, divide $\boldsymbol{v}$ by its length $\|\boldsymbol{v}\| = 5$. For a perpendicular vector $\boldsymbol{V}$ we can choose $(-4,3)$ since the dot product $\boldsymbol{v} \cdot \boldsymbol{V}$ is $(3)(-4) + (4)(3) = 0$. For a *unit* vector $\boldsymbol{U}$, divide $\boldsymbol{V}$ by its length $\|\boldsymbol{V}\|$:

$$\boldsymbol{u} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} = \left(\frac{3}{5}, \frac{4}{5}\right) \qquad \boldsymbol{U} = \frac{\boldsymbol{V}}{\|\boldsymbol{V}\|} = \left(-\frac{4}{5}, \frac{3}{5}\right) \qquad \boldsymbol{u} \cdot \boldsymbol{U} = 0$$

The only other perpendicular unit vector would be $-\boldsymbol{U} = (\frac{4}{5}, -\frac{3}{5})$.

1.2 C Find a vector $\boldsymbol{x} = (c,d)$ that has dot products $\boldsymbol{x} \cdot \boldsymbol{r} = 1$ and $\boldsymbol{x} \cdot \boldsymbol{s} = 0$ with the given vectors $\boldsymbol{r} = (2,-1)$ and $\boldsymbol{s} = (-1,2)$.

How is this question related to Example **1.1 C**, which solved $c\boldsymbol{v} + d\boldsymbol{w} = \boldsymbol{b} = (1,0)$?

Solution Those two dot products give linear equations for c and d. Then $\boldsymbol{x} = (c,d)$.

$$\begin{array}{lrl} \boldsymbol{x} \cdot \boldsymbol{r} = 1 & 2c - \ \ d = 1 & \textbf{The same equations as} \\ \boldsymbol{x} \cdot \boldsymbol{s} = 0 & -\ c + 2d = 0 & \textbf{in Worked Example 1.1 C} \end{array}$$

The second equation makes $\boldsymbol{x}$ perpendicular to $\boldsymbol{s} = (-1,2)$. So I can see the geometry: Go in the perpendicular direction $(2,1)$. When you reach $\boldsymbol{x} = \frac{1}{3}(2,1)$, the dot product with $\boldsymbol{r} = (2,-1)$ has the required value $\boldsymbol{x} \cdot \boldsymbol{r} = 1$.

Comment on n equations for $\boldsymbol{x} = (x_1, \ldots, x_n)$ *in n-dimensional space*

Section 1.1 would start with column vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$. The goal is to combine them to produce a required vector $x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n = \boldsymbol{b}$. This section would start from vectors $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n$. Now the goal is to find $\boldsymbol{x}$ with the required dot products $\boldsymbol{x} \cdot \boldsymbol{r}_i = b_i$.

Soon the $\boldsymbol{v}$'s will be the columns of a matrix A, and the $\boldsymbol{r}$'s will be the rows of A. Then the (one and only) problem will be to solve $A\boldsymbol{x} = \boldsymbol{b}$.

Problem Set 1.2

1 Calculate the dot products $\boldsymbol{u} \cdot \boldsymbol{v}$ and $\boldsymbol{u} \cdot \boldsymbol{w}$ and $\boldsymbol{u} \cdot (\boldsymbol{v} + \boldsymbol{w})$ and $\boldsymbol{w} \cdot \boldsymbol{v}$:

$$\boldsymbol{u} = \begin{bmatrix} -.6 \\ .8 \end{bmatrix} \qquad \boldsymbol{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \qquad \boldsymbol{w} = \begin{bmatrix} 8 \\ 6 \end{bmatrix}.$$

2 Compute the lengths $\|\boldsymbol{u}\|$ and $\|\boldsymbol{v}\|$ and $\|\boldsymbol{w}\|$ of those vectors. Check the Schwarz inequalities $|\boldsymbol{u} \cdot \boldsymbol{v}| \leq \|\boldsymbol{u}\|\,\|\boldsymbol{v}\|$ and $|\boldsymbol{v} \cdot \boldsymbol{w}| \leq \|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$.

3 Find unit vectors in the directions of $\boldsymbol{v}$ and $\boldsymbol{w}$ in Problem 1, and the cosine of the angle θ. Choose vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ that make $0°$, $90°$, and $180°$ angles with $\boldsymbol{w}$.

4 For any *unit* vectors $\boldsymbol{v}$ and $\boldsymbol{w}$, find the dot products (actual numbers) of

(a) $\boldsymbol{v}$ and $-\boldsymbol{v}$ (b) $\boldsymbol{v} + \boldsymbol{w}$ and $\boldsymbol{v} - \boldsymbol{w}$ (c) $\boldsymbol{v} - 2\boldsymbol{w}$ and $\boldsymbol{v} + 2\boldsymbol{w}$

5 Find unit vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ in the directions of $\boldsymbol{v} = (3, 1)$ and $\boldsymbol{w} = (2, 1, 2)$. Find unit vectors $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ that are perpendicular to $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$.

6 (a) Describe every vector $\boldsymbol{w} = (w_1, w_2)$ that is perpendicular to $\boldsymbol{v} = (2, -1)$.

(b) The vectors that are perpendicular to $\boldsymbol{V} = (1, 1, 1)$ lie on a _____.

(c) The vectors that are perpendicular to $(1, 1, 1)$ and $(1, 2, 3)$ lie on a _____.

7 Find the angle θ (from its cosine) between these pairs of vectors:

(a) $\boldsymbol{v} = \begin{bmatrix} 1 \\ \sqrt{3} \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ (b) $\boldsymbol{v} = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}$

(c) $\boldsymbol{v} = \begin{bmatrix} 1 \\ \sqrt{3} \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} -1 \\ \sqrt{3} \end{bmatrix}$ (d) $\boldsymbol{v} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$.

8 True or false (give a reason if true or a counterexample if false):

(a) If $\boldsymbol{u}$ is perpendicular (in three dimensions) to $\boldsymbol{v}$ and $\boldsymbol{w}$, those vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ are parallel.

(b) If $\boldsymbol{u}$ is perpendicular to $\boldsymbol{v}$ and $\boldsymbol{w}$, then $\boldsymbol{u}$ is perpendicular to $\boldsymbol{v} + 2\boldsymbol{w}$.

(c) If $\boldsymbol{u}$ and $\boldsymbol{v}$ are perpendicular unit vectors then $\|\boldsymbol{u} - \boldsymbol{v}\| = \sqrt{2}$.

9 The slopes of the arrows from $(0, 0)$ to (v_1, v_2) and (w_1, w_2) are v_2/v_1 and w_2/w_1. **Suppose the product v_2w_2/v_1w_1 of those slopes is -1.** Show that $\boldsymbol{v} \cdot \boldsymbol{w} = 0$ and the vectors are perpendicular.

10 Draw arrows from $(0, 0)$ to the points $\boldsymbol{v} = (1, 2)$ and $\boldsymbol{w} = (-2, 1)$. Multiply their slopes. That answer is a signal that $\boldsymbol{v} \cdot \boldsymbol{w} = 0$ and the arrows are _____.

11 If $\boldsymbol{v} \cdot \boldsymbol{w}$ is negative, what does this say about the angle between $\boldsymbol{v}$ and $\boldsymbol{w}$? Draw a 3-dimensional vector $\boldsymbol{v}$ (an arrow), and show where to find all $\boldsymbol{w}$'s with $\boldsymbol{v} \cdot \boldsymbol{w} < 0$.

12 With $\boldsymbol{v} = (1,1)$ and $\boldsymbol{w} = (1,5)$ choose a number c so that $\boldsymbol{w} - c\boldsymbol{v}$ is perpendicular to $\boldsymbol{v}$. Then find the formula that gives this number c for any nonzero $\boldsymbol{v}$ and $\boldsymbol{w}$. (*Note*: $c\boldsymbol{v}$ is the "projection" of $\boldsymbol{w}$ onto $\boldsymbol{v}$.)

13 Find two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ that are perpendicular to $(1,0,1)$ and to each other.

14 Find nonzero vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ that are perpendicular to $(1,1,1,1)$ and to each other.

15 The geometric mean of $x = 2$ and $y = 8$ is $\sqrt{xy} = 4$. The arithmetic mean is larger: $\frac{1}{2}(x + y) =$ _____. This would come in Example 6 from the Schwarz inequality for $\boldsymbol{v} = (\sqrt{2}, \sqrt{8})$ and $\boldsymbol{w} = (\sqrt{8}, \sqrt{2})$. Find $\cos\theta$ for this $\boldsymbol{v}$ and $\boldsymbol{w}$.

16 **How long is the vector $\boldsymbol{v} = (1,1,\ldots,1)$ in 9 dimensions**? Find a unit vector $\boldsymbol{u}$ in the same direction as $\boldsymbol{v}$ and a unit vector $\boldsymbol{w}$ that is perpendicular to $\boldsymbol{v}$.

17 What are the cosines of the angles α, β, θ between the vector $(1,0,-1)$ and the unit vectors $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ along the axes? Check the formula $\cos^2\alpha + \cos^2\beta + \cos^2\theta = 1$.

Problems 18–31 lead to the main facts about lengths and angles in triangles.

18 The parallelogram with sides $\boldsymbol{v} = (4,2)$ and $\boldsymbol{w} = (-1,2)$ is a rectangle. Check the Pythagoras formula $a^2 + b^2 = c^2$ which is for ***right triangles only***:

$$\text{(length of } \boldsymbol{v})^2 + (\text{length of } \boldsymbol{w})^2 = (\text{length of } \boldsymbol{v} + \boldsymbol{w})^2.$$

19 (Rules for dot products) These equations are simple but useful:

(1) $\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{w} \cdot \boldsymbol{v}$ (2) $\boldsymbol{u} \cdot (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{u} \cdot \boldsymbol{w}$ (3) $(c\boldsymbol{v}) \cdot \boldsymbol{w} = c(\boldsymbol{v} \cdot \boldsymbol{w})$

Use (2) with $\boldsymbol{u} = \boldsymbol{v} + \boldsymbol{w}$ to prove $\|\boldsymbol{v} + \boldsymbol{w}\|^2 = \boldsymbol{v} \cdot \boldsymbol{v} + 2\boldsymbol{v} \cdot \boldsymbol{w} + \boldsymbol{w} \cdot \boldsymbol{w}$.

20 The "Law of Cosines" comes from $(\boldsymbol{v} - \boldsymbol{w}) \cdot (\boldsymbol{v} - \boldsymbol{w}) = \boldsymbol{v} \cdot \boldsymbol{v} - 2\boldsymbol{v} \cdot \boldsymbol{w} + \boldsymbol{w} \cdot \boldsymbol{w}$:

$$\textbf{Cosine Law} \qquad \|\boldsymbol{v} - \boldsymbol{w}\|^2 = \|\boldsymbol{v}\|^2 - 2\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|\cos\theta + \|\boldsymbol{w}\|^2.$$

If $\theta < 90^\circ$ show that $\|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2$ is larger than $\|\boldsymbol{v} - \boldsymbol{w}\|^2$ (the third side).

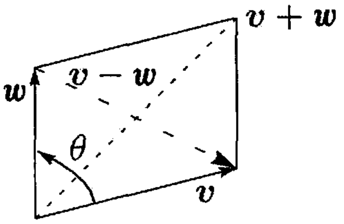
21 The ***triangle inequality*** says: (length of $\boldsymbol{v} + \boldsymbol{w}$) $\leq$ (length of $\boldsymbol{v}$) + (length of $\boldsymbol{w}$). Problem 19 found $\|\boldsymbol{v} + \boldsymbol{w}\|^2 = \|\boldsymbol{v}\|^2 + 2\boldsymbol{v} \cdot \boldsymbol{w} + \|\boldsymbol{w}\|^2$. Use the Schwarz inequality $\boldsymbol{v} \cdot \boldsymbol{w} \leq \|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$ to show that $\|$**side 3**$\|$ can not exceed $\|$**side 1**$\|$ + $\|$**side 2**$\|$:

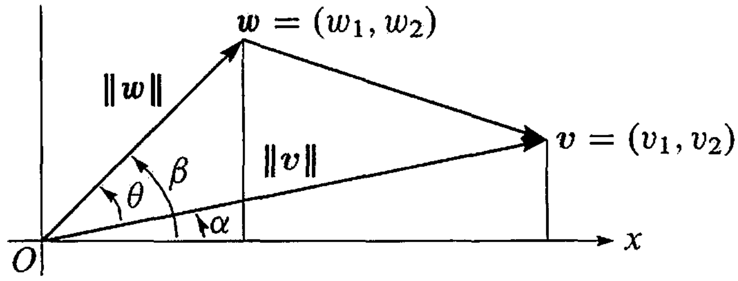
$$\textbf{Triangle inequality} \qquad \|\boldsymbol{v} + \boldsymbol{w}\|^2 \leq (\|\boldsymbol{v}\| + \|\boldsymbol{w}\|)^2 \quad \textbf{or} \quad \|\boldsymbol{v} + \boldsymbol{w}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{w}\|.$$

22 The Schwarz inequality $|\boldsymbol{v} \cdot \boldsymbol{w}| \leq \|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$ by algebra instead of trigonometry:

(a) Multiply out both sides of $(v_1w_1 + v_2w_2)^2 \leq (v_1^2 + v_2^2)(w_1^2 + w_2^2)$.

(b) Show that the difference between those two sides equals $(v_1w_2 - v_2w_1)^2$. This cannot be negative since it is a square—so the inequality is true.





23 The figure shows that $\cos\alpha = v_1/\|v\|$ and $\sin\alpha = v_2/\|v\|$. Similarly $\cos\beta$ is _____ and $\sin\beta$ is _____. The angle θ is $\beta - \alpha$. Substitute into the trigonometry formula $\cos\beta\cos\alpha + \sin\beta\sin\alpha$ for $\cos(\beta - \alpha)$ to find $\cos\theta = \boldsymbol{v}\cdot\boldsymbol{w}/\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$.

24 One-line proof of the Schwarz inequality $|\boldsymbol{u}\cdot\boldsymbol{U}| \le 1$ for unit vectors:

$$|\boldsymbol{u}\cdot\boldsymbol{U}| \le |u_1|\,|U_1| + |u_2|\,|U_2| \le \frac{u_1^2 + U_1^2}{2} + \frac{u_2^2 + U_2^2}{2} = \frac{1+1}{2} = 1.$$

Put $(u_1, u_2) = (.6, .8)$ and $(U_1, U_2) = (.8, .6)$ in that whole line and find $\cos\theta$.

25 Why is $|\cos\theta|$ never greater than 1 in the first place?

26 If $\boldsymbol{v} = (1, 2)$ draw all vectors $\boldsymbol{w} = (x, y)$ in the xy plane with $\boldsymbol{v}\cdot\boldsymbol{w} = x + 2y = 5$. Which is the shortest $\boldsymbol{w}$?

27 (*Recommended*) If $||\boldsymbol{v}|| = 5$ and $||\boldsymbol{w}|| = 3$, what are the smallest and largest values of $||\boldsymbol{v} - \boldsymbol{w}||$? What are the smallest and largest values of $\boldsymbol{v}\cdot\boldsymbol{w}$?

Challenge Problems

28 Can three vectors in the xy plane have $\boldsymbol{u}\cdot\boldsymbol{v} < 0$ and $\boldsymbol{v}\cdot\boldsymbol{w} < 0$ and $\boldsymbol{u}\cdot\boldsymbol{w} < 0$? I don't know how many vectors in xyz space can have all negative dot products. (Four of those vectors in the plane would certainly be impossible ...).

29 Pick any numbers that add to $x + y + z = 0$. Find the angle between your vector $\boldsymbol{v} = (x, y, z)$ and the vector $\boldsymbol{w} = (z, x, y)$. Challenge question: Explain why $\boldsymbol{v}\cdot\boldsymbol{w}/\|\boldsymbol{v}\|\|\boldsymbol{w}\|$ is always $-\frac{1}{2}$.

30 How could you prove $\sqrt[3]{xyz} \le \frac{1}{3}(x+y+z)$ (geometric mean $\le$ arithmetic mean)?

31 Find four perpendicular unit vectors with all components equal to $\frac{1}{2}$ or $-\frac{1}{2}$.

32 Using v = randn(3, 1) in MATLAB, create a random unit vector $\boldsymbol{u} = \boldsymbol{v}/\|\boldsymbol{v}\|$. Using V = randn(3, 30) create 30 more random unit vectors $\boldsymbol{U}_j$. What is the average size of the dot products $|\boldsymbol{u}\cdot\boldsymbol{U}_j|$? In calculus, the average $\int_0^\pi |\cos\theta|\,d\theta/\pi = 2/\pi$.

1.3 Matrices

This section is based on two carefully chosen examples. They both start with three vectors. I will take their combinations using *matrices*. The three vectors in the first example are $\boldsymbol{u}, \boldsymbol{v}$, and $\boldsymbol{w}$:

$$\textbf{First example} \qquad \boldsymbol{u} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \qquad \boldsymbol{v} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \qquad \boldsymbol{w} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Their linear combinations in three-dimensional space are $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$:

$$\textbf{Combinations} \qquad c \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + d \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} + e \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} c \\ d-c \\ e-d \end{bmatrix}. \qquad (1)$$

Now something important: *Rewrite that combination using a matrix.* The vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ go into the columns of the matrix A. That matrix "multiplies" a vector:

$$\begin{array}{l}\textbf{Same combination} \\ \textbf{is now } A \textbf{ times } \boldsymbol{x}\end{array} \qquad \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} c \\ d \\ e \end{bmatrix} = \begin{bmatrix} c \\ d-c \\ e-d \end{bmatrix}. \qquad (2)$$

The numbers c, d, e are the components of a vector $\boldsymbol{x}$. The matrix A times the vector $\boldsymbol{x}$ is the same as the combination $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$ of the three columns:

$$\textbf{Matrix times vector} \qquad A\boldsymbol{x} = \begin{bmatrix} & & \\ \boldsymbol{u} & \boldsymbol{v} & \boldsymbol{w} \\ & & \end{bmatrix} \begin{bmatrix} c \\ d \\ e \end{bmatrix} = c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}. \qquad (3)$$

This is more than a definition of $A\boldsymbol{x}$, because the rewriting brings a crucial change in viewpoint. At first, the numbers c, d, e were multiplying the vectors. Now the matrix is multiplying those numbers. **The matrix A acts on the vector $\boldsymbol{x}$.** The result $A\boldsymbol{x}$ is a combination $\boldsymbol{b}$ of the columns of A.

To see that action, I will write x_1, x_2, x_3 instead of c, d, e. I will write b_1, b_2, b_3 for the components of $A\boldsymbol{x}$. With new letters we see

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \boldsymbol{b}. \qquad (4)$$

The input is $\boldsymbol{x}$ and the output is $\boldsymbol{b} = A\boldsymbol{x}$. This A is a "**difference matrix**" because $\boldsymbol{b}$ contains differences of the input vector $\boldsymbol{x}$. The top difference is $x_1 - x_0 = x_1 - 0$.

Here is an example to show differences of numbers (squares in $\boldsymbol{x}$, odd numbers in $\boldsymbol{b}$):

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ 4 \\ 9 \end{bmatrix} = \text{squares} \qquad A\boldsymbol{x} = \begin{bmatrix} 1-0 \\ 4-1 \\ 9-4 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} = \boldsymbol{b}\,. \tag{5}$$

That pattern would continue for a 4 by 4 difference matrix. The next square would be $x_4 = 16$. The next difference would be $x_4 - x_3 = 16 - 9 = 7$ (this is the next odd number). The matrix finds all the differences at once.

Important Note. You may already have learned about multiplying $A\boldsymbol{x}$, a matrix times a vector. Probably it was explained differently, using the rows instead of the columns. The usual way takes the dot product of each row with $\boldsymbol{x}$:

Dot products with rows

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} (1,0,0) \cdot (x_1,x_2,x_3) \\ (-1,1,0) \cdot (x_1,x_2,x_3) \\ (0,-1,1) \cdot (x_1,x_2,x_3) \end{bmatrix}.$$

Those dot products are the same x_1 and $x_2 - x_1$ and $x_3 - x_2$ that we wrote in equation (4). The new way is to work with $A\boldsymbol{x}$ *a column at a time*. Linear combinations are the key to linear algebra, and the output $A\boldsymbol{x}$ is a linear combination of the **columns** of A.

With numbers, you can multiply $A\boldsymbol{x}$ either way (I admit to using rows). With letters, columns are the good way. Chapter 2 will repeat these rules of matrix multiplication, and explain the underlying ideas. There we will multiply matrices both ways.

Linear Equations

One more change in viewpoint is crucial. Up to now, the numbers x_1, x_2, x_3 were known (called c, d, e at first). The right hand side $\boldsymbol{b}$ was not known. We found that vector of differences by multiplying $A\boldsymbol{x}$. **Now we think of $\boldsymbol{b}$ as known and we look for $\boldsymbol{x}$.**

Old question: Compute the linear combination $x_1\boldsymbol{u} + x_2\boldsymbol{v} + x_3\boldsymbol{w}$ to find $\boldsymbol{b}$.
New question: Which combination of $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ produces a particular vector $\boldsymbol{b}$?

This is the inverse problem—to find the input $\boldsymbol{x}$ that gives the desired output $\boldsymbol{b} = A\boldsymbol{x}$. You have seen this before, as a system of linear equations for x_1, x_2, x_3. The right hand sides of the equations are b_1, b_2, b_3. We can solve that system to find x_1, x_2, x_3:

$$A\boldsymbol{x} = \boldsymbol{b} \quad \begin{aligned} x_1 \quad\quad\;\; &= b_1 \\ -x_1 + x_2 \quad\; &= b_2 \\ -x_2 + x_3 &= b_3 \end{aligned} \qquad \textbf{Solution} \quad \begin{aligned} x_1 &= b_1 \\ x_2 &= b_1 + b_2 \\ x_3 &= b_1 + b_2 + b_3. \end{aligned} \tag{6}$$

Let me admit right away—most linear systems are not so easy to solve. In this example, the first equation decided $x_1 = b_1$. Then the second equation produced $x_2 = b_1 + b_2$. *The equations could be solved in order* (top to bottom) *because the matrix A was selected to be* **lower triangular**.

Look at two specific choices 0, 0, 0 and 1, 3, 5 of the right sides b_1, b_2, b_3:

$$\boldsymbol{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ gives } \boldsymbol{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} \text{ gives } \boldsymbol{x} = \begin{bmatrix} 1 \\ 1+3 \\ 1+3+5 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 9 \end{bmatrix}.$$

The first solution (all zeros) is more important than it looks. In words: *If the output is* $\boldsymbol{b} = \mathbf{0}$, *then the input must be* $\boldsymbol{x} = \mathbf{0}$. That statement is true for this matrix A. It is not true for all matrices. Our second example will show (for a different matrix C) how we can have $C\boldsymbol{x} = \mathbf{0}$ when $C \neq 0$ and $\boldsymbol{x} \neq \mathbf{0}$.

This matrix A is "**invertible**". From $\boldsymbol{b}$ we can recover $\boldsymbol{x}$.

The Inverse Matrix

Let me repeat the solution $\boldsymbol{x}$ in equation (6). A sum matrix will appear!

$$A\boldsymbol{x} = \boldsymbol{b} \text{ is solved by } \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_1 + b_2 \\ b_1 + b_2 + b_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}. \tag{7}$$

If the differences of the x's are the b's, the sums of the b's are the x's. That was true for the odd numbers $\boldsymbol{b} = (1,3,5)$ and the squares $\boldsymbol{x} = (1,4,9)$. It is true for all vectors. **The sum matrix S in equation (7) is the inverse of the difference matrix A.**

Example: The differences of $\boldsymbol{x} = (1,2,3)$ are $\boldsymbol{b} = (1,1,1)$. So $\boldsymbol{b} = A\boldsymbol{x}$ and $\boldsymbol{x} = S\boldsymbol{b}$:

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \\ \mathbf{1} \end{bmatrix} \text{ and } S\boldsymbol{b} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \\ \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \end{bmatrix}$$

Equation (7) for the solution vector $\boldsymbol{x} = (x_1, x_2, x_3)$ tells us two important facts:

1. For every $\boldsymbol{b}$ there is one solution to $A\boldsymbol{x} = \boldsymbol{b}$. **2.** A matrix S produces $\boldsymbol{x} = S\boldsymbol{b}$.

The next chapters ask about other equations $A\boldsymbol{x} = \boldsymbol{b}$. Is there a solution? How is it computed? In linear algebra, the notation for the "inverse matrix" is A^{-1}:

$$A\boldsymbol{x} = \boldsymbol{b} \quad \text{is solved by} \quad \boldsymbol{x} = A^{-1}\boldsymbol{b} = S\boldsymbol{b}.$$

Note on calculus. Let me connect these special matrices A and S to calculus. The vector $\boldsymbol{x}$ changes to a function $x(t)$. The differences $A\boldsymbol{x}$ become the *derivative* $dx/dt = b(t)$. In the inverse direction, the sum $S\boldsymbol{b}$ becomes the *integral* of $b(t)$. The Fundamental Theorem of Calculus says that ***integration S is the inverse of differentiation A***.

$$\boldsymbol{A\boldsymbol{x} = \boldsymbol{b} \text{ and } \boldsymbol{x} = S\boldsymbol{b}} \qquad \frac{dx}{dt} = b \text{ and } x(t) = \int_0^t b. \tag{8}$$

The derivative of distance traveled (x) is the velocity (b). The integral of $b(t)$ is the distance $x(t)$. Instead of adding $+C$, I measured the distance from $x(0) = 0$. In the same way, the differences started at $x_0 = 0$. This zero start makes the pattern complete, when we write $x_1 - x_0$ for the first component of Ax (we just wrote x_1).

Notice another analogy with calculus. The differences of squares $0, 1, 4, 9$ are odd numbers $1, 3, 5$. The derivative of $x(t) = t^2$ is $2t$. A perfect analogy would have produced the even numbers $b = 2, 4, 6$ at times $t = 1, 2, 3$. But differences are not the same as derivatives, and our matrix A produces not $2t$ but $2t - 1$ (these one-sided "backward differences" are centered at $t - \frac{1}{2}$):

$$x(t) - x(t-1) = t^2 - (t-1)^2 = t^2 - (t^2 - 2t + 1) = 2t - 1. \tag{9}$$

The Problem Set will follow up to show that "forward differences" produce $2t + 1$. A better choice (not always seen in calculus courses) is a **centered difference** that uses $x(t+1) - x(t-1)$. Divide Δx by the distance Δt from $t - 1$ to $t + 1$, which is 2:

$$\textbf{Centered difference of } x(t) = t^2 \qquad \frac{(t+1)^2 - (t-1)^2}{2} = 2t \quad \text{exactly.} \tag{10}$$

Difference matrices are great. Centered is best. Our second example is *not invertible*.

Cyclic Differences

This example keeps the same columns $\boldsymbol{u}$ and $\boldsymbol{v}$ but changes $\boldsymbol{w}$ to a new vector $\boldsymbol{w}^*$:

$$\textbf{Second example} \qquad \boldsymbol{u} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \boldsymbol{v} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \quad \boldsymbol{w}^* = \begin{bmatrix} \mathbf{-1} \\ 0 \\ 1 \end{bmatrix}.$$

Now the linear combinations of $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}^*$ lead to a **cyclic difference matrix** C:

$$\textbf{Cyclic} \qquad C\boldsymbol{x} = \begin{bmatrix} 1 & 0 & \mathbf{-1} \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 - x_3 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \boldsymbol{b}. \tag{11}$$

This matrix C is not triangular. It is not so simple to solve for $\boldsymbol{x}$ when we are given $\boldsymbol{b}$. Actually it is impossible to find *the* solution to $C\boldsymbol{x} = \boldsymbol{b}$, because the three equations either have **infinitely many solutions** or else **no solution**:

$$\begin{matrix} C\boldsymbol{x} = \mathbf{0} \\ \textbf{Infinitely} \\ \textbf{many } \boldsymbol{x} \end{matrix} \qquad \begin{bmatrix} x_1 - x_3 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ is solved by all vectors } \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} c \\ c \\ c \end{bmatrix}. \tag{12}$$

Every constant vector (c, c, c) has zero differences when we go cyclically. This undetermined constant c is like the $+C$ that we add to integrals. The cyclic differences have $x_1 - x_3$ in the first component, instead of starting from $x_0 = 0$.

The other very likely possibility for $Cx = b$ is **no solution** at all:

$$\boldsymbol{Cx = b} \qquad \begin{bmatrix} x_1 - x_3 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} \qquad \begin{array}{l} \text{Left sides add to 0} \\ \text{Right sides add to 9} \\ \textit{No solution } x_1, x_2, x_3 \end{array} \tag{13}$$

Look at this example geometrically. No combination of $\boldsymbol{u}, \boldsymbol{v}$, and $\boldsymbol{w}^*$ will produce the vector $\boldsymbol{b} = (1, 3, 5)$. The combinations don't fill the whole three-dimensional space. The right sides must have $b_1 + b_2 + b_3 = 0$ to allow a solution to $Cx = b$, because the left sides $x_1 - x_3$, $x_2 - x_1$, and $x_3 - x_2$ always add to zero.

Put that in different words. **All linear combinations** $x_1\boldsymbol{u} + x_2\boldsymbol{v} + x_3\boldsymbol{w}^* = \boldsymbol{b}$ **lie on the plane given by** $b_1 + b_2 + b_3 = 0$. This subject is suddenly connecting algebra with geometry. Linear combinations can fill all of space, or only a plane. We need a picture to show the crucial difference between $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ (the first example) and $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}^*$.

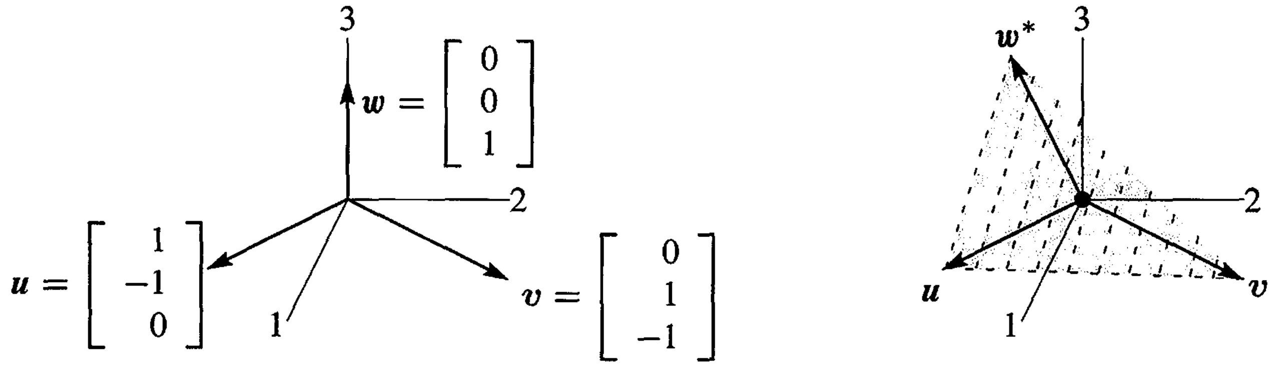


Figure 1.10: Independent vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$. Dependent vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}^*$ in a plane.

Independence and Dependence

Figure 1.10 shows those column vectors, first of the matrix A and then of C. The first two columns $\boldsymbol{u}$ and $\boldsymbol{v}$ are the same in both pictures. If we only look at the combinations of those two vectors, we will get a two-dimensional plane. **The key question is whether the third vector is in that plane:**

Independence $\boldsymbol{w}$ is not in the plane of $\boldsymbol{u}$ and $\boldsymbol{v}$.
Dependence $\boldsymbol{w}^*$ *is* in the plane of $\boldsymbol{u}$ and $\boldsymbol{v}$.

The important point is that the new vector $\boldsymbol{w}^*$ is a linear combination of $\boldsymbol{u}$ and $\boldsymbol{v}$:

$$\boldsymbol{u} + \boldsymbol{v} + \boldsymbol{w}^* = 0 \qquad \boldsymbol{w}^* = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = -\boldsymbol{u} - \boldsymbol{v}. \tag{14}$$

All three vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}^*$ have components adding to zero. Then all their combinations will have $b_1 + b_2 + b_3 = 0$ (as we saw above, by adding the three equations). This is the equation for the plane containing all combinations of $\boldsymbol{u}$ and $\boldsymbol{v}$. By including $\boldsymbol{w}^*$ we get *no new vectors* because $\boldsymbol{w}^*$ is already on that plane.

The original $\boldsymbol{w} = (0, 0, 1)$ is not on the plane: $0 + 0 + 1 \neq 0$. The combinations of $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ fill the whole three-dimensional space. We know this already, because the solution $\boldsymbol{x} = S\boldsymbol{b}$ in equation (6) gave the right combination to produce any $\boldsymbol{b}$.

The two matrices A and C, with third columns $\boldsymbol{w}$ and $\boldsymbol{w}^*$, allowed me to mention two key words of linear algebra: independence and dependence. The first half of the course will develop these ideas much further—I am happy if you see them early in the two examples:

$\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ are **independent**. No combination except $0\boldsymbol{u} + 0\boldsymbol{v} + 0\boldsymbol{w} = \boldsymbol{0}$ gives $\boldsymbol{b} = \boldsymbol{0}$.

$\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}^*$ are **dependent**. Other combinations (specifically $\boldsymbol{u} + \boldsymbol{v} + \boldsymbol{w}^*$) give $\boldsymbol{b} = \boldsymbol{0}$.

You can picture this in three dimensions. The three vectors lie in a plane or they don't. Chapter 2 has n vectors in n-dimensional space. *Independence or dependence* is the key point. The vectors go into the columns of an n by n matrix:

Independent columns: $A\boldsymbol{x} = \boldsymbol{0}$ has one solution. A is an **invertible matrix.**

Dependent columns: $A\boldsymbol{x} = \boldsymbol{0}$ has many solutions. A is a **singular matrix.**

Eventually we will have n vectors in m-dimensional space. The matrix A with those n columns is now *rectangular* (m by n). Understanding $A\boldsymbol{x} = \boldsymbol{b}$ is the problem of Chapter 3.

■ REVIEW OF THE KEY IDEAS ■

1. **Matrix times vector**: $A\boldsymbol{x} =$ combination of the columns of A.
2. The solution to $A\boldsymbol{x} = \boldsymbol{b}$ is $\boldsymbol{x} = A^{-1}\boldsymbol{b}$, when A is an invertible matrix.
3. The difference matrix A is inverted by the sum matrix $S = A^{-1}$.
4. The cyclic matrix C has no inverse. Its three columns lie in the same plane. Those dependent columns add to the zero vector. $C\boldsymbol{x} = \boldsymbol{0}$ has many solutions.
5. This section is looking ahead to key ideas, not fully explained yet.

■ WORKED EXAMPLES ■

1.3 A Change the southwest entry a_{31} of A (row 3, column 1) to $a_{31} = \mathbf{1}$:

$$A\boldsymbol{x} = \boldsymbol{b} \qquad \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ \mathbf{1} & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ -x_1 + x_2 \\ \boldsymbol{x_1} - x_2 + x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

Find the solution $\boldsymbol{x}$ for any $\boldsymbol{b}$. From $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ read off the inverse matrix A^{-1}.

Solution Solve the (linear triangular) system $Ax = b$ from top to bottom:

$$\begin{aligned} &\text{first} & x_1 &= b_1 \\ &\text{then} & x_2 &= b_1 + b_2 \\ &\text{then} & x_3 &= \quad\;\; b_2 + b_3 \end{aligned} \qquad \text{This says that } x = A^{-1}b = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

This is good practice to see the columns of the inverse matrix multiplying b_1, b_2, and b_3. The first column of A^{-1} is the solution for $b = (1,0,0)$. The second column is the solution for $b = (0,1,0)$. The third column x of A^{-1} is the solution for $Ax = b = (0,0,1)$.

The three columns of A are still independent. They don't lie in a plane. The combinations of those three columns, using the right weights x_1, x_2, x_3, can produce any three-dimensional vector $b = (b_1, b_2, b_3)$. Those weights come from $x = A^{-1}b$.

1.3 B This E is an **elimination matrix.** E has a subtraction, E^{-1} has an addition.

$$Ex = b \qquad \begin{bmatrix} 1 & 0 \\ -\ell & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \qquad\qquad E = \begin{bmatrix} 1 & 0 \\ -\ell & 1 \end{bmatrix}$$

The first equation is $x_1 = b_1$. The second equation is $x_2 - \ell x_1 = b_2$. The inverse will *add* $\ell x_1 = \ell b_1$, because the elimination matrix *subtracted* ℓx_1:

$$x = E^{-1}b \qquad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ \ell b_1 + b_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \ell & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \qquad E^{-1} = \begin{bmatrix} 1 & 0 \\ \ell & 1 \end{bmatrix}$$

1.3 C Change C from a cyclic difference to a **centered difference** producing $x_3 - x_1$:

$$Cx = b \qquad \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_2 - 0 \\ x_3 - x_1 \\ 0 - x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}. \tag{15}$$

Show that $Cx = b$ can only be solved when $b_1 + b_3 = 0$. That is a plane of vectors b in three-dimensional space. Each column of C is in the plane, the matrix has no inverse. So this plane contains all combinations of those columns (which are all the vectors Cx).

Solution The first component of $b = Cx$ is x_2, and the last component of b is $-x_2$. So we always have $b_1 + b_3 = 0$, for every choice of x.

If you draw the column vectors in C, the first and third columns fall on the same line. In fact (column 1) $= -$(column 3). So the three columns will lie in a plane, and C is *not* an invertible matrix. We cannot solve $Cx = b$ unless $b_1 + b_3 = 0$.

I included the zeros so you could see that this matrix produces "centered differences". Row i of Cx is x_{i+1} (*right of center*) minus x_{i-1} (*left of center*). Here is the 4 by 4 centered difference matrix:

$$Cx = b \qquad \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_2 - 0 \\ x_3 - x_1 \\ x_4 - x_2 \\ 0 - x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \tag{16}$$

Surprisingly this matrix is now invertible! The first and last rows give x_2 and x_3. Then the middle rows give x_1 and x_4. It is possible to write down the inverse matrix C^{-1}. But 5 by 5 will be singular (*not invertible*) again ...

Problem Set 1.3

1 Find the linear combination $2\boldsymbol{s}_1 + 3\boldsymbol{s}_2 + 4\boldsymbol{s}_3 = \boldsymbol{b}$. Then write $\boldsymbol{b}$ as a matrix-vector multiplication $S\boldsymbol{x}$. Compute the dot products (row of S) $\cdot\, \boldsymbol{x}$:

$$\boldsymbol{s}_1 = \begin{bmatrix}1\\1\\1\end{bmatrix} \quad \boldsymbol{s}_2 = \begin{bmatrix}0\\1\\1\end{bmatrix} \quad \boldsymbol{s}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix} \text{ go into the columns of } S.$$

2 Solve these equations $S\boldsymbol{y} = \boldsymbol{b}$ with $\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3$ in the columns of S:

$$\begin{bmatrix}1&0&0\\1&1&0\\1&1&1\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix} = \begin{bmatrix}1\\1\\1\end{bmatrix} \text{ and } \begin{bmatrix}1&0&0\\1&1&0\\1&1&1\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix} = \begin{bmatrix}1\\4\\9\end{bmatrix}.$$

The sum of the first n odd numbers is _____.

3 Solve these three equations for y_1, y_2, y_3 in terms of B_1, B_2, B_3:

$$S\boldsymbol{y} = \boldsymbol{B} \qquad \begin{bmatrix}1&0&0\\1&1&0\\1&1&1\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix} = \begin{bmatrix}B_1\\B_2\\B_3\end{bmatrix}.$$

Write the solution $\boldsymbol{y}$ as a matrix $A = S^{-1}$ times the vector $\boldsymbol{B}$. Are the columns of S independent or dependent?

4 Find a combination $x_1\boldsymbol{w}_1 + x_2\boldsymbol{w}_2 + x_3\boldsymbol{w}_3$ that gives the zero vector:

$$\boldsymbol{w}_1 = \begin{bmatrix}1\\2\\3\end{bmatrix} \quad \boldsymbol{w}_2 = \begin{bmatrix}4\\5\\6\end{bmatrix} \quad \boldsymbol{w}_3 = \begin{bmatrix}7\\8\\9\end{bmatrix}.$$

Those vectors are (independent) (dependent). The three vectors lie in a _____. The matrix W with those columns is *not invertible.*

5 The rows of that matrix W produce three vectors (I write them as columns):

$$\boldsymbol{r}_1 = \begin{bmatrix}1\\4\\7\end{bmatrix} \quad \boldsymbol{r}_2 = \begin{bmatrix}2\\5\\8\end{bmatrix} \quad \boldsymbol{r}_3 = \begin{bmatrix}3\\6\\9\end{bmatrix}.$$

Linear algebra says that these vectors must also lie in a plane. There must be many combinations with $y_1\boldsymbol{r}_1 + y_2\boldsymbol{r}_2 + y_3\boldsymbol{r}_3 = \boldsymbol{0}$. Find two sets of y's.

6 Which values of c give dependent columns (combination equals zero)?

$$\begin{bmatrix}1&3&5\\1&2&4\\1&1&c\end{bmatrix} \quad \begin{bmatrix}1&0&c\\1&1&0\\0&1&1\end{bmatrix} \quad \begin{bmatrix}c&c&c\\2&1&5\\3&3&6\end{bmatrix}$$

7 If the columns combine into $A\boldsymbol{x} = \mathbf{0}$ then each row has $\boldsymbol{r} \cdot \boldsymbol{x} = 0$:

$$\begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad \textbf{By rows} \begin{bmatrix} \boldsymbol{r}_1 \cdot \boldsymbol{x} \\ \boldsymbol{r}_2 \cdot \boldsymbol{x} \\ \boldsymbol{r}_3 \cdot \boldsymbol{x} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

The three rows also lie in a plane. Why is that plane perpendicular to $\boldsymbol{x}$?

8 Moving to a 4 by 4 difference equation $A\boldsymbol{x} = \boldsymbol{b}$, find the four components x_1, x_2, x_3, x_4. Then write this solution as $\boldsymbol{x} = S\boldsymbol{b}$ to find the inverse matrix $S = A^{-1}$:

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \boldsymbol{b}.$$

9 What is the *cyclic* 4 by 4 difference matrix C? It will have 1 and -1 in each row. Find all solutions $\boldsymbol{x} = (x_1, x_2, x_3, x_4)$ to $C\boldsymbol{x} = \mathbf{0}$. The four columns of C lie in a "three-dimensional hyperplane" inside four-dimensional space.

10 A *forward* difference matrix Δ is *upper* triangular:

$$\Delta \boldsymbol{z} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} z_2 - z_1 \\ z_3 - z_2 \\ 0 - z_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \boldsymbol{b}.$$

Find z_1, z_2, z_3 from b_1, b_2, b_3. What is the inverse matrix in $\boldsymbol{z} = \Delta^{-1}\boldsymbol{b}$?

11 Show that the forward differences $(t+1)^2 - t^2$ are $2t+1 =$ *odd numbers*. As in calculus, the difference $(t+1)^n - t^n$ will begin with the derivative of t^n, which is _____.

12 The last lines of the Worked Example say that the 4 by 4 centered difference matrix in (16) *is* invertible. Solve $C\boldsymbol{x} = (b_1, b_2, b_3, b_4)$ to find its inverse in $\boldsymbol{x} = C^{-1}\boldsymbol{b}$.

Challenge Problems

13 The very last words say that the 5 by 5 centered difference matrix *is not* invertible. Write down the 5 equations $C\boldsymbol{x} = \boldsymbol{b}$. Find a combination of left sides that gives zero. What combination of b_1, b_2, b_3, b_4, b_5 must be zero? (The 5 columns lie on a "4-dimensional hyperplane" in 5-dimensional space.)

14 If (a, b) is a multiple of (c, d) with $abcd \neq 0$, *show that* (a, c) *is a multiple of* (b, d). This is surprisingly important; two columns are falling on one line. You could use numbers first to see how a, b, c, d are related. The question will lead to:

The matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ has dependent columns when it has dependent rows.

Chapter 2

Solving Linear Equations

2.1 Vectors and Linear Equations

The central problem of linear algebra is to solve a system of equations. Those equations are linear, which means that the unknowns are only multiplied by numbers—we never see x times y. Our first linear system is certainly not big. But you will see how far it leads:

$$\begin{array}{ll} \textbf{Two equations} & x - 2y = 1 \\ \textbf{Two unknowns} & 3x + 2y = 11 \end{array} \qquad (1)$$

We begin *a row at a time*. The first equation $\boldsymbol{x - 2y = 1}$ produces a straight line in the xy plane. The point $x = 1, y = 0$ is on the line because it solves that equation. The point $x = 3, y = 1$ is also on the line because $3 - 2 = 1$. If we choose $x = 101$ we find $y = 50$.

The slope of this particular line is $\frac{1}{2}$, because y increases by 1 when x changes by 2. But slopes are important in calculus and this is linear algebra!

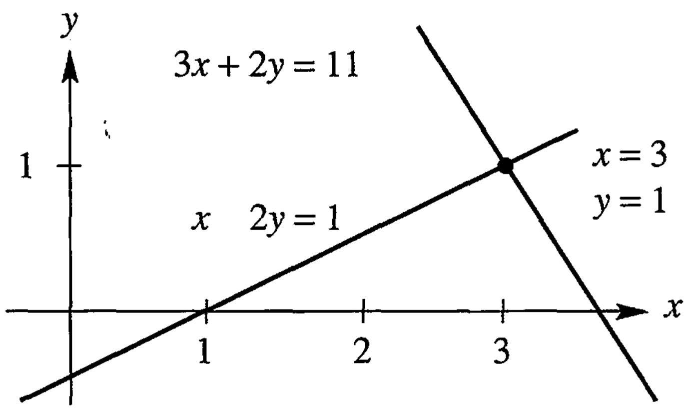


Figure 2.1: *Row picture*: The point (3, 1) where the lines meet is the solution.

Figure 2.1 shows that line $x - 2y = 1$. The second line in this "row picture" comes from the second equation $3x + 2y = 11$. You can't miss the intersection point where the two lines meet. *The point* $x = 3, y = 1$ *lies on both lines.* That point solves both equations at once. This is the solution to our system of linear equations.

ROWS ***The row picture shows two lines meeting at a single point (the solution).***

Turn now to the column picture. I want to recognize the same linear system as a "vector equation". Instead of numbers we need to see *vectors*. If you separate the original system into its columns instead of its rows, you get a vector equation:

$$\textbf{Combination equals } \boldsymbol{b} \qquad x\begin{bmatrix} 1 \\ 3 \end{bmatrix} + y\begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix} = \boldsymbol{b}. \tag{2}$$

This has two column vectors on the left side. The problem is *to find the combination of those vectors that equals the vector on the right.* We are multiplying the first column by x and the second column by y, and adding. With the right choices $x = 3$ and $y = 1$ (the same numbers as before), this produces $3(\textit{\textbf{column 1}}) + 1(\textit{\textbf{column 2}}) = \boldsymbol{b}$.

COLUMNS ***The column picture combines the column vectors on the left side to produce the vector b on the right side.***

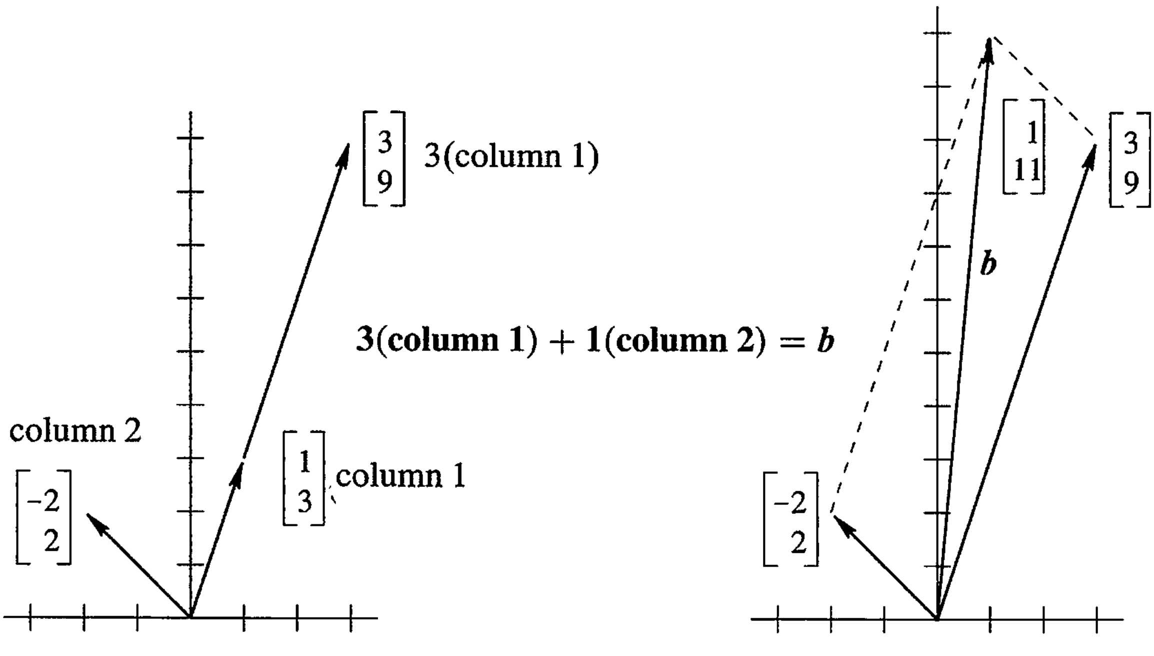


Figure 2.2: *Column picture*: A combination of columns produces the right side (1,11).

Figure 2.2 is the "column picture" of two equations in two unknowns. The first part shows the two separate columns, and that first column multiplied by 3. This multiplication by a *scalar* (a number) is one of the two basic operations in linear algebra:

$$\textbf{Scalar multiplication} \qquad 3\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \end{bmatrix}.$$

If the components of a vector $\boldsymbol{v}$ are v_1 and v_2, then $c\boldsymbol{v}$ has components cv_1 and cv_2.

The other basic operation is *vector addition*. We add the first components and the second components separately. The vector sum is $(1, 11)$ as desired:

$$\textbf{Vector addition} \qquad \begin{bmatrix} 3 \\ 9 \end{bmatrix} + \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}.$$

The right side of Figure 2.2 shows this addition. The sum along the diagonal is the vector $\boldsymbol{b} = (1, 11)$ on the right side of the linear equations.

To repeat: The left side of the vector equation is a ***linear combination*** of the columns. The problem is to find the right coefficients $x = 3$ and $y = 1$. We are combining scalar multiplication and vector addition into one step. That step is crucially important, because it contains both of the basic operations:

$$\textbf{Linear combination} \qquad 3\begin{bmatrix} 1 \\ 3 \end{bmatrix} + \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}.$$

Of course the solution $x = 3, y = 1$ is the same as in the row picture. I don't know which picture you prefer! I suspect that the two intersecting lines are more familiar at first. You may like the row picture better, but only for one day. My own preference is to combine column vectors. It is a lot easier to see a combination of four vectors in four-dimensional space, than to visualize how four hyperplanes might possibly meet at a point. (*Even one hyperplane is hard enough...*)

The ***coefficient matrix*** on the left side of the equations is the 2 by 2 matrix A:

$$\textbf{Coefficient matrix} \qquad A = \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix}.$$

This is very typical of linear algebra, to look at a matrix by rows and by columns. Its rows give the row picture and its columns give the column picture. Same numbers, different pictures, same equations. We write those equations as a matrix problem $A\boldsymbol{x} = \boldsymbol{b}$:

$$\textbf{Matrix equation} \qquad \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}.$$

The row picture deals with the two rows of A. The column picture combines the columns. The numbers $x = 3$ and $y = 1$ go into $\boldsymbol{x}$. Here is matrix-vector multiplication:

$$\begin{array}{l}\textbf{Dot products with rows} \\ \textbf{Combination of columns}\end{array} \qquad A\boldsymbol{x} = \boldsymbol{b} \quad \text{is} \quad \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}.$$

Looking ahead This chapter is going to solve n equations in n unknowns (for any n). I am not going at top speed, because smaller systems allow examples and pictures and a complete understanding. You are free to go faster, as long as **matrix multiplication and inversion** become clear. Those two ideas will be the keys to invertible matrices.

I can list four steps to understanding elimination using matrices.

1. Elimination goes from A to a triangular U by a sequence of matrix steps E_{ij}.

2. The inverse matrices E_{ij}^{-1} in reverse order bring U back to the original A.

3. In matrix language that reverse order is $A = LU =$ (lower triangle) (upper triangle).

4. Elimination succeeds if A is invertible. (It may need row exchanges.).

The most-used algorithm in computational science takes those steps (MATLAB calls it **lu**). But linear algebra goes beyond square invertible matrices! For m by n matrices, $A\boldsymbol{x} = \boldsymbol{0}$ may have many solutions. Those solutions will go into a **vector space**. The **rank** of A leads to the **dimension** of that vector space.

All this comes in Chapter 3, and I don't want to hurry. But I must get there.

Three Equations in Three Unknowns

The three unknowns are x, y, z. We have three linear equations:

$$A\boldsymbol{x} = \boldsymbol{b} \qquad \begin{array}{rcrcrcr} x & + & 2y & + & 3z & = & 6 \\ 2x & + & 5y & + & 2z & = & 4 \\ 6x & - & 3y & + & z & = & 2 \end{array} \qquad (3)$$

We look for numbers x, y, z that solve all three equations at once. Those desired numbers might or might not exist. For this system, they do exist. When the number of unknowns matches the number of equations, there is *usually* one solution. Before solving the problem, we visualize it both ways:

ROW ***The row picture shows three planes meeting at a single point.***

COLUMN ***The column picture combines three columns to produce*** $(6, 4, 2)$.

In the row picture, each equation produces a *plane* in three-dimensional space. The first plane in Figure 2.3 comes from the first equation $x + 2y + 3z = 6$. That plane crosses the x and y and z axes at the points $(6, 0, 0)$ and $(0, 3, 0)$ and $(0, 0, 2)$. Those three points solve the equation and they determine the whole plane.

The vector $(x, y, z) = (0, 0, 0)$ does not solve $x + 2y + 3z = 6$. Therefore that plane does not contain the origin. The plane $x + 2y + 3z = 0$ does pass through the origin, and it is parallel to $x + 2y + 3z = 6$. When the right side increases to 6, the parallel plane moves away from the origin.

The second plane is given by the second equation $2x + 5y + 2z = 4$. *It intersects the first plane in a line* $\boldsymbol{L}$. The usual result of two equations in three unknowns is a line $\boldsymbol{L}$ of solutions. (Not if the equations were $x + 2y + 3z = 6$ and $x + 2y + 3z = 0$.)

The third equation gives a third plane. It cuts the line $\boldsymbol{L}$ at a single point. That point lies on all three planes and it solves all three equations. It is harder to draw this triple intersection point than to imagine it. The three planes meet at the solution (which we haven't found yet). **The column form will now show immediately why** $z = 2$.

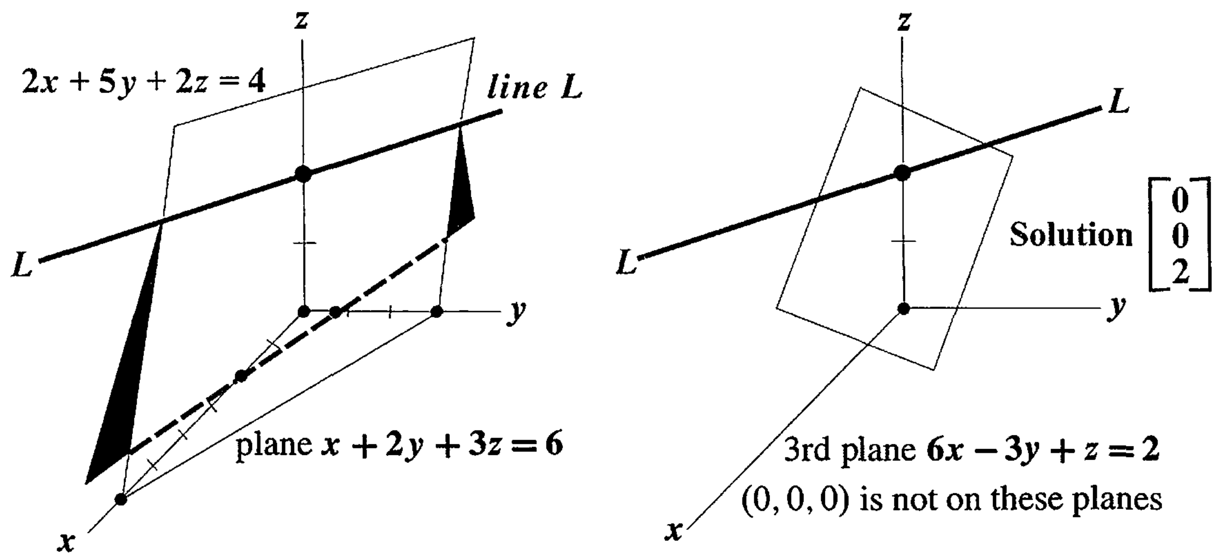


Figure 2.3: *Row picture*: Two planes meet at a line, three planes at a point.

The column picture starts with the vector form of the equations $Ax = b$:

$$\textbf{Combine columns} \qquad x \begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix} + y \begin{bmatrix} 2 \\ 5 \\ -3 \end{bmatrix} + z \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix}. \tag{4}$$

The unknowns are the coefficients x, y, z. We want to multiply the three column vectors by the correct numbers x, y, z to produce $b = (6, 4, 2)$.

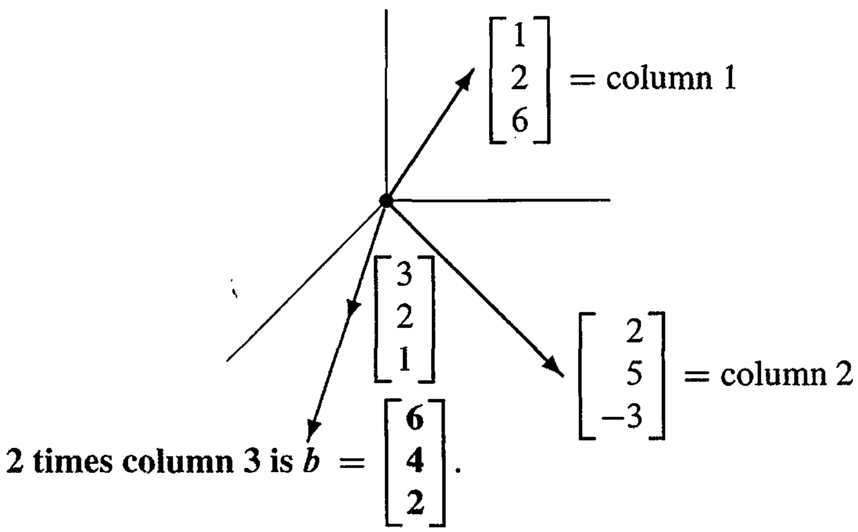


Figure 2.4: *Column picture*: $(x, y, z) = (0, 0, 2)$ because $2(3, 2, 1) = (6, 4, 2) = b$.

Figure 2.4 shows this column picture. Linear combinations of those columns can produce any vector b! The combination that produces $b = (6, 4, 2)$ is just 2 times the third column. *The coefficients we need are* $x = 0$, $y = 0$, *and* $z = 2$.

The three planes in the row picture meet at that same solution point $(0,0,2)$:

Correct combination
$(x, y, z) = (\mathbf{0}, \mathbf{0}, \mathbf{2})$

$$\mathbf{0}\begin{bmatrix}1\\2\\6\end{bmatrix} + \mathbf{0}\begin{bmatrix}2\\5\\-3\end{bmatrix} + \mathbf{2}\begin{bmatrix}3\\2\\1\end{bmatrix} = \begin{bmatrix}6\\4\\2\end{bmatrix}.$$

The Matrix Form of the Equations

We have three rows in the row picture and three columns in the column picture (plus the right side). The three rows and three columns contain nine numbers. *These nine numbers fill a* 3 *by* 3 *matrix* A:

$$\textbf{\textit{The "coefficient matrix" in } } Ax = b \textbf{\textit{ is}} \qquad A = \begin{bmatrix}1 & 2 & 3\\2 & 5 & 2\\6 & -3 & 1\end{bmatrix}.$$

The capital letter A stands for all nine coefficients (in this square array). The letter $\boldsymbol{b}$ denotes the column vector with components $6, 4, 2$. The unknown $\boldsymbol{x}$ is also a column vector, with components x, y, z. (We use boldface because it is a vector, $\boldsymbol{x}$ because it is unknown.) By rows the equations were (3), by columns they were (4), and by matrices they are (5):

$$\textbf{\textit{Matrix equation }} A\boldsymbol{x} = \boldsymbol{b} \qquad \begin{bmatrix}1 & 2 & 3\\2 & 5 & 2\\6 & -3 & 1\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}6\\4\\2\end{bmatrix}. \tag{5}$$

Basic question: What does it mean to "multiply A times $\boldsymbol{x}$"? We can multiply by rows or by columns. Either way, $A\boldsymbol{x} = \boldsymbol{b}$ must be a correct representation of the three equations. You do the same nine multiplications either way.

Multiplication by rows $\quad A\boldsymbol{x}$ comes from *dot products*, each row times the column $\boldsymbol{x}$:

$$A\boldsymbol{x} = \begin{bmatrix}(\textit{row 1}) \cdot \boldsymbol{x}\\(\textit{row 2}) \cdot \boldsymbol{x}\\(\textit{row 3}) \cdot \boldsymbol{x}\end{bmatrix}. \tag{6}$$

Multiplication by columns $\quad A\boldsymbol{x}$ is a ***combination of column vectors***:

$$A\boldsymbol{x} = x\,(\textit{column 1}) + y\,(\textit{column 2}) + z\,(\textit{column 3}). \tag{7}$$

When we substitute the solution $\boldsymbol{x} = (0, 0, 2)$, the multiplication $A\boldsymbol{x}$ produces $\boldsymbol{b}$:

$$\begin{bmatrix}1 & 2 & 3\\2 & 5 & 2\\6 & -3 & 1\end{bmatrix}\begin{bmatrix}0\\0\\2\end{bmatrix} = \textbf{\textit{2 times column 3}} = \begin{bmatrix}6\\4\\2\end{bmatrix}.$$

The dot product from the first row is $(1, 2, 3) \cdot (0, 0, 2) = 6$. The other rows give dot products 4 and 2. ***This book sees*** $A\boldsymbol{x}$ ***as a combination of the columns of*** A.

Example 1 Here are 3 by 3 matrices A and I = identity, with three 1's and six 0's:

$$Ax = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix} \qquad Ix = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

If you are a row person, the dot product of $(1, 0, 0)$ with $(4, 5, 6)$ is 4. If you are a column person, the linear combination Ax is 4 times the first column $(1, 1, 1)$. In that matrix A, the second and third columns are zero vectors.

The other matrix I is special. It has ones on the "main diagonal". *Whatever vector this matrix multiplies, that vector is not changed.* This is like multiplication by 1, but for matrices and vectors. The exceptional matrix in this example is the 3 by 3 ***identity matrix***:

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{always yields the multiplication} \quad Ix = x\,.$$

Matrix Notation

The first row of a 2 by 2 matrix contains a_{11} and a_{12}. The second row contains a_{21} and a_{22}. The first index gives the row number, so that a_{ij} is an entry in row i. The second index j gives the column number. But those subscripts are not very convenient on a keyboard! Instead of a_{ij} we type $A(i, j)$. ***The entry*** $a_{57} = A(5, 7)$ ***would be in row 5, column 7.***

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} A(1,1) & A(1,2) \\ A(2,1) & A(2,2) \end{bmatrix}.$$

For an m by n matrix, the row index i goes from 1 to m. The column index j stops at n. There are mn entries $a_{ij} = A(i, j)$. A square matrix of order n has n^2 entries.

Multiplication in MATLAB

I want to express A and x and their product Ax using MATLAB commands. This is a first step in learning that language. I begin by defining the matrix A and the vector x. This vector is a 3 by 1 matrix, with three rows and one column. Enter matrices a row at a time, and use a semicolon to signal the end of a row:

$$\begin{aligned} A &= [1 \;\; 2 \;\; 3; \;\;\; 2 \;\; 5 \;\; 2; \;\;\; 6 \; -3 \;\; 1] \\ x &= [0\,;\, 0\,;\, 2] \end{aligned}$$

Here are three ways to multiply Ax in MATLAB. In reality, $A * x$ is the good way to do it. MATLAB is a high level language, and it works with matrices:

$$\textit{Matrix multiplication} \quad b = A * x$$

We can also pick out the first row of A (as a smaller matrix!). The notation for that 1 by 3 submatrix is $A(1,:)$. **Here the colon symbol keeps all columns of row 1**:

$$\textit{Row at a time} \quad \boldsymbol{b} = [\,A(1,:) * \boldsymbol{x}\ ;\ A(2,:) * \boldsymbol{x}\ ;\ A(3,:) * \boldsymbol{x}\,]$$

Each entry is a dot product, row times column, 1 by 3 matrix times 3 by 1 matrix.

The other way to multiply uses the columns of A. The first column is the 3 by 1 submatrix $A(:,1)$. Now the colon symbol : is keeping all rows of column 1. This column multiplies $x(1)$ and the other columns multiply $x(2)$ and $x(3)$:

$$\textit{Column at a time} \quad \boldsymbol{b} = A(:,1) * x(1) + A(:,2) * x(2) + A(:,3) * x(3)$$

I think that matrices are stored by columns. Then multiplying a column at a time will be a little faster. So $A * \boldsymbol{x}$ is actually executed by columns.

You can see the same choice in a FORTRAN-type structure, which operates on single entries of A and $\boldsymbol{x}$. This lower level language needs an outer and inner "DO loop". When the outer loop uses the row number I, multiplication is a row at a time. The inner loop $J = 1, 3$ goes along each row I.

When the outer loop uses J, multiplication is a column at a time. I will do that in MATLAB (which really needs two more lines "end" and "end" to close "for i" and "for j").

FORTRAN by rows	MATLAB **by columns**
DO 10 $\;I = 1,3$	for $j = 1:3$
DO 10 $\;J = 1,3$	for $i = 1:3$
10 $\;B(I) = B(I) + A(I,J) * X(J)$	$b(i) = b(i) + A(i,j) * x(j)$

Notice that MATLAB is sensitive to upper case versus lower case (capital letters and small letters). If the matrix is A then its entries are not $a(i,j)$: not recognized.

I think you will prefer the higher level $A * \boldsymbol{x}$. FORTRAN won't appear again in this book. *Maple* and *Mathematica* and graphing calculators also operate at the higher level. Multiplication is $A.\ x$ in *Mathematica*. It is **multiply**(A,x); or equally **evalm**$(A\&*\,x)$; in *Maple*. Those languages allow symbolic entries $a, b, x, \ldots$ and not only real numbers. Like MATLAB's Symbolic Toolbox, they give the symbolic answer.

■ REVIEW OF THE KEY IDEAS ■

1. The basic operations on vectors are multiplication $c\boldsymbol{v}$ and vector addition $\boldsymbol{v} + \boldsymbol{w}$.
2. Together those operations give linear combinations $c\boldsymbol{v} + d\boldsymbol{w}$.
3. Matrix-vector multiplication $A\boldsymbol{x}$ can be computed by dot products, a row at a time. But $A\boldsymbol{x}$ should be understood as a combination of the columns of A.
4. Column picture: $A\boldsymbol{x} = \boldsymbol{b}$ asks for a combination of columns to produce $\boldsymbol{b}$.
5. Row picture: Each equation in $A\boldsymbol{x} = \boldsymbol{b}$ gives a line ($n = 2$) or a plane ($n = 3$) or a "hyperplane" ($n > 3$). They intersect at the solution or solutions, if any.

■ WORKED EXAMPLES ■

2.1 A Describe the column picture of these three equations $Ax = b$. Solve by careful inspection of the columns (instead of elimination):

$$\begin{aligned} x + 3y + 2z &= -3 \\ 2x + 2y + 2z &= -2 \\ 3x + 5y + 6z &= -5 \end{aligned} \qquad \text{which is} \qquad \begin{bmatrix} 1 & 3 & 2 \\ 2 & 2 & 2 \\ 3 & 5 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -3 \\ -2 \\ -5 \end{bmatrix}.$$

Solution The column picture asks for a linear combination that produces b from the three columns of A. In this example b is *minus the second column.* So the solution is $x = 0, y = -1, z = 0$. To show that $(0, -1, 0)$ is the *only* solution we have to know that "A is invertible" and "the columns are independent" and "the determinant isn't zero."

Those words are not yet defined but the test comes from elimination: We need (and for this matrix we find) a full set of three nonzero pivots.

Suppose the right side changes to $b = (4, 4, 8) =$ sum of the first two columns. Then the good combination has $x = 1, y = 1, z = 0$. The solution becomes $x = (1, 1, 0)$.

2.1 B This system has *no solution.* The planes in the row picture don't meet at a point. ***No combination of the three columns produces b. How to show this?***

$$\begin{aligned} x + 3y + 5z &= 4 \\ x + 2y - 3z &= 5 \\ 2x + 5y + 2z &= 8 \end{aligned} \qquad \begin{bmatrix} 1 & 3 & 5 \\ 1 & 2 & -3 \\ 2 & 5 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 8 \end{bmatrix} = b$$

(1) Multiply the equations by $1, 1, -1$ and add to get $0 = 1$. *No solution.* Are any two of the planes parallel? What are the equations of planes parallel to $x + 3y + 5z = 4$?

(2) Take the dot product of each column of A (and also b) with $y = (1, 1, -1)$. How do those dot products show that the system $Ax = b$ has no solution?

(3) Find three right side vectors b^* and b^{**} and b^{***} that *do* allow solutions.

Solution

(1) Multiplying the equations by $1, 1, -1$ and adding gives $0 = 1$:

$$\begin{aligned} x + 3y + 5z &= 4 \\ x + 2y - 3z &= 5 \\ -[2x + 5y + 2z &= 8] \\ \hline 0x + 0y + 0z &= 1 \qquad \textit{No Solution} \end{aligned}$$

The planes don't meet at a point, even though no two planes are parallel. For a plane parallel to $x + 3y + 5z = 4$, change the "4". The parallel plane $x + 3y + 5z = 0$ goes through the origin $(0, 0, 0)$. And the equation multiplied by any nonzero constant still gives the same plane, as in $2x + 6y + 10z = 8$.

(2) The dot product of each column of A with $\boldsymbol{y} = (1, 1, -1)$ is *zero*. On the right side, $\boldsymbol{y} \cdot \boldsymbol{b} = (1, 1, -1) \cdot (4, 5, 8) = 1$ is *not zero*. So a solution is impossible.

(3) There is a solution when $\boldsymbol{b}$ is a combination of the columns. These three choices of $\boldsymbol{b}$ have solutions $\boldsymbol{x}^* = (1, 0, 0)$ and $\boldsymbol{x}^{**} = (1, 1, 1)$ and $\boldsymbol{x}^{***} = (0, 0, 0)$:

$$\boldsymbol{b}^* = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \text{first column} \quad \boldsymbol{b}^{**} = \begin{bmatrix} 9 \\ 0 \\ 9 \end{bmatrix} = \text{sum of columns} \quad \boldsymbol{b}^{***} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Problem Set 2.1

Problems 1–8 are about the row and column pictures of $A\boldsymbol{x} = \boldsymbol{b}$.

1 With $A = I$ (the identity matrix) draw the planes in the row picture. Three sides of a box meet at the solution $\boldsymbol{x} = (x, y, z) = (2, 3, 4)$:

$$\begin{aligned} 1x + 0y + 0z &= 2 \\ 0x + 1y + 0z &= 3 \\ 0x + 0y + 1z &= 4 \end{aligned} \quad \text{or} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}.$$

Draw the vectors in the column picture. Two times column 1 plus three times column 2 plus four times column 3 equals the right side $\boldsymbol{b}$.

2 If the equations in Problem 1 are multiplied by $2, 3, 4$ they become $DX = \boldsymbol{B}$:

$$\begin{aligned} 2x + 0y + 0z &= 4 \\ 0x + 3y + 0z &= 9 \\ 0x + 0y + 4z &= 16 \end{aligned} \quad \text{or} \quad DX = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 4 \\ 9 \\ 16 \end{bmatrix} = \boldsymbol{B}$$

Why is the row picture the same? Is the solution $\boldsymbol{X}$ the same as $\boldsymbol{x}$? What is changed in the column picture—the columns or the right combination to give $\boldsymbol{B}$?

3 If equation 1 is added to equation 2, which of these are changed: the planes in the row picture, the vectors in the column picture, the coefficient matrix, the solution? The new equations in Problem 1 would be $x = 2$, $x + y = 5$, $z = 4$.

4 Find a point with $z = 2$ on the intersection line of the planes $x + y + 3z = 6$ and $x - y + z = 4$. Find the point with $z = 0$. Find a third point halfway between.

5 The first of these equations plus the second equals the third:

$$\begin{aligned} x + y + z &= 2 \\ x + 2y + z &= 3 \\ 2x + 3y + 2z &= 5. \end{aligned}$$

The first two planes meet along a line. The third plane contains that line, because if x, y, z satisfy the first two equations then they also _____. The equations have infinitely many solutions (the whole line **L**). Find three solutions on **L**.

6 Move the third plane in Problem 5 to a parallel plane $2x + 3y + 2z = 9$. Now the three equations have no solution—*why not*? The first two planes meet along the line **L**, but the third plane doesn't _____ that line.

7 In Problem 5 the columns are $(1, 1, 2)$ and $(1, 2, 3)$ and $(1, 1, 2)$. This is a "singular case" because the third column is _____. Find two combinations of the columns that give $\boldsymbol{b} = (2, 3, 5)$. This is only possible for $\boldsymbol{b} = (4, 6, c)$ if $c =$ _____.

8 Normally 4 "planes" in 4-dimensional space meet at a _____. Normally 4 column vectors in 4-dimensional space can combine to produce $\boldsymbol{b}$. What combination of $(1, 0, 0, 0), (1, 1, 0, 0), (1, 1, 1, 0), (1, 1, 1, 1)$ produces $\boldsymbol{b} = (3, 3, 3, 2)$? What 4 equations for x, y, z, t are you solving?

Problems 9–14 are about multiplying matrices and vectors.

9 Compute each $A\boldsymbol{x}$ by dot products of the rows with the column vector:

$$\text{(a)}\quad \begin{bmatrix} 1 & 2 & 4 \\ -2 & 3 & 1 \\ -4 & 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} \qquad \text{(b)}\quad \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 2 \end{bmatrix}$$

10 Compute each $A\boldsymbol{x}$ in Problem 9 as a combination of the columns:

$$\text{9(a) becomes} \quad A\boldsymbol{x} = 2\begin{bmatrix} 1 \\ -2 \\ -4 \end{bmatrix} + 2\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} + 3\begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \\ \\ \\ \end{bmatrix}.$$

How many separate multiplications for $A\boldsymbol{x}$, when the matrix is "3 by 3"?

11 Find the two components of $A\boldsymbol{x}$ by rows or by columns:

$$\begin{bmatrix} 2 & 3 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} \text{ and } \begin{bmatrix} 3 & 6 \\ 6 & 12 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 2 & 4 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}.$$

12 Multiply A times $\boldsymbol{x}$ to find three components of $A\boldsymbol{x}$:

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \text{ and } \begin{bmatrix} 2 & 1 & 3 \\ 1 & 2 & 3 \\ 3 & 3 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \text{ and } \begin{bmatrix} 2 & 1 \\ 1 & 2 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

13 (a) A matrix with m rows and n columns multiplies a vector with _____ components to produce a vector with _____ components.

(b) The planes from the m equations $A\boldsymbol{x} = \boldsymbol{b}$ are in _____-dimensional space. The combination of the columns of A is in _____-dimensional space.

14 Write $2x+3y+z+5t = 8$ as a matrix A (how many rows?) multiplying the column vector $\boldsymbol{x} = (x, y, z, t)$ to produce $\boldsymbol{b}$. The solutions $\boldsymbol{x}$ fill a plane or "hyperplane" in 4-dimensional space. *The plane is 3-dimensional with no 4D volume.*

Problems 15–22 ask for matrices that act in special ways on vectors.

15 (a) What is the 2 by 2 identity matrix? I times $\begin{bmatrix} x \\ y \end{bmatrix}$ equals $\begin{bmatrix} x \\ y \end{bmatrix}$.

(b) What is the 2 by 2 exchange matrix? P times $\begin{bmatrix} x \\ y \end{bmatrix}$ equals $\begin{bmatrix} y \\ x \end{bmatrix}$.

16 (a) What 2 by 2 matrix R rotates every vector by $90°$? R times $\begin{bmatrix} x \\ y \end{bmatrix}$ is $\begin{bmatrix} y \\ -x \end{bmatrix}$.

(b) What 2 by 2 matrix R^2 rotates every vector by $180°$?

17 Find the matrix P that multiplies (x, y, z) to give (y, z, x). Find the matrix Q that multiplies (y, z, x) to bring back (x, y, z).

18 What 2 by 2 matrix E subtracts the first component from the second component? What 3 by 3 matrix does the same?

$$E\begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \quad \text{and} \quad E\begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 7 \end{bmatrix}.$$

19 What 3 by 3 matrix E multiplies (x, y, z) to give $(x, y, z + x)$? What matrix E^{-1} multiplies (x, y, z) to give $(x, y, z - x)$? If you multiply $(3, 4, 5)$ by E and then multiply by E^{-1}, the two results are (_____) and (_____).

20 What 2 by 2 matrix P_1 projects the vector (x, y) onto the x axis to produce $(x, 0)$? What matrix P_2 projects onto the y axis to produce $(0, y)$? If you multiply $(5, 7)$ by P_1 and then multiply by P_2, you get (_____) and (_____).

21 What 2 by 2 matrix R rotates every vector through $45°$? The vector $(1, 0)$ goes to $(\sqrt{2}/2, \sqrt{2}/2)$. The vector $(0, 1)$ goes to $(-\sqrt{2}/2, \sqrt{2}/2)$. Those determine the matrix. Draw these particular vectors in the xy plane and find R.

22 Write the dot product of $(1, 4, 5)$ and (x, y, z) as a matrix multiplication $A\boldsymbol{x}$. The matrix A has one row. The solutions to $A\boldsymbol{x} = \boldsymbol{0}$ lie on a _____ perpendicular to the vector _____. The columns of A are only in _____ -dimensional space.

23 In MATLAB notation, write the commands that define this matrix A and the column vectors $\boldsymbol{x}$ and $\boldsymbol{b}$. What command would test whether or not $A\boldsymbol{x} = \boldsymbol{b}$?

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad \boldsymbol{x} = \begin{bmatrix} 5 \\ -2 \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}$$

24 The MATLAB commands `A = eye(3)` and `v = [3:5]'` produce the 3 by 3 identity matrix and the column vector $(3, 4, 5)$. What are the outputs from `A*v` and `v'*v`? (Computer not needed!) If you ask for `v*A`, what happens?

25 If you multiply the 4 by 4 all-ones matrix A = ones(4) and the column v = ones(4,1), what is A∗v? (Computer not needed.) If you multiply B = eye(4) + ones(4) times w = zeros(4,1) + 2∗ones(4,1), what is B∗w?

Questions 26–28 review the row and column pictures in 2, 3, and 4 dimensions.

26 Draw the row and column pictures for the equations $x - 2y = 0, x + y = 6$.

27 For two linear equations in three unknowns x, y, z, the row picture will show (2 or 3) (lines or planes) in (2 or 3)-dimensional space. The column picture is in (2 or 3)-dimensional space. The solutions normally lie on a _____.

28 For four linear equations in two unknowns x and y, the row picture shows four _____. The column picture is in _____-dimensional space. The equations have no solution unless the vector on the right side is a combination of _____.

29 Start with the vector $\boldsymbol{u}_0 = (1,0)$. Multiply again and again by the same "Markov matrix" $A = [.8\ .3;\ .2\ .7]$. The next three vectors are $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3$:

$$\boldsymbol{u}_1 = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} .8 \\ .2 \end{bmatrix} \qquad \boldsymbol{u}_2 = A\boldsymbol{u}_1 = \underline{\qquad} \qquad \boldsymbol{u}_3 = A\boldsymbol{u}_2 = \underline{\qquad}.$$

What property do you notice for all four vectors $\boldsymbol{u}_0, \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3$?

Challenge Problems

30 Continue Problem 29 from $\boldsymbol{u}_0 = (1,0)$ to $\boldsymbol{u}_7$, and also from $\boldsymbol{v}_0 = (0,1)$ to $\boldsymbol{v}_7$. What do you notice about $\boldsymbol{u}_7$ and $\boldsymbol{v}_7$? Here are two MATLAB codes, with while and for. They plot $\boldsymbol{u}_0$ to $\boldsymbol{u}_7$ and $\boldsymbol{v}_0$ to $\boldsymbol{v}_7$. You can use other languages:

```
u = [1 ; 0]; A = [.8 .3 ; .2 .7];
x = u; k = [0 : 7];
while size(x,2) <= 7
  u = A*u; x = [x u];
end
plot(k, x)
```

```
v = [0 ; 1]; A = [.8 .3 ; .2 .7];
x = v; k = [0 : 7];
for j = 1 : 7
  v = A*v; x = [x v];
end
plot(k, x)
```

The $\boldsymbol{u}$'s and $\boldsymbol{v}$'s are approaching a steady state vector $\boldsymbol{s}$. Guess that vector and check that $A\boldsymbol{s} = \boldsymbol{s}$. If you start with $\boldsymbol{s}$, you stay with $\boldsymbol{s}$.

31 Invent a 3 by 3 **magic matrix** M_3 with entries $1, 2, \ldots, 9$. All rows and columns and diagonals add to 15. The first row could be 8, 3, 4. What is M_3 times $(1,1,1)$? What is M_4 times $(1,1,1,1)$ if a 4 by 4 magic matrix has entries $1, \ldots, 16$?

32 Suppose $\boldsymbol{u}$ and $\boldsymbol{v}$ are the first two columns of a 3 by 3 matrix A. Which third columns $\boldsymbol{w}$ would make this matrix singular? Describe a typical column picture of $A\boldsymbol{x} = \boldsymbol{b}$ in that singular case, and a typical row picture (for a random $\boldsymbol{b}$).

33 **Multiplying by A is a "linear transformation".** Those important words mean:

If w is a combination of u and v, then Aw is the same combination of Au and Av.

It is this "***linearity***" $Aw = cAu + dAv$ that gives us the name *linear algebra*.

Problem: If $u = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $v = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ then Au and Av are the columns of A.

Combine $w = cu + dv$. **If** $w = \begin{bmatrix} 5 \\ 7 \end{bmatrix}$ **how is Aw connected to Au and Av?**

34 Start from the four equations $-x_{i+1} + 2x_i - x_{i-1} = i$ (for $i = 1, 2, 3, 4$ with $x_0 = x_5 = 0$). Write those equations in their matrix form $Ax = b$. Can you solve them for x_1, x_2, x_3, x_4?

35 A 9 by 9 ***Sudoku matrix*** S has the numbers $1, \ldots, 9$ in every row and column, and in every 3 by 3 block. For the all-ones vector $x = (1, \ldots, 1)$, what is Sx?

A better question is: **Which row exchanges will produce another Sudoku matrix?** Also, which exchanges of block rows give another Sudoku matrix?

Section 2.7 will look at all possible permutations (reorderings) of the rows. I can see 6 orders for the first 3 rows, all giving Sudoku matrices. Also 6 permutations of the next 3 rows, and of the last 3 rows. And 6 block permutations of the block rows?

2.2 The Idea of Elimination

This chapter explains a systematic way to solve linear equations. The method is called ***"elimination"***, and you can see it immediately in our 2 by 2 example. Before elimination, x and y appear in both equations. After elimination, the first unknown x has disappeared from the second equation $8y = 8$:

$$\textbf{Before} \quad \begin{aligned} x - 2y &= 1 \\ 3x + 2y &= 11 \end{aligned} \qquad \textbf{After} \quad \begin{aligned} x - 2y &= 1 \\ 8y &= 8 \end{aligned} \qquad \begin{aligned} &(\textit{multiply equation } 1 \textit{ by } 3) \\ &(\textit{subtract to eliminate } 3x) \end{aligned}$$

The new equation $8y = 8$ instantly gives $y = 1$. Substituting $y = 1$ back into the first equation leaves $x - 2 = 1$. Therefore $x = 3$ and the solution $(x, y) = (3, 1)$ is complete.

Elimination produces an ***upper triangular system***—this is the goal. The nonzero coefficients $1, -2, 8$ form a triangle. That system is solved from the bottom upwards—first $y = 1$ and then $x = 3$. This quick process is called ***back substitution***. It is used for upper triangular systems of any size, after elimination gives a triangle.

Important point: The original equations have the same solution $x = 3$ and $y = 1$. Figure 2.5 shows each system as a pair of lines, intersecting at the solution point $(3, 1)$. After elimination, the lines still meet at the same point. Every step worked with correct equations.

How did we get from the first pair of lines to the second pair? We subtracted 3 times the first equation from the second equation. The step that eliminates x from equation 2 is the fundamental operation in this chapter. We use it so often that we look at it closely:

To eliminate x: Subtract a multiple of equation 1 from equation 2.

Three times $x - 2y = 1$ gives $3x - 6y = 3$. When this is subtracted from $3x + 2y = 11$, the right side becomes 8. The main point is that $3x$ cancels $3x$. What remains on the left side is $2y - (-6y)$ or $8y$, and x is eliminated. **The system became triangular.**

Ask yourself how that multiplier $\ell = 3$ was found. The first equation contains $\mathbf{1}x$. ***So the first pivot was* 1** (the coefficient of x). The second equation contains $\mathbf{3}x$, **so the multiplier was 3**. Then subtraction $3x - 3x$ produced the zero and the triangle.

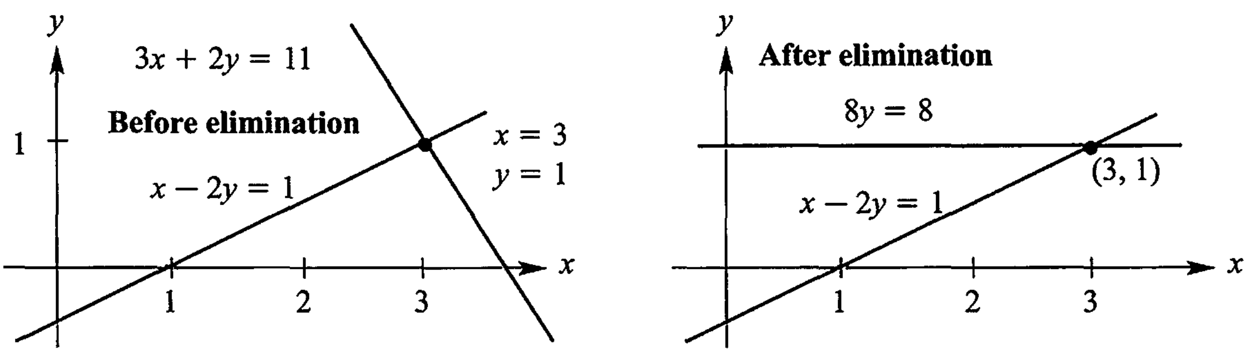


Figure 2.5: Eliminating x makes the second line horizontal. Then $8y = 8$ gives $y = 1$.

You will see the multiplier rule if I change the first equation to $4x - 8y = 4$. (Same straight line but the first pivot becomes 4.) The correct multiplier is now $\ell = \frac{3}{4}$. *To find the multiplier, divide the coefficient "* **3** *" to be eliminated by the pivot "* **4** *"*:

$$\begin{array}{lll} \mathbf{4x - 8y = 4} & \textbf{Multiply equation 1 by } \frac{3}{4} & 4x - 8y = 4 \\ \mathbf{3x + 2y = 11} & \textbf{Subtract from equation 2} & \phantom{4x -{}} 8y = 8. \end{array}$$

The final system is triangular and the last equation still gives $y = 1$. Back substitution produces $4x - 8 = 4$ and $4x = 12$ and $x = 3$. We changed the numbers but not the lines or the solution. ***Divide by the pivot to find that multiplier*** $\ell = \frac{3}{4}$:

Pivot	=	***first nonzero in the row that does the elimination***
Multiplier	=	***(entry to eliminate) divided by (pivot)*** $= \frac{3}{4}$.

The new second equation starts with the second pivot, which is 8. We would use it to eliminate y from the third equation if there were one. *To solve n equations we want n pivots.* ***The pivots are on the diagonal of the triangle after elimination.***

You could have solved those equations for x and y without reading this book. It is an extremely humble problem, but we stay with it a little longer. Even for a 2 by 2 system, elimination might break down. By understanding the possible breakdown (when we can't find a full set of pivots), you will understand the whole process of elimination.

Breakdown of Elimination

Normally, elimination produces the pivots that take us to the solution. But failure is possible. At some point, the method might ask us to *divide by zero.* We can't do it. The process has to stop. There might be a way to adjust and continue—or failure may be unavoidable.

Example 1 fails with ***no solution to*** $0y = 8$. Example 2 fails with ***too many solutions to*** $0y = 0$. Example 3 succeeds by exchanging the equations.

Example 1 ***Permanent failure with no solution.*** Elimination makes this clear:

$$\begin{array}{lll} x - 2y = 1 & \text{Subtract 3 times} & x - 2y = 1 \\ 3x - 6y = 11 & \text{eqn. 1 from eqn. 2} & \phantom{x-{}} \mathbf{0}y = 8. \end{array}$$

There is *no* solution to $0y = 8$. Normally we divide the right side 8 by the second pivot, but *this system has no second pivot.* ***(Zero is never allowed as a pivot!)*** The row and column pictures in Figure 2.6 show why failure was unavoidable. If there is no solution, elimination will discover that fact by reaching an equation like $0y = 8$.

The row picture of failure shows parallel lines—which never meet. A solution must lie on both lines. With no meeting point, the equations have no solution.

The column picture shows the two columns $(1, 3)$ and $(-2, -6)$ in the same direction. *All combinations of the columns lie along a line.* But the column from the right side is in a different direction $(1, 11)$. No combination of the columns can produce this right side—therefore no solution.

When we change the right side to $(1, 3)$, failure shows as a whole line of solution points. Instead of no solution, next comes Example 2 with infinitely many.

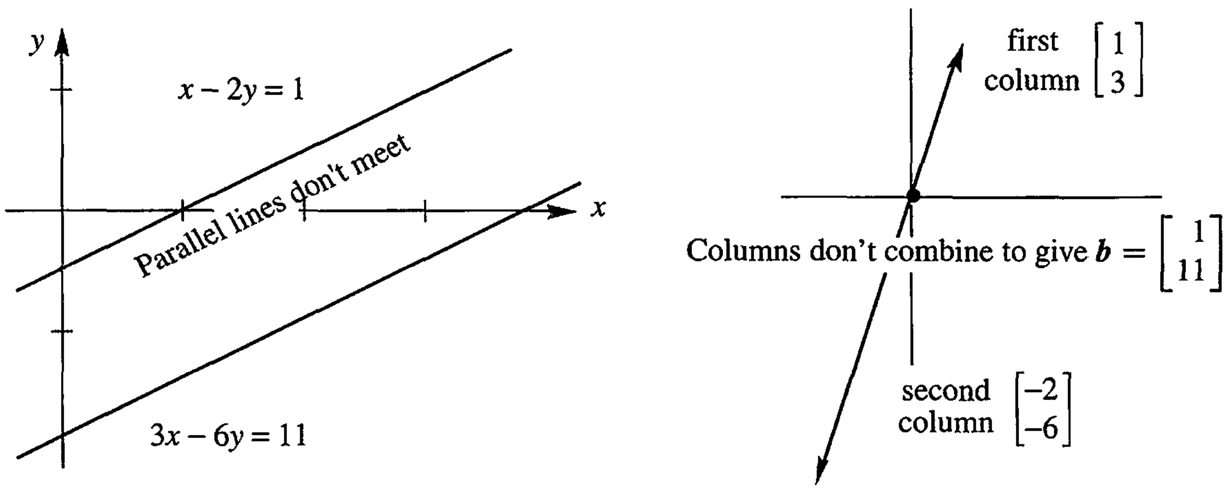


Figure 2.6: Row picture and column picture for Example 1: ***no solution.***

Example 2 ***Failure with infinitely many solutions. Change*** $\boldsymbol{b} = (1, 11)$ ***to*** $(1, 3)$.

$$\begin{array}{rcl} x - 2y = 1 & \text{Subtract 3 times} & x - 2y = 1 \\ 3x - 6y = 3 & \text{eqn. 1 from eqn. 2} & \mathbf{0}y = \mathbf{0}. \end{array} \quad \begin{array}{l}\text{Still only}\\ \textbf{one pivot.}\end{array}$$

Every y satisfies $0y = 0$. There is really only one equation $x - 2y = 1$. The unknown y is *"free"*. After y is freely chosen, x is determined as $x = 1 + 2y$.

In the row picture, the parallel lines have become the same line. Every point on that line satisfies both equations. We have a whole line of solutions in Figure 2.7.

In the column picture, $\boldsymbol{b} = (1, 3)$ is now the same as column 1. So we can choose $x = 1$ and $y = 0$. We can also choose $x = 0$ and $y = -\frac{1}{2}$; column 2 times $-\frac{1}{2}$ equals $\boldsymbol{b}$. Every (x, y) that solves the row problem also solves the column problem.

Failure For n equations we do not get n pivots

Elimination leads to an equation $\mathbf{0} \neq \mathbf{0}$ (no solution) or $\mathbf{0} = \mathbf{0}$ (many solutions)

Success comes with n pivots. But we may have to exchange the n equations.

Elimination can go wrong in a third way—but this time it can be fixed. *Suppose the first pivot position contains zero.* We refuse to allow zero as a pivot. When the first equation has no term involving x, we can exchange it with an equation below:

Example 3 ***Temporary failure (zero in pivot). A row exchange produces two pivots:***

$$\textbf{Permutation} \qquad \begin{array}{rcl} 0x + 2y = 4 & \text{Exchange the} & 3x - 2y = 5 \\ 3x - 2y = 5 & \text{two equations} & 2y = 4. \end{array}$$

The new system is already triangular. This small example is ready for back substitution. The last equation gives $y = 2$, and then the first equation gives $x = 3$. The row picture is normal (two intersecting lines). The column picture is also normal (column vectors not in the same direction). The pivots 3 and 2 are normal—but a ***row exchange*** was required.

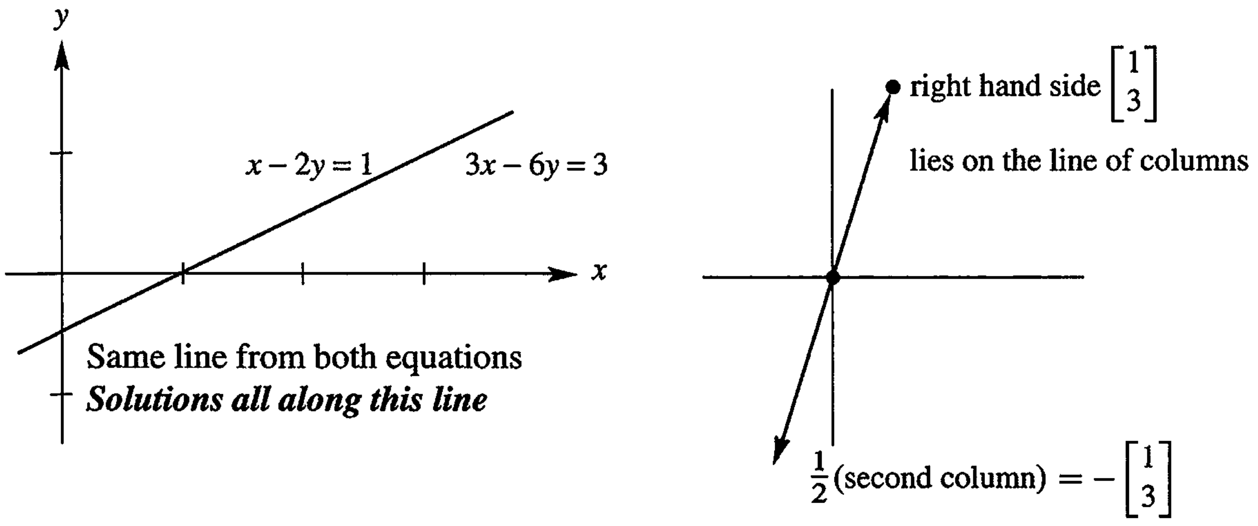


Figure 2.7: Row and column pictures for Example 2: *infinitely many solutions.*

Examples 1 and 2 are ***singular***—there is no second pivot. Example 3 is ***nonsingular***—there is a full set of pivots and exactly one solution. Singular equations have no solution or infinitely many solutions. Pivots must be nonzero because we have to divide by them.

Three Equations in Three Unknowns

To understand Gaussian elimination, you have to go beyond 2 by 2 systems. Three by three is enough to see the pattern. For now the matrices are square—an equal number of rows and columns. Here is a 3 by 3 system, specially constructed so that all steps lead to whole numbers and not fractions:

$$\begin{aligned} \mathbf{2}x + 4y - 2z &= 2 \\ 4x + 9y - 3z &= 8 \\ -2x - 3y + 7z &= 10 \end{aligned} \tag{1}$$

What are the steps? The first pivot is the boldface **2** (upper left). Below that pivot we want to eliminate the 4. *The first multiplier is the ratio* $4/2 = 2$. Multiply the pivot equation by $\ell_{21} = 2$ and subtract. Subtraction removes the $4x$ from the second equation:

Step 1 Subtract 2 times equation 1 from equation 2. This leaves $y + z = 4$.

We also eliminate $-2x$ from equation 3—still using the first pivot. The quick way is to add equation 1 to equation 3. Then $2x$ cancels $-2x$. We do exactly that, but the rule in this book is to *subtract rather than add.* The systematic pattern has multiplier $\ell_{31} = -2/2 = -1$. Subtracting -1 times an equation is the same as adding:

Step 2 Subtract -1 times equation 1 from equation 3. This leaves $y + 5z = 12$.

The two new equations involve only y and z. The second pivot (in boldface) is 1:

$$\textbf{x is eliminated} \qquad \begin{aligned} \mathbf{1}y + 1z &= 4 \\ 1y + 5z &= 12 \end{aligned}$$

We have reached a 2 *by* 2 *system.* The final step eliminates y to make it 1 by 1:

Step 3 Subtract equation 2_{new} from 3_{new}. The multiplier is $1/1 = 1$. Then $4z = 8$.

The original $A\boldsymbol{x} = \boldsymbol{b}$ has been converted into an upper triangular $U\boldsymbol{x} = \boldsymbol{c}$:

$$\begin{array}{r} 2x + 4y - 2z = 2 \\ 4x + 9y - 3z = 8 \\ -2x - 3y + 7z = 10 \end{array} \quad \begin{array}{c} A\boldsymbol{x} = \boldsymbol{b} \\ \text{has become} \\ U\boldsymbol{x} = \boldsymbol{c} \end{array} \quad \begin{array}{r} 2x + 4y - 2z = 2 \\ 1y + 1z = 4 \\ 4z = 8. \end{array} \tag{2}$$

The goal is achieved—forward elimination is complete from A to U. ***Notice the pivots* $2, 1, 4$ *along the diagonal of* U.** The pivots 1 and 4 were hidden in the original system. Elimination brought them out. $U\boldsymbol{x} = \boldsymbol{c}$ is ready for ***back substitution***, which is quick:

$$(4z = 8 \quad \text{gives} \quad z = 2) \qquad (y + z = 4 \quad \text{gives} \quad y = 2) \qquad (\text{equation 1 gives} \quad x = -1)$$

The solution is $(x, y, z) = (-1, 2, 2)$. The row picture has three planes from three equations. All the planes go through this solution. The original planes are sloping, but the last plane $4z = 8$ after elimination is horizontal.

The column picture shows a combination $A\boldsymbol{x}$ of column vectors producing the right side $\boldsymbol{b}$. The coefficients in that combination are $-1, 2, 2$ (the solution):

$$A\boldsymbol{x} = (-1)\begin{bmatrix} 2 \\ 4 \\ -2 \end{bmatrix} + 2\begin{bmatrix} 4 \\ 9 \\ -3 \end{bmatrix} + 2\begin{bmatrix} -2 \\ -3 \\ 7 \end{bmatrix} \text{ equals } \begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix} = \boldsymbol{b}. \tag{3}$$

The numbers x, y, z multiply columns $1, 2, 3$ in $A\boldsymbol{x} = \boldsymbol{b}$ and also in the triangular $U\boldsymbol{x} = \boldsymbol{c}$.

For a 4 by 4 problem, or an n by n problem, elimination proceeds the same way. Here is the whole idea, column by column from A to U, when elimination succeeds.

Column 1. ***Use the first equation to create zeros below the first pivot.***

Column 2. ***Use the new equation 2 to create zeros below the second pivot.***

Columns 3 to *n*. ***Keep going to find all n pivots and the triangular U.***

$$\text{After column 2 we have} \quad \begin{bmatrix} \boldsymbol{x} & x & x & x \\ 0 & \boldsymbol{x} & x & x \\ 0 & 0 & x & x \\ 0 & 0 & x & x \end{bmatrix}. \quad \text{We want} \quad \begin{bmatrix} \boldsymbol{x} & x & x & x \\ & \boldsymbol{x} & x & x \\ & & \boldsymbol{x} & x \\ & & & \boldsymbol{x} \end{bmatrix}. \tag{4}$$

The result of forward elimination is an upper triangular system. It is nonsingular if there is a full set of n pivots (never zero!). *Question*: Which x on the left could be changed to boldface $\boldsymbol{x}$ because the pivot is known? Here is a final example to show the original $A\boldsymbol{x} = \boldsymbol{b}$, the triangular system $U\boldsymbol{x} = \boldsymbol{c}$, and the solution (x, y, z) from back substitution:

$$\begin{array}{rl} x + y + z = 6 & \\ x + 2y + 2z = 9 & \textbf{Forward} \\ x + 2y + 3z = 10 & \textbf{Forward} \end{array} \qquad \begin{array}{r} x + y + z = 6 \\ y + z = 3 \\ z = 1 \end{array} \qquad \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \quad \begin{array}{l} \textbf{Back} \\ \textbf{Back} \\ \end{array}$$

All multipliers are 1. All pivots are 1. All planes meet at the solution $(3, 2, 1)$. The columns of A combine with $3, 2, 1$ to give $\boldsymbol{b} = (6, 9, 10)$. The triangle shows $U\boldsymbol{x} = \boldsymbol{c} = (6, 3, 1)$.

■ REVIEW OF THE KEY IDEAS ■

1. A linear system ($A\boldsymbol{x} = \boldsymbol{b}$) becomes upper triangular ($U\boldsymbol{x} = \boldsymbol{c}$) after elimination.
2. We subtract ℓ_{ij} times equation j from equation i, to make the (i, j) entry zero.
3. The multiplier is $\ell_{ij} = \dfrac{\text{entry to eliminate in row } i}{\text{pivot in row } j}$. Pivots can not be zero!
4. A zero in the pivot position can be repaired if there is a nonzero below it.
5. The upper triangular system is solved by back substitution (starting at the bottom).
6. When breakdown is permanent, the system has no solution or infinitely many.

■ WORKED EXAMPLES ■

2.2 A When elimination is applied to this matrix A, what are the first and second pivots? What is the multiplier ℓ_{21} in the first step (ℓ_{21} times row 1 is *subtracted* from row 2)?

A has a ***first difference*** in row 1 and a ***second difference*** $-1, 2, -1$ in row 2.

$$A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 2 \end{bmatrix} \longrightarrow U = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

What entry in the 2, 2 position (instead of 2) would force an exchange of rows 2 and 3? Why is the lower left multiplier $\ell_{31} = 0$, subtracting zero times row 1 from row 3?
If you change the corner entry from $a_{33} = 2$ ***to*** $a_{33} = 1$***, why does elimination fail?***

Solution The first pivot is 1. The multiplier ℓ_{21} is $-1/1 = -1$. When -1 times row 1 is subtracted (so row 1 is added to row 2), the second pivot is revealed as 1.

If we reduce the middle entry "2" to "1", that would force a row exchange. (Zero will appear in the second pivot position.) The multiplier ℓ_{31} is zero because $a_{31} = 0$. A zero at the start of a row needs no elimination. This A is a "*band matrix*".

The last pivot is 1. So if the original corner entry a_{33} is reduced by 1 (to $a_{33} = 1$), elimination would produce 0. **No third pivot, elimination fails.**

2.2 B Suppose A is already a ***triangular matrix*** (upper triangular or lower triangular). *Where do you see its pivots*? When does $A\boldsymbol{x} = \boldsymbol{b}$ have exactly one solution for every $\boldsymbol{b}$?

Solution The pivots of a triangular matrix are already set along the main diagonal. *Elimination succeeds when all those numbers are nonzero.* Use ***back*** substitution when A is upper triangular, go ***forward*** when A is lower triangular.

2.2 C Use elimination to reach upper triangular matrices U. Solve by back substitution or explain why this is impossible. What are the pivots (never zero)? Exchange equations when necessary. The only difference is the $-x$ in the last equation.

$$\begin{array}{rrr} \textbf{Success} & x+y+z=7 & x+y+z=7 \\ \textbf{then} & x+y-z=5 & x+y-z=5 \\ \textbf{Failure} & x-y+z=3 & -x-y+z=3 \end{array}$$

Solution For the first system, subtract equation 1 from equations 2 and 3 (the multipliers are $\ell_{21}=1$ and $\ell_{31}=1$). The 2, 2 entry becomes zero, so exchange equations:

$$\textbf{Success} \quad \begin{aligned} x+y+z &= 7 \\ \mathbf{0}y-2z &= -2 \\ -2y+0z &= -4 \end{aligned} \quad \text{exchanges into} \quad \begin{aligned} x+y+z &= 7 \\ -\mathbf{2}y+0z &= -4 \\ -2z &= -2 \end{aligned}$$

Then back substitution gives $z=1$ and $y=2$ and $x=4$. The pivots are $1, -2, -2$.

For the second system, subtract equation 1 from equation 2 as before. Add equation 1 to equation 3. This leaves zero in the 2, 2 entry *and also below*:

$$\textbf{Failure} \quad \begin{aligned} x+y+z &= 7 \\ \mathbf{0}y-2z &= -2 \\ \mathbf{0}y+2z &= 10 \end{aligned} \quad \begin{array}{l} \text{There is } \textit{no pivot} \text{ in column 2 (it was – column 1)} \\ \text{A further elimination step gives } 0z=8 \\ \text{The three planes don't meet} \end{array}$$

Plane 1 meets plane 2 in a line. Plane 1 meets plane 3 in a parallel line. *No solution.*

If we change the "3" in the original third equation to "-5" then elimination would lead to $0=0$. There are infinitely many solutions! *The three planes now meet along a whole line.*

Changing 3 to -5 moved the third plane to meet the other two. The second equation gives $z=1$. Then the first equation leaves $x+y=6$. **No pivot in column 2 makes y free** (it can have any value). Then $x=6-y$.

Problem Set 2.2

Problems 1–10 are about elimination on 2 by 2 systems.

1 What multiple ℓ_{21} of equation 1 should be subtracted from equation 2?

$$\begin{aligned} 2x+3y &= 1 \\ 10x+9y &= 11. \end{aligned}$$

After this elimination step, write down the upper triangular system and circle the two pivots. The numbers 1 and 11 have no influence on those pivots.

2 Solve the triangular system of Problem 1 by back substitution, y before x. Verify that x times $(2,10)$ plus y times $(3,9)$ equals $(1,11)$. If the right side changes to $(4,44)$, what is the new solution?

3 What multiple of equation 1 should be *subtracted* from equation 2?

$$\begin{aligned} 2x - 4y &= 6 \\ -x + 5y &= 0. \end{aligned}$$

After this elimination step, solve the triangular system. If the right side changes to $(-6, 0)$, what is the new solution?

4 What multiple ℓ of equation 1 should be subtracted from equation 2 to remove c?

$$\begin{aligned} ax + by &= f \\ cx + dy &= g. \end{aligned}$$

The first pivot is a (assumed nonzero). Elimination produces what formula for the second pivot? What is y? The second pivot is missing when $ad = bc$: singular.

5 Choose a right side which gives no solution and another right side which gives infinitely many solutions. What are two of those solutions?

$$\textbf{Singular system} \qquad \begin{aligned} 3x + 2y &= 10 \\ 6x + 4y &= \end{aligned}$$

6 Choose a coefficient b that makes this system singular. Then choose a right side g that makes it solvable. Find two solutions in that singular case.

$$\begin{aligned} 2x + by &= 16 \\ 4x + 8y &= g. \end{aligned}$$

7 For which numbers a does elimination break down (1) permanently (2) temporarily?

$$\begin{aligned} ax + 3y &= -3 \\ 4x + 6y &= 6. \end{aligned}$$

Solve for x and y after fixing the temporary breakdown by a row exchange.

8 For which three numbers k does elimination break down? Which is fixed by a row exchange? In each case, is the number of solutions 0 or 1 or ∞?

$$\begin{aligned} kx + 3y &= 6 \\ 3x + ky &= -6. \end{aligned}$$

9 What test on b_1 and b_2 decides whether these two equations allow a solution? How many solutions will they have? Draw the column picture for $\boldsymbol{b} = (1, 2)$ and $(1, 0)$.

$$\begin{aligned} 3x - 2y &= b_1 \\ 6x - 4y &= b_2. \end{aligned}$$

10 In the xy plane, draw the lines $x + y = 5$ and $x + 2y = 6$ and the equation $y = $ _____ that comes from elimination. The line $5x - 4y = c$ will go through the solution of these equations if $c = $ _____.

Problems 11–20 study elimination on 3 by 3 systems (and possible failure).

11 (Recommended) A system of linear equations can't have exactly two solutions. *Why*?

(a) If (x, y, z) and (X, Y, Z) are two solutions, what is another solution?

(b) If 25 planes meet at two points, where else do they meet?

12 Reduce this system to upper triangular form by two row operations:

$$\begin{aligned} 2x + 3y + z &= 8 \\ 4x + 7y + 5z &= 20 \\ -2y + 2z &= 0. \end{aligned}$$

Circle the pivots. Solve by back substitution for z, y, x.

13 Apply elimination (circle the pivots) and back substitution to solve

$$\begin{aligned} 2x - 3y \quad\;\; &= 3 \\ 4x - 5y + \;\; z &= 7 \\ 2x - \;\; y - 3z &= 5. \end{aligned}$$

List the three row operations: Subtract ____ times row ____ from row ____.

14 Which number d forces a row exchange, and what is the triangular system (not singular) for that d? Which d makes this system singular (no third pivot)?

$$\begin{aligned} 2x + 5y + z &= 0 \\ 4x + dy + z &= 2 \\ y - z &= 3. \end{aligned}$$

15 Which number b leads later to a row exchange? Which b leads to a missing pivot? In that singular case find a nonzero solution x, y, z.

$$\begin{aligned} x + by \quad\;\; &= 0 \\ x - 2y - z &= 0 \\ y + z &= 0. \end{aligned}$$

16 (a) Construct a 3 by 3 system that needs two row exchanges to reach a triangular form and a solution.

(b) Construct a 3 by 3 system that needs a row exchange to keep going, but breaks down later.

17 If rows 1 and 2 are the same, how far can you get with elimination (allowing row exchange)? If columns 1 and 2 are the same, which pivot is missing?

Equal	$2x - y + z = 0$	$2x + 2y + z = 0$	**Equal**
rows	$2x - y + z = 0$	$4x + 4y + z = 0$	**columns**
	$4x + y + z = 2$	$6x + 6y + z = 2.$	

18 Construct a 3 by 3 example that has 9 different coefficients on the left side, but rows 2 and 3 become zero in elimination. How many solutions to your system with $\boldsymbol{b} = (1, 10, 100)$ and how many with $\boldsymbol{b} = (0, 0, 0)$?

19 Which number q makes this system singular and which right side t gives it infinitely many solutions? Find the solution that has $z = 1$.

$$\begin{aligned} x + 4y - 2z &= 1 \\ x + 7y - 6z &= 6 \\ 3y + qz &= t. \end{aligned}$$

20 Three planes can fail to have an intersection point, even if no planes are parallel. The system is singular if row 3 of A is a ____ of the first two rows. Find a third equation that can't be solved together with $x + y + z = 0$ and $x - 2y - z = 1$.

21 Find the pivots and the solution for both systems ($A\boldsymbol{x} = \boldsymbol{b}$ and $K\boldsymbol{x} = \boldsymbol{b}$):

$$\begin{aligned} 2x + y \quad\quad &= 0 \\ x + 2y + z \quad &= 0 \\ y + 2z + t &= 0 \\ z + 2t &= 5 \end{aligned} \qquad\qquad \begin{aligned} 2x - y \quad\quad &= 0 \\ -x + 2y - z \quad &= 0 \\ -y + 2z - t &= 0 \\ -z + 2t &= 5. \end{aligned}$$

22 If you extend Problem 21 following the $1, 2, 1$ pattern or the $-1, 2, -1$ pattern, what is the fifth pivot? What is the nth pivot? K is my favorite matrix.

23 If elimination leads to $x + y = 1$ and $2y = 3$, find three possible original problems.

24 For which two numbers a will elimination fail on $A = \begin{bmatrix} a & 2 \\ a & a \end{bmatrix}$?

25 For which three numbers a will elimination fail to give three pivots?

$$A = \begin{bmatrix} a & 2 & 3 \\ a & a & 4 \\ a & a & a \end{bmatrix} \text{ is singular for three values of } a.$$

26 Look for a matrix that has row sums 4 and 8, and column sums 2 and s:

$$\text{Matrix } = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \begin{aligned} a + b &= 4 & a + c &= 2 \\ c + d &= 8 & b + d &= s \end{aligned}$$

The four equations are solvable only if $s =$ ____. Then find two different matrices that have the correct row and column sums. *Extra credit*: Write down the 4 by 4 system $A\boldsymbol{x} = \boldsymbol{b}$ with $\boldsymbol{x} = (a, b, c, d)$ and make A triangular by elimination.

27 Elimination in the usual order gives what matrix U and what solution to this "lower triangular" system? We are really solving by *forward substitution*:

$$\begin{aligned} 3x \quad\quad\quad &= 3 \\ 6x + 2y \quad &= 8 \\ 9x - 2y + z &= 9. \end{aligned}$$

28 Create a MATLAB command A(2, :) = ... for the new row 2, to subtract 3 times row 1 from the existing row 2 if the matrix A is already known.

Challenge Problems

29 Find experimentally the average 1st and 2nd and 3rd pivot sizes from MATLAB 's $[L, U] =$ **lu(rand**(3)). The average size **abs**(U(1, 1)) is above $\frac{1}{2}$ because **lu** picks the largest available pivot in column 1. Here $A =$ **rand**(3) has random entries between 0 and 1.

30 If the last corner entry is $A(5,5) = 11$ and the last pivot of A is $U(5,5) = 4$, what different entry $A(5,5)$ would have made A singular?

31 Suppose elimination takes A to U without row exchanges. Then row j of U is a combination of which rows of A? If $A\boldsymbol{x} = \mathbf{0}$, is $U\boldsymbol{x} = \mathbf{0}$? If $A\boldsymbol{x} = \boldsymbol{b}$, is $U\boldsymbol{x} = \boldsymbol{b}$? If A starts out lower triangular, what is the upper triangular U?

32 Start with 100 equations $A\boldsymbol{x} = \mathbf{0}$ for 100 unknowns $\boldsymbol{x} = (x_1, \ldots, x_{100})$. Suppose elimination reduces the 100th equation to $0 = 0$, so the system is "singular".

(a) Elimination takes linear combinations of the rows. So this singular system has the singular property: Some linear combination of the 100 **rows** is _____.

(b) Singular systems $A\boldsymbol{x} = \mathbf{0}$ have infinitely many solutions. This means that some linear combination of the 100 ***columns*** is _____.

(c) Invent a 100 by 100 singular matrix with no zero entries.

(d) For your matrix, describe in words the row picture and the column picture of $A\boldsymbol{x} = \mathbf{0}$. Not necessary to draw 100-dimensional space.

2.3 Elimination Using Matrices

We now combine two ideas—elimination and matrices. The goal is to express all the steps of elimination (and the final result) in the clearest possible way. In a 3 by 3 example, elimination could be described in words. For larger systems, a long list of steps would be hopeless. You will see how to subtract a multiple of row j from row i—*using a matrix E.*

The 3 by 3 example in the previous section has the beautifully short form $A\boldsymbol{x} = \boldsymbol{b}$:

$$\begin{aligned} 2x_1 + 4x_2 - 2x_3 &= 2 \\ 4x_1 + 9x_2 - 3x_3 &= 8 \\ -2x_1 - 3x_2 + 7x_3 &= 10 \end{aligned} \quad \text{is the same as} \quad \begin{bmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix}. \qquad (1)$$

The nine numbers on the left go into the matrix A. That matrix not only sits beside $\boldsymbol{x}$, it *multiplies* $\boldsymbol{x}$. The rule for "A times $\boldsymbol{x}$" is exactly chosen to yield the three equations.

Review of A times $\boldsymbol{x}$. A matrix times a vector gives a vector. The matrix is square when the number of equations (three) matches the number of unknowns (three). Our matrix is 3 by 3. A general square matrix is n by n. Then the vector $\boldsymbol{x}$ is in n-dimensional space.

$$\textbf{\textit{The unknown in }} \mathbf{R}^3 \textbf{\textit{ is}} \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad \textbf{\textit{and the solution is}} \quad \boldsymbol{x} = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}.$$

Key point: $A\boldsymbol{x} = \boldsymbol{b}$ represents the row form and also the column form of the equations.

$$\textbf{Column form} \qquad A\boldsymbol{x} = (-1)\begin{bmatrix} 2 \\ 4 \\ -2 \end{bmatrix} + 2\begin{bmatrix} 4 \\ 9 \\ -3 \end{bmatrix} + 2\begin{bmatrix} -2 \\ -3 \\ 7 \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix} = \boldsymbol{b}.$$

This rule for $A\boldsymbol{x}$ is used so often that we express it once more for emphasis.

$A\boldsymbol{x}$ is a combination of the columns of A. Components of $\boldsymbol{x}$ multiply those columns:

$$A\boldsymbol{x} = x_1 \text{ times (column 1)} + \cdots + x_n \text{ times (column } n).$$

When we compute the components of $A\boldsymbol{x}$, we use the row form of matrix multiplication. The ith component is a dot product with row i of A, which is $[a_{i1} \;\; a_{i2} \;\; \ldots \;\; a_{in}]$. The short formula for that dot product with $\boldsymbol{x}$ uses "sigma notation".

Components of $A\boldsymbol{x}$ are dot products with rows of A.

$$\text{The } i\text{th component of } A\boldsymbol{x} \text{ is} \quad a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n. \quad \text{This is} \quad \sum_{j=1}^{n} a_{ij}x_j.$$

The sigma symbol $\sum$ is an instruction to add.[1] Start with $j = 1$ and stop with $j = n$. Start the sum with $a_{i1}x_1$ and stop with $a_{in}x_n$. That produces (row i) $\cdot\, \boldsymbol{x}$.

[1]Einstein shortened this even more by omitting the $\sum$. The repeated j in $a_{ij}x_j$ automatically meant addition. He also wrote the sum as $a_i^j x_j$. Not being Einstein, we include the $\sum$.

One point to repeat about matrix notation: The entry in row 1, column 1 (the top left corner) is a_{11}. The entry in row 1, column 3 is a_{13}. The entry in row 3, column 1 is a_{31}. (Row number comes before column number.) The word "entry" for a matrix corresponds to "component" for a vector. General rule: $a_{ij} = A(i,j)$ ***is in row i, column j.***

Example 1 This matrix has $a_{ij} = 2i + j$. Then $a_{11} = 3$. Also $a_{12} = 4$ and $a_{21} = 5$. Here is Ax with numbers and letters:

$$\begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \cdot 2 + 4 \cdot 1 \\ 5 \cdot 2 + 6 \cdot 1 \end{bmatrix} \qquad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}.$$

The first component of Ax is $6 + 4 = 10$. ***A row times a column gives a dot product.***

The Matrix Form of One Elimination Step

$Ax = b$ is a convenient form for the original equation. What about the elimination steps? The first step in this example subtracts 2 times the first equation from the second equation. On the right side, 2 times the first component of b is subtracted from the second component:

$$\textbf{First step} \qquad b = \begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix} \quad \text{changes to} \quad b_{\text{new}} = \begin{bmatrix} 2 \\ 4 \\ 10 \end{bmatrix}.$$

We want to do that subtraction with a matrix! The same result $b_{\text{new}} = Eb$ is achieved when we multiply an "elimination matrix" E times b. It subtracts $2b_1$ from b_2:

$$\textit{The elimination matrix is} \qquad E = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Multiplication by E subtracts 2 times row 1 from row 2. Rows 1 and 3 stay the same:

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 10 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 - 2b_1 \\ b_3 \end{bmatrix}$$

The first and third rows of E are rows from the identity matrix I. The new second component is the number 4 that appeared after the elimination step. This is $b_2 - 2b_1$.

It is easy to describe the "elementary matrices" or "elimination matrices" like this E. Start with the identity matrix I. *Change one of its zeros to the multiplier* $-\ell$:

The ***identity matrix*** has 1's on the diagonal and otherwise 0's. Then $Ib = b$ for all b. The ***elementary matrix or elimination matrix*** E_{ij} that subtracts a multiple ℓ of row j from row i has the extra nonzero entry $-\ell$ in the i,j position (still diagonal 1's).

Example 2 The matrix E_{31} has $-\ell$ in the 3, 1 position:

$$\textbf{Identity} \quad I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \textbf{Elimination} \quad E_{31} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\ell & 0 & 1 \end{bmatrix}.$$

When you multiply I times $\boldsymbol{b}$, you get $\boldsymbol{b}$. But E_{31} subtracts ℓ times the first component from the third component. With $\ell = 4$ this example gives $9 - 4 = 5$:

$$I\boldsymbol{b} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix} \quad \text{and} \quad E\boldsymbol{b} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{1} \\ 3 \\ \mathbf{9} \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ \mathbf{5} \end{bmatrix}.$$

What about the left side of $A\boldsymbol{x} = \boldsymbol{b}$? Both sides are multiplied by E_{31}. ***The purpose of E_{31} is to produce a zero in the (3, 1) position of the matrix.***

The notation fits this purpose. Start with A. Apply E's to produce zeros below the pivots (the first E is E_{21}). End with a triangular U. We now look in detail at those steps.

First a small point. The vector $\boldsymbol{x}$ stays the same. The solution is not changed by elimination. (That may be more than a small point.) It is the coefficient matrix that is changed. When we start with $A\boldsymbol{x} = \boldsymbol{b}$ and multiply by E, the result is $EA\boldsymbol{x} = E\boldsymbol{b}$. The new matrix EA is the result of *multiplying E times A.*

Confession The *elimination matrices E_{ij}* are great examples, but you won't see them later. They show how a matrix acts on rows. By taking several elimination steps, we will see how to *multiply matrices* (and the order of the E's becomes important). ***Products and inverses*** are especially clear for E's. It is those two ideas that the book will now use.

Matrix Multiplication

The big question is: ***How do we multiply two matrices?*** When the first matrix is E, we already know what to expect for EA. This particular E subtracts 2 times row 1 from row 2 of this matrix A and any matrix. The multiplier is $\ell = 2$:

$$EA = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{bmatrix} = \begin{bmatrix} 2 & 4 & -2 \\ \mathbf{0} & \mathbf{1} & \mathbf{1} \\ -2 & -3 & 7 \end{bmatrix} \quad (\textit{with the zero}). \qquad (2)$$

This step does not change rows 1 and 3 of A. Those rows are unchanged in EA—only row 2 is different. *Twice the first row has been subtracted from the second row.* Matrix multiplication agrees with elimination—and the new system of equations is $EA\boldsymbol{x} = E\boldsymbol{b}$.

$EA\boldsymbol{x}$ is simple but it involves a subtle idea. Start with $A\boldsymbol{x} = \boldsymbol{b}$. Multiplying both sides by E gives $E(A\boldsymbol{x}) = E\boldsymbol{b}$. With matrix multiplication, this is also $(EA)\boldsymbol{x} = E\boldsymbol{b}$. **The first was E times $A\boldsymbol{x}$, the second is EA times $\boldsymbol{x}$. They are the same.** Parentheses are not needed. We just write $EA\boldsymbol{x}$.

That rule extends to a matrix C with several column vectors like $C = [\boldsymbol{c}_1 \; \boldsymbol{c}_2 \; \boldsymbol{c}_3]$. When multiplying EAC, you can do AC first or EA first. This is the point of an "associative law" like $3 \times (4 \times 5) = (3 \times 4) \times 5$. Multiply 3 times 20, or multiply 12 times 5. Both answers are 60. That law seems so clear that it is hard to imagine it could be false.

The "commutative law" $3 \times 4 = 4 \times 3$ looks even more obvious. But EA is usually different from AE. When E multiplies on the right, it acts on the ***columns*** of A.

Associative law is true	$A(BC) = (AB)C$
Commutative law is false	Often $AB \neq BA$

There is another requirement on matrix multiplication. Suppose B has only one column (this column is $\boldsymbol{b}$). The matrix-matrix law for EB should agree with the matrix-vector law for $E\boldsymbol{b}$. Even more, we should be able to *multiply matrices EB a column at a time*:

If B has several columns $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3$, then the columns of EB are $E\boldsymbol{b}_1, E\boldsymbol{b}_2, E\boldsymbol{b}_3$.

Matrix multiplication $\quad AB = A\,[\boldsymbol{b}_1\ \boldsymbol{b}_2\ \boldsymbol{b}_3] = [A\boldsymbol{b}_1\ A\boldsymbol{b}_2\ A\boldsymbol{b}_3].$ (3)

This holds true for the matrix multiplication in (2). If you multiply column 3 of A by E, you correctly get column 3 of EA:

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ -3 \\ 7 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 7 \end{bmatrix} \qquad E(\text{column } j \text{ of } A) = \text{column } j \text{ of } EA.$$

This requirement deals with columns, while elimination is applied to rows. **The next section describes each entry of every product** AB. The beauty of matrix multiplication is that all three approaches (*rows, columns, whole matrices*) come out right.

The Matrix P_{ij} for a Row Exchange

To subtract row j from row i we use E_{ij}. To exchange or "permute" those rows we use another matrix P_{ij} (a **permutation matrix**). A row exchange is needed when zero is in the pivot position. Lower down, that pivot column may contain a nonzero. By exchanging the two rows, we have a pivot and elimination goes forward.

What matrix P_{23} exchanges row 2 with row 3? We can find it by exchanging rows of the identity matrix I:

$$\textbf{Permutation matrix} \qquad P_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

This is a ***row exchange matrix***. Multiplying by P_{23} exchanges components 2 and 3 of any column vector. Therefore it also exchanges rows 2 and 3 of any matrix:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{3} \\ \mathbf{5} \end{bmatrix} = \begin{bmatrix} 1 \\ \mathbf{5} \\ \mathbf{3} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 4 & 1 \\ \mathbf{0} & \mathbf{0} & \mathbf{3} \\ 0 & 6 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 1 \\ 0 & 6 & 5 \\ \mathbf{0} & \mathbf{0} & \mathbf{3} \end{bmatrix}.$$

On the right, P_{23} is doing what it was created for. With zero in the second pivot position and "6" below it, the exchange puts 6 into the pivot.

Matrices *act*. They don't just sit there. We will soon meet other permutation matrices, which can change the order of several rows. Rows $1, 2, 3$ can be moved to $3, 1, 2$. Our P_{23} is one particular permutation matrix—it exchanges rows 2 and 3.

Row Exchange Matrix P_{ij} is the identity matrix with rows i and j reversed. When this **"permutation matrix"** P_{ij} multiplies a matrix, it exchanges rows i and j.

$$\textit{\textbf{To exchange equations 1 and 3 multiply by}} \quad P_{13} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Usually row exchanges are not required. The odds are good that elimination uses only the E_{ij}. But the P_{ij} are ready if needed, to move a pivot up to the diagonal.

The Augmented Matrix

This book eventually goes far beyond elimination. Matrices have all kinds of practical applications, in which they are multiplied. Our best starting point was a square E times a square A, because we met this in elimination—and we know what answer to expect for EA. The next step is to allow a *rectangular matrix*. It still comes from our original equations, but now it includes the right side $\boldsymbol{b}$.

Key idea: Elimination does the same row operations to A and to $\boldsymbol{b}$. ***We can include b as an extra column and follow it through elimination.*** The matrix A is enlarged or "augmented" by the extra column $\boldsymbol{b}$:

$$\textit{\textbf{Augmented matrix}} \quad \begin{bmatrix} A & \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} 2 & 4 & -2 & \mathbf{2} \\ 4 & 9 & -3 & \mathbf{8} \\ -2 & -3 & 7 & \mathbf{10} \end{bmatrix}.$$

Elimination acts on whole rows of this matrix. The left side and right side are both multiplied by E, to subtract 2 times equation 1 from equation 2. With $\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix}$ those steps happen together:

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & -2 & \mathbf{2} \\ 4 & 9 & -3 & \mathbf{8} \\ -2 & -3 & 7 & \mathbf{10} \end{bmatrix} = \begin{bmatrix} 2 & 4 & -2 & \mathbf{2} \\ 0 & 1 & 1 & \mathbf{4} \\ -2 & -3 & 7 & \mathbf{10} \end{bmatrix}.$$

The new second row contains $0, 1, 1, 4$. The new second equation is $x_2 + x_3 = 4$. Matrix multiplication works by rows and at the same time by columns:

ROWS Each row of E acts on $[\, A \ \ \boldsymbol{b}\,]$ to give a row of $[\, EA \ \ E\boldsymbol{b}\,]$.

COLUMNS E acts on each column of $[\, A \ \ \boldsymbol{b}\,]$ to give a column of $[\, EA \ \ E\boldsymbol{b}\,]$.

Notice again that word "acts." This is essential. Matrices do something! The matrix A acts on $\boldsymbol{x}$ to produce $\boldsymbol{b}$. The matrix E operates on A to give EA. The whole process of elimination is a sequence of row operations, alias matrix multiplications. A goes to $E_{21}A$ which goes to $E_{31}E_{21}A$. Finally $E_{32}E_{31}E_{21}A$ is a triangular matrix.

The right side is included in the augmented matrix. The end result is a triangular system of equations. We stop for exercises on multiplication by E, before writing down the rules for all matrix multiplications (including block multiplication).

■ REVIEW OF THE KEY IDEAS ■

1. $Ax = x_1$ times column $1 + \cdots + x_n$ times column n. And $(Ax)_i = \sum_{j=1}^{n} a_{ij}x_j$.
2. Identity matrix $= I$, elimination matrix $= E_{ij}$ using ℓ_{ij}, exchange matrix $= P_{ij}$.
3. Multiplying $Ax = b$ by E_{21} subtracts a multiple ℓ_{21} of equation 1 from equation 2. The number $-\ell_{21}$ is the $(2, 1)$ entry of the elimination matrix E_{21}.
4. For the augmented matrix $\begin{bmatrix} A & b \end{bmatrix}$, that elimination step gives $\begin{bmatrix} E_{21}A & E_{21}b \end{bmatrix}$.
5. When A multiplies any matrix B, it multiplies each column of B separately.

■ WORKED EXAMPLES ■

2.3 A What 3 by 3 matrix E_{21} subtracts 4 times row 1 from row 2? What matrix P_{32} exchanges row 2 and row 3? If you multiply A on the *right* instead of the left, describe the results AE_{21} and AP_{32}.

Solution By doing those operations on the identity matrix I, we find

$$E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad P_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Multiplying by E_{21} on the right side will subtract 4 times **column 2** from **column 1**. Multiplying by P_{32} on the right will exchange **columns** 2 and 3.

2.3 B Write down the augmented matrix $[A \quad b]$ with an extra column:

$$\begin{aligned} x + 2y + 2z &= 1 \\ 4x + 8y + 9z &= 3 \\ 3y + 2z &= 1 \end{aligned}$$

Apply E_{21} and then P_{32} to reach a triangular system. Solve by back substitution. What combined matrix $P_{32}\,E_{21}$ will do both steps at once?

Solution E_{21} removes the 4 in column 1. But zero appears in column 2:

$$[A \quad b] = \begin{bmatrix} 1 & 2 & 2 & 1 \\ \mathbf{4} & 8 & 9 & 3 \\ 0 & 3 & 2 & 1 \end{bmatrix} \quad \text{and} \quad E_{21}[A \quad b] = \begin{bmatrix} 1 & 2 & 2 & 1 \\ \mathbf{0} & \mathbf{0} & 1 & -1 \\ 0 & 3 & 2 & 1 \end{bmatrix}$$

Now P_{32} exchanges rows 2 and 3. Back substitution produces z then y and x.

$$P_{32}\,E_{21}[A \quad b] = \begin{bmatrix} 1 & 2 & 2 & 1 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

For the matrix $P_{32}\,E_{21}$ that does both steps at once, *apply* P_{32} *to* E_{21}.

One matrix
Both steps

$$P_{32}\,E_{21} = \text{exchange the rows of } E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ -4 & 1 & 0 \end{bmatrix}.$$

2.3 C Multiply these matrices in two ways. First, rows of A times columns of B. Second, ***columns of A times rows of B***. That unusual way produces two matrices that add to AB. How many separate ordinary multiplications are needed?

Both ways

$$AB = \begin{bmatrix} 3 & 4 \\ 1 & 5 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{10} & \mathbf{16} \\ \mathbf{7} & \mathbf{9} \\ \mathbf{4} & \mathbf{8} \end{bmatrix}$$

Solution Rows of A times columns of B are dot products of vectors:

$$(\text{row } 1) \cdot (\text{column } 1) = \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \mathbf{10} \qquad \text{is the } (1,1) \text{ entry of } AB$$

$$(\text{row } 2) \cdot (\text{column } 1) = \begin{bmatrix} 1 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \mathbf{7} \qquad \text{is the } (2,1) \text{ entry of } AB$$

We need 6 dot products, 2 multiplications each, 12 in all $(3 \cdot 2 \cdot 2)$. The same AB comes from *columns of A times rows of B*. A column times a row is a matrix.

$$AB = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} \begin{bmatrix} 2 & 4 \end{bmatrix} + \begin{bmatrix} 4 \\ 5 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{6} & \mathbf{12} \\ \mathbf{2} & \mathbf{4} \\ \mathbf{4} & \mathbf{8} \end{bmatrix} + \begin{bmatrix} \mathbf{4} & \mathbf{4} \\ \mathbf{5} & \mathbf{5} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

Problem Set 2.3

Problems 1–15 are about elimination matrices.

1 Write down the 3 by 3 matrices that produce these elimination steps:

(a) E_{21} subtracts 5 times row 1 from row 2.

(b) E_{32} subtracts -7 times row 2 from row 3.

(c) P exchanges rows 1 and 2, then rows 2 and 3.

2 In Problem 1, applying E_{21} and then E_{32} to $\boldsymbol{b} = (1,0,0)$ gives $E_{32}E_{21}\boldsymbol{b} =$ ____. Applying E_{32} before E_{21} gives $E_{21}E_{32}\boldsymbol{b} =$ ____. When E_{32} comes first, row ____ feels no effect from row ____.

3 Which three matrices E_{21}, E_{31}, E_{32} put A into triangular form U?

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 4 & 6 & 1 \\ -2 & 2 & 0 \end{bmatrix} \quad \text{and} \quad E_{32}E_{31}E_{21}A = U.$$

Multiply those E's to get one matrix M that does elimination: $MA = U$.

4 Include $\boldsymbol{b} = (1,0,0)$ as a fourth column in Problem 3 to produce $[\,A \ \ \boldsymbol{b}\,]$. Carry out the elimination steps on this augmented matrix to solve $A\boldsymbol{x} = \boldsymbol{b}$.

5 Suppose $a_{33} = 7$ and the third pivot is 5. If you change a_{33} to 11, the third pivot is ____. If you change a_{33} to ____, there is no third pivot.

6 If every column of A is a multiple of $(1,1,1)$, then $A\boldsymbol{x}$ is always a multiple of $(1,1,1)$. Do a 3 by 3 example. How many pivots are produced by elimination?

7 Suppose E subtracts 7 times row 1 from row 3.

(a) To *invert* that step you should ____ 7 times row ____ to row ____.

(b) What "inverse matrix" E^{-1} takes that reverse step (so $E^{-1}E = I$)?

(c) If the reverse step is applied first (and then E) show that $EE^{-1} = I$.

8 The ***determinant*** of $M = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ is det $M = ad - bc$. Subtract ℓ times row 1 from row 2 to produce a new M^*. Show that $\det M^* = \det M$ for every ℓ. When $\ell = c/a$, *the product of pivots equals the determinant*: $(a)(d - \ell b)$ equals $ad - bc$.

9 (a) E_{21} subtracts row 1 from row 2 and then P_{23} exchanges rows 2 and 3. What matrix $M = P_{23}E_{21}$ does both steps at once?

(b) P_{23} exchanges rows 2 and 3 and then E_{31} subtracts row 1 from row 3. What matrix $M = E_{31}P_{23}$ does both steps at once? Explain why the M's are the same but the E's are different.

10 (a) What 3 by 3 matrix E_{13} will add row 3 to row 1?

(b) What matrix adds row 1 to row 3 and *at the same time* row 3 to row 1?

(c) What matrix adds row 1 to row 3 and *then* adds row 3 to row 1?

11 Create a matrix that has $a_{11} = a_{22} = a_{33} = 1$ but elimination produces two negative pivots without row exchanges. (The first pivot is 1.)

12 Multiply these matrices:

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 1 \\ 1 & 4 & 0 \end{bmatrix}.$$

13 Explain these facts. If the third column of B is all zero, the third column of EB is all zero (for any E). If the third *row* of B is all zero, the third row of EB might *not* be zero.

14 This 4 by 4 matrix will need elimination matrices E_{21} and E_{32} and E_{43}. What are those matrices?

$$A = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}.$$

15 Write down the 3 by 3 matrix that has $a_{ij} = 2i - 3j$. This matrix has $a_{32} = 0$, but elimination still needs E_{32} to produce a zero in the 3, 2 position. Which previous step destroys the original zero and what is E_{32}?

Problems 16–23 are about creating and multiplying matrices.

16 Write these ancient problems in a 2 by 2 matrix form $A\boldsymbol{x} = \boldsymbol{b}$ and solve them:

(a) X is twice as old as Y and their ages add to 33.

(b) $(x, y) = (2, 5)$ and $(3, 7)$ lie on the line $y = mx + c$. Find m and c.

17 The parabola $y = a + bx + cx^2$ goes through the points $(x, y) = (1, 4)$ and $(2, 8)$ and $(3, 14)$. Find and solve a matrix equation for the unknowns (a, b, c).

18 Multiply these matrices in the orders EF and FE:

$$E = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & 0 & 1 \end{bmatrix} \qquad F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{bmatrix}.$$

Also compute $E^2 = EE$ and $F^3 = FFF$. You can guess F^{100}.

19 Multiply these row exchange matrices in the orders PQ and QP and P^2:

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad Q = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Find another non-diagonal matrix whose square is $M^2 = I$.

20 (a) Suppose all columns of B are the same. Then all columns of EB are the same, because each one is E times _____.

(b) Suppose all rows of B are $[\,1 \;\; 2 \;\; 4\,]$. Show by example that all rows of EB are *not* $[\,1 \;\; 2 \;\; 4\,]$. It is true that those rows are _____.

21 If E adds row 1 to row 2 and F adds row 2 to row 1, does EF equal FE?

22 The entries of A and $\boldsymbol{x}$ are a_{ij} and x_j. So the first component of $A\boldsymbol{x}$ is $\sum a_{1j}x_j = a_{11}x_1 + \cdots + a_{1n}x_n$. If E_{21} subtracts row 1 from row 2, write a formula for

(a) the third component of $A\boldsymbol{x}$

(b) the $(2,1)$ entry of $E_{21}A$

(c) the $(2,1)$ entry of $E_{21}(E_{21}A)$

(d) the first component of $E_{21}A\boldsymbol{x}$.

23 The elimination matrix $E = \left[\begin{smallmatrix} 1 & 0 \\ -2 & 1 \end{smallmatrix}\right]$ subtracts 2 times row 1 of A from row 2 of A. The result is EA. What is the effect of $E(EA)$? In the opposite order AE, we are subtracting 2 times _____ of A from _____. (Do examples.)

Problems 24–27 include the column $\boldsymbol{b}$ in the augmented matrix $[\,A \;\; \boldsymbol{b}\,]$.

24 Apply elimination to the 2 by 3 augmented matrix $[\,A \;\; \boldsymbol{b}\,]$. What is the triangular system $U\boldsymbol{x} = \boldsymbol{c}$? What is the solution $\boldsymbol{x}$?

$$A\boldsymbol{x} = \begin{bmatrix} 2 & 3 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 17 \end{bmatrix}.$$

25 Apply elimination to the 3 by 4 augmented matrix $[\,A \;\; \boldsymbol{b}\,]$. How do you know this system has no solution? Change the last number 6 so there *is* a solution.

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 5 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix}.$$

26 The equations $A\boldsymbol{x} = \boldsymbol{b}$ and $A\boldsymbol{x}^* = \boldsymbol{b}^*$ have the same matrix A. What double augmented matrix should you use in elimination to solve both equations at once?

Solve both of these equations by working on a 2 by 4 matrix:

$$\begin{bmatrix} 1 & 4 \\ 2 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 4 \\ 2 & 7 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

27 Choose the numbers a, b, c, d in this augmented matrix so that there is (a) no solution (b) infinitely many solutions.

$$\begin{bmatrix} A & b \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & a \\ 0 & 4 & 5 & b \\ 0 & 0 & d & c \end{bmatrix}$$

Which of the numbers a, b, c, or d have no effect on the solvability?

28 If $AB = I$ and $BC = I$ use the associative law to prove $A = C$.

Challenge Problems

29 Find the triangular matrix E that reduces "*Pascal's matrix*" to a smaller Pascal:

Eliminate column 1 $$E \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 1 & 3 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 2 & 1 \end{bmatrix}.$$

Which matrix M (multiplying several E's) reduces Pascal all the way to I? Pascal's triangular matrix is exceptional, all of its multipliers are $\ell_{ij} = 1$.

30 Write $M = \left[\begin{smallmatrix} 3 & 4 \\ 5 & 7 \end{smallmatrix}\right]$ as a product of many factors $A = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$.

(a) What matrix E subtracts row 1 from row 2 to make row 2 of EM smaller?

(b) What matrix F subtracts row 2 of EM from row 1 to reduce row 1 of FEM?

(c) Continue E's and F's until (many E's and F's) times (M) is (A or B).

(d) E and F are the inverses of A and B! Moving all E's and F's to the right side will give you the desired result M = ***product of A's and B's.***
This is possible for integer matrices $M = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] > 0$ that have $ad - bc = 1$.

31 Find elimination matrices E_{21} then E_{32} then E_{43} to change K into U:

$$E_{43}\,E_{32}\,E_{21} \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 & 0 \\ 0 & 3/2 & -1 & 0 \\ 0 & 0 & 4/3 & -1 \\ 0 & 0 & 0 & 5/4 \end{bmatrix}.$$

Apply those three steps to the identity matrix I, to multiply $E_{43}E_{32}E_{21}$.

2.4 Rules for Matrix Operations

I will start with basic facts. A matrix is a rectangular array of numbers or "entries". When A has m rows and n columns, it is an "m by n" matrix. Matrices can be added if their shapes are the same. They can be multiplied by any constant c. Here are examples of $A + B$ and $2A$, for 3 by 2 matrices:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 2 & 2 \\ 4 & 4 \\ 9 & 9 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 7 & 8 \\ 9 & 9 \end{bmatrix} \quad \text{and} \quad 2\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 8 \\ 0 & 0 \end{bmatrix}.$$

Matrices are added exactly as vectors are—one entry at a time. We could even regard a column vector as a matrix with only one column (so $n = 1$). The matrix $-A$ comes from multiplication by $c = -1$ (reversing all the signs). Adding A to $-A$ leaves the *zero matrix*, with all entries zero. All this is only common sense.

The entry in row i and column j is called a_{ij} or $A(i, j)$. The n entries along the first row are $a_{11}, a_{12}, \ldots, a_{1n}$. The lower left entry in the matrix is a_{m1} and the lower right is a_{mn}. The row number i goes from 1 to m. The column number j goes from 1 to n.

Matrix addition is easy. The serious question is ***matrix multiplication.*** When can we multiply A times B, and what is the product AB? We cannot multiply when A and B are 3 by 2. They don't pass the following test:

To multiply AB: If A has n columns, B must have n rows.

When A is 3 by 2, the matrix B can be 2 by 1 (a vector) or 2 by 2 (square) or 2 by 20. ***Every column of B is multiplied by A.*** I will begin matrix multiplication the *dot product way*, and then return to this *column way*: A times columns of B. The most important rule is that ***AB times C equals A times BC.*** A Challenge Problem will prove this.

Suppose A is m by n and B is n by p. We can multiply. The product AB is m by p.

$$(m \times n)(n \times p) = (m \times p) \qquad \begin{bmatrix} \mathbf{m\ rows} \\ n\ columns \end{bmatrix} \begin{bmatrix} n\ rows \\ \mathbf{p\ columns} \end{bmatrix} = \begin{bmatrix} \mathbf{m\ rows} \\ \mathbf{p\ columns} \end{bmatrix}.$$

A row times a column is an extreme case. Then 1 by n multiplies n by 1. The result is 1 by 1. That single number is the "dot product".

In every case AB is filled with dot products. For the top corner, the $(1, 1)$ entry of AB is (row 1 of A) $\cdot$ (column 1 of B). To multiply matrices, take the dot product of ***each row of A with each column of B.***

The entry in row i and column j of AB is (row i of A) $\cdot$ (column j of B).

Figure 2.8 picks out the second row ($i = 2$) of a 4 by 5 matrix A. It picks out the third column ($j = 3$) of a 5 by 6 matrix B. Their dot product goes into row 2 and column 3 of AB. The matrix AB has *as many rows as A* (4 rows), and *as many columns as B*.

$$\underset{A \text{ is 4 by 5}}{\begin{bmatrix} * & & & \\ a_{i1} & a_{i2} & \cdots & a_{i5} \\ * & & & \\ * & & & \end{bmatrix}} \underset{B \text{ is 5 by 6}}{\begin{bmatrix} * & * & b_{1j} & * & * & * \\ & & b_{2j} & & & \\ & & \vdots & & & \\ & & b_{5j} & & & \end{bmatrix}} = \underset{AB \text{ is 4 by 6}}{\begin{bmatrix} & & * & & & \\ * & * & (AB)_{ij} & * & * & * \\ & & * & & & \\ & & * & & & \end{bmatrix}}$$

Figure 2.8: Here $i = 2$ and $j = 3$. Then $(AB)_{23}$ is (row 2) $\cdot$ (column 3) $= \Sigma a_{2k} b_{k3}$.

Example 1 Square matrices can be multiplied if and only if they have the same size:

$$\begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 5 & 6 \\ 1 & 0 \end{bmatrix}.$$

The first dot product is $1 \cdot 2 + 1 \cdot 3 = 5$. Three more dot products give 6, 1, and 0. Each dot product requires two multiplications—thus eight in all.

If A and B are n by n, so is AB. It contains n^2 dot products, row of A times column of B. Each dot product needs n multiplications, so ***the computation of AB uses n^3 separate multiplications***. For $n = 100$ we multiply a million times. For $n = 2$ we have $n^3 = 8$.

Mathematicians thought until recently that AB absolutely needed $2^3 = 8$ multiplications. Then somebody found a way to do it with 7 (and extra additions). By breaking n by n matrices into 2 by 2 blocks, this idea also reduced the count for large matrices. Instead of n^3 it went below $n^{2.8}$, and the exponent keeps falling.[1] The best at this moment is $n^{2.376}$. But the algorithm is so awkward that scientific computing is done the regular way: n^2 dot products in AB, and n multiplications for each one.

Example 2 Suppose A is a row vector (1 by 3) and B is a column vector (3 by 1). Then AB is 1 by 1 (only one entry, the dot product). On the other hand B times A (***a column times a row***) is a full 3 by 3 matrix. This multiplication is allowed!

Column times row
$(\boldsymbol{n \times 1})(\boldsymbol{1 \times n}) = (\boldsymbol{n \times n})$

$$\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix}.$$

A row times a column is an *"inner"* product—that is another name for dot product. A column times a row is an *"outer"* product. These are extreme cases of matrix multiplication.

Rows and Columns of AB

In the big picture, A multiplies each column of B. The result is a column of AB. In that column, we are combining the columns of A. ***Each column of AB is a combination of***

[1] Maybe 2.376 will drop to 2. No other number looks special, but no change for 10 years.

the columns of $\boldsymbol{A}$. That is the column picture of matrix multiplication:

$$\textbf{Matrix } A \textbf{ times column of } B \quad A\begin{bmatrix} \boldsymbol{b}_1 \cdots \boldsymbol{b}_p \end{bmatrix} = \begin{bmatrix} A\boldsymbol{b}_1 \cdots A\boldsymbol{b}_p \end{bmatrix}.$$

The row picture is reversed. Each row of A multiplies the whole matrix B. The result is a row of AB. It is a combination of the rows of B:

$$\textbf{Row times matrix} \qquad \begin{bmatrix} \text{row } i \text{ of } A \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} \text{row } i \text{ of } AB \end{bmatrix}.$$

We see row operations in elimination (E times A). We see columns in A times $\boldsymbol{x}$. The "row-column picture" has the dot products of rows with columns. Believe it or not, ***there is also a column-row picture.*** Not everybody knows that columns $1, \ldots, n$ of A multiply rows $1, \ldots, n$ of B and add up to the same answer AB. Worked Example **2.3 C** had numbers for $n = 2$. ***Example 3 will show how to multiply*** AB ***using columns times rows.***

The Laws for Matrix Operations

May I put on record six laws that matrices do obey, while emphasizing an equation they don't obey? The matrices can be square or rectangular, and the laws involving $A + B$ are all simple and all obeyed. Here are three addition laws:

$$\begin{array}{rl} A + B = B + A & \text{(commutative law)} \\ c(A + B) = cA + cB & \text{(distributive law)} \\ A + (B + C) = (A + B) + C & \text{(associative law).} \end{array}$$

Three more laws hold for multiplication, but $AB = BA$ is not one of them:

$$\begin{array}{rl} AB \neq BA & \text{(the commutative "law" is } \textit{usually broken}\text{)} \\ C(A + B) = CA + CB & \text{(distributive law from the left)} \\ (A + B)C = AC + BC & \text{(distributive law from the right)} \\ A(BC) = (AB)C & \text{(associative law for } ABC\text{) } (\textbf{\textit{parentheses not needed}}). \end{array}$$

When A and B are not square, AB is a different size from BA. These matrices can't be equal—even if both multiplications are allowed. For square matrices, almost any example shows that AB is different from BA:

$$AB = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{1} \end{bmatrix} \quad \text{but} \quad BA = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{1} & 0 \\ 0 & 0 \end{bmatrix}.$$

It is true that $AI = IA$. All square matrices commute with I and also with cI. Only these matrices cI commute with all other matrices.

The law $A(B + C) = AB + AC$ is proved a column at a time. Start with $A(\boldsymbol{b} + \boldsymbol{c}) = A\boldsymbol{b} + A\boldsymbol{c}$ for the first column. That is the key to everything—***linearity.*** Say no more.

The law $A(BC) = (AB)C$ ***means that you can multiply*** BC ***first or else*** AB ***first.*** The direct proof is sort of awkward (Problem 37) but this law is extremely useful. We highlighted it above; it is the key to the way we multiply matrices.

Look at the special case when $A = B = C =$ square matrix. Then (A *times* A^2) is equal to (A^2 *times* A). The product in either order is A^3. The matrix powers A^p follow the same rules as numbers:

$$A^p = AAA\cdots A \text{ (p factors)} \qquad (A^p)(A^q) = A^{p+q} \qquad (A^p)^q = A^{pq}.$$

Those are the ordinary laws for exponents. A^3 times A^4 is A^7 (seven factors). A^3 to the fourth power is A^{12} (twelve A's). When p and q are zero or negative these rules still hold, provided A has a "-1 power"—which is the *inverse matrix* A^{-1}. Then $A^0 = I$ is the identity matrix (no factors).

For a number, a^{-1} is $1/a$. For a matrix, the inverse is written A^{-1}. (It is *never* I/A, except this is allowed in MATLAB.) Every number has an inverse except $a = 0$. To decide when A has an inverse is a central problem in linear algebra. Section 2.5 will start on the answer. This section is a Bill of Rights for matrices, to say when A and B can be multiplied and how.

Block Matrices and Block Multiplication

We have to say one more thing about matrices. They can be cut into ***blocks*** (which are smaller matrices). This often happens naturally. Here is a 4 by 6 matrix broken into blocks of size 2 by 2—in this example each block is just I:

4 by 6 matrix
2 by 2 blocks

$$A = \left[\begin{array}{cc|cc|cc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right] = \begin{bmatrix} I & I & I \\ I & I & I \end{bmatrix}.$$

If B is also 4 by 6 and the block sizes match, you can add $A + B$ *a block at a time.*

We have seen block matrices before. The right side vector $\boldsymbol{b}$ was placed next to A in the "augmented matrix". Then $[\,A \;\; \boldsymbol{b}\,]$ has two blocks of different sizes. Multiplying by an elimination matrix gave $[\,EA \;\; E\boldsymbol{b}\,]$. No problem to multiply blocks times blocks, when their shapes permit.

Block multiplication If the cuts between columns of A match the cuts between rows of B, then block multiplication of AB is allowed:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & \cdots \\ B_{21} & \cdots \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & \cdots \\ A_{21}B_{11} + A_{22}B_{21} & \cdots \end{bmatrix}. \qquad (1)$$

This equation is the same as if the blocks were numbers (which are 1 by 1 blocks). We are careful to keep A's in front of B's, because BA can be different.

Main point When matrices split into blocks, it is often simpler to see how they act. The block matrix of I's above is much clearer than the original 4 by 6 matrix A.

Example 3 **(Important special case)** Let the blocks of A be its n columns. Let the blocks of B be its n rows. Then block multiplication AB adds up ***columns times rows***:

Columns times rows
$$\begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix} \begin{bmatrix} - & b_1 & - \\ & \vdots & \\ - & b_n & - \end{bmatrix} = \begin{bmatrix} a_1 b_1 + \cdots + a_n b_n \end{bmatrix}. \tag{2}$$

This is another way to multiply matrices. Compare it with the usual rows times columns. Row 1 of A times column 1 of B gave the $(1,1)$ entry in AB. Now *column* 1 of A times *row* 1 of B gives a full matrix—not just a single number. Look at this example:

$$\begin{bmatrix} 1 & 4 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 3 & 2 \end{bmatrix} + \begin{bmatrix} 4 \\ 5 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix}$$

$$\begin{matrix} \text{Column 1 times row 1} \\ + \text{Column 2 times row 2} \end{matrix} = \begin{bmatrix} 3 & 2 \\ 3 & 2 \end{bmatrix} + \begin{bmatrix} 4 & 0 \\ 5 & 0 \end{bmatrix}. \tag{3}$$

We stop there so you can see columns multiplying rows. If a 2 by 1 matrix (a column) multiplies a 1 by 2 matrix (a row), the result is 2 by 2. That is what we found. Dot products are *inner* products and these are *outer* products. In the top left corner the answer is $3 + 4 = 7$. This agrees with the row-column dot product of $(1,4)$ with $(3,1)$.

Summary The usual way, rows times columns, gives four dot products (8 multiplications). The new way, columns times rows, gives two full matrices (the same 8 multiplications). The 8 multiplications, and the 4 additions, are just executed in a different order.

Example 4 **(Elimination by blocks)** Suppose the first column of A contains $1, 3, 4$. To change 3 and 4 to 0 and 0, multiply the pivot row by 3 and 4 and subtract. Those row operations are really multiplications by elimination matrices E_{21} and E_{31}:

One at a time
$$E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad E_{31} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}.$$

The "block idea" is to do both eliminations with one matrix E. That matrix clears out the whole first column of A below the pivot $a = 1$:

$$E = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix} \quad \text{multiplies} \quad \begin{bmatrix} \mathbf{1} & x & x \\ \mathbf{3} & x & x \\ \mathbf{4} & x & x \end{bmatrix} \quad \text{to give} \quad EA = \begin{bmatrix} \mathbf{1} & x & x \\ \mathbf{0} & x & x \\ \mathbf{0} & x & x \end{bmatrix}.$$

Using inverses from 2.5, a block matrix E can do elimination on a whole (block) column of A. Suppose A has four blocks A, B, C, D. Watch how E multiplies A by blocks:

Block elimination
$$\left[\begin{array}{c|c} I & \mathbf{0} \\ \hline -CA^{-1} & I \end{array}\right] \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = \left[\begin{array}{c|c} A & B \\ \hline \mathbf{0} & D - CA^{-1}B \end{array}\right]. \tag{4}$$

Elimination multiplies the first row $[\,A \quad B\,]$ by CA^{-1} (previously c/a). It subtracts from C to get a zero block in the first column. It subtracts from D to get $S = D - CA^{-1}B$.

This is ordinary elimination, a column at a time—written in blocks. That final block S is $D - CA^{-1}B$, just like $d - cb/a$. This is called the ***Schur complement.***

■ REVIEW OF THE KEY IDEAS ■

1. The (i, j) entry of AB is (row i of A) $\cdot$ (column j of B).
2. An m by n matrix times an n by p matrix uses mnp separate multiplications.
3. A times BC equals AB times C (surprisingly important).
4. AB is also the sum of these matrices: (column j of A) times (row j of B).
5. Block multiplication is allowed when the block shapes match correctly.
6. Block elimination produces the *Schur complement* $D - CA^{-1}B$.

■ WORKED EXAMPLES ■

2.4 A Put yourself in the position of the author! I want to show you matrix multiplications that are *special*, but mostly I am stuck with small matrices. There is one terrific family of **Pascal matrices**, and they come in all sizes, and above all they have real meaning. I think 4 by 4 is a good size to show some of their amazing patterns.

Here is the lower triangular Pascal matrix L. Its entries come from *"Pascal's triangle"*. I will multiply L times the **ones** vector, and the **powers** vector:

$$\begin{array}{c}\textbf{Pascal}\\\textbf{matrix}\end{array} \begin{bmatrix}1&&&\\1&1&&\\1&2&1&\\1&3&3&1\end{bmatrix}\begin{bmatrix}\mathbf{1}\\\mathbf{1}\\\mathbf{1}\\\mathbf{1}\end{bmatrix}=\begin{bmatrix}\mathbf{1}\\\mathbf{2}\\\mathbf{4}\\\mathbf{8}\end{bmatrix} \qquad \begin{bmatrix}1&&&\\1&1&&\\1&2&1&\\1&3&3&1\end{bmatrix}\begin{bmatrix}\mathbf{1}\\\mathbf{x}\\\mathbf{x}^2\\\mathbf{x}^3\end{bmatrix}=\begin{bmatrix}\mathbf{1}\\\mathbf{1+x}\\\mathbf{(1+x)^2}\\\mathbf{(1+x)^3}\end{bmatrix}.$$

Each row of L leads to the next row: *Add an entry to the one on its left to get the entry below.* In symbols $\ell_{i\,j} + \ell_{i\,j-1} = \ell_{i+1\,j}$. The numbers after 1, 3, 3, 1 would be 1, 4, 6, 4, 1. Pascal lived in the 1600's, long before matrices, but his triangle fits perfectly into L.

Multiplying by **ones** is the same as adding up each row, to get powers of 2. By writing out L times **powers of x**, you see the entries of L as the "binomial coefficients" that are so essential to gamblers:

$$\mathbf{1} + \mathbf{2}x + \mathbf{1}x^2 = (1+x)^2 \qquad \mathbf{1} + \mathbf{3}x + \mathbf{3}x^2 + \mathbf{1}x^3 = (1+x)^3$$

The number "3" counts the ways to get Heads once and Tails twice in three coin flips: HTT and THT and TTH. The other "3" counts the ways to get Heads twice: HHT and

HTH and THH. Those are examples of "i choose j" = the number of ways to get j heads in i coin flips. That number is exactly ℓ_{ij}, if we start counting rows and columns of L at $i = 0$ and $j = 0$ (and remember $0! = 1$):

$$\ell_{ij} = \binom{i}{j} = i \text{ choose } j = \frac{i!}{j!\,(i-j)!} \qquad \binom{4}{2} = \frac{4!}{2!2!} = \frac{24}{(2)(2)} = 6$$

There are six ways to choose two aces out of four aces. We will see Pascal's triangle and these matrices again. Here are the questions I want to ask now:

1. What is $H = L^2$? This is the "hypercube matrix".
2. Multiply H times **ones** and **powers**.
3. The last row of H is $8, 12, 6, 1$. A cube has 8 corners, 12 edges, 6 faces, 1 box. *What would the next row of H tell about a hypercube in* **4D**?

Solution Multiply L times L to get the hypercube matrix $H = L^2$:

$$\begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ 1 & 2 & 1 & \\ 1 & 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ 1 & 2 & 1 & \\ 1 & 3 & 3 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{1} & & & \\ \mathbf{2} & \mathbf{1} & & \\ \mathbf{4} & \mathbf{4} & \mathbf{1} & \\ \mathbf{8} & \mathbf{12} & \mathbf{6} & \mathbf{1} \end{bmatrix} = H.$$

Now multiply H times the vectors of **ones** and **powers**:

$$\begin{bmatrix} 1 & & & \\ 2 & 1 & & \\ 4 & 4 & 1 & \\ 8 & 12 & 6 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{3} \\ \mathbf{9} \\ \mathbf{27} \end{bmatrix} \qquad \begin{bmatrix} 1 & & & \\ 2 & 1 & & \\ 4 & 4 & 1 & \\ 8 & 12 & 6 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \\ x^3 \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{2+x} \\ \mathbf{(2+x)^2} \\ \mathbf{(2+x)^3} \end{bmatrix}$$

If $x = 1$ we get the powers of 3. If $x = 0$ we get powers of 2. When L produces powers of $1 + x$, applying L again produces powers of $2 + x$.

How do the rows of H count corners and edges and faces of a cube? A square in **2D** has 4 corners, 4 edges, 1 face. Add one dimension at a time:

Connect two squares to get a **3D** *cube. Connect two cubes to get a* **4D** *hypercube.*

The cube has 8 corners and 12 edges: 4 edges in each square and 4 between the squares. The cube has 6 faces: 1 in each square and 4 faces between the squares. This row $8, 12, 6, 1$ will lead to the next row **16, 32, 24, 8, 1**. The rule is $2h_{i\,j} + h_{i\,j-1} = h_{i+1\,j}$.

Can you see this in four dimensions? The hypercube has 16 corners, no problem. It has 12 edges from one cube, 12 from the other cube, 8 that connect corners of those cubes: total 32 edges. It has 6 faces from each separate cube and 12 more from connecting pairs of edges: total $2 \times 6 + 12 = 24$ faces. It has one box from each cube and 6 more from connecting pairs of faces: total 8 boxes. And finally 1 hypercube.

2.4 B For these matrices, when does $AB = BA$? When does $BC = CB$? When does A times BC equal AB times C? Give the conditions on their entries p, q, r, z:

$$A = \begin{bmatrix} p & 0 \\ q & r \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad C = \begin{bmatrix} 0 & z \\ 0 & 0 \end{bmatrix}$$

If $p, q, r, 1, z$ are 4 by 4 blocks instead of numbers, do the answers change?

Solution First of all, A times BC *always* equals AB times C. Parentheses are not needed in $A(BC) = (AB)C = ABC$. But we must keep the matrices in this order:

$$\textbf{Usually } \boldsymbol{AB \neq BA} \qquad AB = \begin{bmatrix} p & p \\ q & q+r \end{bmatrix} \qquad BA = \begin{bmatrix} p+q & r \\ q & r \end{bmatrix}.$$

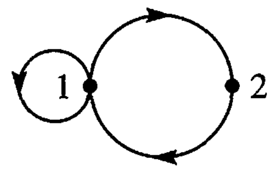
$$\textbf{By chance } \boldsymbol{BC = CB} \qquad BC = \begin{bmatrix} 0 & z \\ 0 & 0 \end{bmatrix} \qquad CB = \begin{bmatrix} 0 & z \\ 0 & 0 \end{bmatrix}.$$

B and C happen to commute. Part of the explanation is that the diagonal of B is I, which commutes with all 2 by 2 matrices. When p, q, r, z are 4 by 4 blocks and 1 changes to I, all these products remain correct. So the answers are the same.

2.4 C A **directed graph** starts with n nodes. The n by n **adjacency matrix** has $a_{ij} = 1$ when an edge leaves node i and enters node j; if no edge then $a_{ij} = 0$.

node 1 to node 2

node 1 to node 1



node 2 to node 1

$$A = \begin{bmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{bmatrix} = \text{adjacency matrix}$$

*The i, j entry of A^2 is $\sum a_{ik}a_{kj}$. **This is $a_{i1}a_{1j} + \cdots + a_{in}a_{nj}$**.* Why does that sum count the *two-step paths* from i to any node to j? The i, j entry of A^k counts k-step paths:

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^2 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \qquad \begin{matrix}\textbf{Count paths} \\ \textbf{with two edges}\end{matrix} \qquad \begin{bmatrix} \text{1 to 2 to 1, 1 to 1 to 1} & \text{1 to 1 to 2} \\ \text{2 to 1 to 1} & \text{2 to 1 to 2} \end{bmatrix}$$

List all of the 3-step paths between each pair of nodes and compare with A^3.

Solution The number $a_{ik}a_{kj}$ will be "1" if there is an edge from node i to k and an edge from k to j. This is a 2-step path. The number $a_{ik}a_{kj}$ will be "0" if either of those edges (i to k, k to j) is missing. So the sum of $a_{ik}a_{kj}$ is the number of 2-step paths leaving i and entering j. Matrix multiplication is just right for this count.

The 3-step paths are counted by A^3; we look at paths to node 2:

$$A^3 = \begin{bmatrix} 3 & \mathbf{2} \\ 2 & \mathbf{1} \end{bmatrix} \quad \text{counts the paths with three steps} \quad \begin{bmatrix} \cdots & 1 \text{ to } 1 \text{ to } 1 \text{ to } 2,\ 1 \text{ to } 2 \text{ to } 1 \text{ to } 2 \\ \cdots & 2 \text{ to } 1 \text{ to } 1 \text{ to } 2 \end{bmatrix}$$

These A^k contain the Fibonacci numbers $0, 1, 1, 2, 3, 5, 8, 13, \ldots$ coming in Section 6.2. Multiplying A by A^k involves Fibonacci's rule $F_{k+2} = F_{k+1} + F_k$ (as in $13 = 8 + 5$):

$$(A)(A^k) = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_{k+1} & F_k \\ F_k & F_{k-1} \end{bmatrix} = \begin{bmatrix} F_{k+2} & F_{k+1} \\ F_{k+1} & F_k \end{bmatrix} = A^{k+1}.$$

There are 13 six-step paths from node 1 to node 1, but I can't find them all.

A^k also counts words. A path like 1 to 1 to 2 to 1 corresponds to the word **aaba**. The letter **b** can't repeat because there is no edge from 2 to 2. The i, j entry of A^k counts the words of length $k + 1$ that start with the ith letter and end with the jth.

Problem Set 2.4

Problems 1–16 are about the laws of matrix multiplication.

1 A is 3 by 5, B is 5 by 3, C is 5 by 1, and D is 3 by 1. *All entries are* 1. Which of these matrix operations are allowed, and what are the results?

$$BA \qquad AB \qquad ABD \qquad DBA \qquad A(B + C).$$

2 What rows or columns or matrices do you multiply to find

(a) the third column of AB?

(b) the first row of AB?

(c) the entry in row 3, column 4 of AB?

(d) the entry in row 1, column 1 of CDE?

3 Add AB to AC and compare with $A(B + C)$:

$$A = \begin{bmatrix} 1 & 5 \\ 2 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 2 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 3 & 1 \\ 0 & 0 \end{bmatrix}.$$

4 In Problem 3, multiply A times BC. Then multiply AB times C.

5 Compute A^2 and A^3. Make a prediction for A^5 and A^n:

$$A = \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 2 & 2 \\ 0 & 0 \end{bmatrix}.$$

6 Show that $(A+B)^2$ is different from $A^2 + 2AB + B^2$, when

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 3 & 0 \end{bmatrix}.$$

Write down the correct rule for $(A+B)(A+B) = A^2 + \underline{\qquad} + B^2$.

7 True or false. Give a specific example when false:

(a) If columns 1 and 3 of B are the same, so are columns 1 and 3 of AB.

(b) If rows 1 and 3 of B are the same, so are rows 1 and 3 of AB.

(c) If rows 1 and 3 of A are the same, so are rows 1 and 3 of ABC.

(d) $(AB)^2 = A^2B^2$.

8 How is each row of DA and EA related to the rows of A, when

$$D = \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix} \quad \text{and} \quad E = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}?$$

How is each column of AD and AE related to the columns of A?

9 Row 1 of A is added to row 2. This gives EA below. Then column 1 of EA is added to column 2 to produce $(EA)F$:

$$EA = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ a+c & b+d \end{bmatrix}$$

$$\text{and} \quad (EA)F = (EA) \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a & a+b \\ a+c & a+c+b+d \end{bmatrix}.$$

(a) Do those steps in the opposite order. First add column 1 of A to column 2 by AF, then add row 1 of AF to row 2 by $E(AF)$.

(b) Compare with $(EA)F$. What law is obeyed by matrix multiplication?

10 Row 1 of A is again added to row 2 to produce EA. Then F adds row 2 of EA to row 1. The result is $F(EA)$:

$$F(EA) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ a+c & b+d \end{bmatrix} = \begin{bmatrix} 2a+c & 2b+d \\ a+c & b+d \end{bmatrix}.$$

(a) Do those steps in the opposite order: first add row 2 to row 1 by FA, then add row 1 of FA to row 2.

(b) What law is or is not obeyed by matrix multiplication?

11 (3 by 3 matrices) Choose the only B so that for every matrix A

(a) $BA = 4A$

(b) $BA = 4B$

(c) BA has rows 1 and 3 of A reversed and row 2 unchanged

(d) All rows of BA are the same as row 1 of A.

12 Suppose $AB = BA$ and $AC = CA$ for these two particular matrices B and C:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{commutes with} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Prove that $a = d$ and $b = c = 0$. Then A is a multiple of I. The only matrices that commute with B and C and all other 2 by 2 matrices are A = multiple of I.

13 Which of the following matrices are guaranteed to equal $(A - B)^2$: $A^2 - B^2$, $(B - A)^2$, $A^2 - 2AB + B^2$, $A(A - B) - B(A - B)$, $A^2 - AB - BA + B^2$?

14 True or false:

(a) If A^2 is defined then A is necessarily square.

(b) If AB and BA are defined then A and B are square.

(c) If AB and BA are defined then AB and BA are square.

(d) If $AB = B$ then $A = I$.

15 If A is m by n, how many separate multiplications are involved when

(a) A multiplies a vector $\boldsymbol{x}$ with n components?

(b) A multiplies an n by p matrix B?

(c) A multiplies itself to produce A^2? Here $m = n$.

16 For $A = \begin{bmatrix} 2 & -1 \\ 3 & -2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 & 4 \\ 1 & 0 & 6 \end{bmatrix}$, compute these answers *and nothing more*:

(a) column 2 of AB

(b) row 2 of AB

(c) row 2 of $AA = A^2$

(d) row 2 of $AAA = A^3$.

Problems 17–19 use a_{ij} for the entry in row i, column j of A.

17 Write down the 3 by 3 matrix A whose entries are

(a) a_{ij} = minimum of i and j

(b) $a_{ij} = (-1)^{i+j}$

(c) $a_{ij} = i/j$.

18 What words would you use to describe each of these classes of matrices? Give a 3 by 3 example in each class. Which matrix belongs to all four classes?

(a) $a_{ij} = 0$ if $i \neq j$

(b) $a_{ij} = 0$ if $i < j$

(c) $a_{ij} = a_{ji}$

(d) $a_{ij} = a_{1j}$.

19 The entries of A are a_{ij}. Assuming that zeros don't appear, what is

(a) the first pivot?

(b) the multiplier ℓ_{31} of row 1 to be subtracted from row 3?

(c) the new entry that replaces a_{32} after that subtraction?

(d) the second pivot?

Problems 20–24 involve powers of A.

20 Compute A^2, A^3, A^4 and also Av, A^2v, A^3v, A^4v for

$$A = \begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad v = \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix}.$$

21 Find all the powers $A^2, A^3, \ldots$ and $AB, (AB)^2, \ldots$ for

$$A = \begin{bmatrix} .5 & .5 \\ .5 & .5 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

22 By trial and error find real nonzero 2 by 2 matrices such that

$$A^2 = -I \qquad BC = 0 \qquad DE = -ED \text{ (not allowing } DE = 0\text{)}.$$

23 (a) Find a nonzero matrix A for which $A^2 = 0$.

(b) Find a matrix that has $A^2 \neq 0$ but $A^3 = 0$.

24 By experiment with $n = 2$ and $n = 3$ predict A^n for these matrices:

$$A_1 = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad A_3 = \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix}.$$

Problems 25–31 use column-row multiplication and block multiplication.

25 Multiply A times I using columns of A (3 by 3) times rows of I.

26 Multiply AB using columns times rows:

$$AB = \begin{bmatrix} 1 & 0 \\ 2 & 4 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 3 & 3 & 0 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \begin{bmatrix} 3 & 3 & 0 \end{bmatrix} + \underline{\quad} = \underline{\quad}.$$

27 Show that the product of upper triangular matrices is always upper triangular:

$$AB = \begin{bmatrix} x & x & x \\ 0 & x & x \\ 0 & 0 & x \end{bmatrix} \begin{bmatrix} x & x & x \\ 0 & x & x \\ 0 & 0 & x \end{bmatrix} = \begin{bmatrix} & & \\ 0 & & \\ 0 & 0 & \end{bmatrix}.$$

Proof using dot products (Row times column) (Row 2 of A) $\cdot$ (column 1 of B) $= 0$. Which other dot products give zeros?

Proof using full matrices (Column times row) Draw x's and 0's in (column 2 of A) times (row 2 of B). Also show (column 3 of A) times (row 3 of B).

28 Draw the cuts in A (2 by 3) and B (3 by 4) and AB to show how each of the four multiplication rules is really a block multiplication:

(1) Matrix A times columns of B. **Columns of AB**

(2) Rows of A times the matrix B. **Rows of AB**

(3) Rows of A times columns of B. **Inner products** (numbers in AB)

(4) Columns of A times rows of B. **Outer products** (matrices add to AB)

29 Which matrices E_{21} and E_{31} produce zeros in the $(2,1)$ and $(3,1)$ positions of $E_{21}A$ and $E_{31}A$?

$$A = \begin{bmatrix} 2 & 1 & 0 \\ -2 & 0 & 1 \\ 8 & 5 & 3 \end{bmatrix}.$$

Find the single matrix $E = E_{31}E_{21}$ that produces both zeros at once. Multiply EA.

30 Block multiplication says that column 1 is eliminated by

$$EA = \begin{bmatrix} 1 & \mathbf{0} \\ -c/a & I \end{bmatrix} \begin{bmatrix} a & \boldsymbol{b} \\ \boldsymbol{c} & D \end{bmatrix} = \begin{bmatrix} a & \boldsymbol{b} \\ \mathbf{0} & D - \boldsymbol{cb}/a \end{bmatrix}.$$

In Problem 29, what are $\boldsymbol{c}$ and D and what is $D - \boldsymbol{cb}/a$?

31 With $i^2 = -1$, the product of $(A + iB)$ and $(\boldsymbol{x} + i\boldsymbol{y})$ is $A\boldsymbol{x} + iB\boldsymbol{x} + iA\boldsymbol{y} - B\boldsymbol{y}$. Use blocks to separate the real part without i from the imaginary part that multiplies i:

$$\begin{bmatrix} A & -B \\ ? & ? \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} = \begin{bmatrix} A\boldsymbol{x} - B\boldsymbol{y} \\ ? \end{bmatrix} \begin{matrix} \text{real part} \\ \text{imaginary part} \end{matrix}$$

32 (*Very important*) Suppose you solve $A\boldsymbol{x} = \boldsymbol{b}$ for three special right sides $\boldsymbol{b}$:

$$A\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad A\boldsymbol{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad A\boldsymbol{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

If the three solutions $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ are the columns of a matrix X, what is A times X?

33 If the three solutions in Question 32 are $\boldsymbol{x}_1 = (1, 1, 1)$ and $\boldsymbol{x}_2 = (0, 1, 1)$ and $\boldsymbol{x}_3 = (0, 0, 1)$, solve $A\boldsymbol{x} = \boldsymbol{b}$ when $\boldsymbol{b} = (3, 5, 8)$. Challenge problem: What is A?

34 Find all matrices $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ that satisfy $A\left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right]A$.

35 Suppose a "circle graph" has 4 nodes connected (in both directions) by edges around a circle. What is its adjacency matrix from Worked Example **2.4 C**? What is A^2? Find all the 2-step paths (or 3-letter words) predicted by A^2.

Challenge Problems

36 **Practical question** Suppose A is m by n, B is n by p, and C is p by q. Then the multiplication count for $(AB)C$ is $mnp + mpq$. The same answer comes from A times BC with $mnq + npq$ separate multiplications. Notice npq for BC.

(a) If A is 2 by 4, B is 4 by 7, and C is 7 by 10, do you prefer $(AB)C$ or $A(BC)$?

(b) With N-component vectors, would you choose $(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v})\boldsymbol{w}^{\mathrm{T}}$ or $\boldsymbol{u}^{\mathrm{T}}(\boldsymbol{v}\boldsymbol{w}^{\mathrm{T}})$?

(c) Divide by $mnpq$ to show that $(AB)C$ is faster when $n^{-1} + q^{-1} < m^{-1} + p^{-1}$.

37 To prove that $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$, use the column vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ of B. First suppose that C has only one column $\boldsymbol{c}$ with entries $c_1, \ldots, c_n$:

AB has columns $A\boldsymbol{b}_1, \ldots, A\boldsymbol{b}_n$ and then $(AB)\boldsymbol{c}$ equals $c_1 A\boldsymbol{b}_1 + \cdots + c_n A\boldsymbol{b}_n$.

$B\boldsymbol{c}$ has one column $c_1\boldsymbol{b}_1 + \cdots + c_n\boldsymbol{b}_n$ and then $A(B\boldsymbol{c})$ equals $A(c_1\boldsymbol{b}_1 + \cdots + c_n\boldsymbol{b}_n)$.

Linearity gives equality of those two sums. This proves $(AB)\boldsymbol{c} = A(B\boldsymbol{c})$. The same is true for all other _____ of C. Therefore $(AB)C = A(BC)$. Apply to inverses:

If $BA = I$ and $AC = I$, prove that the left-inverse B equals the right-inverse C.

2.5 Inverse Matrices

Suppose A is a square matrix. We look for an "***inverse matrix***" A^{-1} of the same size, such that A^{-1} *times* A *equals* I. Whatever A does, A^{-1} undoes. Their product is the identity matrix—which does nothing to a vector, so $A^{-1}A\boldsymbol{x} = \boldsymbol{x}$. *But* A^{-1} *might not exist.*

What a matrix mostly does is to multiply a vector $\boldsymbol{x}$. Multiplying $A\boldsymbol{x} = \boldsymbol{b}$ by A^{-1} gives $A^{-1}A\boldsymbol{x} = A^{-1}\boldsymbol{b}$. ***This is*** $\boldsymbol{x} = A^{-1}\boldsymbol{b}$. The product $A^{-1}A$ is like multiplying by a number and then dividing by that number. A number has an inverse if it is not zero—matrices are more complicated and more interesting. The matrix A^{-1} is called "A inverse."

DEFINITION The matrix A is ***invertible*** if there exists a matrix A^{-1} such that

$$A^{-1}A = I \quad \text{and} \quad AA^{-1} = I. \tag{1}$$

Not all matrices have inverses. This is the first question we ask about a square matrix: Is A invertible? We don't mean that we immediately calculate A^{-1}. In most problems we never compute it! Here are six "notes" about A^{-1}.

Note 1 ***The inverse exists if and only if elimination produces n pivots*** (row exchanges are allowed). Elimination solves $A\boldsymbol{x} = \boldsymbol{b}$ without explicitly using the matrix A^{-1}.

Note 2 The matrix A cannot have two different inverses. Suppose $BA = I$ and also $AC = I$. Then $B = C$, according to this "proof by parentheses":

$$B(AC) = (BA)C \quad \text{gives} \quad BI = IC \quad \text{or} \quad B = C. \tag{2}$$

This shows that a *left-inverse* B (multiplying from the left) and a *right-inverse* C (multiplying A from the right to give $AC = I$) must be the *same matrix.*

Note 3 If A is invertible, the one and only solution to $A\boldsymbol{x} = \boldsymbol{b}$ is $\boldsymbol{x} = A^{-1}\boldsymbol{b}$:

$$\textit{\textbf{Multiply}} \quad A\boldsymbol{x} = \boldsymbol{b} \quad \textit{\textbf{by}} \quad A^{-1}. \quad \textit{\textbf{Then}} \quad \boldsymbol{x} = A^{-1}A\boldsymbol{x} = A^{-1}\boldsymbol{b}.$$

Note 4 (Important) ***Suppose there is a nonzero vector x such that*** $A\boldsymbol{x} = \boldsymbol{0}$. *Then A cannot have an inverse.* No matrix can bring $\boldsymbol{0}$ back to $\boldsymbol{x}$.

If A is invertible, then $A\boldsymbol{x} = \boldsymbol{0}$ can only have the zero solution $\boldsymbol{x} = A^{-1}\boldsymbol{0} = \boldsymbol{0}$.

Note 5 A 2 by 2 matrix is invertible if and only if $ad - bc$ is not zero:

$$\textbf{2 by 2 Inverse:} \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{3}$$

This number $ad - bc$ is the *determinant* of A. A matrix is invertible if its determinant is not zero (Chapter 5). The test for n pivots is usually decided before the determinant appears.

Note 6 A diagonal matrix has an inverse provided no diagonal entries are zero:

$$\text{If} \quad A = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix} \quad \text{then} \quad A^{-1} = \begin{bmatrix} 1/d_1 & & \\ & \ddots & \\ & & 1/d_n \end{bmatrix}.$$

Example 1 The 2 by 2 matrix $A = \left[\begin{smallmatrix} 1 & 2 \\ 1 & 2 \end{smallmatrix}\right]$ is not invertible. It fails the test in Note 5, because $ad - bc$ equals $2 - 2 = 0$. It fails the test in Note 3, because $Ax = \mathbf{0}$ when $x = (2, -1)$. It fails to have two pivots as required by Note 1.

Elimination turns the second row of this matrix A into a zero row.

The Inverse of a Product AB

For two nonzero numbers a and b, the sum $a + b$ might or might not be invertible. The numbers $a = 3$ and $b = -3$ have inverses $\frac{1}{3}$ and $-\frac{1}{3}$. Their sum $a + b = 0$ has no inverse. But the product $ab = -9$ does have an inverse, which is $\frac{1}{3}$ times $-\frac{1}{3}$.

For two matrices A and B, the situation is similar. It is hard to say much about the invertibility of $A + B$. But the *product* AB has an inverse, if and only if the two factors A and B are separately invertible (and the same size). The important point is that A^{-1} and B^{-1} come in ***reverse order***:

If A and B are invertible then so is AB. The inverse of a product AB is

$$(AB)^{-1} = B^{-1}A^{-1}. \tag{4}$$

To see why the order is reversed, multiply AB times $B^{-1}A^{-1}$. Inside that is $BB^{-1} = I$:

$$\textbf{Inverse of } AB \quad (AB)(B^{-1}A^{-1}) = AIA^{-1} = AA^{-1} = I.$$

We moved parentheses to multiply BB^{-1} first. Similarly $B^{-1}A^{-1}$ times AB equals I. This illustrates a basic rule of mathematics: Inverses come in reverse order. It is also common sense: If you put on socks and then shoes, the first to be taken off are the ______. The same reverse order applies to three or more matrices:

$$\textbf{Reverse order} \qquad (ABC)^{-1} = C^{-1}B^{-1}A^{-1}. \tag{5}$$

Example 2 ***Inverse of an elimination matrix.*** If E subtracts 5 times row 1 from row 2, then E^{-1} *adds* 5 times row 1 to row 2:

$$E = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Multiply EE^{-1} to get the identity matrix I. Also multiply $E^{-1}E$ to get I. We are adding and subtracting the same 5 times row 1. Whether we add and then subtract (this is EE^{-1}) or subtract and then add (this is $E^{-1}E$), we are back at the start.

For square matrices, an inverse on one side is automatically an inverse on the other side. If $AB = I$ then automatically $BA = I$. In that case B is A^{-1}. This is very useful to know but we are not ready to prove it.

Example 3 Suppose F subtracts 4 times row 2 from row 3, and F^{-1} adds it back:

$$F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -4 & 1 \end{bmatrix} \quad \text{and} \quad F^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix}.$$

Now multiply F by the matrix E in Example 2 to find FE. Also multiply E^{-1} times F^{-1} to find $(FE)^{-1}$. Notice the orders FE and $E^{-1}F^{-1}$!

$$FE = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 20 & -4 & 1 \end{bmatrix} \quad \text{is inverted by} \quad E^{-1}F^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix}. \tag{6}$$

The result is beautiful and correct. The product FE contains "20" but its inverse doesn't. E subtracts 5 times row 1 from row 2. Then F subtracts 4 times the *new* row 2 (changed by row 1) from row 3. ***In this order FE, row 3 feels an effect from row 1.***

In the order $E^{-1}F^{-1}$, that effect does not happen. First F^{-1} adds 4 times row 2 to row 3. After that, E^{-1} adds 5 times row 1 to row 2. There is no 20, because row 3 doesn't change again. ***In this order $E^{-1}F^{-1}$, row 3 feels no effect from row 1.***

In elimination order F follows E. In reverse order E^{-1} follows F^{-1}.
$E^{-1}F^{-1}$ is quick. The multipliers 5, 4 fall into place below the diagonal of 1's.

This special multiplication $E^{-1}F^{-1}$ and $E^{-1}F^{-1}G^{-1}$ will be useful in the next section. We will explain it again, more completely. In this section our job is A^{-1}, and we expect some serious work to compute it. Here is a way to organize that computation.

Calculating A^{-1} by Gauss-Jordan Elimination

I hinted that A^{-1} might not be explicitly needed. The equation $Ax = b$ is solved by $x = A^{-1}b$. But it is not necessary or efficient to compute A^{-1} and multiply it times b. *Elimination goes directly to x.* Elimination is also the way to calculate A^{-1}, as we now show. The Gauss-Jordan idea is to solve $AA^{-1} = I$, *finding each column of A^{-1}.*

A multiplies the first column of A^{-1} (call that x_1) to give the first column of I (call that e_1). This is our equation $Ax_1 = e_1 = (1, 0, 0)$. There will be two more equations. Each of the columns x_1, x_2, x_3 of A^{-1} is multiplied by A to produce a column of I:

3 columns of A^{-1} $$AA^{-1} = A\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} = \begin{bmatrix} e_1 & e_2 & e_3 \end{bmatrix} = I. \tag{7}$$

To invert a 3 by 3 matrix A, we have to solve three systems of equations: $Ax_1 = e_1$ and $Ax_2 = e_2 = (0, 1, 0)$ and $Ax_3 = e_3 = (0, 0, 1)$. Gauss-Jordan finds A^{-1} this way.

The ***Gauss-Jordan method*** computes A^{-1} by solving ***all n equations together.*** Usually the "augmented matrix" $[A \ \ \boldsymbol{b}]$ has one extra column $\boldsymbol{b}$. Now we have three right sides $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ (when A is 3 by 3). They are the columns of I, so the augmented matrix is really the block matrix $[\,A \ \ I\,]$. I take this chance to invert my favorite matrix K, with 2's on the main diagonal and -1's next to the 2's:

$$\begin{bmatrix} K & e_1 & e_2 & e_3 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{bmatrix} \quad \textbf{Start Gauss-Jordan on } \boldsymbol{K}$$

$$\to \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ \mathbf{0} & \frac{3}{2} & \mathbf{-1} & \frac{1}{2} & \mathbf{1} & \mathbf{0} \\ 0 & -1 & 2 & 0 & 0 & 1 \end{bmatrix} \quad (\tfrac{1}{2}\ \textit{row}\ \mathbf{1} + \textit{row}\ \mathbf{2})$$

$$\to \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & -1 & \frac{1}{2} & 1 & 0 \\ \mathbf{0} & \mathbf{0} & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & \mathbf{1} \end{bmatrix} \quad (\tfrac{2}{3}\ \textit{row}\ \mathbf{2} + \textit{row}\ \mathbf{3})$$

We are halfway to K^{-1}. The matrix in the first three columns is U (upper triangular). The pivots $2, \frac{3}{2}, \frac{4}{3}$ are on its diagonal. Gauss would finish by back substitution. The contribution of Jordan is *to continue with elimination*! He goes all the way to the ***"reduced echelon form"***. Rows are added to rows above them, to produce ***zeros above the pivots***:

$$\begin{pmatrix} \text{Zero above} \\ \text{third pivot} \end{pmatrix} \to \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ \mathbf{0} & \frac{3}{2} & \mathbf{0} & \frac{3}{4} & \frac{3}{2} & \frac{3}{4} \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{bmatrix} \quad (\tfrac{3}{4}\ \textit{row}\ \mathbf{3} + \textit{row}\ \mathbf{2})$$

$$\begin{pmatrix} \text{Zero above} \\ \text{second pivot} \end{pmatrix} \to \begin{bmatrix} \mathbf{2} & \mathbf{0} & \mathbf{0} & \frac{3}{2} & \mathbf{1} & \frac{1}{2} \\ 0 & \frac{3}{2} & 0 & \frac{3}{4} & \frac{3}{2} & \frac{3}{4} \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{bmatrix} \quad (\tfrac{2}{3}\ \textit{row}\ \mathbf{2} + \textit{row}\ \mathbf{1})$$

The last Gauss-Jordan step is to divide each row by its pivot. The new pivots are 1. We have reached I in the first half of the matrix, because K is invertible. ***The three columns of K^{-1} are in the second half of*** $[\,I \ \ K^{-1}\,]$:

$$\begin{matrix} \text{(divide by 2)} \\ \text{(divide by } \frac{3}{2}) \\ \text{(divide by } \frac{4}{3}) \end{matrix} \quad \begin{bmatrix} \mathbf{1} & 0 & 0 & \frac{3}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \mathbf{1} & 0 & \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & 0 & \mathbf{1} & \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{bmatrix} = \begin{bmatrix} I & x_1 & x_2 & x_3 \end{bmatrix} = \begin{bmatrix} I & K^{-1} \end{bmatrix}.$$

Starting from the 3 by 6 matrix $[\,K \ \ I\,]$, we ended with $[\,I \ \ K^{-1}\,]$. Here is the whole Gauss-Jordan process on one line for any invertible matrix A:

Gauss-Jordan ***Multiply*** $[\,A \ \ I\,]$ ***by*** A^{-1} ***to get*** $[\,I \ \ A^{-1}\,]$.

The elimination steps create the inverse matrix while changing A to I. For large matrices, we probably don't want A^{-1} at all. But for small matrices, it can be very worthwhile to know the inverse. We add three observations about this particular K^{-1} because it is an important example. We introduce the words *symmetric, tridiagonal,* and *determinant:*

1. K is ***symmetric*** across its main diagonal. So is K^{-1}.

2. K is ***tridiagonal*** (only three nonzero diagonals). But K^{-1} is a dense matrix with no zeros. That is another reason we don't often compute inverse matrices. The inverse of a band matrix is generally a dense matrix.

3. The *product of pivots* is $2(\frac{3}{2})(\frac{4}{3}) = 4$. This number 4 is the ***determinant*** of K.

K^{-1} involves division by the determinant
$$K^{-1} = \frac{1}{4}\begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}. \tag{8}$$

This is why an invertible matrix cannot have a zero determinant.

Example 4 Find A^{-1} by Gauss-Jordan elimination starting from $A = \left[\begin{smallmatrix} 2 & 3 \\ 4 & 7 \end{smallmatrix}\right]$. There are two row operations and then a division to put 1's in the pivots:

$$\begin{bmatrix} A & I \end{bmatrix} = \begin{bmatrix} \mathbf{2} & 3 & 1 & 0 \\ \mathbf{4} & 7 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 3 & 1 & 0 \\ \mathbf{0} & \mathbf{1} & \mathbf{-2} & \mathbf{1} \end{bmatrix} \quad (\text{this is } \begin{bmatrix} U & L^{-1} \end{bmatrix})$$

$$\rightarrow \begin{bmatrix} 2 & \mathbf{0} & \mathbf{7} & \mathbf{-3} \\ 0 & 1 & -2 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & \frac{7}{2} & -\frac{3}{2} \\ 0 & 1 & -2 & 1 \end{bmatrix} \quad (\text{this is } \begin{bmatrix} I & A^{-1} \end{bmatrix}).$$

That A^{-1} involves division by the determinant $ad - bc = 2 \cdot 7 - 3 \cdot 4 = 2$. The code for $X =$ inverse(A) can use **rref**, the "row reduced echelon form" from Chapter 3:

```
I = eye (n);                % Define the n by n identity matrix
R = rref ([A I]);           % Eliminate on the augmented matrix [A I]
X = R(:, n + 1 : n + n)     % Pick A^-1 from the last n columns of R
```

A must be invertible, or elimination cannot reduce it to I (in the left half of R).

Gauss-Jordan shows why A^{-1} is expensive. We must solve n equations for its n columns.

To solve $Ax = b$ without A^{-1}, we deal with *one* column b to find one column x.

In defense of A^{-1}, we want to say that its cost is not n times the cost of one system $Ax = b$. Surprisingly, the cost for n columns is only multiplied by 3. This saving is because the n equations $Ax_i = e_i$ all involve the same matrix A. Working with the right sides is relatively cheap, because elimination only has to be done once on A.

The complete A^{-1} needs n^3 elimination steps, where a single x needs $n^3/3$. The next section calculates these costs.

Singular versus Invertible

We come back to the central question. Which matrices have inverses? The start of this section proposed the pivot test: A^{-1} ***exists exactly when A has a full set of n pivots.*** (Row exchanges are allowed.) Now we can prove that by Gauss-Jordan elimination:

1. With n pivots, elimination solves all the equations $Ax_i = e_i$. The columns x_i go into A^{-1}. Then $AA^{-1} = I$ and A^{-1} is at least a ***right-inverse.***

2. Elimination is really a sequence of multiplications by E's and P's and D^{-1}:

Left-inverse $$(D^{-1} \cdots E \cdots P \cdots E)A = I. \tag{9}$$

D^{-1} divides by the pivots. The matrices E produce zeros below and above the pivots. P will exchange rows if needed (see Section 2.7). The product matrix in equation (9) is evidently a ***left-inverse.*** With n pivots we have reached $A^{-1}A = I$.

The right-inverse equals the left-inverse. That was Note 2 at the start of in this section. So a square matrix with a full set of pivots will always have a two-sided inverse.

Reasoning in reverse will now show that A ***must have n pivots if*** $AC = I$. (Then we deduce that C is also a left-inverse and $CA = I$.) Here is one route to those conclusions:

1. If A doesn't have n pivots, elimination will lead to a *zero row.*
2. Those elimination steps are taken by an invertible M. *So a row of MA is zero.*
3. If $AC = I$ had been possible, then $MAC = M$. The zero row of MA, times C, gives a zero row of M itself.
4. An invertible matrix M can't have a zero row! A *must* have n pivots if $AC = I$.

That argument took four steps, but the outcome is short and important.

Elimination gives a complete test for invertibility of a square matrix. A^{-1} ***exists (and Gauss-Jordan finds it) exactly when A has n pivots.*** The argument above shows more:

$$\text{If} \quad AC = I \quad \text{then} \quad CA = I \quad \text{and} \quad C = A^{-1}$$

Example 5 If L is lower triangular with 1's on the diagonal, so is L^{-1}.

A triangular matrix is invertible if and only if no diagonal entries are zero.

Here L has 1's so L^{-1} also has 1's. Use the Gauss-Jordan method to construct L^{-1}. Start by subtracting multiples of pivot rows from rows *below.* Normally this gets us halfway to the inverse, but for L it gets us all the way. L^{-1} appears on the right when I appears on the left. Notice how L^{-1} contains 11, from 3 times 5 minus 4.

Gauss-Jordan on triangular L

$$\begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & 1 & 0 & 0 \\ \mathbf{3} & \mathbf{1} & \mathbf{0} & 0 & 1 & 0 \\ \mathbf{4} & \mathbf{5} & \mathbf{1} & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} L & I \end{bmatrix}$$

$$\begin{matrix} \\ \rightarrow \\ \rightarrow \end{matrix} \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -3 & 1 & 0 \\ 0 & 5 & 1 & -4 & 0 & 1 \end{bmatrix} \qquad \begin{matrix} \text{(3 times row 1 from row 2)} \\ \text{(4 times row 1 from row 3)} \\ \text{(then 5 times row 2 from row 3)} \end{matrix}$$

$$\begin{matrix} \\ \\ \rightarrow \end{matrix} \begin{bmatrix} 1 & 0 & 0 & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ 0 & 1 & 0 & \mathbf{-3} & \mathbf{1} & \mathbf{0} \\ 0 & 0 & 1 & \mathbf{11} & \mathbf{-5} & \mathbf{1} \end{bmatrix} = \begin{bmatrix} I & \boldsymbol{L^{-1}} \end{bmatrix}.$$

L goes to I by a product of elimination matrices $E_{32}E_{31}E_{21}$. So that product is L^{-1}. All pivots are 1's (a full set). L^{-1} is lower triangular, with the strange entry "11".

That 11 does not appear to spoil 3, 4, 5 in the good order $E_{21}^{-1}E_{31}^{-1}E_{32}^{-1} = L$.

■ REVIEW OF THE KEY IDEAS ■

1. The inverse matrix gives $AA^{-1} = I$ and $A^{-1}A = I$.
2. A is invertible if and only if it has n pivots (row exchanges allowed).
3. If $A\boldsymbol{x} = \mathbf{0}$ for a nonzero vector $\boldsymbol{x}$, then A has no inverse.
4. The inverse of AB is the reverse product $B^{-1}A^{-1}$. And $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$.
5. The Gauss-Jordan method solves $AA^{-1} = I$ to find the n columns of A^{-1}. The augmented matrix $\begin{bmatrix} A & I \end{bmatrix}$ is row-reduced to $\begin{bmatrix} I & A^{-1} \end{bmatrix}$.

■ WORKED EXAMPLES ■

2.5 A The inverse of a triangular **difference matrix** A is a triangular **sum matrix** S:

$$\begin{bmatrix} A & I \end{bmatrix} = \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{array}\right] \rightarrow \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right] = \begin{bmatrix} I & A^{-1} \end{bmatrix} = \begin{bmatrix} I & \textit{sum matrix} \end{bmatrix}.$$

If I change a_{13} to -1, then all rows of A add to zero. The equation $A\boldsymbol{x} = \mathbf{0}$ will now have the nonzero solution $\boldsymbol{x} = (1, 1, 1)$. A clear signal: ***This new A can't be inverted.***

2.5 B Three of these matrices are invertible, and three are singular. Find the inverse when it exists. Give reasons for noninvertibility (zero determinant, too few pivots, nonzero solution to $Ax = \mathbf{0}$) for the other three. The matrices are in the order A, B, C, D, S, E:

$$\begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix} \quad \begin{bmatrix} 4 & 3 \\ 8 & 7 \end{bmatrix} \quad \begin{bmatrix} 6 & 6 \\ 6 & 0 \end{bmatrix} \quad \begin{bmatrix} 6 & 6 \\ 6 & 6 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Solution

$$B^{-1} = \frac{1}{4}\begin{bmatrix} 7 & -3 \\ -8 & 4 \end{bmatrix} \quad C^{-1} = \frac{1}{36}\begin{bmatrix} 0 & 6 \\ 6 & -6 \end{bmatrix} \quad S^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

A is not invertible because its determinant is $4 \cdot 6 - 3 \cdot 8 = 24 - 24 = 0$. D is not invertible because there is only one pivot; the second row becomes zero when the first row is subtracted. E is not invertible because a combination of the columns (the second column minus the first column) is zero—in other words $Ex = \mathbf{0}$ has the solution $x = (-1, 1, 0)$.

Of course all three reasons for noninvertibility would apply to each of A, D, E.

2.5 C Apply the Gauss-Jordan method to invert this triangular "Pascal matrix" L. You see **Pascal's triangle**—adding each entry to the entry on its left gives the entry below. The entries of L are "binomial coefficients". The next row would be $1, 4, 6, 4, 1$.

$$\textbf{Triangular Pascal matrix} \quad L = \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 \\ \mathbf{1} & \mathbf{1} & 0 & 0 \\ \mathbf{1} & \mathbf{2} & \mathbf{1} & 0 \\ \mathbf{1} & \mathbf{3} & \mathbf{3} & \mathbf{1} \end{bmatrix} = \texttt{abs(pascal (4,1))}$$

Solution Gauss-Jordan starts with $[\,L \;\; I\,]$ and produces zeros by subtracting row 1:

$$[L \;\; I] = \left[\begin{array}{cccc|cccc} \mathbf{1} & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \mathbf{1} & \mathbf{1} & 0 & 0 & 0 & 1 & 0 & 0 \\ \mathbf{1} & \mathbf{2} & \mathbf{1} & 0 & 0 & 0 & 1 & 0 \\ \mathbf{1} & \mathbf{3} & \mathbf{3} & \mathbf{1} & 0 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \mathbf{0} & 1 & 0 & 0 & \mathbf{-1} & 1 & 0 & 0 \\ \mathbf{0} & 2 & 1 & 0 & \mathbf{-1} & 0 & 1 & 0 \\ \mathbf{0} & 3 & 3 & 1 & \mathbf{-1} & 0 & 0 & 1 \end{array}\right].$$

The next stage creates zeros below the second pivot, using multipliers 2 and 3. Then the last stage subtracts 3 times the new row 3 from the new row 4:

$$\to \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & \mathbf{0} & 1 & 0 & \mathbf{1} & \mathbf{-2} & 1 & 0 \\ 0 & \mathbf{0} & 3 & 1 & \mathbf{2} & \mathbf{-3} & 0 & 1 \end{array}\right] \to \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \mathbf{-1} & \mathbf{1} & 0 & 0 \\ 0 & 0 & 1 & 0 & \mathbf{1} & \mathbf{-2} & \mathbf{1} & 0 \\ 0 & 0 & 0 & 1 & \mathbf{-1} & \mathbf{3} & \mathbf{-3} & \mathbf{1} \end{array}\right] = [I \;\; L^{-1}].$$

All the pivots were 1! So we didn't need to divide rows by pivots to get I. The inverse matrix L^{-1} looks like L itself, except odd-numbered diagonals have minus signs.

The same pattern continues to n by n Pascal matrices, L^{-1} has "alternating diagonals".

Problem Set 2.5

1 Find the inverses (directly or from the 2 by 2 formula) of A, B, C:

$$A = \begin{bmatrix} 0 & 3 \\ 4 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & 0 \\ 4 & 2 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 3 & 4 \\ 5 & 7 \end{bmatrix}.$$

2 For these "permutation matrices" find P^{-1} by trial and error (with 1's and 0's):

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

3 Solve for the first column (x, y) and second column (t, z) of A^{-1}:

$$\begin{bmatrix} 10 & 20 \\ 20 & 50 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 10 & 20 \\ 20 & 50 \end{bmatrix} \begin{bmatrix} t \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

4 Show that $\left[\begin{smallmatrix} 1 & 2 \\ 3 & 6 \end{smallmatrix}\right]$ is not invertible by trying to solve $AA^{-1} = I$ for column 1 of A^{-1}:

$$\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \left(\begin{array}{l} \textit{For a different } A, \text{ could column 1 of } A^{-1} \\ \text{be possible to find but not column 2?} \end{array}\right)$$

5 Find an upper triangular U (not diagonal) with $U^2 = I$ which gives $U = U^{-1}$.

6 (a) If A is invertible and $AB = AC$, prove quickly that $B = C$.

(b) If $A = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right]$, find two different matrices such that $AB = AC$.

7 (Important) If A has row 1 + row 2 = row 3, show that A is not invertible:

(a) Explain why $A\boldsymbol{x} = (1, 0, 0)$ cannot have a solution.

(b) Which right sides (b_1, b_2, b_3) might allow a solution to $A\boldsymbol{x} = \boldsymbol{b}$?

(c) What happens to row 3 in elimination?

8 If A has column 1 + column 2 = column 3, show that A is not invertible:

(a) Find a nonzero solution $\boldsymbol{x}$ to $A\boldsymbol{x} = \boldsymbol{0}$. The matrix is 3 by 3.

(b) Elimination keeps column 1 + column 2 = column 3. Explain why there is no third pivot.

9 Suppose A is invertible and you exchange its first two rows to reach B. Is the new matrix B invertible and how would you find B^{-1} from A^{-1}?

10 Find the inverses (in any legal way) of

$$A = \begin{bmatrix} 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 \\ 0 & 4 & 0 & 0 \\ 5 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 & 2 & 0 & 0 \\ 4 & 3 & 0 & 0 \\ 0 & 0 & 6 & 5 \\ 0 & 0 & 7 & 6 \end{bmatrix}.$$

11 (a) Find invertible matrices A and B such that $A + B$ is not invertible.

(b) Find singular matrices A and B such that $A + B$ is invertible.

12 If the product $C = AB$ is invertible (A and B are square), then A itself is invertible. Find a formula for A^{-1} that involves C^{-1} and B.

13 If the product $M = ABC$ of three square matrices is invertible, then B is invertible. (So are A and C.) Find a formula for B^{-1} that involves M^{-1} and A and C.

14 If you add row 1 of A to row 2 to get B, how do you find B^{-1} from A^{-1}?

$$\text{Notice the order.}\quad \text{The inverse of}\quad B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} & A & \end{bmatrix} \quad \text{is} \quad \underline{\qquad}.$$

15 Prove that a matrix with a column of zeros cannot have an inverse.

16 Multiply $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ times $\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. What is the inverse of each matrix if $ad \neq bc$?

17 (a) What 3 by 3 matrix E has the same effect as these three steps? Subtract row 1 from row 2, subtract row 1 from row 3, then subtract row 2 from row 3.

(b) What single matrix L has the same effect as these three reverse steps? Add row 2 to row 3, add row 1 to row 3, then add row 1 to row 2.

18 If B is the inverse of A^2, show that AB is the inverse of A.

19 Find the numbers a and b that give the inverse of 5 ∗ eye(4) − ones(4,4):

$$\begin{bmatrix} 4 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 \\ -1 & -1 & 4 & -1 \\ -1 & -1 & -1 & 4 \end{bmatrix}^{-1} = \begin{bmatrix} a & b & b & b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \end{bmatrix}.$$

What are a and b in the inverse of 6 ∗ eye(5) − ones(5,5)?

20 Show that A = 4 ∗ eye(4) − ones(4,4) is *not* invertible: Multiply A ∗ ones(4,1).

21 There are sixteen 2 by 2 matrices whose entries are 1's and 0's. How many of them are invertible?

Questions 22–28 are about the Gauss-Jordan method for calculating A^{-1}.

22 Change I into A^{-1} as you reduce A to I (by row operations):

$$\begin{bmatrix} A & I \end{bmatrix} = \begin{bmatrix} 1 & 3 & 1 & 0 \\ 2 & 7 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} A & I \end{bmatrix} = \begin{bmatrix} 1 & 4 & 1 & 0 \\ 3 & 9 & 0 & 1 \end{bmatrix}$$

23 Follow the 3 by 3 text example but with plus signs in A. Eliminate above and below the pivots to reduce $[\,A \quad I\,]$ to $[\,I \quad A^{-1}\,]$:

$$\begin{bmatrix} A & I \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{bmatrix}.$$

24 Use Gauss-Jordan elimination on $[\,U \quad I\,]$ to find the upper triangular U^{-1}:

$$UU^{-1} = I \qquad \begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} & & \\ x_1 & x_2 & x_3 \\ & & \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

25 Find A^{-1} and B^{-1} (*if they exist*) by elimination on $[\,A \quad I\,]$ and $[\,B \quad I\,]$:

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

26 What three matrices E_{21} and E_{12} and D^{-1} reduce $A = \left[\begin{smallmatrix} 1 & 2 \\ 2 & 6 \end{smallmatrix}\right]$ to the identity matrix? Multiply $D^{-1}E_{12}E_{21}$ to find A^{-1}.

27 Invert these matrices A by the Gauss-Jordan method starting with $[\,A \quad I\,]$:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

28 Exchange rows and continue with Gauss-Jordan to find A^{-1}:

$$[\,A \quad I\,] = \begin{bmatrix} 0 & 2 & 1 & 0 \\ 2 & 2 & 0 & 1 \end{bmatrix}.$$

29 True or false (with a counterexample if false and a reason if true):

(a) A 4 by 4 matrix with a row of zeros is not invertible.

(b) Every matrix with 1's down the main diagonal is invertible.

(c) If A is invertible then A^{-1} and A^2 are invertible.

30 For which three numbers c is this matrix not invertible, and why not?

$$A = \begin{bmatrix} 2 & c & c \\ c & c & c \\ 8 & 7 & c \end{bmatrix}.$$

31 Prove that A is invertible if $a \neq 0$ and $a \neq b$ (find the pivots or A^{-1}):

$$A = \begin{bmatrix} a & b & b \\ a & a & b \\ a & a & a \end{bmatrix}.$$

32 This matrix has a remarkable inverse. Find A^{-1} by elimination on $[A \;\; I]$. Extend to a 5 by 5 "alternating matrix" and guess its inverse; then multiply to confirm.

$$\text{Invert } A = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and solve } A\boldsymbol{x} = (1,1,1,1).$$

33 Suppose the matrices P and Q have the same rows as I but in any order. They are "permutation matrices". Show that $P - Q$ is singular by solving $(P - Q)\boldsymbol{x} = \mathbf{0}$.

34 Find and check the inverses (assuming they exist) of these block matrices:

$$\begin{bmatrix} I & 0 \\ C & I \end{bmatrix} \quad \begin{bmatrix} A & 0 \\ C & D \end{bmatrix} \quad \begin{bmatrix} 0 & I \\ I & D \end{bmatrix}.$$

35 Could a 4 by 4 matrix A be invertible if every row contains the numbers $0, 1, 2, 3$ in some order? What if every row of B contains $0, 1, 2, -3$ in some order?

36 In the Worked Example **2.5 C**, the triangular Pascal matrix L has an inverse with "alternating diagonals". Check that this L^{-1} is DLD, where the diagonal matrix D has alternating entries $1, -1, 1, -1$. Then $LDLD = I$, so what is the inverse of $LD =$ pascal (4,1)?

37 The Hilbert matrices have $H_{ij} = 1/(i + j - 1)$. Ask MATLAB for the exact 6 by 6 inverse invhilb(6). Then ask it to compute inv(hilb(6)). How can these be different, when the computer never makes mistakes?

38 (a) Use inv(P) to invert MATLAB's 4 by 4 symmetric matrix $P =$ pascal(4).

(b) Create Pascal's lower triangular $L =$ abs(pascal(4,1)) and test $P = LL^{\mathrm{T}}$.

39 If $A =$ ones(4) and $\boldsymbol{b} =$ rand(4,1), how does MATLAB tell you that $A\boldsymbol{x} = \boldsymbol{b}$ has no solution? For the special $\boldsymbol{b} =$ ones(4,1), which solution to $A\boldsymbol{x} = \boldsymbol{b}$ is found by $A\backslash\boldsymbol{b}$?

Challenge Problems

40 (Recommended) A is a 4 by 4 matrix with 1's on the diagonal and $-a, -b, -c$ on the diagonal above. Find A^{-1} for this bidiagonal matrix.

41 Suppose E_1, E_2, E_3 are 4 by 4 identity matrices, except E_1 has a, b, c in column 1 and E_2 has d, e in column 2 and E_3 has f in column 3 (below the 1's). Multiply $L = E_1E_2E_3$ to show that all these nonzeros are copied into L.

$E_1E_2E_3$ is in the *opposite* order from elimination (because E_3 is acting first). But $E_1E_2E_3 = L$ is in the *correct* order to invert elimination and recover A.

42 Direct multiplications **1–4** give $MM^{-1} = I$, and I would recommend doing #3. M^{-1} shows the change in A^{-1} (useful to know) when a matrix is subtracted from A:

1	$M = I - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$	and	$M^{-1} = I + \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}/(1 - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{u})$	(rank 1 change in I)
2	$M = A - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$	and	$M^{-1} = A^{-1} + A^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}A^{-1}/(1 - \boldsymbol{v}^{\mathrm{T}}A^{-1}\boldsymbol{u})$	
3	$M = I - UV$	and	$M^{-1} = I_n + U(I_m - VU)^{-1}V$	
4	$M = A - UW^{-1}V$	and	$M^{-1} = A^{-1} + A^{-1}U(W - VA^{-1}U)^{-1}VA^{-1}$	

The Woodbury-Morrison formula **4** is the "matrix inversion lemma" in engineering. The ***Kalman filter*** for solving block tridiagonal systems uses formula **4** at each step. The four matrices M^{-1} are in diagonal blocks when inverting these block matrices ($\boldsymbol{v}^{\mathrm{T}}$ is 1 by n, $\boldsymbol{u}$ is n by 1, V is m by n, U is n by m).

$$\begin{bmatrix} I & \boldsymbol{u} \\ \boldsymbol{v}^{\mathrm{T}} & 1 \end{bmatrix} \qquad \begin{bmatrix} A & \boldsymbol{u} \\ \boldsymbol{v}^{\mathrm{T}} & 1 \end{bmatrix} \qquad \begin{bmatrix} I_n & U \\ V & I_m \end{bmatrix} \qquad \begin{bmatrix} A & U \\ V & W \end{bmatrix}$$

43 Second difference matrices have beautiful inverses if they start with $T_{11} = 1$ (instead of $K_{11} = 2$). Here is the 3 by 3 tridiagonal matrix T and its inverse:

$$\boldsymbol{T_{11} = 1} \qquad T = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \qquad T^{-1} = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

One approach is Gauss-Jordan elimination on $[\,T \quad I\,]$. That seems too mechanical. I would rather write T as the product of first differences L times U. The inverses of L and U in Worked Example **2.5 A** are **sum matrices**, so here are T and T^{-1}:

$$LU = \underset{\textbf{difference}}{\begin{bmatrix} 1 & & \\ -1 & 1 & \\ 0 & -1 & 1 \end{bmatrix}} \underset{\textbf{difference}}{\begin{bmatrix} 1 & -1 & 0 \\ & 1 & -1 \\ & & 1 \end{bmatrix}} \qquad U^{-1}L^{-1} = \underset{\textbf{sum}}{\begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{bmatrix}} \underset{\textbf{sum}}{\begin{bmatrix} 1 & & \\ 1 & 1 & \\ 1 & 1 & 1 \end{bmatrix}}$$

Question. (**4 by 4**) What are the pivots of T? What is its 4 by 4 inverse? The reverse order UL gives what matrix T^*? What is the inverse of T^*?

44 Here are two more difference matrices, both important. ***But are they invertible***?

$$\textbf{Cyclic } C = \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix} \qquad \textbf{Free ends } F = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}.$$

One test is elimination—the fourth pivot fails. Another test is the determinant, we don't want that. The best way is much faster, and independent of matrix size:

Produce $\boldsymbol{x} \neq \mathbf{0}$ so that $C\boldsymbol{x} = \mathbf{0}$. Do the same for $F\boldsymbol{x} = \mathbf{0}$. Not invertible.

Show how both equations $C\boldsymbol{x} = \boldsymbol{b}$ and $F\boldsymbol{x} = \boldsymbol{b}$ lead to $0 = b_1 + b_2 + \cdots + b_n$. There is no solution for other $\boldsymbol{b}$.

45 *Elimination for a 2 by 2 block matrix*: When you multiply the first block row by CA^{-1} and subtract from the second row, the "*Schur complement*" S appears:

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & S \end{bmatrix} \qquad \begin{matrix} A \text{ and } D \text{ are square} \\ S = D - CA^{-1}B. \end{matrix}$$

Multiply on the right to subtract $A^{-1}B$ times block column 1 from block column 2.

$$\begin{bmatrix} A & B \\ 0 & S \end{bmatrix}\begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} = ? \quad \text{Find } S \text{ for} \quad \begin{bmatrix} A & B \\ C & I \end{bmatrix} = \begin{bmatrix} 2 & 3 & 3 \\ 4 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix}.$$

The block pivots are A ***and*** S***. If they are invertible, so is*** $[\,A \quad B\,;\;\; C \quad D\,]$.

46 How does the identity $A(I + BA) = (I + AB)A$ connect the inverses of $I + BA$ and $I + AB$? Those are both invertible or both singular: not obvious.

2.6 Elimination = Factorization: $A = LU$

Students often say that mathematics courses are too theoretical. Well, not this section. It is almost purely practical. The goal is to describe Gaussian elimination in the most useful way. Many key ideas of linear algebra, when you look at them closely, are really *factorizations* of a matrix. The original matrix A becomes the product of two or three special matrices. The first factorization—also the most important in practice—comes now from elimination. ***The factors L and U are triangular matrices. The factorization that comes from elimination is*** $A = LU$.

We already know U, the upper triangular matrix with the pivots on its diagonal. The elimination steps take A to U. We will show how reversing those steps (taking U back to A) is achieved by a lower triangular L. ***The entries of L are exactly the multipliers*** ℓ_{ij}—which multiplied the pivot row j when it was subtracted from row i.

Start with a 2 by 2 example. The matrix A contains $2, 1, 6, 8$. The number to eliminate is 6. ***Subtract 3 times row*** 1 ***from row*** 2. That step is E_{21} in the forward direction with multiplier $\ell_{21} = 3$. The return step from U to A is $L = E_{21}^{-1}$ (an addition using $+3$):

$$\textit{Forward from } A \textit{ to } U: \quad E_{21}A = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 6 & 8 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 5 \end{bmatrix} = U$$

$$\textit{Back from } U \textit{ to } A: \quad E_{21}^{-1}U = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 6 & 8 \end{bmatrix} = A.$$

The second line is our factorization $LU = A$. Instead of E_{21}^{-1} we write L. Move now to larger matrices with many E's. ***Then L will include all their inverses.***

Each step from A to U multiplies by a matrix E_{ij} to produce zero in the (i, j) position. To keep this clear, we stay with the most frequent case—***when no row exchanges are involved.*** If A is 3 by 3, we multiply by E_{21} and E_{31} and E_{32}. The multipliers ℓ_{ij} produce zeros in the $(2, 1)$ and $(3, 1)$ and $(3, 2)$ positions—all below the diagonal. Elimination ends with the upper triangular U.

Now move those E's onto the other side, *where their inverses multiply* U:

$$(E_{32}E_{31}E_{21})A = U \quad \textbf{becomes} \quad A = (E_{21}^{-1}E_{31}^{-1}E_{32}^{-1})U \quad \textbf{which is} \quad A = LU. \tag{1}$$

The inverses go in opposite order, as they must. That product of three inverses is L. ***We have reached*** $A = LU$. Now we stop to understand it.

Explanation and Examples

First point: Every inverse matrix E^{-1} is *lower triangular.* Its off-diagonal entry is ℓ_{ij}, to undo the subtraction produced by $-\ell_{ij}$. The main diagonals of E and E^{-1} contain 1's. Our example above had $\ell_{21} = 3$ and $E = \left[\begin{smallmatrix} 1 & 0 \\ -3 & 1 \end{smallmatrix}\right]$ and $L = E^{-1} = \left[\begin{smallmatrix} 1 & 0 \\ 3 & 1 \end{smallmatrix}\right]$.

Second point: Equation (1) shows a lower triangular matrix (the product of the E_{ij}) multiplying A. It also shows all the E_{ij}^{-1} multiplying U to bring back A. ***This lower triangular product of inverses is L.***

One reason for working with the inverses is that we want to factor A, not U. The "inverse form" gives $A = LU$. Another reason is that we get something extra, almost more than we deserve. This is the third point, showing that L is exactly right.

Third point: Each multiplier ℓ_{ij} goes directly into its i, j position—*unchanged*—in the product of inverses which is L. Usually matrix multiplication will mix up all the numbers. Here that doesn't happen. The order is right for the inverse matrices, to keep the ℓ's unchanged. The reason is given below in equation (3).

Since each E^{-1} has 1's down its diagonal, the final good point is that L does too.

($A = LU$) ***This is elimination without row exchanges.*** The upper triangular U has the pivots on its diagonal. The lower triangular L has all 1's on its diagonal. ***The multipliers ℓ_{ij} are below the diagonal of L.***

Example 1 Elimination subtracts $\frac{1}{2}$ times row 1 from row 2. The last step subtracts $\frac{2}{3}$ times row 2 from row 3. The lower triangular L has $\ell_{21} = \frac{1}{2}$ and $\ell_{32} = \frac{2}{3}$. Multiplying LU produces A:

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & \frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 0 & \frac{3}{2} & 1 \\ 0 & 0 & \frac{4}{3} \end{bmatrix} = LU.$$

The (3, 1) multiplier is zero because the (3, 1) entry in A is zero. No operation needed.

Example 2 Change the top left entry from 2 to 1. The pivots all become 1. The multipliers are all 1. That pattern continues when A is 4 by 4:

$$\textbf{Special pattern} \qquad A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ 0 & 1 & 1 & \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ & 1 & 1 & 0 \\ & & 1 & 1 \\ & & & 1 \end{bmatrix}.$$

These LU examples are showing something extra, which is very important in practice. Assume no row exchanges. When can we predict *zeros* in L and U?

When a row of A starts with zeros, so does that row of L.

When a column of A starts with zeros, so does that column of U.

If a row starts with zero, we don't need an elimination step. L has a zero, which saves computer time. Similarly, zeros at the *start* of a column survive into U. But please realize: Zeros in the *middle* of a matrix are likely to be filled in, while elimination sweeps forward. We now explain why L has the multipliers ℓ_{ij} in position, with no mix-up.

The key reason why A equals LU: Ask yourself about the pivot rows that are subtracted from lower rows. Are they the original rows of A? *No*, elimination probably changed them. Are they rows of U? *Yes*, the pivot rows never change again. When computing the third row of U, we subtract multiples of earlier rows of U (*not rows of* A!):

$$\text{Row 3 of } U = (\text{Row 3 of } A) - \ell_{31}(\text{Row 1 of } U) - \ell_{32}(\text{Row 2 of } U). \tag{2}$$

Rewrite this equation to see that the row $[\,\ell_{31} \quad \ell_{32} \quad 1\,]$ is multiplying U:

$$(\text{Row 3 of } A) = \ell_{31}(\text{Row 1 of } U) + \ell_{32}(\text{Row 2 of } U) + 1(\text{Row 3 of } U). \tag{3}$$

This is exactly row 3 ***of*** $A = LU$. That row of L holds $\ell_{31}, \ell_{32}, 1$. All rows look like this, whatever the size of A. With no row exchanges, we have $A = LU$.

Better balance The LU factorization is "unsymmetric" because U has the pivots on its diagonal where L has 1's. This is easy to change. ***Divide U by a diagonal matrix D that contains the pivots.*** That leaves a new matrix with 1's on the diagonal:

$$\text{Split } U \text{ into } \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix} \begin{bmatrix} 1 & u_{12}/d_1 & u_{13}/d_1 & \cdot \\ & 1 & u_{23}/d_2 & \cdot \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}.$$

It is convenient (but a little confusing) to keep the same letter U for this new upper triangular matrix. It has 1's on the diagonal (like L). Instead of the normal LU, the new form has D in the middle: ***Lower triangular L times diagonal D times upper triangular U.***

The triangular factorization can be written $\boldsymbol{A = LU}$ ***or*** $\boldsymbol{A = LDU}$.

Whenever you see LDU, it is understood that U has 1's on the diagonal. *Each row is divided by its first nonzero entry—the pivot.* Then L and U are treated evenly in LDU:

$$\begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 2 & 8 \\ 0 & 5 \end{bmatrix} \quad \text{splits further into} \quad \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 2 & \\ & 5 \end{bmatrix}\begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}. \tag{4}$$

The pivots 2 and 5 went into D. Dividing the rows by 2 and 5 left the rows $[\,1 \quad 4\,]$ and $[\,0 \quad 1\,]$ in the new U with diagonal ones. The multiplier 3 is still in L.

My own lectures sometimes stop at this point. The next paragraphs show how elimination codes are organized, and how long they take. If MATLAB (or any software) is available, you can measure the computing time by just counting the seconds.

One Square System = Two Triangular Systems

The matrix L contains our memory of Gaussian elimination. It holds the numbers that multiplied the pivot rows, before subtracting them from lower rows. When do we need this record and how do we use it in solving $A\boldsymbol{x} = \boldsymbol{b}$?

We need L as soon as there is a *right side* $\boldsymbol{b}$. The factors L and U were completely decided by the left side (the matrix A). On the right side of $A\boldsymbol{x} = \boldsymbol{b}$, we use L^{-1} and then U^{-1}. That ***Solve*** step deals with two triangular matrices.

1 ***Factor*** (into L and U, by elimination on the left side matrix A)

2 ***Solve*** (forward elimination on $\boldsymbol{b}$ using L, then back substitution for $\boldsymbol{x}$ using U).

Earlier, we worked on A and $\boldsymbol{b}$ at the same time. No problem with that—just augment to $[\,A \quad \boldsymbol{b}\,]$. But most computer codes keep the two sides separate. The memory of elimination is held in L and U, to process $\boldsymbol{b}$ whenever we want to. The User's Guide to LAPACK remarks that "This situation is so common and the savings are so important that no provision has been made for solving a single system with just one subroutine."

How does ***Solve*** work on $\boldsymbol{b}$? First, apply forward elimination to the right side (the multipliers are stored in L, use them now). This changes $\boldsymbol{b}$ to a new right side $\boldsymbol{c}$. *We are really solving* $L\boldsymbol{c} = \boldsymbol{b}$. Then back substitution solves $U\boldsymbol{x} = \boldsymbol{c}$ as always. The original system $A\boldsymbol{x} = \boldsymbol{b}$ is factored into ***two triangular systems***:

$$\textbf{Forward and backward} \qquad \textit{Solve} \quad L\boldsymbol{c} = \boldsymbol{b} \quad \textit{and then solve} \quad U\boldsymbol{x} = \boldsymbol{c}\,. \tag{5}$$

To see that $\boldsymbol{x}$ is correct, multiply $U\boldsymbol{x} = \boldsymbol{c}$ by L. Then $LU\boldsymbol{x} = L\boldsymbol{c}$ is just $A\boldsymbol{x} = \boldsymbol{b}$.

To emphasize: There is *nothing new* about those steps. This is exactly what we have done all along. We were really solving the triangular system $L\boldsymbol{c} = \boldsymbol{b}$ as elimination went forward. Then back substitution produced $\boldsymbol{x}$. An example shows what we actually did.

Example 3 Forward elimination (downward) on $A\boldsymbol{x} = \boldsymbol{b}$ ends at $U\boldsymbol{x} = \boldsymbol{c}$:

$$A\boldsymbol{x} = \boldsymbol{b} \quad \begin{aligned} u + 2v &= 5 \\ 4u + 9v &= 21 \end{aligned} \quad \text{becomes} \quad \begin{aligned} u + 2v &= 5 \\ v &= 1 \end{aligned} \quad U\boldsymbol{x} = \boldsymbol{c}$$

The multiplier was 4, which is saved in L. The right side used it to change 21 to 1:

$$L\boldsymbol{c} = \boldsymbol{b} \quad \text{The lower triangular system} \quad \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} 5 \\ 21 \end{bmatrix} \quad \text{gave} \quad \boldsymbol{c} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

$$U\boldsymbol{x} = \boldsymbol{c} \quad \text{The upper triangular system} \quad \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \end{bmatrix} \quad \text{gives} \quad \boldsymbol{x} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

L and U can go into the n^2 storage locations that originally held A (now forgettable).

The Cost of Elimination

A very practical question is cost—or computing time. We can solve 1000 equations on a PC. What if $n = 100{,}000$? (*Not if A is dense.*) Large systems come up all the time in scientific computing, where a three-dimensional problem can easily lead to a million unknowns. We can let the calculation run overnight, but we can't leave it for 100 years.

The first stage of elimination, on column 1, produces zeros below the first pivot. To find each new entry below the pivot row requires one multiplication and one subtraction. *We will count this first stage as n^2 multiplications and n^2 subtractions.* It is actually less, $n^2 - n$, because row 1 does not change.

The next stage clears out the second column below the second pivot. The working matrix is now of size $n-1$. Estimate this stage by $(n-1)^2$ multiplications and subtractions. The matrices are getting smaller as elimination goes forward. The rough count to reach U is the sum of squares $n^2 + (n-1)^2 + \cdots + 2^2 + 1^2$.

There is an exact formula $\frac{1}{3}n(n+\frac{1}{2})(n+1)$ for this sum of squares. When n is large, the $\frac{1}{2}$ and the 1 are not important. *The number that matters is* $\frac{1}{3}n^3$. The sum of squares is like the integral of x^2! The integral from 0 to n is $\frac{1}{3}n^3$:

Elimination on A requires about $\frac{1}{3}n^3$ multiplications and $\frac{1}{3}n^3$ subtractions.

What about the right side $\boldsymbol{b}$? Going forward, we subtract multiples of b_1 from the lower components $b_2, \ldots, b_n$. This is $n-1$ steps. The second stage takes only $n-2$ steps, because b_1 is not involved. The last stage of forward elimination takes one step.

Now start back substitution. Computing x_n uses one step (divide by the last pivot). The next unknown uses two steps. When we reach x_1 it will require n steps ($n-1$ substitutions of the other unknowns, then division by the first pivot). The total count on the right side, from $\boldsymbol{b}$ to $\boldsymbol{c}$ to $\boldsymbol{x}$—*forward to the bottom and back to the top*—is exactly n^2:

$$[(n-1)+(n-2)+\cdots+1] \quad + \quad [1+2+\cdots+(n-1)+n] = n^2. \tag{6}$$

To see that sum, pair off $(n-1)$ with 1 and $(n-2)$ with 2. The pairings leave n terms, each equal to n. That makes n^2. The right side costs a lot less than the left side!

Solve ***Each right side needs n^2 multiplications and n^2 subtractions.***

A **band matrix** B has only w nonzero diagonals below and also above its main diagonal. The zero entries outside the band stay zero in elimination (zeros in L and U). Clearing out the first column needs w^2 multiplications and subtractions (w zeros to be produced below the pivot, each one using a pivot row of length w). Then clearing out all n columns, to reach U, needs no more than nw^2. This saves a lot of time:

Band matrices **Factor** change $\dfrac{1}{3}n^3$ to $\boldsymbol{nw^2}$ **Solve** change n^2 to $\boldsymbol{2nw}$

Here are codes to factor A into LU and to solve $A\boldsymbol{x} = \boldsymbol{b}$. The Teaching code **slu** stops right away if a number smaller than the tolerance "*tol*" appears in a pivot position. The Teaching Codes are on **web.mit.edu/18.06/www**. Professional codes will look down each column for the largest available pivot, to exchange rows and continue solving.

MATLAB's backslash command $\boldsymbol{x} = A\backslash\boldsymbol{b}$ combines **Factor** and **Solve** to reach $\boldsymbol{x}$.

```
function [L, U] = slu(A)
%     Square LU factorization with no row exchanges!
[n, n] = size(A);   tol = 1.e - 6;
for k = 1 : n
  if abs(A(k, k)) < tol
  end       % Cannot proceed without a row exchange: stop
  L(k, k) = 1;
  for i = k + 1 : n
    L(i, k) = A(i, k)/A(k, k);   % Multipliers for column k are put into L
    for j = k + 1 : n   % Elimination beyond row k and column k
      A(i, j) = A(i, j) - L(i, k) * A(k, j);   % Matrix still called A
    end
  end
  for j = k : n
     U(k, j) = A(k, j);        % row k is settled, now name it U
  end
end
```

```
function x = slv(A, b)
%     Solve Ax = b using L and U from slu(A).
[L, U] = slu(A); s = 0;   % No row exchanges!
for k = 1 : n   % Forward elimination to solve Lc = b
  for j = 1 : k - 1
    s = s + L(k, j) * c(j);   % Add L times earlier c(j) before c(k)
  end
  c(k) = b(k) - s; s = 0;   % Find c(k) and reset s for next k
end
for k = n : -1 : 1   % Going backwards from x(n) to x(1)
  for j = k + 1 : n   % Back substitution
    t = t + U(k, j) * x(j);   % U times later x(j)
  end
  x(k) = (c(k) - t)/U(k, k);   % Divide by pivot
end
x = x';   % Transpose to column vector
```

How long does it take to solve $Ax = b$? For a random matrix of order $n = 1000$, a typical time is 1 second. See **web.mit.edu/18.06** and **math.mit.edu/linearalgebra** for the times in MATLAB, Maple, Mathematica, SciLab, Python, and R. The time is multiplied by about 8 when n is multiplied by 2. For professional codes go to **netlib.org**.

According to this n^3 rule, matrices that are 10 times as large (order 10,000) will take a thousand seconds. Matrices of order 100,000 will take a million seconds. This is too expensive without a supercomputer, but remember that these matrices are full. Most matrices in practice are sparse (many zero entries). In that case $A = LU$ is much faster.

For tridiagonal matrices of order 10,000, storing only the nonzeros, solving $Ax = b$ is a breeze. Provided the code recognizes that A is tridiagonal.

■ REVIEW OF THE KEY IDEAS ■

1. Gaussian elimination (with no row exchanges) factors A into L times U.
2. The lower triangular L contains the numbers ℓ_{ij} that multiply pivot rows, going from A to U. The product LU adds those rows back to recover A.
3. On the right side we solve $L\boldsymbol{c} = \boldsymbol{b}$ (forward) and $U\boldsymbol{x} = \boldsymbol{c}$ (backward).
4. ***Factor*** : There are $\frac{1}{3}(n^3 - n)$ multiplications and subtractions on the left side.
5. ***Solve*** : There are n^2 multiplications and subtractions on the right side.
6. For a band matrix, change $\frac{1}{3}n^3$ to nw^2 and change n^2 to $2wn$.

■ WORKED EXAMPLES ■

2.6 A The lower triangular Pascal matrix L contains the famous "*Pascal triangle*". Gauss-Jordan found its inverse in the worked example **2.5 C**. This problem connects L to the *symmetric* Pascal matrix P and the upper triangular U. The symmetric P has Pascal's triangle tilted, so each entry is the sum of the entry above and the entry to the left. The n by n symmetric P is pascal(n) in MATLAB.

Problem: *Establish the amazing lower-upper factorization* $P = LU$.

$$\text{pascal(4)} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{bmatrix} = \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 \\ \mathbf{1} & \mathbf{1} & 0 & 0 \\ \mathbf{1} & \mathbf{2} & \mathbf{1} & 0 \\ \mathbf{1} & \mathbf{3} & \mathbf{3} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ 0 & \mathbf{1} & \mathbf{2} & \mathbf{3} \\ 0 & 0 & \mathbf{1} & \mathbf{3} \\ 0 & 0 & 0 & \mathbf{1} \end{bmatrix} = LU.$$

Then predict and check the next row and column for 5 by 5 Pascal matrices.

Solution You could multiply LU to get P. Better to start with the symmetric P and reach the upper triangular U by elimination:

$$P = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 2 & 5 & 9 \\ 0 & 3 & 9 & 19 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 3 & 10 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix} = U.$$

The multipliers ℓ_{ij} that entered these steps go perfectly into L. Then $P = LU$ is a particularly neat example. *Notice that every pivot is* 1 on the diagonal of U.

The next section will show how symmetry produces a special relationship between the triangular L and U. For Pascal, U is the "**transpose**" of L.

You might expect the MATLAB command lu(pascal(4)) to produce these L and U. That doesn't happen because the **lu** subroutine chooses the largest available pivot in each column. The second pivot will change from 1 to 3. But a "Cholesky factorization" does no row exchanges: U = chol(pascal(4))

The full proof of $P = LU$ for all Pascal sizes is quite fascinating. The paper "*Pascal Matrices*" is on the course web page **web.mit.edu/18.06** which is also available through MIT's *OpenCourseWare* at **ocw.mit.edu**. These Pascal matrices have so many remarkable properties—we will see them again.

2.6 B The problem is: *Solve* $Px = b = (1,0,0,0)$. This right side = column of I means that x will be the first column of P^{-1}. That is Gauss-Jordan, matching the columns of $PP^{-1} = I$. We already know the Pascal matrices L and U as factors of P:

$$\textbf{Two triangular systems} \qquad Lc = b \text{ (forward)} \qquad Ux = c \text{ (back)}.$$

Solution The lower triangular system $Lc = b$ is solved *top to bottom*:

$$\begin{aligned} c_1 \qquad\qquad\quad &= 1 \\ c_1 + c_2 \qquad\quad &= 0 \\ c_1 + 2c_2 + c_3 \quad &= 0 \\ c_1 + 3c_2 + 3c_3 + c_4 &= 0 \end{aligned} \quad \text{gives} \quad \begin{aligned} c_1 &= +1 \\ c_2 &= -1 \\ c_3 &= +1 \\ c_4 &= -1 \end{aligned}$$

Forward elimination is multiplication by L^{-1}. It produces the upper triangular system $Ux = c$. The solution x comes as always by back substitution, *bottom to top*:

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 1 \\ x_2 + 2x_3 + 3x_4 &= -1 \\ x_3 + 3x_4 &= 1 \\ x_4 &= -1 \end{aligned} \quad \text{gives} \quad \begin{aligned} x_1 &= +4 \\ x_2 &= -6 \\ x_3 &= +4 \\ x_4 &= -1 \end{aligned}$$

I see a pattern in that x, but I don't know where it comes from. Try **inv(pascal**(4)).

Problem Set 2.6

Problems 1–14 compute the factorization $A = LU$ (and also $A = LDU$).

1 (Important) Forward elimination changes $\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}x = b$ to a triangular $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}x = c$:

$$\begin{aligned} x + y &= 5 \\ x + 2y &= 7 \end{aligned} \quad \longrightarrow \quad \begin{aligned} x + y &= 5 \\ y &= 2 \end{aligned} \qquad \begin{bmatrix} 1 & 1 & 5 \\ 1 & 2 & 7 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & 5 \\ 0 & 1 & 2 \end{bmatrix}$$

That step subtracted $\ell_{21} =$ ____ times row 1 from row 2. The reverse step *adds* ℓ_{21} times row 1 to row 2. The matrix for that reverse step is $L =$ ____. Multiply this L times the triangular system $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}x_1 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$ to get ____ = ____. In letters, L multiplies $Ux = c$ to give ____.

2 Write down the 2 by 2 triangular systems $Lc = b$ and $Ux = c$ from Problem 1. Check that $c = (5,2)$ solves the first one. Find x that solves the second one.

3 (Move to 3 by 3) Forward elimination changes $A\boldsymbol{x} = \boldsymbol{b}$ to a triangular $U\boldsymbol{x} = \boldsymbol{c}$:

$$\begin{aligned} x + \ y + \ z &= 5 \\ x + 2y + 3z &= 7 \\ x + 3y + 6z &= 11 \end{aligned} \qquad \begin{aligned} x + \ y + \ z &= 5 \\ y + 2z &= 2 \\ 2y + 5z &= 6 \end{aligned} \qquad \begin{aligned} x + \ y + \ z &= 5 \\ y + 2z &= 2 \\ z &= 2 \end{aligned}$$

The equation $z = 2$ in $U\boldsymbol{x} = \boldsymbol{c}$ comes from the original $x + 3y + 6z = 11$ in $A\boldsymbol{x} = \boldsymbol{b}$ by subtracting $\ell_{31} =$ ____ times equation 1 and $\ell_{32} =$ ____ times the *final* equation 2. Reverse that to recover $[1 \ \ 3 \ \ 6 \ \ 11]$ in the last row of A and $\boldsymbol{b}$ from the final $[1 \ \ 1 \ \ 1 \ \ 5]$ and $[0 \ \ 1 \ \ 2 \ \ 2]$ and $[0 \ \ 0 \ \ 1 \ \ 2]$ in U and $\boldsymbol{c}$:

$$\text{Row 3 of } \begin{bmatrix} A & \boldsymbol{b} \end{bmatrix} = (\ell_{31} \text{ Row 1} + \ell_{32} \text{ Row 2} + 1 \text{ Row 3}) \text{ of } \begin{bmatrix} U & \boldsymbol{c} \end{bmatrix}.$$

In matrix notation this is multiplication by L. So $A = LU$ and $\boldsymbol{b} = L\boldsymbol{c}$.

4 What are the 3 by 3 triangular systems $L\boldsymbol{c} = \boldsymbol{b}$ and $U\boldsymbol{x} = \boldsymbol{c}$ from Problem 3? Check that $\boldsymbol{c} = (5, 2, 2)$ solves the first one. Which $\boldsymbol{x}$ solves the second one?

5 What matrix E puts A into triangular form $EA = U$? Multiply by $E^{-1} = L$ to factor A into LU:

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 4 & 2 \\ 6 & 3 & 5 \end{bmatrix}.$$

6 What two elimination matrices E_{21} and E_{32} put A into upper triangular form $E_{32}E_{21}A = U$? Multiply by E_{32}^{-1} and E_{21}^{-1} to factor A into $LU = E_{21}^{-1}E_{32}^{-1}U$:

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 4 & 5 \\ 0 & 4 & 0 \end{bmatrix}.$$

7 What three elimination matrices E_{21}, E_{31}, E_{32} put A into its upper triangular form $E_{32}E_{31}E_{21}A = U$? Multiply by E_{32}^{-1}, E_{31}^{-1} and E_{21}^{-1} to factor A into L times U:

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 2 & 2 \\ 3 & 4 & 5 \end{bmatrix} \qquad L = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}.$$

8 Suppose A is already lower triangular with 1's on the diagonal. Then $U = I$!

$$A = L = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{bmatrix}.$$

The elimination matrices E_{21}, E_{31}, E_{32} contain $-a$ then $-b$ then $-c$.

(a) Multiply $E_{32}E_{31}E_{21}$ to find the single matrix E that produces $EA = I$.

(b) Multiply $E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}$ to bring back L (nicer than E).

9 *When zero appears in a pivot position, $A = LU$ is not possible!* (We are requiring nonzero pivots in U.) Show directly why these are both impossible:

$$\begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \ell & 1 \end{bmatrix} \begin{bmatrix} d & e \\ 0 & f \end{bmatrix} \qquad \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ \ell & 1 & \\ m & n & 1 \end{bmatrix} \begin{bmatrix} d & e & g \\ & f & h \\ & & i \end{bmatrix}.$$

This difficulty is fixed by a row exchange. That needs a "permutation" P.

10 Which number c leads to zero in the second pivot position? A row exchange is needed and $A = LU$ will not be possible. Which c produces zero in the third pivot position? Then a row exchange can't help and elimination fails:

$$A = \begin{bmatrix} 1 & c & 0 \\ 2 & 4 & 1 \\ 3 & 5 & 1 \end{bmatrix}.$$

11 What are L and D (the diagonal ***pivot matrix***) for this matrix A? What is U in $A = LU$ and what is the new U in $A = LDU$?

Already triangular $$A = \begin{bmatrix} 2 & 4 & 8 \\ 0 & 3 & 9 \\ 0 & 0 & 7 \end{bmatrix}.$$

12 A and B are symmetric across the diagonal (because $4 = 4$). Find their triple factorizations LDU and say how U is related to L for these symmetric matrices:

Symmetric $$A = \begin{bmatrix} 2 & 4 \\ 4 & 11 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 4 & 0 \\ 4 & 12 & 4 \\ 0 & 4 & 0 \end{bmatrix}.$$

13 (*Recommended*) Compute L and U for the symmetric matrix A:

$$A = \begin{bmatrix} a & a & a & a \\ a & b & b & b \\ a & b & c & c \\ a & b & c & d \end{bmatrix}.$$

Find four conditions on a, b, c, d to get $A = LU$ with four pivots.

14 This nonsymmetric matrix will have the same L as in Problem **13**:

Find L and U for $$A = \begin{bmatrix} a & r & r & r \\ a & b & s & s \\ a & b & c & t \\ a & b & c & d \end{bmatrix}.$$

Find the four conditions on a, b, c, d, r, s, t to get $A = LU$ with four pivots.

Problems 15-16 use L and U (without needing A) to solve $Ax = b$.

15 Solve the triangular system $Lc = b$ to find c. Then solve $Ux = c$ to find x:

$$L = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} 2 & 4 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 2 \\ 11 \end{bmatrix}.$$

For safety multiply LU and solve $Ax = b$ as usual. Circle c when you see it.

16 Solve $Lc = b$ to find c. Then solve $Ux = c$ to find x. **What was** A?

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}.$$

17 (a) When you apply the usual elimination steps to L, what matrix do you reach?

$$L = \begin{bmatrix} 1 & 0 & 0 \\ \ell_{21} & 1 & 0 \\ \ell_{31} & \ell_{32} & 1 \end{bmatrix}.$$

(b) When you apply the same steps to I, what matrix do you get?

(c) When you apply the same steps to LU, what matrix do you get?

18 If $A = LDU$ and also $A = L_1D_1U_1$ with all factors invertible, then $L = L_1$ and $D = D_1$ and $U = U_1$. *"The three factors are unique."*

Derive the equation $L_1^{-1}LD = D_1U_1U^{-1}$. Are the two sides triangular or diagonal? Deduce $L = L_1$ and $U = U_1$ (they all have diagonal 1's). Then $D = D_1$.

19 *Tridiagonal matrices* have zero entries except on the main diagonal and the two adjacent diagonals. Factor these into $A = LU$ and $A = LDL^{\mathrm{T}}$:

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} a & a & 0 \\ a & a+b & b \\ 0 & b & b+c \end{bmatrix}.$$

20 When T is tridiagonal, its L and U factors have only two nonzero diagonals. How would you take advantage of knowing the zeros in T, in a code for Gaussian elimination? Find L and U.

Tridiagonal
$$T = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 2 & 3 & 1 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 3 & 4 \end{bmatrix}.$$

21 If A and B have nonzeros in the positions marked by x, which zeros (marked by 0) *stay zero* in their factors L and U?

$$A = \begin{bmatrix} x & x & x & x \\ x & x & x & 0 \\ 0 & x & x & x \\ 0 & 0 & x & x \end{bmatrix} \qquad B = \begin{bmatrix} x & x & x & 0 \\ x & x & 0 & x \\ x & 0 & x & x \\ 0 & x & x & x \end{bmatrix}.$$

22 Suppose you eliminate upwards (almost unheard of). Use the last row to produce zeros in the last column (the pivot is 1). Then use the second row to produce zero above the second pivot. Find the factors in the unusual order $A = UL$.

Upper times lower $$A = \begin{bmatrix} 5 & 3 & 1 \\ 3 & 3 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

23 *Easy but important.* If A has pivots $5, 9, 3$ with no row exchanges, what are the pivots for the upper left 2 by 2 submatrix A_2 (without row 3 and column 3)?

Challenge Problems

24 Which invertible matrices allow $A = LU$ (elimination without row exchanges)? *Good question!* Look at each of the square upper left submatrices of A.

All upper left k by k submatrices A_k must be invertible (sizes $k = 1, \ldots, n$).

Explain that answer: A_k factors into _____ because $LU = \begin{bmatrix} L_k & 0 \\ * & * \end{bmatrix}\begin{bmatrix} U_k & * \\ 0 & * \end{bmatrix}$.

25 For the 6 by 6 second difference constant-diagonal matrix K, put the pivots and multipliers into $K = LU$. (L and U will have only two nonzero diagonals, because K has three.) Find a formula for the i, j entry of L^{-1}, by software like MATLAB using inv(L) or by looking for a nice pattern.

$-1, 2, -1$ matrix $$K = \begin{bmatrix} 2 & -1 & & & & \\ -1 & \cdot & \cdot & & & \\ & \cdot & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & -1 \\ & & & & -1 & 2 \end{bmatrix} = \text{toeplitz([2 } -1 \ 0 \ 0 \ 0 \ 0\text{])}$$

26 If you print K^{-1}, it doesn't look so good. But if you print $7K^{-1}$ (when K is 6 by 6), that matrix looks wonderful. Write down $7K^{-1}$ by hand, following this pattern:

1 Row 1 and column 1 are $(6, 5, 4, 3, 2, 1)$.

2 On and above the main diagonal, row i is i times row 1.

3 On and below the main diagonal, column j is j times column 1.

Multiply K times that $7K^{-1}$ to produce $7I$. Here is that pattern for $n = 3$:

3 by 3 case
The determinant
of this K is 4
$$(K)(4K^{-1}) = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}\begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 4 & & \\ & 4 & \\ & & 4 \end{bmatrix}.$$

2.7 Transposes and Permutations

We need one more matrix, and fortunately it is much simpler than the inverse. It is the ***"transpose"*** of A, which is denoted by A^{T}. *The columns of A^{T} are the rows of A.*

When A is an m by n matrix, the transpose is n by m:

$$\textbf{Transpose} \qquad \text{If} \quad A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 4 \end{bmatrix} \quad \text{then} \quad A^{\mathrm{T}} = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 3 & 4 \end{bmatrix}.$$

You can write the rows of A into the columns of A^{T}. Or you can write the columns of A into the rows of A^{T}. The matrix "flips over" its main diagonal. The entry in row i, column j of A^{T} comes from row j, column i of the original A:

$$\textbf{Exchange rows and columns} \qquad (A^{\mathrm{T}})_{ij} = A_{ji}.$$

The transpose of a lower triangular matrix is upper triangular. (But the inverse is still lower triangular.) The transpose of A^{T} is A.

Note MATLAB's symbol for the transpose of A is A'. Typing $[1\ \ 2\ \ 3]$ gives a row vector and the column vector is $\boldsymbol{v} = [1\ \ 2\ \ 3]'$. To enter a matrix M with second column $\boldsymbol{w} = [\ 4\ \ 5\ \ 6\ ]'$ you could define $M = [\ \boldsymbol{v}\ \ \boldsymbol{w}\ ]$. Quicker to enter by rows and then transpose the whole matrix: $M = [1\ \ 2\ \ 3;\ \ 4\ \ 5\ \ 6]'$.

The rules for transposes are very direct. We can transpose $A+B$ to get $(A+B)^{\mathrm{T}}$. Or we can transpose A and B separately, and then add $A^{\mathrm{T}} + B^{\mathrm{T}}$—with the same result. The serious questions are about the transpose of a product AB and an inverse A^{-1}:

$$\textbf{Sum} \qquad \text{The transpose of} \quad A+B \quad \text{is} \quad A^{\mathrm{T}} + B^{\mathrm{T}}. \tag{1}$$

$$\textbf{Product} \qquad \text{The transpose of} \quad AB \quad \text{is} \quad (AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}. \tag{2}$$

$$\textbf{Inverse} \qquad \text{The transpose of} \quad A^{-1} \quad \text{is} \quad (A^{-1})^{\mathrm{T}} = (A^{\mathrm{T}})^{-1}. \tag{3}$$

Notice especially how $B^{\mathrm{T}}A^{\mathrm{T}}$ comes in reverse order. For inverses, this reverse order was quick to check: $B^{-1}A^{-1}$ times AB produces I. To understand $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$, start with $(A\boldsymbol{x})^{\mathrm{T}} = \boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}$:

$$A\boldsymbol{x} \ \ \textit{combines the columns of } A \textit{ while } \ \boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}} \ \textit{combines the rows of } A^{\mathrm{T}}.$$

It is the same combination of the same vectors! In A they are columns, in A^{T} they are rows. So the transpose of the column $A\boldsymbol{x}$ is the row $\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}$. That fits our formula $(A\boldsymbol{x})^{\mathrm{T}} = \boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}$. Now we can prove the formula $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$, when B has several columns.

If $B = [\boldsymbol{x}_1\ \ \boldsymbol{x}_2]$ has two columns, apply the same idea to each column. The columns of AB are $A\boldsymbol{x}_1$ and $A\boldsymbol{x}_2$. Their transposes are the rows of $B^{\mathrm{T}}A^{\mathrm{T}}$:

$$\text{Transposing} \quad AB = \begin{bmatrix} \\ A\boldsymbol{x}_1 & A\boldsymbol{x}_2 & \cdots \\ \\ \end{bmatrix} \quad \text{gives} \quad \begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}}A^{\mathrm{T}} \\ \boldsymbol{x}_2^{\mathrm{T}}A^{\mathrm{T}} \\ \vdots \end{bmatrix} \quad \text{which is} \quad B^{\mathrm{T}}A^{\mathrm{T}}. \tag{4}$$

The right answer $B^{\mathrm{T}}A^{\mathrm{T}}$ comes out a row at a time. Here are numbers in $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$:

$$AB = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 5 & 0 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{5} & \mathbf{0} \\ \mathbf{9} & \mathbf{1} \end{bmatrix} \quad \text{and} \quad B^{\mathrm{T}}A^{\mathrm{T}} = \begin{bmatrix} 5 & 4 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{5} & \mathbf{9} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}.$$

The reverse order rule extends to three or more factors: $(ABC)^{\mathrm{T}}$ equals $C^{\mathrm{T}}B^{\mathrm{T}}A^{\mathrm{T}}$.

If $A = LDU$ ***then*** $A^{\mathrm{T}} = U^{\mathrm{T}}D^{\mathrm{T}}L^{\mathrm{T}}$. ***The pivot matrix has*** $D = D^{\mathrm{T}}$.

Now apply this product rule to both sides of $A^{-1}A = I$. On one side, I^{T} is I. We confirm the rule that $(A^{-1})^{\mathrm{T}}$ *is the inverse of* A^{T}, because their product is I:

Transpose of inverse $\quad A^{-1}A = I \quad$ is transposed to $\quad A^{\mathrm{T}}(A^{-1})^{\mathrm{T}} = I. \qquad (5)$

Similarly $AA^{-1} = I$ leads to $(A^{-1})^{\mathrm{T}}A^{\mathrm{T}} = I$. We can invert the transpose or we can transpose the inverse. Notice especially: A^{T} ***is invertible exactly when*** A ***is invertible.***

Example 1 The inverse of $A = \begin{bmatrix} 1 & 0 \\ 6 & 1 \end{bmatrix}$ is $A^{-1} = \begin{bmatrix} 1 & 0 \\ -6 & 1 \end{bmatrix}$. The transpose is $A^{\mathrm{T}} = \begin{bmatrix} 1 & 6 \\ 0 & 1 \end{bmatrix}$.

$$(A^{-1})^{\mathrm{T}} \quad \textit{and} \quad (A^{\mathrm{T}})^{-1} \quad \textit{are both equal to} \quad \begin{bmatrix} 1 & -6 \\ 0 & 1 \end{bmatrix}.$$

The Meaning of Inner Products

We know the dot product (inner product) of $\boldsymbol{x}$ and $\boldsymbol{y}$. It is the sum of numbers $x_i y_i$. Now we have a better way to write $\boldsymbol{x} \cdot \boldsymbol{y}$, without using that unprofessional dot. Use matrix notation instead:

$^{\mathrm{T}}$ **is inside**	***The dot product or inner product is*** $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$	$(1 \times n)(n \times 1)$
$^{\mathrm{T}}$ **is outside**	***The rank one product or outer product is*** $\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}$	$(n \times 1)(1 \times n)$

$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ is a number, $\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}$ is a matrix. Quantum mechanics would write those as $< \boldsymbol{x}|\boldsymbol{y} >$ (inner) and $|\boldsymbol{x} >< \boldsymbol{y}|$ (outer). I think the world is governed by linear algebra, but physics disguises it well. Here are examples where the inner product has meaning:

From mechanics	Work = (Movements) (Forces) = $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{f}$
From circuits	Heat loss = (Voltage drops) (Currents) = $\boldsymbol{e}^{\mathrm{T}}\boldsymbol{y}$
From economics	Income = (Quantities) (Prices) = $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{p}$

We are really close to the heart of applied mathematics, and there is one more point to explain. It is the deeper connection between inner products and the transpose of A.

We defined A^{T} by flipping the matrix across its main diagonal. That's not mathematics. There is a better way to approach the transpose. A^{T} *is the matrix that makes these two inner products equal for every* $\boldsymbol{x}$ *and* $\boldsymbol{y}$:

$(A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{y}) \qquad$ **Inner product of** $A\boldsymbol{x}$ **with** $\boldsymbol{y}$ = **Inner product of** $\boldsymbol{x}$ **with** $A^{\mathrm{T}}\boldsymbol{y}$

Example 2 Start with $A = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$ $\quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ $\quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$

On one side we have Ax multiplying y: $(x_2 - x_1)y_1 + (x_3 - x_2)y_2$
That is the same as $x_1(-y_1) + x_2(y_1 - y_2) + x_3(y_2)$. Now x is multiplying $A^{\mathrm{T}}y$.

$$A^{\mathrm{T}}y \text{ must be } \begin{bmatrix} -y_1 \\ y_1 - y_2 \\ y_2 \end{bmatrix} \text{ which produces } A^{\mathrm{T}} = \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \text{ as expected.}$$

Example 3 Will you allow me a little calculus? It is extremely important or I wouldn't leave linear algebra. (This is really linear algebra for functions $x(t)$.) **The difference matrix changes to a derivative** $A = d/dt$. Its transpose will now come from $(dx/dt, y) = (x, -dy/dt)$.

The inner product changes from a finite sum of $x_k y_k$ to an integral of $x(t)y(t)$.

Inner product of functions $$x^{\mathrm{T}}y = (x, y) = \int_{-\infty}^{\infty} x(t)\, y(t)\, dt \quad \text{by definition}$$

Transpose rule $(Ax)^{\mathrm{T}}y = x^{\mathrm{T}}(A^{\mathrm{T}}y)$ $$\int_{-\infty}^{\infty} \frac{dx}{dt} y(t)\, dt = \int_{-\infty}^{\infty} x(t) \left(-\frac{dy}{dt}\right) dt \quad \text{shows } A^{\mathrm{T}} \tag{6}$$

I hope you recognize "***integration by parts***". The derivative moves from the first function $x(t)$ to the second function $y(t)$. During that move, a minus sign appears. This tells us that ***the "transpose" of the derivative is minus the derivative.***

The derivative is *anti-symmetric*: $A = d/dt$ ***and*** $A^{\mathrm{T}} = -d/dt$. Symmetric matrices have $A^{\mathrm{T}} = A$, anti-symmetric matrices have $A^{\mathrm{T}} = -A$. In some way, the 2 by 3 difference matrix above followed this pattern. The 3 by 2 matrix A^{T} was *minus* a difference matrix. It produced $y_1 - y_2$ in the middle component of $A^{\mathrm{T}}y$ instead of the difference $y_2 - y_1$.

Symmetric Matrices

For a *symmetric matrix*, transposing A to A^{T} produces no change. Then $A^{\mathrm{T}} = A$. Its (j, i) entry across the main diagonal equals its (i, j) entry. In my opinion, these are the most important matrices of all.

DEFINITION A *symmetric matrix* has $A^{\mathrm{T}} = A$. This means that $a_{ji} = a_{ij}$.

$$\textbf{Symmetric matrices} \quad A = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} = A^{\mathrm{T}} \quad \text{and} \quad D = \begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix} = D^{\mathrm{T}}.$$

The inverse of a symmetric matrix is also symmetric. The transpose of A^{-1} is $(A^{-1})^{\mathrm{T}} = (A^{\mathrm{T}})^{-1} = A^{-1}$. That says A^{-1} is symmetric (when A is invertible):

$$\textbf{Symmetric inverses} \quad A^{-1} = \begin{bmatrix} 5 & -2 \\ -2 & 1 \end{bmatrix} \quad \text{and} \quad D^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 0.1 \end{bmatrix}.$$

Now we produce symmetric matrices by ***multiplying any matrix R by*** R^{T}.

Symmetric Products $R^{\mathrm{T}}R$ and RR^{T} and LDL^{T}

Choose any matrix R, probably rectangular. Multiply R^{T} times R. Then the product $R^{\mathrm{T}}R$ is automatically a square symmetric matrix:

$$\textit{The transpose of} \quad R^{\mathrm{T}}R \quad \textit{is} \quad R^{\mathrm{T}}(R^{\mathrm{T}})^{\mathrm{T}} \quad \textit{which is} \quad R^{\mathrm{T}}R. \tag{7}$$

That is a quick proof of symmetry for $R^{\mathrm{T}}R$. We could also look at the (i, j) entry of $R^{\mathrm{T}}R$. It is the dot product of row i of R^{T} (column i of R) with column j of R. The (j, i) entry is the same dot product, column j with column i. So $R^{\mathrm{T}}R$ is symmetric.

The matrix RR^{T} is also symmetric. (The shapes of R and R^{T} allow multiplication.) But RR^{T} is a different matrix from $R^{\mathrm{T}}R$. In our experience, most scientific problems that start with a rectangular matrix R end up with $R^{\mathrm{T}}R$ or RR^{T} or both. As in least squares.

Example 4 Multiply $R = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$ and $R^{\mathrm{T}} = \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}$ in both orders.

$$RR^{\mathrm{T}} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \text{ and } R^{\mathrm{T}}R = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \text{ are both symmetric matrices.}$$

The product $R^{\mathrm{T}}R$ is n by n. In the opposite order, RR^{T} is m by m. Both are symmetric, with positive diagonal (*why*?). But even if $m = n$, it is not very likely that $R^{\mathrm{T}}R = RR^{\mathrm{T}}$. Equality can happen, but it is abnormal.

Symmetric matrices in elimination $A^{\mathrm{T}} = A$ makes elimination faster, because we can work with half the matrix (plus the diagonal). It is true that the upper triangular U is probably not symmetric. ***The symmetry is in the triple product*** $A = LDU$. Remember how the diagonal matrix D of pivots can be divided out, to leave 1's on the diagonal of both L and U:

$$\begin{bmatrix} 1 & 2 \\ 2 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \qquad LU \text{ misses the symmetry of } A$$

$$= \begin{bmatrix} \mathbf{1} & 0 \\ \mathbf{2} & \mathbf{1} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{2} \\ 0 & \mathbf{1} \end{bmatrix} \qquad \begin{array}{l} LDU \text{ captures the symmetry} \\ \textit{Now } U \textit{ is the transpose of } L. \end{array}$$

When A is symmetric, the usual form $A = LDU$ becomes $A = LDL^{\mathrm{T}}$. The final U (with 1's on the diagonal) is the transpose of L (also with 1's on the diagonal). The diagonal matrix D containing the pivots is symmetric by itself.

If $A = A^{\mathrm{T}}$ ***is factored into*** LDU ***with no row exchanges, then*** U ***is exactly*** L^{T}.

The symmetric factorization of a symmetric matrix is $A = LDL^{\mathrm{T}}$.

Notice that the transpose of LDL^{T} is automatically $(L^{\mathrm{T}})^{\mathrm{T}}D^{\mathrm{T}}L^{\mathrm{T}}$ which is LDL^{T} again. The work of elimination is cut in half, from $n^3/3$ multiplications to $n^3/6$. The storage is also cut essentially in half. We only keep L and D, not U which is just L^{T}.

Permutation Matrices

The transpose plays a special role for a *permutation matrix*. This matrix P has a single "1" in every row and every column. Then P^{T} is also a permutation matrix—maybe the same or maybe different. Any product $P_1 P_2$ is again a permutation matrix. We now create every P from the identity matrix, by reordering the rows of I.

The simplest permutation matrix is $P = I$ (*no exchanges*). The next simplest are the row exchanges P_{ij}. Those are constructed by exchanging two rows i and j of I. Other permutations reorder more rows. By doing all possible row exchanges to I, we get all possible permutation matrices:

DEFINITION ***A permutation matrix P has the rows of the identity I in any order.***

Example 5 There are six 3 by 3 permutation matrices. Here they are without the zeros:

$$I = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \quad P_{21} = \begin{bmatrix} & 1 & \\ 1 & & \\ & & 1 \end{bmatrix} \quad P_{32}P_{21} = \begin{bmatrix} & 1 & \\ & & 1 \\ 1 & & \end{bmatrix}$$

$$P_{31} = \begin{bmatrix} & & 1 \\ & 1 & \\ 1 & & \end{bmatrix} \quad P_{32} = \begin{bmatrix} 1 & & \\ & & 1 \\ & 1 & \end{bmatrix} \quad P_{21}P_{32} = \begin{bmatrix} & & 1 \\ 1 & & \\ & 1 & \end{bmatrix}.$$

There are $n!$ permutation matrices of order n. The symbol $n!$ means "n factorial," the product of the numbers $(1)(2)\cdots(n)$. Thus $3! = (1)(2)(3)$ which is 6. There will be 24 permutation matrices of order $n = 4$. And 120 permutations of order 5.

There are only two permutation matrices of order 2, namely $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$.

Important: P^{-1} is also a permutation matrix. Among the six 3 by 3 P's displayed above, the four matrices on the left are their own inverses. The two matrices on the right are inverses of each other. In all cases, a single row exchange is its own inverse. If we repeat the exchange we are back to I. But for $P_{32}P_{21}$, the inverses go in opposite order as always. The inverse is $P_{21}P_{32}$.

More important: ***P^{-1} is always the same as P^{T}.*** The two matrices on the right are transposes—and inverses—of each other. When we multiply PP^{T}, the "1" in the first row of P hits the "1" in the first column of P^{T} (since the first row of P is the first column of P^{T}). It misses the ones in all the other columns. So $PP^{\mathrm{T}} = I$.

Another proof of $P^{\mathrm{T}} = P^{-1}$ looks at P as a product of row exchanges. Every row exchange is its own transpose and its own inverse. P^{T} and P^{-1} both come from the product of row exchanges *in reverse order*. So P^{T} and P^{-1} are the same.

Symmetric matrices led to $A = LDL^{\mathrm{T}}$. Now permutations lead to $PA = LU$.

The $PA = LU$ Factorization with Row Exchanges

We sure hope you remember $A = LU$. It started with $A = (E_{21}^{-1} \cdots E_{ij}^{-1} \cdots)U$. Every elimination step was carried out by an E_{ij} and it was inverted by E_{ij}^{-1}. Those inverses were compressed into one matrix L, bringing U back to A. The lower triangular L has 1's on the diagonal, and the result is $A = LU$.

This is a great factorization, but it doesn't always work. Sometimes row exchanges are needed to produce pivots. Then $A = (E^{-1} \cdots P^{-1} \cdots E^{-1} \cdots P^{-1} \cdots)U$. Every row exchange is carried out by a P_{ij} and inverted by that P_{ij}. We now compress those row exchanges into a *single permutation matrix* P. This gives a factorization for every invertible matrix A—which we naturally want.

The main question is where to collect the P_{ij}'s. There are two good possibilities—do all the exchanges before elimination, or do them after the E_{ij}'s. The first way gives $PA = LU$. The second way has a permutation matrix P_1 in the middle.

1. The row exchanges can be done *in advance*. Their product P puts the rows of A in the right order, so that no exchanges are needed for PA. ***Then*** $\boldsymbol{PA = LU}$.

2. If we hold row exchanges until *after elimination*, the pivot rows are in a strange order. P_1 puts them in the correct triangular order in U_1. ***Then*** $\boldsymbol{A = L_1 P_1 U_1}$.

$PA = LU$ is constantly used in all computing (and in MATLAB). ***We will concentrate on this form.*** Most numerical analysts have never seen the other form.

The factorization $A = L_1 P_1 U_1$ might be more elegant. If we mention both, it is because the difference is not well known. Probably you will not spend a long time on either one. Please don't. The most important case has $P = I$, when A equals LU with no exchanges.

For this matrix A, exchange rows 1 and 2 to put the first pivot in its usual place. Then go through elimination on PA:

$$\underset{A}{\begin{bmatrix} 0 & 1 & 1 \\ 1 & 2 & 1 \\ 2 & 7 & 9 \end{bmatrix}} \rightarrow \underset{PA}{\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 2 & 7 & 9 \end{bmatrix}} \rightarrow \underset{\ell_{31} = 2}{\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 3 & 7 \end{bmatrix}} \rightarrow \underset{\ell_{32} = 3}{\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{bmatrix}}.$$

The matrix PA has its rows in good order, and it factors as usual into LU:

$$\boldsymbol{P} = \begin{bmatrix} 0 & \mathbf{1} & 0 \\ \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{1} \end{bmatrix} \quad PA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{bmatrix} = LU. \tag{8}$$

We started with A and ended with U. *The only requirement is invertibility of* A.

If A is invertible, a permutation P will put its rows in the right order to factor $\boldsymbol{PA = LU}$.
There must be a full set of pivots after row exchanges for A to be invertible.

In MATLAB, $A([r\ k],:) = A([k\ r],:)$ exchanges row k with row r below it (where the kth pivot has been found). Then the **lu** code updates L and P and the sign of P:

This is part of $[L, U, P] = \mathsf{lu}(A)$

$$\begin{aligned} &A([r\ k],:) = A([k\ r],:);\\ &L([r\ k], 1:k-1) = L([k\ r], 1:k-1);\\ &P([r\ k],:) = P([k\ r],:);\\ &\text{sign} = -\text{sign} \end{aligned}$$

The "**sign**" of P tells whether the number of row exchanges is even (sign $= +1$). An odd number of row exchanges will produce sign $= -1$. At the start, P is I and sign $= +1$. When there is a row exchange, the sign is reversed. The final value of sign is the **determinant of** P and it does not depend on the order of the row exchanges.

For PA we get back to the familiar LU. This is the usual factorization. In reality, $\mathsf{lu}(A)$ often does not use the first available pivot. Mathematically we accept a small pivot—anything but zero. It is better if the computer looks down the column for the largest pivot. (Section 9.1 explains why this "***partial pivoting***" reduces the roundoff error.) Then P may contain row exchanges that are not algebraically necessary. Still $PA = LU$.

Our advice is to understand permutations but let the computer do the work. Calculations of $A = LU$ are enough to do by hand, without P. The Teaching Code splu(A) factors $PA = LU$ and splv(A, b) solves $Ax = b$ for any invertible A. The program splu stops if no pivot can be found in column k. Then A is not invertible.

■ REVIEW OF THE KEY IDEAS ■

1. The transpose puts the rows of A into the columns of A^{T}. Then $(A^{\mathrm{T}})_{ij} = A_{ji}$.

2. The transpose of AB is $B^{\mathrm{T}}A^{\mathrm{T}}$. The transpose of A^{-1} is the inverse of A^{T}.

3. The dot product is $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$. Then $(A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y}$ equals the dot product $\boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{y})$.

4. When A is symmetric ($A^{\mathrm{T}} = A$), its LDU factorization is symmetric: $A = LDL^{\mathrm{T}}$.

5. A permutation matrix P has a 1 in each row and column, and $\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P}^{-1}$.

6. There are $n!$ permutation matrices of size n. *Half even, half odd.*

7. If A is invertible then a permutation P will reorder its rows for $PA = LU$.

■ WORKED EXAMPLES ■

2.7 A Applying the permutation P to the rows of A destroys its symmetry:

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 4 & 5 \\ 4 & \mathbf{2} & 6 \\ 5 & 6 & \mathbf{3} \end{bmatrix} \qquad PA = \begin{bmatrix} 4 & \mathbf{2} & 6 \\ 5 & 6 & \mathbf{3} \\ 1 & 4 & 5 \end{bmatrix}$$

What permutation Q applied to the *columns* of PA will recover symmetry in PAQ? The numbers $1, 2, 3$ must come back to the main diagonal (not necessarily in order). Show that Q is P^{T}, so that symmetry is saved by $PAQ = PAP^{\mathrm{T}}$.

Solution To recover symmetry and put "2" back on the diagonal, column 2 of PA must move to column 1. Column 3 of PA (containing "3") must move to column 2. Then the "1" moves to the $3, 3$ position. The matrix that permutes columns is Q:

$$PA = \begin{bmatrix} 4 & \mathbf{2} & 6 \\ 5 & 6 & \mathbf{3} \\ 1 & 4 & 5 \end{bmatrix} \qquad Q = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \qquad PAQ = \begin{bmatrix} \mathbf{2} & 6 & 4 \\ 6 & \mathbf{3} & 5 \\ 4 & 5 & 1 \end{bmatrix} \quad \text{is symmetric.}$$

The matrix Q is P^{T}. This choice always recovers symmetry, because PAP^{T} is guaranteed to be symmetric. (Its transpose is again PAP^{T}.) *The matrix Q is also P^{-1}, because the inverse of every permutation matrix is its transpose.*

If D is a diagonal matrix, we are finding that PDP^{T} is also diagonal. When P moves row 1 down to row 3, P^{T} on the right will move column 1 to column 3. The $(1, 1)$ entry moves down to $(3, 1)$ and over to $(3, 3)$.

2.7 B Find the symmetric factorization $A = LDL^{\mathrm{T}}$ for the matrix A above. Is this A invertible? Find also the $PQ = LU$ factorization for Q, which needs row exchanges.

Solution To factor A into LDL^{T} we eliminate below the pivots:

$$A = \begin{bmatrix} 1 & 4 & 5 \\ 4 & 2 & 6 \\ 5 & 6 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 4 & 5 \\ 0 & -14 & -14 \\ 0 & -14 & -22 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 4 & 5 \\ 0 & -14 & -14 \\ 0 & 0 & -8 \end{bmatrix} = U.$$

The multipliers were $\ell_{21} = 4$ and $\ell_{31} = 5$ and $\ell_{32} = 1$. The pivots $1, -14, -8$ go into D. When we divide the rows of U by those pivots, L^{T} should appear:

Symmetric factorization when $A = A^{\mathrm{T}}$

$$A = LDL^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 5 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & -14 & \\ & & -8 \end{bmatrix} \begin{bmatrix} 1 & 4 & 5 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

This matrix A is invertible because *it has three pivots.* Its inverse is $(L^{\mathrm{T}})^{-1}D^{-1}L^{-1}$ and A^{-1} is also symmetric. The numbers 14 and 8 will turn up in the denominators of A^{-1}. The "determinant" of A is the product of the pivots $(1)(-14)(-8) = 112$.

Any permutation matrix Q is invertible. Here elimination needs two row exchanges:

$$Q = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \underset{1 \leftrightarrow 2}{\xrightarrow{\text{rows}}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \underset{2 \leftrightarrow 3}{\xrightarrow{\text{rows}}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I.$$

With $A = Q$, the $PQ = (L)(U)$ factorization is the same as $Q^{-1}Q = (I)(I)$.

2.7 C For a rectangular A, this ***saddle-point matrix*** S is symmetric and important:

Block matrix from least squares $\quad S = \begin{bmatrix} I & A \\ A^T & \mathbf{0} \end{bmatrix} = S^T$ has size $m + n$.

Apply block elimination to find a **block factorization** $S = LDL^T$. Then test invertibility:

S is invertible $\iff$ A^TA is invertible $\iff$ $Ax \neq \mathbf{0}$ whenever $x \neq \mathbf{0}$

Solution The first block pivot is I. The matrix to multiply row 1 is certainly A^T:

Block elimination $\quad S = \begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix}$ goes to $\begin{bmatrix} I & A \\ 0 & -A^TA \end{bmatrix}$. This is U.

The block pivot matrix D contains I and $-A^TA$. Then L and L^T contain A^T and A:

Block factorization $\quad S = LDL^T = \begin{bmatrix} I & 0 \\ A^T & I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & -A^TA \end{bmatrix} \begin{bmatrix} I & A \\ 0 & I \end{bmatrix}.$

L is certainly invertible, with diagonal 1's from I. The inverse of the middle matrix involves $(A^TA)^{-1}$. Section 4.2 answers a key question about the matrix A^TA:

When is A^TA invertible? *Answer: A must have independent columns.*
Then $Ax = \mathbf{0}$ only if $x = \mathbf{0}$. Otherwise $Ax = \mathbf{0}$ will lead to $A^TAx = \mathbf{0}$.

Problem Set 2.7

Questions 1–7 are about the rules for transpose matrices.

1 Find A^T and A^{-1} and $(A^{-1})^T$ and $(A^T)^{-1}$ for

$$A = \begin{bmatrix} 1 & 0 \\ 9 & 3 \end{bmatrix} \quad \text{and also} \quad A = \begin{bmatrix} 1 & c \\ c & 0 \end{bmatrix}.$$

2 Verify that $(AB)^T$ equals B^TA^T but those are different from A^TB^T:

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix} \quad AB = \begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix}.$$

In case $AB = BA$ (not generally true!) how do you prove that $B^TA^T = A^TB^T$?

3 (a) The matrix $((AB)^{-1})^{\mathrm{T}}$ comes from $(A^{-1})^{\mathrm{T}}$ and $(B^{-1})^{\mathrm{T}}$. *In what order*?

(b) If U is upper triangular then $(U^{-1})^{\mathrm{T}}$ is _____ triangular.

4 Show that $A^2 = 0$ is possible but $A^{\mathrm{T}}A = 0$ is not possible (unless $A =$ zero matrix).

5 (a) The row vector $\boldsymbol{x}^{\mathrm{T}}$ times A times the column $\boldsymbol{y}$ produces what number?

$$\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{y} = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \underline{\qquad}.$$

(b) This is the row $\boldsymbol{x}^{\mathrm{T}}A =$ _____ times the column $\boldsymbol{y} = (0, 1, 0)$.

(c) This is the row $\boldsymbol{x}^{\mathrm{T}} = [0 \quad 1]$ times the column $A\boldsymbol{y} =$ _____.

6 The transpose of a block matrix $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ is $M^{\mathrm{T}} =$ _____. Test an example. Under what conditions on A, B, C, D is the block matrix symmetric?

7 True or false:

(a) The block matrix $\begin{bmatrix} 0 & A \\ A & 0 \end{bmatrix}$ is automatically symmetric.

(b) If A and B are symmetric then their product AB is symmetric.

(c) If A is not symmetric then A^{-1} is not symmetric.

(d) When A, B, C are symmetric, the transpose of ABC is CBA.

Questions 8–15 are about permutation matrices.

8 Why are there $n!$ permutation matrices of order n?

9 If P_1 and P_2 are permutation matrices, so is P_1P_2. This still has the rows of I in some order. Give examples with $P_1P_2 \neq P_2P_1$ and $P_3P_4 = P_4P_3$.

10 There are 12 "*even*" permutations of $(1, 2, 3, 4)$, with an *even number of exchanges*. Two of them are $(1, 2, 3, 4)$ with no exchanges and $(4, 3, 2, 1)$ with two exchanges. List the other ten. Instead of writing each 4 by 4 matrix, just order the numbers.

11 Which permutation makes PA upper triangular? Which permutations make P_1AP_2 lower triangular? ***Multiplying A on the right by P_2 exchanges the*** _____ ***of A.***

$$A = \begin{bmatrix} 0 & 0 & 6 \\ 1 & 2 & 3 \\ 0 & 4 & 5 \end{bmatrix}.$$

12 Explain why the dot product of $\boldsymbol{x}$ and $\boldsymbol{y}$ equals the dot product of $P\boldsymbol{x}$ and $P\boldsymbol{y}$. Then from $(P\boldsymbol{x})^{\mathrm{T}}(P\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ deduce that $P^{\mathrm{T}}P = I$ for any permutation. With $\boldsymbol{x} = (1, 2, 3)$ and $\boldsymbol{y} = (1, 4, 2)$ choose P to show that $P\boldsymbol{x} \cdot \boldsymbol{y}$ is not always $\boldsymbol{x} \cdot P\boldsymbol{y}$.

13 (a) Find a 3 by 3 permutation matrix with $P^3 = I$ (but not $P = I$).

(b) Find a 4 by 4 permutation $\widehat{P}$ with $\widehat{P}^4 \neq I$.

14 If P has 1's on the antidiagonal from $(1,n)$ to $(n,1)$, describe PAP. Note $P = P^{\mathrm{T}}$.

15 All row exchange matrices are symmetric: $P^{\mathrm{T}} = P$. Then $P^{\mathrm{T}}P = I$ becomes $P^2 = I$. Other permutation matrices may or may not be symmetric.

(a) If P sends row 1 to row 4, then P^{T} sends row _____ to row _____.
When $P^{\mathrm{T}} = P$ the row exchanges come in pairs with no overlap.

(b) Find a 4 by 4 example with $P^{\mathrm{T}} = P$ that moves all four rows.

Questions 16–21 are about symmetric matrices and their factorizations.

16 If $A = A^{\mathrm{T}}$ and $B = B^{\mathrm{T}}$, which of these matrices are certainly symmetric?

(a) $A^2 - B^2$ (b) $(A+B)(A-B)$ (c) ABA (d) $ABAB$.

17 Find 2 by 2 symmetric matrices $A = A^{\mathrm{T}}$ with these properties:

(a) A is not invertible.

(b) A is invertible but cannot be factored into LU (row exchanges needed).

(c) A can be factored into LDL^{T} but not into LL^{T} (because of negative D).

18 (a) How many entries of A can be chosen independently, if $A = A^{\mathrm{T}}$ is 5 by 5?

(b) How do L and D (still 5 by 5) give the same number of choices in LDL^{T}?

(c) How many entries can be chosen if A is *skew-symmetric*? ($A^{\mathrm{T}} = -A$).

19 Suppose R is rectangular (m by n) and A is symmetric (m by m).

(a) Transpose $R^{\mathrm{T}}AR$ to show its symmetry. What shape is this matrix?

(b) Show why $R^{\mathrm{T}}R$ has no negative numbers on its diagonal.

20 Factor these symmetric matrices into $A = LDL^{\mathrm{T}}$. The pivot matrix D is diagonal:

$$A = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & b \\ b & c \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

21 After elimination clears out column 1 below the first pivot, find the symmetric 2 by 2 matrix that appears in the lower right corner:

$$\text{Start from } A = \begin{bmatrix} 2 & 4 & 8 \\ 4 & 3 & 9 \\ 8 & 9 & 0 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & b & c \\ b & d & e \\ c & e & f \end{bmatrix}.$$

Questions 22–24 are about the factorizations $PA = LU$ and $A = L_1P_1U_1$.

22 Find the $PA = LU$ factorizations (and check them) for

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 2 & 3 & 4 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

23 Find a 4 by 4 permutation matrix (call it A) that needs 3 row exchanges to reach the end of elimination. For this matrix, what are its factors P, L, and U?

24 Factor the following matrix into $PA = LU$. Factor it also into $A = L_1P_1U_1$ (hold the exchange of row 3 until 3 times row 1 is subtracted from row 2):

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 3 & 8 \\ 2 & 1 & 1 \end{bmatrix}.$$

25 Extend the **slu** code in Section 2.6 to a code **splu** that factors PA into LU.

26 Prove that the identity matrix cannot be the product of three row exchanges (or five). It can be the product of two exchanges (or four).

27 (a) Choose E_{21} to remove the 3 below the first pivot. Then multiply $E_{21}AE_{21}^{\mathrm{T}}$ to remove both 3's:

$$A = \begin{bmatrix} 1 & 3 & 0 \\ 3 & 11 & 4 \\ 0 & 4 & 9 \end{bmatrix} \quad \text{is going toward} \quad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

(b) Choose E_{32} to remove the 4 below the second pivot. Then A is reduced to D by $E_{32}E_{21}AE_{21}^{\mathrm{T}}E_{32}^{\mathrm{T}} = D$. Invert the E's to find L in $A = LDL^{\mathrm{T}}$.

28 If every row of a 4 by 4 matrix contains the numbers $0, 1, 2, 3$ in some order, can the matrix be symmetric?

29 Prove that no reordering of rows and reordering of columns can transpose a typical matrix. (Watch the diagonal entries.)

The next three questions are about applications of the identity $(Ax)^{\mathrm{T}}y = x^{\mathrm{T}}(A^{\mathrm{T}}y)$.

30 Wires go between Boston, Chicago, and Seattle. Those cities are at voltages x_B, x_C, x_S. With unit resistances between cities, the currents between cities are in y:

$$y = Ax \quad \text{is} \quad \begin{bmatrix} y_{BC} \\ y_{CS} \\ y_{BS} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_B \\ x_C \\ x_S \end{bmatrix}.$$

(a) Find the total currents $A^{\mathrm{T}}y$ out of the three cities.

(b) Verify that $(Ax)^{\mathrm{T}}y$ agrees with $x^{\mathrm{T}}(A^{\mathrm{T}}y)$—six terms in both.

31 Producing x_1 trucks and x_2 planes needs $x_1 + 50x_2$ tons of steel, $40x_1 + 1000x_2$ pounds of rubber, and $2x_1 + 50x_2$ months of labor. If the unit costs y_1, y_2, y_3 are \$700 per ton, \$3 per pound, and \$3000 per month, what are the values of one truck and one plane? Those are the components of $A^{\mathrm{T}}\boldsymbol{y}$.

32 $A\boldsymbol{x}$ gives the amounts of steel, rubber, and labor to produce $\boldsymbol{x}$ in Problem 31. Find A. Then $A\boldsymbol{x} \cdot \boldsymbol{y}$ is the _____ of inputs while $\boldsymbol{x} \cdot A^{\mathrm{T}}\boldsymbol{y}$ is the value of _____.

33 The matrix P that multiplies (x, y, z) to give (z, x, y) is also a rotation matrix. Find P and P^3. The rotation axis $\boldsymbol{a} = (1, 1, 1)$ doesn't move, it equals $P\boldsymbol{a}$. What is the angle of rotation from $\boldsymbol{v} = (2, 3, -5)$ to $P\boldsymbol{v} = (-5, 2, 3)$?

34 Write $A = \left[\begin{smallmatrix} 1 & 2 \\ 4 & 9 \end{smallmatrix}\right]$ as the product EH of an elementary row operation matrix E and a symmetric matrix H.

35 Here is a new factorization of A into *triangular* (with 1's) *times symmetric*:

$$\text{Start from } A = LDU. \text{ Then } A = L(U^{\mathrm{T}})^{-1} \text{ times } U^{\mathrm{T}}DU.$$

Why is $L(U^{\mathrm{T}})^{-1}$ triangular? Its diagonal is all 1's. Why is $U^{\mathrm{T}}DU$ symmetric?

36 A *group* of matrices includes AB and A^{-1} if it includes A and B. "Products and inverses stay in the group." Which of these sets are groups?
Lower triangular matrices L with 1's on the diagonal, symmetric matrices S, positive matrices M, diagonal invertible matrices D, permutation matrices P, matrices with $Q^{\mathrm{T}} = Q^{-1}$. ***Invent two more matrix groups.***

Challenge Problems

37 A square ***northwest matrix B*** is zero in the southeast corner, below the antidiagonal that connects $(1, n)$ to $(n, 1)$. Will B^{T} and B^2 be northwest matrices? Will B^{-1} be northwest or southeast? What is the shape of BC = ***northwest times southeast***?

38 If you take powers of a permutation matrix, why is some P^k eventually equal to I?
Find a 5 by 5 permutation P so that the smallest power to equal I is P^6.

39 (a) Write down any 3 by 3 matrix A. Split A into $B + C$ where $B = B^{\mathrm{T}}$ is symmetric and $C = -C^{\mathrm{T}}$ is anti-symmetric.

(b) Find formulas for B and C involving A and A^{T}. We want $A = B + C$ with $B = B^{\mathrm{T}}$ and $C = -C^{\mathrm{T}}$.

40 Suppose Q^{T} *equals* Q^{-1} (transpose equals inverse, so $Q^{\mathrm{T}}Q = I$).

(a) Show that the columns $q_1, \ldots, q_n$ are unit vectors: $\|\boldsymbol{q}_i\|^2 = 1$.

(b) Show that every two columns of Q are perpendicular: $\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{q}_2 = 0$.

(c) Find a 2 by 2 example with first entry $q_{11} = \cos\theta$.

Chapter 3

Vector Spaces and Subspaces

3.1 Spaces of Vectors

To a newcomer, matrix calculations involve a lot of numbers. To you, they involve vectors. The columns of Ax and AB are linear combinations of n vectors—the columns of A. This chapter moves from numbers and vectors to a third level of understanding (the highest level). Instead of individual columns, we look at "spaces" of vectors. Without seeing ***vector spaces*** and especially their ***subspaces***, you haven't understood everything about $Ax = b$.

Since this chapter goes a little deeper, it may seem a little harder. That is natural. We are looking inside the calculations, to find the mathematics. The author's job is to make it clear. The chapter ends with the "*Fundamental Theorem of Linear Algebra*".

We begin with the most important vector spaces. They are denoted by $\mathbf{R}^1$, $\mathbf{R}^2$, $\mathbf{R}^3$, $\mathbf{R}^4$, ... Each space $\mathbf{R}^n$ consists of a whole collection of vectors. $\mathbf{R}^5$ contains all column vectors with five components. This is called "5-dimensional space".

DEFINITION *The space* $\mathbf{R}^n$ *consists of all column vectors* v *with* n *components.*

The components of v are real numbers, which is the reason for the letter $\mathbf{R}$. A vector whose n components are complex numbers lies in the space $\mathbf{C}^n$.

The vector space $\mathbf{R}^2$ is represented by the usual xy plane. Each vector v in $\mathbf{R}^2$ has two components. The word "*space*" asks us to think of all those vectors—the whole plane. Each vector gives the x and y coordinates of a point in the plane: $v = (x, y)$.

Similarly the vectors in $\mathbf{R}^3$ correspond to points (x, y, z) in three-dimensional space. The one-dimensional space $\mathbf{R}^1$ is a line (like the x axis). As before, we print vectors as a column between brackets, or along a line using commas and parentheses:

$$\begin{bmatrix} 4 \\ \pi \end{bmatrix} \text{ is in } \mathbf{R}^2, \quad (1,1,0,1,1) \text{ is in } \mathbf{R}^5, \quad \begin{bmatrix} 1+i \\ 1-i \end{bmatrix} \text{ is in } \mathbf{C}^2.$$

The great thing about linear algebra is that it deals easily with five-dimensional space. We don't draw the vectors, we just need the five numbers (or n numbers).

To multiply v by 7, multiply every component by 7. Here 7 is a "scalar". To add vectors in $\mathbf{R}^5$, add them a component at a time. The two essential vector operations go on *inside the vector space*, and they produce ***linear combinations***:

We can add any vectors in $\mathbf{R}^n$***, and we can multiply any vector*** v ***by any scalar*** c.

"Inside the vector space" means that ***the result stays in the space.*** If v is the vector in $\mathbf{R}^4$ with components $1, 0, 0, 1$, then $2v$ is the vector in $\mathbf{R}^4$ with components $2, 0, 0, 2$. (In this case 2 is the scalar.) A whole series of properties can be verified in $\mathbf{R}^n$. The commutative law is $v + w = w + v$; the distributive law is $c(v + w) = cv + cw$. There is a unique "zero vector" satisfying $\mathbf{0} + v = v$. Those are three of the eight conditions listed at the start of the problem set.

These eight conditions are required of every vector space. There are vectors other than column vectors, and vector spaces other than $\mathbf{R}^n$, and all vector spaces have to obey the eight reasonable rules.

A real vector space is a set of "vectors" *together with rules for vector addition and for multiplication by real numbers.* The addition and the multiplication must produce vectors that are in the space. And the eight conditions must be satisfied (which is usually no problem). Here are three vector spaces other than $\mathbf{R}^n$:

M	The vector space of ***all real 2 by 2 matrices.***
F	The vector space of ***all real functions*** $f(x)$.
Z	The vector space that consists only of a ***zero vector.***

In **M** the "vectors" are really matrices. In **F** the vectors are functions. In **Z** the only addition is $\mathbf{0} + \mathbf{0} = \mathbf{0}$. In each case we can add: matrices to matrices, functions to functions, zero vector to zero vector. We can multiply a matrix by 4 or a function by 4 or the zero vector by 4. The result is still in **M** or **F** or **Z**. The eight conditions are all easily checked.

The function space **F** is infinite-dimensional. A smaller function space is **P**, or $\mathbf{P}_n$, containing all polynomials $a_0 + a_1x + \cdots + a_nx^n$ of degree n.

The space **Z** is zero-dimensional (by any reasonable definition of dimension). It is the smallest possible vector space. We hesitate to call it $\mathbf{R}^0$, which means no components—you might think there was no vector. The vector space **Z** contains exactly *one vector* (zero). No space can do without that zero vector. Each space has its own zero vector—the zero matrix, the zero function, the vector $(0, 0, 0)$ in $\mathbf{R}^3$.

Subspaces

At different times, we will ask you to think of matrices and functions as vectors. But at all times, the vectors that we need most are ordinary column vectors. They are vectors with n components—but *maybe not all* of the vectors with n components. There are important vector spaces ***inside*** $\mathbf{R}^n$. Those are ***subspaces*** of $\mathbf{R}^n$.

Start with the usual three-dimensional space $\mathbf{R}^3$. Choose a plane through the origin $(0, 0, 0)$. ***That plane is a vector space in its own right.*** If we add two vectors in the plane, their sum is in the plane. If we multiply an in-plane vector by 2 or -5, it is still in the plane.

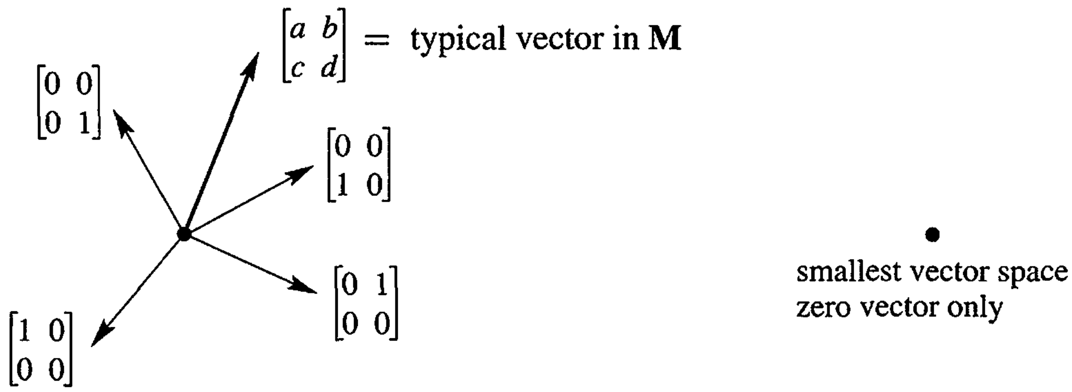


Figure 3.1: "Four-dimensional" matrix space $\mathbf{M}$. The "zero-dimensional" space $\mathbf{Z}$.

A plane in three-dimensional space is not $\mathbf{R}^2$ (even if it looks like $\mathbf{R}^2$). The vectors have three components and they belong to $\mathbf{R}^3$. The plane is a vector space *inside* $\mathbf{R}^3$.

This illustrates one of the most fundamental ideas in linear algebra. The plane going through $(0, 0, 0)$ is a ***subspace*** of the full vector space $\mathbf{R}^3$.

DEFINITION A ***subspace*** of a vector space is a set of vectors (including $\mathbf{0}$) that satisfies two requirements: ***If v and w are vectors in the subspace and c is any scalar, then***

(i) $v + w$ is in the subspace

(ii) cv is in the subspace.

In other words, the set of vectors is "closed" under addition $v + w$ and multiplication cv (and cw). Those operations leave us in the subspace. We can also subtract, because $-w$ is in the subspace and its sum with v is $v - w$. In short, ***all linear combinations stay in the subspace.***

All these operations follow the rules of the host space, so the eight required conditions are automatic. We just have to check the requirements for a subspace, so that we can take linear combinations.

First fact: ***Every subspace contains the zero vector.*** The plane in $\mathbf{R}^3$ has to go through $(0, 0, 0)$. We mention this separately, for extra emphasis, but it follows directly from rule (**ii**). Choose $c = 0$, and the rule requires $0v$ to be in the subspace.

Planes that don't contain the origin fail those tests. When v is on such a plane, $-v$ and $0v$ are *not* on the plane. A plane that misses the origin is not a subspace.

Lines through the origin are also subspaces. When we multiply by 5, or add two vectors on the line, we stay on the line. But the line must go through $(0, 0, 0)$.

Another subspace is all of $\mathbf{R}^3$. The whole space is a subspace (*of itself*). Here is a list of all the possible subspaces of $\mathbf{R}^3$:

(**L**) Any line through $(0, 0, 0)$	($\mathbf{R}^3$) The whole space
(**P**) Any plane through $(0, 0, 0)$	(**Z**) The single vector $(0, 0, 0)$

If we try to keep only *part* of a plane or line, the requirements for a subspace don't hold. Look at these examples in $\mathbf{R}^2$.

Example 1 Keep only the vectors (x, y) whose components are positive or zero (this is a quarter-plane). The vector $(2, 3)$ is included but $(-2, -3)$ is not. So rule (**ii**) is violated when we try to multiply by $c = -1$. ***The quarter-plane is not a subspace.***

Example 2 Include also the vectors whose components are both negative. Now we have two quarter-planes. Requirement (**ii**) is satisfied; we can multiply by any c. But rule (**i**) now fails. The sum of $\boldsymbol{v} = (2, 3)$ and $\boldsymbol{w} = (-3, -2)$ is $(-1, 1)$, which is outside the quarter-planes. ***Two quarter-planes don't make a subspace.***

Rules (i) and (ii) involve vector addition $\boldsymbol{v} + \boldsymbol{w}$ and multiplication by scalars like c and d. The rules can be combined into a single requirement—*the rule for subspaces*:

A subspace containing $\boldsymbol{v}$ and $\boldsymbol{w}$ must contain all linear combinations $c\boldsymbol{v} + d\boldsymbol{w}$.

Example 3 Inside the vector space **M** of all 2 by 2 matrices, here are two subspaces:

$$(\mathbf{U}) \text{ All upper triangular matrices } \begin{bmatrix} a & b \\ 0 & d \end{bmatrix} \quad (\mathbf{D}) \text{ All diagonal matrices } \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}.$$

Add any two matrices in **U**, and the sum is in **U**. Add diagonal matrices, and the sum is diagonal. In this case **D** is also a subspace of **U**! Of course the zero matrix is in these subspaces, when a, b, and d all equal zero.

To find a smaller subspace of diagonal matrices, we could require $a = d$. The matrices are multiples of the identity matrix I. The sum $2I + 3I$ is in this subspace, and so is 3 times $4I$. The matrices cI form a "line of matrices" inside **M** and **U** and **D**.

Is the matrix I a subspace by itself? Certainly not. Only the zero matrix is. Your mind will invent more subspaces of 2 by 2 matrices—write them down for Problem 5.

The Column Space of A

The most important subspaces are tied directly to a matrix A. We are trying to solve $A\boldsymbol{x} = \boldsymbol{b}$. If A is not invertible, the system is solvable for some $\boldsymbol{b}$ and not solvable for other $\boldsymbol{b}$. We want to describe the good right sides $\boldsymbol{b}$—the vectors that *can* be written as A times some vector $\boldsymbol{x}$. Those $\boldsymbol{b}'s$ form the "*column space*" of A.

Remember that $A\boldsymbol{x}$ is a combination of the columns of A. To get every possible $\boldsymbol{b}$, we use every possible $\boldsymbol{x}$. So start with the columns of A, and ***take all their linear combinations. This produces the column space of A.*** **It is a vector space made up of column vectors.**

$\boldsymbol{C}(A)$ contains not just the n columns of A, but all their combinations $A\boldsymbol{x}$.

DEFINITION The ***column space*** consists of ***all linear combinations of the columns.*** The combinations are all possible vectors Ax. They fill the column space $C(A)$.

This column space is crucial to the whole book, and here is why. ***To solve $Ax = b$ is to express b as a combination of the columns.*** The right side b has to be *in the column space* produced by A on the left side, or no solution!

The system $Ax = b$ is solvable if and only if b is in the column space of A.

When b is in the column space, it is a combination of the columns. The coefficients in that combination give us a solution x to the system $Ax = b$.

Suppose A is an m by n matrix. Its columns have m components (not n). So the columns belong to $\mathbf{R}^m$. ***The column space of A is a subspace of $\mathbf{R}^m$ (not $\mathbf{R}^n$).*** The set of all column combinations Ax satisfies rules (i) and (ii) for a subspace: When we add linear combinations or multiply by scalars, we still produce combinations of the columns. The word "subspace" is justified *by taking all linear combinations.*

Here is a 3 by 2 matrix A, whose column space is a subspace of $\mathbf{R}^3$. The column space of A is a plane in Figure 3.2.

Example 4

$$Ax \quad \text{is} \quad \begin{bmatrix} 1 & 0 \\ 4 & 3 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{which is} \quad x_1 \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 3 \\ 3 \end{bmatrix}.$$

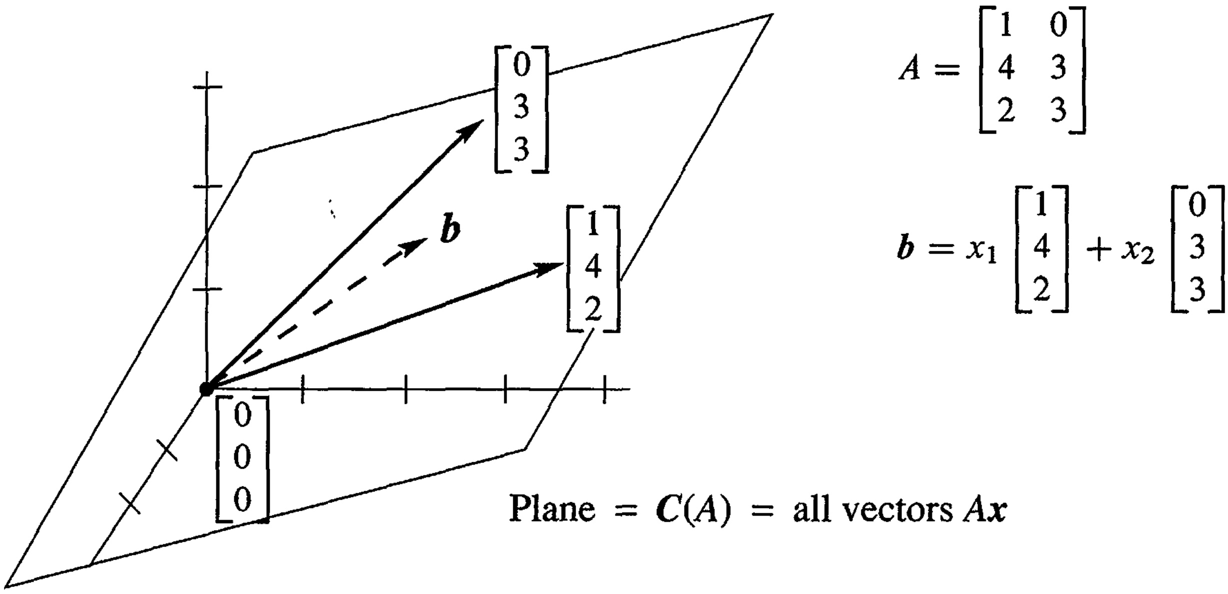


Figure 3.2: The column space $C(A)$ is a plane containing the two columns. $Ax = b$ is solvable when b is on that plane. Then b is a combination of the columns.

The column space of all combinations of the two columns *fills up a plane in* $\mathbf{R}^3$. We drew one particular $\boldsymbol{b}$ (a combination of the columns). This $\boldsymbol{b} = A\boldsymbol{x}$ lies on the plane. The plane has zero thickness, so most right sides $\boldsymbol{b}$ in $\mathbf{R}^3$ are *not* in the column space. For most $\boldsymbol{b}$ there is no solution to our 3 equations in 2 unknowns.

Of course $(0, 0, 0)$ is in the column space. The plane passes through the origin. There is certainly a solution to $A\boldsymbol{x} = \mathbf{0}$. That solution, always available, is $\boldsymbol{x} =$ ____.

To repeat, the attainable right sides $\boldsymbol{b}$ are exactly the vectors in the column space. One possibility is the first column itself—take $x_1 = 1$ and $x_2 = 0$. Another combination is the second column—take $x_1 = 0$ and $x_2 = 1$. The new level of understanding is to see *all* combinations—the whole subspace is generated by those two columns.

Notation The column space of A is denoted by $\boldsymbol{C}(A)$. Start with the columns and take all their linear combinations. We might get the whole $\mathbf{R}^m$ or only a subspace.

Important Instead of columns in $\mathbf{R}^m$, we could start with any set $\mathbf{S}$ of vectors in a vector space $\mathbf{V}$. To get a subspace $\mathbf{SS}$ of $\mathbf{V}$, we take *all combinations* of the vectors in that set:

$$\begin{aligned} \mathbf{S} &= \text{set of vectors in } \mathbf{V} \text{ (probably } \textit{not} \text{ a subspace)} \\ \mathbf{SS} &= \text{all combinations of vectors in } \mathbf{S} \end{aligned}$$

$$\mathbf{SS} = \text{all } c_1\boldsymbol{v}_1 + \cdots + c_N\boldsymbol{v}_N = \textbf{the subspace of V "spanned" by S}$$

When $\mathbf{S}$ is the set of columns, $\mathbf{SS}$ is the column space. When there is only one nonzero vector $\boldsymbol{v}$ in $\mathbf{S}$, the subspace $\mathbf{SS}$ is the line through $\boldsymbol{v}$. *Always* $\mathbf{SS}$ *is the smallest subspace containing* $\mathbf{S}$. This is a fundamental way to create subspaces and we will come back to it.

The subspace SS is the "span" of S, containing all combinations of vectors in S.

Example 5 Describe the column spaces (they are subspaces of $\mathbf{R}^2$) for

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 4 \end{bmatrix}.$$

Solution The column space of I is the *whole space* $\mathbf{R}^2$. Every vector is a combination of the columns of I. In vector space language, $\boldsymbol{C}(I)$ is $\mathbf{R}^2$.

The column space of A is only a line. The second column $(2, 4)$ is a multiple of the first column $(1, 2)$. Those vectors are different, but our eye is on vector ***spaces***. The column space contains $(1, 2)$ and $(2, 4)$ and all other vectors $(c, 2c)$ along that line. The equation $A\boldsymbol{x} = \boldsymbol{b}$ is only solvable when $\boldsymbol{b}$ is on the line.

For the third matrix (with three columns) the column space $\boldsymbol{C}(B)$ is all of $\mathbf{R}^2$. Every $\boldsymbol{b}$ is attainable. The vector $\boldsymbol{b} = (5, 4)$ is column 2 plus column 3, so $\boldsymbol{x}$ can be $(0, 1, 1)$. The same vector $(5, 4)$ is also 2(column 1) + column 3, so another possible $\boldsymbol{x}$ is $(2, 0, 1)$. This matrix has the same column space as I—any $\boldsymbol{b}$ is allowed. But now $\boldsymbol{x}$ has extra components and there are more solutions—more combinations that give $\boldsymbol{b}$.

The next section creates a vector space $\boldsymbol{N}(A)$, to describe all the solutions of $A\boldsymbol{x} = \mathbf{0}$. This section created the column space $\boldsymbol{C}(A)$, to describe all the attainable right sides $\boldsymbol{b}$.

■ REVIEW OF THE KEY IDEAS ■

1. $\mathbf{R}^n$ contains all column vectors with n real components.
2. $\mathbf{M}$ (2 by 2 matrices) and $\mathbf{F}$ (functions) and $\mathbf{Z}$ (zero vector alone) are vector spaces.
3. A subspace containing $\boldsymbol{v}$ and $\boldsymbol{w}$ must contain all their combinations $c\boldsymbol{v} + d\boldsymbol{w}$.
4. The combinations of the columns of A form the ***column space*** $\boldsymbol{C(A)}$. Then the column space is "spanned" by the columns.
5. $A\boldsymbol{x} = \boldsymbol{b}$ has a solution exactly when $\boldsymbol{b}$ is in the column space of A.

■ WORKED EXAMPLES ■

3.1 A We are given three different vectors $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3$. Construct a matrix so that the equations $A\boldsymbol{x} = \boldsymbol{b}_1$ and $A\boldsymbol{x} = \boldsymbol{b}_2$ are solvable, but $A\boldsymbol{x} = \boldsymbol{b}_3$ is not solvable. How can you decide if this is possible? How could you construct A?

Solution We want to have $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ in the column space of A. Then $A\boldsymbol{x} = \boldsymbol{b}_1$ and $A\boldsymbol{x} = \boldsymbol{b}_2$ will be solvable. *The quickest way is to make $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ the two columns of A.* Then the solutions are $\boldsymbol{x} = (1, 0)$ and $\boldsymbol{x} = (0, 1)$.

Also, we don't want $A\boldsymbol{x} = \boldsymbol{b}_3$ to be solvable. So don't make the column space any larger! Keeping only the columns of $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$, the question is:

$$\text{Is } A\boldsymbol{x} = \begin{bmatrix} & & \\ \boldsymbol{b}_1 & \boldsymbol{b}_2 \\ & & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \boldsymbol{b}_3 \text{ solvable?} \qquad \text{Is } \boldsymbol{b}_3 \text{ a combination of } \boldsymbol{b}_1 \text{ and } \boldsymbol{b}_2\text{?}$$

If the answer is *no*, we have the desired matrix A. If the answer is *yes*, then it is *not possible* to construct A. When the column space contains $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$, it will have to contain all their linear combinations. So $\boldsymbol{b}_3$ would necessarily be in that column space and $A\boldsymbol{x} = \boldsymbol{b}_3$ would necessarily be solvable.

3.1 B Describe a subspace $\mathbf{S}$ of each vector space $\mathbf{V}$, and then a subspace $\mathbf{SS}$ of $\mathbf{S}$.

$$\begin{aligned}
\mathbf{V}_1 &= \text{all combinations of } (1,1,0,0) \text{ and } (1,1,1,0) \text{ and } (1,1,1,1) \\
\mathbf{V}_2 &= \text{all vectors perpendicular to } \boldsymbol{u} = (1,2,1), \text{ so } \boldsymbol{u} \cdot \boldsymbol{v} = 0 \\
\mathbf{V}_3 &= \text{all symmetric 2 by 2 matrices (a subspace of } \mathbf{M}) \\
\mathbf{V}_4 &= \text{all solutions to the equation } d^4y/dx^4 = 0 \text{ (a subspace of } \mathbf{F})
\end{aligned}$$

Describe each $\mathbf{V}$ two ways: *All combinations of ...., all solutions of the equations ....*

Solution $\mathbf{V}_1$ starts with three vectors. A subspace $\mathbf{S}$ comes from all combinations of the first two vectors $(1,1,0,0)$ and $(1,1,1,0)$. A subspace $\mathbf{SS}$ of $\mathbf{S}$ comes from all multiples $(c,c,0,0)$ of the first vector. So many possibilities.

A subspace $\mathbf{S}$ of $\mathbf{V}_2$ is the line through $(1,-1,1)$. This line is perpendicular to $\boldsymbol{u}$. The vector $\boldsymbol{x} = (0,0,0)$ is in $\mathbf{S}$ and all its multiples $c\boldsymbol{x}$ give the smallest subspace $\mathbf{SS} = \mathbf{Z}$.

The diagonal matrices are a subspace $\mathbf{S}$ of the symmetric matrices. The multiples cI are a subspace $\mathbf{SS}$ of the diagonal matrices.

$\mathbf{V}_4$ contains all cubic polynomials $y = a + bx + cx^2 + dx^3$, with $d^4y/dx^4 = 0$. The quadratic polynomials give a subspace $\mathbf{S}$. The linear polynomials are one choice of $\mathbf{SS}$. The constants could be $\mathbf{SSS}$.

In all four parts we could take $\mathbf{S} = \mathbf{V}$ itself, and $\mathbf{SS}$ = the zero subspace $\mathbf{Z}$.

Each $\mathbf{V}$ can be described as *all combinations of ....* and as *all solutions of ....*:

$\mathbf{V}_1$ = all combinations of the 3 vectors $\qquad$ $\mathbf{V}_1$ = all solutions of $v_1 - v_2 = 0$

$\mathbf{V}_2$ = all combinations of $(1,0,-1)$ and $(1,-1,1)$ are solutions of $\boldsymbol{u}\cdot\boldsymbol{v} = 0$.

$\mathbf{V}_3$ = all combinations of $\begin{bmatrix}1&0\\0&0\end{bmatrix}, \begin{bmatrix}0&1\\1&0\end{bmatrix}, \begin{bmatrix}0&0\\0&1\end{bmatrix}$. $\mathbf{V}_3$ = all solutions $\begin{bmatrix}a&b\\c&d\end{bmatrix}$ of $b = c$

$\mathbf{V}_4$ = all combinations of $1, x, x^2, x^3$ $\qquad$ $\mathbf{V}_4$ = all solutions to $d^4y/dx^4 = 0$.

Problem Set 3.1

The first problems 1–8 are about vector spaces in general. The vectors in those spaces are not necessarily column vectors. In the definition of a *vector space*, vector addition $\boldsymbol{x}+\boldsymbol{y}$ and scalar multiplication $c\boldsymbol{x}$ must obey the following eight rules:

(1) $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$

(2) $\boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z}) = (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z}$

(3) There is a unique "zero vector" such that $\boldsymbol{x} + \mathbf{0} = \boldsymbol{x}$ for all $\boldsymbol{x}$

(4) For each $\boldsymbol{x}$ there is a unique vector $-\boldsymbol{x}$ such that $\boldsymbol{x} + (-\boldsymbol{x}) = \mathbf{0}$

(5) 1 times $\boldsymbol{x}$ equals $\boldsymbol{x}$

(6) $(c_1c_2)\boldsymbol{x} = c_1(c_2\boldsymbol{x})$

(7) $c(\boldsymbol{x} + \boldsymbol{y}) = c\boldsymbol{x} + c\boldsymbol{y}$

(8) $(c_1 + c_2)\boldsymbol{x} = c_1\boldsymbol{x} + c_2\boldsymbol{x}$.

1 Suppose $(x_1, x_2) + (y_1, y_2)$ is defined to be $(x_1 + y_2, x_2 + y_1)$. With the usual multiplication $c\boldsymbol{x} = (cx_1, cx_2)$, which of the eight conditions are not satisfied?

2 Suppose the multiplication $c\boldsymbol{x}$ is defined to produce $(cx_1, 0)$ instead of (cx_1, cx_2). With the usual addition in $\mathbf{R}^2$, are the eight conditions satisfied?

3 (a) Which rules are broken if we keep only the positive numbers $x > 0$ in $\mathbf{R}^1$? Every c must be allowed. The half-line is not a subspace.

(b) The positive numbers with $\boldsymbol{x} + \boldsymbol{y}$ and $c\boldsymbol{x}$ redefined to equal the usual xy and x^c *do* satisfy the eight rules. Test rule 7 when $c = 3, x = 2, y = 1$. (Then $\boldsymbol{x} + \boldsymbol{y} = 2$ and $c\boldsymbol{x} = 8$.) Which number acts as the "zero vector"?

4 The matrix $A = \begin{bmatrix} 2 & -2 \\ 2 & -2 \end{bmatrix}$ is a "vector" in the space $\mathbf{M}$ of all 2 by 2 matrices. Write down the zero vector in this space, the vector $\frac{1}{2}A$, and the vector $-A$. What matrices are in the smallest subspace containing A?

5 (a) Describe a subspace of $\mathbf{M}$ that contains $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ but not $B = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$.

(b) If a subspace of $\mathbf{M}$ contains A and B, must it contain I?

(c) Describe a subspace of $\mathbf{M}$ that contains no nonzero diagonal matrices.

6 The functions $f(x) = x^2$ and $g(x) = 5x$ are "vectors" in $\mathbf{F}$. This is the vector space of all real functions. (The functions are defined for $-\infty < x < \infty$.) The combination $3f(x) - 4g(x)$ is the function $h(x) =$ _____.

7 Which rule is broken if multiplying $f(x)$ by c gives the function $f(cx)$? Keep the usual addition $f(x) + g(x)$.

8 If the sum of the "vectors" $f(x)$ and $g(x)$ is defined to be the function $f(g(x))$, then the "zero vector" is $g(x) = x$. Keep the usual scalar multiplication $c f(x)$ and find two rules that are broken.

Questions 9–18 are about the "subspace requirements": $\boldsymbol{x} + \boldsymbol{y}$ and $c\boldsymbol{x}$ (and then all linear combinations $c\boldsymbol{x} + d\boldsymbol{y}$) stay in the subspace.

9 One requirement can be met while the other fails. Show this by finding

(a) A set of vectors in $\mathbf{R}^2$ for which $\boldsymbol{x} + \boldsymbol{y}$ stays in the set but $\frac{1}{2}\boldsymbol{x}$ may be outside.

(b) A set of vectors in $\mathbf{R}^2$ (other than two quarter-planes) for which every $c\boldsymbol{x}$ stays in the set but $\boldsymbol{x} + \boldsymbol{y}$ may be outside.

10 Which of the following subsets of $\mathbf{R}^3$ are actually subspaces?

(a) The plane of vectors (b_1, b_2, b_3) with $b_1 = b_2$.

(b) The plane of vectors with $b_1 = 1$.

(c) The vectors with $b_1 b_2 b_3 = 0$.

(d) All linear combinations of $\boldsymbol{v} = (1, 4, 0)$ and $\boldsymbol{w} = (2, 2, 2)$.

(e) All vectors that satisfy $b_1 + b_2 + b_3 = 0$.

(f) All vectors with $b_1 \le b_2 \le b_3$.

11 Describe the smallest subspace of the matrix space $\mathbf{M}$ that contains

(a) $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

12 Let P be the plane in $\mathbf{R}^3$ with equation $x + y - 2z = 4$. The origin $(0, 0, 0)$ is not in P! Find two vectors in P and check that their sum is not in P.

13 Let $\mathbf{P}_0$ be the plane through $(0, 0, 0)$ parallel to the previous plane P. What is the equation for $\mathbf{P}_0$? Find two vectors in $\mathbf{P}_0$ and check that their sum is in $\mathbf{P}_0$.

14 The subspaces of $\mathbf{R}^3$ are planes, lines, $\mathbf{R}^3$ itself, or $\mathbf{Z}$ containing only $(0, 0, 0)$.

(a) Describe the three types of subspaces of $\mathbf{R}^2$.

(b) Describe all subspaces of $\mathbf{D}$, the space of 2 by 2 diagonal matrices.

15 (a) The intersection of two planes through $(0, 0, 0)$ is probably a _____ but it could be a _____. It can't be $\mathbf{Z}$!

(b) The intersection of a plane through $(0, 0, 0)$ with a line through $(0, 0, 0)$ is probably a _____ but it could be a _____.

(c) If $\mathbf{S}$ and $\mathbf{T}$ are subspaces of $\mathbf{R}^5$, prove that their intersection $\mathbf{S} \cap \mathbf{T}$ is a subspace of $\mathbf{R}^5$. Here $\mathbf{S} \cap \mathbf{T}$ consists of the vectors that lie in both subspaces. *Check the requirements on* $\boldsymbol{x} + \boldsymbol{y}$ *and* $c\boldsymbol{x}$.

16 Suppose $\mathbf{P}$ is a plane through $(0, 0, 0)$ and $\mathbf{L}$ is a line through $(0, 0, 0)$. The smallest vector space containing both $\mathbf{P}$ and $\mathbf{L}$ is either _____ or _____.

17 (a) Show that the set of *invertible* matrices in $\mathbf{M}$ is not a subspace.

(b) Show that the set of *singular* matrices in $\mathbf{M}$ is not a subspace.

18 True or false (check addition in each case by an example):

(a) The symmetric matrices in $\mathbf{M}$ (with $A^{\mathrm{T}} = A$) form a subspace.

(b) The skew-symmetric matrices in $\mathbf{M}$ (with $A^{\mathrm{T}} = -A$) form a subspace.

(c) The unsymmetric matrices in $\mathbf{M}$ (with $A^{\mathrm{T}} \neq A$) form a subspace.

Questions 19–27 are about column spaces $C(A)$ and the equation $A\boldsymbol{x} = \boldsymbol{b}$.

19 Describe the column spaces (lines or planes) of these particular matrices:

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 0 \end{bmatrix}.$$

20 For which right sides (find a condition on b_1, b_2, b_3) are these systems solvable?

$$\text{(a)} \quad \begin{bmatrix} 1 & 4 & 2 \\ 2 & 8 & 4 \\ -1 & -4 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \qquad \text{(b)} \quad \begin{bmatrix} 1 & 4 \\ 2 & 9 \\ -1 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

21 Adding row 1 of A to row 2 produces B. Adding column 1 to column 2 produces C. A combination of the columns of (B or C ?) is also a combination of the columns of A. Which two matrices have the same column _____ ?

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}.$$

22 For which vectors (b_1, b_2, b_3) do these systems have a solution?

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$$\text{and} \quad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

23 (Recommended) If we add an extra column $\boldsymbol{b}$ to a matrix A, then the column space gets larger unless _____. Give an example where the column space gets larger and an example where it doesn't. Why is $A\boldsymbol{x} = \boldsymbol{b}$ solvable exactly when the column space *doesn't* get larger—it is the same for A and $\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix}$?

24 The columns of AB are combinations of the columns of A. This means: *The column space of AB is contained in* (possibly equal to) *the column space of A*. Give an example where the column spaces of A and AB are not equal.

25 Suppose $A\boldsymbol{x} = \boldsymbol{b}$ and $A\boldsymbol{y} = \boldsymbol{b}^*$ are both solvable. Then $A\boldsymbol{z} = \boldsymbol{b} + \boldsymbol{b}^*$ is solvable. What is $\boldsymbol{z}$? This translates into: If $\boldsymbol{b}$ and $\boldsymbol{b}^*$ are in the column space $\boldsymbol{C}(A)$, *then* $\boldsymbol{b} + \boldsymbol{b}^*$ is in $\boldsymbol{C}(A)$.

26 If A is any 5 by 5 invertible matrix, then its column space is _____. Why?

27 True or false (with a counterexample if false):

(a) The vectors $\boldsymbol{b}$ that are not in the column space $\boldsymbol{C}(A)$ form a subspace.

(b) If $\boldsymbol{C}(A)$ contains only the zero vector, then A is the zero matrix.

(c) The column space of $2A$ equals the column space of A.

(d) The column space of $A - I$ equals the column space of A (test this).

28 Construct a 3 by 3 matrix whose column space contains $(1, 1, 0)$ and $(1, 0, 1)$ but not $(1, 1, 1)$. Construct a 3 by 3 matrix whose column space is only a line.

29 If the 9 by 12 system $A\boldsymbol{x} = \boldsymbol{b}$ is solvable for every $\boldsymbol{b}$, then $\boldsymbol{C}(A) =$ _____.

Challenge Problems

30 Suppose **S** and **T** are two subspaces of a vector space **V**.

(a) **Definition**: The **sum S + T** contains all sums $s + t$ of a vector s in **S** and a vector t in **T**. Show that **S + T** satisfies the requirements (addition and scalar multiplication) for a vector space.

(b) If **S** and **T** are lines in $\mathbf{R}^m$, what is the difference between **S + T** and **S ∪ T**? That union contains all vectors from **S** or **T** or both. Explain this statement: *The span of* **S ∪ T** *is* **S + T**. (Section 3.5 returns to this word "span".)

31 If **S** is the column space of A and **T** is $\boldsymbol{C}(B)$, then **S + T** is the column space of what matrix M? The columns of A and B and M are all in $\mathbf{R}^m$. (I don't think $A + B$ is always a correct M.)

32 Show that the matrices A and $\begin{bmatrix} A & AB \end{bmatrix}$ (with extra columns) have the same column space. But find a square matrix with $\boldsymbol{C}(A^2)$ smaller than $\boldsymbol{C}(A)$. Important point:

An n by n matrix has $\boldsymbol{C}(A) = \mathbf{R}^n$ exactly when A is an _____ matrix.

3.2 The Nullspace of A: Solving $Ax = 0$

This section is about the subspace containing all solutions to $A\boldsymbol{x} = \boldsymbol{0}$. The m by n matrix A can be square or rectangular. *One immediate solution is $\boldsymbol{x} = \boldsymbol{0}$.* For invertible matrices this is the only solution. For other matrices, not invertible, there are nonzero solutions to $A\boldsymbol{x} = \boldsymbol{0}$. *Each solution $\boldsymbol{x}$ belongs to the nullspace of A.*

Elimination will find all solutions and identify this very important subspace.

The nullspace of A consists of all solutions to $A\boldsymbol{x} = \boldsymbol{0}$. **These vectors $\boldsymbol{x}$ are in $\mathbf{R}^n$.** The nullspace containing all solutions of $A\boldsymbol{x} = \boldsymbol{0}$ is denoted by $\boldsymbol{N}(A)$.

Check that the solution vectors form a subspace. Suppose $\boldsymbol{x}$ and $\boldsymbol{y}$ are in the nullspace (this means $A\boldsymbol{x} = \boldsymbol{0}$ and $A\boldsymbol{y} = \boldsymbol{0}$). The rules of matrix multiplication give $A(\boldsymbol{x} + \boldsymbol{y}) = \boldsymbol{0} + \boldsymbol{0}$. The rules also give $A(c\boldsymbol{x}) = c\boldsymbol{0}$. The right sides are still zero. Therefore $\boldsymbol{x} + \boldsymbol{y}$ and $c\boldsymbol{x}$ are also in the nullspace $\boldsymbol{N}(A)$. Since we can add and multiply without leaving the nullspace, it is a subspace.

To repeat: The solution vectors $\boldsymbol{x}$ have n components. They are vectors in $\mathbf{R}^n$, so *the nullspace is a subspace of* $\mathbf{R}^n$. The column space $\boldsymbol{C}(A)$ is a subspace of $\mathbf{R}^m$.

If the right side $\boldsymbol{b}$ is not zero, the solutions of $A\boldsymbol{x} = \boldsymbol{b}$ do *not* form a subspace. The vector $\boldsymbol{x} = \boldsymbol{0}$ is only a solution if $\boldsymbol{b} = \boldsymbol{0}$. When the set of solutions does not include $\boldsymbol{x} = \boldsymbol{0}$, it cannot be a subspace. Section 3.4 will show how the solutions to $A\boldsymbol{x} = \boldsymbol{b}$ (if there are any solutions) are shifted away from the origin by one particular solution.

Example 1 $x + 2y + 3z = 0$ comes from the 1 by 3 matrix $A = [1 \quad 2 \quad 3]$. This equation $A\boldsymbol{x} = \boldsymbol{0}$ produces a plane through the origin $(0, 0, 0)$. The plane is a subspace of $\mathbf{R}^3$. *It is the nullspace of A.*

The solutions to $x + 2y + 3z = 6$ also form a plane, but not a subspace.

Example 2 Describe the nullspace of $A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$. This matrix is singular!

Solution Apply elimination to the linear equations $A\boldsymbol{x} = \boldsymbol{0}$:

$$\begin{matrix} x_1 + 2x_2 = 0 \\ 3x_1 + 6x_2 = 0 \end{matrix} \quad \to \quad \begin{matrix} x_1 + 2x_2 = 0 \\ \boldsymbol{0 = 0} \end{matrix}$$

There is really only one equation. The second equation is the first equation multiplied by 3. In the row picture, the line $x_1 + 2x_2 = 0$ is the same as the line $3x_1 + 6x_2 = 0$. That line is the nullspace $\boldsymbol{N}(A)$. It contains all solutions (x_1, x_2).

To describe this line of solutions, here is an efficient way. Choose one point on the line (one "***special solution***"). Then all points on the line are multiples of this one. We choose the second component to be $x_2 = 1$ (a special choice). From the equation $x_1 + 2x_2 = 0$, the first component must be $x_1 = -2$. The special solution $\boldsymbol{s}$ is $(-2, 1)$:

Special solution The nullspace of $A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$ contains all multiples of $\boldsymbol{s} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$.

This is the best way to describe the nullspace, by computing special solutions to $Ax = \mathbf{0}$. This example has one special solution and the nullspace is a line.

The nullspace consists of all combinations of the special solutions.

The plane $x + 2y + 3z = 0$ in Example 1 had *two* special solutions:

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0 \text{ has the special solutions } s_1 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \text{ and } s_2 = \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}.$$

Those vectors s_1 and s_2 lie on the plane $x + 2y + 3z = 0$, which is the nullspace of $A = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$. All vectors on the plane are combinations of s_1 and s_2.

Notice what is special about s_1 and s_2. They have ones and zeros in the last two components. *Those components are "free" and we choose them specially.* Then the first components -2 and -3 are determined by the equation $Ax = \mathbf{0}$.

The first column of $A = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$ contains the *pivot*, so the first component of x is *not free*. The free components correspond to columns without pivots. This description of special solutions will be completed after one more example.

The special choice (one or zero) is only for the free variables.

Example 3 Describe the nullspaces of these three matrices A, B, C:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 8 \end{bmatrix} \quad B = \begin{bmatrix} A \\ 2A \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 8 \\ 2 & 4 \\ 6 & 16 \end{bmatrix} \quad C = \begin{bmatrix} A & 2A \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 3 & 8 & 6 & 16 \end{bmatrix}.$$

Solution The equation $Ax = \mathbf{0}$ has only the zero solution $x = \mathbf{0}$. *The nullspace is* $\mathbf{Z}$. It contains only the single point $x = \mathbf{0}$ in $\mathbf{R}^2$. This comes from elimination:

$$\begin{bmatrix} 1 & 2 \\ 3 & 8 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ yields } \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} x_1 = 0 \\ x_2 = 0 \end{bmatrix}.$$

A is invertible. There are no special solutions. All columns of this A have pivots.

The rectangular matrix B has the same nullspace $\mathbf{Z}$. The first two equations in $Bx = \mathbf{0}$ again require $x = \mathbf{0}$. The last two equations would also force $x = \mathbf{0}$. When we add extra equations, the nullspace certainly cannot become larger. The extra rows impose more conditions on the vectors x in the nullspace.

The rectangular matrix C is different. It has extra columns instead of extra rows. The solution vector x has *four* components. Elimination will produce pivots in the first two columns of C, but the last two columns are "free". They don't have pivots:

$$C = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 3 & 8 & 6 & 16 \end{bmatrix} \text{ becomes } U = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 2 & 0 & 4 \end{bmatrix}$$
$$\begin{array}{cccc} \uparrow & \uparrow & \uparrow & \uparrow \\ \end{array}$$
$$\textbf{pivot columns} \quad \textbf{free columns}$$

For the free variables x_3 and x_4, we make special choices of ones and zeros. First $x_3 = 1$, $x_4 = 0$ and second $x_3 = 0$, $x_4 = 1$. The pivot variables x_1 and x_2 are determined by the equation $Ux = \mathbf{0}$. We get two special solutions in the nullspace of C (which is also the nullspace of U). The special solutions are s_1 and s_2:

$$s_1 = \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{and } s_2 = \begin{bmatrix} 0 \\ -2 \\ 0 \\ 1 \end{bmatrix} \begin{array}{l} \leftarrow \text{ pivot} \\ \leftarrow \quad \text{variables} \\ \leftarrow \text{ free} \\ \leftarrow \quad \text{variables} \end{array}$$

One more comment to anticipate what is coming soon. Elimination will not stop at the upper triangular U! We can continue to make this matrix simpler, in two ways:

1. ***Produce zeros above the pivots,*** **by eliminating upward.**
2. ***Produce ones in the pivots,*** **by dividing the whole row by its pivot.**

Those steps don't change the zero vector on the right side of the equation. The nullspace stays the same. This nullspace becomes easiest to see when we reach the ***reduced row echelon form*** R. It has I in the pivot columns:

Reduced form R

$$U = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 2 & 0 & 4 \end{bmatrix} \quad \text{becomes} \quad R = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{bmatrix}.$$
$$\begin{array}{cc} \uparrow & \uparrow \end{array}$$

now the pivot columns contain I

I subtracted row 2 of U from row 1, and then multiplied row 2 by $\frac{1}{2}$. The original two equations have simplified to $x_1 + 2x_3 = 0$ and $x_2 + 2x_4 = 0$.

The first special solution is still $s_1 = (-2, 0, 1, 0)$, and s_2 is also unchanged. Special solutions are much easier to find from the reduced system $Rx = \mathbf{0}$.

Before moving to m by n matrices A and their nullspaces $N(A)$ and special solutions, allow me to repeat one comment. For many matrices, the only solution to $Ax = \mathbf{0}$ is $x = \mathbf{0}$. Their nullspaces $N(A) = \mathbf{Z}$ contain only that zero vector. The only combination of the columns that produces $b = \mathbf{0}$ is then the "zero combination" or "trivial combination". The solution is trivial (just $x = \mathbf{0}$) but the idea is not trivial.

This case of a zero nullspace $\mathbf{Z}$ is of the greatest importance. It says that the columns of A are **independent**. No combination of columns gives the zero vector (except the zero combination). All columns have pivots, and no columns are free. You will see this idea of independence again ...

Solving $Ax = \mathbf{0}$ by Elimination

This is important. ***A is rectangular and we still use elimination.*** We solve m equations in n unknowns when $b = \mathbf{0}$. After A is simplified by row operations, we read off the solution (or solutions). Remember the two stages (forward and back) in solving $Ax = \mathbf{0}$:

1. **Forward elimination** takes A to a triangular U (or its reduced form R).

2. **Back substitution** in $Ux = \mathbf{0}$ or $Rx = \mathbf{0}$ produces x.

You will notice a difference in back substitution, when A and U have fewer than n pivots. *We are allowing all matrices in this chapter*, not just the nice ones (which are square matrices with inverses).

Pivots are still nonzero. The columns below the pivots are still zero. But it might happen that a column has no pivot. That free column doesn't stop the calculation. ***Go on to the next column.*** The first example is a 3 by 4 matrix with two pivots:

$$A = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 2 & 2 & 8 & 10 \\ 3 & 3 & 10 & 13 \end{bmatrix}.$$

Certainly $a_{11} = 1$ is the first pivot. Clear out the 2 and 3 below that pivot:

$$A \to \begin{bmatrix} 1 & 1 & 2 & 3 \\ 0 & 0 & 4 & 4 \\ 0 & 0 & 4 & 4 \end{bmatrix} \quad \begin{matrix} \\ \text{(subtract } 2 \times \text{ row 1)} \\ \text{(subtract } 3 \times \text{ row 1)} \end{matrix}$$

The second column has a zero in the pivot position. We look below the zero for a nonzero entry, ready to do a row exchange. *The entry below that position is also zero.* Elimination can do nothing with the second column. This signals trouble, which we expect anyway for a rectangular matrix. There is no reason to quit, and we go on to the third column.

The second pivot is 4 (but it is in the third column). Subtracting row 2 from row 3 clears out that column below the pivot. **The pivot columns are 1 and 3:**

$$\textbf{Triangular } U: \quad U = \begin{bmatrix} \mathbf{1} & 1 & 2 & 3 \\ 0 & 0 & \mathbf{4} & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{matrix} \textit{Only two pivots} \\ \textit{The last equation} \\ \textit{became } 0 = 0 \end{matrix}$$

The fourth column also has a zero in the pivot position—but nothing can be done. There is no row below it to exchange, and forward elimination is complete. The matrix has three rows, four columns, and *only two pivots*. The original $Ax = \mathbf{0}$ seemed to involve three different equations, but the third equation is the sum of the first two. It is automatically satisfied $(0 = 0)$ when the first two equations are satisfied. Elimination reveals the inner truth about a system of equations. Soon we push on from U to R.

Now comes back substitution, to find all solutions to $Ux = \mathbf{0}$. With four unknowns and only two pivots, there are many solutions. The question is how to write them all down. A good method is to separate the ***pivot variables*** from the ***free variables***.

P	The ***pivot*** variables are x_1 and x_3.	**Columns 1 and 3 contain pivots.**
F	The ***free*** variables are x_2 and x_4.	**Columns 2 and 4 have no pivots.**

The free variables x_2 and x_4 can be given any values whatsoever. Then back substitution finds the pivot variables x_1 and x_3. (In Chapter 2 no variables were free. When A is invertible, all variables are pivot variables.) The simplest choices for the free variables are ones and zeros. Those choices give the *special solutions*.

Special solutions to $x_1 + x_2 + 2x_3 + 3x_4 = 0$ and $4x_3 + 4x_4 = 0$

- Set $x_2 = 1$ and $x_4 = 0$. By back substitution $x_3 = 0$. Then $x_1 = -1$.
- Set $x_2 = 0$ and $x_4 = 1$. By back substitution $x_3 = -1$. Then $x_1 = -1$.

These special solutions solve $Ux = \mathbf{0}$ and therefore $Ax = \mathbf{0}$. They are in the nullspace. The good thing is that *every solution is a combination of the special solutions.*

Complete solution to $Ax = 0$

$$x = x_2 \underbrace{\begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}}_{\textbf{special}} + x_4 \underbrace{\begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix}}_{\textbf{special}} = \underbrace{\begin{bmatrix} -x_2 - x_4 \\ x_2 \\ -x_4 \\ x_4 \end{bmatrix}}_{\textbf{complete}}. \qquad (1)$$

Please look again at that answer. It is the main goal of this section. The vector $s_1 = (-1, 1, 0, 0)$ is the special solution when $x_2 = 1$ and $x_4 = 0$. The second special solution has $x_2 = 0$ and $x_4 = 1$. ***All solutions are linear combinations of s_1 and s_2.*** The special solutions are in the nullspace $N(A)$, and their combinations fill out the whole nullspace.

The MATLAB code **nullbasis** computes these special solutions. They go into the columns of a ***nullspace matrix*** N. The complete solution to $Ax = \mathbf{0}$ is a combination of those columns. Once we have the special solutions, we have the whole nullspace.

There is a special solution for each free variable. If no variables are free—this means there are n pivots—then the only solution to $Ux = \mathbf{0}$ and $Ax = \mathbf{0}$ is the trivial solution $x = \mathbf{0}$. All variables are pivot variables. In that case the nullspaces of A and U contain only the zero vector. With no free variables, and pivots in every column, the output from **nullbasis** is an empty matrix. The nullspace with n pivots is $\mathbf{Z}$.

Example 4 Find the nullspace of $U = \begin{bmatrix} 1 & 5 & 7 \\ 0 & 0 & 9 \end{bmatrix}$.

The second column of U has no pivot. So x_2 is free. The special solution has $x_2 = 1$. Back substitution into $9x_3 = 0$ gives $x_3 = 0$. Then $x_1 + 5x_2 = 0$ or $x_1 = -5$. The solutions to $Ux = \mathbf{0}$ are multiples of one special solution:

$$x = x_2 \begin{bmatrix} -5 \\ 1 \\ 0 \end{bmatrix}$$

The nullspace of U is a line in $\mathbf{R}^3$.
It contains multiples of the special solution $s = (-5, 1, 0)$.
One variable is free, and $N = \textbf{nullbasis}\,(U)$ has one column s.

In a minute elimination will get zeros above the pivots and ones in the pivots. By continuing elimination on U, the 7 is removed and the pivot changes from 9 to 1.

The final result will be the **reduced row echelon form R**:

$$U = \begin{bmatrix} 1 & 5 & 7 \\ 0 & 0 & 9 \end{bmatrix} \text{ reduces to } R = \begin{bmatrix} 1 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \text{rref}(U).$$

This makes it even clearer that the special solution (column of N) is $s = (-5, 1, 0)$.

Echelon Matrices

Forward elimination goes from A to U. It acts by row operations, including row exchanges. It goes on to the next column when no pivot is available in the current column. The m by n "staircase" U is an ***echelon matrix***.

Here is a 4 by 7 echelon matrix with the three pivots $\boldsymbol{p}$ highlighted in boldface:

$$U = \begin{bmatrix} \boldsymbol{p} & x & x & x & x & x & x \\ 0 & \boldsymbol{p} & x & x & x & x & x \\ 0 & 0 & 0 & 0 & 0 & \boldsymbol{p} & x \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Three pivot variables $\mathbf{x_1, x_2, x_6}$
Four free variables $\mathbf{x_3, x_4, x_5, x_7}$
Four special solutions in $\boldsymbol{N(U)}$

Question What are the column space and the nullspace for this matrix?

Answer The columns have four components so they lie in $\mathbf{R}^4$. (Not in $\mathbf{R}^3$!) The fourth component of every column is zero. Every combination of the columns—every vector in the column space—has fourth component zero. *The column space $C(U)$ consists of all vectors of the form* $(b_1, b_2, b_3, 0)$. For those vectors we can solve $Ux = \boldsymbol{b}$ by back substitution. These vectors $\boldsymbol{b}$ are all possible combinations of the seven columns.

The nullspace $N(U)$ is a subspace of $\mathbf{R}^7$. The solutions to $Ux = \mathbf{0}$ are all the combinations of the four special solutions—*one for each free variable*:

1. Columns 3, 4, 5, 7 have no pivots. So the free variables are x_3, x_4, x_5, x_7.
2. Set one free variable to 1 and set the other free variables to zero.
3. Solve $Ux = \mathbf{0}$ for the pivot variables x_1, x_2, x_6.
4. This gives one of the four special solutions in the nullspace matrix N.

The nonzero rows of an echelon matrix go down in a staircase pattern. The pivots are the first nonzero entries in those rows. There is a column of zeros below every pivot.

Counting the pivots leads to an extremely important theorem. Suppose A has more columns than rows. ***With $n > m$ there is at least one free variable.*** The system $Ax = \mathbf{0}$ has at least one special solution. This solution is *not zero*!

Suppose $Ax = \mathbf{0}$ has more unknowns than equations (**n > m**, more columns than rows). Then there are **nonzero solutions**. There must be free columns, without pivots.

A short wide matrix ($n > m$) always has nonzero vectors in its nullspace. There must be at least $n - m$ free variables, since the number of pivots cannot exceed m. (The matrix only has m rows, and a row never has two pivots.) Of course a row might have *no* pivot—which means an extra free variable. But here is the point: When there is a free variable, it can be set to 1. Then the equation $Ax = \mathbf{0}$ has a nonzero solution.

To repeat: There are at most m pivots. With $n > m$, the system $Ax = \mathbf{0}$ has a nonzero solution. Actually there are infinitely many solutions, since any multiple cx is also a solution. The nullspace contains at least a line of solutions. With two free variables, there are two special solutions and the nullspace is even larger.

The nullspace is a subspace. Its "dimension" is the number of free variables. This central idea—the ***dimension*** of a subspace—is defined and explained in this chapter.

The Reduced Row Echelon Matrix R

From an echelon matrix U we go one more step. Continue with a 3 by 4 example:

$$U = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

We can divide the second row by 4. Then both pivots equal 1. *We can subtract* 2 *times this new row* $\begin{bmatrix} 0 & 0 & 1 & 1 \end{bmatrix}$ *from the row above.* ***The reduced row echelon matrix R has zeros above the pivots as well as below:***

Reduced row echelon matrix $$R = \text{rref}(A) = \begin{bmatrix} \mathbf{1} & 1 & \mathbf{0} & 1 \\ \mathbf{0} & 0 & \mathbf{1} & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$ **Pivot rows contain I**

R ***has 1's as pivots. Zeros above pivots come from upward elimination.***

Important *If A is invertible, its reduced row echelon form is the identity matrix $R = I$.* This is the ultimate in row reduction. Of course the nullspace is then **Z**.

The zeros in R make it easy to find the special solutions (the same as before):

1. Set $x_2 = 1$ and $x_4 = 0$. Solve $Rx = \mathbf{0}$. Then $x_1 = -1$ and $x_3 = 0$.

Those numbers -1 and 0 are sitting in column 2 of R (with plus signs).

2. Set $x_2 = 0$ and $x_4 = 1$. Solve $Rx = \mathbf{0}$. Then $x_1 = -1$ and $x_3 = -1$.

Those numbers -1 and -1 are sitting in column 4 (with plus signs).

By reversing signs we can read off the special solutions directly from R. The nullspace $N(A) = N(U) = N(R)$ contains all combinations of the special solutions:

$$x = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix} = \textit{(complete solution of } Ax = \mathbf{0}).$$

The next section of the book moves firmly from U to the row reduced form R. The MATLAB command $[\,R, pivcol\,] = \text{rref}(A)$ produces R and also a list of the pivot columns.

■ REVIEW OF THE KEY IDEAS ■

1. The nullspace $N(A)$ is a subspace of $\mathbf{R}^n$. It contains all solutions to $Ax = \mathbf{0}$.

2. Elimination produces an echelon matrix U, and then a row reduced R, with pivot columns and free columns.

3. Every free column of U or R leads to a special solution. The free variable equals 1 and the other free variables equal 0. Back substitution solves $Ax = \mathbf{0}$.

4. The complete solution to $Ax = \mathbf{0}$ is a combination of the special solutions.

5. If $n > m$ then A has at least one column without pivots, giving a special solution. So there are nonzero vectors x in the nullspace of this rectangular A.

■ WORKED EXAMPLES ■

3.2 A Create a 3 by 4 matrix whose special solutions to $Ax = \mathbf{0}$ are s_1 and s_2:

$$s_1 = \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad s_2 = \begin{bmatrix} -2 \\ 0 \\ -6 \\ 1 \end{bmatrix} \qquad \begin{array}{l} \text{pivot columns 1 and 3} \\ \text{free variables } x_2 \text{ and } x_4 \end{array}$$

You could create the matrix A in row reduced form R. Then describe all possible matrices A with the required nullspace $N(A) =$ all combinations of s_1 and s_2.

Solution The reduced matrix R has pivots $= 1$ in columns 1 and 3. There is no third pivot, so the third row of R is all zeros. The free columns 2 and 4 will be combinations of the pivot columns:

$$R = \begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{has} \quad Rs_1 = \mathbf{0} \quad \text{and} \quad Rs_2 = \mathbf{0}.$$

The entries $3, 2, 6$ in R are the negatives of $-3, -2, -6$ in the special solutions!

R is only one matrix (one possible A) with the required nullspace. We could do any elementary operations on R—exchange rows, multiply a row by any $c \neq 0$, subtract any multiple of one row from another. **R can be multiplied** (*on the left*) **by any invertible matrix, without changing its nullspace.**

Every 3 by 4 matrix has at least one special solution. *These matrices have two.*

3.2 B Find the special solutions and describe the *complete solution* to $Ax = \mathbf{0}$ for

$$A_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 3 & 6 \\ 1 & 2 \end{bmatrix} \qquad A_3 = \begin{bmatrix} A_2 & A_2 \end{bmatrix}$$

Which are the pivot columns? Which are the free variables? What is R in each case?

Solution $A_1x = \mathbf{0}$ has four special solutions. They are the columns s_1, s_2, s_3, s_4 of the 4 by 4 identity matrix. The nullspace is all of $\mathbf{R}^4$. The complete solution to $A_1x = \mathbf{0}$ is any $x = c_1s_1 + c_2s_2 + c_3s_3 + c_4s_4$ in $\mathbf{R}^4$. There are no pivot columns; all variables are free; the reduced R is the same zero matrix as A_1.

$A_2x = \mathbf{0}$ has only one special solution $s = (-2, 1)$. The multiples $x = cs$ give the complete solution. The first column of A_2 is its pivot column, and x_2 is the free variable. The row reduced matrices R_2 for A_2 and R_3 for $A_3 = [\,A_2 \;\; A_2\,]$ have 1's in the pivot:

$$A_2 = \begin{bmatrix} 3 & 6 \\ 1 & 2 \end{bmatrix} \to R_2 = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} A_2 & A_2 \end{bmatrix} \to R_3 = \begin{bmatrix} 1 & 2 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Notice that R_3 has only one pivot column (the first column). All the variables x_2, x_3, x_4 are free. There are three special solutions to $A_3x = \mathbf{0}$ (and also $R_3x = \mathbf{0}$):

$s_1 = (-2, 1, 0, 0)$ $s_2 = (-1, 0, 1, 0)$ $s_3 = (-2, 0, 0, 1)$ **Complete** $x = c_1s_1 + c_2s_2 + c_3s_3$.

With r pivots, A has $n - r$ free variables. $Ax = \mathbf{0}$ has $n - r$ special solutions.

Problem Set 3.2

Questions 1–4 and 5–8 are about the matrices in Problems 1 and 5.

1 Reduce these matrices to their ordinary echelon forms U:

$$\text{(a)}\; A = \begin{bmatrix} 1 & 2 & 2 & 4 & 6 \\ 1 & 2 & 3 & 6 & 9 \\ 0 & 0 & 1 & 2 & 3 \end{bmatrix} \qquad \text{(b)}\quad B = \begin{bmatrix} 2 & 4 & 2 \\ 0 & 4 & 4 \\ 0 & 8 & 8 \end{bmatrix}.$$

Which are the free variables and which are the pivot variables?

2 For the matrices in Problem 1, find a special solution for each free variable. (Set the free variable to 1. Set the other free variables to zero.)

3 By combining the special solutions in Problem 2, describe every solution to $Ax = \mathbf{0}$ and $Bx = \mathbf{0}$. The nullspace contains only $x = \mathbf{0}$ when there are no ______.

4 By further row operations on each U in Problem 1, find the reduced echelon form R. *True or false*: The nullspace of R equals the nullspace of U.

5 By row operations reduce each matrix to its echelon form U. Write down a 2 by 2 lower triangular L such that $B = LU$.

(a) $A = \begin{bmatrix} -1 & 3 & 5 \\ -2 & 6 & 10 \end{bmatrix}$ (b) $B = \begin{bmatrix} -1 & 3 & 5 \\ -2 & 6 & 7 \end{bmatrix}$.

6 For the same A and B, find the special solutions to $A\boldsymbol{x} = \mathbf{0}$ and $B\boldsymbol{x} = \mathbf{0}$. For an m by n matrix, the number of pivot variables plus the number of free variables is _____.

7 In Problem 5, describe the nullspaces of A and B in two ways. Give the equations for the plane or the line, and give all vectors $\boldsymbol{x}$ that satisfy those equations as combinations of the special solutions.

8 Reduce the echelon forms U in Problem 5 to R. For each R draw a box around the identity matrix that is in the pivot rows and pivot columns.

Questions 9–17 are about free variables and pivot variables.

9 True or false (with reason if true or example to show it is false):

(a) A square matrix has no free variables.

(b) An invertible matrix has no free variables.

(c) An m by n matrix has no more than n pivot variables.

(d) An m by n matrix has no more than m pivot variables.

10 Construct 3 by 3 matrices A to satisfy these requirements (if possible):

(a) A has no zero entries but $U = I$.

(b) A has no zero entries but $R = I$.

(c) A has no zero entries but $R = U$.

(d) $A = U = 2R$.

11 Put as many 1's as possible in a 4 by 7 echelon matrix U whose pivot columns are

(a) $2, 4, 5$

(b) $1, 3, 6, 7$

(c) 4 and 6.

12 Put as many 1's as possible in a 4 by 8 *reduced* echelon matrix R so that the free columns are

(a) $2, 4, 5, 6$

(b) $1, 3, 6, 7, 8$.

13 Suppose column 4 of a 3 by 5 matrix is all zero. Then x_4 is certainly a _____ variable. The special solution for this variable is the vector $\boldsymbol{x} =$ _____.

14 Suppose the first and last columns of a 3 by 5 matrix are the same (not zero). Then _____ is a free variable. Find the special solution for this variable.

15 Suppose an m by n matrix has r pivots. The number of special solutions is _____. The nullspace contains only $\boldsymbol{x} = \mathbf{0}$ when $r =$ _____. The column space is all of $\mathbf{R}^m$ when $r =$ _____.

16 The nullspace of a 5 by 5 matrix contains only $\boldsymbol{x} = \mathbf{0}$ when the matrix has _____ pivots. The column space is $\mathbf{R}^5$ when there are _____ pivots. Explain why.

17 The equation $x - 3y - z = 0$ determines a plane in $\mathbf{R}^3$. What is the matrix A in this equation? Which are the free variables? The special solutions are $(3, 1, 0)$ and _____.

18 (Recommended) The plane $x-3y-z = 12$ is parallel to the plane $x-3y-z = 0$ in Problem 17. One particular point on this plane is $(12, 0, 0)$. All points on the plane have the form (fill in the first components)

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \\ 0 \\ 0 \end{bmatrix} + y \begin{bmatrix} \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} \\ 0 \\ 1 \end{bmatrix}.$$

19 Prove that U and $A = LU$ have the same nullspace when L is invertible:

$$\text{If } U\boldsymbol{x} = \mathbf{0} \text{ then } LU\boldsymbol{x} = \mathbf{0}. \quad \text{If } LU\boldsymbol{x} = \mathbf{0}, \text{ how do you know } U\boldsymbol{x} = \mathbf{0}?$$

20 Suppose column 1 + column 3 + column 5 = $\mathbf{0}$ in a 4 by 5 matrix with four pivots. Which column is sure to have no pivot (and which variable is free)? What is the special solution? What is the nullspace?

Questions 21–28 ask for matrices (if possible) with specific properties.

21 Construct a matrix whose nullspace consists of all combinations of $(2, 2, 1, 0)$ and $(3, 1, 0, 1)$.

22 Construct a matrix whose nullspace consists of all multiples of $(4, 3, 2, 1)$.

23 Construct a matrix whose column space contains $(1, 1, 5)$ and $(0, 3, 1)$ and whose nullspace contains $(1, 1, 2)$.

24 Construct a matrix whose column space contains $(1, 1, 0)$ and $(0, 1, 1)$ and whose nullspace contains $(1, 0, 1)$ and $(0, 0, 1)$.

25 Construct a matrix whose column space contains $(1, 1, 1)$ and whose nullspace is the line of multiples of $(1, 1, 1, 1)$.

26 Construct a 2 by 2 matrix whose nullspace equals its column space. This is possible.

27 Why does no 3 by 3 matrix have a nullspace that equals its column space?

28 If $AB = 0$ then the column space of B is contained in the _____ of A. Give an example of A and B.

29 The reduced form R of a 3 by 3 matrix with randomly chosen entries is almost sure to be ____. What R is virtually certain if the random A is 4 by 3?

30 Show by example that these three statements are generally *false*:

(a) A and A^{T} have the same nullspace.

(b) A and A^{T} have the same free variables.

(c) If R is the reduced form $\mathbf{rref}(A)$ then R^{T} is $\mathbf{rref}(A^{\mathrm{T}})$.

31 If the nullspace of A consists of all multiples of $x = (2, 1, 0, 1)$, how many pivots appear in U? What is R?

32 If the special solutions to $Rx = \mathbf{0}$ are in the columns of these N, go backward to find the nonzero rows of the reduced matrices R:

$$N = \begin{bmatrix} 2 & 3 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad N = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad N = \begin{bmatrix} \\ \\ \\ \end{bmatrix} \text{ (empty 3 by 1).}$$

33 (a) What are the five 2 by 2 reduced echelon matrices R whose entries are all 0's and 1's?

(b) What are the eight 1 by 3 matrices containing only 0's and 1's? Are all eight of them reduced echelon matrices R?

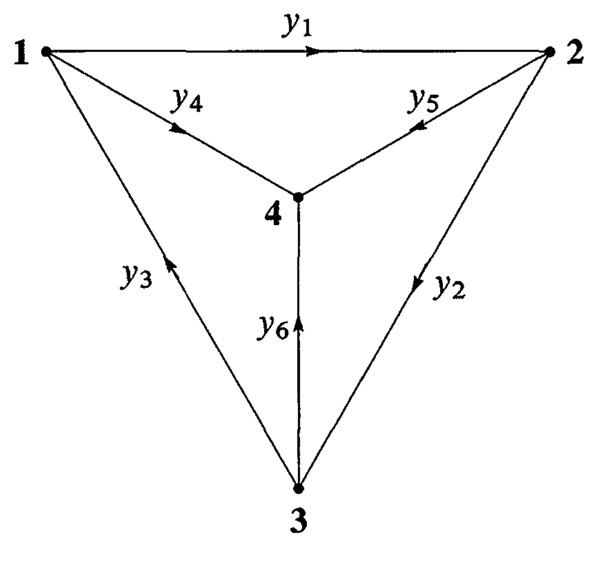
34 Explain why A and $-A$ always have the same reduced echelon form R.

Challenge Problems

35 If A is 4 by 4 and invertible, describe all vectors in the nullspace of the 4 by 8 matrix $B = [A \;\; A]$.

36 How is the nullspace $\mathbf{N}(C)$ related to the spaces $\mathbf{N}(A)$ and $\mathbf{N}(B)$, if $C = \begin{bmatrix} A \\ B \end{bmatrix}$?

37 Kirchhoff's Law says that *current in* = *current out* at every node. This network has six currents $y_1, \ldots, y_6$ (the arrows show the positive direction, each y_i could be positive or negative). Find the four equations $Ay = \mathbf{0}$ for Kirchhoff's Law at the four nodes. Find three special solutions in the nullspace of A.



3.3 The Rank and the Row Reduced Form

The numbers m and n give the size of a matrix—but not necessarily the *true size* of a linear system. An equation like $0 = 0$ should not count. If there are two identical rows in A, the second one disappears in elimination. Also if row 3 is a combination of rows 1 and 2, then row 3 will become all zeros in the triangular U and the reduced echelon form R. We don't want to count rows of zeros. ***The true size of A is given by its rank***:

DEFINITION *The rank of A is the number of pivots. This number is r.*

That definition is computational, and I would like to say more about the rank r. The matrix will eventually be reduced to r nonzero rows. Start with a 3 by 4 example.

Four columns
How many pivots?

$$A = \begin{bmatrix} 1 & 1 & 2 & 4 \\ 1 & 2 & 2 & 5 \\ 1 & 3 & 2 & 6 \end{bmatrix}. \tag{1}$$

The first two columns are $(1, 1, 1)$ and $(1, 2, 3)$, going in different directions. Those will be pivot columns. The third column $(2, 2, 2)$ is a multiple of the first. We won't see a pivot in that third column. The fourth column $(4, 5, 6)$ is a combination of the first three (their sum). That column will also be without a pivot.

The fourth column is actually a combination $3(1, 1, 1) + (1, 2, 3)$ of the two pivot columns. ***Every "free column" is a combination of earlier pivot columns.*** It is the *special solutions* s that tell us those combinations of pivot columns:

$$\begin{array}{lll} \text{Column 3} = \mathbf{2}\ (\text{column 1}) & s_1 = (-\mathbf{2}, 0, 1, 0) & As_1 = \mathbf{0} \\ \text{Column 4} = \mathbf{3}\ (\text{column 1}) + \mathbf{1}\ (\text{column 2}) & s_2 = (-\mathbf{3}, -\mathbf{1}, 0, 1) & As_2 = \mathbf{0} \end{array}$$

With nice numbers we can see the right combinations. The systematic way to find s is by elimination! This will change the columns but it won't change the combinations, because $Ax = \mathbf{0}$ is equivalent to $Ux = \mathbf{0}$ and also $Rx = \mathbf{0}$. I will go from A to U and then to R:

$$\begin{bmatrix} 1 & 1 & 2 & 4 \\ 1 & 2 & 2 & 5 \\ 1 & 3 & 2 & 6 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 2 & 4 \\ 0 & 1 & 0 & 1 \\ 0 & 2 & 0 & 2 \end{bmatrix} \to \begin{bmatrix} \mathbf{1} & 1 & 2 & 4 \\ 0 & \mathbf{1} & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = U$$

U already shows the two pivots in the pivot columns. **The rank of A (and U) is 2.** Continuing to R we see the combinations of pivot columns that produce the free columns:

$$U = \begin{bmatrix} 1 & 1 & 2 & 4 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \underset{\text{row 1} - \text{row 2}}{\overset{\text{Subtract}}{\longrightarrow}} \quad R = \begin{bmatrix} \mathbf{1} & 0 & 2 & 3 \\ 0 & \mathbf{1} & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{2}$$

Clearly the $(3, 1, 0)$ column equals 3 (column 1) + column 2. Moving all columns to the "left side" will reverse signs to -3 and -1, which go in the special solution s:

$$-3\ (\text{column 1}) - (\text{column 2}) + (\text{column 4}) = \mathbf{0} \qquad s = (-3, -1, 0, 1).$$

Rank One

Matrices of ***rank one*** have only *one pivot.* When elimination produces zero in the first column, it produces zero in all the columns. *Every row is a multiple of the pivot row.* At the same time, every column is a multiple of the pivot column!

$$\textbf{Rank one matrix} \qquad A = \begin{bmatrix} \mathbf{1} & 3 & 10 \\ \mathbf{2} & 6 & 20 \\ \mathbf{3} & 9 & 30 \end{bmatrix} \longrightarrow R = \begin{bmatrix} \mathbf{1} & 3 & 10 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The column space of a rank one matrix is "one-dimensional". Here all columns are on the line through $\boldsymbol{u} = (1,2,3)$. The columns of A are $\boldsymbol{u}$ and $3\boldsymbol{u}$ and $10\boldsymbol{u}$. Put those numbers into the row $\boldsymbol{v}^{\mathrm{T}} = [\ 1 \quad 3 \quad 10\ ]$ and you have the special rank one form $A = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$:

$$A = \textbf{column times row} = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} \qquad \begin{bmatrix} 1 & 3 & 10 \\ 2 & 6 & 20 \\ 3 & 9 & 30 \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \end{bmatrix} \begin{bmatrix} 1 & 3 & 10 \end{bmatrix}. \qquad (3)$$

With rank one, the solutions to $A\boldsymbol{x} = \mathbf{0}$ are easy to understand. That equation $\boldsymbol{u}(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}) = \mathbf{0}$ leads us to $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x} = 0$. All vectors $\boldsymbol{x}$ in the nullspace must be orthogonal to $\boldsymbol{v}$ in the row space. This is the geometry: *row space = line, nullspace = perpendicular plane.* Now describe the special solutions with numbers:

$$\begin{array}{l} \textbf{Pivot row } [1\ \ 3\ \ 10] \\ \textbf{Pivot variable } x_1 \\ \textbf{Free variables } x_2 \textbf{ and } x_3 \end{array} \qquad s_1 = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix} \qquad s_2 = \begin{bmatrix} -10 \\ 0 \\ 1 \end{bmatrix}$$

The nullspace contains all combinations of s_1 and s_2. This produces the plane $x + 3y + 10z = 0$, perpendicular to the row $(1,3,10)$. **Nullspace** (plane) **perpendicular to row space** (line).

Example 1 When all rows are multiples of one pivot row, the rank is $r = 1$:

$$\begin{bmatrix} 1 & 3 & 4 \\ 2 & 6 & 8 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 3 \\ 0 & 5 \end{bmatrix} \text{ and } \begin{bmatrix} 5 \\ 2 \end{bmatrix} \text{ and } \begin{bmatrix} 6 \end{bmatrix} \text{ all have rank 1.}$$

For those matrices, the reduced row echelon $R = \textbf{rref}\ (A)$ can be checked by eye:

$$R = \begin{bmatrix} 1 & 3 & 4 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 1 \end{bmatrix} \text{ have only one pivot.}$$

Our second definition of rank will be at a higher level. It deals with entire rows and entire columns—vectors and not just numbers. The matrices A and U and R have r *independent* rows (the pivot rows). They also have r *independent* columns (the pivot columns). Section 3.5 says what it means for rows or columns to be independent.

A third definition of rank, at the top level of linear algebra, will deal with *spaces* of vectors. ***The rank r is the "dimension" of the column space.*** It is also the dimension of the row space. The great thing is that r also reveals the dimension of the nullspace.

The Pivot Columns

The pivot columns of R have 1's in the pivots and 0's everywhere else. The r pivot columns taken together contain an r by r identity matrix I. It sits above $m - r$ rows of zeros. The numbers of the pivot columns are in the list *pivcol.*

The pivot columns of A are probably not obvious from A itself. But their column numbers are given by the *same list pivcol.* The r columns of A that eventually have pivots (in U and R) are the pivot columns of A. This example has *pivcol* $= (1, 3)$:

Pivot Columns

$$A = \begin{bmatrix} \mathbf{1} & 3 & \mathbf{0} & 2 & -1 \\ \mathbf{0} & 0 & \mathbf{1} & 4 & -3 \\ \mathbf{1} & 3 & \mathbf{1} & 6 & -4 \end{bmatrix} \text{ yields } R = \begin{bmatrix} \mathbf{1} & 3 & \mathbf{0} & 2 & -1 \\ \mathbf{0} & 0 & \mathbf{1} & 4 & -3 \\ \mathbf{0} & 0 & \mathbf{0} & 0 & 0 \end{bmatrix}.$$

The column spaces of A and R are different! All columns of this R end with zeros. Elimination subtracts rows 1 and 2 of A from row 3, to produce that zero row in R:

$\boldsymbol{EA = R}$
$\boldsymbol{A = E^{-1}R}$

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix} \quad \text{and} \quad E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

The r pivot columns of A are also the first r columns of E^{-1}. The r by r identity matrix inside R just picks out the first r columns of E^{-1} as columns of $A = E^{-1}R$.

One more fact about pivot columns. Their definition has been purely computational, based on R. Here is a direct mathematical description of the pivot columns of A:

The pivot columns are not combinations of earlier columns. The free columns ***are*** combinations of earlier columns. These combinations are the special solutions!

A pivot column of R (with 1 in the pivot row) cannot be a combination of earlier columns (with 0's in that row). The same column of A can't be a combination of earlier columns, because $A\boldsymbol{x} = \mathbf{0}$ *exactly when* $R\boldsymbol{x} = \mathbf{0}$.

Now we look at the special solution $\boldsymbol{x}$ from each free column.

The Special Solutions

Each special solution to $A\boldsymbol{x} = \mathbf{0}$ and $R\boldsymbol{x} = \mathbf{0}$ has one free variable equal to 1. The other free variables in $\boldsymbol{x}$ are all zero. The solutions come directly from the echelon form R:

Free columns
Free variables
in boldface

$$R\boldsymbol{x} = \begin{bmatrix} 1 & \mathbf{3} & 0 & \mathbf{2} & \mathbf{-1} \\ 0 & \mathbf{0} & 1 & \mathbf{4} & \mathbf{-3} \\ 0 & \mathbf{0} & 0 & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} x_1 \\ \boldsymbol{x_2} \\ x_3 \\ \boldsymbol{x_4} \\ \boldsymbol{x_5} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Set the first free variable to $x_2 = 1$ with $x_4 = x_5 = 0$. The equations give the pivot variables $x_1 = -3$ and $x_3 = 0$. The special solution is $s_1 = (-3, 1, 0, 0, 0)$.

The next special solution has $x_4 = 1$. The other free variables are $x_2 = x_5 = 0$. The solution is $s_2 = (-2, 0, -4, 1, 0)$. Notice -2 and -4 in R, with plus signs.

The third special solution has $x_5 = 1$. With $x_2 = 0$ and $x_4 = 0$ we find $s_3 = (1, 0, 3, 0, 1)$. The numbers $x_1 = 1$ and $x_3 = 3$ are in column 5 of R, again with opposite signs. This is a general rule as we soon verify. The nullspace matrix N contains the three special solutions in its columns, so $AN =$ *zero matrix*:

Nullspace matrix
$n - r = 5 - 2$
3 special solutions

$$N = \begin{bmatrix} -3 & -2 & 1 \\ 1 & 0 & 0 \\ 0 & -4 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{array}{l} \text{not free} \\ \text{free} \\ \text{not free} \\ \text{free} \\ \text{free} \end{array}$$

The linear combinations of these three columns give all vectors in the nullspace. This is the complete solution to $A\boldsymbol{x} = \mathbf{0}$ (and $R\boldsymbol{x} = \mathbf{0}$). Where R had the identity matrix (2 by 2) in its pivot columns, N has the identity matrix (3 by 3) in its free rows.

There is a special solution for every free variable. Since r columns have pivots, that leaves $n - r$ free variables. This is the key to $A\boldsymbol{x} = \mathbf{0}$ and the nullspace:

$A\boldsymbol{x} = \mathbf{0}$ ***has r pivots and $n - r$ free variables***: n columns minus r pivot columns. The ***nullspace matrix*** N contains the $n - r$ special solutions. Then $AN = 0$.

When we introduce the idea of "independent" vectors, we will show that the special solutions are independent. You can see in N that no column is a combination of the other columns. The beautiful thing is that the count is exactly right:

$A\boldsymbol{x} = \mathbf{0}$ ***has r independent equations so it has*** $\mathbf{n - r}$ ***independent solutions.***

The special solutions are easy for $R\boldsymbol{x} = \mathbf{0}$. Suppose that the first r columns are the pivot columns. Then the reduced row echelon form looks like

$$R = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{array}{l} r \text{ pivot rows} \\ m - r \text{ zero rows} \end{array} \qquad (4)$$

r pivot columns $\quad n - r$ free columns

The pivot variables in the $n - r$ special solutions come by changing F to $-F$:

$$\textbf{Nullspace matrix} \quad N = \begin{bmatrix} -\boldsymbol{F} \\ \boldsymbol{I} \end{bmatrix} \begin{array}{l} r \text{ pivot variables} \\ n - r \text{ free variables} \end{array} \qquad (5)$$

Check $RN = 0$. The first block row of RN is (I times $-F$) + (F times I) = zero. The columns of N solve $R\boldsymbol{x} = \mathbf{0}$. When the free part of $R\boldsymbol{x} = \mathbf{0}$ moves to the right side, the left side just holds the identity matrix:

$$Rx = \mathbf{0} \quad \text{means} \quad I \begin{bmatrix} \text{pivot} \\ \text{variables} \end{bmatrix} = -F \begin{bmatrix} \text{free} \\ \text{variables} \end{bmatrix}. \tag{6}$$

In each special solution, the free variables are a column of I. Then the pivot variables are a column of $-F$. Those special solutions give the nullspace matrix N.

The idea is still true if the pivot columns are mixed in with the free columns. Then I and F are mixed together. You can still see $-F$ in the solutions. Here is an example where $I = [1]$ comes first and $F = \begin{bmatrix} 2 & 3 \end{bmatrix}$ comes last.

Example 2 The special solutions of $Rx = x_1 + 2x_2 + 3x_3 = 0$ are the columns of N:

$$R = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \qquad N = \begin{bmatrix} -F \\ I \end{bmatrix} = \begin{bmatrix} -2 & -3 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The rank is one. There are $n - r = 3 - 1$ special solutions $(-2, 1, 0)$ and $(-3, 0, 1)$.

Final Note How can I write confidently about R not knowing which steps MATLAB will take? A could be reduced to R in different ways. Very likely you and Mathematica and Maple would do the elimination differently. The key is that ***the final R is always the same.*** *The original A completely determines the I and F and zero rows in R.*

For proof I will determine the pivot columns (which locate I) and free columns (which contain F) in an "algebra way"—two rules that have nothing to do with any particular elimination steps. Here are those rules:

1. The pivot columns *are not* combinations of earlier columns of A.

2. The free columns *are* combinations of earlier columns (F tells the combinations).

A small example with rank one will show two E's that produce the correct $EA = R$:

$$A = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \quad \text{reduces to} \quad R = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = \text{rref}(A) \quad \text{and no other } R.$$

You could multiply row 1 of A by $\frac{1}{2}$, and subtract row 1 from row 2:

$$\textbf{Two steps give } E \qquad \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 \\ -1/2 & 1 \end{bmatrix} = E.$$

Or you could exchange rows in A, and then subtract 2 times row 1 from row 2:

$$\textbf{Two different steps give } E_{\text{new}} \qquad \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & -2 \end{bmatrix} = E_{\text{new}}.$$

Multiplication gives $EA = R$ and also $E_{\text{new}}A = R$. ***Different E's but the same R.***

Codes for Row Reduction

There is no way that rref will ever come close in importance to lu. The Teaching Code elim for this book uses rref. Of course rref(R) would give R again!

MATLAB: $[R, pivcol] =$ rref(A) Teaching Code: $[E, R] =$ elim(A)

The extra output *pivcol* gives the numbers of the pivot columns. They are the same in A and R. The extra output E in the Teaching Code is an m by m ***elimination matrix*** that puts the original A (whatever it was) into its row reduced form R:

$$EA = R.$$

The square matrix E is the product of elementary matrices E_{ij} and also P_{ij} and D^{-1}. P_{ij} exchanges rows. The diagonal D^{-1} divides rows by their pivots to produce 1's.

If we want E, we can apply row reduction to the matrix $\begin{bmatrix} A & I \end{bmatrix}$ with $n + m$ columns. All the elementary matrices that multiply A (to produce R) will also multiply I (to produce E). The whole augmented matrix is being multiplied by E:

$$E\,[A \;\; I] \quad = \quad [R \;\; E] \tag{7}$$

This is exactly what "Gauss-Jordan" did in Chapter 2 to compute A^{-1}. ***When A is square and invertible, its reduced row echelon form is I.*** Then $EA = R$ becomes $EA = I$. In this invertible case, E is A^{-1}. This chapter is going further, to every A.

■ REVIEW OF THE KEY IDEAS ■

1. The rank r of A is the number of pivots (which are 1's in $R =$ rref(A)).
2. The r pivot columns of A and R are in the same list *pivcol*.
3. Those r pivot columns are not combinations of earlier columns.
4. The $n - r$ free columns *are* combinations of earlier columns (pivot columns).
5. Those combinations (using $-F$ taken from R) give the $n - r$ special solutions to $A\boldsymbol{x} = \mathbf{0}$ and $R\boldsymbol{x} = \mathbf{0}$. They are the $n - r$ columns of the nullspace matrix N.

■ WORKED EXAMPLES ■

3.3 A Find the reduced echelon form of A. What is the rank? What is the special solution to $A\boldsymbol{x} = \mathbf{0}$?

Second differences $-1, 2, -1$
Notice $A_{11} = A_{44} = 1$

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

Solution Add row 1 to row 2. Then add row 2 to row 3. Then add row 3 to row 4:

First differences 1, −1 $$U = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Now add row 3 to row 2. Then add row 2 to row 1:

Reduced form $$R = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}.$$

The rank is $r = 3$. There is one free variable ($n - r = 1$). *The special solution is* $s = (\mathbf{1}, \mathbf{1}, \mathbf{1}, \mathbf{1})$. *Every row adds to* 0. Notice $-F = (1, 1, 1)$ in the pivot variables of s.

3.3 B Factor these rank one matrices into $A = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ = column times row:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix} \qquad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{(find } d \text{ from } a, b, c \text{ if } a \neq 0)$$

Split this rank two matrix into $\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$ = (3 by 2) times (2 by 4) using R:

$$A = \begin{bmatrix} 1 & 1 & 0 & 2 \\ 1 & 2 & 0 & 3 \\ 2 & 3 & 0 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = E^{-1}R.$$

Solution For the 3 by 3 matrix A, all rows are multiples of $\boldsymbol{v}^{\mathrm{T}} = [1 \;\; 2 \;\; 3]$. All columns are multiples of the column $\boldsymbol{u} = (1, 2, 3)$. This symmetric matrix has $\boldsymbol{u} = \boldsymbol{v}$ and A is $\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$. Every rank one symmetric matrix will have this form or else $-\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$.

If the 2 by 2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ has rank one, it must be singular. In Chapter 5, its determinant is $ad - bc = 0$. In this chapter, row 2 is c/a times row 1.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 \\ c/a \end{bmatrix} \begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} a & b \\ c & bc/a \end{bmatrix}. \quad \text{So} \quad d = \frac{bc}{a}.$$

The 3 by 4 matrix of rank two is a sum of *two matrices of rank one*. All columns of A are combinations of the pivot columns 1 and 2. All rows are combinations of the nonzero rows of R. The pivot columns are $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ and those rows are $\boldsymbol{v}_1^{\mathrm{T}}$ and $\boldsymbol{v}_2^{\mathrm{T}}$. Then A is $\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$, multiplying r columns of E^{-1} times r rows of R:

Columns times rows $$\begin{bmatrix} 1 & 1 & 0 & 2 \\ 1 & 2 & 0 & 3 \\ 2 & 3 & 0 & 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 1 \end{bmatrix}$$

3.3 C Find the row reduced form R and the rank r of A and B (*those depend on* c). Which are the pivot columns of A? What are the special solutions and the matrix N?

Find special solutions $\quad A = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 6 & 3 \\ 4 & 8 & c \end{bmatrix}$ and $B = \begin{bmatrix} c & c \\ c & c \end{bmatrix}$.

Solution The matrix A has rank $r = 2$ ***except if*** $c = 4$. The pivots are in columns 1 and 3. The second variable x_2 is free. Notice the form of R:

$$c \neq 4 \quad R = \begin{bmatrix} \mathbf{1} & 2 & 0 \\ 0 & 0 & \mathbf{1} \\ 0 & 0 & 0 \end{bmatrix} \qquad c = 4 \quad R = \begin{bmatrix} \mathbf{1} & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Two pivots leave one free variable x_2. But when $c = 4$, the only pivot is in column 1 (rank one). The second and third variables are free, producing two special solutions:

$$c \neq 4 \quad \text{Special solution with } x_2 = 1 \text{ goes into } \quad N = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}.$$

$$c = 4 \quad \text{Another special solution goes into } \quad N = \begin{bmatrix} -2 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The 2 by 2 matrix $\left[\begin{smallmatrix} c & c \\ c & c \end{smallmatrix}\right]$ has rank $r = 1$ ***except if*** $c = 0$, when the rank is zero!

$$c \neq 0 \quad R = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad N = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \text{Nullspace} = \text{line}$$

The matrix has *no pivot columns* if $c = 0$. Then both variables are free:

$$c = 0 \quad R = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad N = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{Nullspace} = \mathbf{R}^2.$$

Problem Set 3.3

1 Which of these rules gives a correct definition of the *rank* of A?

(a) The number of nonzero rows in R.

(b) The number of columns minus the total number of rows.

(c) The number of columns minus the number of free columns.

(d) The number of 1's in the matrix R.

2 Find the reduced row echelon forms R and the rank of these matrices:

(a) The 3 by 4 matrix with all entries equal to 4.
(b) The 3 by 4 matrix with $a_{ij} = i + j - 1$.
(c) The 3 by 4 matrix with $a_{ij} = (-1)^j$.

3 Find the reduced R for each of these (block) matrices:

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 3 \\ 2 & 4 & 6 \end{bmatrix} \quad B = \begin{bmatrix} A & A \end{bmatrix} \quad C = \begin{bmatrix} A & A \\ A & 0 \end{bmatrix}$$

4 Suppose all the pivot variables come *last* instead of first. Describe all four blocks in the reduced echelon form (the block B should be r by r):

$$R = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

What is the nullspace matrix N containing the special solutions?

5 (Silly problem) Describe all 2 by 3 matrices A_1 and A_2, with row echelon forms R_1 and R_2, such that $R_1 + R_2$ is the row echelon form of $A_1 + A_2$. Is is true that $R_1 = A_1$ and $R_2 = A_2$ in this case? Does $R_1 - R_2$ equal $\mathbf{rref}(A_1 - A_2)$?

6 If A has r pivot columns, how do you know that A^{T} has r pivot columns? Give a 3 by 3 example with different column numbers in *pivcol* for A and A^{T}.

7 What are the special solutions to $R\boldsymbol{x} = \mathbf{0}$ and $\boldsymbol{y}^{\mathrm{T}}R = \mathbf{0}$ for these R?

$$R = \begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & 4 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad R = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Problems 8–11 are about matrices of rank $r = 1$.

8 Fill out these matrices so that they have rank 1:

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 2 & & \\ 4 & & \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} & 9 & \\ 1 & & \\ 2 & 6 & -3 \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} a & b \\ c & \end{bmatrix}.$$

9 If A is an m by n matrix with $r = 1$, its columns are multiples of one column and its rows are multiples of one row. The column space is a _____ in $\mathbf{R}^m$. The nullspace is a _____ in $\mathbf{R}^n$. The nullspace matrix N has shape _____.

10 Choose vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ so that $A = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ = column times row:

$$A = \begin{bmatrix} 3 & 6 & 6 \\ 1 & 2 & 2 \\ 4 & 8 & 8 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 2 & 2 & 6 & 4 \\ -1 & -1 & -3 & -2 \end{bmatrix}.$$

$A = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ is the natural form for every matrix that has rank $r = 1$.

11 If A is a rank one matrix, the second row of U is _____. Do an example.

Problems 12–14 are about r by r invertible matrices inside A.

12 *If A has rank r, then it has an r by r submatrix S that is invertible.* Remove $m-r$ rows and $n-r$ columns to find an invertible submatrix S inside A, B, and C. You could keep the pivot rows and pivot columns:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix} \qquad C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

13 Suppose P contains only the r pivot columns of an m by n matrix. Explain why this m by r submatrix P has rank r.

14 Transpose P in problem 13. Then find the r pivot columns of P^{T}. Transposing back, **this produces an r by r invertible submatrix S inside P and A:**

$$\text{For } A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 2 & 4 & 7 \end{bmatrix} \text{ find } P \text{ (3 by 2) and then the invertible } S \text{ (2 by 2).}$$

Problems 15–20 show that rank(AB) is not greater than rank(A) or rank(B).

15 Find the ranks of AB and AC (rank one matrix times rank one matrix):

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & 1 & 4 \\ 3 & 1.5 & 6 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 1 & b \\ c & bc \end{bmatrix}.$$

16 The rank one matrix $\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ times the rank one matrix $\boldsymbol{w}\boldsymbol{z}^{\mathrm{T}}$ is $\boldsymbol{u}\boldsymbol{z}^{\mathrm{T}}$ times the number ______. This product $\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w}\boldsymbol{z}^{\mathrm{T}}$ also has rank one unless ______ $= 0$.

17 (a) Suppose column j of B is a combination of previous columns of B. Show that column j of AB is the same combination of previous columns of AB. Then AB cannot have new pivot columns, so **rank(AB) $\leq$ rank(B).**

(b) Find A_1 and A_2 so that rank$(A_1B) = 1$ and rank$(A_2B) = 0$ for $B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

18 Problem 17 proved that rank$(AB) \leq$ rank(B). Then the same reasoning gives **rank($B^{\mathrm{T}}A^{\mathrm{T}}$)** $\leq$ rank(A^{T}). How do you deduce that **rank(AB) $\leq$ rank A**?

19 *(Important)* Suppose A and B are n by n matrices, and $AB = I$. Prove from rank$(AB) \leq$ rank(A) that the rank of A is n. So A is invertible and B must be its two-sided inverse (Section 2.5). Therefore $BA = I$ *(which is not so obvious!)*.

20 If A is 2 by 3 and B is 3 by 2 and $AB = I$, show from its rank that $BA \neq I$. Give an example of A and B with $AB = I$. For $m < n$, a right inverse is not a left inverse.

21 Suppose A and B have the *same* reduced row echelon form R.

(a) Show that A and B have the same nullspace and the same row space.

(b) We know $E_1A = R$ and $E_2B = R$. So A equals an _____ matrix times B.

22 Express A and then B as the sum of two rank one matrices:

$$\textbf{rank} = \mathbf{2} \qquad A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 4 \\ 1 & 1 & 8 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}.$$

23 Answer the same questions as in Worked Example **3.3 C** for

$$A = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 2 & 2 & 4 & 4 \\ 1 & c & 2 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1-c & 2 \\ 0 & 2-c \end{bmatrix}.$$

24 What is the nullspace matrix N (containing the special solutions) for A, B, C?

$$A = \begin{bmatrix} I & I \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} I & I \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} I & I & I \end{bmatrix}.$$

25 *Neat fact* ***Every m by n matrix of rank r reduces to*** (m by r) ***times*** (r by n):

$$A = (\text{pivot columns of } A)\,(\text{first } r \text{ rows of } R) = (\mathbf{COL})(\mathbf{ROW}).$$

Write the 3 by 4 matrix A in equation (1) at the start of this section as the product of the 3 by 2 matrix from the pivot columns and the 2 by 4 matrix from R.

Challenge Problems

26 Suppose A is an m by n matrix of rank r. Its reduced echelon form is R. Describe exactly the matrix Z (its shape and all its entries) that comes from *transposing the reduced row echelon form of* R' (prime means transpose):

$$R = \text{rref}(A) \quad \text{and} \quad Z = (\text{rref}(R'))'.$$

27 Suppose R is m by n of rank r, with pivot columns first:

$$R = \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}.$$

(a) What are the shapes of those four blocks?

(b) Find a *right-inverse* B with $RB = I$ if $r = m$.

(c) Find a *left-inverse* C with $CR = I$ if $r = n$.

(d) What is the reduced row echelon form of R^{T} (with shapes)?

(e) What is the reduced row echelon form of $R^{\mathrm{T}}R$ (with shapes)?

Prove that $R^{\mathrm{T}}R$ has the same nullspace as R. Later we show that $A^{\mathrm{T}}A$ always has the same nullspace as A (a valuable fact).

28 Suppose you allow elementary *column* operations on A as well as elementary row operations (which get to R). What is the "row-and-column reduced form" for an m by n matrix of rank r?

3.4 The Complete Solution to $Ax = b$

The last sections totally solved $Ax = \mathbf{0}$. Elimination converted the problem to $Rx = \mathbf{0}$. The free variables were given special values (one and zero). Then the pivot variables were found by back substitution. We paid no attention to the right side $\boldsymbol{b}$ because it started and ended as zero. The solution $\boldsymbol{x}$ was in the nullspace of A.

Now $\boldsymbol{b}$ is not zero. Row operations on the left side must act also on the right side. $Ax = b$ is reduced to a simpler system $Rx = d$. One way to organize that is to ***add b as an extra column of the matrix.*** I will *"augment"* A with the right side $(b_1, b_2, b_3) = (1, 6, 7)$ and reduce the bigger matrix $\begin{bmatrix} A & b \end{bmatrix}$:

$$\begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 7 \end{bmatrix} \quad \text{has the augmented matrix} \quad \begin{bmatrix} 1 & 3 & 0 & 2 & \mathbf{1} \\ 0 & 0 & 1 & 4 & \mathbf{6} \\ 1 & 3 & 1 & 6 & \mathbf{7} \end{bmatrix} = \begin{bmatrix} A & b \end{bmatrix}.$$

The augmented matrix is just $\begin{bmatrix} A & b \end{bmatrix}$. When we apply the usual elimination steps to A, we also apply them to $\boldsymbol{b}$. That keeps all the equations correct.

In this example we subtract row 1 from row 3 and then subtract row 2 from row 3. This produces a *complete row of zeros* in R, and it changes $\boldsymbol{b}$ to a new right side $\boldsymbol{d} = (1, 6, 0)$:

$$\begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix} \quad \text{has the augmented matrix} \quad \begin{bmatrix} 1 & 3 & 0 & 2 & \mathbf{1} \\ 0 & 0 & 1 & 4 & \mathbf{6} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix} = \begin{bmatrix} R & d \end{bmatrix}.$$

That very last zero is crucial. The third equation has become $0 = 0$ and the equations can be solved. In the original matrix A, the first row plus the second row equals the third row. If the equations are consistent, this must be true on the right side of the equations also! The all-important property on the right side was $1 + 6 = 7$.

Here are the same augmented matrices for a general $\boldsymbol{b} = (b_1, b_2, b_3)$:

$$\begin{bmatrix} A & b \end{bmatrix} = \begin{bmatrix} 1 & 3 & 0 & 2 & b_1 \\ 0 & 0 & 1 & 4 & b_2 \\ 1 & 3 & 1 & 6 & b_3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 3 & 0 & 2 & b_1 \\ 0 & 0 & 1 & 4 & b_2 \\ 0 & 0 & 0 & 0 & b_3 - b_1 - b_2 \end{bmatrix} = \begin{bmatrix} R & d \end{bmatrix}$$

Now we get $0 = 0$ in the third equation provided $b_3 - b_1 - b_2 = 0$. This is $b_1 + b_2 = b_3$.

One Particular Solution

For an easy solution $\boldsymbol{x}$, *choose the free variables to be* $x_2 = x_4 = 0$. Then the two nonzero equations give the two pivot variables $x_1 = 1$ and $x_3 = 6$. Our particular solution to $Ax = b$ (and also $Rx = d$) is $x_p = (1, 0, 6, 0)$. This particular solution is my favorite: *free variables = zero, pivot variables from* $\boldsymbol{d}$. The method always works.

For a solution to exist, zero rows in R must also be zero in d. Since I is in the pivot rows and pivot columns of R, the pivot variables in $x_{\text{particular}}$ come from d:

$$Rx_p = \begin{bmatrix} \mathbf{1} & 3 & \mathbf{0} & 2 \\ \mathbf{0} & 0 & \mathbf{1} & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{1} \\ 0 \\ \mathbf{6} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{6} \\ 0 \end{bmatrix} \qquad \begin{matrix} \textbf{Pivot variables 1, 6} \\ \textbf{Free variables } 0, 0 \end{matrix}.$$

Notice how we *choose* the free variables (as zero) and *solve* for the pivot variables. After the row reduction to R, those steps are quick. When the free variables are zero, the pivot variables for x_p are already seen already seen in the right side vector d.

$x_{\text{particular}}$	***The particular solution solves***	$Ax_p = b$
$x_{\text{nullspace}}$	***The $n - r$ special solutions solve***	$Ax_n = 0.$

That particular solution is $(1, 0, 6, 0)$. The two special (nullspace) solutions to $Rx = 0$ come from the two free columns of R, by reversing signs of $3, 2$, and 4. ***Please notice how I write the complete solution $x_p + x_n$ to $Ax = b$:***

Complete solution one x_p many x_n

$$x = x_p + x_n = \begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -2 \\ 0 \\ -4 \\ 1 \end{bmatrix}.$$

Question Suppose A is a square invertible matrix, $m = n = r$. What are x_p and x_n?
Answer The particular solution is the one and *only* solution $A^{-1}b$. There are no special solutions or free variables. $R = I$ has no zero rows. The only vector in the nullspace is $x_n = 0$. The complete solution is $x = x_p + x_n = A^{-1}b + 0$.

This was the situation in Chapter 2. We didn't mention the nullspace in that chapter. $N(A)$ contained only the zero vector. Reduction goes from $\begin{bmatrix} A & b \end{bmatrix}$ to $\begin{bmatrix} I & A^{-1}b \end{bmatrix}$. The original $Ax = b$ is reduced all the way to $x = A^{-1}b$ which is d. This is a special case here, but square invertible matrices are the ones we see most often in practice. So they got their own chapter at the start of the book.

For small examples we can reduce $\begin{bmatrix} A & b \end{bmatrix}$ to $\begin{bmatrix} R & d \end{bmatrix}$. For a large matrix, MATLAB does it better. One particular solution (not necessarily ours) is $A\backslash b$ from backslash. Here is an example with *full column rank*. Both columns have pivots.

Example 1 Find the condition on (b_1, b_2, b_3) for $Ax = b$ to be solvable, if

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ -2 & -3 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

This condition puts b in the column space of A. Find the complete $x = x_p + x_n$.

Solution Use the augmented matrix, with its extra column $\boldsymbol{b}$. Subtract row 1 of $\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix}$ from row 2, and add 2 times row 1 to row 3 to reach $\begin{bmatrix} R & \boldsymbol{d} \end{bmatrix}$:

$$\begin{bmatrix} 1 & 1 & b_1 \\ 1 & 2 & b_2 \\ -2 & -3 & b_3 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & b_1 \\ 0 & 1 & b_2 - b_1 \\ 0 & -1 & b_3 + 2b_1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 2b_1 - b_2 \\ 0 & 1 & b_2 - b_1 \\ 0 & 0 & b_3 + b_1 + b_2 \end{bmatrix}.$$

The last equation is $0 = 0$ provided $b_3 + b_1 + b_2 = 0$. This is the condition to put $\boldsymbol{b}$ in the column space; then $A\boldsymbol{x} = \boldsymbol{b}$ will be solvable. The rows of A add to the zero row. So for consistency (these are equations!) the entries of $\boldsymbol{b}$ must also add to zero.

This example has no free variables since $n - r = 2 - 2$. Therefore no special solutions. The nullspace solution is $\boldsymbol{x}_n = \boldsymbol{0}$. The particular solution to $A\boldsymbol{x} = \boldsymbol{b}$ and $R\boldsymbol{x} = \boldsymbol{d}$ is at the top of the augmented column $\boldsymbol{d}$:

$$\textbf{Only solution} \qquad \boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_n = \begin{bmatrix} 2b_1 - b_2 \\ b_2 - b_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

If $b_3 + b_1 + b_2$ is not zero, there is no solution to $A\boldsymbol{x} = \boldsymbol{b}$ ($\boldsymbol{x}_p$ doesn't exist).

This example is typical of an extremely important case: A has *full column rank*. Every column has a pivot. *The rank is $r = n$.* The matrix is tall and thin ($m \geq n$). Row reduction puts I at the top, when A is reduced to R with rank n:

$$\textbf{Full column rank} \quad R = \begin{bmatrix} I \\ 0 \end{bmatrix} = \begin{bmatrix} n \text{ by } n \text{ identity matrix} \\ m - n \text{ rows of zeros} \end{bmatrix} \tag{1}$$

There are no free columns or free variables. The nullspace matrix is empty!

We will collect together the different ways of recognizing this type of matrix.

Every matrix A with **full column rank** ($r = n$) has all these properties:

1. All columns of A are pivot columns.
2. There are no free variables or special solutions.
3. The nullspace $\boldsymbol{N}(A)$ contains only the zero vector $\boldsymbol{x} = \boldsymbol{0}$.
4. If $A\boldsymbol{x} = \boldsymbol{b}$ has a solution (it might not) then it has only *one solution*.

In the essential language of the next section, **this A has *independent columns*.** $A\boldsymbol{x} = \boldsymbol{0}$ only happens when $\boldsymbol{x} = \boldsymbol{0}$. In Chapter 4 we will add one more fact to the list: *The square matrix $A^{\mathrm{T}}A$ is invertible when the rank is n.*

In this case the nullspace of A (and R) has shrunk to the zero vector. The solution to $A\boldsymbol{x} = \boldsymbol{b}$ is *unique* (if it exists). There will be $m - n$ (here $3 - 2$) zero rows in R. So there are $m - n$ conditions in order to have $0 = 0$ in those rows, and $\boldsymbol{b}$ in the column space. With full column rank, $A\boldsymbol{x} = \boldsymbol{b}$ has *one* solution or *no* solution ($m > n$ is overdetermined).

The Complete Solution

The other extreme case is full row rank. Now $Ax = b$ has *one or infinitely many* solutions. In this case A must be *short and wide* ($m \le n$). A matrix has ***full row rank*** if $r = m$ ("*independent rows*"). Every row has a pivot, and here is an example.

Example 2 There are $n = 3$ unknowns but only $m = 2$ equations:

$$\textbf{Full row rank} \qquad \begin{array}{rcrcrcl} x & + & y & + & z & = & 3 \\ x & + & 2y & - & z & = & 4 \end{array} \qquad (\text{rank } r = m = 2)$$

These are two planes in xyz space. The planes are not parallel so they intersect in a line. This line of solutions is exactly what elimination will find. ***The particular solution will be one point on the line. Adding the nullspace vectors x_n will move us along the line.*** Then $x = x_p + x_n$ gives the whole line of solutions.

We find x_p and x_n by elimination on $\begin{bmatrix} A & b \end{bmatrix}$. Subtract row 1 from row 2 and then subtract row 2 from row 1:

$$\begin{bmatrix} 1 & 1 & 1 & \mathbf{3} \\ 1 & 2 & -1 & \mathbf{4} \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 1 & \mathbf{3} \\ 0 & 1 & -2 & \mathbf{1} \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 3 & \mathbf{2} \\ 0 & 1 & -2 & \mathbf{1} \end{bmatrix} = \begin{bmatrix} R & d \end{bmatrix}.$$

The particular solution has free variable $x_3 = 0$. The special solution has $x_3 = 1$:

$x_{\textbf{particular}}$ comes directly from d on the right side: $x_p = (2, 1, 0)$
$x_{\textbf{special}}$ comes from the third column (free column) of R: $s = (-3, 2, 1)$

It is wise to check that x_p and s satisfy the original equations $Ax_p = b$ and $As = 0$:

$$\begin{array}{rcl} 2+1 & = & 3 \\ 2+2 & = & 4 \end{array} \qquad\qquad \begin{array}{rcl} -3+2+1 & = & 0 \\ -3+4-1 & = & 0 \end{array}$$

The nullspace solution x_n is any multiple of s. It moves along the line of solutions, starting at $x_{\textbf{particular}}$. ***Please notice again how to write the answer:***

$$\textbf{Complete solution} \qquad x = x_p + x_n = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}.$$

This line is drawn in Figure 3.3. Any point on the line *could* have been chosen as the particular solution; we chose the point with $x_3 = 0$.

The particular solution is *not* multiplied by an arbitrary constant! The special solution is, and you understand why.

Now we summarize this short wide case of *full row rank*. If $m < n$ the equation $Ax = b$ is **underdetermined** (many solutions).

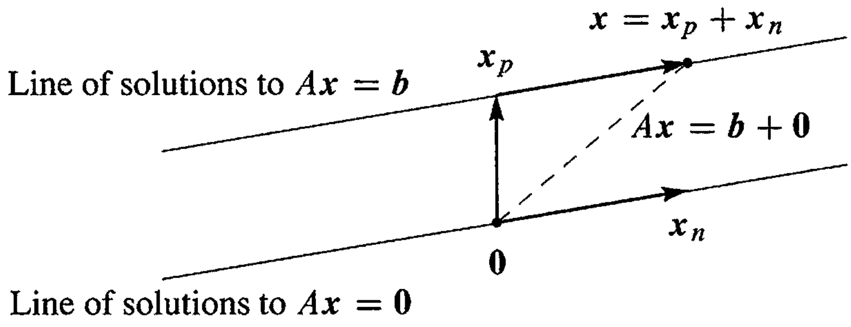


Figure 3.3: Complete solution = *one* particular solution + *all* nullspace solutions.

Every matrix A with ***full row rank*** $(r = m)$ has all these properties:

1. All rows have pivots, and R has no zero rows.

2. $Ax = b$ has a solution for every right side b.

3. The column space is the whole space $\mathbf{R}^m$.

4. There are $n - r = n - m$ special solutions in the nullspace of A.

In this case with m pivots, the rows are "***linearly independent***". So the columns of A^{T} are linearly independent. We are more than ready for the definition of linear independence, as soon as we summarize the four possibilities—which depend on the rank. Notice how r, m, n are the critical numbers.

The four possibilities for linear equations depend on the rank r:

$\boldsymbol{r = m}$	and	$\boldsymbol{r = n}$	*Square and invertible*	$Ax = b$	has 1 solution
$\boldsymbol{r = m}$	and	$r < n$	*Short and wide*	$Ax = b$	has ∞ solutions
$r < m$	and	$\boldsymbol{r = n}$	*Tall and thin*	$Ax = b$	has 0 or 1 solution
$r < m$	and	$r < n$	*Not full rank*	$Ax = b$	has 0 or ∞ solutions

The reduced R will fall in the same category as the matrix A. In case the pivot columns happen to come first, we can display these four possibilities for R. For $Rx = d$ (and the original $Ax = b$) to be solvable, d must end in $m - r$ zeros.

Four types	$R = \begin{bmatrix} I \end{bmatrix}$	$\begin{bmatrix} I & F \end{bmatrix}$	$\begin{bmatrix} I \\ 0 \end{bmatrix}$	$\begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}$
Their ranks	$\boldsymbol{r = m = n}$	$\boldsymbol{r = m < n}$	$\boldsymbol{r = n < m}$	$\boldsymbol{r < m, r < n}$

Cases 1 and 2 have full row rank $r = m$. Cases 1 and 3 have full column rank $r = n$. Case 4 is the most general in theory and it is the least common in practice.

Note My classes used to stop at U before reaching R. Instead of reading the complete solution directly from $Rx = d$, we found it by back substitution from $Ux = c$. This reduction to U and back substitution for x is slightly faster. Now we prefer the complete reduction: a single "1" in each pivot column. Everything is so clear in R (and the computer should do the hard work anyway) that we reduce all the way.

■ REVIEW OF THE KEY IDEAS ■

1. The rank r is the number of pivots. The matrix R has $m - r$ zero rows.
2. $Ax = b$ is solvable if and only if the last $m - r$ equations reduce to $0 = 0$.
3. One particular solution x_p has all free variables equal to zero.
4. The pivot variables are determined after the free variables are chosen.
5. Full column rank $r = n$ means no free variables: one solution or none.
6. Full row rank $r = m$ means one solution if $m = n$ or infinitely many if $m < n$.

■ WORKED EXAMPLES ■

3.4 A This question connects elimination (**pivot columns and back substitution**) to **column space-nullspace-rank-solvability** (the full picture). A has rank 2:

$$Ax = b \quad \text{is} \quad \begin{aligned} x_1 + 2x_2 + 3x_3 + 5x_4 &= b_1 \\ 2x_1 + 4x_2 + 8x_3 + 12x_4 &= b_2 \\ 3x_1 + 6x_2 + 7x_3 + 13x_4 &= b_3 \end{aligned}$$

1. Reduce $[\, A \;\; b \,]$ to $[\, U \;\; c \,]$, so that $Ax = b$ becomes a triangular system $Ux = c$.
2. Find the condition on b_1, b_2, b_3 for $Ax = b$ to have a solution.
3. Describe the column space of A. Which plane in $\mathbf{R}^3$?
4. Describe the nullspace of A. Which special solutions in $\mathbf{R}^4$?
5. Find a particular solution to $Ax = (0, 6, -6)$ and then the complete solution.
6. Reduce $[\, U \;\; c \,]$ to $[\, R \;\; d \,]$: Special solutions from R, particular solution from d.

Solution

1. The multipliers in elimination are 2 and 3 and -1. They take $[\, A \;\; b \,]$ into $[\, U \;\; c \,]$.

$$\begin{bmatrix} 1 & 2 & 3 & 5 & \mathbf{b_1} \\ 2 & 4 & 8 & 12 & \mathbf{b_2} \\ 3 & 6 & 7 & 13 & \mathbf{b_3} \end{bmatrix} \rightarrow \left[\begin{array}{cccc|c} 1 & 2 & 3 & 5 & \mathbf{b_1} \\ 0 & 0 & 2 & 2 & \mathbf{b_2 - 2b_1} \\ 0 & 0 & -2 & -2 & \mathbf{b_3 - 3b_1} \end{array}\right] \rightarrow \left[\begin{array}{cccc|c} 1 & 2 & 3 & 5 & \mathbf{b_1} \\ 0 & 0 & 2 & 2 & \mathbf{b_2 - 2b_1} \\ 0 & 0 & 0 & 0 & \mathbf{b_3 + b_2 - 5b_1} \end{array}\right]$$

2. The last equation shows the solvability condition $b_3 + b_2 - 5b_1 = 0$. Then $0 = 0$.
3. **First description**: The column space is the plane containing all combinations of the pivot columns $(1, 2, 3)$ and $(3, 8, 7)$. The pivots are in columns 1 and 3. **Second description**: The column space contains all vectors with $b_3 + b_2 - 5b_1 = 0$. That makes $\boldsymbol{Ax} = \boldsymbol{b}$ solvable, so $\boldsymbol{b}$ is in the column space. *All columns of A pass this test $b_3 + b_2 - 5b_1 = 0$. This is the equation for the plane in the first description*!
4. The special solutions have free variables $x_2 = 1, x_4 = 0$ and then $x_2 = 0, x_4 = 1$:

Special solutions to $A\boldsymbol{x} = \boldsymbol{0}$
Back substitution in $U\boldsymbol{x} = \boldsymbol{0}$

$$s_1 = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad s_2 = \begin{bmatrix} -2 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

The nullspace $\boldsymbol{N}(A)$ in $\mathbf{R}^4$ contains all $\boldsymbol{x}_n = c_1 s_1 + c_2 s_2$.

5. One particular solution $\boldsymbol{x}_p$ has free variables = zero. Back substitute in $U\boldsymbol{x} = \boldsymbol{c}$:

Particular solution to $A\boldsymbol{x}_p = \boldsymbol{b} = (0, 6, -6)$
This vector $\boldsymbol{b}$ satisfies $b_3 + b_2 - 5b_1 = 0$

$$\boldsymbol{x}_p = \begin{bmatrix} -9 \\ 0 \\ 3 \\ 0 \end{bmatrix}$$

The complete solution to $A\boldsymbol{x} = (0, 6, -6)$ is $\boldsymbol{x} = \boldsymbol{x}_p + \text{all } \boldsymbol{x}_n$.

6. In the reduced form R, the third column changes from $(3, 2, 0)$ in U to $(0, 1, 0)$. The right side $\boldsymbol{c} = (0, 6, 0)$ becomes $\boldsymbol{d} = (-9, 3, 0)$ showing -9 and 3 in $\boldsymbol{x}_p$:

$$[U \quad \boldsymbol{c}] = \begin{bmatrix} 1 & 2 & 3 & 5 & \mathbf{0} \\ 0 & 0 & 2 & 2 & \mathbf{6} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix} \longrightarrow [R \quad \boldsymbol{d}] = \begin{bmatrix} 1 & 2 & 0 & 2 & \mathbf{-9} \\ 0 & 0 & 1 & 1 & \mathbf{3} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix}$$

3.4 B If you have this information about the solutions to $A\boldsymbol{x} = \boldsymbol{b}$ for a specific $\boldsymbol{b}$, what does that tell you about the shape of A (and A itself)? And possibly about $\boldsymbol{b}$.

1. There is exactly one solution.
2. All solutions to $A\boldsymbol{x} = \boldsymbol{b}$ have the form $\boldsymbol{x} = \left[\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right] + c\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$.
3. There are no solutions.
4. All solutions to $A\boldsymbol{x} = \boldsymbol{b}$ have the form $\boldsymbol{x} = \left[\begin{smallmatrix} 1 \\ 1 \\ 0 \end{smallmatrix}\right] + c\left[\begin{smallmatrix} 1 \\ 0 \\ 1 \end{smallmatrix}\right]$
5. There are infinitely many solutions.

Solution In case **1**, with exactly one solution, A must have full column rank $r = n$. The nullspace of A contains only the zero vector. Necessarily $m \geq n$.

In case **2**, A must have $n = 2$ columns (and m is arbitrary). With $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ in the nullspace of A, column 2 is the *negative* of column 1. Also $A \neq 0$: the rank is 1. With $\boldsymbol{x} = \left[\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right]$ as a solution, $\boldsymbol{b} = 2(\text{column } 1) + (\text{column } 2)$. My choice for $\boldsymbol{x}_p$ would be $(1, 0)$.

In case **3** we only know that $\boldsymbol{b}$ is not in the column space of A. The rank of A must be less than m. I guess we know $\boldsymbol{b} \neq \boldsymbol{0}$, otherwise $\boldsymbol{x} = \boldsymbol{0}$ would be a solution.

In case **4**, A must have $n = 3$ columns. With $(1, 0, 1)$ in the nullspace of A, column 3 is the negative of column 1. Column 2 must *not* be a multiple of column 1, or the nullspace would contain another special solution. So the rank of A is $3 - 1 = 2$. Necessarily A has $m \geq 2$ rows. The right side $\boldsymbol{b}$ is column 1 + column 2.

In case **5** with infinitely many solutions, the nullspace must contain nonzero vectors. The rank r must be less than n (not full column rank), and $\boldsymbol{b}$ must be in the column space of A. We don't know if *every* $\boldsymbol{b}$ is in the column space, so we don't know if $r = m$.

3.4 C Find the complete solution $\boldsymbol{x} = \boldsymbol{x_p} + \boldsymbol{x_n}$ by forward elimination on $[A \ \ \boldsymbol{b}]$:

$$\begin{bmatrix} 1 & 2 & 1 & 0 \\ 2 & 4 & 4 & 8 \\ 4 & 8 & 6 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 10 \end{bmatrix}.$$

Find numbers y_1, y_2, y_3 so that y_1 (row 1) + y_2 (row 2) + y_3 (row 3) = ***zero row***. Check that $\boldsymbol{b} = (4, 2, 10)$ satisfies the condition $y_1b_1 + y_2b_2 + y_3b_3 = 0$. Why is this the condition for the equations to be solvable and $\boldsymbol{b}$ to be in the column space?

Solution Forward elimination on $[A \ \ \boldsymbol{b}]$ produces a zero row in $[U \ \ \boldsymbol{c}]$. The third equation becomes $0 = 0$ and the equations are consistent (and solvable):

$$\begin{bmatrix} 1 & 2 & 1 & 0 & \mathbf{4} \\ 2 & 4 & 4 & 8 & \mathbf{2} \\ 4 & 8 & 6 & 8 & \mathbf{10} \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 1 & 0 & \mathbf{4} \\ 0 & 0 & 2 & 8 & \mathbf{-6} \\ 0 & 0 & 2 & 8 & \mathbf{-6} \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 1 & 0 & \mathbf{4} \\ 0 & 0 & 2 & 8 & \mathbf{-6} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix}.$$

Columns 1 and 3 contain pivots. The variables x_2 and x_4 are free. If we set those to zero we can solve (back substitution) for the particular solution $\boldsymbol{x_p} = (7, 0, -3, 0)$. We see 7 and -3 again if elimination continues all the way to $[R \ \ \boldsymbol{d}]$:

$$\begin{bmatrix} 1 & 2 & 1 & 0 & \mathbf{4} \\ 0 & 0 & 2 & 8 & \mathbf{-6} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 1 & 0 & \mathbf{4} \\ 0 & 0 & 1 & 4 & \mathbf{-3} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 0 & -4 & \mathbf{7} \\ 0 & 0 & 1 & 4 & \mathbf{-3} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix}.$$

For the nullspace part $\boldsymbol{x_n}$ with $\boldsymbol{b} = \boldsymbol{0}$, set the free variables x_2, x_4 to $1, 0$ and also $0, 1$:

Special solutions $\qquad s_1 = (-2, 1, 0, 0)$ **and** $s_2 = (4, 0, -4, 1)$

Then the complete solution to $A\boldsymbol{x} = \boldsymbol{b}$ (and $R\boldsymbol{x} = \boldsymbol{d}$) is $\boldsymbol{x}_{\text{complete}} = \boldsymbol{x_p} + c_1s_1 + c_2s_2$.

The rows of A produced the zero row from 2(row 1) + (row 2) − (row 3) = $(0, 0, 0, 0)$. Thus $\boldsymbol{y} = (2, 1, -1)$. The same combination for $\boldsymbol{b} = (4, 2, 10)$ gives $2(4) + (2) - (10) = 0$.

If a combination of the rows (on the left side) gives the zero row, then the same combination must give zero on the right side. Of course! *Otherwise no solution.*

Later we will say this again in different words: If every column of A is perpendicular to $\boldsymbol{y} = (2, 1, -1)$, then any combination $\boldsymbol{b}$ of those columns must also be perpendicular to $\boldsymbol{y}$. Otherwise $\boldsymbol{b}$ is not in the column space and $A\boldsymbol{x} = \boldsymbol{b}$ is not solvable.

And again: If $\boldsymbol{y}$ is in the nullspace of A^{T} then $\boldsymbol{y}$ must be perpendicular to every $\boldsymbol{b}$ in the column space of A. Just looking ahead...

Problem Set 3.4

1 (Recommended) Execute the six steps of Worked Example **3.4 A** to describe the column space and nullspace of A and the complete solution to $A\boldsymbol{x} = \boldsymbol{b}$:

$$A = \begin{bmatrix} 2 & 4 & 6 & 4 \\ 2 & 5 & 7 & 6 \\ 2 & 3 & 5 & 2 \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 5 \end{bmatrix}$$

2 Carry out the same six steps for this matrix A with rank one. You will find *two* conditions on b_1, b_2, b_3 for $A\boldsymbol{x} = \boldsymbol{b}$ to be solvable. Together these two conditions put $\boldsymbol{b}$ into the _____ space (two planes give a line):

$$A = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \begin{bmatrix} 2 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 3 \\ 6 & 3 & 9 \\ 4 & 2 & 6 \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 30 \\ 20 \end{bmatrix}$$

Questions 3–15 are about the solution of $A\boldsymbol{x} = \boldsymbol{b}$. Follow the steps in the text to $\boldsymbol{x}_p$ and $\boldsymbol{x}_n$. Use the augmented matrix with last column $\boldsymbol{b}$.

3 Write the complete solution as $\boldsymbol{x}_p$ plus any multiple of $\boldsymbol{s}$ in the nullspace:

$$\begin{aligned} x + 3y + 3z &= 1 \\ 2x + 6y + 9z &= 5 \\ -x - 3y + 3z &= 5. \end{aligned}$$

4 Find the complete solution (also called the *general solution*) to

$$\begin{bmatrix} 1 & 3 & 1 & 2 \\ 2 & 6 & 4 & 8 \\ 0 & 0 & 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}.$$

5 Under what condition on b_1, b_2, b_3 is this system solvable? Include $\boldsymbol{b}$ as a fourth column in elimination. Find all solutions when that condition holds:

$$\begin{aligned} x + 2y - 2z &= b_1 \\ 2x + 5y - 4z &= b_2 \\ 4x + 9y - 8z &= b_3. \end{aligned}$$

6 What conditions on b_1, b_2, b_3, b_4 make each system solvable? Find $\boldsymbol{x}$ in that case:

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 2 & 5 \\ 3 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 2 & 5 & 7 \\ 3 & 9 & 12 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}.$$

7 Show by elimination that (b_1, b_2, b_3) is in the column space if $b_3 - 2b_2 + 4b_1 = 0$.

$$A = \begin{bmatrix} 1 & 3 & 1 \\ 3 & 8 & 2 \\ 2 & 4 & 0 \end{bmatrix}.$$

What combination of the rows of A gives the zero row?

8 Which vectors (b_1, b_2, b_3) are in the column space of A? Which combinations of the rows of A give zero?

$$\text{(a)}\ A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 6 & 3 \\ 0 & 2 & 5 \end{bmatrix} \qquad \text{(b)}\ A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 2 & 4 & 8 \end{bmatrix}.$$

9 (a) The Worked Example **3.4 A** reached $[\,U \ \ c\,]$ from $[\,A \ \ b\,]$. Put the multipliers into L and verify that LU equals A and Lc equals $\boldsymbol{b}$.

(b) Combine the pivot columns of A with the numbers -9 and 3 in the particular solution $\boldsymbol{x}_p$. What is that linear combination and why?

10 Construct a 2 by 3 system $A\boldsymbol{x} = \boldsymbol{b}$ with particular solution $\boldsymbol{x}_p = (2, 4, 0)$ and homogeneous solution $\boldsymbol{x}_n =$ any multiple of $(1, 1, 1)$.

11 Why can't a 1 by 3 system have $\boldsymbol{x}_p = (2, 4, 0)$ and $\boldsymbol{x}_n =$ any multiple of $(1, 1, 1)$?

12 (a) If $A\boldsymbol{x} = \boldsymbol{b}$ has two solutions $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, find two solutions to $A\boldsymbol{x} = \boldsymbol{0}$.

(b) Then find another solution to $A\boldsymbol{x} = \boldsymbol{0}$ and another solution to $A\boldsymbol{x} = \boldsymbol{b}$.

13 Explain why these are all false:

(a) The complete solution is any linear combination of $\boldsymbol{x}_p$ and $\boldsymbol{x}_n$.

(b) A system $A\boldsymbol{x} = \boldsymbol{b}$ has at most one particular solution.

(c) The solution $\boldsymbol{x}_p$ with all free variables zero is the shortest solution (minimum length $\|\boldsymbol{x}\|$). Find a 2 by 2 counterexample.

(d) If A is invertible there is no solution $\boldsymbol{x}_n$ in the nullspace.

14 Suppose column 5 of U has no pivot. Then x_5 is a ____ variable. The zero vector (is) (is not) the only solution to $A\boldsymbol{x} = \boldsymbol{0}$. If $A\boldsymbol{x} = \boldsymbol{b}$ has a solution, then it has ____ solutions.

15 Suppose row 3 of U has no pivot. Then that row is ____. The equation $U\boldsymbol{x} = \boldsymbol{c}$ is only solvable provided ____. The equation $A\boldsymbol{x} = \boldsymbol{b}$ (*is*) (*is not*) (*might not be*) solvable.

Questions 16–20 are about matrices of "full rank" $r = m$ or $r = n$.

16 The largest possible rank of a 3 by 5 matrix is ____. Then there is a pivot in every ____ of U and R. The solution to $A\boldsymbol{x} = \boldsymbol{b}$ (*always exists*) (*is unique*). The column space of A is ____. An example is $A =$ ____.

17 The largest possible rank of a 6 by 4 matrix is _____. Then there is a pivot in every _____ of U and R. The solution to $Ax = b$ (*always exists*) (*is unique*). The nullspace of A is _____. An example is $A =$ _____.

18 Find by elimination the rank of A and also the rank of A^{T}:

$$A = \begin{bmatrix} 1 & 4 & 0 \\ 2 & 11 & 5 \\ -1 & 2 & 10 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 1 & 1 & q \end{bmatrix} \text{ (rank depends on } q\text{).}$$

19 Find the rank of A and also of $A^{\mathrm{T}}A$ and also of AA^{T}:

$$A = \begin{bmatrix} 1 & 1 & 5 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

20 Reduce A to its echelon form U. Then find a triangular L so that $A = LU$.

$$A = \begin{bmatrix} 3 & 4 & 1 & 0 \\ 6 & 5 & 2 & 1 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 2 & 2 & 0 & 3 \\ 0 & 6 & 5 & 4 \end{bmatrix}.$$

21 Find the complete solution in the form $x_p + x_n$ to these full rank systems:

(a) $x + y + z = 4$ (b) $\begin{aligned} x + y + z &= 4 \\ x - y + z &= 4. \end{aligned}$

22 If $Ax = b$ has infinitely many solutions, why is it impossible for $Ax = B$ (new right side) to have only one solution? Could $Ax = B$ have no solution?

23 Choose the number q so that (if possible) the ranks are (a) 1, (b) 2, (c) 3:

$$A = \begin{bmatrix} 6 & 4 & 2 \\ -3 & -2 & -1 \\ 9 & 6 & q \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 & 1 & 3 \\ q & 2 & q \end{bmatrix}.$$

24 Give examples of matrices A for which the number of solutions to $Ax = b$ is

(a) 0 or 1, depending on b

(b) ∞, regardless of b

(c) 0 or ∞, depending on b

(d) 1, regardless of b.

25 Write down all known relations between r and m and n if $Ax = b$ has

(a) no solution for some b

(b) infinitely many solutions for every $\boldsymbol{b}$

(c) exactly one solution for some $\boldsymbol{b}$, no solution for other $\boldsymbol{b}$

(d) exactly one solution for every $\boldsymbol{b}$.

Questions 26–33 are about Gauss-Jordan elimination (upwards as well as downwards) and the reduced echelon matrix R.

26 Continue elimination from U to R. Divide rows by pivots so the new pivots are all 1. Then produce zeros *above* those pivots to reach R:

$$U = \begin{bmatrix} 2 & 4 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} 2 & 4 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 5 \end{bmatrix}.$$

27 Suppose U is square with n pivots (an invertible matrix). *Explain why $R = I$.*

28 Apply Gauss-Jordan elimination to $U\boldsymbol{x} = \boldsymbol{0}$ and $U\boldsymbol{x} = \boldsymbol{c}$. Reach $R\boldsymbol{x} = \boldsymbol{0}$ and $R\boldsymbol{x} = \boldsymbol{d}$:

$$\begin{bmatrix} U & \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & \mathbf{0} \\ 0 & 0 & 4 & \mathbf{0} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} U & \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & \mathbf{5} \\ 0 & 0 & 4 & \mathbf{8} \end{bmatrix}.$$

Solve $R\boldsymbol{x} = \boldsymbol{0}$ to find $\boldsymbol{x}_n$ (its free variable is $x_2 = 1$). Solve $R\boldsymbol{x} = \boldsymbol{d}$ to find $\boldsymbol{x}_p$ (its free variable is $x_2 = 0$).

29 Apply Gauss-Jordan elimination to reduce to $R\boldsymbol{x} = \boldsymbol{0}$ and $R\boldsymbol{x} = \boldsymbol{d}$:

$$\begin{bmatrix} U & 0 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 6 & \mathbf{0} \\ 0 & 0 & 2 & \mathbf{0} \\ 0 & 0 & 0 & \mathbf{0} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} U & \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} 3 & 0 & 6 & \mathbf{9} \\ 0 & 0 & 2 & \mathbf{4} \\ 0 & 0 & 0 & \mathbf{5} \end{bmatrix}.$$

Solve $U\boldsymbol{x} = \boldsymbol{0}$ or $R\boldsymbol{x} = \boldsymbol{0}$ to find $\boldsymbol{x}_n$ (free variable $= 1$). What are the solutions to $R\boldsymbol{x} = \boldsymbol{d}$?

30 Reduce to $U\boldsymbol{x} = \boldsymbol{c}$ (Gaussian elimination) and then $R\boldsymbol{x} = \boldsymbol{d}$ (Gauss-Jordan):

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 2 & 3 \\ 1 & 3 & 2 & 0 \\ 2 & 0 & 4 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 10 \end{bmatrix} = \boldsymbol{b}.$$

Find a particular solution $\boldsymbol{x}_p$ and all homogeneous solutions $\boldsymbol{x}_n$.

31 Find matrices A and B with the given property or explain why you can't:

(a) The only solution of $A\boldsymbol{x} = \begin{bmatrix} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \end{bmatrix}$ is $\boldsymbol{x} = \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}$.

(b) The only solution of $B\boldsymbol{x} = \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}$ is $\boldsymbol{x} = \begin{bmatrix} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \end{bmatrix}$.

32 Find the LU factorization of A and the complete solution to $Ax = b$:

$$A = \begin{bmatrix} 1 & 3 & 1 \\ 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 1 & 5 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 1 \\ 3 \\ 6 \\ 5 \end{bmatrix} \quad \text{and then} \quad b = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

33 The complete solution to $Ax = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ is $x = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Find A.

Challenge Problems

34 Suppose you know that the 3 by 4 matrix A has the vector $s = (2, 3, 1, 0)$ as the only special solution to $Ax = 0$.

(a) What is the *rank* of A and the complete solution to $Ax = 0$?

(b) What is the exact row reduced echelon form R of A?

(c) How do you know that $Ax = b$ can be solved for all b?

35 Suppose K is the 9 by 9 second difference matrix (2's on the diagonal, -1's on the diagonal above and also below). Solve the equation $Kx = b = (10, \ldots, 10)$. If you graph $x_1, \ldots, x_9$ above the points $1, \ldots, 9$ on the x axis, I think the nine points fall on a parabola.

36 Suppose $Ax = b$ and $Cx = b$ have the same (complete) solutions for every b. Is it true that $A = C$?

3.5 Independence, Basis and Dimension

This important section is about the true size of a subspace. There are n columns in an m by n matrix. But the true "dimension" of the column space is not necessarily n. The dimension is measured by counting ***independent columns***—and we have to say what that means. We will see that *the true dimension of the column space is the rank* r.

The idea of independence applies to any vectors $v_1, \ldots, v_n$ in any vector space. Most of this section concentrates on the subspaces that we know and use—especially the column space and the nullspace of A. In the last part we also study "vectors" that are not column vectors. They can be matrices and functions; they can be linearly independent (or dependent). First come the key examples using column vectors.

The goal is to understand a ***basis***: **independent vectors that "span the space".**

Every vector in the space is a unique combination of the basis vectors.

We are at the heart of our subject, and we cannot go on without a basis. The four essential ideas in this section (with first hints at their meaning) are:

1. Independent vectors	*(no extra vectors)*
2. Spanning a space	*(enough vectors to produce the rest)*
3. Basis for a space	*(not too many or too few)*
4. Dimension of a space	*(the number of vectors in a basis)*

Linear Independence

Our first definition of independence is not so conventional, but you are ready for it.

DEFINITION The columns of A are *linearly independent* when the only solution to $Ax = \mathbf{0}$ is $x = \mathbf{0}$. *No other combination Ax of the columns gives the zero vector.*

The columns are independent when the nullspace $N(A)$ contains only the zero vector. Let me illustrate linear independence (and dependence) with three vectors in $\mathbf{R}^3$:

1. If three vectors are *not* in the same plane, they are independent. No combination of v_1, v_2, v_3 in Figure 3.4 gives zero except $0v_1 + 0v_2 + 0v_3$.

2. If three vectors w_1, w_2, w_3 are *in the same plane*, they are dependent.

This idea of independence applies to 7 vectors in 12-dimensional space. If they are the columns of A, and independent, the nullspace only contains $x = \mathbf{0}$. None of the vectors is a combination of the other six vectors.

Now we choose different words to express the same idea. The following definition of independence will apply to any sequence of vectors in any vector space. When the vectors are the columns of A, the two definitions say exactly the same thing.

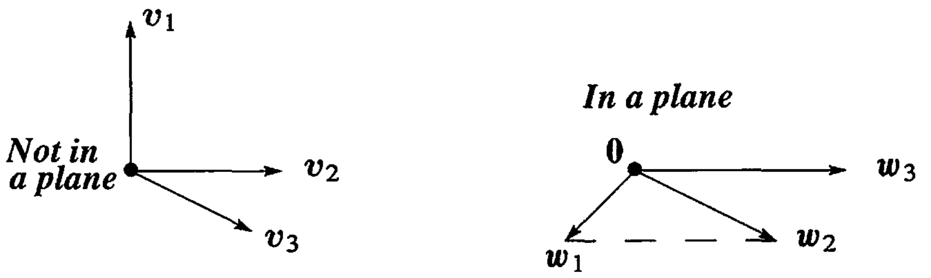


Figure 3.4: Independent vectors v_1, v_2, v_3. Only $0v_1 + 0v_2 + 0v_3$ gives the vector $\mathbf{0}$. Dependent vectors w_1, w_2, w_3. The combination $w_1 - w_2 + w_3$ is $(0, 0, 0)$.

DEFINITION The sequence of vectors $v_1, \ldots, v_n$ is ***linearly independent*** if the only combination that gives the zero vector is $0v_1 + 0v_2 + \cdots + 0v_n$.

Linear independence

$$x_1v_1 + x_2v_2 + \cdots + x_nv_n = \mathbf{0} \quad \text{only happens when all } x\text{'s are zero.} \tag{1}$$

If a combination gives $\mathbf{0}$, when the x's are not all zero, the vectors are *dependent*.

Correct language: "The sequence of vectors is linearly independent." *Acceptable shortcut*: "The vectors are independent." *Unacceptable*: "The matrix is independent."

A sequence of vectors is either dependent or independent. They can be combined to give the zero vector (with nonzero x's) or they can't. So the key question is: Which combinations of the vectors give zero? We begin with some small examples in $\mathbf{R}^2$:

(a) The vectors $(1, 0)$ and $(0, 1)$ are independent.

(b) The vectors $(1, 0)$ and $(1, 0.00001)$ are independent.

(c) The vectors $(1, 1)$ and $(-1, -1)$ are *dependent*.

(d) The vectors $(1, 1)$ and $(0, 0)$ are *dependent* because of the zero vector.

(e) In $\mathbf{R}^2$, any three vectors (a, b) and (c, d) and (e, f) are *dependent*.

Geometrically, $(1, 1)$ and $(-1, -1)$ are on a line through the origin. They are dependent. To use the definition, find numbers x_1 and x_2 so that $x_1(1, 1) + x_2(-1, -1) = (0, 0)$. This is the same as solving $Ax = \mathbf{0}$:

$$\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{for } x_1 = 1 \text{ and } x_2 = 1.$$

The columns are dependent exactly when *there is a nonzero vector in the nullspace.*

If one of the v's is the zero vector, independence has no chance. Why not?

Three vectors in $\mathbf{R}^2$ cannot be independent! One way to see this: the matrix A with those three columns must have a free variable and then a special solution to $Ax = \mathbf{0}$. Another way: If the first two vectors are independent, some combination will produce the third vector. See the second highlight below.

Now move to three vectors in $\mathbf{R}^3$. If one of them is a multiple of another one, these vectors are dependent. But the complete test involves all three vectors at once. We put them in a matrix and try to solve $Ax = \mathbf{0}$.

Example 1 The columns of this A are dependent. $Ax = \mathbf{0}$ has a nonzero solution:

$$Ax = \begin{bmatrix} 1 & 0 & 3 \\ 2 & 1 & 5 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \quad \text{is} \quad -3 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 3 \\ 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

The rank is only $r = 2$. *Independent columns produce full column rank* $r = n = 3$.

In that matrix the rows are also dependent. Row 1 minus row 3 is the zero row. For a *square matrix*, we will show that dependent columns imply dependent rows.

Question How to find that solution to $Ax = \mathbf{0}$? The systematic way is elimination.

$$A = \begin{bmatrix} 1 & 0 & 3 \\ 2 & 1 & 5 \\ 1 & 0 & 3 \end{bmatrix} \text{ reduces to } R = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}.$$

The solution $x = (-3, 1, 1)$ was exactly the special solution. It shows how the free column (column 3) is a combination of the pivot columns. That kills independence!

Full column rank The columns of A are independent exactly when the rank is $r = n$. There are n pivots and no free variables. Only $x = \mathbf{0}$ is in the nullspace.

One case is of special importance because it is clear from the start. Suppose seven columns have five components each ($m = 5$ is less than $n = 7$). Then the columns *must be dependent*. Any seven vectors from $\mathbf{R}^5$ are dependent. The rank of A cannot be larger than 5. There cannot be more than five pivots in five rows. $Ax = \mathbf{0}$ has at least $7 - 5 = 2$ free variables, so it has nonzero solutions—which means that the columns are dependent.

Any set of n vectors in $\mathbf{R}^m$ must be linearly dependent if $n > m$.

This type of matrix has more columns than rows—it is short and wide. The columns are certainly dependent if $n > m$, because $Ax = \mathbf{0}$ has a nonzero solution.

The columns might be dependent or might be independent if $n \leq m$. Elimination will reveal the r pivot columns. *It is those r pivot columns that are independent.*

Note Another way to describe linear dependence is this: "*One vector is a combination of the other vectors.*" That sounds clear. Why don't we say this from the start? Our definition was longer: "*Some combination gives the zero vector, other than the trivial combination with every* $x = 0$." We must rule out the easy way to get the zero vector.

That trivial combination of zeros gives every author a headache. If one vector is a combination of the others, that vector has coefficient $x = 1$.

The point is, our definition doesn't pick out one particular vector as guilty. All columns of A are treated the same. We look at $A\boldsymbol{x} = \mathbf{0}$, and it has a nonzero solution or it hasn't. In the end that is better than asking if the last column (or the first, or a column in the middle) is a combination of the others.

Vectors that Span a Subspace

The first subspace in this book was the column space. Starting with columns $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$, the subspace was filled out by including all combinations $x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n$. *The column space consists of all combinations $A\boldsymbol{x}$ of the columns.* We now introduce the single word "span" to describe this: The column space is ***spanned*** by the columns.

DEFINITION A set of vectors ***spans*** a space if their linear combinations fill the space.

The columns of a matrix span its column space. They might be dependent.

Example 2 $\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ span the full two-dimensional space $\mathbf{R}^2$.

Example 3 $\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \boldsymbol{v}_3 = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$ also span the full space $\mathbf{R}^2$.

Example 4 $\boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\boldsymbol{w}_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ only span a line in $\mathbf{R}^2$. So does $\boldsymbol{w}_1$ by itself.

Think of two vectors coming out from $(0, 0, 0)$ in 3-dimensional space. Generally they span a plane. Your mind fills in that plane by taking linear combinations. Mathematically you know other possibilities: two vectors could span a line, three vectors could span all of $\mathbf{R}^3$, or only a plane. It is even possible that three vectors span only a line, or ten vectors span only a plane. They are certainly not independent!

The columns span the column space. Here is a new subspace—*which is spanned by the rows.* ***The combinations of the rows produce the "row space".***

DEFINITION The ***row space*** of a matrix is the subspace of $\mathbf{R}^n$ spanned by the rows. ***The row space of A is $C(A^{\mathrm{T}})$. It is the column space of A^{T}.***

The rows of an m by n matrix have n components. They are vectors in $\mathbf{R}^n$—or they would be if they were written as column vectors. There is a quick way to fix that: *Transpose the matrix.* Instead of the rows of A, look at the columns of A^{T}. Same numbers, but now in the column space $C(A^{\mathrm{T}})$. This row space of A is a subspace of $\mathbf{R}^n$.

Example 5 Describe the column space and the row space of A.

$$A = \begin{bmatrix} 1 & 4 \\ 2 & 7 \\ 3 & 5 \end{bmatrix} \text{ and } A^{\mathrm{T}} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 7 & 5 \end{bmatrix}. \text{ Here } m = 3 \text{ and } n = 2.$$

The column space of A is the plane in $\mathbf{R}^3$ spanned by the two columns of A. *The row space of A is spanned by the three rows of A* (which are columns of A^{T}). This row space is all of $\mathbf{R}^2$. Remember: The rows are in $\mathbf{R}^n$ spanning the row space. The columns are in $\mathbf{R}^m$ spanning the column space. Same numbers, different vectors, different spaces.

A Basis for a Vector Space

Two vectors can't span all of $\mathbf{R}^3$, even if they are independent. Four vectors can't be independent, even if they span $\mathbf{R}^3$. We want ***enough independent vectors to span the space*** (and not more). A "***basis***" is just right.

DEFINITION A ***basis*** for a vector space is a sequence of vectors with two properties:

The basis vectors are linearly independent and they span the space.

This combination of properties is fundamental to linear algebra. Every vector v in the space is a combination of the basis vectors, because they span the space. More than that, the combination that produces v is *unique*, because the basis vectors $v_1, \ldots, v_n$ are independent:

There is one and only one way to write v as a combination of the basis vectors.

Reason: Suppose $v = a_1v_1 + \cdots + a_nv_n$ and also $v = b_1v_1 + \cdots + b_nv_n$. By subtraction $(a_1 - b_1)v_1 + \cdots + (a_n - b_n)v_n$ is the zero vector. From the independence of the v's, each $a_i - b_i = 0$. Hence $a_i = b_i$, and there are not two ways to produce v.

Example 6 The columns of $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ produce the "standard basis" for $\mathbf{R}^2$.

$$\text{The basis vectors } \quad \boldsymbol{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \text{ are independent. They span } \mathbf{R}^2.$$

Everybody thinks of this basis first. The vector $\boldsymbol{i}$ goes across and $\boldsymbol{j}$ goes straight up. The columns of the 3 by 3 identity matrix are the standard basis $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$. The columns of the n by n identity matrix give the "**standard basis**" for $\mathbf{R}^n$.

Now we find many other bases (infinitely many). The basis is not unique!

Example 7 (Important) The columns of *every invertible n by n matrix* give a basis for $\mathbf{R}^n$:

$$\begin{array}{ll} \textbf{Invertible matrix} & \\ \text{Independent columns} & A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \\ \text{Column space is } \mathbf{R}^3 & \end{array} \qquad \begin{array}{ll} \textbf{Singular matrix} & \\ \text{Dependent columns} & B = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 1 & 1 & 2 \end{bmatrix}. \\ \text{Column space} \neq \mathbf{R}^3 & \end{array}$$

The only solution to $Ax = \mathbf{0}$ is $x = A^{-1}\mathbf{0} = \mathbf{0}$. The columns are independent. They span the whole space $\mathbf{R}^n$—because every vector $\boldsymbol{b}$ is a combination of the columns. $Ax = \boldsymbol{b}$ can always be solved by $x = A^{-1}\boldsymbol{b}$. Do you see how everything comes together for invertible matrices? Here it is in one sentence:

The vectors $v_1, \ldots, v_n$ are a ***basis for*** $\mathbf{R}^n$ exactly when they are ***the columns of an n by n invertible matrix***. Thus $\mathbf{R}^n$ has infinitely many different bases.

When the columns are dependent, we keep only the *pivot columns*—the first two columns of B above, with its two pivots. They are independent and they span the column space.

The pivot columns of A are a basis for its column space. The pivot rows of A are a basis for its row space. So are the pivot rows of its echelon form R.

Example 8 This matrix is not invertible. Its columns are not a basis for anything!

$$\begin{array}{l} \textbf{One pivot column} \\ \textbf{One pivot row } (r = 1) \end{array} \qquad A = \begin{bmatrix} 2 & 4 \\ 3 & 6 \end{bmatrix} \text{ reduces to } R = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}.$$

Column 1 of A is the pivot column. That column alone is a basis for its column space. The second column of A would be a different basis. So would any nonzero multiple of that column. There is no shortage of bases. One definite choice is the pivot columns.

Notice that the pivot column $(1, 0)$ of this R ends in zero. That column is a basis for the column space of R, but it doesn't belong to the column space of A. The column spaces of A and R are different. Their bases are different. (Their dimensions are the same.)

The row space of A is the *same* as the row space of R. It contains $(2, 4)$ and $(1, 2)$ and all other multiples of those vectors. As always, there are infinitely many bases to choose from. One natural choice is to pick the nonzero rows of R (rows with a pivot). So this matrix A with rank one has only one vector in the basis:

$$\text{Basis for the column space: } \begin{bmatrix} 2 \\ 3 \end{bmatrix}. \quad \text{Basis for the row space: } \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

The next chapter will come back to these bases for the column space and row space. We are happy first with examples where the situation is clear (and the idea of a basis is still new). The next example is larger but still clear.

Example 9 Find bases for the column and row spaces of this rank two matrix:

$$R = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Columns 1 and 3 are the pivot columns. They are a basis for the column space (of R!). The vectors in that column space all have the form $\boldsymbol{b} = (x, y, 0)$. The column space of R is the "xy plane" inside the full 3-dimensional xyz space. That plane is not $\mathbf{R}^2$, it is a subspace of $\mathbf{R}^3$. Columns 2 and 3 are also a basis for the same column space. Which pairs of columns of R are *not* a basis for its column space?

The row space of R is a subspace of $\mathbf{R}^4$. The simplest basis for that row space is the two nonzero rows of R. The third row (the zero vector) is in the row space too. But it is not in a *basis* for the row space. The basis vectors must be independent.

Question Given five vectors in $\mathbf{R}^7$, ***how do you find a basis for the space they span?***

First answer Make them the rows of A, and eliminate to find the nonzero rows of R.
Second answer Put the five vectors into the columns of A. Eliminate to find the pivot columns (of A not R). The program **colbasis** uses the column numbers from *pivcol.*

Could another basis have more vectors, or fewer? This is a crucial question with a good answer: *No.* ***All bases for a vector space contain the same number of vectors.***

The number of vectors, in any and every basis, is the "dimension" of the space.

Dimension of a Vector Space

We have to prove what was just stated. There are many choices for the basis vectors, but the *number* of basis vectors doesn't change.

If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ and $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$ are both bases for the same vector space, then $m = n$.

Proof Suppose that there are more $\boldsymbol{w}$'s than $\boldsymbol{v}$'s. From $n > m$ we want to reach a contradiction. The $\boldsymbol{v}$'s are a basis, so $\boldsymbol{w}_1$ must be a combination of the $\boldsymbol{v}$'s. If $\boldsymbol{w}_1$ equals $a_{11}\boldsymbol{v}_1 + \cdots + a_{m1}\boldsymbol{v}_m$, this is the first column of a matrix multiplication VA:

Each $\boldsymbol{w}$ is a combination of the $\boldsymbol{v}$'s

$$W = \begin{bmatrix} & & & \\ \boldsymbol{w}_1 & \boldsymbol{w}_2 & \ldots & \boldsymbol{w}_n \\ & & & \end{bmatrix} = \begin{bmatrix} & & \\ \boldsymbol{v}_1 & \ldots & \boldsymbol{v}_m \\ & & \end{bmatrix} \begin{bmatrix} a_{11} & & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & & a_{mn} \end{bmatrix} = VA.$$

We don't know each a_{ij}, but we know the shape of A (it is m by n). The second vector $\boldsymbol{w}_2$ is also a combination of the $\boldsymbol{v}$'s. The coefficients in that combination fill the second column of A. The key is that A has a row for every $\boldsymbol{v}$ and a column for every $\boldsymbol{w}$. A is a short wide matrix, since we assumed $n > m$. So $A\boldsymbol{x} = \boldsymbol{0}$ ***has a nonzero solution.***

$A\boldsymbol{x} = 0$ gives $VA\boldsymbol{x} = \boldsymbol{0}$ which is $W\boldsymbol{x} = \boldsymbol{0}$. *A combination of the $\boldsymbol{w}$'s gives zero*! Then the $\boldsymbol{w}$'s could not be a basis—our assumption $n > m$ is **not possible** for two bases.

If $m > n$ we exchange the $\boldsymbol{v}$'s and $\boldsymbol{w}$'s and repeat the same steps. The only way to avoid a contradiction is to have $m = n$. This completes the proof that $m = n$.

The number of basis vectors depends on the space—not on a particular basis. The number is the same for every basis, and it counts the "degrees of freedom" in the space.

The dimension of the space $\mathbf{R}^n$ is n. We now introduce the important word ***dimension*** for other vector spaces too.

DEFINITION The ***dimension of a space*** is the ***number of vectors*** in every basis.

This matches our intuition. The line through $v = (1, 5, 2)$ has dimension one. It is a subspace with this one vector v in its basis. Perpendicular to that line is the plane $x + 5y + 2z = 0$. This plane has dimension 2. To prove it, we find a basis $(-5, 1, 0)$ and $(-2, 0, 1)$. The dimension is 2 because the basis contains two vectors.

The plane is the nullspace of the matrix $A = \begin{bmatrix} 1 & 5 & 2 \end{bmatrix}$, which has two free variables. Our basis vectors $(-5, 1, 0)$ and $(-2, 0, 1)$ are the "special solutions" to $Ax = \mathbf{0}$. The next section shows that the $n - r$ special solutions always give *a basis for the nullspace.* $\boldsymbol{C}(A)$ has dimension r and the nullspace $\boldsymbol{N}(A)$ has dimension $n - r$.

Note about the language of linear algebra We never say "the rank of a space" or "the dimension of a basis" or "the basis of a matrix". Those terms have no meaning. It is the ***dimension of the column space*** that equals the ***rank of the matrix.***

Bases for Matrix Spaces and Function Spaces

The words "independence" and "basis" and "dimension" are not at all restricted to column vectors. We can ask whether three matrices A_1, A_2, A_3 are independent. When they are in the space of all 3 by 4 matrices, some combination might give the zero matrix. We can also ask the dimension of the full 3 by 4 matrix space. (It is 12.)

In differential equations, $d^2y/dx^2 = y$ has a space of solutions. One basis is $y = e^x$ and $y = e^{-x}$. Counting the basis functions gives the dimension 2 for the space of all solutions. (The dimension is 2 because of the second derivative.)

Matrix spaces and function spaces may look a little strange after $\mathbf{R}^n$. But in some way, you haven't got the ideas of basis and dimension straight until you can apply them to "vectors" other than column vectors.

Matrix spaces The vector space $\mathbf{M}$ contains all 2 by 2 matrices. Its dimension is 4.

$$\textbf{One basis is} \quad A_1, A_2, A_3, A_4 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Those matrices are linearly independent. We are not looking at their columns, but at the whole matrix. Combinations of those four matrices can produce any matrix in $\mathbf{M}$, so they span the space:

$$\begin{matrix}\textbf{Every } A \textbf{ combines} \\ \textbf{the basis matrices}\end{matrix} \qquad c_1A_1 + c_2A_2 + c_3A_3 + c_4A_4 = \begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix} = A.$$

A is zero only if the c's are all zero—this proves independence of A_1, A_2, A_3, A_4.

The three matrices A_1, A_2, A_4 are a basis for a subspace—the upper triangular matrices. Its dimension is 3. A_1 and A_4 are a basis for the diagonal matrices. What is a basis for the symmetric matrices? Keep A_1 and A_4, and throw in $A_2 + A_3$.

To push this further, think about the space of all n by n matrices. One possible basis uses matrices that have only a single nonzero entry (that entry is 1). There are n^2 positions for that 1, so there are n^2 basis matrices:

The dimension of the whole n by n matrix space is n^2.

The dimension of the subspace of *upper triangular* matrices is $\frac{1}{2}n^2 + \frac{1}{2}n$.

The dimension of the subspace of *diagonal* matrices is n.

The dimension of the subspace of *symmetric* matrices is $\frac{1}{2}n^2 + \frac{1}{2}n$ (why?).

Function spaces The equations $d^2y/dx^2 = 0$ and $d^2y/dx^2 = -y$ and $d^2y/dx^2 = y$ involve the second derivative. In calculus we solve to find the functions $y(x)$:

$$\begin{array}{ll} y'' = 0 & \text{is solved by any linear function } y = cx + d \\ y'' = -y & \text{is solved by any combination } y = c\sin x + d\cos x \\ y'' = y & \text{is solved by any combination } y = ce^x + de^{-x}. \end{array}$$

That solution space for $y'' = -y$ has two basis functions: $\sin x$ and $\cos x$. The space for $y'' = 0$ has x and 1. It is the "nullspace" of the second derivative! The dimension is 2 in each case (these are second-order equations).

The solutions of $y'' = 2$ don't form a subspace—the right side $b = 2$ is not zero. A particular solution is $y(x) = x^2$. The complete solution is $y(x) = x^2 + cx + d$. All those functions satisfy $y'' = 2$. Notice the particular solution plus any function $cx + d$ in the nullspace. A linear differential equation is like a linear matrix equation $Ax = \boldsymbol{b}$. But we solve it by calculus instead of linear algebra.

We end here with the space $\mathbf{Z}$ that contains only the zero vector. The dimension of this space is *zero*. ***The empty set*** (containing no vectors) ***is a basis for*** $\mathbf{Z}$. We can never allow the zero vector into a basis, because then linear independence is lost.

■ REVIEW OF THE KEY IDEAS ■

1. The columns of A are ***independent*** if $\boldsymbol{x} = \mathbf{0}$ is the only solution to $A\boldsymbol{x} = \mathbf{0}$.

2. The vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ ***span*** a space if their combinations fill that space.

3. ***A basis consists of linearly independent vectors that span the space.*** Every vector in the space is a *unique* combination of the basis vectors.

4. All bases for a space have the same number of vectors. This number of vectors in a basis is the ***dimension*** of the space.

5. The pivot columns are one basis for the column space. The dimension is r.

■ WORKED EXAMPLES ■

3.5 A Start with the vectors $v_1 = (1,2,0)$ and $v_2 = (2,3,0)$. **(a)** Are they linearly independent? **(b)** Are they a basis for any space? **(c)** What space $\mathbf{V}$ do they span? **(d)** What is the dimension of $\mathbf{V}$? **(e)** Which matrices A have $\mathbf{V}$ as their column space? **(f)** Which matrices have $\mathbf{V}$ as their nullspace? **(g)** Describe all vectors v_3 that complete a basis v_1, v_2, v_3 for $\mathbf{R}^3$.

Solution

(a) v_1 and v_2 are independent—the only combination to give $\mathbf{0}$ is $0v_1 + 0v_2$.

(b) Yes, they are a basis for the space they span.

(c) That space $\mathbf{V}$ contains all vectors $(x, y, 0)$. It is the xy plane in $\mathbf{R}^3$.

(d) The dimension of $\mathbf{V}$ is 2 since the basis contains two vectors.

(e) This $\mathbf{V}$ is the column space of any 3 by n matrix A of rank 2, if every column is a combination of v_1 and v_2. In particular A could just have columns v_1 and v_2.

(f) This $\mathbf{V}$ is the nullspace of any m by 3 matrix B of rank 1, if every row is a multiple of $(0,0,1)$. In particular take $B = [0\ \ 0\ \ 1]$. Then $Bv_1 = \mathbf{0}$ and $Bv_2 = \mathbf{0}$.

(g) Any third vector $v_3 = (a, b, c)$ will complete a basis for $\mathbf{R}^3$ provided $c \neq 0$.

3.5 B Start with three independent vectors w_1, w_2, w_3. Take combinations of those vectors to produce v_1, v_2, v_3. Write the combinations in matrix form as $V = WM$:

$$\begin{aligned} v_1 &= w_1 + w_2 \\ v_2 &= w_1 + 2w_2 + w_3 \\ v_3 &= \phantom{w_1 +{}} w_2 + cw_3 \end{aligned} \quad \text{which is} \quad \begin{bmatrix} & & \\ v_1 & v_2 & v_3 \\ & & \end{bmatrix} = \begin{bmatrix} & & \\ w_1 & w_2 & w_3 \\ & & \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & c \end{bmatrix}$$

What is the test on a matrix V to see if its columns are linearly independent? If $c \neq 1$ show that v_1, v_2, v_3 are linearly independent. If $c = 1$ show that the v's are linearly *dependent*.

Solution The test on V for independence of its columns was in our first definition: *The nullspace of V must contain only the zero vector.* Then $x = (0,0,0)$ is the only combination of the columns that gives $Vx =$ zero vector.

If $c = 1$ in our problem, we can see *dependence* in two ways. First, $v_1 + v_3$ will be the same as v_2. (If you add $w_1 + w_2$ to $w_2 + w_3$ you get $w_1 + 2w_2 + w_3$ which is v_2.) In other words $v_1 - v_2 + v_3 = \mathbf{0}$—which says that the v's are not independent.

The other way is to look at the nullspace of M. If $c = 1$, the vector $x = (1,-1,1)$ is in that nullspace, and $Mx = \mathbf{0}$. Then certainly $WMx = \mathbf{0}$ which is the same as $Vx = 0$. So the v's are dependent. This specific $x = (1,-1,1)$ from the nullspace tells us again that $v_1 - v_2 + v_3 = \mathbf{0}$.

Now suppose $c \neq 1$. Then the matrix M is invertible. So if $\boldsymbol{x}$ is *any nonzero vector* we know that $M\boldsymbol{x}$ is nonzero. Since the $\boldsymbol{w}$'s are given as independent, we further know that $WM\boldsymbol{x}$ is nonzero. Since $V = WM$, this says that $\boldsymbol{x}$ is *not* in the nullspace of V. In other words $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ are independent.

The general rule is "independent $\boldsymbol{v}$'s from independent $\boldsymbol{w}$'s when M is invertible". And if these vectors are in $\mathbf{R}^3$, they are not only independent—they are a basis for $\mathbf{R}^3$. "Basis of $\boldsymbol{v}$'s from basis of $\boldsymbol{w}$'s when the change of basis matrix M is invertible."

3.5 C (***Important example***) Suppose $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ is a basis for $\mathbf{R}^n$ and the n by n matrix A is invertible. Show that $A\boldsymbol{v}_1, \ldots, A\boldsymbol{v}_n$ is also a basis for $\mathbf{R}^n$.

Solution In *matrix language*: Put the basis vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ in the columns of an invertible(!) matrix V. Then $A\boldsymbol{v}_1, \ldots, A\boldsymbol{v}_n$ are the columns of AV. Since A is invertible, so is AV and its columns give a basis.

In *vector language*: Suppose $c_1 A\boldsymbol{v}_1 + \cdots + c_n A\boldsymbol{v}_n = \mathbf{0}$. This is $A\boldsymbol{v} = \mathbf{0}$ with $\boldsymbol{v} = c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n$. Multiply by A^{-1} to reach $\boldsymbol{v} = \mathbf{0}$. By linear independence of the $\boldsymbol{v}$'s, all $c_i = 0$. This shows that the $A\boldsymbol{v}$'s are independent.

To show that the $A\boldsymbol{v}$'s span $\mathbf{R}^n$, solve $c_1 A\boldsymbol{v}_1 + \cdots + c_n A\boldsymbol{v}_n = \boldsymbol{b}$ which is the same as $c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n = A^{-1}\boldsymbol{b}$. Since the $\boldsymbol{v}$'s are a basis, this must be solvable.

Problem Set 3.5

Questions 1–10 are about linear independence and linear dependence.

1 Show that $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ are independent but $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ are dependent:

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \boldsymbol{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \boldsymbol{v}_4 = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}.$$

Solve $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + c_4\boldsymbol{v}_4 = \mathbf{0}$ or $A\boldsymbol{x} = \mathbf{0}$. The $\boldsymbol{v}$'s go in the columns of A.

2 (Recommended) Find the largest possible number of independent vectors among

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} \quad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} \quad \boldsymbol{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} \quad \boldsymbol{v}_4 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} \quad \boldsymbol{v}_5 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix} \quad \boldsymbol{v}_6 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$$

3 Prove that if $a = 0$ or $d = 0$ or $f = 0$ (3 cases), the columns of U are dependent:

$$U = \begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix}.$$

4 If a, d, f in Question 3 are all nonzero, show that the only solution to $Ux = \mathbf{0}$ is $x = \mathbf{0}$. Then the upper triangular U has independent columns.

5 Decide the dependence or independence of

(a) the vectors $(1,3,2)$ and $(2,1,3)$ and $(3,2,1)$

(b) the vectors $(1,-3,2)$ and $(2,1,-3)$ and $(-3,2,1)$.

6 Choose three independent columns of U. Then make two other choices. Do the same for A.

$$U = \begin{bmatrix} 2 & 3 & 4 & 1 \\ 0 & 6 & 7 & 0 \\ 0 & 0 & 0 & 9 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 2 & 3 & 4 & 1 \\ 0 & 6 & 7 & 0 \\ 0 & 0 & 0 & 9 \\ 4 & 6 & 8 & 2 \end{bmatrix}.$$

7 If $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3$ are independent vectors, show that the differences $\boldsymbol{v}_1 = \boldsymbol{w}_2 - \boldsymbol{w}_3$ and $\boldsymbol{v}_2 = \boldsymbol{w}_1 - \boldsymbol{w}_3$ and $\boldsymbol{v}_3 = \boldsymbol{w}_1 - \boldsymbol{w}_2$ are *dependent*. Find a combination of the $\boldsymbol{v}$'s that gives zero. Which matrix A in $[\,\boldsymbol{v}_1 \ \boldsymbol{v}_2 \ \boldsymbol{v}_3\,] = [\,\boldsymbol{w}_1 \ \boldsymbol{w}_2 \ \boldsymbol{w}_3\,]\,A$ is singular?

8 If $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3$ are independent vectors, show that the sums $\boldsymbol{v}_1 = \boldsymbol{w}_2 + \boldsymbol{w}_3$ and $\boldsymbol{v}_2 = \boldsymbol{w}_1 + \boldsymbol{w}_3$ and $\boldsymbol{v}_3 = \boldsymbol{w}_1 + \boldsymbol{w}_2$ are *independent*. (Write $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = \mathbf{0}$ in terms of the $\boldsymbol{w}$'s. Find and solve equations for the c's, to show they are zero.)

9 Suppose $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ are vectors in $\mathbf{R}^3$.

(a) These four vectors are dependent because _____.

(b) The two vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ will be dependent if _____.

(c) The vectors $\boldsymbol{v}_1$ and $(0,0,0)$ are dependent because _____.

10 Find two independent vectors on the plane $x + 2y - 3z - t = 0$ in $\mathbf{R}^4$. Then find three independent vectors. Why not four? This plane is the nullspace of what matrix?

Questions 11–15 are about the space *spanned* by a set of vectors. Take all linear combinations of the vectors.

11 Describe the subspace of $\mathbf{R}^3$ (is it a line or plane or $\mathbf{R}^3$?) spanned by

(a) the two vectors $(1,1,-1)$ and $(-1,-1,1)$

(b) the three vectors $(0,1,1)$ and $(1,1,0)$ and $(0,0,0)$

(c) all vectors in $\mathbf{R}^3$ with whole number components

(d) all vectors with positive components.

12 The vector $\boldsymbol{b}$ is in the subspace spanned by the columns of A when _____ has a solution. The vector $\boldsymbol{c}$ is in the row space of A when _____ has a solution.

True or false: If the zero vector is in the row space, the rows are dependent.

13 Find the dimensions of these 4 spaces. Which two of the spaces are the same? (a) column space of A, (b) column space of U, (c) row space of A, (d) row space of U:

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 3 & 1 \\ 3 & 1 & -1 \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

14 $\boldsymbol{v} + \boldsymbol{w}$ and $\boldsymbol{v} - \boldsymbol{w}$ are combinations of $\boldsymbol{v}$ and $\boldsymbol{w}$. Write $\boldsymbol{v}$ and $\boldsymbol{w}$ as combinations of $\boldsymbol{v} + \boldsymbol{w}$ and $\boldsymbol{v} - \boldsymbol{w}$. The two pairs of vectors _____ the same space. When are they a basis for the same space?

Questions 15–25 are about the requirements for a basis.

15 If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are linearly independent, the space they span has dimension _____. These vectors are a _____ for that space. If the vectors are the columns of an m by n matrix, then m is _____ than n. If $m = n$, that matrix is _____.

16 Find a basis for each of these subspaces of $\mathbf{R}^4$:

(a) All vectors whose components are equal.

(b) All vectors whose components add to zero.

(c) All vectors that are perpendicular to $(1, 1, 0, 0)$ and $(1, 0, 1, 1)$.

(d) The column space and the nullspace of I (4 by 4).

17 Find three different bases for the column space of $U = \left[\begin{smallmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{smallmatrix}\right]$. Then find two different bases for the row space of U.

18 Suppose $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_6$ are six vectors in $\mathbf{R}^4$.

(a) Those vectors (do)(do not)(might not) span $\mathbf{R}^4$.

(b) Those vectors (are)(are not)(might be) linearly independent.

(c) Any four of those vectors (are)(are not)(might be) a basis for $\mathbf{R}^4$.

19 The columns of A are n vectors from $\mathbf{R}^m$. If they are linearly independent, what is the rank of A? If they span $\mathbf{R}^m$, what is the rank? If they are a basis for $\mathbf{R}^m$, what then? *Looking ahead*: The rank r counts the number of _____ columns.

20 Find a basis for the plane $x - 2y + 3z = 0$ in $\mathbf{R}^3$. Then find a basis for the intersection of that plane with the xy plane. Then find a basis for all vectors perpendicular to the plane.

21 Suppose the columns of a 5 by 5 matrix A are a basis for $\mathbf{R}^5$.

(a) The equation $A\boldsymbol{x} = \boldsymbol{0}$ has only the solution $\boldsymbol{x} = \boldsymbol{0}$ because _____.

(b) If $\boldsymbol{b}$ is in $\mathbf{R}^5$ then $A\boldsymbol{x} = \boldsymbol{b}$ is solvable because the basis vectors _____ $\mathbf{R}^5$.

Conclusion: A is invertible. Its rank is 5. Its rows are also a basis for $\mathbf{R}^5$.

22 Suppose S is a 5-dimensional subspace of $\mathbf{R}^6$. True or false (example if false):

(a) Every basis for S can be extended to a basis for $\mathbf{R}^6$ by adding one more vector.

(b) Every basis for $\mathbf{R}^6$ can be reduced to a basis for S by removing one vector.

23 U comes from A by subtracting row 1 from row 3:

$$A = \begin{bmatrix} 1 & 3 & 2 \\ 0 & 1 & 1 \\ 1 & 3 & 2 \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} 1 & 3 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Find bases for the two column spaces. Find bases for the two row spaces. Find bases for the two nullspaces. Which spaces stay fixed in elimination?

24 True or false (give a good reason):

(a) If the columns of a matrix are dependent, so are the rows.

(b) The column space of a 2 by 2 matrix is the same as its row space.

(c) The column space of a 2 by 2 matrix has the same dimension as its row space.

(d) The columns of a matrix are a basis for the column space.

25 For which numbers c and d do these matrices have rank 2?

$$A = \begin{bmatrix} 1 & 2 & 5 & 0 & 5 \\ 0 & 0 & c & 2 & 2 \\ 0 & 0 & 0 & d & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} c & d \\ d & c \end{bmatrix}.$$

Questions 26–30 are about spaces where the "vectors" are matrices.

26 Find a basis (and the dimension) for each of these subspaces of 3 by 3 matrices:

(a) All diagonal matrices.

(b) All symmetric matrices ($A^{\mathrm{T}} = A$).

(c) All skew-symmetric matrices ($A^{\mathrm{T}} = -A$).

27 Construct six linearly independent 3 by 3 echelon matrices $U_1, \ldots, U_6$.

28 Find a basis for the space of all 2 by 3 matrices whose columns add to zero. Find a basis for the subspace whose rows also add to zero.

29 What subspace of 3 by 3 matrices is spanned (take all combinations) by

(a) the invertible matrices?

(b) the rank one matrices?

(c) the identity matrix?

30 Find a basis for the space of 2 by 3 matrices whose nullspace contains $(2, 1, 1)$.

Questions 31–35 are about spaces where the "vectors" are functions.

31 (a) Find all functions that satisfy $\frac{dy}{dx} = 0$.

(b) Choose a particular function that satisfies $\frac{dy}{dx} = 3$.

(c) Find all functions that satisfy $\frac{dy}{dx} = 3$.

32 The cosine space $\mathbf{F}_3$ contains all combinations $y(x) = A\cos x + B\cos 2x + C\cos 3x$. Find a basis for the subspace with $y(0) = 0$.

33 Find a basis for the space of functions that satisfy

(a) $\frac{dy}{dx} - 2y = 0$

(b) $\frac{dy}{dx} - \frac{y}{x} = 0$.

34 Suppose $y_1(x), y_2(x), y_3(x)$ are three different functions of x. The vector space they span could have dimension 1, 2, or 3. Give an example of y_1, y_2, y_3 to show each possibility.

35 Find a basis for the space of polynomials $p(x)$ of degree ≤ 3. Find a basis for the subspace with $p(1) = 0$.

36 Find a basis for the space $\mathbf{S}$ of vectors (a, b, c, d) with $a + c + d = 0$ and also for the space $\mathbf{T}$ with $a + b = 0$ and $c = 2d$. What is the dimension of the intersection $\mathbf{S} \cap \mathbf{T}$?

37 If $AS = SA$ for the *shift matrix* S, show that A must have this special form:

$$\text{If } \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \text{ then } A = \begin{bmatrix} a & b & c \\ 0 & a & b \\ 0 & 0 & a \end{bmatrix}.$$

"The subspace of matrices that commute with the shift S has dimension _____."

38 Which of the following are bases for $\mathbf{R}^3$?

(a) $(1, 2, 0)$ and $(0, 1, -1)$

(b) $(1, 1, -1), (2, 3, 4), (4, 1, -1), (0, 1, -1)$

(c) $(1, 2, 2), (-1, 2, 1), (0, 8, 0)$

(d) $(1, 2, 2), (-1, 2, 1), (0, 8, 6)$

39 Suppose A is 5 by 4 with rank 4. Show that $A\boldsymbol{x} = \boldsymbol{b}$ has no solution when the 5 by 5 matrix $[\,A \;\; \boldsymbol{b}\,]$ is invertible. Show that $A\boldsymbol{x} = \boldsymbol{b}$ is solvable when $[\,A \;\; \boldsymbol{b}\,]$ is singular.

40 (a) Find a basis for all solutions to $d^4y/dx^4 = y(x)$.
(b) Find a particular solution to $d^4y/dx^4 = y(x) + 1$. Find the complete solution.

Challenge Problems

41 Write the 3 by 3 identity matrix as a combination of the other five permutation matrices! Then show that those five matrices are linearly independent. (Assume a combination gives $c_1 P_1 + \cdots + c_5 P_5 =$ zero matrix, and check entries to prove c_i is zero.) The five permutations are a basis for the subspace of 3 by 3 matrices with row and column sums all equal.

42 Choose $\boldsymbol{x} = (x_1, x_2, x_3, x_4)$ in $\mathbf{R}^4$. It has 24 rearrangements like (x_2, x_1, x_3, x_4) and (x_4, x_3, x_1, x_2). Those 24 vectors, including $\boldsymbol{x}$ itself, span a subspace $\mathbf{S}$. Find specific vectors $\boldsymbol{x}$ so that the dimension of $\mathbf{S}$ is: (a) zero, (b) one, (c) three, (d) four.

43 Intersections and sums have $\dim(\mathbf{V}) + \dim(\mathbf{W}) = \dim(\mathbf{V} \cap \mathbf{W}) + \dim(\mathbf{V} + \mathbf{W})$. Start with a basis $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ for the intersection $\mathbf{V} \cap \mathbf{W}$. Extend with $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_s$ to a basis for $\mathbf{V}$, and separately with $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_t$ to a basis for $\mathbf{W}$. Prove that the $\boldsymbol{u}$'s, $\boldsymbol{v}$'s and $\boldsymbol{w}$'s together are ***independent***. The dimensions have $(r + s) + (r + t) = (r) + (r + s + t)$ as desired.

44 Mike Artin suggested a neat higher-level proof of that dimension formula in Problem 43. From all inputs $\boldsymbol{v}$ in $\mathbf{V}$ and $\boldsymbol{w}$ in $\mathbf{W}$, the "sum transformation" produces $\boldsymbol{v}+\boldsymbol{w}$. Those outputs fill the space $\mathbf{V}+\mathbf{W}$. The nullspace contains all pairs $\boldsymbol{v} = \boldsymbol{u}$, $\boldsymbol{w} = -\boldsymbol{u}$ for vectors $\boldsymbol{u}$ in $\mathbf{V} \cap \mathbf{W}$. (Then $\boldsymbol{v} + \boldsymbol{w} = \boldsymbol{u} - \boldsymbol{u} = \mathbf{0}$.) So $\dim(\mathbf{V}+\mathbf{W}) + \dim(\mathbf{V} \cap \mathbf{W})$ equals $\dim(\mathbf{V}) + \dim(\mathbf{W})$ (*input dimension from* $\mathbf{V}$ *and* $\mathbf{W}$) by the crucial formula

dimension of outputs + dimension of nullspace = dimension of inputs.

Problem For an m by n matrix of rank r, what are those 3 dimensions? Outputs = column space. This question will be answered in Section 3.6, can you do it now?

45 Inside $\mathbf{R}^n$, suppose dimension $(\mathbf{V})$ + dimension $(\mathbf{W}) > n$. Show that some nonzero vector is in both $\mathbf{V}$ and $\mathbf{W}$.

46 Suppose A is 10 by 10 and $A^2 = 0$ (zero matrix). This means that the column space of A is contained in the _____. If A has rank r, those subspaces have dimension $r \leq 10 - r$. So the rank is $r \leq 5$.

(This problem was added to the second printing: If $A^2 = 0$ it says that $r \leq n/2$.)

3.6 Dimensions of the Four Subspaces

The main theorem in this chapter connects ***rank*** and ***dimension***. The ***rank*** of a matrix is the number of pivots. The ***dimension*** of a subspace is the number of vectors in a basis. We count pivots or we count basis vectors. *The rank of A reveals the dimensions of all four fundamental subspaces.* Here are the subspaces, including the new one.

Two subspaces come directly from A, and the other two from A^{T}:

Four Fundamental Subspaces

1. The ***row space*** is $C(A^{\mathrm{T}})$, a subspace of $\mathbf{R}^n$.
2. The ***column space*** is $C(A)$, a subspace of $\mathbf{R}^m$.
3. The ***nullspace*** is $N(A)$, a subspace of $\mathbf{R}^n$.
4. The ***left nullspace*** is $N(A^{\mathrm{T}})$, a subspace of $\mathbf{R}^m$. This is our new space.

In this book the column space and nullspace came first. We know $C(A)$ and $N(A)$ pretty well. Now the other two subspaces come forward. The row space contains all combinations of the rows. *This is the column space of* A^{T}.

For the left nullspace we solve $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$—that system is n by m. *This is the nullspace of* A^{T}. The vectors $\boldsymbol{y}$ go on the *left* side of A when the equation is written as $\boldsymbol{y}^{\mathrm{T}}A = \mathbf{0}^{\mathrm{T}}$. The matrices A and A^{T} are usually different. So are their column spaces and their nullspaces. But those spaces are connected in an absolutely beautiful way.

Part 1 of the Fundamental Theorem finds the dimensions of the four subspaces. One fact stands out: ***The row space and column space have the same dimension*** r (the rank of the matrix). The other important fact involves the two nullspaces:

$N(A)$ ***and*** $N(A^{\mathrm{T}})$ ***have dimensions*** $n - r$ ***and*** $m - r$***, to make up the full*** n ***and*** m***.***

Part 2 of the Fundamental Theorem will describe how the four subspaces fit together (two in $\mathbf{R}^n$ and two in $\mathbf{R}^m$). That completes the "right way" to understand every $A\boldsymbol{x} = \boldsymbol{b}$. Stay with it—you are doing real mathematics.

The Four Subspaces for R

Suppose A is reduced to its row echelon form R. For that special form, the four subspaces are easy to identify. We will find a basis for each subspace and check its dimension. Then we watch how the subspaces change (two of them don't change!) as we look back at A. The main point is that *the four dimensions are the same for A and R.*

As a specific 3 by 5 example, look at the four subspaces for the echelon matrix R:

$$\begin{matrix} m=3 \\ n=5 \\ r=2 \end{matrix} \qquad \begin{bmatrix} 1 & 3 & 5 & 0 & 7 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{matrix} \textbf{pivot rows 1 and 2} \\ \\ \textbf{pivot columns 1 and 4} \end{matrix}$$

The rank of this matrix R is $r = 2$ (*two pivots*). Take the four subspaces in order.

1. The ***row space*** of R has dimension 2, matching the rank.

Reason: The first two rows are a basis. The row space contains combinations of all three rows, but the third row (the zero row) adds nothing new. So rows 1 and 2 span the row space $\boldsymbol{C}(R^{\mathrm{T}})$.

The pivot rows 1 and 2 are independent. That is obvious for this example, and it is always true. If we look only at the pivot columns, we see the r by r identity matrix. There is no way to combine its rows to give the zero row (except by the combination with all coefficients zero). So the r pivot rows are a basis for the row space.

The dimension of the row space is the rank r. The nonzero rows of R form a basis.

2. The ***column space*** of R also has dimension $r = 2$.

Reason: The pivot columns 1 and 4 form a basis for $\boldsymbol{C}(R)$. They are independent because they start with the r by r identity matrix. No combination of those pivot columns can give the zero column (except the combination with all coefficients zero). And they also span the column space. Every other (free) column is a combination of the pivot columns. Actually the combinations we need are the three special solutions!

Column 2 is 3 (column 1). The special solution is $(-3, 1, 0, 0, 0)$.

Column 3 is 5 (column 1). The special solution is $(-5, 0, 1, 0, 0,)$.

Column 5 is 7 (column 1) $+\,2$ (column 4). That solution is $(-7, 0, 0, -2, 1)$.

The pivot columns are independent, and they span, so they are a basis for $\boldsymbol{C}(R)$.

The dimension of the column space is the rank r. The pivot columns form a basis.

3. The ***nullspace*** has dimension $n - r = 5 - 2$. There are $n - r = 3$ free variables. Here x_2, x_3, x_5 are free (no pivots in those columns). They yield the three special solutions to $R\boldsymbol{x} = \boldsymbol{0}$. Set a free variable to 1, and solve for x_1 and x_4:

$$s_2 = \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad s_3 = \begin{bmatrix} -5 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad s_5 = \begin{bmatrix} -7 \\ 0 \\ 0 \\ -2 \\ 1 \end{bmatrix} \quad \begin{array}{l} R\boldsymbol{x} = \boldsymbol{0} \text{ has the} \\ \text{complete solution} \\ \boldsymbol{x} = x_2 s_2 + x_3 s_3 + x_5 s_5 \end{array}.$$

There is a special solution for each free variable. With n variables and r pivot variables, that leaves $n - r$ free variables and special solutions. $\boldsymbol{N}(R)$ has dimension $n - r$.

The nullspace has dimension* $n - r$. *The special solutions form a basis.

The special solutions are independent, because they contain the identity matrix in rows 2, 3, 5. All solutions are combinations of special solutions, $x = x_2 s_2 + x_3 s_3 + x_5 s_5$, because this puts x_2, x_3 and x_5 in the correct positions. Then the pivot variables x_1 and x_4 are totally determined by the equations $Rx = \mathbf{0}$.

4. The ***nullspace of*** R^{T} ***(left nullspace of*** R***)*** has dimension $m - r = 3 - 2$.

Reason: The equation $R^{\mathrm{T}}y = \mathbf{0}$ looks for combinations of the columns of R^{T} (*the rows of* R) that produce zero. This equation $R^{\mathrm{T}}y = \mathbf{0}$ or $y^{\mathrm{T}}R = \mathbf{0}^{\mathrm{T}}$ is

$$\textbf{Left nullspace} \qquad \begin{array}{r} y_1\,[1,\ 3,\ 5,\ 0,\ 7] \\ +y_2\,[0,\ 0,\ 0,\ 1,\ 2] \\ +y_3\,[0,\ 0,\ 0,\ 0,\ 0] \\ \hline [0,\ 0,\ 0,\ 0,\ 0] \end{array} \tag{1}$$

The solutions y_1, y_2, y_3 are pretty clear. We need $y_1 = 0$ and $y_2 = 0$. The variable y_3 is free (it can be anything). The nullspace of R^{T} contains all vectors $y = (0, 0, y_3)$. It is the line of all multiples of the basis vector $(0, 0, 1)$.

In all cases R ends with $m - r$ zero rows. Every combination of these $m - r$ rows gives zero. These are the *only* combinations of the rows of R that give zero, because the pivot rows are linearly independent. The left nullspace of R contains all these solutions $y = (0, \cdots, 0, y_{r+1}, \cdots, y_m)$ to $R^{\mathrm{T}}y = 0$.

If A is m by n of rank r, its left nullspace has dimension $m - r$.

To produce a zero combination, y must start with r zeros. This leaves dimension $m - r$.

Why is this a "*left* nullspace"? The reason is that $R^{\mathrm{T}}y = \mathbf{0}$ can be transposed to $y^{\mathrm{T}}R = \mathbf{0}^{\mathrm{T}}$. Now y^{T} is a row vector to the *left* of R. You see the y's in equation (1) multiplying the rows. This subspace came fourth, and some linear algebra books omit it—but that misses the beauty of the whole subject.

In $\mathbf{R}^n$ ***the row space and nullspace have dimensions*** r and $n - r$ (adding to n).
In $\mathbf{R}^m$ ***the column space and left nullspace have dimensions*** r and $m - r$ (total m).

So far this is proved for echelon matrices R. Figure 3.5 shows the same for A.

The Four Subspaces for A

We have a job still to do. ***The subspace dimensions for A are the same as for R.*** The job is to explain why. A is now any matrix that reduces to $R = \text{rref}(A)$.

$$\textbf{A reduces to R} \qquad A = \begin{bmatrix} 1 & 3 & 5 & 0 & 7 \\ 0 & 0 & 0 & 1 & 2 \\ 1 & 3 & 5 & 1 & 9 \end{bmatrix} \qquad \text{Notice } C(A) \neq C(R) \tag{2}$$

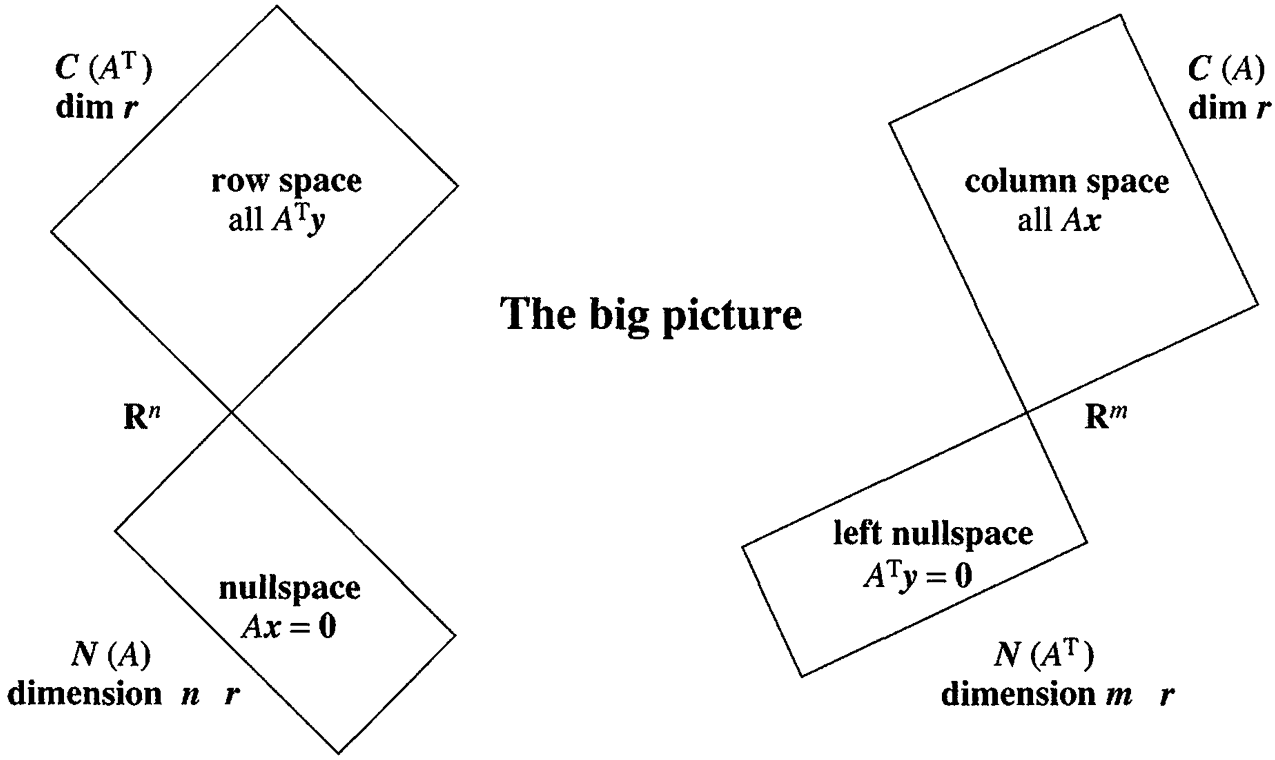


Figure 3.5: The dimensions of the Four Fundamental Subspaces (for R and for A).

An elimination matrix takes A to R. The big picture (Figure 3.5) applies to both. The invertible matrix E is the product of the elementary matrices that reduce A to R:

$$\textbf{A to R and back} \qquad EA = R \quad \text{and} \quad A = E^{-1}R \tag{3}$$

1 ***A has the same row space as R*. Same dimension r and same basis.**

Reason: Every row of A is a combination of the rows of R. Also every row of R is a combination of the rows of A. Elimination changes rows, but not row *spaces*.

Since A has the same row space as R, we can choose the first r rows of R as a basis. Or we could choose r suitable rows of the original A. They might not always be the *first* r rows of A, because those could be dependent. The good r rows of A are the ones that end up as pivot rows in R.

2 ***The column space of A has dimension r*.** For every matrix this is essential:

The number of independent columns equals the number of independent rows.

Wrong reason: "A and R have the same column space." This is false. The columns of R often end in zeros. The columns of A don't often end in zeros. The column spaces are different, but their *dimensions* are the same—equal to r.

Right reason: The ***same*** combinations of the columns are zero (or nonzero) for A and R. Say that another way: $Ax = \mathbf{0}$ *exactly when* $Rx = \mathbf{0}$. The r pivot columns (of both) are independent.

Conclusion The r pivot columns of A are a basis for *its* column space.

3 ***A has the same nullspace as R.*** **Same dimension $n - r$ and same basis.**

Reason: The elimination steps don't change the solutions. The special solutions are a basis for this nullspace (as we always knew). There are $n - r$ free variables, so the dimension of the nullspace is $n - r$. Notice that $r + (n - r)$ equals n:

(dimension of column space) + (dimension of nullspace) = dimension of $\mathbf{R}^n$.

4 ***The left nullspace of A*** (the nullspace of A^{T}) ***has dimension m − r.***

Reason: A^{T} is just as good a matrix as A. When we know the dimensions for every A, we also know them for A^{T}. Its column space was proved to have dimension r. Since A^{T} is n by m, the "whole space" is now $\mathbf{R}^m$. The counting rule for A was $r + (n - r) = n$. The counting rule for A^{T} is $r + (m - r) = m$. We now have all details of the main theorem:

***Fundamental Theorem of Linear Algebra,* Part 1**

The column space and row space both have dimension r.
The nullspaces have dimensions n − r and m − r.

By concentrating on *spaces* of vectors, not on individual numbers or vectors, we get these clean rules. You will soon take them for granted—eventually they begin to look obvious. But if you write down an 11 by 17 matrix with 187 nonzero entries, I don't think most people would see why these facts are true:

Two key facts $\quad$ dimension of $\mathbf{C}(A)$ = dimension of $\mathbf{C}(A^{\mathrm{T}})$ = rank of A
dimension of $\mathbf{C}(A)$ + dimension of $\mathbf{N}(A) = 17$.

Example 1 $A = [\,1 \;\; 2 \;\; 3\,]$ has $m = 1$ and $n = 3$ and rank $r = 1$.

The row space is a line in $\mathbf{R}^3$. The nullspace is the plane $A\boldsymbol{x} = x_1 + 2x_2 + 3x_3 = 0$. This plane has dimension 2 (which is $3 - 1$). The dimensions add to $\mathbf{1 + 2 = 3}$.

The columns of this 1 by 3 matrix are in $\mathbf{R}^1$! The column space is all of $\mathbf{R}^1$. The left nullspace contains only the zero vector. The only solution to $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$ is $\boldsymbol{y} = \mathbf{0}$, no other multiple of $[\,1 \;\; 2 \;\; 3\,]$ gives the zero row. Thus $\mathbf{N}(A^{\mathrm{T}})$ is $\mathbf{Z}$, the zero space with dimension 0 (which is $m - r$). In $\mathbf{R}^m$ the dimensions add to $\mathbf{1 + 0 = 1}$.

Example 2 $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix}$ has $m = 2$ with $n = 3$ and rank $r = 1$.

The row space is the same line through $(1, 2, 3)$. The nullspace must be the same plane $x_1 + 2x_2 + 3x_3 = 0$. Their dimensions still add to $1 + 2 = 3$.

All columns are multiples of the first column $(1, 2)$. Twice the first row minus the second row is the zero row. Therefore $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$ has the solution $\boldsymbol{y} = (2, -1)$. The column space and left nullspace are **perpendicular lines** in $\mathbf{R}^2$. Dimensions $1 + 1 = 2$.

$$\text{Column space} = \text{line through } \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \text{Left nullspace} = \text{line through } \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

If A has three equal rows, its rank is _____. What are two of the $\boldsymbol{y}$'s in its left nullspace?

The y's in the left nullspace combine the rows to give the zero row.

Matrices of Rank One

That last example had rank $r = 1$—and rank one matrices are special. We can describe them all. You will see again that dimension of row space = dimension of column space. When $r = 1$, every row is a multiple of the same row:

$$A = uv^{\mathrm{T}} \quad A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ -3 & -6 & -9 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{equals} \quad \begin{bmatrix} 1 \\ 2 \\ -3 \\ 0 \end{bmatrix} \quad \text{times} \quad \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = v^{\mathrm{T}}.$$

A column times a row (4 by 1 times 1 by 3) produces a matrix (4 by 3). All rows are multiples of the row $(1, 2, 3)$. All columns are multiples of the column $(1, 2, -3, 0)$. The row space is a line in $\mathbf{R}^n$, and the column space is a line in $\mathbf{R}^m$.

Every rank one matrix has the special form $A = uv^{\mathrm{T}}$ = column times row.

The columns are multiples of u. The rows are multiples of v^{T}. *The nullspace is the plane perpendicular to v.* ($Ax = \mathbf{0}$ means that $u(v^{\mathrm{T}}x) = \mathbf{0}$ and then $v^{\mathrm{T}}x = 0$.) It is this **perpendicularity** of the subspaces that will be Part 2 of the Fundamental Theorem.

■ REVIEW OF THE KEY IDEAS ■

1. The r pivot rows of R are a basis for the row spaces of R and A (same space).
2. The r pivot columns of A (!) are a basis for its column space.
3. The $n - r$ special solutions are a basis for the nullspaces of A and R (same space).
4. The last $m - r$ rows of I are a basis for the left nullspace of R.
5. The last $m - r$ rows of E are a basis for the left nullspace of A.

Note about the four subspaces The Fundamental Theorem looks like pure algebra, but it has very important applications. My favorites are the networks in Chapter 8 (often I go there for my next lecture). The equation for y in the left nullspace is $A^{\mathrm{T}}y = 0$:

Flow into a node equals flow out. Kirchhoff's Current Law is the "balance equation".

This is (in my opinion) the most important equation in applied mathematics. All models in science and engineering and economics involve a balance—of force or heat flow or charge or momentum or money. That balance equation, plus Hooke's Law or Ohm's Law or some law connecting "potentials" to "flows", gives a clear framework for applied mathematics.

My textbook on *Computational Science and Engineering* develops that framework, together with algorithms to solve the equations: Finite differences, finite elements, spectral methods, iterative methods, and multigrid.

■ WORKED EXAMPLES ■

3.6 A Find bases and dimensions for all four fundamental subspaces if you know that

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 5 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 0 & 5 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 \end{bmatrix} = LU = E^{-1}R.$$

By changing only *one* number in R, change the dimensions of all four subspaces.

Solution This matrix has pivots in columns 1 and 3. Its rank is $r = 2$.

Row space	Basis $(1, 3, 0, 5)$ and $(0, 0, 1, 6)$ from R. Dimension 2.
Column space	Basis $(1, 2, 5)$ and $(0, 1, 0)$ from E^{-1} (and A). Dimension 2.
Nullspace	Basis $(-3, 1, 0, 0)$ and $(-5, 0, -6, 1)$ from R. Dimension 2.
Nullspace of A^{T}	Basis $(-5, 0, 1)$ from row 3 of E. Dimension $3 - 2 = 1$.

We need to comment on that left nullspace $N(A^{\mathrm{T}})$. $EA = R$ says that the last row of E combines the three rows of A into the zero row of R. So that last row of E is a basis vector for the left nullspace. If R had *two* zero rows, then the last *two* rows of E would be a basis. (Just like elimination, $y^{\mathrm{T}}A = \mathbf{0}^{\mathrm{T}}$ combines rows of A to give zero rows in R.)

To change all these dimensions we need to change the rank r. One way to do that is to change an entry (*any entry*) in the zero row of R.

3.6 B Put four 1's into a 5 by 6 matrix of zeros, keeping the dimension of its *row space* as small as possible. Describe all the ways to make the dimension of its *column space* as small as possible. Describe all the ways to make the dimension of its *nullspace* as small as possible. How to make the *sum of the dimensions of all four subspaces small*?

Solution The rank is 1 if the four 1's go into the same row, or into the same column. They can also go into *two rows and two columns* (so $a_{ii} = a_{ij} = a_{ji} = a_{jj} = 1$). Since the column space and row space always have the same dimensions, this answers the first two questions: Dimension 1.

The nullspace has its smallest possible dimension $6 - 4 = 2$ when the rank is $r = 4$. To achieve rank 4, the 1's must go into four different rows and columns.

You can't do anything about the sum $r + (n - r) + r + (m - r) = n + m$. It will be $6 + 5 = 11$ no matter how the 1's are placed. The sum is 11 even if there aren't any 1's...

If all the other entries of A are 2's instead of 0's, how do these answers change?

Problem Set 3.6

1 (a) If a 7 by 9 matrix has rank 5, what are the dimensions of the four subspaces? What is the sum of all four dimensions?

(b) If a 3 by 4 matrix has rank 3, what are its column space and left nullspace?

2 Find bases and dimensions for the four subspaces associated with A and B:

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 5 & 8 \end{bmatrix}.$$

3 Find a basis for each of the four subspaces associated with A:

$$A = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 4 & 6 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

4 Construct a matrix with the required property or explain why this is impossible:

(a) Column space contains $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, row space contains $\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix}$.

(b) Column space has basis $\begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$, nullspace has basis $\begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$.

(c) Dimension of nullspace $= 1 +$ dimension of left nullspace.

(d) Left nullspace contains $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$, row space contains $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$.

(e) Row space $=$ column space, nullspace $\neq$ left nullspace.

5 If $\mathbf{V}$ is the subspace spanned by $(1,1,1)$ and $(2,1,0)$, find a matrix A that has $\mathbf{V}$ as its row space. Find a matrix B that has $\mathbf{V}$ as its nullspace.

6 Without elimination, find dimensions and bases for the four subspaces for

$$A = \begin{bmatrix} 0 & 3 & 3 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}.$$

7 Suppose the 3 by 3 matrix A is invertible. Write down bases for the four subspaces for A, and also for the 3 by 6 matrix $B = [\,A \quad A\,]$.

8 What are the dimensions of the four subspaces for A, B, and C, if I is the 3 by 3 identity matrix and 0 is the 3 by 2 zero matrix?

$$A = \begin{bmatrix} I & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} I & I \\ 0^{\mathrm{T}} & 0^{\mathrm{T}} \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 0 \end{bmatrix}.$$

9 Which subspaces are the same for these matrices of different sizes?

(a) $[\,A\,]$ and $\begin{bmatrix} A \\ A \end{bmatrix}$ (b) $\begin{bmatrix} A \\ A \end{bmatrix}$ and $\begin{bmatrix} A & A \\ A & A \end{bmatrix}$.

Prove that all three of those matrices have the *same rank r*.

10 If the entries of a 3 by 3 matrix are chosen randomly between 0 and 1, what are the most likely dimensions of the four subspaces? What if the matrix is 3 by 5?

11 (Important) A is an m by n matrix of rank r. Suppose there are right sides $\boldsymbol{b}$ for which $A\boldsymbol{x} = \boldsymbol{b}$ has *no solution.*

(a) What are all inequalities ($<$ or $\leq$) that must be true between m, n, and r?

(b) How do you know that $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$ has solutions other than $\boldsymbol{y} = \mathbf{0}$?

12 Construct a matrix with $(1,0,1)$ and $(1,2,0)$ as a basis for its row space and its column space. Why can't this be a basis for the row space and nullspace?

13 True or false (with a reason or a counterexample):

(a) If $m = n$ then the row space of A equals the column space.

(b) The matrices A and $-A$ share the same four subspaces.

(c) If A and B share the same four subspaces then A is a multiple of B.

14 Without computing A, find bases for its four fundamental subspaces:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 6 & 1 & 0 \\ 9 & 8 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \end{bmatrix}.$$

15 If you exchange the first two rows of A, which of the four subspaces stay the same? If $\boldsymbol{v} = (1,2,3,4)$ is in the left nullspace of A, write down a vector in the left nullspace of the new matrix.

16 Explain why $\boldsymbol{v} = (1,0,-1)$ *cannot be a row of A and also in the nullspace.*

17 Describe the four subspaces of $\mathbf{R}^3$ associated with

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad I + A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

18 (Left nullspace) Add the extra column $\boldsymbol{b}$ and reduce A to echelon form:

$$\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & b_1 \\ 4 & 5 & 6 & b_2 \\ 7 & 8 & 9 & b_3 \end{bmatrix} \quad \rightarrow \quad \begin{bmatrix} 1 & 2 & 3 & b_1 \\ 0 & -3 & -6 & b_2 - 4b_1 \\ 0 & 0 & 0 & b_3 - 2b_2 + b_1 \end{bmatrix}.$$

A combination of the rows of A has produced the zero row. What combination is it? (Look at $b_3 - 2b_2 + b_1$ on the right side.) Which vectors are in the nullspace of A^{T} and which are in the nullspace of A?

19 Following the method of Problem 18, reduce A to echelon form and look at zero rows. The $\boldsymbol{b}$ column tells which combinations you have taken of the rows:

$$\text{(a)} \begin{bmatrix} 1 & 2 & b_1 \\ 3 & 4 & b_2 \\ 4 & 6 & b_3 \end{bmatrix} \qquad \text{(b)} \begin{bmatrix} 1 & 2 & b_1 \\ 2 & 3 & b_2 \\ 2 & 4 & b_3 \\ 2 & 5 & b_4 \end{bmatrix}$$

From the $\boldsymbol{b}$ column after elimination, read off $m-r$ basis vectors in the left nullspace. Those y's are combinations of rows that give zero rows.

20 (a) Check that the solutions to $A\boldsymbol{x} = \mathbf{0}$ are perpendicular to the rows:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4 & 1 \end{bmatrix} \begin{bmatrix} 4 & 2 & 0 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} = ER.$$

(b) How many independent solutions to $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$? Why is $\boldsymbol{y}^{\mathrm{T}}$ the last row of E^{-1}?

21 Suppose A is the sum of two matrices of rank one: $A = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} + \boldsymbol{w}\boldsymbol{z}^{\mathrm{T}}$.

(a) Which vectors span the column space of A?

(b) Which vectors span the row space of A?

(c) The rank is less than 2 if _____ or if _____.

(d) Compute A and its rank if $\boldsymbol{u} = \boldsymbol{z} = (1,0,0)$ and $\boldsymbol{v} = \boldsymbol{w} = (0,0,1)$.

22 Construct $A = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} + \boldsymbol{w}\boldsymbol{z}^{\mathrm{T}}$ whose column space has basis $(1,2,4), (2,2,1)$ and whose row space has basis $(1,0), (1,1)$. Write A as (3 by 2) times (2 by 2).

23 Without multiplying matrices, find bases for the row and column spaces of A:

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 2 & 7 \end{bmatrix} \begin{bmatrix} 3 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix}.$$

How do you know from these shapes that A cannot be invertible?

24 (Important) $A^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{d}$ is solvable when $\boldsymbol{d}$ is in which of the four subspaces? The solution $\boldsymbol{y}$ is unique when the _____ contains only the zero vector.

25 True or false (with a reason or a counterexample):

(a) A and A^{T} have the same number of pivots.

(b) A and A^{T} have the same left nullspace.

(c) If the row space equals the column space then $A^{\mathrm{T}} = A$.

(d) If $A^{\mathrm{T}} = -A$ then the row space of A equals the column space.

26 (***Rank of*** AB) If $AB = C$, the rows of C are combinations of the rows of ____. So the rank of C is not greater than the rank of ____. Since $B^{\mathrm{T}}A^{\mathrm{T}} = C^{\mathrm{T}}$, the rank of C is also not greater than the rank of ____.

27 If a, b, c are given with $a \neq 0$, how would you choose d so that $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ has rank 1? Find a basis for the row space and nullspace. Show they are perpendicular!

28 Find the ranks of the 8 by 8 checkerboard matrix B and the chess matrix C:

$$B = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} r & n & b & q & k & b & n & r \\ p & p & p & p & p & p & p & p \\ & & & \multicolumn{2}{c}{\text{four zero rows}} & & & \\ p & p & p & p & p & p & p & p \\ r & n & b & q & k & b & n & r \end{bmatrix}$$

The numbers r, n, b, q, k, p are all different. Find bases for the row space and left nullspace of B and C. Challenge problem: Find a basis for the nullspace of C.

29 Can tic-tac-toe be completed (5 ones and 4 zeros in A) so that rank $(A) = 2$ but neither side passed up a winning move?

Challenge Problems

30 If $A = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ is a 2 by 2 matrix of rank 1, redraw Figure 3.5 to show clearly the Four Fundamental Subspaces. If B produces those same four subspaces, what is the exact relation of B to A?

31 $\mathbf{M}$ is the space of 3 by 3 matrices. Multiply every matrix X in $\mathbf{M}$ by

$$A = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}. \quad \text{Notice: } A \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

(a) Which matrices X lead to $AX =$ zero matrix?

(b) Which matrices have the form AX for some matrix X?

(a) finds the "nullspace" of that operation AX and (b) finds the "column space". What are the dimensions of those two subspaces of $\mathbf{M}$? Why do the dimensions add to $(n - r) + r = 9$?

32 Suppose the m by n matrices A and B have *the same four subspaces*. If they are both in row reduced echelon form, prove that F must equal G:

$$A = \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} I & G \\ 0 & 0 \end{bmatrix}.$$

Chapter 4

Orthogonality

4.1 Orthogonality of the Four Subspaces

Two vectors are orthogonal when their dot product is zero: $v \cdot w = 0$ or $v^{\mathrm{T}}w = 0$. This chapter moves to **orthogonal subspaces** and **orthogonal bases** and **orthogonal matrices**. The vectors in two subspaces, and the vectors in a basis, and the vectors in the columns, all pairs will be orthogonal. Think of $a^2 + b^2 = c^2$ for a *right triangle* with sides v and w.

Orthogonal vectors $\quad v^{\mathrm{T}}w = 0 \quad$ and $\quad \|v\|^2 + \|w\|^2 = \|v + w\|^2.$

The right side is $(v + w)^{\mathrm{T}}(v + w)$. This equals $v^{\mathrm{T}}v + w^{\mathrm{T}}w$ when $v^{\mathrm{T}}w = w^{\mathrm{T}}v = 0$.

Subspaces entered Chapter 3 to throw light on $Ax = b$. Right away we needed the column space (for b) and the nullspace (for x). Then the light turned onto A^{T}, uncovering two more subspaces. Those four fundamental subspaces reveal what a matrix really does.

A matrix multiplies a vector: A *times* x. At the first level this is only numbers. At the second level Ax is a combination of column vectors. The third level shows subspaces. But I don't think you have seen the whole picture until you study Figure 4.2. It fits the subspaces together, to show the hidden reality of A times x. The 90° angles between subspaces are new—and we have to say what those right angles mean.

The row space is perpendicular to the nullspace. Every row of A is perpendicular to every solution of $Ax = \mathbf{0}$. That gives the 90° angle on the left side of the figure. This perpendicularity of subspaces is Part 2 of the Fundamental Theorem of Linear Algebra.

The column space is perpendicular to the nullspace of A^{T}. When b is outside the column space—when we want to solve $Ax = b$ and can't do it—then this nullspace of A^{T} comes into its own. It contains the error $e = b - Ax$ in the "least-squares" solution. Least squares is the key application of linear algebra in this chapter.

Part 1 of the Fundamental Theorem gave the dimensions of the subspaces. The row and column spaces have the same dimension r (they are drawn the same size). The two nullspaces have the remaining dimensions $n - r$ and $m - r$. Now we will show that ***the row space and nullspace are orthogonal subspaces inside*** $\mathbf{R}^n$.

DEFINITION Two subspaces V and W of a vector space are ***orthogonal*** if every vector v in V is perpendicular to every vector w in W:

Orthogonal subspaces $\quad v^{\mathrm{T}} w = 0$ ***for all*** v ***in*** V ***and all*** w ***in*** W.

Example 1 The floor of your room (extended to infinity) is a subspace V. The line where two walls meet is a subspace W (one-dimensional). Those subspaces are orthogonal. Every vector up the meeting line is perpendicular to every vector in the floor.

Example 2 Two walls look perpendicular but they are not orthogonal subspaces! The meeting line is in both V and W—and this line is not perpendicular to itself. Two planes (dimensions 2 and 2 in $\mathbf{R}^3$) cannot be orthogonal subspaces.

When a vector is in two orthogonal subspaces, it *must* be zero. It is perpendicular to itself. It is v and it is w, so $v^{\mathrm{T}} v = 0$. This has to be the zero vector.

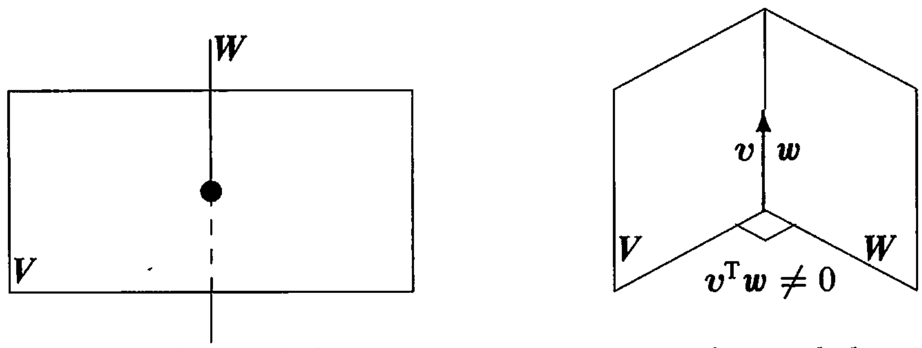


Figure 4.1: Orthogonality is impossible when dim V +dim W > dimension of whole space.

The crucial examples for linear algebra come from the fundamental subspaces. Zero is the only point where the nullspace meets the row space. More than that, the nullspace and row space of A meet at 90°. This key fact comes directly from $Ax = \mathbf{0}$:

Every vector x in the nullspace is perpendicular to every row of A, because $Ax = \mathbf{0}$.
The nullspace $N(A)$ ***and the row space*** $C(A^{\mathrm{T}})$ ***are orthogonal subspaces of*** $\mathbf{R}^n$.

To see why x is perpendicular to the rows, look at $Ax = \mathbf{0}$. Each row multiplies x:

$$Ax = \begin{bmatrix} \text{row } 1 \\ \vdots \\ \text{row } m \end{bmatrix} \begin{bmatrix} \\ x \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \longleftarrow \ (\text{row } 1) \cdot x \text{ is zero} \\ \\ \longleftarrow \ (\text{row } m) \cdot x \text{ is zero} \end{matrix} \qquad (1)$$

The first equation says that row 1 is perpendicular to x. The last equation says that row m is perpendicular to x. *Every row has a zero dot product with* x. Then x is also perpendicular to every *combination* of the rows. The whole row space $C(A^{\mathrm{T}})$ is orthogonal to $N(A)$.

Here is a second proof of that orthogonality for readers who like matrix shorthand. The vectors in the row space are combinations $A^{\mathrm{T}}\boldsymbol{y}$ of the rows. Take the dot product of $A^{\mathrm{T}}\boldsymbol{y}$ with any $\boldsymbol{x}$ in the nullspace. *These vectors are perpendicular*:

Nullspace and Row space $$\boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{y}) = (A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{0}^{\mathrm{T}}\boldsymbol{y} = 0. \qquad (2)$$

We like the first proof. You can see those rows of A multiplying $\boldsymbol{x}$ to produce zeros in equation (1). The second proof shows why A and A^{T} are both in the Fundamental Theorem. A^{T} goes with $\boldsymbol{y}$ and A goes with $\boldsymbol{x}$. At the end we used $A\boldsymbol{x} = \boldsymbol{0}$.

Example 3 The rows of A are perpendicular to $\boldsymbol{x} = (1, 1, -1)$ in the nullspace:

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 3 & 4 \\ 5 & 2 & 7 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{gives the dot products} \quad \begin{matrix} 1+3-4=0 \\ 5+2-7=0 \end{matrix}$$

Now we turn to the other two subspaces. In this example, the column space is all of $\mathbf{R}^2$. The nullspace of A^{T} is only the zero vector (orthogonal to every vector). The columns of A and nullspace of A^{T} are always orthogonal subspaces.

Every vector $\boldsymbol{y}$ in the nullspace of A^{T} is perpendicular to every column of A. ***The left nullspace $N(A^{\mathrm{T}})$ and the column space $C(A)$ are orthogonal in $\mathbf{R}^m$.***

Apply the original proof to A^{T}. Its nullspace is orthogonal to its row space—and the row space of A^{T} is the column space of A. Q.E.D.

For a visual proof, look at $A^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{0}$. Each column of A multiplies $\boldsymbol{y}$ to give 0:

$$C(A) \perp N(A^{\mathrm{T}}) \qquad A^{\mathrm{T}}\boldsymbol{y} = \begin{bmatrix} (\textbf{column } 1)^{\mathrm{T}} \\ \cdots \\ (\textbf{column } n)^{\mathrm{T}} \end{bmatrix}\begin{bmatrix} \\ \boldsymbol{y} \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ \cdot \\ 0 \end{bmatrix}. \qquad (3)$$

The dot product of $\boldsymbol{y}$ with every column of A is zero. Then $\boldsymbol{y}$ in the left nullspace is perpendicular to each column—and to the whole column space.

Orthogonal Complements

Important The fundamental subspaces are more than just orthogonal (in pairs). Their dimensions are also right. Two lines could be perpendicular in $\mathbf{R}^3$, but those lines *could not be* the row space and nullspace of a 3 by 3 matrix. The lines have dimensions 1 and 1, adding to 2. The correct dimensions r and $n-r$ must add to $n = 3$.

The fundamental subspaces have dimensions 2 and 1, or 3 and 0. Those subspaces are not only orthogonal, they are *orthogonal complements.*

DEFINITION The ***orthogonal complement*** of a subspace V contains ***every*** vector that is perpendicular to V. This orthogonal subspace is denoted by $V^{\perp}$ (pronounced "V perp").

By this definition, the nullspace is the orthogonal complement of the row space. ***Every*** $\boldsymbol{x}$ that is perpendicular to the rows satisfies $A\boldsymbol{x} = \boldsymbol{0}$.

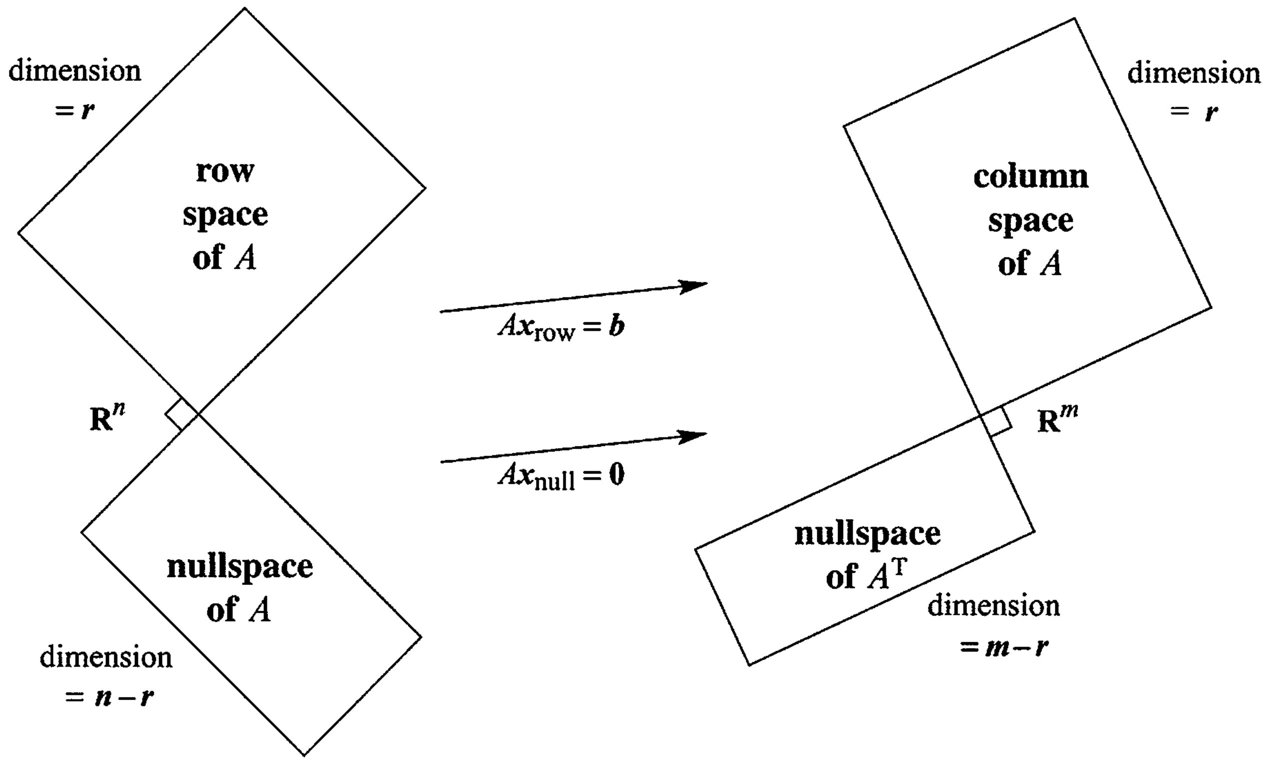


Figure 4.2: Two pairs of orthogonal subspaces. The dimensions add to n and add to m. **This is an important picture**—one pair of subspaces is in $\mathbf{R}^n$ and one pair is in $\mathbf{R}^m$.

The reverse is also true. ***If v is orthogonal to the nullspace, it must be in the row space.*** Otherwise we could add this v as an extra row of the matrix, without changing its nullspace. The row space would grow, which breaks the law $r + (n - r) = n$. We conclude that the nullspace complement $N(A)^{\perp}$ is exactly the row space $C(A^{\mathrm{T}})$.

The left nullspace and column space are orthogonal in $\mathbf{R}^m$, and they are orthogonal complements. Their dimensions r and $m - r$ add to the full dimension m.

***Fundamental Theorem of Linear Algebra*, Part 2**

***$N(A)$ is the orthogonal complement of the row space $C(A^{\mathrm{T}})$* (in $\mathbf{R}^n$).**

***$N(A^{\mathrm{T}})$ is the orthogonal complement of the column space $C(A)$* (in $\mathbf{R}^m$).**

Part 1 gave the dimensions of the subspaces. Part 2 gives the 90° angles between them. The point of "complements" is that every x can be split into a *row space component* x_r and a *nullspace component* x_n. When A multiplies $x = x_r + x_n$, Figure 4.3 shows what happens:

The nullspace component goes to zero: $Ax_n = \mathbf{0}$.

The row space component goes to the column space: $Ax_r = Ax$.

Every vector goes to the column space! Multiplying by A cannot do anything else.

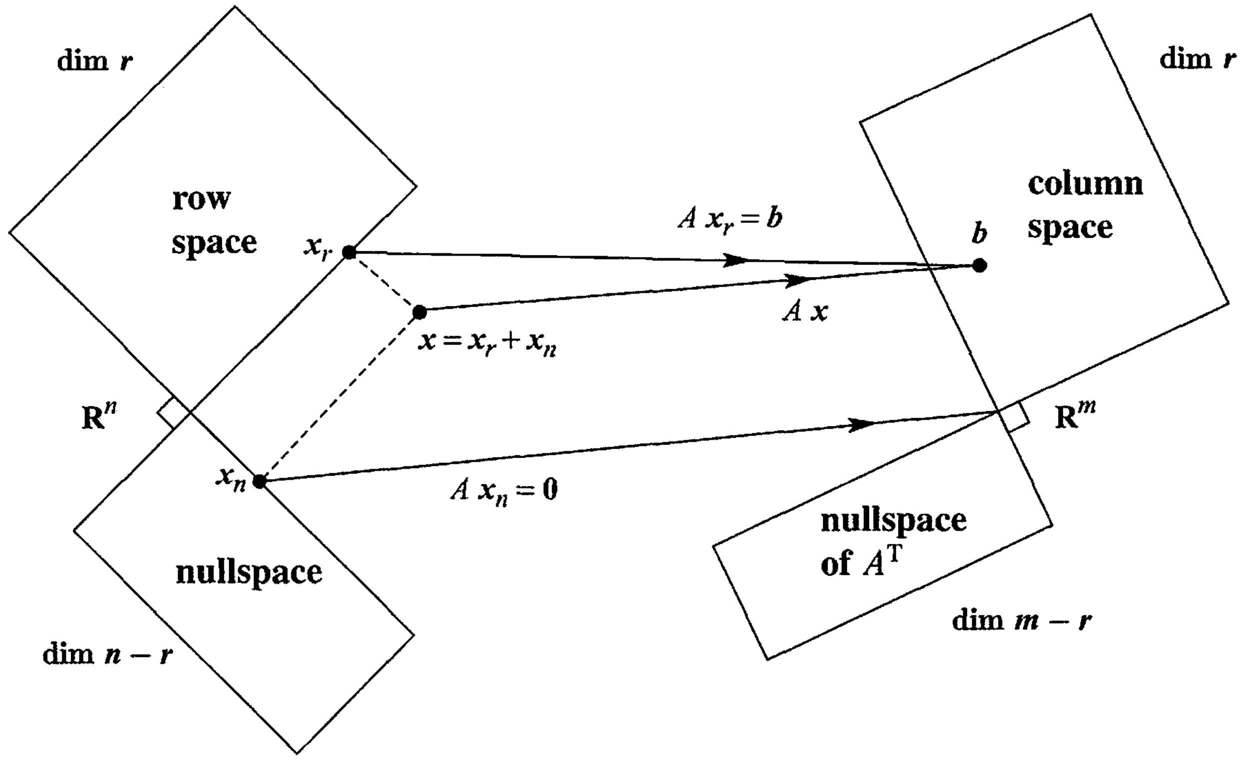


Figure 4.3: This update of Figure 4.2 shows the true action of A on $\boldsymbol{x} = \boldsymbol{x}_r + \boldsymbol{x}_n$. Row space vector $\boldsymbol{x}_r$ to column space, nullspace vector $\boldsymbol{x}_n$ to zero.

More than that: *Every vector $\boldsymbol{b}$ in the column space comes from one and only one vector in the row space.* Proof: If $A\boldsymbol{x}_r = A\boldsymbol{x}'_r$, the difference $\boldsymbol{x}_r - \boldsymbol{x}'_r$ is in the nullspace. It is also in the row space, where $\boldsymbol{x}_r$ and $\boldsymbol{x}'_r$ came from. This difference must be the zero vector, because the nullspace and row space are perpendicular. Therefore $\boldsymbol{x}_r = \boldsymbol{x}'_r$.

There is an r by r invertible matrix hiding inside A, if we throw away the two nullspaces. ***From the row space to the column space, A is invertible.*** The "pseudoinverse" will invert it in Section 7.3.

Example 4 Every diagonal matrix has an r by r invertible submatrix:

$$A = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{contains the submatrix} \quad \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}.$$

The other eleven zeros are responsible for the nullspaces. The rank of B is also $r = 2$:

$$B = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 4 & 5 & 6 \\ 1 & 2 & 4 & 5 & 6 \end{bmatrix} \quad \text{contains} \quad \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix} \quad \text{in the pivot rows and columns.}$$

Every A becomes a diagonal matrix, when we choose the right bases for $\mathbf{R}^n$ and $\mathbf{R}^m$. This ***Singular Value Decomposition*** has become extremely important in applications.

Combining Bases from Subspaces

What follows are some valuable facts about bases. They were saved until now—when we are ready to use them. After a week you have a clearer sense of what a basis is (*linearly independent* vectors that *span* the space). Normally we have to check both of these properties. When the count is right, one property implies the other:

> Any n independent vectors in $\mathbf{R}^n$ must span $\mathbf{R}^n$. So they are a basis.
> Any n vectors that span $\mathbf{R}^n$ must be independent. So they are a basis.

Starting with the correct number of vectors, one property of a basis produces the other. This is true in any vector space, but we care most about $\mathbf{R}^n$. When the vectors go into the columns of an n by n *square* matrix A, here are the same two facts:

> If the n columns of A are independent, they span $\mathbf{R}^n$. So $A\boldsymbol{x} = \boldsymbol{b}$ is solvable.
> If the n columns span $\mathbf{R}^n$, they are independent. So $A\boldsymbol{x} = \boldsymbol{b}$ has only one solution.

Uniqueness implies existence and existence implies uniqueness. ***Then A is invertible.*** If there are no free variables, the solution $\boldsymbol{x}$ is unique. There must be n pivots. Then back substitution solves $A\boldsymbol{x} = \boldsymbol{b}$ (the solution exists).

Starting in the opposite direction, suppose $A\boldsymbol{x} = \boldsymbol{b}$ can be solved for every $\boldsymbol{b}$ (*existence of solutions*). Then elimination produced no zero rows. There are n pivots and no free variables. The nullspace contains only $\boldsymbol{x} = \mathbf{0}$ (*uniqueness* of solutions).

With bases for the row space and the nullspace, we have $r + (n - r) = n$ vectors, This is the right number. Those n vectors are independent.[2] *Therefore they span* $\mathbf{R}^n$.

Each $\boldsymbol{x}$ is the sum $\boldsymbol{x}_r + \boldsymbol{x}_n$ of a row space vector $\boldsymbol{x}_r$ and a nullspace vector $\boldsymbol{x}_n$.

The splitting in Figure 4.3 shows the key point of orthogonal complements—the dimensions add to n and all vectors are fully accounted for.

Example 5 For $A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$ split $\boldsymbol{x} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$ into $\boldsymbol{x}_r + \boldsymbol{x}_n = \begin{bmatrix} 2 \\ 4 \end{bmatrix} + \begin{bmatrix} 2 \\ -1 \end{bmatrix}$.

The vector $(2, 4)$ is in the row space. The orthogonal vector $(2, -1)$ is in the nullspace. The next section will compute this splitting for any A and $\boldsymbol{x}$, by a projection.

---

[2] If a combination of all n vectors gives $\boldsymbol{x}_r + \boldsymbol{x}_n = \mathbf{0}$, then $\boldsymbol{x}_r = -\boldsymbol{x}_n$ is in both subspaces. So $\boldsymbol{x}_r = \boldsymbol{x}_n = 0$. All coefficients of the row space basis and nullspace basis must be zero—which proves independence of the n vectors together.

■ REVIEW OF THE KEY IDEAS ■

1. Subspaces $\boldsymbol{V}$ and $\boldsymbol{W}$ are orthogonal if every $\boldsymbol{v}$ in $\boldsymbol{V}$ is orthogonal to every $\boldsymbol{w}$ in $\boldsymbol{W}$.

2. $\boldsymbol{V}$ and $\boldsymbol{W}$ are "orthogonal complements" if $\boldsymbol{W}$ contains **all** vectors perpendicular to $\boldsymbol{V}$ (and vice versa). Inside $\mathbf{R}^n$, the dimensions of complements $\boldsymbol{V}$ and $\boldsymbol{W}$ add to n.

3. The nullspace $\boldsymbol{N}(A)$ and the row space $\boldsymbol{C}(A^{\mathrm{T}})$ are orthogonal complements, from $A\boldsymbol{x} = \mathbf{0}$. Similarly $\boldsymbol{N}(A^{\mathrm{T}})$ and $\boldsymbol{C}(A)$ are orthogonal complements.

4. Any n independent vectors in $\mathbf{R}^n$ will span $\mathbf{R}^n$.

5. Every $\boldsymbol{x}$ in $\mathbf{R}^n$ has a nullspace component $\boldsymbol{x}_n$ and a row space component $\boldsymbol{x}_r$.

■ WORKED EXAMPLES ■

4.1 A Suppose $\boldsymbol{S}$ is a six-dimensional subspace of nine-dimensional space $\mathbf{R}^9$.

(a) What are the possible dimensions of subspaces orthogonal to $\boldsymbol{S}$?

(b) What are the possible dimensions of the orthogonal complement $\boldsymbol{S}^{\perp}$ of $\boldsymbol{S}$?

(c) What is the smallest possible size of a matrix A that has row space $\boldsymbol{S}$?

(d) What is the shape of its nullspace matrix N?

Solution

(a) If $\boldsymbol{S}$ is six-dimensional in $\mathbf{R}^9$, subspaces orthogonal to $\boldsymbol{S}$ can have dimensions 0, 1, 2, 3.

(b) The complement $\boldsymbol{S}^{\perp}$ is the largest orthogonal subspace, with dimension 3.

(c) The smallest matrix A is 6 by 9 (its six rows are a basis for $\boldsymbol{S}$).

(d) Its nullspace matrix N is 9 by 3. The columns of N contain a basis for $\boldsymbol{S}^{\perp}$.

If a new row 7 of B is a combination of the six rows of A, then B has the same row space as A. It also has the same nullspace matrix N. The special solutions $\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3$ will be the same. Elimination will change row 7 of B to all zeros.

4.1 B The equation $x - 3y - 4z = 0$ describes a plane $\boldsymbol{P}$ in $\mathbf{R}^3$ (actually a subspace).

(a) The plane $\boldsymbol{P}$ is the nullspace $\boldsymbol{N}(A)$ of what 1 by 3 matrix A?

(b) Find a basis $\boldsymbol{s}_1, \boldsymbol{s}_2$ of special solutions of $x - 3y - 4z = 0$ (these would be the columns of the nullspace matrix N).

(c) Also find a basis for the line $P^\perp$ that is perpendicular to P.

(d) Split $v = (6,4,5)$ into its nullspace component v_n in P and its row space component v_r in $P^\perp$.

Solution

(a) The equation $x - 3y - 4z = 0$ is $Ax = \mathbf{0}$ for the 1 by 3 matrix $A = [1 \; -3 \; -4]$.

(b) Columns 2 and 3 are free (the only pivot is 1). The special solutions with free variables 1 and 0 are $s_1 = (3,1,0)$ and $s_2 = (4,0,1)$ in the plane $P = N(A)$.

(c) The row space of A is the line $P^\perp$ in the direction of the row $z = (1,-3,-4)$.

(d) To split v into $v_n + v_r = (c_1 s_1 + c_2 s_2) + c_3 z$, solve for $c_1 = 1, c_2 = 1, c_3 = -1$.

$$\begin{bmatrix} 6 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 3 & 4 & 1 \\ 1 & 0 & -3 \\ 0 & 1 & -4 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \qquad \begin{array}{llll} v_n & = s_1 + s_2 = (7,1,1) & \text{is in} & P = N(A) \\ v_r & = -s_3 = (-1,3,4) & \text{is in} & P^\perp = C(A^{\mathrm{T}}). \\ v & = (6,4,5) & \text{equals} & (7,1,1) + (-1,3,4) \end{array}$$

This method used a basis for each subspace combined into an overall basis s_1, s_2, z. Section 4.2 will also project v onto a subspace S. There we will not need a basis for the perpendicular subspace $S^\perp$.

Problem Set 4.1

Questions 1–12 grow out of Figures 4.2 and 4.3 with four subspaces.

1 Construct any 2 by 3 matrix of rank one. Copy Figure 4.2 and put one vector in each subspace (two in the nullspace). Which vectors are orthogonal?

2 Redraw Figure 4.3 for a 3 by 2 matrix of rank $r = 2$. Which subspace is Z (zero vector only)? The nullspace part of any vector x in $\mathbf{R}^2$ is $x_n =$ _____.

3 Construct a matrix with the required property or say why that is impossible:

(a) Column space contains $\begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix}$, nullspace contains $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

(b) Row space contains $\begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix}$, nullspace contains $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

(c) $Ax = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ has a solution and $A^{\mathrm{T}} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

(d) Every row is orthogonal to every column (A is not the zero matrix)

(e) Columns add up to a column of zeros, rows add to a row of 1's.

4 If $AB = 0$ then the columns of B are in the _____ of A. The rows of A are in the _____ of B. Why can't A and B be 3 by 3 matrices of rank 2?

5 (a) If $Ax = b$ has a solution and $A^{\mathrm{T}}y = 0$, *is* $(y^{\mathrm{T}}x = 0)$ *or* $(y^{\mathrm{T}}b = 0)$?

(b) If $A^{\mathrm{T}}y = (1, 1, 1)$ has a solution and $Ax = 0$, then _____.

6 This system of equations $Ax = b$ has *no solution* (they lead to $0 = 1$):

$$\begin{aligned} x + 2y + 2z &= 5 \\ 2x + 2y + 3z &= 5 \\ 3x + 4y + 5z &= 9 \end{aligned}$$

Find numbers y_1, y_2, y_3 to multiply the equations so they add to $0 = 1$. You have found a vector y in which subspace? Its dot product $y^{\mathrm{T}}b$ is 1, so no solution x.

7 Every system with no solution is like the one in Problem 6. There are numbers $y_1, \ldots, y_m$ that multiply the m equations so they add up to $0 = 1$. This is called **Fredholm's Alternative**:

Exactly one of these problems has a solution

$$Ax = b \quad \textbf{OR} \quad A^{\mathrm{T}}y = 0 \quad \text{with} \quad y^{\mathrm{T}}b = 1.$$

If b is not in the column space of A, it is not orthogonal to the nullspace of A^{T}. Multiply the equations $x_1 - x_2 = 1$ and $x_2 - x_3 = 1$ and $x_1 - x_3 = 1$ by numbers y_1, y_2, y_3 chosen so that the equations add up to $0 = 1$.

8 In Figure 4.3, how do we know that Ax_r is equal to Ax? How do we know that this vector is in the column space? If $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ what is x_r?

9 If $A^{\mathrm{T}}Ax = 0$ then $Ax = 0$. Reason: Ax is in the nullspace of A^{T} and also in the _____ of A and those spaces are _____. *Conclusion*: $A^{\mathrm{T}}A$ *has the same nullspace as* A. *This key fact is repeated in the next section.*

10 Suppose A is a symmetric matrix ($A^{\mathrm{T}} = A$).

(a) Why is its column space perpendicular to its nullspace?

(b) If $Ax = 0$ and $Az = 5z$, which subspaces contain these "eigenvectors" x and z? **Symmetric matrices have perpendicular eigenvectors** $x^{\mathrm{T}}z = 0$.

11 (Recommended) Draw Figure 4.2 to show each subspace correctly for

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 3 & 0 \end{bmatrix}.$$

12 Find the pieces x_r and x_n and draw Figure 4.3 properly if

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad x = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

Questions 13–23 are about orthogonal subspaces.

13 Put bases for the subspaces $\boldsymbol{V}$ and $\boldsymbol{W}$ into the columns of matrices V and W. Explain why the test for orthogonal subspaces can be written $V^{\mathrm{T}}W =$ zero matrix. This matches $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w} = 0$ for orthogonal vectors.

14 The floor $\boldsymbol{V}$ and the wall $\boldsymbol{W}$ are not orthogonal subspaces, because they share a nonzero vector (along the line where they meet). No planes $\boldsymbol{V}$ and $\boldsymbol{W}$ in $\mathbf{R}^3$ can be orthogonal! Find a vector in the column spaces of both matrices:

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 5 & 4 \\ 6 & 3 \\ 5 & 1 \end{bmatrix}$$

This will be a vector $A\boldsymbol{x}$ and also $B\widehat{\boldsymbol{x}}$. Think 3 by 4 with the matrix $[\,A \;\; B\,]$.

15 Extend Problem 14 to a p-dimensional subspace $\boldsymbol{V}$ and a q-dimensional subspace $\boldsymbol{W}$ of $\mathbf{R}^n$. What inequality on $p+q$ guarantees that $\boldsymbol{V}$ intersects $\boldsymbol{W}$ in a nonzero vector? These subspaces cannot be orthogonal.

16 Prove that every $\boldsymbol{y}$ in $\boldsymbol{N}(A^{\mathrm{T}})$ is perpendicular to every $A\boldsymbol{x}$ in the column space, using the matrix shorthand of equation (2). Start from $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$.

17 If $\boldsymbol{S}$ is the subspace of $\mathbf{R}^3$ containing only the zero vector, what is $\boldsymbol{S}^{\perp}$? If $\boldsymbol{S}$ is spanned by $(1,1,1)$, what is $\boldsymbol{S}^{\perp}$? If $\boldsymbol{S}$ is spanned by $(1,1,1)$ and $(1,1,-1)$, what is a basis for $\boldsymbol{S}^{\perp}$?

18 Suppose $\boldsymbol{S}$ only contains two vectors $(1,5,1)$ and $(2,2,2)$ (not a subspace). Then $\boldsymbol{S}^{\perp}$ is the nullspace of the matrix $A =$ _____. $\boldsymbol{S}^{\perp}$ is a subspace even if $\boldsymbol{S}$ is not.

19 Suppose $\boldsymbol{L}$ is a one-dimensional subspace (a line) in $\mathbf{R}^3$. Its orthogonal complement $\boldsymbol{L}^{\perp}$ is the _____ perpendicular to $\boldsymbol{L}$. Then $(\boldsymbol{L}^{\perp})^{\perp}$ is a _____ perpendicular to $\boldsymbol{L}^{\perp}$. In fact $(\boldsymbol{L}^{\perp})^{\perp}$ is the same as _____.

20 Suppose $\boldsymbol{V}$ is the whole space $\mathbf{R}^4$. Then $\boldsymbol{V}^{\perp}$ contains only the vector _____. Then $(\boldsymbol{V}^{\perp})^{\perp}$ is _____. So $(\boldsymbol{V}^{\perp})^{\perp}$ is the same as _____.

21 Suppose $\boldsymbol{S}$ is spanned by the vectors $(1,2,2,3)$ and $(1,3,3,2)$. Find two vectors that span $\boldsymbol{S}^{\perp}$. This is the same as solving $A\boldsymbol{x} = \mathbf{0}$ for which A?

22 If $\boldsymbol{P}$ is the plane of vectors in $\mathbf{R}^4$ satisfying $x_1 + x_2 + x_3 + x_4 = 0$, write a basis for $\boldsymbol{P}^{\perp}$. Construct a matrix that has $\boldsymbol{P}$ as its nullspace.

23 If a subspace $\boldsymbol{S}$ is contained in a subspace $\boldsymbol{V}$, prove that $\boldsymbol{S}^{\perp}$ contains $\boldsymbol{V}^{\perp}$.

Questions 24–30 are about perpendicular columns and rows.

24 Suppose an n by n matrix is invertible: $AA^{-1} = I$. Then the first column of A^{-1} is orthogonal to the space spanned by which rows of A?

25 Find $A^{\mathrm{T}}A$ if the columns of A are unit vectors, all mutually perpendicular.

26 Construct a 3 by 3 matrix A with no zero entries whose columns are mutually perpendicular. Compute $A^{\mathrm{T}}A$. Why is it a diagonal matrix?

27 The lines $3x + y = b_1$ and $6x + 2y = b_2$ are _____. They are the same line if _____. In that case (b_1, b_2) is perpendicular to the vector _____. The nullspace of the matrix is the line $3x + y =$ _____. One particular vector in that nullspace is _____.

28 Why is each of these statements false?

(a) $(1, 1, 1)$ is perpendicular to $(1, 1, -2)$ so the planes $x + y + z = 0$ and $x + y - 2z = 0$ are orthogonal subspaces.

(b) The subspace spanned by $(1, 1, 0, 0, 0)$ and $(0, 0, 0, 1, 1)$ is the orthogonal complement of the subspace spanned by $(1, -1, 0, 0, 0)$ and $(2, -2, 3, 4, -4)$.

(c) Two subspaces that meet only in the zero vector are orthogonal.

29 Find a matrix with $\boldsymbol{v} = (1, 2, 3)$ in the row space and column space. Find another matrix with $\boldsymbol{v}$ in the nullspace and column space. Which pairs of subspaces can $\boldsymbol{v}$ *not* be in?

Challenge Problems

30 Suppose A is 3 by 4 and B is 4 by 5 and $AB = 0$. So $\boldsymbol{N}(A)$ contains $\boldsymbol{C}(B)$. Prove from the dimensions of $\boldsymbol{N}(A)$ and $\boldsymbol{C}(B)$ that $\operatorname{rank}(A) + \operatorname{rank}(B) \leq 4$.

31 The command $N = \mathsf{null}(A)$ will produce a basis for the nullspace of A. Then the command $B = \mathsf{null}(N')$ will produce a basis for the _____ of A.

32 Suppose I give you four nonzero vectors $\boldsymbol{r}, \boldsymbol{n}, \boldsymbol{c}, \boldsymbol{l}$ in $\mathbf{R}^2$.

(a) What are the conditions for those to be bases for the four fundamental subspaces $C(A^{\mathrm{T}})$, $N(A)$, $C(A)$, $N(A^{\mathrm{T}})$ of a 2 by 2 matrix?

(b) What is one possible matrix A?

33 Suppose I give you eight vectors $\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{l}_1, \boldsymbol{l}_2$ in $\mathbf{R}^4$.

(a) What are the conditions for those pairs to be bases for the four fundamental subspaces of a 4 by 4 matrix?

(b) What is one possible matrix A?

4.2 Projections

May we start this section with two questions? (In addition to that one.) The first question aims to show that projections are easy to visualize. The second question is about "projection matrices"—symmetric matrices with $P^2 = P$. *The projection of b is Pb.*

1 What are the projections of $b = (2, 3, 4)$ onto the z axis and the xy plane?

2 What matrices produce those projections onto a line and a plane?

When b is projected onto a line, ***its projection p is the part of b along that line.*** If b is projected onto a plane, p is the part in that plane. ***The projection p is Pb.***

The projection matrix P multiplies b to give p. This section finds p and P.

The projection onto the z axis we call p_1. The second projection drops straight down to the xy plane. The picture in your mind should be Figure 4.4. Start with $b = (2, 3, 4)$. One projection gives $p_1 = (0, 0, 4)$ and the other gives $p_2 = (2, 3, 0)$. Those are the parts of b along the z axis and in the xy plane.

The projection matrices P_1 and P_2 are 3 by 3. They multiply b with 3 components to produce p with 3 components. Projection onto a line comes from a rank one matrix. Projection onto a plane comes from a rank two matrix:

$$\text{Onto the } z \text{ axis:} \quad P_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{Onto the } xy \text{ plane:} \quad P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

P_1 picks out the z component of every vector. P_2 picks out the x and y components. To find the projections p_1 and p_2 of b, multiply b by P_1 and P_2 (small p for the vector, capital P for the matrix that produces it):

$$p_1 = P_1 b = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ z \end{bmatrix} \qquad p_2 = P_2 b = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}.$$

In this case the projections p_1 and p_2 are perpendicular. The xy plane and the z axis are ***orthogonal subspaces***, like the floor of a room and the line between two walls.

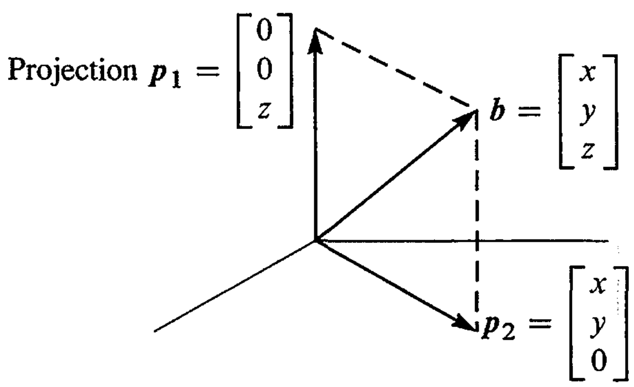


Figure 4.4: The projections $p_1 = P_1 b$ and $p_2 = P_2 b$ onto the z axis and the xy plane.

More than that, the line and plane are orthogonal ***complements***. Their dimensions add to $1+2=3$. Every vector $\boldsymbol{b}$ in the whole space is the sum of its parts in the two subspaces. The projections $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are exactly those parts:

$$\text{The vectors give } \boldsymbol{p}_1+\boldsymbol{p}_2=\boldsymbol{b}. \qquad \text{The matrices give } P_1+P_2=I. \tag{1}$$

This is perfect. Our goal is reached—for this example. We have the same goal for any line and any plane and any n-dimensional subspace. The object is to find the part $\boldsymbol{p}$ in each subspace, and the projection matrix P that produces that part $\boldsymbol{p}=P\boldsymbol{b}$. Every subspace of $\mathbf{R}^m$ has its own m by m projection matrix. To compute P, we absolutely need a good description of the subspace that it projects onto.

The best description of a subspace is a basis. We put the basis vectors into the columns of A. ***Now we are projecting onto the column space of*** A! Certainly the z axis is the column space of the 3 by 1 matrix A_1. The xy plane is the column space of A_2. That plane is also the column space of A_3 (a subspace has many bases):

$$A_1=\begin{bmatrix}0\\0\\1\end{bmatrix} \quad\text{and}\quad A_2=\begin{bmatrix}1&0\\0&1\\0&0\end{bmatrix} \quad\text{and}\quad A_3=\begin{bmatrix}1&2\\2&3\\0&0\end{bmatrix}.$$

Our problem is ***to project any b onto the column space of any*** m ***by*** n ***matrix.*** Start with a line (dimension $n=1$). The matrix A has only one column. Call it $\boldsymbol{a}$.

Projection Onto a Line

A line goes through the origin in the direction of $\boldsymbol{a}=(a_1,\ldots,a_m)$. Along that line, we want the point $\boldsymbol{p}$ closest to $\boldsymbol{b}=(b_1,\ldots,b_m)$. The key to projection is orthogonality: ***The line from b to p is perpendicular to the vector a.*** This is the dotted line marked $\boldsymbol{e}$ for error in Figure 4.5—which we now compute by algebra.

The projection $\boldsymbol{p}$ is some multiple of $\boldsymbol{a}$. Call it $\boldsymbol{p}=\widehat{x}\boldsymbol{a}=$ "x hat" times $\boldsymbol{a}$. Computing this number $\widehat{x}$ will give the vector $\boldsymbol{p}$. Then from the formula for $\boldsymbol{p}$, we read off the projection matrix P. These three steps will lead to all projection matrices: *find* $\widehat{x}$, *then find the* ***vector p***, *then find the* ***matrix*** P.

The dotted line $\boldsymbol{b}-\boldsymbol{p}$ is $\boldsymbol{e}=\boldsymbol{b}-\widehat{x}\boldsymbol{a}$. It is perpendicular to $\boldsymbol{a}$—this will determine $\widehat{x}$. Use the fact that $\boldsymbol{b}-\boldsymbol{p}$ is perpendicular to $\boldsymbol{a}$ when their dot product is zero:

Projecting $\boldsymbol{b}$ onto $\boldsymbol{a}$, error $\boldsymbol{e}=\boldsymbol{b}-\widehat{x}\boldsymbol{a}$

$$\boldsymbol{a}\cdot(\boldsymbol{b}-\widehat{x}\boldsymbol{a})=0 \quad\text{or}\quad \boldsymbol{a}\cdot\boldsymbol{b}-\widehat{x}\boldsymbol{a}\cdot\boldsymbol{a}=0 \qquad \widehat{x}=\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{\boldsymbol{a}\cdot\boldsymbol{a}}=\frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}. \tag{2}$$

The multiplication $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}$ is the same as $\boldsymbol{a}\cdot\boldsymbol{b}$. Using the transpose is better, because it applies also to matrices. Our formula $\widehat{x}=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ gives the projection $\boldsymbol{p}=\widehat{x}\boldsymbol{a}$.

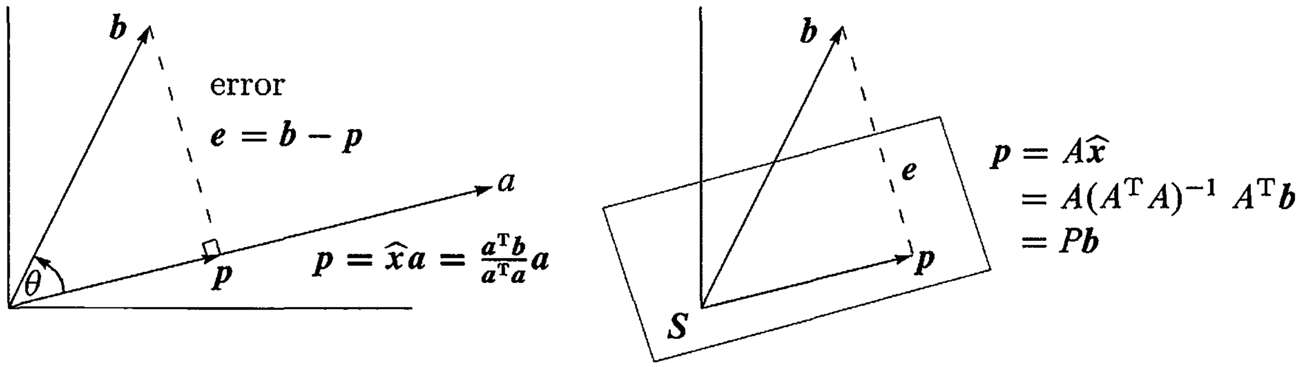


Figure 4.5: The projection p of b onto a line and onto S = column space of A.

The projection of b onto the line through a is the vector $p = \widehat{x}a = \frac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a}a$.

Special case 1: If $b = a$ then $\widehat{x} = 1$. The projection of a onto a is itself. $Pa = a$.

Special case 2: If b is perpendicular to a then $a^{\mathrm{T}}b = 0$. The projection is $p = \mathbf{0}$.

Example 1 Project $b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ onto $a = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ to find $p = \widehat{x}a$ in Figure 4.5.

Solution The number $\widehat{x}$ is the ratio of $a^{\mathrm{T}}b = 5$ to $a^{\mathrm{T}}a = 9$. So the projection is $p = \frac{5}{9}a$. The error vector between b and p is $e = b - p$. Those vectors p and e will add to $b = (1, 1, 1)$:

$$p = \frac{5}{9}a = \left(\frac{5}{9}, \frac{10}{9}, \frac{10}{9}\right) \quad \text{and} \quad e = b - p = \left(\frac{4}{9}, -\frac{1}{9}, -\frac{1}{9}\right).$$

The error e should be perpendicular to $a = (1, 2, 2)$ and it is: $e^{\mathrm{T}}a = \frac{4}{9} - \frac{2}{9} - \frac{2}{9} = 0$.

Look at the right triangle of b, p, and e. The vector b is split into two parts—its component along the line is p, its perpendicular part is e. Those two sides of a right triangle have length $\|b\| \cos\theta$ and $\|b\| \sin\theta$. Trigonometry matches the dot product:

$$p = \frac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a}a \quad \text{has length} \quad \|p\| = \frac{\|a\|\,\|b\|\cos\theta}{\|a\|^2}\|a\| = \|b\|\cos\theta. \tag{3}$$

The dot product is a lot simpler than getting involved with $\cos\theta$ and the length of b. The example has square roots in $\cos\theta = 5/3\sqrt{3}$ and $\|b\| = \sqrt{3}$. There are no square roots in the projection $p = 5a/9$. The good way to $5/9$ is $b^{\mathrm{T}}a/a^{\mathrm{T}}a$.

Now comes the ***projection matrix***. In the formula for p, what matrix is multiplying b? You can see the matrix better if the number $\widehat{x}$ is on the right side of a:

Projection matrix P $\quad p = a\widehat{x} = a\frac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a} = Pb$ ***when the matrix is*** $P = \frac{aa^{\mathrm{T}}}{a^{\mathrm{T}}a}$.

P is a column times a row! The column is $\boldsymbol{a}$, the row is $\boldsymbol{a}^{\mathrm{T}}$. Then divide by the number $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$. The projection matrix P is m by m, but ***its rank is one***. We are projecting onto a one-dimensional subspace, the line through $\boldsymbol{a}$. That is the column space of P.

Example 2 Find the projection matrix $P = \dfrac{\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}$ onto the line through $\boldsymbol{a} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$.

Solution Multiply column $\boldsymbol{a}$ times row $\boldsymbol{a}^{\mathrm{T}}$ and divide by $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a} = 9$:

$$\text{Projection matrix} \qquad P = \frac{\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}} = \frac{1}{9}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}\begin{bmatrix} 1 & 2 & 2 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}.$$

This matrix projects *any* vector $\boldsymbol{b}$ onto $\boldsymbol{a}$. Check $\boldsymbol{p} = P\boldsymbol{b}$ for $\boldsymbol{b} = (1, 1, 1)$ in Example 1:

$$\boldsymbol{p} = P\boldsymbol{b} = \frac{1}{9}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 5 \\ 10 \\ 10 \end{bmatrix} \qquad \text{which is correct.}$$

If the vector $\boldsymbol{a}$ is doubled, the matrix P stays the same. It still projects onto the same line. If the matrix is squared, P^2 equals P. ***Projecting a second time doesn't change anything***, so $\boldsymbol{P^2 = P}$. The diagonal entries of P add up to $\frac{1}{9}(1 + 4 + 4) = 1$.

The matrix $I - P$ should be a projection too. It produces the other side $\boldsymbol{e}$ of the triangle—the perpendicular part of $\boldsymbol{b}$. Note that $(I - P)\boldsymbol{b}$ equals $\boldsymbol{b} - \boldsymbol{p}$ which is $\boldsymbol{e}$ in the left nullspace. ***When P projects onto one subspace, $I - P$ projects onto the perpendicular subspace***. Here $I - P$ projects onto the plane perpendicular to $\boldsymbol{a}$.

Now we move beyond projection onto a line. Projecting onto an n-dimensional subspace of $\mathbf{R}^m$ takes more effort. The crucial formulas will be collected in equations (5)–(6)–(7). Basically you need to remember those three equations.

Projection Onto a Subspace

Start with n vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ in $\mathbf{R}^m$. Assume that these $\boldsymbol{a}$'s are linearly independent.

Problem: Find the combination $\boldsymbol{p} = \widehat{x}_1\boldsymbol{a}_1 + \cdots + \widehat{x}_n\boldsymbol{a}_n$ closest to a given vector $\boldsymbol{b}$. We are projecting each $\boldsymbol{b}$ in $\mathbf{R}^m$ onto the subspace spanned by the $\boldsymbol{a}$'s, to get $\boldsymbol{p}$.

With $n = 1$ (only one vector $\boldsymbol{a}_1$) this is projection onto a line. The line is the column space of A, which has just one column. In general the matrix A has n columns $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$.

The combinations in $\mathbf{R}^m$ are the vectors $A\boldsymbol{x}$ in the column space. We are looking for the particular combination $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$ (***the projection***) that is closest to $\boldsymbol{b}$. The hat over $\widehat{\boldsymbol{x}}$ indicates the *best* choice $\widehat{\boldsymbol{x}}$, to give the closest vector in the column space. That choice is $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ when $n = 1$. For $n > 1$, the best $\widehat{\boldsymbol{x}}$ is to be found now.

We compute projections onto n-dimensional subspaces in three steps as before: ***Find the vector $\widehat{x}$, find the projection $p = A\widehat{x}$, find the matrix P.***

The key is in the geometry! The dotted line in Figure 4.5 goes from $\boldsymbol{b}$ to the nearest point $A\widehat{\boldsymbol{x}}$ in the subspace. ***This error vector $\boldsymbol{b} - A\widehat{\boldsymbol{x}}$ is perpendicular to the subspace.***

The error $b - A\widehat{x}$ makes a right angle with all the vectors $a_1, \ldots, a_n$. The n right angles give the n equations for $\widehat{x}$:

$$\begin{matrix} a_1^{\mathrm{T}}(b - A\widehat{x}) = 0 \\ \vdots \\ a_n^{\mathrm{T}}(b - A\widehat{x}) = 0 \end{matrix} \qquad \text{or} \qquad \begin{bmatrix} - a_1^{\mathrm{T}} - \\ \vdots \\ - a_n^{\mathrm{T}} - \end{bmatrix} \begin{bmatrix} \\ b - A\widehat{x} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \mathbf{0} \\ \\ \end{bmatrix}. \tag{4}$$

The matrix with those rows a_i^{T} is A^{T}. The n equations are exactly $A^{\mathrm{T}}(b - A\widehat{x}) = \mathbf{0}$.

Rewrite $A^{\mathrm{T}}(b - A\widehat{x}) = \mathbf{0}$ in its famous form $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$. This is the equation for $\widehat{x}$, and the coefficient matrix is $A^{\mathrm{T}}A$. Now we can find $\widehat{x}$ and p and P, in that order:

The combination $p = \widehat{x}_1 a_1 + \cdots + \widehat{x}_n a_n = A\widehat{x}$ that is closest to b comes from

$$A^{\mathrm{T}}(b - A\widehat{x}) = \mathbf{0} \quad \text{or} \quad \boxed{A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b.} \tag{5}$$

This symmetric matrix $A^{\mathrm{T}}A$ is n by n. It is invertible if the a's are independent. The solution is $\widehat{x} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b$. The ***projection*** of b onto the subspace is p:

$$p = A\widehat{x} = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b. \tag{6}$$

This formula shows the n by n ***projection matrix*** that produces $p = Pb$:

$$P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}. \tag{7}$$

Compare with projection onto a line, when the matrix A has only one column a:

For $n = 1$ $\quad \widehat{x} = \dfrac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a}$ **and** $\quad p = a\dfrac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a}$ **and** $\quad P = \dfrac{aa^{\mathrm{T}}}{a^{\mathrm{T}}a}$.

Those formulas are identical with (5) and (6) and (7). The number $a^{\mathrm{T}}a$ becomes the matrix $A^{\mathrm{T}}A$. When it is a number, we divide by it. When it is a matrix, we invert it. The new formulas contain $(A^{\mathrm{T}}A)^{-1}$ instead of $1/a^{\mathrm{T}}a$. The linear independence of the columns $a_1, \ldots, a_n$ will guarantee that this inverse matrix exists.

The key step was $A^{\mathrm{T}}(b - A\widehat{x}) = \mathbf{0}$. We used geometry ($e$ is perpendicular to all the a's). Linear algebra gives this "normal equation" too, in a very quick way:

1. Our subspace is the column space of A.
2. The error vector $b - A\widehat{x}$ is perpendicular to that column space.
3. Therefore $b - A\widehat{x}$ is in the nullspace of A^{T}. This means $A^{\mathrm{T}}(b - A\widehat{x}) = \mathbf{0}$.

The left nullspace is important in projections. That nullspace of A^{T} contains the error vector $e = b - A\widehat{x}$. The vector b is being split into the projection p and the error $e = b - p$. Projection produces a right triangle (Figure 4.5) with sides p, e, and b.

Example 3 If $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$ find $\widehat{\boldsymbol{x}}$ and $\boldsymbol{p}$ and P.

Solution Compute the square matrix $A^{\mathrm{T}}A$ and also the vector $A^{\mathrm{T}}\boldsymbol{b}$:

$$A^{\mathrm{T}}A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \quad \text{and} \quad A^{\mathrm{T}}\boldsymbol{b} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}.$$

Now solve the normal equation $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ to find $\widehat{\boldsymbol{x}}$:

$$\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \quad \text{gives} \quad \widehat{\boldsymbol{x}} = \begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{5} \\ \mathbf{-3} \end{bmatrix}. \tag{8}$$

The combination $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$ is the projection of $\boldsymbol{b}$ onto the column space of A:

$$\boldsymbol{p} = \mathbf{5} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \mathbf{-3} \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}. \quad \text{The error is} \quad \boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}. \tag{9}$$

Two checks on the calculation. First, the error $\boldsymbol{e} = (1, -2, 1)$ is perpendicular to both columns $(1, 1, 1)$ and $(0, 1, 2)$. Second, the final P times $\boldsymbol{b} = (6, 0, 0)$ correctly gives $\boldsymbol{p} = (5, 2, -1)$. That solves the problem for one particular $\boldsymbol{b}$.

To find $\boldsymbol{p} = P\boldsymbol{b}$ for every $\boldsymbol{b}$, compute $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$. The determinant of $A^{\mathrm{T}}A$ is $15 - 9 = 6$; then $(A^{\mathrm{T}}A)^{-1}$ is easy. Multiply A times $(A^{\mathrm{T}}A)^{-1}$ times A^{T} to reach P:

$$(A^{\mathrm{T}}A)^{-1} = \frac{1}{6} \begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix} \quad \text{and} \quad P = \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}. \tag{10}$$

We must have $P^2 = P$, because a second projection doesn't change the first projection.

Warning The matrix $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ is deceptive. You might try to split $(A^{\mathrm{T}}A)^{-1}$ into A^{-1} times $(A^{\mathrm{T}})^{-1}$. If you make that mistake, and substitute it into P, you will find $P = AA^{-1}(A^{\mathrm{T}})^{-1}A^{\mathrm{T}}$. Apparently everything cancels. This looks like $P = I$, the identity matrix. We want to say why this is wrong.

The matrix A is rectangular. It has no inverse matrix. We cannot split $(A^{\mathrm{T}}A)^{-1}$ into A^{-1} times $(A^{\mathrm{T}})^{-1}$ because there is no A^{-1} in the first place.

In our experience, a problem that involves a rectangular matrix almost always leads to $A^{\mathrm{T}}A$. When A has independent columns, $A^{\mathrm{T}}A$ is invertible. This fact is so crucial that we state it clearly and give a proof.

$A^{\mathrm{T}}A$ is invertible if and only if A has linearly independent columns.

Proof $A^{\mathrm{T}}A$ is a square matrix (n by n). For every matrix A, we will now show that ***$A^{\mathrm{T}}A$ has the same nullspace as A.*** When the columns of A are linearly independent, its nullspace contains only the zero vector. Then $A^{\mathrm{T}}A$, with this same nullspace, is invertible.

Let A be any matrix. If $\boldsymbol{x}$ is in its nullspace, then $A\boldsymbol{x} = \boldsymbol{0}$. Multiplying by A^{T} gives $A^{\mathrm{T}}A\boldsymbol{x} = \boldsymbol{0}$. So $\boldsymbol{x}$ is also in the nullspace of $A^{\mathrm{T}}A$.

Now start with the nullspace of $A^{\mathrm{T}}A$. **From $A^{\mathrm{T}}A\boldsymbol{x} = \boldsymbol{0}$ we must prove $A\boldsymbol{x} = \boldsymbol{0}$.** We can't multiply by $(A^{\mathrm{T}})^{-1}$, which generally doesn't exist. Just multiply by $\boldsymbol{x}^{\mathrm{T}}$:

$$(\boldsymbol{x}^{\mathrm{T}})A^{\mathrm{T}}A\boldsymbol{x} = 0 \quad \text{or} \quad (A\boldsymbol{x})^{\mathrm{T}}(A\boldsymbol{x}) = 0 \quad \text{or} \quad \|A\boldsymbol{x}\|^2 = 0.$$

This says: If $A^{\mathrm{T}}A\boldsymbol{x} = \boldsymbol{0}$ then $A\boldsymbol{x}$ has length zero. Therefore $A\boldsymbol{x} = \boldsymbol{0}$. Every vector $\boldsymbol{x}$ in one nullspace is in the other nullspace. If $A^{\mathrm{T}}A$ has dependent columns, so has A. If $A^{\mathrm{T}}A$ has independent columns, so has A. This is the good case:

When A has independent columns, $A^{\mathrm{T}}A$ is square, symmetric, and invertible.

To repeat for emphasis: $A^{\mathrm{T}}A$ is (n by m) times (m by n). Then $A^{\mathrm{T}}A$ is square (n by n). It is symmetric, because its transpose is $(A^{\mathrm{T}}A)^{\mathrm{T}} = A^{\mathrm{T}}(A^{\mathrm{T}})^{\mathrm{T}}$ which equals $A^{\mathrm{T}}A$. We just proved that $A^{\mathrm{T}}A$ is invertible—provided A has independent columns. Watch the difference between dependent and independent columns:

$$\underset{A^{\mathrm{T}}}{\begin{bmatrix} 1 & 1 & 0 \\ 2 & 2 & 0 \end{bmatrix}} \underset{\text{dependent}}{\overset{A}{\begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 0 \end{bmatrix}}} = \underset{\text{singular}}{\overset{A^{\mathrm{T}}A}{\begin{bmatrix} 2 & 4 \\ 4 & 8 \end{bmatrix}}} \qquad \underset{A^{\mathrm{T}}}{\begin{bmatrix} 1 & 1 & 0 \\ 2 & 2 & 1 \end{bmatrix}} \underset{\text{indep.}}{\overset{A}{\begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 1 \end{bmatrix}}} = \underset{\text{invertible}}{\overset{A^{\mathrm{T}}A}{\begin{bmatrix} 2 & 4 \\ 4 & 9 \end{bmatrix}}}$$

Very brief summary To find the projection $\boldsymbol{p} = \widehat{x}_1\boldsymbol{a}_1 + \cdots + \widehat{x}_n\boldsymbol{a}_n$, solve $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$. This gives $\widehat{\boldsymbol{x}}$. The projection is $A\widehat{\boldsymbol{x}}$ and the error is $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = \boldsymbol{b} - A\widehat{\boldsymbol{x}}$. The projection matrix $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ gives $\boldsymbol{p} = P\boldsymbol{b}$.

This matrix satisfies $P^2 = P$. ***The distance from $\boldsymbol{b}$ to the subspace is $\|\boldsymbol{e}\|$.***

■ REVIEW OF THE KEY IDEAS ■

1. The projection of $\boldsymbol{b}$ onto the line through $\boldsymbol{a}$ is $\boldsymbol{p} = \boldsymbol{a}\widehat{x} = \boldsymbol{a}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a})$.

2. The rank one projection matrix $P = \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ multiplies $\boldsymbol{b}$ to produce $\boldsymbol{p}$.

3. Projecting $\boldsymbol{b}$ onto a subspace leaves $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ perpendicular to the subspace.

4. When A has full rank n, the equation $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ leads to $\widehat{\boldsymbol{x}}$ and $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$.

5. The projection matrix $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ has $P^{\mathrm{T}} = P$ and $P^2 = P$.

■ WORKED EXAMPLES ■

4.2 A Project the vector $\boldsymbol{b} = (3,4,4)$ onto the line through $\boldsymbol{a} = (2,2,1)$ and then onto the plane that also contains $\boldsymbol{a}^* = (1,0,0)$. Check that the first error vector $\boldsymbol{b} - \boldsymbol{p}$ is perpendicular to $\boldsymbol{a}$, and the second error vector $\boldsymbol{e}^* = \boldsymbol{b} - \boldsymbol{p}^*$ is also perpendicular to $\boldsymbol{a}^*$.

Find the 3 by 3 projection matrix P onto that plane of $\boldsymbol{a}$ and $\boldsymbol{a}^*$. Find a vector whose projection onto the plane is the zero vector.

Solution The projection of $\boldsymbol{b} = (3,4,4)$ onto the line through $\boldsymbol{a} = (2,2,1)$ is $\boldsymbol{p} = 2\boldsymbol{a}$:

Onto a line $$\boldsymbol{p} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}\boldsymbol{a} = \frac{18}{9}(2,2,1) = (4,4,2).$$

The error vector $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = (-1,0,2)$ is perpendicular to $\boldsymbol{a}$. So $\boldsymbol{p}$ is correct.

The plane of $\boldsymbol{a} = (2,2,1)$ and $\boldsymbol{a}^* = (1,0,0)$ is the column space of $A = [\boldsymbol{a}\ \boldsymbol{a}^*]$:

$$A = \begin{bmatrix} 2 & 1 \\ 2 & 0 \\ 1 & 0 \end{bmatrix} \quad A^{\mathrm{T}}A = \begin{bmatrix} 9 & 2 \\ 2 & 1 \end{bmatrix} \quad (A^{\mathrm{T}}A)^{-1} = \frac{1}{5}\begin{bmatrix} 1 & -2 \\ -2 & 9 \end{bmatrix} \quad P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & .8 & .4 \\ 0 & .4 & .2 \end{bmatrix}$$

Then $\boldsymbol{p}^* = P\boldsymbol{b} = (3, 4.8, 2.4)$. The error $\boldsymbol{e}^* = \boldsymbol{b} - \boldsymbol{p}^* = (0, -.8, 1.6)$ is perpendicular to $\boldsymbol{a}$ and $\boldsymbol{a}^*$. This $\boldsymbol{e}^*$ is in the nullspace of P and its projection is zero! Note $P^2 = P$.

4.2 B Suppose your pulse is measured at $x = 70$ beats per minute, then at $x = 80$, then at $x = 120$. Those three equations $Ax = \boldsymbol{b}$ in one unknown have $A^{\mathrm{T}} = [1\ 1\ 1]$ and $\boldsymbol{b} = (70, 80, 120)$. ***The best $\widehat{x}$ is the _____ of 70, 80, 120.*** Use calculus and projection:

1. Minimize $E = (x-70)^2 + (x-80)^2 + (x-120)^2$ by solving $dE/dx = 0$.
2. Project $\boldsymbol{b} = (70, 80, 120)$ onto $\boldsymbol{a} = (1,1,1)$ to find $\widehat{x} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$.

Solution The closest horizontal line to the heights 70, 80, 120 is the *average* $\widehat{x} = 90$:

$$\frac{dE}{dx} = 2(x-70) + 2(x-80) + 2(x-120) = 0 \quad \text{gives} \quad \widehat{x} = \frac{70+80+120}{3}$$

Projection : $$\widehat{x} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}} = \frac{(1,1,1)^{\mathrm{T}}(70,80,120)}{(1,1,1)^{\mathrm{T}}(1,1,1)} = \frac{70+80+120}{3} = 90.$$

4.2 C In *recursive* least squares, a fourth measurement 130 changes $\widehat{x}_{\text{old}}$ to $\widehat{x}_{\text{new}}$. Compute $\widehat{x}_{\text{new}}$ and verify the *update formula* $\widehat{x}_{\text{new}} = \widehat{x}_{\text{old}} + \frac{1}{4}(130 - \widehat{x}_{\text{old}})$.

Going from 999 to 1000 measurements, $\widehat{x}_{\text{new}} = \widehat{x}_{\text{old}} + \frac{1}{1000}(b_{1000} - \widehat{x}_{\text{old}})$ would only need $\widehat{x}_{\text{old}}$ and the latest value b_{1000}. We don't have to average all 1000 numbers!

Solution The new measurement $b_4 = 130$ adds a fourth equation and $\widehat{x}$ is updated to 100. You can average b_1, b_2, b_3, b_4 or combine the average of b_1, b_2, b_3 with b_4:

$$\frac{70+80+120+130}{4} = 100 \quad \text{is also} \quad \widehat{x}_{\text{old}} + \frac{1}{4}(b_4 - \widehat{x}_{\text{old}}) = 90 + \frac{1}{4}(40).$$

The update from 999 to 1000 measurements shows the "gain matrix" $\frac{1}{1000}$ in a **Kalman filter** multiplying the prediction error $b_{\text{new}} - \widehat{x}_{\text{old}}$. Notice $\frac{1}{1000} = \frac{1}{999} - \frac{1}{999000}$:

$$\widehat{x}_{\text{new}} = \frac{b_1 + \cdots + b_{1000}}{1000} = \frac{b_1 + \cdots + b_{999}}{999} + \frac{1}{1000}\left(b_{1000} - \frac{b_1 + \cdots + b_{999}}{999}\right).$$

Problem Set 4.2

Questions 1–9 ask for projections onto lines. Also errors $e = b - p$ and matrices P.

1 Project the vector $\boldsymbol{b}$ onto the line through $\boldsymbol{a}$. Check that $\boldsymbol{e}$ is perpendicular to $\boldsymbol{a}$:

$$\text{(a)}\ \boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \text{ and } \boldsymbol{a} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \text{(b)}\ \boldsymbol{b} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \text{ and } \boldsymbol{a} = \begin{bmatrix} -1 \\ -3 \\ -1 \end{bmatrix}.$$

2 *Draw* the projection of $\boldsymbol{b}$ onto $\boldsymbol{a}$ and also compute it from $\boldsymbol{p} = \widehat{x}\boldsymbol{a}$:

$$\text{(a)}\ \boldsymbol{b} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \text{ and } \boldsymbol{a} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \text{(b)}\ \boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ and } \boldsymbol{a} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

3 In Problem 1, find the projection matrix $P = \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ onto the line through each vector $\boldsymbol{a}$. Verify in both cases that $P^2 = P$. Multiply $P\boldsymbol{b}$ in each case to compute the projection $\boldsymbol{p}$.

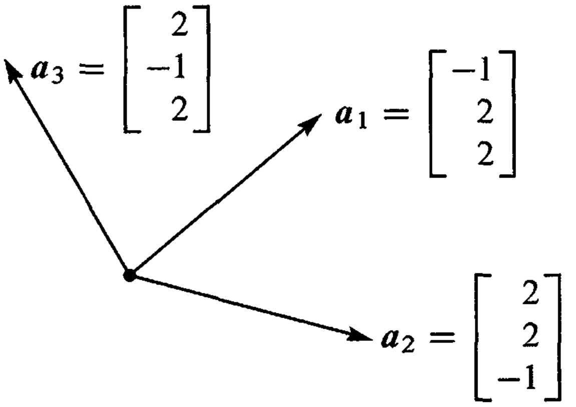
4 Construct the projection matrices P_1 and P_2 onto the lines through the $\boldsymbol{a}$'s in Problem 2. Is it true that $(P_1 + P_2)^2 = P_1 + P_2$? This *would* be true if $P_1P_2 = 0$.

5 Compute the projection matrices $\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ onto the lines through $\boldsymbol{a}_1 = (-1, 2, 2)$ and $\boldsymbol{a}_2 = (2, 2, -1)$. Multiply those projection matrices and explain why their product P_1P_2 is what it is.

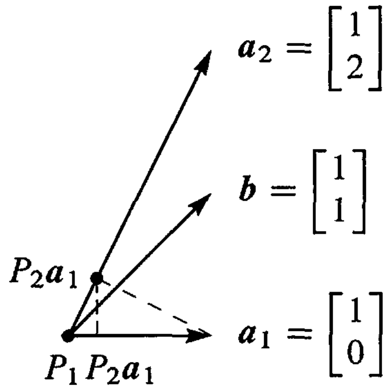
6 Project $\boldsymbol{b} = (1, 0, 0)$ onto the lines through $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ in Problem 5 and also onto $\boldsymbol{a}_3 = (2, -1, 2)$. Add up the three projections $\boldsymbol{p}_1 + \boldsymbol{p}_2 + \boldsymbol{p}_3$.

7 Continuing Problems 5–6, find the projection matrix P_3 onto $\boldsymbol{a}_3 = (2, -1, 2)$. Verify that $P_1 + P_2 + P_3 = I$. The basis $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ is orthogonal!

8 Project the vector $\boldsymbol{b} = (1, 1)$ onto the lines through $\boldsymbol{a}_1 = (1, 0)$ and $\boldsymbol{a}_2 = (1, 2)$. Draw the projections $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ and add $\boldsymbol{p}_1 + \boldsymbol{p}_2$. The projections do not add to $\boldsymbol{b}$ because the $\boldsymbol{a}$'s are not orthogonal.



Questions 5–6–7



Questions 8–9–10

9 In Problem 8, the projection of $\boldsymbol{b}$ onto the *plane* of $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ will equal $\boldsymbol{b}$. Find $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ for $A = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 \end{bmatrix} = \left[\begin{smallmatrix} 1 & 1 \\ 0 & 2 \end{smallmatrix}\right]$.

10 Project $\boldsymbol{a}_1 = (1, 0)$ onto $\boldsymbol{a}_2 = (1, 2)$. Then project the result back onto $\boldsymbol{a}_1$. Draw these projections and multiply the projection matrices $P_1 P_2$: Is this a projection?

Questions 11–20 ask for projections, and projection matrices, onto subspaces.

11 Project $\boldsymbol{b}$ onto the column space of A by solving $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ and $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$:

$$\text{(a)}\ A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \text{ and } \boldsymbol{b} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \qquad \text{(b)}\ A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \text{ and } \boldsymbol{b} = \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix}.$$

Find $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$. It should be perpendicular to the columns of A.

12 Compute the projection matrices P_1 and P_2 onto the column spaces in Problem 11. Verify that $P_1\boldsymbol{b}$ gives the first projection $\boldsymbol{p}_1$. Also verify $P_2^2 = P_2$.

13 (Quick and Recommended) Suppose A is the 4 by 4 identity matrix with its last column removed. A is 4 by 3. Project $\boldsymbol{b} = (1, 2, 3, 4)$ onto the column space of A. What shape is the projection matrix P and what is P?

14 Suppose $\boldsymbol{b}$ equals 2 times the first column of A. What is the projection of $\boldsymbol{b}$ onto the column space of A? Is $P = I$ for sure in this case? Compute $\boldsymbol{p}$ and P when $\boldsymbol{b} = (0, 2, 4)$ and the columns of A are $(0, 1, 2)$ and $(1, 2, 0)$.

15 If A is doubled, then $P = 2A(4A^{\mathrm{T}}A)^{-1}2A^{\mathrm{T}}$. This is the same as $A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$. The column space of $2A$ is the same as _____. Is $\widehat{\boldsymbol{x}}$ the same for A and $2A$?

16 What linear combination of $(1, 2, -1)$ and $(1, 0, 1)$ is closest to $\boldsymbol{b} = (2, 1, 1)$?

17 (*Important*) If $P^2 = P$ show that $(I - P)^2 = I - P$. When P projects onto the column space of A, $I - P$ projects onto the _____.

18 (a) If P is the 2 by 2 projection matrix onto the line through $(1,1)$, then $I - P$ is the projection matrix onto ______.

(b) If P is the 3 by 3 projection matrix onto the line through $(1,1,1)$, then $I - P$ is the projection matrix onto ______.

19 To find the projection matrix onto the plane $x - y - 2z = 0$, choose two vectors in that plane and make them the columns of A. The plane should be the column space. Then compute $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$.

20 To find the projection matrix P onto the same plane $x - y - 2z = 0$, write down a vector $\boldsymbol{e}$ that is perpendicular to that plane. Compute the projection $Q = \boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}/\boldsymbol{e}^{\mathrm{T}}\boldsymbol{e}$ and then $P = I - Q$.

Questions 21–26 show that projection matrices satisfy $P^2 = P$ and $P^{\mathrm{T}} = P$.

21 Multiply the matrix $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ by itself. Cancel to prove that $P^2 = P$. Explain why $P(P\boldsymbol{b})$ always equals $P\boldsymbol{b}$: The vector $P\boldsymbol{b}$ is in the column space so its projection is ______.

22 Prove that $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ is symmetric by computing P^{T}. Remember that the inverse of a symmetric matrix is symmetric.

23 If A is square and invertible, the warning against splitting $(A^{\mathrm{T}}A)^{-1}$ does not apply. It is true that $AA^{-1}(A^{\mathrm{T}})^{-1}A^{\mathrm{T}} = I$. *When A is invertible, why is $P = I$?* ***What is the error $\boldsymbol{e}$?***

24 The nullspace of A^{T} is ______ to the column space $\boldsymbol{C}(A)$. So if $A^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{0}$, the projection of $\boldsymbol{b}$ onto $\boldsymbol{C}(A)$ should be $\boldsymbol{p} =$ ______. Check that $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ gives this answer.

25 The projection matrix P onto an n-dimensional subspace has rank $r = n$. ***Reason:*** The projections $P\boldsymbol{b}$ fill the subspace S. So S is the ______ of P.

26 If an m by m matrix has $A^2 = A$ and its rank is m, prove that $A = I$.

27 The important fact that ends the section is this: ***If $A^{\mathrm{T}}A\boldsymbol{x} = \boldsymbol{0}$ then $A\boldsymbol{x} = \boldsymbol{0}$.*** *New Proof*: The vector $A\boldsymbol{x}$ is in the nullspace of ______. $A\boldsymbol{x}$ is always in the column space of ______. To be in both of those perpendicular spaces, $A\boldsymbol{x}$ must be zero.

28 Use $P^{\mathrm{T}} = P$ and $P^2 = P$ to prove that the length squared of column 2 always equals the diagonal entry P_{22}. This number is $\frac{2}{6} = \frac{4}{36} + \frac{4}{36} + \frac{4}{36}$ for

$$P = \frac{1}{6}\begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}.$$

29 If B has rank m (full row rank, independent rows) show that BB^{T} is invertible.

Challenge Problems

30 (a) Find the projection matrix P_C onto the column space of A (after looking closely at the matrix!)

$$A = \begin{bmatrix} 3 & 6 & 6 \\ 4 & 8 & 8 \end{bmatrix}$$

(b) Find the 3 by 3 projection matrix P_R onto the row space of A. Multiply $B = P_C A P_R$. Your answer B should be a little surprising—can you explain it?

31 In $\mathbf{R}^m$, suppose I give you $\boldsymbol{b}$ and $\boldsymbol{p}$, and $\boldsymbol{p}$ is a combination of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. How would you test to see if $\boldsymbol{p}$ is the projection of $\boldsymbol{b}$ onto the subspace spanned by the $\boldsymbol{a}$'s?

32 Suppose P_1 is the projection matrix onto the 1-dimensional subspace spanned by the first column of A. Suppose P_2 is the projection matrix onto the 2-dimensional column space of A. After thinking a little, compute the product $P_2 P_1$.

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 1 \end{bmatrix}.$$

33 P_1 and P_2 are projections onto subspaces $\boldsymbol{S}$ and $\boldsymbol{T}$. What is the requirement on those subspaces to have $P_1 P_2 = P_2 P_1$?

34 **If A has r independent columns and B has r independent rows, AB is invertible.**

Proof: When A is m by r with independent columns, we know that $A^{\mathrm{T}} A$ is invertible. If B is r by n with independent rows, show that $B B^{\mathrm{T}}$ is invertible. (Take $A = B^{\mathrm{T}}$.)

Now show that AB has rank r. Hint: Why does $A^{\mathrm{T}} A B B^{\mathrm{T}}$ have rank r? That matrix multiplication by A^{T} and B^{T} cannot increase the rank of AB, by Problem 3.6:26.

4.3 Least Squares Approximations

It often happens that $Ax = b$ has no solution. The usual reason is: *too many equations.* The matrix has more rows than columns. There are more equations than unknowns (m is greater than n). The n columns span a small part of m-dimensional space. Unless all measurements are perfect, b is outside that column space. Elimination reaches an impossible equation and stops. But we can't stop just because measurements include noise.

To repeat: We cannot always get the error $e = b - Ax$ down to zero. When e is zero, x is an exact solution to $Ax = b$. *When the length of e is as small as possible, $\widehat{x}$ is a least squares solution.* Our goal in this section is to compute $\widehat{x}$ and use it. These are real problems and they need an answer.

The previous section emphasized p (the projection). This section emphasizes $\widehat{x}$ (the least squares solution). They are connected by $p = A\widehat{x}$. The fundamental equation is still $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$. Here is a short unofficial way to reach this equation:

When $Ax = b$ has no solution, multiply by A^{T} and solve $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$.

Example 1 A crucial application of least squares is fitting a straight line to m points. Start with three points: *Find the closest line to the points* $(0,6)$, $(1,0)$, *and* $(2,0)$.

No straight line $b = C + Dt$ goes through those three points. We are asking for two numbers C and D that satisfy three equations. Here are the equations at $t = 0, 1, 2$ to match the given values $b = 6, 0, 0$:

$t = 0$	The first point is on the line $b = C + Dt$ if	$C + D \cdot 0 = 6$
$t = 1$	The second point is on the line $b = C + Dt$ if	$C + D \cdot 1 = 0$
$t = 2$	The third point is on the line $b = C + Dt$ if	$C + D \cdot 2 = 0.$

This 3 by 2 system has *no solution*: $b = (6,0,0)$ is not a combination of the columns $(1,1,1)$ and $(0,1,2)$. Read off A, x, and b from those equations:

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad x = \begin{bmatrix} C \\ D \end{bmatrix} \quad b = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} \quad Ax = b \text{ is } \textit{not} \text{ solvable.}$$

The same numbers were in Example 3 in the last section. We computed $\widehat{x} = (5, -3)$. Those numbers are the best C and D, so $5 - 3t$ will be the best line for the 3 points. We must connect projections to least squares, by explaining why $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$.

In practical problems, there could easily be $m = 100$ points instead of $m = 3$. They don't exactly match any straight line $C + Dt$. Our numbers 6, 0, 0 exaggerate the error so you can see e_1, e_2, and e_3 in Figure 4.6.

Minimizing the Error

How do we make the error $e = b - Ax$ as small as possible? This is an important question with a beautiful answer. The best x (called $\widehat{x}$) can be found by geometry or algebra or calculus: 90° angle or project using P or set the derivative of the error to zero.

By geometry Every Ax lies in the plane of the columns $(1, 1, 1)$ and $(0, 1, 2)$. In that plane, we look for the point closest to $\boldsymbol{b}$. *The nearest point is the projection* $\boldsymbol{p}$.

The best choice for $A\widehat{\boldsymbol{x}}$ is $\boldsymbol{p}$. The smallest possible error is $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$. The three points at heights (p_1, p_2, p_3) *do lie on a line*, because $\boldsymbol{p}$ is in the column space. In fitting a straight line, $\widehat{\boldsymbol{x}}$ gives the best choice for (C, D).

By algebra Every vector $\boldsymbol{b}$ splits into two parts. The part in the column space is $\boldsymbol{p}$. The perpendicular part in the nullspace of A^{T} is $\boldsymbol{e}$. There is an equation we cannot solve $(A\boldsymbol{x} = \boldsymbol{b})$. There is an equation $A\widehat{\boldsymbol{x}} = \boldsymbol{p}$ we do solve (by removing $\boldsymbol{e}$):

$$A\boldsymbol{x} = \boldsymbol{b} = \boldsymbol{p} + \boldsymbol{e} \quad \text{is impossible;} \qquad A\widehat{\boldsymbol{x}} = \boldsymbol{p} \quad \text{is solvable.} \tag{1}$$

The solution to $A\widehat{\boldsymbol{x}} = \boldsymbol{p}$ leaves the least possible error (which is $\boldsymbol{e}$):

$$\textbf{Squared length for any } \boldsymbol{x} \qquad \|A\boldsymbol{x} - \boldsymbol{b}\|^2 = \|A\boldsymbol{x} - \boldsymbol{p}\|^2 + \|\boldsymbol{e}\|^2. \tag{2}$$

This is the law $c^2 = a^2 + b^2$ for a right triangle. The vector $A\boldsymbol{x} - \boldsymbol{p}$ in the column space is perpendicular to $\boldsymbol{e}$ in the left nullspace. We reduce $A\boldsymbol{x} - \boldsymbol{p}$ to zero by choosing $\boldsymbol{x}$ to be $\widehat{\boldsymbol{x}}$. That leaves the smallest possible error $\boldsymbol{e} = (e_1, e_2, e_3)$.

Notice what "smallest" means. The *squared length* of $A\boldsymbol{x} - \boldsymbol{b}$ is minimized:

The least squares solution $\widehat{\boldsymbol{x}}$ makes $E = \|A\boldsymbol{x} - \boldsymbol{b}\|^2$ as small as possible.

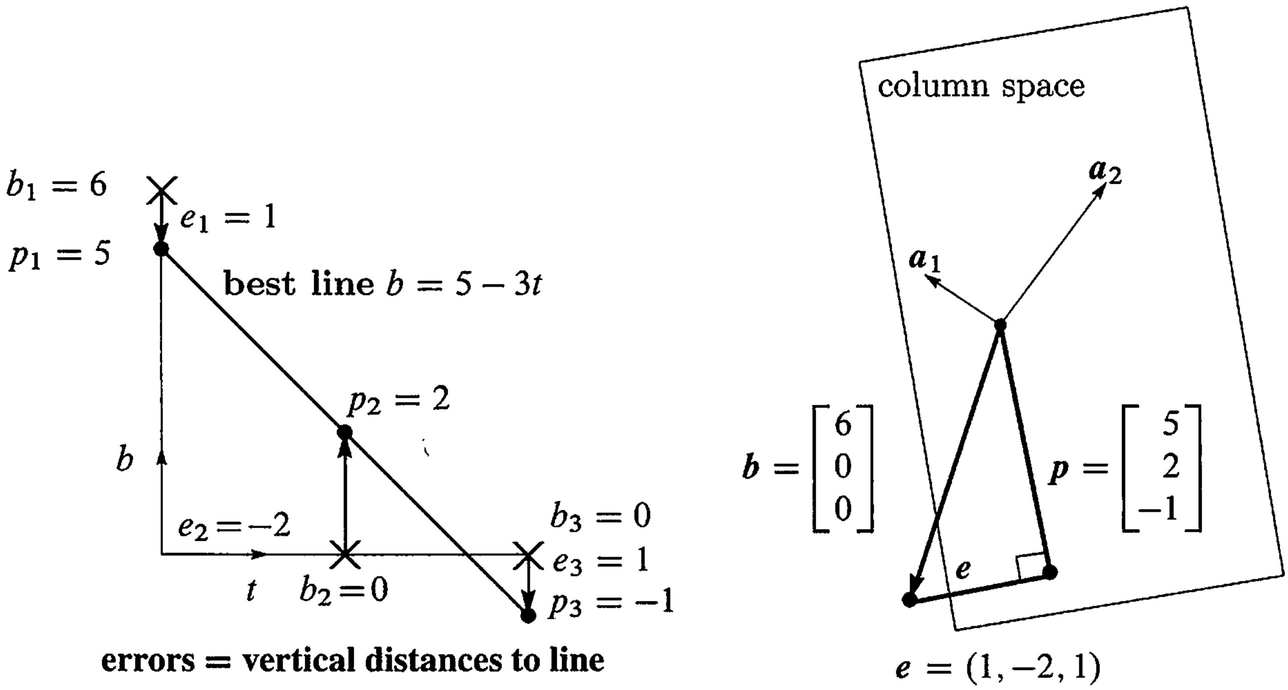


Figure 4.6: **Best line and projection: Two pictures, same problem.** The line has heights $\boldsymbol{p} = (5, 2, -1)$ with errors $\boldsymbol{e} = (1, -2, 1)$. The equations $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ give $\widehat{\boldsymbol{x}} = (5, -3)$. The best line is $b = 5 - 3t$ and the projection is $\boldsymbol{p} = 5\boldsymbol{a}_1 - 3\boldsymbol{a}_2$.

Figure 4.6a shows the closest line. It misses by distances $e_1, e_2, e_3 = 1, -2, 1$. *Those are vertical distances.* The least squares line minimizes $E = e_1^2 + e_2^2 + e_3^2$.

Figure 4.6b shows the same problem in 3-dimensional space ($\boldsymbol{b}$ $\boldsymbol{p}$ $\boldsymbol{e}$ space). The vector $\boldsymbol{b}$ is not in the column space of A. That is why we could not solve $A\boldsymbol{x} = \boldsymbol{b}$. No line goes through the three points. The smallest possible error is the perpendicular vector $\boldsymbol{e}$. This is $\boldsymbol{e} = \boldsymbol{b} - A\widehat{\boldsymbol{x}}$, the vector of errors $(1, -2, 1)$ in the three equations. Those are the distances from the best line. Behind both figures is the fundamental equation $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$.

Notice that the errors $1, -2, 1$ add to zero. The error $\boldsymbol{e} = (e_1, e_2, e_3)$ is perpendicular to the first column $(1, 1, 1)$ in A. The dot product gives $e_1 + e_2 + e_3 = 0$.

By calculus Most functions are minimized by calculus! The graph bottoms out and the derivative in every direction is zero. Here the error function E to be minimized is a *sum of squares* $e_1^2 + e_2^2 + e_3^2$ (the square of the error in each equation):

$$E = \|A\boldsymbol{x} - \boldsymbol{b}\|^2 = (C + D \cdot 0 - 6)^2 + (C + D \cdot 1)^2 + (C + D \cdot 2)^2. \tag{3}$$

The unknowns are C and D. With two unknowns there are *two derivatives*—both zero at the minimum. They are "partial derivatives" because $\partial E/\partial C$ treats D as constant and $\partial E/\partial D$ treats C as constant:

$$\begin{aligned} \partial E/\partial C &= 2(C + D \cdot 0 - 6) \quad + 2(C + D \cdot 1) \quad + 2(C + D \cdot 2) \quad = 0 \\ \partial E/\partial D &= 2(C + D \cdot 0 - 6)(\mathbf{0}) + 2(C + D \cdot 1)(\mathbf{1}) + 2(C + D \cdot 2)(\mathbf{2}) = 0. \end{aligned}$$

$\partial E/\partial D$ contains the extra factors $\mathbf{0}, \mathbf{1}, \mathbf{2}$ from the chain rule. (The last derivative from $(C + 2\mathrm{D})^2$ was 2 times $C + 2\mathrm{D}$ times that extra 2.) In the C derivative the corresponding factors are $1, 1, 1$, because C is always multiplied by 1. It is no accident that 1, 1, 1 and 0, 1, 2 are the columns of A.

Now cancel 2 from every term and collect all C's and all D's:

$$\begin{array}{ll} \text{The } C \text{ derivative is zero:} & 3C + 3D = 6 \\ \text{The } D \text{ derivative is zero:} & 3C + 5D = 0 \end{array} \qquad \textbf{This matrix} \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \textbf{ is } A^{\mathrm{T}}A \tag{4}$$

These equations are identical with $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$. The best C and D are the components of $\widehat{\boldsymbol{x}}$. The equations from calculus are the same as the "normal equations" from linear algebra. These are the key equations of least squares:

The partial derivatives of $\|A\boldsymbol{x} - \boldsymbol{b}\|^2$ ***are zero when*** $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$.

The solution is $C = 5$ and $D = -3$. Therefore $b = 5 - 3t$ is the best line—it comes closest to the three points. At $t = 0, 1, 2$ this line goes through $\boldsymbol{p} = 5, 2, -1$. It could not go through $\boldsymbol{b} = 6, 0, 0$. The errors are $1, -2, 1$. This is the vector $\boldsymbol{e}$!

The Big Picture

The key figure of this book shows the four subspaces and the true action of a matrix. The vector $\boldsymbol{x}$ on the left side of Figure 4.3 went to $\boldsymbol{b} = A\boldsymbol{x}$ on the right side. In that figure $\boldsymbol{x}$ was split into $\boldsymbol{x}_r + \boldsymbol{x}_n$. There were many solutions to $A\boldsymbol{x} = \boldsymbol{b}$.

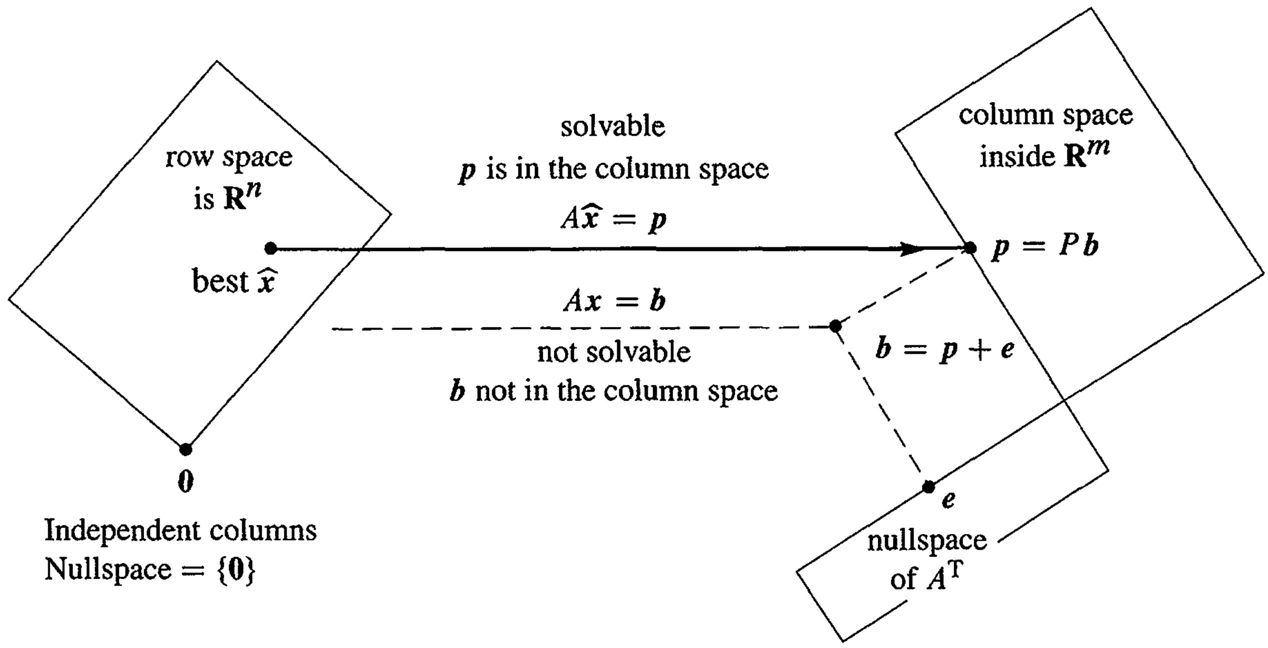


Figure 4.7: The projection $p = A\widehat{x}$ is closest to b, so $\widehat{x}$ minimizes $E = \|b - Ax\|^2$.

In this section the situation is just the opposite. There are *no* solutions to $Ax = b$. *Instead of splitting up* x *we are splitting up* b. Figure 4.3 shows the big picture for least squares. Instead of $Ax = b$ we solve $A\widehat{x} = p$. The error $e = b - p$ is unavoidable.

Notice how the nullspace $N(A)$ is very small—just one point. With independent columns, the only solution to $Ax = \mathbf{0}$ is $x = \mathbf{0}$. Then $A^{\mathrm{T}}A$ is invertible. The equation $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$ fully determines the best vector $\widehat{x}$. The error has $A^{\mathrm{T}}e = \mathbf{0}$.

Chapter 7 will have the complete picture—all four subspaces included. Every x splits into $x_r + x_n$, and every b splits into $p + e$. The best solution is $\widehat{x}_r$ in the row space. We can't help e and we don't want x_n—this leaves $A\widehat{x} = p$.

Fitting a Straight Line

Fitting a line is the clearest application of least squares. It starts with $m > 2$ points, hopefully near a straight line. At times $t_1, \ldots, t_m$ those m points are at heights $b_1, \ldots, b_m$. The best line $C + Dt$ misses the points by vertical distances $e_1, \ldots, e_m$. No line is perfect, and the least squares line minimizes $E = e_1^2 + \cdots + e_m^2$.

The first example in this section had three points in Figure 4.6. Now we allow m points (and m can be large). The two components of $\widehat{x}$ are still C and D.

A line goes through the m points when we exactly solve $Ax = b$. Generally we can't do it. Two unknowns C and D determine a line, so A has only $n = 2$ columns. To fit the m points, we are trying to solve m equations (and we only want two!):

$$Ax = b \quad \text{is} \quad \begin{matrix} C + Dt_1 = b_1 \\ C + Dt_2 = b_2 \\ \vdots \\ C + Dt_m = b_m \end{matrix} \quad \text{with} \quad A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}. \tag{5}$$

The column space is so thin that almost certainly $\boldsymbol{b}$ is outside of it. When $\boldsymbol{b}$ happens to lie in the column space, the points happen to lie on a line. In that case $\boldsymbol{b} = \boldsymbol{p}$. Then $A\boldsymbol{x} = \boldsymbol{b}$ is solvable and the errors are $\boldsymbol{e} = (0, \ldots, 0)$.

The closest line $C + Dt$ has heights $p_1, \ldots, p_m$ with errors $e_1, \ldots, e_m$.
Solve $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ for $\widehat{\boldsymbol{x}} = (C, D)$. The errors are $e_i = b_i - C - Dt_i$.

Fitting points by a straight line is so important that we give the two equations $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$, once and for all. The two columns of A are independent (unless all times t_i are the same). So we turn to least squares and solve $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$.

$$\textbf{Dot-product matrix} \quad A^{\mathrm{T}}A = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} = \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix}. \tag{6}$$

On the right side of the normal equation is the 2 by 1 vector $A^{\mathrm{T}}\boldsymbol{b}$:

$$A^{\mathrm{T}}\boldsymbol{b} = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_m \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}. \tag{7}$$

In a specific problem, these numbers are given. The best $\widehat{\boldsymbol{x}} = (C, D)$ is in equation (9).

The line $C + Dt$ minimizes $e_1^2 + \cdots + e_m^2 = \|A\boldsymbol{x} - \boldsymbol{b}\|^2$ when $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$:

$$\begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}. \tag{8}$$

The vertical errors at the m points on the line are the components of $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$. This error vector (the *residual*) $\boldsymbol{b} - A\widehat{\boldsymbol{x}}$ is perpendicular to the columns of A (geometry). The error is in the nullspace of A^{T} (linear algebra). The best $\widehat{\boldsymbol{x}} = (C, D)$ minimizes the total error E, the sum of squares:

$$E(x) = \|A\boldsymbol{x} - \boldsymbol{b}\|^2 = (C + Dt_1 - b_1)^2 + \cdots + (C + Dt_m - b_m)^2.$$

When calculus sets the derivatives $\partial E/\partial C$ and $\partial E/\partial D$ to zero, it produces $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$.

Other least squares problems have more than two unknowns. Fitting by the best parabola has $n = 3$ coefficients C, D, E (see below). In general we are fitting m data points by n parameters $x_1, \ldots, x_n$. The matrix A has n columns and $n < m$. The derivatives of $\|A\boldsymbol{x} - \boldsymbol{b}\|^2$ give the n equations $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$. **The derivative of a square is linear.** This is why the method of least squares is so popular.

Example 2 A has *orthogonal columns* when the measurement times t_i add to zero.

Suppose $b = 1, 2, 4$ at times $t = -2, 0, 2$. Those times add to zero. The columns of A have *zero dot product*:

$$\begin{aligned} C + D(-2) &= 1 \\ C + \quad D(0) &= 2 \\ C + \quad D(2) &= 4 \end{aligned} \qquad \text{or} \qquad Ax = \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}.$$

Look at the zeros in $A^{\mathrm{T}}A$:

$$A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}\boldsymbol{b} \qquad \text{is} \qquad \begin{bmatrix} 3 & 0 \\ 0 & 8 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}.$$

Main point: Now $A^{\mathrm{T}}A$ *is diagonal.* We can solve separately for $C = \frac{7}{3}$ and $D = \frac{6}{8}$. The zeros in $A^{\mathrm{T}}A$ are dot products of perpendicular columns in A. The diagonal matrix $A^{\mathrm{T}}A$, with entries $m = 3$ and $t_1^2 + t_2^2 + t_3^2 = 8$, is virtually as good as the identity matrix.

Orthogonal columns are so helpful that it is worth moving the time origin to produce them. To do that, subtract away the average time $\widehat{t} = (t_1 + \cdots + t_m)/m$. The shifted times $T_i = t_i - \widehat{t}$ add to $\sum T_i = m\widehat{t} - m\widehat{t} = 0$. With the columns now orthogonal, $A^{\mathrm{T}}A$ is diagonal. Its entries are m and $T_1^2 + \cdots + T_m^2$. The best C and D have direct formulas:

$$\boldsymbol{T} \textbf{ is } \boldsymbol{t - \widehat{t}} \qquad C = \frac{b_1 + \cdots + b_m}{m} \qquad \text{and} \qquad D = \frac{b_1T_1 + \cdots + b_mT_m}{T_1^2 + \cdots + T_m^2}. \tag{9}$$

The best line is $C + DT$ ***or*** $C + D(t - \widehat{t})$. The time shift that makes $A^{\mathrm{T}}A$ diagonal is an example of the Gram-Schmidt process: *orthogonalize the columns in advance.*

Fitting by a Parabola

If we throw a ball, it would be crazy to fit the path by a straight line. A parabola $b = C + Dt + Et^2$ allows the ball to go up and come down again (b is the height at time t). The actual path is not a perfect parabola, but the whole theory of projectiles starts with that approximation.

When Galileo dropped a stone from the Leaning Tower of Pisa, it accelerated. The distance contains a quadratic term $\frac{1}{2}gt^2$. (Galileo's point was that the stone's mass is not involved.) Without that t^2 term we could never send a satellite into the right orbit. But even with a nonlinear function like t^2, the unknowns C, D, E appear linearly! Choosing the best parabola is still a problem in linear algebra.

Problem Fit heights $b_1, \ldots, b_m$ at times $t_1, \ldots, t_m$ by a parabola $C + Dt + Et^2$.

Solution With $m > 3$ points, the m equations for an exact fit are generally unsolvable:

$$\begin{aligned} C + Dt_1 + Et_1^2 &= b_1 \\ &\vdots \\ C + Dt_m + Et_m^2 &= b_m \end{aligned} \qquad \text{has the } m \text{ by 3 matrix} \qquad A = \begin{bmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 \end{bmatrix}. \tag{10}$$

Least squares The closest parabola $C + Dt + Et^2$ chooses $\widehat{x} = (C, D, E)$ to satisfy the three normal equations $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}\boldsymbol{b}$.

May I ask you to convert this to a problem of projection? The column space of A has dimension ____. The projection of $\boldsymbol{b}$ is $\boldsymbol{p} = A\widehat{x}$, which combines the three columns using the coefficients C, D, E. The error at the first data point is $e_1 = b_1 - C - Dt_1 - Et_1^2$. The total squared error is $e_1^2 +$ ____. If you prefer to minimize by calculus, take the partial derivatives of E with respect to ____, ____, ____. These three derivatives will be zero when $\widehat{x} = (C, D, E)$ solves the 3 by 3 system of equations ____.

Section 8.5 has more least squares applications. The big one is Fourier series—approximating functions instead of vectors. The function to be minimized changes from a sum of squared errors $e_1^2 + \cdots + e_m^2$ to an integral of the squared error.

Example 3 For a parabola $b = C + Dt + Et^2$ to go through the three heights $b = 6, 0, 0$ when $t = 0, 1, 2$, the equations are

$$\begin{aligned} C + D \cdot 0 + E \cdot 0^2 &= 6 \\ C + D \cdot 1 + E \cdot 1^2 &= 0 \\ C + D \cdot 2 + E \cdot 2^2 &= 0. \end{aligned} \tag{11}$$

This is $A\boldsymbol{x} = \boldsymbol{b}$. We can solve it exactly. Three data points give three equations and a square matrix. The solution is $\boldsymbol{x} = (C, D, E) = (6, -9, 3)$. The parabola through the three points in Figure 4.8a is $b = 6 - 9t + 3t^2$.

What does this mean for projection? The matrix has three columns, which span the whole space $\mathbf{R}^3$. The projection matrix is the identity. The projection of $\boldsymbol{b}$ is $\boldsymbol{b}$. The error is zero. We didn't need $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}\boldsymbol{b}$, because we solved $A\boldsymbol{x} = \boldsymbol{b}$. Of course we could multiply by A^{T}, but there is no reason to do it.

Figure 4.8 also shows a fourth point b_4 at time t_4. If that falls on the parabola, the new $A\boldsymbol{x} = \boldsymbol{b}$ (four equations) is still solvable. When the fourth point is not on the parabola, we turn to $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}\boldsymbol{b}$. Will the least squares parabola stay the same, with all the error at the fourth point? Not likely!

The smallest error vector (e_1, e_2, e_3, e_4) is perpendicular to $(1, 1, 1, 1)$, the first column of A. Least squares balances out the four errors, and they add to zero.

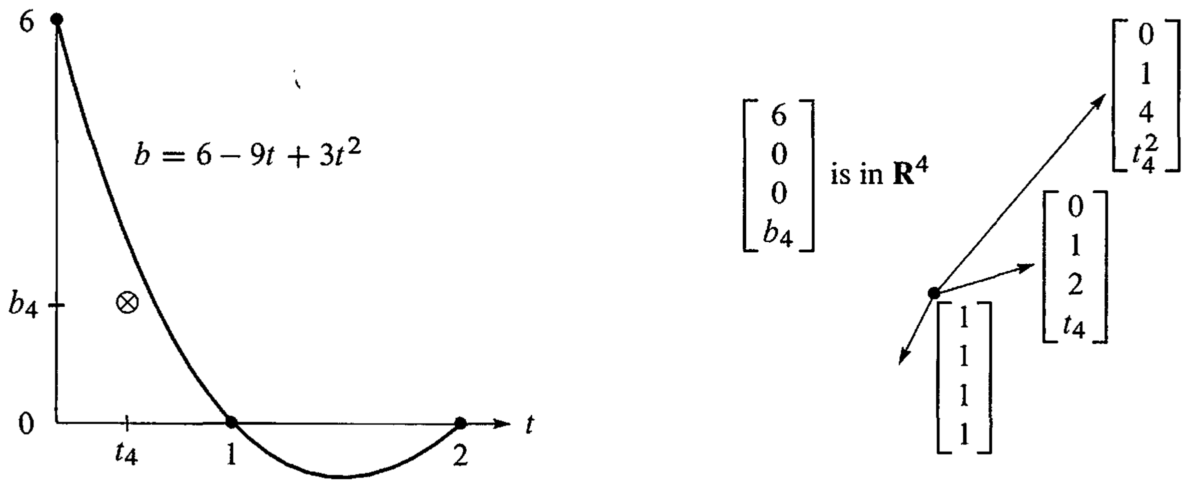


Figure 4.8: From Example 3: An exact fit of the parabola at $t = 0, 1, 2$ means that $\boldsymbol{p} = \boldsymbol{b}$ and $\boldsymbol{e} = \boldsymbol{0}$. The point b_4 off the parabola makes $m > n$ and we need least squares.

■ REVIEW OF THE KEY IDEAS ■

1. The least squares solution $\widehat{x}$ minimizes $E = \|Ax - b\|^2$. This is the sum of squares of the errors in the m equations ($m > n$).

2. The best $\widehat{x}$ comes from the normal equations $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$.

3. To fit m points by a line $b = C + Dt$, the normal equations give C and D.

4. The heights of the best line are $p = (p_1, \ldots, p_m)$. The vertical distances to the data points are the errors $e = (e_1, \ldots, e_m)$.

5. If we try to fit m points by a combination of $n < m$ functions, the m equations $Ax = b$ are generally unsolvable. The n equations $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$ give the least squares solution—the combination with smallest MSE (mean square error).

■ WORKED EXAMPLES ■

4.3 A Start with nine measurements b_1 to b_9, *all zero*, at times $t = 1, \ldots, 9$. The tenth measurement $b_{10} = 40$ is an outlier. Find the best *horizontal line* $y = C$ to fit the ten points $(1,0), (2,0), \ldots, (9,0), (10,40)$ using three measures for the error E:
(1) Least squares $E_2 = e_1^2 + \cdots + e_{10}^2$ (then the normal equation for C is linear)
(2) Least maximum error $E_\infty = |e_{\max}|$ (3) Least sum of errors $E_1 = |e_1| + \cdots + |e_{10}|$.

Solution (1) The least squares fit to $0, 0, \ldots, 0, 40$ by a horizontal line is $C = 4$:

$$A = \text{column of 1's} \quad A^{\mathrm{T}}A = 10 \quad A^{\mathrm{T}}b = \text{sum of } b_i = 40. \quad \text{So } 10C = 40.$$

(2) The least maximum error requires $C = 20$, halfway between 0 and 40.

(3) The least sum requires $C = 0$ (!!). The sum of errors $9|C| + |40 - C|$ would increase if C moves up from zero.

The least sum comes from the *median* measurement (the median of $0, \ldots, 0, 40$ is zero). Many statisticians feel that the least squares solution is too heavily influenced by outliers like $b_{10} = 40$, and they prefer least sum. But the equations become nonlinear.

Now find the least squares straight line $C + Dt$ through those ten points.

$$A^{\mathrm{T}}A = \begin{bmatrix} 10 & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} = \begin{bmatrix} 10 & 55 \\ 55 & 385 \end{bmatrix} \qquad A^{\mathrm{T}}b = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix} = \begin{bmatrix} 40 \\ 400 \end{bmatrix}$$

Those come from equation (8). Then $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$ gives $C = -8$ and $D = 24/11$.

What happens to C and D if you multiply the b_i by 3 and then add 30 to get $b_{\text{new}} = (30, 30, \ldots, 150)$? Linearity allows us to rescale $b = (0, 0, \ldots, 40)$. Multiplying b by 3 will multiply C and D by 3. Adding 30 to all b_i will add 30 to C.

4.3 B Find the parabola $C + Dt + Et^2$ that comes closest (least squares error) to the values $\boldsymbol{b} = (0, 0, 1, 0, 0)$ at the times $t = -2, -1, 0, 1, 2$. First write down the five equations $A\boldsymbol{x} = \boldsymbol{b}$ in three unknowns $\boldsymbol{x} = (C, D, E)$ for a parabola to go through the five points. No solution because no such parabola exists. Solve $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$.

I would predict $D = 0$. Why should the best parabola be symmetric around $t = 0$? In $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$, equation 2 for D should uncouple from equations 1 and 3.

Solution The five equations $A\boldsymbol{x} = \boldsymbol{b}$ have a rectangular "Vandermonde" matrix A:

$$\begin{array}{l} C + D\,(-2) + E\,(-2)^2 = 0 \\ C + D\,(-1) + E\,(-1)^2 = 0 \\ C + D\;\;(0) + E\;\;(0)^2 = 1 \\ C + D\;\;(1) + E\;\;(1)^2 = 0 \\ C + D\;\;(2) + E\;\;(2)^2 = 0 \end{array} \qquad A = \begin{bmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \qquad A^{\mathrm{T}}A = \begin{bmatrix} 5 & \mathbf{0} & 10 \\ \mathbf{0} & 10 & \mathbf{0} \\ 10 & \mathbf{0} & 34 \end{bmatrix}$$

Those zeros in $A^{\mathrm{T}}A$ mean that column 2 of A is orthogonal to columns 1 and 3. We see this directly in A (the times $-2, -1, 0, 1, 2$ are symmetric). The best C, D, E in the parabola $C + Dt + Et^2$ come from $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$, and D is uncoupled:

$$\begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{leads to} \quad \begin{array}{l} C = 34/70 \\ D = 0 \ \text{ as predicted} \\ E = -10/70 \end{array}$$

Problem Set 4.3

Problems 1–11 use four data points $\boldsymbol{b} = (0, 8, 8, 20)$ to bring out the key ideas.

1 With $b = 0, 8, 8, 20$ at $t = 0, 1, 3, 4$, set up and solve the normal equations $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$. For the best straight line in Figure 4.9a, find its four heights p_i and four errors e_i. What is the minimum value $E = e_1^2 + e_2^2 + e_3^2 + e_4^2$?

2 (Line $C + Dt$ does go through p's) With $b = 0, 8, 8, 20$ at times $t = 0, 1, 3, 4$, write down the four equations $A\boldsymbol{x} = \boldsymbol{b}$ (unsolvable). Change the measurements to $p = 1, 5, 13, 17$ and find an exact solution to $A\widehat{\boldsymbol{x}} = \boldsymbol{p}$.

3 Check that $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = (-1, 3, -5, 3)$ is perpendicular to both columns of the same matrix A. What is the shortest distance $\|\boldsymbol{e}\|$ from $\boldsymbol{b}$ to the column space of A?

4 (By calculus) Write down $E = \|A\boldsymbol{x} - \boldsymbol{b}\|^2$ as a sum of four squares—the last one is $(C + 4D - 20)^2$. Find the derivative equations $\partial E/\partial C = 0$ and $\partial E/\partial D = 0$. Divide by 2 to obtain the normal equations $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$.

5 Find the height C of the best *horizontal line* to fit $\boldsymbol{b} = (0, 8, 8, 20)$. An exact fit would solve the unsolvable equations $C = 0$, $C = 8$, $C = 8$, $C = 20$. Find the 4 by 1 matrix A in these equations and solve $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$. Draw the horizontal line at height $\widehat{x} = C$ and the four errors in $\boldsymbol{e}$.

6 Project $\boldsymbol{b} = (0, 8, 8, 20)$ onto the line through $\boldsymbol{a} = (1, 1, 1, 1)$. Find $\widehat{x} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ and the projection $\boldsymbol{p} = \widehat{x}\boldsymbol{a}$. Check that $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ is perpendicular to $\boldsymbol{a}$, and find the shortest distance $\|\boldsymbol{e}\|$ from $\boldsymbol{b}$ to the line through $\boldsymbol{a}$.

7 Find the closest line $b = Dt$, *through the origin*, to the same four points. An exact fit would solve $D \cdot 0 = 0, D \cdot 1 = 8, D \cdot 3 = 8, D \cdot 4 = 20$. Find the 4 by 1 matrix and solve $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}\boldsymbol{b}$. Redraw Figure 4.9a showing the best line $b = Dt$ and the e's.

8 Project $\boldsymbol{b} = (0, 8, 8, 20)$ onto the line through $\boldsymbol{a} = (0, 1, 3, 4)$. Find $\widehat{x} = D$ and $\boldsymbol{p} = \widehat{x}\boldsymbol{a}$. The best C in Problems 5–6 and the best D in Problems 7–8 do *not* agree with the best (C, D) in Problems 1–4. That is because $(1, 1, 1, 1)$ and $(0, 1, 3, 4)$ are _____ perpendicular.

9 For the closest parabola $b = C + Dt + Et^2$ to the same four points, write down the unsolvable equations $Ax = \boldsymbol{b}$ in three unknowns $\boldsymbol{x} = (C, D, E)$. Set up the three normal equations $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}\boldsymbol{b}$ (solution not required). In Figure 4.9a you are now fitting a parabola to 4 points—what is happening in Figure 4.9b?

10 For the closest cubic $b = C + Dt + Et^2 + Ft^3$ to the same four points, write down the four equations $Ax = \boldsymbol{b}$. Solve them by elimination. In Figure 4.9a this cubic now goes exactly through the points. What are $\boldsymbol{p}$ and $\boldsymbol{e}$?

11 The average of the four times is $\widehat{t} = \frac{1}{4}(0 + 1 + 3 + 4) = 2$. The average of the four b's is $\widehat{b} = \frac{1}{4}(0 + 8 + 8 + 20) = 9$.

(a) Verify that the best line goes through the center point $(\widehat{t}, \widehat{b}) = (2, 9)$.

(b) Explain why $C + D\widehat{t} = \widehat{b}$ comes from the first equation in $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}\boldsymbol{b}$.

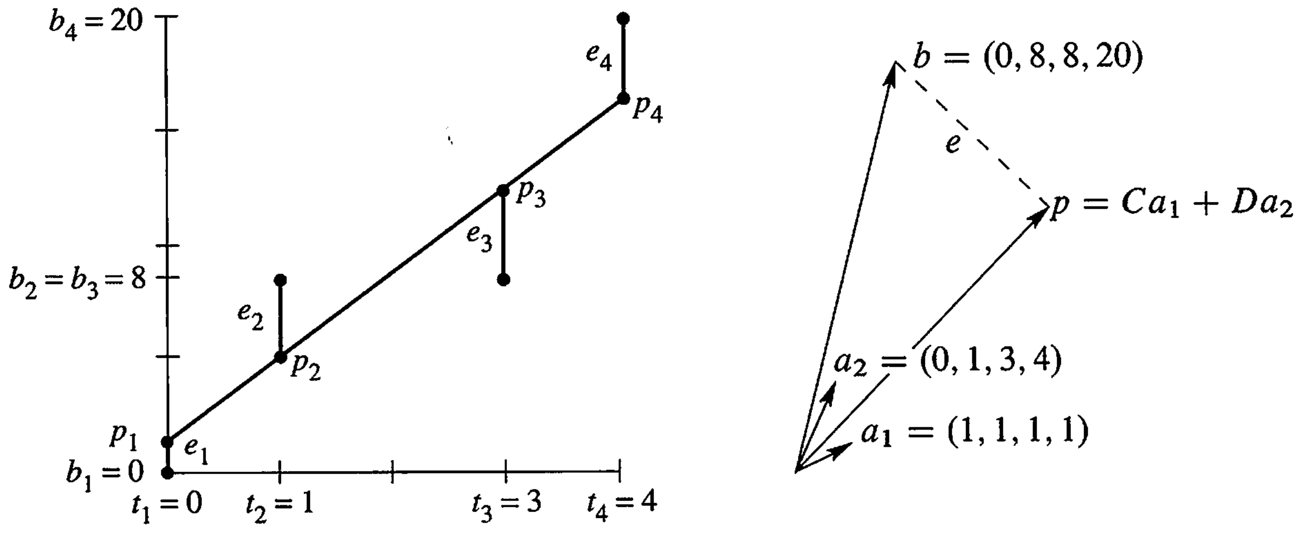


Figure 4.9: **Problems 1–11**: The closest line $C + Dt$ matches $C\boldsymbol{a}_1 + D\boldsymbol{a}_2$ in $\mathbf{R}^4$.

Questions 12–16 introduce basic ideas of statistics—the foundation for least squares.

12 (Recommended) This problem projects $\boldsymbol{b} = (b_1, \ldots, b_m)$ onto the line through $\boldsymbol{a} = (1, \ldots, 1)$. We solve m equations $\boldsymbol{a}x = \boldsymbol{b}$ in 1 unknown (by least squares).

(a) Solve $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}\widehat{x} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}$ to show that $\widehat{x}$ is the *mean* (the average) of the $\boldsymbol{b}$'s.

(b) Find $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{a}\widehat{x}$ and the *variance* $\|\boldsymbol{e}\|^2$ and the *standard deviation* $\|\boldsymbol{e}\|$.

(c) The horizontal line $\widehat{\boldsymbol{b}} = 3$ is closest to $\boldsymbol{b} = (1, 2, 6)$. Check that $\boldsymbol{p} = (3, 3, 3)$ is perpendicular to $\boldsymbol{e}$ and find the 3 by 3 projection matrix P.

13 First assumption behind least squares: $A\boldsymbol{x} = \boldsymbol{b}-$ ***(noise e with mean zero).*** Multiply the error vectors $\boldsymbol{e} = \boldsymbol{b} - A\boldsymbol{x}$ by $(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ to get $\widehat{\boldsymbol{x}} - \boldsymbol{x}$ on the right. The estimation errors $\widehat{\boldsymbol{x}} - \boldsymbol{x}$ also average to zero. The estimate $\widehat{\boldsymbol{x}}$ is *unbiased.*

14 Second assumption behind least squares: The m errors $\boldsymbol{e}_i$ are independent with variance σ^2, so the average of $(\boldsymbol{b} - A\boldsymbol{x})(\boldsymbol{b} - A\boldsymbol{x})^{\mathrm{T}}$ is $\sigma^2 I$. Multiply on the left by $(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ and on the right by $A(A^{\mathrm{T}}A)^{-1}$ to show that the average matrix $(\widehat{\boldsymbol{x}} - \boldsymbol{x})(\widehat{\boldsymbol{x}} - \boldsymbol{x})^{\mathrm{T}}$ is $\sigma^2(A^{\mathrm{T}}A)^{-1}$. This is the ***covariance matrix*** P in section 8.6.

15 A doctor takes 4 readings of your heart rate. The best solution to $x = b_1, \ldots, x = b_4$ is the average $\widehat{x}$ of $b_1, \ldots, b_4$. The matrix A is a column of 1's. Problem 14 gives the expected error $(\widehat{x} - x)^2$ as $\sigma^2(A^{\mathrm{T}}A)^{-1} =$ _____. ***By averaging, the variance drops from*** σ^2 ***to*** $\sigma^2/4$.

16 If you know the average $\widehat{x}_9$ of 9 numbers $b_1, \ldots, b_9$, how can you quickly find the average $\widehat{x}_{10}$ with one more number b_{10}? The idea of *recursive* least squares is to avoid adding 10 numbers. What number multiplies $\widehat{x}_9$ in computing $\widehat{x}_{10}$?

$$\widehat{x}_{10} = \tfrac{1}{10}b_{10} + \underline{\qquad}\,\widehat{x}_9 = \tfrac{1}{10}(b_1 + \cdots + b_{10}) \quad \text{as in Worked Example 4.2 C.}$$

Questions 17–24 give more practice with $\widehat{\boldsymbol{x}}$ and $\boldsymbol{p}$ and $\boldsymbol{e}$.

17 Write down three equations for the line $b = C + Dt$ to go through $b = 7$ at $t = -1$, $b = 7$ at $t = 1$, and $b = 21$ at $t = 2$. Find the least squares solution $\widehat{\boldsymbol{x}} = (C, D)$ and draw the closest line.

18 Find the projection $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$ in Problem 17. This gives the three heights of the closest line. Show that the error vector is $\boldsymbol{e} = (2, -6, 4)$. Why is $P\boldsymbol{e} = \boldsymbol{0}$?

19 Suppose the measurements at $t = -1, 1, 2$ are the errors $2, -6, 4$ in Problem 18. Compute $\widehat{\boldsymbol{x}}$ and the closest line to these new measurements. Explain the answer: $\boldsymbol{b} = (2, -6, 4)$ is perpendicular to _____ so the projection is $\boldsymbol{p} = \boldsymbol{0}$.

20 Suppose the measurements at $t = -1, 1, 2$ are $\boldsymbol{b} = (5, 13, 17)$. Compute $\widehat{\boldsymbol{x}}$ and the closest line and $\boldsymbol{e}$. The error is $\boldsymbol{e} = \boldsymbol{0}$ because this $\boldsymbol{b}$ is _____.

21 Which of the four subspaces contains the error vector $\boldsymbol{e}$? Which contains $\boldsymbol{p}$? Which contains $\widehat{\boldsymbol{x}}$? What is the nullspace of A?

22 Find the best line $C + Dt$ to fit $b = 4, 2, -1, 0, 0$ at times $t = -2, -1, 0, 1, 2$.

23 Is the error vector $\boldsymbol{e}$ orthogonal to $\boldsymbol{b}$ or $\boldsymbol{p}$ or $\boldsymbol{e}$ or $\widehat{\boldsymbol{x}}$? Show that $\|\boldsymbol{e}\|^2$ equals $\boldsymbol{e}^{\mathrm{T}}\boldsymbol{b}$ which equals $\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b} - \boldsymbol{p}^{\mathrm{T}}\boldsymbol{b}$. This is the smallest total error E.

24 The partial derivatives of $\|A\boldsymbol{x}\|^2$ with respect to $x_1, \ldots, x_n$ fill the vector $2A^{\mathrm{T}}A\boldsymbol{x}$. The derivatives of $2\boldsymbol{b}^{\mathrm{T}}A\boldsymbol{x}$ fill the vector $2A^{\mathrm{T}}\boldsymbol{b}$. So the derivatives of $\|A\boldsymbol{x} - \boldsymbol{b}\|^2$ are zero when _____.

Challenge Problems

25 *What condition on $(t_1, b_1), (t_2, b_2), (t_3, b_3)$ puts those three points onto a straight line?* A column space answer is: (b_1, b_2, b_3) must be a combination of $(1,\ 1,\ 1)$ and $(t_1,\ t_2,\ t_3)$. Try to reach a specific equation connecting the t's and b's. I should have thought of this question sooner!

26 Find the *plane* that gives the best fit to the 4 values $\boldsymbol{b} = (0, 1, 3, 4)$ at the corners $(1, 0)$ and $(0, 1)$ and $(-1, 0)$ and $(0, -1)$ of a square. The equations $C + Dx + Ey = b$ at those 4 points are $A\boldsymbol{x} = \boldsymbol{b}$ with 3 unknowns $\boldsymbol{x} = (C, D, E)$. What is A? At the center $(0, 0)$ of the square, show that $C + Dx + Ey =$ average of the b's.

27 (Distance between lines) The points $P = (x, x, x)$ and $Q = (y, 3y, -1)$ are on two lines in space that don't meet. Choose x and y to minimize the squared distance $\|P - Q\|^2$. The line connecting the closest P and Q is perpendicular to _____.

28 Suppose the columns of A are not independent. How could you find a matrix B so that $P = B(B^{\mathrm{T}}B)^{-1}B^{\mathrm{T}}$ does give the projection onto the column space of A? (The usual formula will fail when $A^{\mathrm{T}}A$ is not invertible.)

29 Usually there will be exactly one hyperplane in $\mathbf{R}^n$ that contains the n given points $\boldsymbol{x} = \mathbf{0}, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-1}$. (Example for $n = 3$: There will be one plane containing $\mathbf{0}, \boldsymbol{a}_1, \boldsymbol{a}_2$ unless _____.) What is the test to have exactly one plane in $\mathbf{R}^n$?

4.4 Orthogonal Bases and Gram-Schmidt

This section has two goals. The first is to see how orthogonality makes it easy to find $\widehat{x}$ and $\boldsymbol{p}$ and P. Dot products are zero—so $A^{\mathrm{T}}A$ becomes a diagonal matrix. ***The second goal is to construct orthogonal vectors.*** We will pick combinations of the original vectors to produce right angles. Those original vectors are the columns of A, probably *not* orthogonal. ***The orthogonal vectors will be the columns of a new matrix Q.***

From Chapter 3, a basis consists of independent vectors that span the space. The basis vectors could meet at any angle (except 0° and 180°). But every time we visualize axes, they are perpendicular. *In our imagination, the coordinate axes are practically always orthogonal.* This simplifies the picture and it greatly simplifies the computations.

The vectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ are ***orthogonal*** when their dot products $\boldsymbol{q}_i \cdot \boldsymbol{q}_j$ are zero. More exactly $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_j = 0$ whenever $i \neq j$. With one more step—just divide each vector by its length—the vectors become ***orthogonal unit vectors***. Their lengths are all 1. Then the basis is called ***orthonormal***.

DEFINITION The vectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ are ***orthonormal*** if

$$\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_j = \begin{cases} 0 & \text{when } i \neq j \quad (\textit{\textbf{orthogonal}} \text{ vectors}) \\ 1 & \text{when } i = j \quad (\textit{\textbf{unit}} \text{ vectors: } \|\boldsymbol{q}_i\| = 1) \end{cases}$$

A matrix with orthonormal columns is assigned the special letter Q.

The matrix Q is easy to work with because $Q^{\mathrm{T}}Q = I$. This repeats in matrix language that the columns $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ are orthonormal. Q is not required to be square.

A matrix Q with orthonormal columns satisfies $\boxed{Q^{\mathrm{T}}Q = I}$ **:**

$$\boxed{Q^{\mathrm{T}}Q} = \begin{bmatrix} - \boldsymbol{q}_1^{\mathrm{T}} - \\ - \boldsymbol{q}_2^{\mathrm{T}} - \\ \\ - \boldsymbol{q}_n^{\mathrm{T}} - \end{bmatrix} \begin{bmatrix} | & | & & | \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & & \boldsymbol{q}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \boxed{I}. \qquad (1)$$

When row i of Q^{T} multiplies column j of Q, the dot product is $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_j$. Off the diagonal $(i \neq j)$ that dot product is zero by orthogonality. On the diagonal $(i = j)$ the unit vectors give $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_i = \|\boldsymbol{q}_i\|^2 = 1$. Often Q is rectangular $(m > n)$. Sometimes $m = n$.

When Q is square, $Q^{\mathrm{T}}Q = I$ means that $Q^{\mathrm{T}} = Q^{-1}$: transpose = inverse.

If the columns are only orthogonal (not unit vectors), dot products still give a diagonal matrix (not the identity matrix). But this matrix is almost as good. The important thing is orthogonality—then it is easy to produce unit vectors.

To repeat: $Q^{\mathrm{T}}Q = I$ even when Q is rectangular. In that case Q^{T} is only an inverse from the left. For square matrices we also have $QQ^{\mathrm{T}} = I$, so Q^{T} is the two-sided inverse of Q. The rows of a square Q are orthonormal like the columns. ***The inverse is the transpose***. In this square case we call Q an ***orthogonal matrix***.[1]

Here are three examples of orthogonal matrices—rotation and permutation and reflection. The quickest test is to check $Q^{\mathrm{T}}Q = I$.

Example 1 **(Rotation)** Q rotates every vector in the plane clockwise by the angle θ:

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{and} \quad Q^{\mathrm{T}} = Q^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

The columns of Q are orthogonal (take their dot product). They are unit vectors because $\sin^2\theta + \cos^2\theta = 1$. Those columns give an ***orthonormal basis*** for the plane $\mathbf{R}^2$. The standard basis vectors $\boldsymbol{i}$ and $\boldsymbol{j}$ are rotated through θ (see Figure 4.10a). Q^{-1} rotates vectors back through $-\theta$. It agrees with Q^{T}, because the cosine of $-\theta$ is the cosine of θ, and $\sin(-\theta) = -\sin\theta$. We have $Q^{\mathrm{T}}Q = I$ and $QQ^{\mathrm{T}} = I$.

Example 2 **(Permutation)** These matrices change the order to (y, z, x) and (y, x):

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y \\ z \\ x \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}.$$

All columns of these Q's are unit vectors (their lengths are obviously 1). They are also orthogonal (the 1's appear in different places). *The inverse of a permutation matrix is its transpose*. The inverse puts the components back into their original order:

$$\textbf{Inverse = transpose:} \quad \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ z \\ x \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Every permutation matrix is an orthogonal matrix.

Example 3 **(Reflection)** If $\boldsymbol{u}$ is any unit vector, set $Q = I - 2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$. Notice that $\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ is a matrix while $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}$ is the number $\|\boldsymbol{u}\|^2 = 1$. Then Q^{T} and Q^{-1} both equal Q:

$$Q^{\mathrm{T}} = I - 2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}} = Q \quad \text{and} \quad Q^{\mathrm{T}}Q = I - 4\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}} + 4\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}} = I. \tag{2}$$

Reflection matrices $I - 2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ are symmetric and also orthogonal. If you square them, you get the identity matrix: $Q^2 = Q^{\mathrm{T}}Q = I$. Reflecting twice through a mirror brings back the original. Notice $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = 1$ inside $4\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ in equation (2).

[1]"Orthonormal matrix" would have been a better name for Q, but it's not used. Any matrix with orthonormal columns has the letter Q, but we only call it an *orthogonal matrix* when it is square.

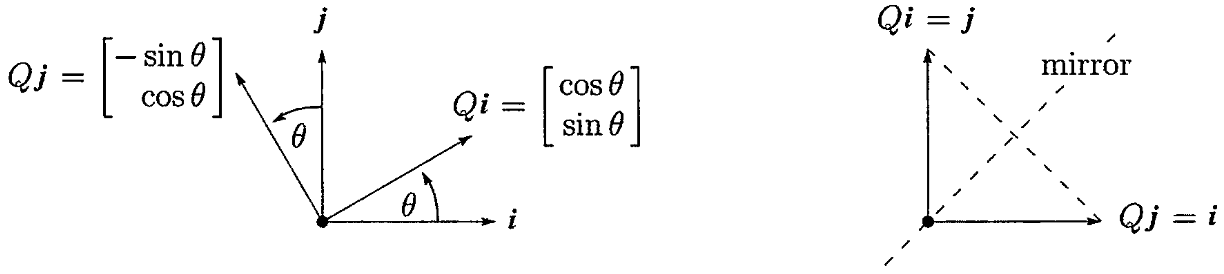


Figure 4.10: Rotation by $Q = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}$ and reflection across 45° by $Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

As examples choose two unit vectors, $\boldsymbol{u} = (1,0)$ and then $\boldsymbol{u} = (1/\sqrt{2}, -1/\sqrt{2})$. Compute $2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ (column times row) and subtract from I to get reflections Q_1 and Q_2:

$$Q_1 = I - 2\begin{bmatrix} 1 \\ 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad Q_2 = I - 2\begin{bmatrix} .5 & -.5 \\ -.5 & .5 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Q_1 reflects $(x,0)$ across the y axis to $(-x,0)$. Every vector (x,y) goes into its image $(-x,y)$, and the y axis is the mirror. Q_2 is reflection across the 45° line:

$$\textbf{Reflections} \quad \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -x \\ y \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}.$$

When (x,y) goes to (y,x), a vector like $(3,3)$ doesn't move. It is on the mirror line. Figure 4.10b shows the 45° mirror.

Rotations preserve the length of a vector. So do reflections. So do permutations. So does multiplication by any orthogonal matrix—***lengths and angles don't change***.

If Q has orthonormal columns ($Q^{\mathrm{T}}Q = I$), it leaves lengths unchanged:

Same length $\quad \|Q\boldsymbol{x}\| = \|\boldsymbol{x}\|$ **for every vector** $\boldsymbol{x}$. (3)

Q also preserves dot products: $(Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}}Q^{\mathrm{T}}Q\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$. Just use $Q^{\mathrm{T}}Q = I$!

Proof $\|Q\boldsymbol{x}\|^2$ equals $\|\boldsymbol{x}\|^2$ because $(Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}Q^{\mathrm{T}}Q\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}I\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$. Orthogonal matrices are excellent for computations—numbers can never grow too large when lengths of vectors are fixed. Stable computer codes use Q's as much as possible.

Projections Using Orthogonal Bases: Q Replaces A

This chapter is about projections onto subspaces. We developed the equations for $\widehat{\boldsymbol{x}}$ and $\boldsymbol{p}$ and the matrix P. When the columns of A were a basis for the subspace, all formulas involved $A^{\mathrm{T}}A$. The entries of $A^{\mathrm{T}}A$ are the dot products $\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{a}_j$.

Suppose the basis vectors are actually orthonormal. The $\boldsymbol{a}$'s become $\boldsymbol{q}$'s. Then $A^{\mathrm{T}}A$ simplifies to $Q^{\mathrm{T}}Q = I$. Look at the improvements in $\widehat{x}$ and p and P. Instead of $Q^{\mathrm{T}}Q$ we print a blank for the identity matrix:

$$\underline{\qquad}\,\widehat{x} = Q^{\mathrm{T}}\boldsymbol{b} \quad \text{and} \quad \boldsymbol{p} = Q\widehat{x} \quad \text{and} \quad P = Q\,\underline{\qquad}\,Q^{\mathrm{T}}. \tag{4}$$

The least squares solution of $Qx = b$ is $\widehat{x} = Q^{\mathrm{T}}b$. The projection matrix is $P = QQ^{\mathrm{T}}$.

There are no matrices to invert. This is the point of an orthonormal basis. The best $\widehat{x} = Q^{\mathrm{T}}\boldsymbol{b}$ just has dot products of $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ with $\boldsymbol{b}$. We have n 1-dimensional projections! The "coupling matrix" or "correlation matrix" $A^{\mathrm{T}}A$ is now $Q^{\mathrm{T}}Q = I$. There is no coupling. When A is Q, with orthonormal columns, here is $\boldsymbol{p} = Q\widehat{x} = QQ^{\mathrm{T}}\boldsymbol{b}$:

Projection onto q's

$$\boldsymbol{p} = \begin{bmatrix} | & & | \\ \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_n \\ | & & | \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b} \\ \vdots \\ \boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{b} \end{bmatrix} = \boldsymbol{q}_1(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b}) + \cdots + \boldsymbol{q}_n(\boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{b}). \tag{5}$$

Important case: When Q is square and $m = n$, the subspace is the whole space. Then $Q^{\mathrm{T}} = Q^{-1}$ and $\widehat{x} = Q^{\mathrm{T}}\boldsymbol{b}$ is the same as $\boldsymbol{x} = Q^{-1}\boldsymbol{b}$. The solution is exact! The projection of $\boldsymbol{b}$ onto the whole space is $\boldsymbol{b}$ itself. In this case $P = QQ^{\mathrm{T}} = I$.

You may think that projection onto the whole space is not worth mentioning. But when $\boldsymbol{p} = \boldsymbol{b}$, our formula assembles $\boldsymbol{b}$ out of its 1-dimensional projections. If $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ is an orthonormal basis for the whole space, so Q is square, then every $\boldsymbol{b} = QQ^{\mathrm{T}}\boldsymbol{b}$ is the sum of its components along the $\boldsymbol{q}$'s:

$$\boldsymbol{b} = \boldsymbol{q}_1(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b}) + \boldsymbol{q}_2(\boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b}) + \cdots + \boldsymbol{q}_n(\boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{b}). \tag{6}$$

That is $QQ^{\mathrm{T}} = I$. It is the foundation of Fourier series and all the great "transforms" of applied mathematics. They break vectors or functions into perpendicular pieces. Then by adding the pieces, the inverse transform puts the function back together.

Example 4 The columns of this orthogonal Q are orthonormal vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$:

$$Q = \frac{1}{3}\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix} \quad \text{has} \quad Q^{\mathrm{T}}Q = QQ^{\mathrm{T}} = I.$$

The separate projections of $\boldsymbol{b} = (0, 0, 1)$ onto $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ and $\boldsymbol{q}_3$ are $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ and $\boldsymbol{p}_3$:

$$\boldsymbol{q}_1(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b}) = \tfrac{2}{3}\boldsymbol{q}_1 \quad \text{and} \quad \boldsymbol{q}_2(\boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b}) = \tfrac{2}{3}\boldsymbol{q}_2 \quad \text{and} \quad \boldsymbol{q}_3(\boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{b}) = -\tfrac{1}{3}\boldsymbol{q}_3.$$

The sum of the first two is the projection of $\boldsymbol{b}$ onto the *plane* of $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$. The sum of all three is the projection of $\boldsymbol{b}$ onto the *whole space*—which is $\boldsymbol{b}$ itself:

Reconstruct $\boldsymbol{b} = \boldsymbol{p}_1 + \boldsymbol{p}_2 + \boldsymbol{p}_3$

$$\tfrac{2}{3}\boldsymbol{q}_1 + \tfrac{2}{3}\boldsymbol{q}_2 - \tfrac{1}{3}\boldsymbol{q}_3 = \tfrac{1}{9}\begin{bmatrix} -2+4-2 \\ 4-2-2 \\ 4+4+1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \boldsymbol{b}.$$

The Gram-Schmidt Process

The point of this section is that "orthogonal is good." Projections and least squares always involve $A^{\mathrm{T}}A$. When this matrix becomes $Q^{\mathrm{T}}Q = I$, the inverse is no problem. The one-dimensional projections are uncoupled. The best $\widehat{x}$ is $Q^{\mathrm{T}}\boldsymbol{b}$ (just n separate dot products). For this to be true, we had to say "*If* the vectors are orthonormal". ***Now we find a way to create orthonormal vectors.***

Start with three independent vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$. We intend to construct three orthogonal vectors $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$. Then (at the end is easiest) we divide $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ by their lengths. That produces three orthonormal vectors $\boldsymbol{q}_1 = \boldsymbol{A}/\|\boldsymbol{A}\|$, $\boldsymbol{q}_2 = \boldsymbol{B}/\|\boldsymbol{B}\|$, $\boldsymbol{q}_3 = \boldsymbol{C}/\|\boldsymbol{C}\|$.

Gram-Schmidt Begin by choosing $\boldsymbol{A} = \boldsymbol{a}$. This first direction is accepted. The next direction $\boldsymbol{B}$ must be perpendicular to $\boldsymbol{A}$. ***Start with b and subtract its projection along A.*** This leaves the perpendicular part, which is the orthogonal vector $\boldsymbol{B}$:

First Gram-Schmidt step $$\boldsymbol{B} = \boldsymbol{b} - \frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}}\boldsymbol{A}. \tag{7}$$

$\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal in Figure 4.11. Take the dot product with $\boldsymbol{A}$ to verify that $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = 0$. This vector $\boldsymbol{B}$ is what we have called the error vector $\boldsymbol{e}$, perpendicular to $\boldsymbol{A}$. Notice that $\boldsymbol{B}$ in equation (7) is not zero (otherwise $\boldsymbol{a}$ and $\boldsymbol{b}$ would be dependent). The directions $\boldsymbol{A}$ and $\boldsymbol{B}$ are now set.

The third direction starts with $\boldsymbol{c}$. This is not a combination of $\boldsymbol{A}$ and $\boldsymbol{B}$ (because $\boldsymbol{c}$ is not a combination of $\boldsymbol{a}$ and $\boldsymbol{b}$). But most likely $\boldsymbol{c}$ is not perpendicular to $\boldsymbol{A}$ and $\boldsymbol{B}$. So subtract off its components in those two directions to get $\boldsymbol{C}$:

Next Gram-Schmidt step $$\boldsymbol{C} = \boldsymbol{c} - \frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c}}{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}}\boldsymbol{A} - \frac{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{c}}{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}}\boldsymbol{B}. \tag{8}$$

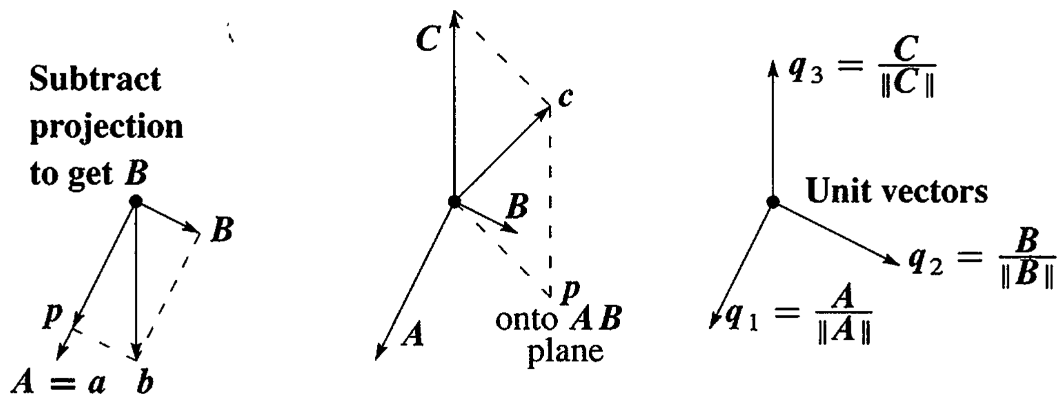


Figure 4.11: First project $\boldsymbol{b}$ onto the line through $\boldsymbol{a}$ and find the orthogonal $\boldsymbol{B}$ as $\boldsymbol{b} - \boldsymbol{p}$. Then project $\boldsymbol{c}$ onto the $\boldsymbol{AB}$ plane and find $\boldsymbol{C}$ as $\boldsymbol{c} - \boldsymbol{p}$. Divide by $\|\boldsymbol{A}\|$, $\|\boldsymbol{B}\|$, $\|\boldsymbol{C}\|$.

This is the one and only idea of the Gram-Schmidt process. ***Subtract from every new vector its projections in the directions already set.*** That idea is repeated at every step.[2] If we had a fourth vector $\boldsymbol{d}$, we would subtract three projections onto $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ to get $\boldsymbol{D}$. At the end, *or immediately when each one is found*, divide the orthogonal vectors $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ by their lengths. The resulting vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3, \boldsymbol{q}_4$ are orthonormal.

Example 5 Suppose the independent non-orthogonal vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ are

$$\boldsymbol{a} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{c} = \begin{bmatrix} 3 \\ -3 \\ 3 \end{bmatrix}.$$

Then $\boldsymbol{A} = \boldsymbol{a}$ has $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = 2$. Subtract from $\boldsymbol{b}$ its projection along $\boldsymbol{A} = (1, -1, 0)$:

$$\textbf{First step} \qquad \boldsymbol{B} = \boldsymbol{b} - \frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}}\boldsymbol{A} = \boldsymbol{b} - \tfrac{2}{2}\boldsymbol{A} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}.$$

Check: $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B} = 0$ as required. Now subtract two projections from $\boldsymbol{c}$ to get $\boldsymbol{C}$:

$$\textbf{Next step} \qquad \boldsymbol{C} = \boldsymbol{c} - \frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c}}{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}}\boldsymbol{A} - \frac{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{c}}{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}}\boldsymbol{B} = \boldsymbol{c} - \tfrac{6}{2}\boldsymbol{A} + \tfrac{6}{6}\boldsymbol{B} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Check: $\boldsymbol{C} = (1, 1, 1)$ is perpendicular to $\boldsymbol{A}$ and $\boldsymbol{B}$. Finally convert $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ to unit vectors (length 1, orthonormal). The lengths of $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ are $\sqrt{2}$ and $\sqrt{6}$ and $\sqrt{3}$. Divide by those lengths, for an orthonormal basis:

$$\boldsymbol{q}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{q}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{q}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Usually $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ contain fractions. Almost always $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ contain square roots.

The Factorization $A = QR$

We started with a matrix A, whose columns were $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$. We ended with a matrix Q, whose columns are $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$. How are those matrices related? Since the vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ are combinations of the $\boldsymbol{q}$'s (and vice versa), there must be a third matrix connecting A to Q. This third matrix is the triangular R in $A = QR$.

The first step was $\boldsymbol{q}_1 = \boldsymbol{a}/\|\boldsymbol{a}\|$ (other vectors not involved). The second step was equation (7), where $\boldsymbol{b}$ is a combination of $\boldsymbol{A}$ and $\boldsymbol{B}$. At that stage $\boldsymbol{C}$ and $\boldsymbol{q}_3$ were not involved. This non-involvement of later vectors is the key point of Gram-Schmidt:

[2]I think Gram had the idea. I don't really know where Schmidt came in.

- The vectors $\boldsymbol{a}$ and A and $\boldsymbol{q}_1$ are all along a single line.
- The vectors $\boldsymbol{a}, \boldsymbol{b}$ and A, B and $\boldsymbol{q}_1, \boldsymbol{q}_2$ are all in the same plane.
- The vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ and A, B, C and $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ are in one subspace (dimension 3).

At every step $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ are combinations of $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_k$. Later $\boldsymbol{q}$'s are not involved. The connecting matrix R is ***triangular***, and we have $A = QR$:

$$\begin{bmatrix} \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{c} \\ & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{c} \\ & & \boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{c} \end{bmatrix} \quad \text{or} \quad \boxed{A = QR.} \tag{9}$$

$A = QR$ is Gram-Schmidt in a nutshell. Multiply by Q^{T} to see why $R = Q^{\mathrm{T}}A$.

(Gram-Schmidt) From independent vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, Gram-Schmidt constructs orthonormal vectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$. The matrices with these columns satisfy $A = QR$. Then $R = Q^{\mathrm{T}}A$ is ***upper triangular*** because later $\boldsymbol{q}$'s are orthogonal to earlier $\boldsymbol{a}$'s.

Here are the $\boldsymbol{a}$'s and $\boldsymbol{q}$'s from the example. The i, j entry of $R = Q^{\mathrm{T}}A$ is row i of Q^{T} times column j of A. This is the dot product of $\boldsymbol{q}_i$ with $\boldsymbol{a}_j$:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & -3 \\ 0 & -2 & 3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & -2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{18} \\ \mathbf{0} & \sqrt{6} & -\sqrt{6} \\ \mathbf{0} & \mathbf{0} & \sqrt{3} \end{bmatrix} = QR.$$

The lengths of $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ are the numbers $\sqrt{2}$, $\sqrt{6}$, $\sqrt{3}$ on the diagonal of R. Because of the square roots, QR looks less beautiful than LU. Both factorizations are absolutely central to calculations in linear algebra.

Any m by n matrix A with independent columns can be factored into QR. The m by n matrix Q has orthonormal columns, and the square matrix R is upper triangular with positive diagonal. We must not forget why this is useful for least squares: $\boldsymbol{A^{\mathrm{T}}A}$ ***equals*** $\boldsymbol{R^{\mathrm{T}}Q^{\mathrm{T}}QR = R^{\mathrm{T}}R}$. The least squares equation $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}\boldsymbol{b}$ simplifies to $Rx = Q^{\mathrm{T}}\boldsymbol{b}$:

$$\textbf{Least squares} \qquad R^{\mathrm{T}}R\widehat{x} = R^{\mathrm{T}}Q^{\mathrm{T}}\boldsymbol{b} \ \text{ or } \ R\widehat{x} = Q^{\mathrm{T}}\boldsymbol{b} \ \text{ or } \ \boxed{\widehat{x} = R^{-1}Q^{\mathrm{T}}\boldsymbol{b}} \tag{10}$$

Instead of solving $Ax = \boldsymbol{b}$, which is impossible, we solve $R\widehat{x} = Q^{\mathrm{T}}\boldsymbol{b}$ by back substitution—which is very fast. The real cost is the mn^2 multiplications in the Gram-Schmidt process, which are needed to construct the orthogonal Q and the triangular R.

Below is an informal code. It executes equations (11) and (12), for $k = 1$ then $k = 2$ and eventually $k = n$. The last line of that code normalizes to unit vectors $\boldsymbol{q}_j$:

$$\begin{array}{l}\textbf{Divide by length} \\ \boldsymbol{q_j} \textbf{ = unit vector}\end{array} \qquad r_{jj} = \left(\sum_{i=1}^{m} v_{ij}^2\right)^{1/2} \quad \textbf{and} \quad q_{ij} = \frac{v_{ij}}{r_{jj}} \quad \textbf{for} \quad i = 1, \ldots, m. \tag{11}$$

The important lines subtract from $v = a_j$ its projection onto each q_i:

$$r_{kj} = \sum_{i=1}^{m} q_{ik} v_{ij} \qquad \text{and} \qquad v_{ij} = v_{ij} - q_{ik} r_{kj}. \tag{12}$$

Starting from $a, b, c = a_1, a_2, a_3$ this code will construct q_1, B, q_2, C, q_3:

$$q_1 = a_1/\|a_1\| \qquad B = a_2 - (q_1^{\mathrm{T}} a_2) q_1 \qquad q_2 = B/\|B\|$$

$$C^* = a_3 - (q_1^{\mathrm{T}} a_3) q_1 \quad C = C^* - (q_2^{\mathrm{T}} C^*) q_2 \quad q_3 = C/\|C\|$$

Equation (12) subtracts off projections as soon as the new vector q_k is found. This change to "subtract one projection at a time" is called ***modified Gram-Schmidt.*** That is numerically more stable than equation (8) which subtracts all projections at once.

```
for j = 1:n                      % modified Gram-Schmidt
  v = A(:, j);                   % v begins as column j of A
  for i = 1:j-1                  % columns up to j-1, already settled in Q
    R(i, j) = Q(:, i)'*v;        % compute R_ij = q_i^T a_j which is q_i^T v
    v = v-R(i, j)*Q(:, i);       % subtract the projection (q_i^T a_j)q_i = (q_i^T v)q_i
  end                            % v is now perpendicular to all of q_1, ..., q_(j-1)
  R(j, j) = norm(v);             % diagonal entries of R
  Q(:, j) = v/R(j, j);           % normalize v to be the next unit vector q_j
end
```

To recover column j of A, undo the last step and the middle steps of the code:

$$R(j,j)q_j = (v \textbf{ \textit{minus its projections}}) = (\text{column } j \text{ of } A) - \sum_{i=1}^{j-1} R(i,j) q_i\,. \tag{13}$$

Moving the sum to the far left, this is column j in the multiplication $A = QR$.

Confession Good software like LAPACK, used in good systems like MATLAB and Octave and Python, will not use this Gram-Schmidt code. There is now a better way. "Householder reflections" produce the upper triangular R, one column at a time, exactly as elimination produces the upper triangular U.

Those reflection matrices $I - 2uu^{\mathrm{T}}$ will be described in Chapter 9 on numerical linear algebra. If A is tridiagonal we can simplify even more to use 2 by 2 rotations. The result is always $A = QR$ and the MATLAB command is $[Q, R] = \mathrm{qr}(A)$. I believe that Gram-Schmidt is still the good process to understand, even if the reflections or rotations lead to a more perfect Q.

■ REVIEW OF THE KEY IDEAS ■

1. If the orthonormal vectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ are the columns of Q, then $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_j = 0$ and $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_i = 1$ translate into $Q^{\mathrm{T}}Q = I$.

2. If Q is square (an ***orthogonal matrix***) then $Q^{\mathrm{T}} = Q^{-1}$: ***transpose = inverse***.

3. The length of $Q\boldsymbol{x}$ equals the length of $\boldsymbol{x}$: $\|Q\boldsymbol{x}\| = \|\boldsymbol{x}\|$.

4. The projection onto the column space spanned by the $\boldsymbol{q}$'s is $P = QQ^{\mathrm{T}}$.

5. If Q is square then $P = I$ and every $\boldsymbol{b} = \boldsymbol{q}_1(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b}) + \cdots + \boldsymbol{q}_n(\boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{b})$.

6. Gram-Schmidt produces orthonormal vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ from independent $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$. In matrix form this is the factorization $A = QR =$ (orthogonal Q)(triangular R).

■ WORKED EXAMPLES ■

4.4 A Add two more columns with all entries 1 or -1, so the columns of this 4 by 4 "Hadamard matrix" are orthogonal. How do you turn H_4 into an *orthogonal matrix* Q?

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad H_4 = \begin{bmatrix} 1 & 1 & x & x \\ 1 & -1 & x & x \\ 1 & 1 & x & x \\ 1 & -1 & x & x \end{bmatrix} \quad \text{and} \quad Q_4 = \begin{bmatrix} & & & \\ & & & \\ & & & \\ & & & \end{bmatrix}$$

The block matrix $H_8 = \begin{bmatrix} H_4 & H_4 \\ H_4 & -H_4 \end{bmatrix}$ is the next Hadamard matrix with 1's and -1's. What is the product $H_8^{\mathrm{T}}H_8$?

The projection of $\boldsymbol{b} = (6, 0, 0, 2)$ onto the first column of H_4 is $\boldsymbol{p}_1 = (2, 2, 2, 2)$. The projection onto the second column is $\boldsymbol{p}_2 = (1, -1, 1, -1)$. What is the projection $\boldsymbol{p}_{1,2}$ of $\boldsymbol{b}$ onto the 2-dimensional space spanned by the first two columns?

Solution H_4 can be built from H_2 just as H_8 is built from H_4:

$$H_4 = \begin{bmatrix} H_2 & H_2 \\ H_2 & -H_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \text{ has orthogonal columns.}$$

Then $Q = H/2$ has orthonormal columns. Dividing by 2 gives unit vectors in Q. Orthogonality for 5 by 5 is impossible because the dot product of columns would have five 1's and/or -1's and could not add to zero. H_8 has orthogonal columns of length $\sqrt{8}$.

$$H_8^{\mathrm{T}} H_8 = \begin{bmatrix} H^{\mathrm{T}} & H^{\mathrm{T}} \\ H^{\mathrm{T}} & -H^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} H & H \\ H & -H \end{bmatrix} = \begin{bmatrix} 2H^{\mathrm{T}}H & 0 \\ 0 & 2H^{\mathrm{T}}H \end{bmatrix} = \begin{bmatrix} 8I & 0 \\ 0 & 8I \end{bmatrix}. \quad Q_8 = \frac{H_8}{\sqrt{8}}$$

Key point of orthogonal columns: We can project $(6,0,0,2)$ onto $(1,1,1,1)$ and $(1,-1,1,-1)$ and **add**. There is no $A^{\mathrm{T}}A$ matrix to invert:

Add p's Projection $p_{1,2} = p_1 + p_2 = (2,2,2,2) + (1,-1,1,-1) = (3,1,3,1)$.

Check that columns a_1 and a_2 of H are perpendicular to the error $e = b - p_1 - p_2$:

$$e = b - \frac{a_1^{\mathrm{T}} b}{a_1^{\mathrm{T}} a_1} a_1 - \frac{a_2^{\mathrm{T}} b}{a_2^{\mathrm{T}} a_2} a_2 \quad \text{and} \quad a_1^{\mathrm{T}} e = a_1^{\mathrm{T}} b - \frac{a_1^{\mathrm{T}} b}{a_1^{\mathrm{T}} a_1} a_1^{\mathrm{T}} a_1 = 0 \quad \text{and also} \quad a_2^{\mathrm{T}} e = 0.$$

So $p_1 + p_2$ is in the space of a_1 and a_2, and *its error e is perpendicular to that space.*

The Gram-Schmidt process on those orthogonal columns a_1 and a_2 would not change their directions. It would only divide by their lengths. *But if a_1 and a_2 are not orthogonal, the projection $p_{1,2}$ is not generally $p_1 + p_2$.* For example, if b is the same as a_1, then $p_1 = b$ and $p_{1,2} = b$ but $p_2 \neq \mathbf{0}$.

Problem Set 4.4

Problems 1–12 are about orthogonal vectors and orthogonal matrices.

1 Are these pairs of vectors orthonormal or only orthogonal or only independent?

(a) $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ (b) $\begin{bmatrix} .6 \\ .8 \end{bmatrix}$ and $\begin{bmatrix} .4 \\ -.3 \end{bmatrix}$ (c) $\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$ and $\begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$.

Change the second vector when necessary to produce orthonormal vectors.

2 The vectors $(2,2,-1)$ and $(-1,2,2)$ are orthogonal. Divide them by their lengths to find orthonormal vectors q_1 and q_2. Put those into the columns of Q and multiply $Q^{\mathrm{T}}Q$ and QQ^{T}.

3 (a) If A has three orthogonal columns each of length 4, what is $A^{\mathrm{T}}A$?

(b) If A has three orthogonal columns of lengths $1,2,3$, what is $A^{\mathrm{T}}A$?

4 Give an example of each of the following:

(a) A matrix Q that has orthonormal columns but $QQ^{\mathrm{T}} \neq I$.

(b) Two orthogonal vectors that are not linearly independent.

(c) An orthonormal basis for $\mathbf{R}^3$, including the vector $q_1 = (1,1,1)/\sqrt{3}$.

5 Find two orthogonal vectors in the plane $x + y + 2z = 0$. Make them orthonormal.

6 If Q_1 and Q_2 are orthogonal matrices, show that their product Q_1Q_2 is also an orthogonal matrix. (Use $Q^TQ = I$.)

7 If Q has orthonormal columns, what is the least squares solution $\widehat{x}$ to $Qx = b$?

8 If q_1 and q_2 are orthonormal vectors in $\mathbf{R}^5$, what combination ____ $q_1 +$ ____ q_2 is closest to a given vector b?

9 (a) Compute $P = QQ^T$ when $q_1 = (.8, .6, 0)$ and $q_2 = (-.6, .8, 0)$. Verify that $P^2 = P$.

(b) Prove that always $(QQ^T)^2 = QQ^T$ by using $Q^TQ = I$. Then $P = QQ^T$ is the projection matrix onto the column space of Q.

10 Orthonormal vectors are automatically linearly independent.

(a) Vector proof: When $c_1q_1 + c_2q_2 + c_3q_3 = \mathbf{0}$, what dot product leads to $c_1 = 0$? Similarly $c_2 = 0$ and $c_3 = 0$. Thus the q's are independent.

(b) Matrix proof: Show that $Qx = \mathbf{0}$ leads to $x = \mathbf{0}$. Since Q may be rectangular, you can use Q^T but not Q^{-1}.

11 (a) Gram-Schmidt: Find orthonormal vectors q_1 and q_2 in the plane spanned by $a = (1, 3, 4, 5, 7)$ and $b = (-6, 6, 8, 0, 8)$.

(b) Which vector in this plane is closest to $(1, 0, 0, 0, 0)$?

12 If a_1, a_2, a_3 is a basis for $\mathbf{R}^3$, any vector b can be written as

$$b = x_1a_1 + x_2a_2 + x_3a_3 \qquad \text{or} \qquad \begin{bmatrix} & & \\ a_1 & a_2 & a_3 \\ & & \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = b.$$

(a) Suppose the a's are orthonormal. Show that $x_1 = a_1^Tb$.

(b) Suppose the a's are orthogonal. Show that $x_1 = a_1^Tb/a_1^Ta_1$.

(c) If the a's are independent, x_1 is the first component of ____ times b.

Problems 13–25 are about the Gram-Schmidt process and $A = QR$.

13 What multiple of $a = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ should be subtracted from $b = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$ to make the result B orthogonal to a? Sketch a figure to show a, b, and B.

14 Complete the Gram-Schmidt process in Problem 13 by computing $q_1 = a/\|a\|$ and $q_2 = B/\|B\|$ and factoring into QR:

$$\begin{bmatrix} 1 & 4 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} q_1 & q_2 \end{bmatrix}\begin{bmatrix} \|a\| & ? \\ 0 & \|B\| \end{bmatrix}.$$

15 (a) Find orthonormal vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ such that $\boldsymbol{q}_1, \boldsymbol{q}_2$ span the column space of

$$A = \begin{bmatrix} 1 & 1 \\ 2 & -1 \\ -2 & 4 \end{bmatrix}.$$

(b) Which of the four fundamental subspaces contains $\boldsymbol{q}_3$?

(c) Solve $A\boldsymbol{x} = (1, 2, 7)$ by least squares.

16 What multiple of $\boldsymbol{a} = (4, 5, 2, 2)$ is closest to $\boldsymbol{b} = (1, 2, 0, 0)$? Find orthonormal vectors $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ in the plane of $\boldsymbol{a}$ and $\boldsymbol{b}$.

17 Find the projection of $\boldsymbol{b}$ onto the line through $\boldsymbol{a}$:

$$\boldsymbol{a} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} \quad \text{and} \quad \boldsymbol{p} = ? \quad \text{and} \quad \boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = ?$$

Compute the orthonormal vectors $\boldsymbol{q}_1 = \boldsymbol{a}/\|\boldsymbol{a}\|$ and $\boldsymbol{q}_2 = \boldsymbol{e}/\|\boldsymbol{e}\|$.

18 (Recommended) Find orthogonal vectors $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ by Gram-Schmidt from $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$:

$$\boldsymbol{a} = (1, -1, 0, 0) \qquad \boldsymbol{b} = (0, 1, -1, 0) \qquad \boldsymbol{c} = (0, 0, 1, -1).$$

$\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ and $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ are bases for the vectors perpendicular to $\boldsymbol{d} = (1, 1, 1, 1)$.

19 If $A = QR$ then $A^{\mathrm{T}}A = R^{\mathrm{T}}R =$ _____ triangular times _____ triangular. *Gram-Schmidt on A corresponds to elimination on $A^{\mathrm{T}}A$.* The pivots for $A^{\mathrm{T}}A$ must be the squares of diagonal entries of R. Find Q and R by Gram-Schmidt for this A:

$$A = \begin{bmatrix} -1 & 1 \\ 2 & 1 \\ 2 & 4 \end{bmatrix} \quad \text{and} \quad A^{\mathrm{T}}A = \begin{bmatrix} 9 & 9 \\ 9 & 18 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 9 & \\ & 9 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

20 True or false (give an example in either case):

(a) Q^{-1} is an orthogonal matrix when Q is an orthogonal matrix.

(b) If Q (3 by 2) has orthonormal columns then $\|Q\boldsymbol{x}\|$ always equals $\|\boldsymbol{x}\|$.

21 Find an orthonormal basis for the column space of A:

$$A = \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 1 & 1 \\ 1 & 3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} -4 \\ -3 \\ 3 \\ 0 \end{bmatrix}.$$

Then compute the projection of $\boldsymbol{b}$ onto that column space.

22 Find orthogonal vectors $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ by Gram-Schmidt from

$$\boldsymbol{a} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{c} = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}.$$

23 Find $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ (orthonormal) as combinations of $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ (independent columns). Then write A as QR:

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{bmatrix}.$$

24 (a) Find a basis for the subspace $\boldsymbol{S}$ in $\mathbf{R}^4$ spanned by all solutions of

$$x_1 + x_2 + x_3 - x_4 = 0.$$

(b) Find a basis for the orthogonal complement $\boldsymbol{S}^{\perp}$.

(c) Find $\boldsymbol{b}_1$ in $\boldsymbol{S}$ and $\boldsymbol{b}_2$ in $\boldsymbol{S}^{\perp}$ so that $\boldsymbol{b}_1 + \boldsymbol{b}_2 = \boldsymbol{b} = (1, 1, 1, 1)$.

25 If $ad - bc > 0$, the entries in $A = QR$ are

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{\begin{bmatrix} a & -c \\ c & a \end{bmatrix}}{\sqrt{a^2 + c^2}} \frac{\begin{bmatrix} a^2 + c^2 & ab + cd \\ 0 & ad - bc \end{bmatrix}}{\sqrt{a^2 + c^2}}.$$

Write $A = QR$ when $a, b, c, d = 2, 1, 1, 1$ and also $1, 1, 1, 1$. Which entry of R becomes zero when the columns are dependent and Gram-Schmidt breaks down?

Problems 26–29 use the QR code in equations (11–12). It executes Gram-Schmidt.

26 Show why $\boldsymbol{C}$ (found via $\boldsymbol{C}^*$ in the steps after (12)) is equal to $\boldsymbol{C}$ in equation (8).

27 Equation (8) subtracts from $\boldsymbol{c}$ its components along $\boldsymbol{A}$ and $\boldsymbol{B}$. Why not subtract the components along $\boldsymbol{a}$ and along $\boldsymbol{b}$?

28 Where are the mn^2 multiplications in equations (11) and (12)?

29 Apply the MATLAB qr code to $\boldsymbol{a} = (2, 2, -1)$, $\boldsymbol{b} = (0, -3, 3)$, $\boldsymbol{c} = (1, 0, 0)$. What are the $\boldsymbol{q}$'s?

Problems 30–35 involve orthogonal matrices that are special.

30 The first four ***wavelets*** are in the columns of this wavelet matrix W:

$$W = \frac{1}{2} \begin{bmatrix} 1 & 1 & \sqrt{2} & 0 \\ 1 & 1 & -\sqrt{2} & 0 \\ 1 & -1 & 0 & \sqrt{2} \\ 1 & -1 & 0 & -\sqrt{2} \end{bmatrix}.$$

What is special about the columns? Find the inverse wavelet transform W^{-1}.

31 (a) Choose c so that Q is an orthogonal matrix:

$$Q = c\begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}.$$

Project $\boldsymbol{b} = (1, 1, 1, 1)$ onto the first column. Then project $\boldsymbol{b}$ onto the plane of the first two columns.

32 If $\boldsymbol{u}$ is a unit vector, then $Q = I - 2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ is a reflection matrix (Example 3). Find Q_1 from $\boldsymbol{u} = (0, 1)$ and Q_2 from $\boldsymbol{u} = (0, \sqrt{2}/2, \sqrt{2}/2)$. Draw the reflections when Q_1 and Q_2 multiply the vectors $(1, 2)$ and $(1, 1, 1)$.

33 Find all matrices that are both orthogonal and lower triangular.

34 $Q = I - 2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ is a reflection matrix when $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = 1$. Two reflections give $Q^2 = I$.

(a) Show that $Q\boldsymbol{u} = -\boldsymbol{u}$. The mirror is perpendicular to $\boldsymbol{u}$.

(b) Find $Q\boldsymbol{v}$ when $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} = 0$. The mirror contains $\boldsymbol{v}$. It reflects to itself.

Challenge Problems

35 (MATLAB) Factor $[\,Q, R\,] = \mathbf{qr}(A)$ for $A = \mathbf{eye}(4) - \mathbf{diag}([1\;\;1\;\;1], -1)$. You are orthogonalizing the columns $(1, -1, 0, 0)$ and $(0, 1, -1, 0)$ and $(0, 0, 1, -1)$ and $(0, 0, 0, 1)$ of A. Can you scale the orthogonal columns of Q to get nice integer components?

36 If A is m by n with rank n, $\mathbf{qr}(A)$ produces a *square* Q and zeros below R:

The factors from MATLAB are $(m \text{ by } m)(m \text{ by } n)$ $\qquad A = [Q_1 \quad Q_2]\begin{bmatrix} R \\ 0 \end{bmatrix}.$

The n columns of Q_1 are an orthonormal basis for which fundamental subspace? The $m-n$ columns of Q_2 are an orthonormal basis for which fundamental subspace?

37 We know that $P = QQ^{\mathrm{T}}$ is the projection onto the column space of $Q(m \text{ by } n)$. Now add another column $\boldsymbol{a}$ to produce $A = [Q \quad \boldsymbol{a}]$. What is the new orthonormal vector $\boldsymbol{q}$ from Gram-Schmidt: start with $\boldsymbol{a}$, subtract _____, divide by _____.

Chapter 5

Determinants

5.1 The Properties of Determinants

The determinant of a square matrix is a single number. That number contains an amazing amount of information about the matrix. It tells immediately whether the matrix is invertible. ***The determinant is zero when the matrix has no inverse.*** When A is invertible, the determinant of A^{-1} is $1/(\det A)$. If $\det A = 2$ then $\det A^{-1} = \frac{1}{2}$. In fact the determinant leads to a formula for every entry in A^{-1}.

This is one use for determinants—to find formulas for inverse matrices and pivots and solutions $A^{-1}\boldsymbol{b}$. For a large matrix we seldom use those formulas, because elimination is faster. For a 2 by 2 matrix with entries a, b, c, d, its determinant $ad - bc$ shows how A^{-1} changes as A changes:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{has inverse} \quad A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{1}$$

Multiply those matrices to get I. When the determinant is $ad - bc = 0$, we are asked to divide by zero and we can't—then A has no inverse. (The rows are parallel when $a/c = b/d$. This gives $ad = bc$ and $\det A = 0$). Dependent rows always lead to $\det A = 0$.

The determinant is also connected to the pivots. For a 2 by 2 matrix the pivots are a and $d - (c/a)b$. ***The product of the pivots is the determinant***:

$$\textbf{Product of pivots} \qquad a\left(d - \frac{c}{a}b\right) = ad - bc \quad \textbf{which is} \quad \det A.$$

After a row exchange the pivots change to c and $b - (a/c)d$. Those new pivots multiply to give $bc - ad$. The row exchange to $\left[\begin{smallmatrix} c & d \\ a & b \end{smallmatrix}\right]$ reversed the sign of the determinant.

Looking ahead The determinant of an n by n matrix can be found in three ways:

1 Multiply the n pivots (times 1 or -1)	This is the **pivot formula.**
2 Add up $n!$ terms (times 1 or -1)	This is the **"big" formula.**
3 Combine n smaller determinants (times 1 or -1)	This is the **cofactor formula.**

You see that *plus or minus signs*—the decisions between 1 and -1—play a big part in determinants. That comes from the following rule for n by n matrices:

The determinant changes sign when two rows* (or two columns) *are exchanged.

The identity matrix has determinant $+1$. Exchange two rows and $\det P = -1$. Exchange two more rows and the new permutation has $\det P = +1$. Half of all permutations are *even* ($\det P = 1$) and half are *odd* ($\det P = -1$). Starting from I, half of the P's involve an even number of exchanges and half require an odd number. In the 2 by 2 case, ad has a plus sign and bc has minus—coming from the row exchange:

$$\det\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} = 1 \quad\text{and}\quad \det\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} = -1.$$

The other essential rule is linearity—but a warning comes first. Linearity does not mean that $\det(A+B) = \det A + \det B$. ***This is absolutely false.*** That kind of linearity is not even true when $A = I$ and $B = I$. The false rule would say that $\det(I + I) = 1 + 1 = 2$. The true rule is $\det 2I = 2^n$. Determinants are multiplied by 2^n (not just by 2) when matrices are multiplied by 2.

We don't intend to define the determinant by its formulas. It is better to start with its properties—*sign reversal and linearity*. The properties are simple (Section 5.1). They prepare for the formulas (Section 5.2). Then come the applications, including these three:

(1) Determinants give A^{-1} and $A^{-1}\boldsymbol{b}$ (this formula is called **Cramer's Rule**).

(2) When the edges of a box are the rows of A, the **volume** is $|\det A|$.

(3) For n special numbers λ, called **eigenvalues**, the determinants of $A - \lambda I$ is zero. This is a truly important application and it fills Chapter 6.

The Properties of the Determinant

Determinants have three basic properties (rules 1, 2, 3). By using those rules we can compute the determinant of any square matrix A. ***This number is written in two ways,* $\det A$ *and* $|A|$.** Notice: Brackets for the matrix, straight bars for its determinant. When A is a 2 by 2 matrix, the three properties lead to the answer we expect:

$$\text{The determinant of}\quad \begin{bmatrix}a & b\\ c & d\end{bmatrix} \quad\text{is}\quad \begin{vmatrix}a & b\\ c & d\end{vmatrix} = ad - bc.$$

The last rules are $\det(AB) = (\det A)(\det B)$ and $\det A^{\mathrm{T}} = \det A$. We will check all rules with the 2 by 2 formula, but do not forget: The rules apply to any n by n matrix. We will show how rules $4 - 10$ always follow from $1 - 3$.

Rule 1 (the easiest) matches $\det I = 1$ with the volume $= 1$ for a unit cube.

1 ***The determinant of the n by n identity matrix is* 1.**

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \qquad \text{and} \qquad \begin{vmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{vmatrix} = 1.$$

2 ***The determinant changes sign when two rows are exchanged*** (sign reversal):

$$\text{Check:} \quad \begin{vmatrix} c & d \\ a & b \end{vmatrix} = - \begin{vmatrix} a & b \\ c & d \end{vmatrix} \quad \text{(both sides equal } bc - ad\text{).}$$

Because of this rule, we can find $\det P$ for any permutation matrix. Just exchange rows of I until you reach P. Then $\det P = +1$ for an ***even*** number of row exchanges and $\det P = -1$ for an ***odd*** number.

The third rule has to make the big jump to the determinants of all matrices.

3 ***The determinant is a linear function of each row separately*** (all other rows stay fixed). If the first row is multiplied by t, the determinant is multiplied by t. If first rows are added, determinants are added. This rule only applies when the other rows do not change! Notice how c and d stay the same:

multiply row 1 by any number t $\quad \begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix}$

add row 1 of A to row 1 of A': $\quad \begin{vmatrix} a+a' & b+b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}.$

In the first case, both sides are $tad - tbc$. Then t factors out. In the second case, both sides are $ad + a'd - bc - b'c$. These rules still apply when A is n by n, and the last $n-1$ rows don't change. May we emphasize rule 3 with numbers:

$$\begin{vmatrix} 4 & 8 & 8 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix} = 4 \begin{vmatrix} 1 & 2 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} 4 & 8 & 8 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 4 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix} + \begin{vmatrix} 0 & 8 & 8 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix}.$$

By itself, rule 3 does not say what those determinants are (the first one is 4).

Combining multiplication and addition, we get any linear combination in one row (the other rows must stay the same). Any row can be the one that changes, since rule 2 for row exchanges can put it up into the first row and back again.

This rule does not mean that $\det 2I = 2 \det I$. To obtain $2I$ we have to multiply *both* rows by 2, and the factor 2 comes out both times:

$$\begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 2^2 = 4 \qquad \text{and} \qquad \begin{vmatrix} t & 0 \\ 0 & t \end{vmatrix} = t^2.$$

This is just like area and volume. Expand a rectangle by 2 and its area increases by 4. Expand an n-dimensional box by t and its volume increases by t^n. The connection is no accident—we will see how *determinants equal volumes*.

Pay special attention to rules 1–3. They completely determine the number $\det A$. We could stop here to find a formula for n by n determinants. (a little complicated) We prefer to go gradually, with other properties that follow directly from the first three. These extra rules $4 - 10$ make determinants much easier to work with.

4 ***If two rows of A are equal, then*** $\det A = 0$.

Equal rows Check 2 by 2 : $\begin{vmatrix} a & b \\ a & b \end{vmatrix} = 0.$

Rule 4 follows from rule 2. (Remember we must use the rules and not the 2 by 2 formula.) *Exchange the two equal rows.* The determinant D is supposed to change sign. But also D has to stay the same, because the matrix is not changed. The only number with $-D = D$ is $D = 0$—this must be the determinant. (Note: In Boolean algebra the reasoning fails, because $-1 = 1$. Then D is defined by rules 1, 3, 4.)

A matrix with two equal rows has no inverse. Rule 4 makes $\det A = 0$. But matrices can be singular and determinants can be zero without having equal rows! Rule 5 will be the key. We can do row operations without changing $\det A$.

5 ***Subtracting a multiple of one row from another row leaves*** $\det A$ ***unchanged.***

ℓ times row 1 from row 2 $\begin{vmatrix} a & b \\ c - \ell a & d - \ell b \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$

Rule 3 (linearity) splits the left side into the right side plus another term $-\ell\begin{vmatrix} a & b \\ a & b \end{vmatrix}$. This extra term is zero by rule 4. Therefore rule 5 is correct (not just 2 by 2).

Conclusion *The determinant is not changed by the usual elimination steps from A to U.* Thus $\det A$ *equals* $\det U$. If we can find determinants of triangular matrices U, we can find determinants of all matrices A. Every row exchange reverses the sign, so always $\det A = \pm \det U$. Rule 5 has narrowed the problem to triangular matrices.

6 ***A matrix with a row of zeros has*** $\det A = 0$.

Row of zeros $\begin{vmatrix} 0 & 0 \\ c & d \end{vmatrix} = 0$ and $\begin{vmatrix} a & b \\ 0 & 0 \end{vmatrix} = 0.$

For an easy proof, add some other row to the zero row. The determinant is not changed (rule 5). But the matrix now has two equal rows. So $\det A = 0$ by rule 4.

7 ***If A is triangular then*** $\det A = a_{11}a_{22}\cdots a_{nn}$ ***= product of diagonal entries.***

Triangular $\begin{vmatrix} a & b \\ 0 & d \end{vmatrix} = ad$ and also $\begin{vmatrix} a & 0 \\ c & d \end{vmatrix} = ad.$

Suppose all diagonal entries of A are nonzero. Eliminate the off-diagonal entries by the usual steps. (If A is lower triangular, subtract multiples of each row from lower rows. If A

is upper triangular, subtract from higher rows.) By rule 5 the determinant is not changed—and now the matrix is diagonal:

Diagonal matrix $$\det \begin{bmatrix} a_{11} & & & 0 \\ & a_{22} & & \\ & & \ddots & \\ 0 & & & a_{nn} \end{bmatrix} = a_{11}a_{22}\cdots a_{nn}.$$

Factor a_{11} from the first row by rule 3. Then factor a_{22} from the second row. Eventually factor a_{nn} from the last row. The determinant is a_{11} times a_{22} times $\cdots$ times a_{nn} times $\det I$. Then rule 1 (used at last!) is $\det I = 1$.

What if a diagonal entry a_{ii} is zero? Then the triangular A is singular. Elimination produces a *zero row*. By rule 5 the determinant is unchanged, and by rule 6 a zero row means $\det A = 0$. Triangular matrices have easy determinants.

8 ***If A is singular then* $\det A = 0$. *If A is invertible then* $\det A \neq 0$.**

Singular $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is singular if and only if $ad - bc = 0$.

Proof Elimination goes from A to U. If A is singular then U has a zero row. The rules give $\det A = \det U = 0$. If A is invertible then U has the pivots along its diagonal. The product of nonzero pivots (using rule 7) gives a nonzero determinant:

Multiply pivots $$\det A = \pm \det U = \pm \text{ (product of the pivots).} \tag{2}$$

The pivots of a 2 by 2 matrix (if $a \neq 0$) are a and $d - (bc/a)$:

The determinant is $$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ 0 & d - (bc/a) \end{vmatrix} = ad - bc.$$

This is the first formula for the determinant. MATLAB uses it to find $\det A$ from the pivots. The sign in $\pm \det u$ depends on whether the number of row exchanges is even or odd. In other words, $+1$ or -1 is the determinant of the permutation matrix P that exchanges rows. With no row exchanges, the number zero is even and $P = I$ and $\det A = \det U =$ *product of pivots*. Always $\det L = 1$, because L is triangular with 1's on the diagonal. What we have is this:

$$\text{If} \quad PA = LU \quad \text{then} \quad \det P \ \det A = \det L \ \det U. \tag{3}$$

Again, $\det P = \pm 1$ and $\det A = \pm \det U$. Equation (3) is our first case of rule 9.

9 ***The determinant of AB is* $\det A$ *times* $\det B$: $|AB| = |A|\,|B|$.**

Product rule $$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \begin{vmatrix} p & q \\ r & s \end{vmatrix} = \begin{vmatrix} ap + br & aq + bs \\ cp + dr & cq + ds \end{vmatrix}.$$

When the matrix B is A^{-1}, *this rule says that the determinant of* A^{-1} is $\mathbf{1/\det A}$:

$$\textbf{A times } A^{-1} \qquad \boxed{AA^{-1} = I \quad \text{so} \quad (\det A)(\det A^{-1}) = \det I = 1.}$$

This product rule is the most intricate so far. Even the 2 by 2 case needs some algebra:

$$|A|\,|B| = (ad - bc)(ps - qr) = (ap + br)(cq + ds) - (aq + bs)(cp + dr) = |AB|.$$

For the n by n case, here is a snappy proof that $|AB| = |A|\,|B|$. When $|B|$ is not zero, consider the ratio $D(A) = |AB|/|B|$. *Check that this ratio has properties* 1,2,3. Then $D(A)$ has to be the determinant and we have $|A| = |AB|/|B|$: good.

Property 1 *(Determinant of I)* If $A = I$ then the ratio becomes $|B|/|B| = 1$.

Property 2 *(Sign reversal)* When two rows of A are exchanged, so are the same two rows of AB. Therefore $|AB|$ changes sign and so does the ratio $|AB|/|B|$.

Property 3 *(Linearity)* When row 1 of A is multiplied by t, so is row 1 of AB. This multiplies $|AB|$ by t and multiplies the ratio by t—as desired.
Add row 1 of A to row 1 of A'. Then row 1 of AB adds to row 1 of $A'B$. By rule 3, determinants add. After dividing by $|B|$, the ratios add—as desired.

Conclusion This ratio $|AB|/|B|$ has the same three properties that define $|A|$. Therefore it equals $|A|$. This proves the product rule $|AB| = |A|\,|B|$. The case $|B| = 0$ is separate and easy, because AB is singular when B is singular. Then $|AB| = |A|\,|B|$ is $0 = 0$.

10 ***The transpose A^{T} has the same determinant as A.***

$$\textbf{Transpose} \quad \begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix} \quad \text{since both sides equal} \quad ad - bc.$$

The equation $|A^{\mathrm{T}}| = |A|$ becomes $0 = 0$ when A is singular (we know that A^{T} is also singular). Otherwise A has the usual factorization $PA = LU$. Transposing both sides gives $A^{\mathrm{T}}P^{\mathrm{T}} = U^{\mathrm{T}}L^{\mathrm{T}}$. The proof of $|A| = |A^{\mathrm{T}}|$ comes by using rule 9 for products:

$$\text{Compare} \quad \det P \det A = \det L \det U \quad \text{with} \quad \det A^{\mathrm{T}} \det P^{\mathrm{T}} = \det U^{\mathrm{T}} \det L^{\mathrm{T}}.$$

First, $\det L = \det L^{\mathrm{T}} = 1$ (both have 1's on the diagonal). Second, $\det U = \det U^{\mathrm{T}}$ (those triangular matrices have the same diagonal). Third, $\det P = \det P^{\mathrm{T}}$ (permutations have $P^{\mathrm{T}}P = I$, so $|P^{\mathrm{T}}|\,|P| = 1$ by rule 9; thus $|P|$ and $|P^{\mathrm{T}}|$ both equal 1 or both equal -1). So L, U, P have the same determinants as $L^{\mathrm{T}}, U^{\mathrm{T}}, P^{\mathrm{T}}$ and this leaves $\det A = \det A^{\mathrm{T}}$.

Important comment on columns Every rule for the rows can apply to the columns (just by transposing, since $|A| = |A^{\mathrm{T}}|$). The determinant changes sign when two columns are exchanged. *A zero column or two equal columns will make the determinant zero.* If a column is multiplied by t, so is the determinant. The determinant is a linear function of each column separately.

It is time to stop. The list of properties is long enough. Next we find and use an explicit formula for the determinant.

■ REVIEW OF THE KEY IDEAS ■

1. The determinant is defined by $\det I = 1$, sign reversal, and linearity in each row.
2. After elimination $\det A$ is $\pm$ (product of the pivots).
3. The determinant is zero exactly when A is not invertible.
4. Two remarkable properties are $\det AB = (\det A)(\det B)$ and $\det A^{\mathrm{T}} = \det A$.

■ WORKED EXAMPLES ■

5.1 A Apply these operations to A and find the determinants of M_1, M_2, M_3, M_4:
In M_1, multiplying each a_{ij} by $(-1)^{i+j}$ gives a checkerboard sign pattern.
In M_2, rows $1, 2, 3$ of A are *subtracted* from rows $2, 3, 1$.
In M_3, rows $1, 2, 3$ of A are *added* to rows $2, 3, 1$.
How are the determinants of M_1, M_2, M_3 related to the determinant of A?

$$\begin{bmatrix} a_{11} & -a_{12} & a_{13} \\ -a_{21} & a_{22} & -a_{23} \\ a_{31} & -a_{32} & a_{33} \end{bmatrix} \quad \begin{bmatrix} \text{row } 1 - \text{row } 3 \\ \text{row } 2 - \text{row } 1 \\ \text{row } 3 - \text{row } 2 \end{bmatrix} \quad \begin{bmatrix} \text{row } 1 + \text{row } 3 \\ \text{row } 2 + \text{row } 1 \\ \text{row } 3 + \text{row } 2 \end{bmatrix}$$

Solution The three determinants are $\det A$, 0, and $2\det A$. Here are reasons:

$$M_1 = \begin{bmatrix} 1 & & \\ & -1 & \\ & & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} 1 & & \\ & -1 & \\ & & 1 \end{bmatrix} \quad \text{so } \det M_1 = (-1)(\det A)(-1).$$

M_2 is singular because its rows add to the zero row. Its determinant is zero.
M_3 can be split into *eight matrices* by Rule 3 (linearity in each row seperately):

$$\begin{vmatrix} \text{row } 1 + \text{row } 3 \\ \text{row } 2 + \text{row } 1 \\ \text{row } 3 + \text{row } 3 \end{vmatrix} = \begin{vmatrix} \text{row } 1 \\ \text{row } 2 \\ \text{row } 3 \end{vmatrix} + \begin{vmatrix} \text{row } 3 \\ \text{row } 2 \\ \text{row } 3 \end{vmatrix} + \begin{vmatrix} \text{row } 1 \\ \text{row } 1 \\ \text{row } 3 \end{vmatrix} + \cdots + \begin{vmatrix} \text{row } 3 \\ \text{row } 1 \\ \text{row } 2 \end{vmatrix}.$$

All but the first and last have repeated rows and zero determinant. The first is A and the last has *two* row exchanges. So $\det M_3 = \det A + \det A$. (Try $A = I$.)

5.1 B Explain how to reach this determinant by row operations:

$$\det\begin{bmatrix} 1-a & 1 & 1 \\ 1 & 1-a & 1 \\ 1 & 1 & 1-a \end{bmatrix} = a^2(3-a). \tag{4}$$

Solution Subtract row 3 from row 1 and then from row 2. This leaves

$$\det\begin{bmatrix} -a & 0 & a \\ 0 & -a & a \\ 1 & 1 & 1-a \end{bmatrix}.$$

Now add column 1 to column 3, and also column 2 to column 3. This leaves a lower triangular matrix with $-a, -a, 3-a$ on the diagonal: $\det = (-a)(-a)(3-a)$.

The determinant is zero if $a = 0$ or $a = 3$. For $a = 0$ we have the *all-ones matrix*—certainly singular. For $a = 3$, each row adds to zero - again singular. Those numbers 0 and 3 are the eigenvalues of the all-ones matrix. This example is revealing and important, leading toward Chapter 6.

Problem Set 5.1

Questions 1–12 are about the rules for determinants.

1 If a 4 by 4 matrix has $\det A = \frac{1}{2}$, find $\det(2A)$ and $\det(-A)$ and $\det(A^2)$ and $\det(A^{-1})$.

2 If a 3 by 3 matrix has $\det A = -1$, find $\det(\frac{1}{2}A)$ and $\det(-A)$ and $\det(A^2)$ and $\det(A^{-1})$.

3 True or false, with a reason if true or a counterexample if false:

(a) The determinant of $I + A$ is $1 + \det A$.

(b) The determinant of ABC is $|A|\,|B|\,|C|$.

(c) The determinant of $4A$ is $4|A|$.

(d) The determinant of $AB - BA$ is zero. Try an example with $A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$.

4 Which row exchanges show that these "reverse identity matrices" J_3 and J_4 have $|J_3| = -1$ but $|J_4| = +1$?

$$\det\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = -1 \quad \text{but} \quad \det\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} = +1.$$

5 For $n = 5, 6, 7$, count the row exchanges to permute the reverse identity J_n to the identity matrix I_n. Propose a rule for every size n and predict whether J_{101} has determinant $+1$ or -1.

6 Show how Rule 6 (determinant $= 0$ if a row is all zero) comes from Rule 3.

7 Find the determinants of rotations and reflections:

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{and} \quad Q = \begin{bmatrix} 1-2\cos^2\theta & -2\cos\theta\sin\theta \\ -2\cos\theta\sin\theta & 1-2\sin^2\theta \end{bmatrix}.$$

8 Prove that every orthogonal matrix ($Q^{\mathrm{T}}Q = I$) has determinant 1 or -1.

(a) Use the product rule $|AB| = |A|\,|B|$ and the transpose rule $|Q| = |Q^{\mathrm{T}}|$.

(b) Use only the product rule. If $|\det Q| > 1$ then $\det Q^n = (\det Q)^n$ blows up. How do you know this can't happen to Q^n?

9 Do these matrices have determinant 0, 1, 2, or 3?

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

10 If the entries in every row of A add to zero, solve $Ax = \mathbf{0}$ to prove $\det A = 0$. If those entries add to one, show that $\det(A - I) = 0$. Does this mean $\det A = 1$?

11 Suppose that $CD = -DC$ and find the flaw in this reasoning: Taking determinants gives $|C|\,|D| = -|D|\,|C|$. Therefore $|C| = 0$ or $|D| = 0$. One or both of the matrices must be singular. (That is not true.)

12 The inverse of a 2 by 2 matrix seems to have determinant $= 1$:

$$\det A^{-1} = \det \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{ad-bc}{ad-bc} = 1.$$

What is wrong with this calculation? What is the correct $\det A^{-1}$?

Questions 13–27 use the rules to compute specific determinants.

13 Reduce A to U and find $\det A =$ product of the pivots:

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 3 \\ 3 & 3 & 3 \end{bmatrix}.$$

14 By applying row operations to produce an upper triangular U, compute

$$\det \begin{bmatrix} 1 & 2 & 3 & 0 \\ 2 & 6 & 6 & 1 \\ -1 & 0 & 0 & 3 \\ 0 & 2 & 0 & 7 \end{bmatrix} \quad \text{and} \quad \det \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}.$$

15 Use row operations to simplify and compute these determinants:

$$\det\begin{bmatrix}101 & 201 & 301\\ 102 & 202 & 302\\ 103 & 203 & 303\end{bmatrix} \quad\text{and}\quad \det\begin{bmatrix}1 & t & t^2\\ t & 1 & t\\ t^2 & t & 1\end{bmatrix}.$$

16 Find the determinants of a rank one matrix and a skew-symmetric matrix:

$$A=\begin{bmatrix}1\\2\\3\end{bmatrix}\begin{bmatrix}1 & -4 & 5\end{bmatrix} \quad\text{and}\quad K=\begin{bmatrix}0 & 1 & 3\\ -1 & 0 & 4\\ -3 & -4 & 0\end{bmatrix}.$$

17 A skew-symmetric matrix has $K^{\mathrm{T}} = -K$. Insert a, b, c for $1, 3, 4$ in Question 16 and show that $|K| = 0$. Write down a 4 by 4 example with $|K| = 1$.

18 Use row operations to show that the 3 by 3 "Vandermonde determinant" is

$$\det\begin{bmatrix}1 & a & a^2\\ 1 & b & b^2\\ 1 & c & c^2\end{bmatrix} = (b-a)(c-a)(c-b).$$

19 Find the determinants of U and U^{-1} and U^2:

$$U=\begin{bmatrix}1 & 4 & 6\\ 0 & 2 & 5\\ 0 & 0 & 3\end{bmatrix} \quad\text{and}\quad U=\begin{bmatrix}a & b\\ 0 & d\end{bmatrix}.$$

20 Suppose you do two row operations at once, going from

$$\begin{bmatrix}a & b\\ c & d\end{bmatrix} \quad\text{to}\quad \begin{bmatrix}a-Lc & b-Ld\\ c-la & d-lb\end{bmatrix}.$$

Find the second determinant. Does it equal $ad - bc$?

21 *Row exchange*: Add row 1 of A to row 2, then subtract row 2 from row 1. Then add row 1 to row 2 and multiply row 1 by -1 to reach B. Which rules show

$$\det B = \begin{vmatrix}c & d\\ a & b\end{vmatrix} \quad\text{equals}\quad -\det A = -\begin{vmatrix}a & b\\ c & d\end{vmatrix}?$$

Those rules could replace Rule 2 in the definition of the determinant.

22 From $ad - bc$, find the determinants of A and A^{-1} and $A - \lambda I$:

$$A=\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix} \quad\text{and}\quad A^{-1}=\frac{1}{3}\begin{bmatrix}2 & -1\\ -1 & 2\end{bmatrix} \quad\text{and}\quad A-\lambda I=\begin{bmatrix}2-\lambda & 1\\ 1 & 2-\lambda\end{bmatrix}.$$

Which two numbers λ lead to $\det(A - \lambda I) = 0$? Write down the matrix $A - \lambda I$ for each of those numbers λ—it should not be invertible.

23 From $A = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$ find A^2 and A^{-1} and $A - \lambda I$ and their determinants. Which two numbers λ lead to $\det(A - \lambda I) = 0$?

24 Elimination reduces A to U. Then $A = LU$:

$$A = \begin{bmatrix} 3 & 3 & 4 \\ 6 & 8 & 7 \\ -3 & 5 & -9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 4 & 1 \end{bmatrix} \begin{bmatrix} 3 & 3 & 4 \\ 0 & 2 & -1 \\ 0 & 0 & -1 \end{bmatrix} = LU.$$

Find the determinants of L, U, A, $U^{-1}L^{-1}$, and $U^{-1}L^{-1}A$.

25 If the i, j entry of A is i times j, show that $\det A = 0$. (Exception when $A = [\,1\,]$.)

26 If the i, j entry of A is $i + j$, show that $\det A = 0$. (Exception when $n = 1$ or 2.)

27 Compute the determinants of these matrices by row operations:

$$A = \begin{bmatrix} 0 & a & 0 \\ 0 & 0 & b \\ c & 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & a & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & 0 & 0 & c \\ d & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} a & a & a \\ a & b & b \\ a & b & c \end{bmatrix}.$$

28 True or false (give a reason if true or a 2 by 2 example if false):

(a) If A is not invertible then AB is not invertible.
(b) The determinant of A is always the product of its pivots.
(c) The determinant of $A - B$ equals $\det A - \det B$.
(d) AB and BA have the same determinant.

29 What is wrong with this proof that projection matrices have $\det P = 1$?

$$P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}} \qquad \text{so} \qquad |P| = |A|\frac{1}{|A^{\mathrm{T}}||A|}|A^{\mathrm{T}}| = 1.$$

30 (Calculus question) Show that the partial derivatives of $\ln(\det A)$ give A^{-1}!

$$f(a,b,c,d) = \ln(ad - bc) \qquad \text{leads to} \qquad \begin{bmatrix} \partial f/\partial a & \partial f/\partial c \\ \partial f/\partial b & \partial f/\partial d \end{bmatrix} = A^{-1}.$$

31 (MATLAB) The Hilbert matrix **hilb**(n) has i, j entry equal to $1/(i + j - 1)$. Print the determinants of **hilb**(1), **hilb**$(2), \ldots,$ **hilb**(10). Hilbert matrices are hard to work with! What are the pivots of **hilb** (5)?

32 (MATLAB) What is a typical determinant (experimentally) of **rand**(n) and **randn**(n) for $n = 50, 100, 200, 400$? (And what does "Inf" mean in MATLAB?)

33 (MATLAB) Find the largest determinant of a 6 by 6 matrix of 1's and -1's.

34 If you know that $\det A = 6$, what is the determinant of B?

$$\text{From } \det A = \begin{vmatrix} \text{row 1} \\ \text{row 2} \\ \text{row 3} \end{vmatrix} = 6 \text{ find } \det B = \begin{vmatrix} \text{row 3 + row 2 + row 1} \\ \text{row 2 + row 1} \\ \text{row 1} \end{vmatrix}.$$

5.2 Permutations and Cofactors

A computer finds the determinant from the pivots. This section explains two other ways to do it. There is a "big formula" using all $n!$ permutations. There is a "cofactor formula" using determinants of size $n-1$. The best example is my favorite 4 by 4 matrix:

$$A = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \quad \text{has} \quad \det A = 5.$$

We can find this determinant in all three ways: ***pivots, big formula, cofactors.***

1. The product of the pivots is $2 \cdot \frac{3}{2} \cdot \frac{4}{3} \cdot \frac{5}{4}$. Cancellation produces 5.
2. The "big formula" in equation (8) has $4! = 24$ terms. Only five terms are nonzero:
$$\det A = 16 - 4 - 4 - 4 + 1 = 5.$$
The 16 comes from $2 \cdot 2 \cdot 2 \cdot 2$ on the diagonal of A. Where do -4 and $+1$ come from? When you can find those five terms, you have understood formula (8).
3. The numbers $2, -1, 0, 0$ in the first row multiply their cofactors $4, 3, 2, 1$ from the other rows. That gives $2 \cdot 4 - 1 \cdot 3 = 5$. Those cofactors are 3 by 3 determinants. Cofactors use the rows and columns that are *not* used by the entry in the first row. ***Every term in a determinant uses each row and column once!***

The Pivot Formula

Elimination leaves the pivots $d_1, \ldots, d_n$ on the diagonal of the upper triangular U. If no row exchanges are involved, ***multiply those pivots*** to find the determinant:

$$\det A = (\det L)(\det U) = (1)(d_1 d_2 \cdots d_n). \tag{1}$$

This formula for $\det A$ appeared in the previous section, with the further possibility of row exchanges. The permutation matrix in $PA = LU$ has determinant -1 or $+1$. This factor $\det P = \pm 1$ enters the determinant of A:

$$(\det P)(\det A) = (\det L)(\det U) \quad \text{gives} \quad \det A = \pm(d_1 d_2 \cdots d_n). \tag{2}$$

When A has fewer than n pivots, $\det A = 0$ by Rule 8. The matrix is singular.

Example 1 A row exchange produces pivots 4, 2, 1 and that important minus sign:

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \qquad PA = \begin{bmatrix} 4 & 5 & 6 \\ 0 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix} \qquad \det A = -(4)(2)(1) = -8.$$

The odd number of row exchanges (namely one exchange) means that $\det P = -1$.

The next example has no row exchanges. It may be the first matrix we factored into LU (when it was 3 by 3). What is remarkable is that we can go directly to n by n. Pivots give the determinant. We will also see how determinants give the pivots.

Example 2 The first pivots of this tridiagonal matrix A are $2, \frac{3}{2}, \frac{4}{3}$. The next are $\frac{5}{4}$ and $\frac{6}{5}$ and eventually $\frac{n+1}{n}$. Factoring this n by n matrix reveals its determinant:

$$\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & \cdot & \\ & & \cdot & \cdot & -1 \\ & & & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ -\frac{1}{2} & 1 & & & \\ & -\frac{2}{3} & 1 & & \\ & & \cdot & \cdot & \\ & & & -\frac{n-1}{n} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{2} & -1 & & & \\ & \frac{3}{2} & -1 & & \\ & & \frac{4}{3} & -1 & \\ & & & \cdot & \cdot \\ & & & & \frac{n+1}{n} \end{bmatrix}$$

The pivots are on the diagonal of U (the last matrix). When 2 and $\frac{3}{2}$ and $\frac{4}{3}$ and $\frac{5}{4}$ are multiplied, the fractions cancel. The determinant of the 4 by 4 matrix is 5. The 3 by 3 determinant is 4. *The n by n determinant is $n+1$*:

$$\mathbf{-1, 2, -1\ matrix} \qquad \det A = (2)\left(\tfrac{3}{2}\right)\left(\tfrac{4}{3}\right)\cdots\left(\tfrac{n+1}{n}\right) = n+1.$$

Important point: The first pivots depend only on the *upper left corner* of the original matrix A. This is a rule for all matrices without row exchanges.

> The first k pivots come from the k by k matrix A_k in the top left corner of A.
> ***The determinant of that corner submatrix A_k is $d_1 d_2 \cdots d_k$.***

The 1 by 1 matrix A_1 contains the very first pivot d_1. This is $\det A_1$. The 2 by 2 matrix in the corner has $\det A_2 = d_1 d_2$. Eventually the n by n determinant uses the product of all n pivots to give $\det A_n$ which is $\det A$.

Elimination deals with the corner matrix A_k while starting on the whole matrix. We assume no row exchanges—then $A = LU$ and $A_k = L_k U_k$. Dividing one determinant by the previous determinant ($\det A_k$ divided by $\det A_{k-1}$) cancels everything but the latest pivot d_k. ***This gives a ratio of determinants formula for the pivots***:

$$\textbf{Pivots from determinants} \qquad \textbf{The } k\textbf{th pivot is } d_k = \frac{d_1 d_2 \cdots d_k}{d_1 d_2 \cdots d_{k-1}} = \frac{\det A_k}{\det A_{k-1}}. \tag{3}$$

In the $-1, 2, -1$ matrices this ratio correctly gives the pivots $\frac{2}{1}, \frac{3}{2}, \frac{4}{3}, \ldots, \frac{n+1}{n}$. The Hilbert matrices in Problem 5.1.31 also build from the upper left corner.

We don't need row exchanges when all these corner submatrices have $\det A_k \neq 0$.

The Big Formula for Determinants

Pivots are good for computing. They concentrate a lot of information—enough to find the determinant. But it is hard to connect them to the original a_{ij}. That part will be clearer if we go back to rules 1-2-3, linearity and sign reversal and $\det I = 1$. We want to derive a single explicit formula for the determinant, directly from the entries a_{ij}.

The formula has $n!$ terms. Its size grows fast because $n! = 1, 2, 6, 24, 120, \ldots$. For $n = 11$ there are about forty million terms. For $n = 2$, the two terms are ad and bc. Half the terms have minus signs (as in $-bc$). The other half have plus signs (as in ad). For $n = 3$ there are $3! = (3)(2)(1)$ terms. Here are those six terms:

$$\text{3 by 3 determinant} \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{matrix} +a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ -a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}. \end{matrix} \tag{4}$$

Notice the pattern. Each product like $a_{11}a_{23}a_{32}$ has ***one entry from each row***. It also has ***one entry from each column***. The column order 1, 3, 2 means that this particular term comes with a minus sign. The column order 3, 1, 2 in $a_{13}a_{21}a_{32}$ has a plus sign. It will be "permutations" that tell us the sign.

The next step ($n = 4$) brings $4! = 24$ terms. There are 24 ways to choose one entry from each row and column. Down the main diagonal, $a_{11}a_{22}a_{33}a_{44}$ with column order $1, 2, 3, 4$ always has a plus sign. That is the "identity permutation".

To derive the big formula I start with $n = 2$. The goal is to reach $ad-bc$ in a systematic way. Break each row into two simpler rows:

$$\begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} a & 0 \end{bmatrix} + \begin{bmatrix} 0 & b \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} c & d \end{bmatrix} = \begin{bmatrix} c & 0 \end{bmatrix} + \begin{bmatrix} 0 & d \end{bmatrix}.$$

Now apply linearity, first in row 1 (with row 2 fixed) and then in row 2 (with row 1 fixed):

$$\begin{aligned} \begin{vmatrix} a & b \\ c & d \end{vmatrix} &= \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix} \\ &= \begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix} + \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix}. \end{aligned} \tag{5}$$

The last line has $2^2 = 4$ determinants. The first and fourth are zero because their rows are dependent—one row is a multiple of the other row. We are left with $2! = 2$ determinants to compute:

$$\begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} = ad \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} + bc \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = ad - bc.$$

The splitting led to permutation matrices. Their determinants give a plus or minus sign. The 1's are multiplied by numbers that come from A. The permutation tells the column sequence, in this case $(1, 2)$ or $(2, 1)$.

Now try $n = 3$. Each row splits into 3 simpler rows like $[a_{11} \;\; 0 \;\; 0]$. Using linearity in each row, $\det A$ splits into $3^3 = 27$ simple determinants. If a column choice is repeated—for example if we also choose $[a_{21} \;\; 0 \;\; 0]$—then the simple determinant is zero. We pay attention only when ***the nonzero terms come from different columns***.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & & \\ & a_{22} & \\ & & a_{33} \end{vmatrix} + \begin{vmatrix} & a_{12} & \\ & & a_{23} \\ a_{31} & & \end{vmatrix} + \begin{vmatrix} & & a_{13} \\ a_{21} & & \\ & a_{32} & \end{vmatrix}$$

$$\textit{\textbf{Six terms}} \quad + \begin{vmatrix} a_{11} & & \\ & & a_{23} \\ & a_{32} & \end{vmatrix} + \begin{vmatrix} & a_{12} & \\ a_{21} & & \\ & & a_{33} \end{vmatrix} + \begin{vmatrix} & & a_{13} \\ & a_{22} & \\ a_{31} & & \end{vmatrix}.$$

There are $3! = 6$ ***ways to order the columns, so six determinants.*** The six permutations of $(1, 2, 3)$ include the identity permutation $(1, 2, 3)$ from $P = I$:

$$\textbf{Column numbers} = (1,2,3), (2,3,1), (3,1,2), (1,3,2), (2,1,3), (3,2,1). \qquad (6)$$

The last three are *odd permutations* (one exchange). The first three are *even permutations* (0 or 2 exchanges). When the column sequence is (α, β, ω), we have chosen the entries $a_{1\alpha}a_{2\beta}a_{3\omega}$—and the column sequence comes with a plus or minus sign. The determinant of A is now split into six simple terms. Factor out the a_{ij}:

$$\begin{aligned} \det A = a_{11}a_{22}a_{33}\begin{vmatrix} 1 & & \\ & 1 & \\ & & 1 \end{vmatrix} + a_{12}a_{23}a_{31}\begin{vmatrix} & 1 & \\ & & 1 \\ 1 & & \end{vmatrix} + a_{13}a_{21}a_{32}\begin{vmatrix} & & 1 \\ 1 & & \\ & 1 & \end{vmatrix} \\ + a_{11}a_{23}a_{32}\begin{vmatrix} 1 & & \\ & & 1 \\ & 1 & \end{vmatrix} + a_{12}a_{21}a_{33}\begin{vmatrix} & 1 & \\ 1 & & \\ & & 1 \end{vmatrix} + a_{13}a_{22}a_{31}\begin{vmatrix} & & 1 \\ & 1 & \\ 1 & & \end{vmatrix}. \end{aligned} \qquad (7)$$

The first three (even) permutations have $\det P = +1$, the last three (odd) permutations have $\det P = -1$. We have proved the 3 by 3 formula in a systematic way.

Now you can see the n by n formula. There are $n!$ orderings of the columns. The columns $(1, 2, \ldots, n)$ go in each possible order $(\alpha, \beta, \ldots, \omega)$. Taking $a_{1\alpha}$ from row 1 and $a_{2\beta}$ from row 2 and eventually $a_{n\omega}$ from row n, the determinant contains the product $a_{1\alpha}a_{2\beta}\cdots a_{n\omega}$ times $+1$ or -1. Half the column orderings have sign -1.

The complete determinant of A is the sum of these $n!$ simple determinants, times 1 or -1. The simple determinants $a_{1\alpha}a_{2\beta}\cdots a_{n\omega}$ choose ***one entry from every row and column***:

$$\begin{aligned} \det A &= \text{sum over all } \mathbf{n!} \text{ column permutations } P = (\alpha, \beta, \ldots, \omega) \\ &= \sum (\det P) a_{1\alpha}a_{2\beta}\cdots a_{n\omega} = \textbf{BIG FORMULA}. \end{aligned} \qquad (8)$$

The 2 by 2 case is $+a_{11}a_{22} - a_{12}a_{21}$ (which is $ad - bc$). Here P is $(1, 2)$ or $(2, 1)$.

The 3 by 3 case has three products "down to the right" (see Problem 28) and three products "down to the left". Warning: Many people believe they should follow this pattern in the 4 by 4 case. They only take 8 products—but we need 24.

Example 3 (Determinant of U) When U is upper triangular, only one of the $n!$ products can be nonzero. This one term comes from the diagonal: $\det U = +u_{11}u_{22}\cdots u_{nn}$. All other column orderings pick at least one entry below the diagonal, where U has zeros. As soon as we pick a number like $u_{21} = 0$ from below the diagonal, that term in equation (8) is sure to be zero.

Of course $\det I = 1$. The only nonzero term is $+(1)(1)\cdots(1)$ from the diagonal.

Example 4 Suppose Z is the identity matrix except for column 3. Then

$$\text{determinant of } Z = \begin{vmatrix} 1 & 0 & a & 0 \\ 0 & 1 & b & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & d & 1 \end{vmatrix} = c. \tag{9}$$

The term $(1)(1)(c)(1)$ comes from the main diagonal with a plus sign. There are 23 other products (choosing one factor from each row and column) but they are all zero. Reason: If we pick a, b, or d from column 3, that column is used up. Then the only available choice from row 3 is zero.

Here is a different reason for the same answer. If $c = 0$, then Z has a row of zeros and $\det Z = c = 0$ is correct. If c is not zero, *use elimination.* Subtract multiples of row 3 from the other rows, to knock out a, b, d. That leaves a diagonal matrix and $\det Z = c$.

This example will soon be used for "Cramer's Rule". If we move a, b, c, d into the first column of Z, the determinant is $\det Z = a$. (*Why*?) Changing one column of I leaves Z with an easy determinant, coming from its main diagonal only.

Example 5 Suppose A has 1's just above and below the main diagonal. Here $n = 4$:

$$A_4 = \begin{bmatrix} 0 & \mathbf{1} & 0 & 0 \\ \mathbf{1} & 0 & 1 & 0 \\ 0 & 1 & 0 & \mathbf{1} \\ 0 & 0 & \mathbf{1} & 0 \end{bmatrix} \quad \text{and} \quad P_4 = \begin{bmatrix} 0 & \mathbf{1} & 0 & 0 \\ \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} \\ 0 & 0 & \mathbf{1} & 0 \end{bmatrix} \quad \text{have determinant 1.}$$

The only nonzero choice in the first row is column 2. The only nonzero choice in row 4 is column 3. Then rows 2 and 3 must choose columns 1 and 4. In other words P_4 is the only permutation that picks out nonzeros in A_4. The determinant of P_4 is $+1$ (two exchanges to reach $2, 1, 4, 3$). Therefore $\det A_4 = +1$.

Determinant by Cofactors

Formula (8) is a direct definition of the determinant. It gives you everything at once—but you have to digest it. Somehow this sum of $n!$ terms must satisfy rules 1-2-3 (then all the other properties follow). The easiest is $\det I = 1$, already checked. The rule of linearity becomes clear, if you ***separate out the factor*** a_{11} ***or*** a_{12} ***or*** $a_{1\alpha}$ ***that comes from the first row.*** For 3 by 3, separate the usual 6 terms of the determinant into 3 pairs:

$$\det A = a_{11}\,(a_{22}a_{33} - a_{23}a_{32}) + a_{12}\,(a_{23}a_{31} - a_{21}a_{33}) + a_{13}\,(a_{21}a_{32} - a_{22}a_{31}). \tag{10}$$

Those three quantities in parentheses are called ***"cofactors"***. They are 2 by 2 determinants, coming from matrices in rows 2 and 3. The first row contributes the factors a_{11}, a_{12}, a_{13}. The lower rows contribute the cofactors C_{11}, C_{12}, C_{13}. Certainly the determinant $a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13}$ depends linearly on a_{11}, a_{12}, a_{13}—this is rule 3.

The cofactor of a_{11} is $C_{11} = a_{22}a_{33} - a_{23}a_{32}$. You can see it in this splitting:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & & \\ & a_{22} & a_{23} \\ & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} & a_{12} & \\ a_{21} & & a_{23} \\ a_{31} & & a_{33} \end{vmatrix} + \begin{vmatrix} & & a_{13} \\ a_{21} & a_{22} & \\ a_{31} & a_{32} & \end{vmatrix}.$$

We are still choosing ***one entry from each row and column***. Since a_{11} uses up row 1 and column 1, that leaves a 2 by 2 determinant as its cofactor.

As always, we have to watch signs. The 2 by 2 determinant that goes with a_{12} looks like $a_{21}a_{33} - a_{23}a_{31}$. But in the cofactor C_{12}, *its sign is reversed.* Then $a_{12}C_{12}$ is the correct 3 by 3 determinant. The sign pattern for cofactors along the first row is plus-minus-plus-minus. ***You cross out row*** 1 ***and column*** j ***to get a submatrix*** M_{1j} ***of size*** $n-1$. Multiply its determinant by $(-1)^{1+j}$ to get the cofactor:

$$\text{The cofactors along row 1 are } C_{1j} = (-1)^{1+j} \det M_{1j}.$$

$$\textbf{The cofactor expansion is } \det A = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}. \qquad (11)$$

In the big formula (8), the terms that multiply a_{11} combine to give $\det M_{11}$. The sign is $(-1)^{1+1}$, meaning *plus*. Equation (11) is another form of equation (8) and also equation (10), with factors from row 1 multiplying cofactors that use the other rows.

Note Whatever is possible for row 1 is possible for row i. The entries a_{ij} in that row also have cofactors C_{ij}. Those are determinants of order $n-1$, multiplied by $(-1)^{i+j}$. Since a_{ij} accounts for row i and column j, ***the submatrix*** M_{ij} ***throws out row*** i ***and column*** j. The display shows a_{43} and M_{43} (with row 4 and column 3 removed). The sign $(-1)^{4+3}$ multiplies the determinant of M_{43} to give C_{43}. The sign matrix shows the $\pm$ pattern:

$$A = \begin{bmatrix} \bullet & \bullet & & \bullet \\ \bullet & \bullet & & \bullet \\ \bullet & \bullet & & \bullet \\ & & a_{43} & \end{bmatrix} \qquad \text{signs } (-1)^{i+j} = \begin{bmatrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{bmatrix}.$$

The determinant is the dot product of any row i of A with its cofactors using other rows:

$$\textbf{COFACTOR FORMULA} \qquad \det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}. \qquad (12)$$

Each cofactor C_{ij} (order $n-1$, without row i and column j) includes its correct sign:

$$\textbf{Cofactor} \qquad C_{ij} = (-1)^{i+j} \det M_{ij}.$$

A determinant of order n is a combination of determinants of order $n-1$. A recursive person would keep going. Each subdeterminant breaks into determinants of order $n-2$. *We could define all determinants via equation* (12). This rule goes from order n to $n-1$

to $n-2$ and eventually to order 1. Define the 1 by 1 determinant $|a|$ to be the number a. Then the cofactor method is complete.

We preferred to construct $\det A$ from its properties (linearity, sign reversal, $\det I = 1$). The big formula (8) and the cofactor formulas (10)–(12) follow from those properties. One last formula comes from the rule that $\det A = \det A^{\mathrm{T}}$. We can expand in cofactors, *down a column* instead of across a row. Down column j the entries are a_{1j} to a_{nj}. The cofactors are C_{1j} to C_{nj}. The determinant is the dot product:

$$\textbf{Cofactors down column } j\textbf{:} \qquad \det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}. \tag{13}$$

Cofactors are useful when matrices have many zeros—as in the next examples.

Example 6 The $-1, 2, -1$ matrix has only two nonzeros in its first row. So only two cofactors C_{11} and C_{12} are involved in the determinant. I will highlight C_{12}:

$$\begin{vmatrix} 2 & -1 & & \\ \mathbf{-1} & 2 & \mathbf{-1} & \\ & -1 & \mathbf{2} & \mathbf{-1} \\ & & \mathbf{-1} & \mathbf{2} \end{vmatrix} = 2\begin{vmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{vmatrix} - (-1)\begin{vmatrix} \mathbf{-1} & \mathbf{-1} & \\ & \mathbf{2} & \mathbf{-1} \\ & \mathbf{-1} & \mathbf{2} \end{vmatrix}. \tag{14}$$

You see 2 times C_{11} first on the right, from crossing out row 1 and column 1. This cofactor has exactly the same $-1, 2, -1$ pattern as the original A—but one size smaller.

To compute the boldface C_{12}, *use cofactors down its first column.* The only nonzero is at the top. That contributes another -1 (so we are back to minus). Its cofactor is the $-1, 2, -1$ determinant which is 2 by 2, *two sizes smaller* than the original A.

Summary ***Each determinant D_n of order n comes from D_{n-1} and D_{n-2}:***

$$D_4 = 2D_3 - D_2 \qquad \text{and generally} \qquad D_n = 2D_{n-1} - D_{n-2}. \tag{15}$$

Direct calculation gives $D_2 = 3$ and $D_3 = 4$. Equation (14) has $D_4 = 2(4) - 3 = 5$. These determinants 3, 4, 5 fit the formula $\boldsymbol{D_n = n + 1}$. That "special tridiagonal answer" also came from the product of pivots in Example 2.

The idea behind cofactors is to reduce the order one step at a time. The determinants $D_n = n + 1$ obey the recursion formula $n + 1 = 2n - (n - 1)$. As they must.

Example 7 This is the same matrix, except the first entry (upper left) is now 1:

$$B_4 = \begin{bmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{bmatrix}.$$

All pivots of this matrix turn out to be 1. So its determinant is 1. How does that come from cofactors? Expanding on row 1, the cofactors all agree with Example 6. Just change $a_{11} = 2$ to $b_{11} = 1$:

$$\det B_4 = D_3 - D_2 \qquad \text{instead of} \qquad \det A_4 = 2D_3 - D_2.$$

The determinant of B_4 is $4 - 3 = 1$. The determinant of every B_n is $n - (n - 1) = 1$. Problem 13 asks you to use cofactors of the *last* row. You still find $\det B_n = 1$.

■ REVIEW OF THE KEY IDEAS ■

1. With no row exchanges, $\det A$ = (*product of pivots*). In the upper left corner, $\det A_k$ = (*product of the first k pivots*).

2. Every term in the big formula (8) uses each row and column once. Half of the $n!$ terms have plus signs (when $\det P = +1$) and half have minus signs.

3. The cofactor C_{ij} is $(-1)^{i+j}$ times the smaller determinant that omits row i and column j (because a_{ij} uses that row and column).

4. The determinant is the dot product of any row of A with its row of cofactors. When a row of A has a lot of zeros, we only need a few cofactors.

■ WORKED EXAMPLES ■

5.2 A A *Hessenberg matrix* is a triangular matrix with one extra diagonal. Use cofactors of row 1 to show that the 4 by 4 determinant satisfies Fibonacci's rule $|H_4| = |H_3| + |H_2|$. The same rule will continue for all sizes, $|H_n| = |H_{n-1}| + |H_{n-2}|$. Which Fibonacci number is $|H_n|$?

$$H_2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \qquad H_3 = \begin{bmatrix} 2 & 1 & \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \qquad H_4 = \begin{bmatrix} 2 & 1 & & \\ \mathbf{1} & 2 & \mathbf{1} & \\ \mathbf{1} & 1 & \mathbf{2} & \mathbf{1} \\ \mathbf{1} & 1 & \mathbf{1} & \mathbf{2} \end{bmatrix}$$

Solution The cofactor C_{11} for H_4 is the determinant $|H_3|$. We also need C_{12} (in boldface):

$$C_{12} = -\begin{vmatrix} \mathbf{1} & \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{2} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{2} \end{vmatrix} = -\begin{vmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{vmatrix} + \begin{vmatrix} 1 & 0 & 0 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{vmatrix}$$

Rows 2 and 3 stayed the same and we used linearity in row 1. The two determinants on the right are $-|H_3|$ and $+|H_2|$. Then the 4 by 4 determinant is

$$|H_4| = 2C_{11} + 1C_{12} = 2|H_3| - |H_3| + |H_2| = |H_3| + |H_2|.$$

The actual numbers are $|H_2| = 3$ and $|H_3| = 5$ (and of course $|H_1| = 2$). Since $|H_n|$ follows Fibonacci's rule $|H_{n-1}| + |H_{n-2}|$, it must be $|H_n| = F_{n+2}$.

5.2 B These questions use the $\pm$ signs (even and odd P's) in the big formula for $\det A$:

1. If A is the 10 by 10 all-ones matrix, how does the big formula give $\det A = 0$?

2. If you multiply all $n!$ permutations together into a single P, is P odd or even?

3. If you multiply each a_{ij} by the fraction i/j, why is $\det A$ unchanged?

Solution In Question **1**, with all $a_{ij} = 1$, all the products in the big formula (8) will be 1. Half of them come with a plus sign, and half with minus. So they cancel to leave $\det A = 0$. (Of course the all-ones matrix is singular.)

In Question **2**, multiplying $\left[\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right]\left[\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right]$ gives an odd permutation. Also for 3 by 3, the three odd permutations multiply (in any order) to give *odd*. But for $n > 3$ the product of all permutations will be *even*. There are $n!/2$ odd permutations and that is an even number as soon as it includes the factor 4.

In Question **3**, each a_{ij} is multiplied by i/j. So each product $a_{1\alpha}a_{2\beta}\cdots a_{n\omega}$ in the big formula is multiplied by all the row numbers $i = 1, 2, \ldots, n$ and divided by all the column numbers $j = 1, 2, \ldots, n$. (The columns come in some permuted order!) Then each product is unchanged and $\det A$ stays the same.

Another approach to Question **3**: We are multiplying the matrix A by the diagonal matrix $D = \mathbf{diag}(1 : n)$ when row i is multiplied by i. And we are postmultiplying by D^{-1} when column j is divided by j. The determinant of DAD^{-1} is the same as $\det A$ by the product rule.

Problem Set 5.2

Problems 1–10 use the big formula with $n!$ terms: $|A| = \sum \pm a_{1\alpha}a_{2\beta}\cdots a_{n\omega}$.

1 Compute the determinants of A, B, C from six terms. Are their rows independent?

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 3 & 2 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 4 & 4 \\ 5 & 6 & 7 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

2 Compute the determinants of A, B, C, D. Are their columns independent?

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \quad C = \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix} \quad D = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}.$$

3 Show that $\det A = 0$, regardless of the five nonzeros marked by x's:

$$A = \begin{bmatrix} x & x & x \\ 0 & 0 & x \\ 0 & 0 & x \end{bmatrix}.$$

What are the cofactors of row 1?
What is the rank of A?
What are the 6 terms in $\det A$?

4 Find two ways to choose nonzeros from four different rows and columns:

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 3 & 4 & 5 \\ 5 & 4 & 0 & 3 \\ 2 & 0 & 0 & 1 \end{bmatrix} \quad (B \text{ has the same zeros as } A).$$

Is $\det A$ equal to $1+1$ or $1-1$ or $-1-1$? What is $\det B$?

5 Place the smallest number of zeros in a 4 by 4 matrix that will guarantee $\det A = 0$. Place as many zeros as possible while still allowing $\det A \neq 0$.

6 (a) If $a_{11} = a_{22} = a_{33} = 0$, how many of the six terms in $\det A$ will be zero?

(b) If $a_{11} = a_{22} = a_{33} = a_{44} = 0$, how many of the 24 products $a_{1j}a_{2k}a_{3l}a_{4m}$ are sure to be zero?

7 How many 5 by 5 permutation matrices have $\det P = +1$? Those are even permutations. Find one that needs four exchanges to reach the identity matrix.

8 If $\det A$ is not zero, at least one of the $n!$ terms in formula (8) is not zero. Deduce from the big formula that some ordering of the rows of A leaves no zeros on the diagonal. (Don't use P from elimination; that PA can have zeros on the diagonal.)

9 Show that 4 is the largest determinant for a 3 by 3 matrix of 1's and -1's.

10 How many permutations of $(1, 2, 3, 4)$ are even and what are they? Extra credit: What are all the possible 4 by 4 determinants of $I + P_{\text{even}}$?

Problems 11–22 use cofactors $C_{ij} = (-1)^{i+j} \det M_{ij}$. Remove row i and column j.

11 Find all cofactors and put them into cofactor matrices C, D. Find AC and $\det B$.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 0 & 0 \end{bmatrix}.$$

12 Find the cofactor matrix C and multiply A times C^{T}. Compare AC^{T} with A^{-1}:

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \qquad A^{-1} = \frac{1}{4}\begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

13 The n by n determinant C_n has 1's above and below the main diagonal:

$$C_1 = |0| \qquad C_2 = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} \qquad C_3 = \begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{vmatrix} \qquad C_4 = \begin{vmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{vmatrix}.$$

(a) What are these determinants C_1, C_2, C_3, C_4?

(b) By cofactors find the relation between C_n and C_{n-1} and C_{n-2}. Find C_{10}.

14 The matrices in Problem 13 have 1's just above and below the main diagonal. Going down the matrix, which order of columns (if any) gives all 1's? Explain why that permutation is *even* for $n = 4, 8, 12, \ldots$ and *odd* for $n = 2, 6, 10, \ldots$. Then

$$C_n = 0 \text{ (odd } n) \qquad C_n = 1 \ (n = 4, 8, \cdots) \qquad C_n = -1 \ (n = 2, 6, \cdots).$$

15 The tridiagonal 1, 1, 1 matrix of order n has determinant E_n:

$$E_1 = |1| \quad E_2 = \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} \quad E_3 = \begin{vmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{vmatrix} \quad E_4 = \begin{vmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix}.$$

(a) By cofactors show that $E_n = E_{n-1} - E_{n-2}$.

(b) Starting from $E_1 = 1$ and $E_2 = 0$ find $E_3, E_4, \ldots, E_8$.

(c) By noticing how these numbers eventually repeat, find E_{100}.

16 F_n is the determinant of the $1, 1, -1$ tridiagonal matrix of order n:

$$F_2 = \begin{vmatrix} 1 & -1 \\ 1 & 1 \end{vmatrix} = 2 \quad F_3 = \begin{vmatrix} 1 & -1 & 0 \\ 1 & 1 & -1 \\ 0 & 1 & 1 \end{vmatrix} = 3 \quad F_4 = \begin{vmatrix} 1 & -1 & & \\ 1 & 1 & -1 & \\ & 1 & 1 & -1 \\ & & 1 & 1 \end{vmatrix} \neq 4.$$

Expand in cofactors to show that $F_n = F_{n-1} + F_{n-2}$. These determinants are *Fibonacci numbers* $1, 2, 3, 5, 8, 13, \ldots$. The sequence usually starts $1, 1, 2, 3$ (with two 1's) so our F_n is the usual F_{n+1}.

17 The matrix B_n is the $-1, 2, -1$ matrix A_n except that $b_{11} = 1$ instead of $a_{11} = 2$. Using cofactors of the *last* row of B_4 show that $|B_4| = 2|B_3| - |B_2| = 1$.

$$B_4 = \begin{bmatrix} \mathbf{1} & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{bmatrix} \quad B_3 = \begin{bmatrix} \mathbf{1} & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{bmatrix} \quad B_2 = \begin{bmatrix} \mathbf{1} & -1 \\ -1 & 2 \end{bmatrix}.$$

The recursion $|B_n| = 2|B_{n-1}| - |B_{n-2}|$ is satisfied when every $|B_n| = 1$. This recursion is the same as for the A's in Example 6. The difference is in the starting values 1, 1, 1 for the determinants of sizes $n = 1, 2, 3$.

18 Go back to B_n in Problem 17. It is the same as A_n except for $b_{11} = 1$. So use linearity in the first row, where $[1 \ \ -1 \ \ 0]$ equals $[2 \ \ -1 \ \ 0]$ minus $[1 \ \ 0 \ \ 0]$:

$$|B_n| = \begin{vmatrix} 1 & -1 & & 0 \\ -1 & & & \\ & & A_{n-1} & \\ 0 & & & \end{vmatrix} = \begin{vmatrix} 2 & -1 & & 0 \\ -1 & & & \\ & & A_{n-1} & \\ 0 & & & \end{vmatrix} - \begin{vmatrix} 1 & 0 & & 0 \\ -1 & & & \\ & & A_{n-1} & \\ 0 & & & \end{vmatrix}.$$

Linearity gives $|B_n| = |A_n| - |A_{n-1}| =$ ____.

19 Explain why the 4 by 4 Vandermonde determinant contains x^3 but not x^4 or x^5:

$$V_4 = \det \begin{bmatrix} 1 & a & a^2 & a^3 \\ 1 & b & b^2 & b^3 \\ 1 & c & c^2 & c^3 \\ 1 & x & x^2 & x^3 \end{bmatrix}.$$

The determinant is zero at $x =$ ____, ____, and ____. The cofactor of x^3 is $V_3 = (b-a)(c-a)(c-b)$. Then $V_4 = (b-a)(c-a)(c-b)(x-a)(x-b)(x-c)$.

20 Find G_2 and G_3 and then by row operations G_4. Can you predict G_n?

$$G_2 = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} \qquad G_3 = \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{vmatrix} \qquad G_4 = \begin{vmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{vmatrix}.$$

21 Compute S_1, S_2, S_3 for these 1, 3, 1 matrices. By Fibonacci guess and check S_4.

$$S_1 = |3| \qquad S_2 = \begin{vmatrix} 3 & 1 \\ 1 & 3 \end{vmatrix} \qquad S_3 = \begin{vmatrix} 3 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 3 \end{vmatrix}$$

22 Change 3 to 2 in the upper left corner of the matrices in Problem 21. Why does that subtract S_{n-1} from the determinant S_n? Show that the determinants of the new matrices become the Fibonacci numbers 2, 5, 13 (always F_{2n+1}).

Problems 23–26 are about block matrices and block determinants.

23 With 2 by 2 blocks in 4 by 4 matrices, you cannot always use block determinants:

$$\begin{vmatrix} A & B \\ 0 & D \end{vmatrix} = |A|\,|D| \qquad \text{but} \qquad \begin{vmatrix} A & B \\ C & D \end{vmatrix} \neq |A|\,|D| - |C|\,|B|.$$

(a) Why is the first statement true? Somehow B doesn't enter.

(b) Show by example that equality fails (as shown) when C enters.

(c) Show by example that the answer $\det(AD - CB)$ is also wrong.

24 With block multiplication, $A = LU$ has $A_k = L_k U_k$ in the top left corner:

$$A = \begin{bmatrix} A_k & * \\ * & * \end{bmatrix} = \begin{bmatrix} L_k & 0 \\ * & * \end{bmatrix} \begin{bmatrix} U_k & * \\ 0 & * \end{bmatrix}.$$

(a) Suppose the first three pivots of A are $2, 3, -1$. What are the determinants of L_1, L_2, L_3 (with diagonal 1's) and U_1, U_2, U_3 and A_1, A_2, A_3?

(b) If A_1, A_2, A_3 have determinants $5, 6, 7$ find the three pivots from equation (3).

25 Block elimination subtracts CA^{-1} times the first row $[\,A \quad B\,]$ from the second row $[C \quad D]$. This leaves the *Schur complement* $D - CA^{-1}B$ in the corner:

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & D - CA^{-1}B \end{bmatrix}.$$

Take determinants of these block matrices to prove correct rules if A^{-1} exists:

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A|\,|D - CA^{-1}B| \quad = \quad |AD - CB| \text{ provided } AC = CA.$$

26 If A is m by n and B is n by m, block multiplication gives $\det M = \det AB$:

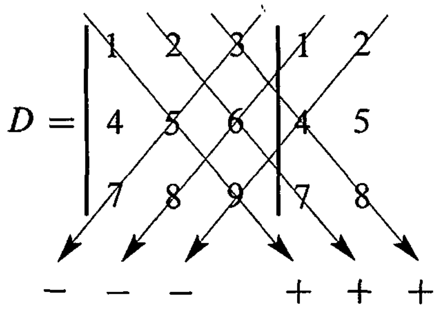
$$M = \begin{bmatrix} 0 & A \\ -B & I \end{bmatrix} = \begin{bmatrix} AB & A \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ -B & I \end{bmatrix}.$$

If A is a single row and B is a single column what is $\det M$? If A is a column and B is a row what is $\det M$? Do a 3 by 3 example of each.

27 (A calculus question) Show that the derivative of $\det A$ with respect to a_{11} is the cofactor C_{11}. The other entries are fixed—we are only changing a_{11}.

Problems 28–33 are about the "big formula" with $n!$ terms.

28 A 3 by 3 determinant has three products "down to the right" and three "down to the left" with minus signs. Compute the six terms like $(1)(5)(9) = 45$ to find D.



Explain without determinants why this particular matrix is or is not invertible.

29 For E_4 in Problem 15, five of the $4! = 24$ terms in the big formula (8) are nonzero. Find those five terms to show that $E_4 = -1$.

30 For the 4 by 4 tridiagonal second difference matrix (entries -1, 2, -1) find the five terms in the big formula that give $\det A = 16 - 4 - 4 - 4 + 1$.

31 Find the determinant of this cyclic P by cofactors of row 1 and then the "big formula". How many exchanges reorder $4, 1, 2, 3$ into $1, 2, 3, 4$? Is $|P^2| = 1$ or -1?

$$P = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad P^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}.$$

Challenge Problems

32 Cofactors of the $1, 3, 1$ matrices in Problem 21 give a recursion $S_n = 3S_{n-1} - S_{n-2}$. Amazingly that recursion produces every second Fibonacci number. Here is the challenge.

Show that S_n is the Fibonacci number F_{2n+2} by proving $F_{2n+2} = 3F_{2n} - F_{2n-2}$. Keep using Fibonacci's rule $F_k = F_{k-1} + F_{k-2}$ starting with $k = 2n + 2$.

33 The symmetric Pascal matrices have determinant 1. If I subtract 1 from the n, n entry, why does the determinant become zero? (Use rule 3 or cofactors.)

$$\det \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{bmatrix} = 1 \text{ (known)} \qquad \det \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & \mathbf{19} \end{bmatrix} = \mathbf{0} \text{ (to explain).}$$

34 This problem shows in two ways that $\det A = 0$ (the x's are any numbers):

$$A = \begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \\ 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & x & x \end{bmatrix}.$$

(a) How do you know that the rows are linearly dependent?

(b) Explain why all 120 terms are zero in the big formula for $\det A$.

35 If $|\det(A)| > 1$, prove that the powers A^n cannot stay bounded. But if $|\det(A)| \leq 1$, show that some entries of A^n might still grow large. Eigenvalues will give the right test for stability, determinants tell us only one number.

5.3 Cramer's Rule, Inverses, and Volumes

This section solves $Ax = b$—by algebra and not by elimination. We also invert A. In the entries of A^{-1}, you will see $\det A$ in every denominator—we divide by it. (If $\det A = 0$ then we can't divide and A^{-1} doesn't exist.) Each entry in A^{-1} and $A^{-1}b$ is a determinant divided by the determinant of A.

Cramer's Rule solves $Ax = b$. A neat idea gives the first component x_1. Replacing the first column of I by x gives a matrix with determinant x_1. When you multiply it by A, the first column becomes Ax which is b. The other columns are copied from A:

$$\textbf{Key idea} \qquad \begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} x_1 & 0 & 0 \\ x_2 & 1 & 0 \\ x_3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{bmatrix} = B_1. \tag{1}$$

We multiplied a column at a time. *Take determinants of the three matrices*:

$$\textbf{Product rule} \qquad (\det A)(x_1) = \det B_1 \qquad \text{or} \qquad \boldsymbol{x_1 = \frac{\det B_1}{\det A}}. \tag{2}$$

This is the first component of x in Cramer's Rule! Changing a column of A gives B_1.

To find x_2, put the vector x into the *second* column of the identity matrix:

$$\textbf{Same idea} \qquad \begin{bmatrix} & & \\ a_1 & a_2 & a_3 \\ & & \end{bmatrix} \begin{bmatrix} 1 & x_1 & 0 \\ 0 & x_2 & 0 \\ 0 & x_3 & 1 \end{bmatrix} = \begin{bmatrix} & & \\ a_1 & b & a_3 \\ & & \end{bmatrix} = B_2. \tag{3}$$

Take determinants to find $(\det A)(x_2) = \det B_2$. This gives x_2 in Cramer's Rule:

CRAMER's RULE If $\det A$ is not zero, $Ax = b$ is solved by determinants:

$$x_1 = \frac{\det B_1}{\det A} \qquad x_2 = \frac{\det B_2}{\det A} \qquad \cdots \qquad x_n = \frac{\det B_n}{\det A} \tag{4}$$

The matrix B_j has the jth column of A replaced by the vector b.

Example 1 Solving $3x_1 + 4x_2 = \mathbf{2}$ and $5x_1 + 6x_2 = \mathbf{4}$ needs three determinants:

$$\det A = \begin{vmatrix} 3 & 4 \\ 5 & 6 \end{vmatrix} \qquad \det B_1 = \begin{vmatrix} \mathbf{2} & 4 \\ \mathbf{4} & 6 \end{vmatrix} \qquad \det B_2 = \begin{vmatrix} 3 & \mathbf{2} \\ 5 & \mathbf{4} \end{vmatrix}$$

Those determinants are -2 and -4 and 2. All ratios divide by $\det A$:

$$\textbf{Cramer's Rule} \quad x_1 = \frac{-4}{-2} = 2 \quad x_2 = \frac{2}{-2} = -1 \quad \textbf{check} \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}.$$

To solve an n by n system, Cramer's Rule evaluates $n + 1$ determinants (of A and the n different B's). When each one is the sum of $n!$ terms—applying the "big formula" with all permutations—this makes a total of $(n + 1)!$ terms. *It would be crazy to solve equations that way.* But we do finally have an explicit formula for the solution x.

Example 2 Cramer's Rule is inefficient for numbers but it is well suited to letters. For $n = 2$, find the columns of A^{-1} by solving $AA^{-1} = I$:

$$\textbf{Columns of } I \qquad \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Those share the same A. We need five determinants for $x_1,\ x_2,\ y_1,\ y_2$:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \text{ and } \begin{vmatrix} \mathbf{1} & b \\ \mathbf{0} & d \end{vmatrix} \quad \begin{vmatrix} a & \mathbf{1} \\ c & \mathbf{0} \end{vmatrix} \quad \begin{vmatrix} \mathbf{0} & b \\ \mathbf{1} & d \end{vmatrix} \quad \begin{vmatrix} a & \mathbf{0} \\ c & \mathbf{1} \end{vmatrix}$$

The last four are $d, -c, -b$, and a. (They are the cofactors!) Here is A^{-1}:

$$x_1 = \frac{d}{|A|},\ x_2 = \frac{-c}{|A|},\ y_1 = \frac{-b}{|A|},\ y_2 = \frac{a}{|A|}, \text{ and then } A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

I chose 2 by 2 so that the main points could come through clearly. The new idea is the appearance of the cofactors. When the right side is a column of the identity matrix I, the determinant of each matrix B_j in Cramer's Rule is a cofactor.

You can see those cofactors for $n = 3$. Solve $AA^{-1} = I$ (first column only):

$$\begin{matrix}\textbf{Determinants} \\ \textbf{= Cofactors of } A\end{matrix} \qquad \begin{vmatrix} \mathbf{1} & a_{12} & a_{13} \\ \mathbf{0} & a_{22} & a_{23} \\ \mathbf{0} & a_{32} & a_{33} \end{vmatrix} \quad \begin{vmatrix} a_{11} & \mathbf{1} & a_{13} \\ a_{21} & \mathbf{0} & a_{23} \\ a_{31} & \mathbf{0} & a_{33} \end{vmatrix} \quad \begin{vmatrix} a_{11} & a_{12} & \mathbf{1} \\ a_{21} & a_{22} & \mathbf{0} \\ a_{31} & a_{32} & \mathbf{0} \end{vmatrix} \qquad (5)$$

That first determinant $|B_1|$ is the cofactor C_{11}. The second determinant $|B_2|$ is the cofactor C_{12}. *Notice that* the correct minus sign appears in $-(a_{21}a_{33} - a_{23}a_{31})$. This cofactor C_{12} goes into the $2, 1$ entry of A^{-1}—the first column! So we transpose the cofactor matrix, and as always we divide by $\det A$.

The i, j entry of A^{-1} is the cofactor C_{ji} (not C_{ij}) divided by $\det A$:

$$\textbf{FORMULA FOR } A^{-1} \qquad (A^{-1})_{ij} = \frac{C_{ji}}{\det A} \quad \text{and} \quad A^{-1} = \frac{C^{\mathrm{T}}}{\det A}. \qquad (6)$$

The cofactors C_{ij} go into the *"cofactor matrix"* C. ***Its transpose leads to*** A^{-1}. To compute the i, j entry of A^{-1}, cross out row j and column i of A. Multiply the determinant by $(-1)^{i+j}$ to get the cofactor, and divide by $\det A$.

Check this rule for the $3, 1$ entry of A^{-1}. This is in column 1 so we solve $Ax = (1, 0, 0)$. The third component x_3 needs the third determinant in equation (5), divided by $\det A$. That third determinant is exactly the cofactor $C_{13} = a_{21}a_{32} - a_{22}a_{31}$. So $(A^{-1})_{31} = C_{13}/\det A$ (2 by 2 determinant divided by 3 by 3).

Summary In solving $AA^{-1} = I$, the columns of I lead to the columns of A^{-1}. Then Cramer's Rule using $\boldsymbol{b}$ = columns of I gives the short formula (6) for A^{-1}.

Direct proof of the formula $A^{-1} = C^{\mathrm{T}}/\det A$ The idea is to multiply A times C^{T}:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix} = \begin{bmatrix} \det A & 0 & 0 \\ 0 & \det A & 0 \\ 0 & 0 & \det A \end{bmatrix}. \tag{7}$$

Row 1 of A times column 1 of the cofactors yields the first $\det A$ on the right:

$$a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} = \det A \quad \text{by the cofactor rule.}$$

Similarly row 2 of A times column 2 of C^{T}(*transpose*) yields $\det A$. The entries a_{2j} are multiplying cofactors C_{2j} as they should, to give the determinant.

How to explain the zeros off the main diagonal in equation (7)? Rows of A are multiplying cofactors from *different* rows. Why is the answer zero?

Row 2 of A
Row 1 of C

$$a_{21}C_{11} + a_{22}C_{12} + a_{23}C_{13} = 0. \tag{8}$$

Answer: This is the cofactor rule for a new matrix, when the second row of A is copied into its first row. The new matrix A^* has two equal rows, so $\det A^* = 0$ in equation (8). Notice that A^* has the same cofactors C_{11}, C_{12}, C_{13} as A—because all rows agree after the first row. Thus the remarkable multiplication (7) is correct:

$$AC^{\mathrm{T}} = (\det A)I \qquad \text{or} \qquad A^{-1} = \frac{C^{\mathrm{T}}}{\det A}.$$

Example 3 The "sum matrix" A has determinant 1. Then A^{-1} contains cofactors:

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \text{has inverse} \quad A^{-1} = \frac{C^{\mathrm{T}}}{1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}.$$

Cross out row 1 and column 1 of A to see the 3 by 3 cofactor $C_{11} = 1$. Now cross out row 1 and column 2 for C_{12}. The 3 by 3 submatrix is still triangular with determinant 1. But the cofactor C_{12} is -1 because of the sign $(-1)^{1+2}$. This number -1 goes into the $(2, 1)$ entry of A^{-1}—don't forget to transpose C.

The inverse of a triangular matrix is triangular. Cofactors give a reason why.

Example 4 If all cofactors are nonzero, is A sure to be invertible? *No way.*

Area of a Triangle

Everybody knows the area of a rectangle—base times height. The area of a triangle is *half* the base times the height. But here is a question that those formulas don't answer. ***If we know the corners (x_1, y_1) and (x_2, y_2) and (x_3, y_3) of a triangle, what is the area?*** Using the corners to find the base and height is not a good way.

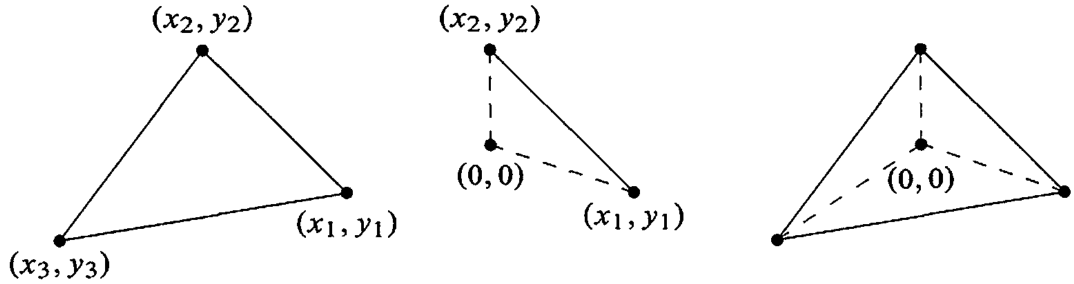


Figure 5.1: General triangle; special triangle from $(0,0)$; general from three specials.

Determinants are much better. The square roots in the base and height cancel out in the good formula. ***The area of a triangle is half of a 3 by 3 determinant.*** If one corner is at the origin, say $(x_3, y_3) = (0,0)$, the determinant is only 2 by 2.

The triangle with corners (x_1, y_1) and (x_2, y_2) and (x_3, y_3) has area $= \dfrac{\text{determinant}}{2}$:

Area of triangle $\frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$ Area $= \frac{1}{2}\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}$ when $(x_3, y_3) = (0,0)$.

When you set $x_3 = y_3 = 0$ in the 3 by 3 determinant, you get the 2 by 2 determinant. These formulas have no square roots—they are reasonable to memorize. The 3 by 3 determinant breaks into a sum of three 2 by 2's, just as the third triangle in Figure 5.1 breaks into three special triangles from $(0,0)$:

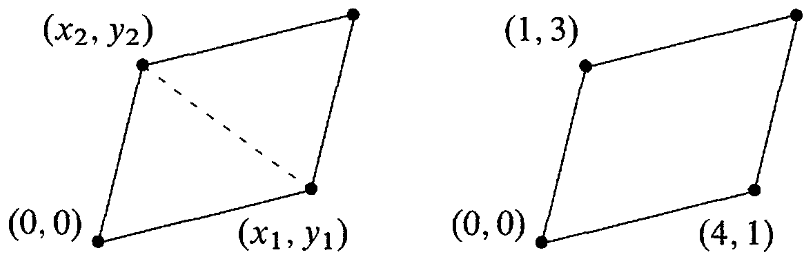
Cofactors of column 3

$$\textbf{Area} = \frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = \begin{matrix} +\frac{1}{2}(x_1y_2 - x_2y_1) \\ +\frac{1}{2}(x_2y_3 - x_3y_2) \\ +\frac{1}{2}(x_3y_1 - x_1y_3). \end{matrix} \tag{9}$$

If $(0,0)$ is outside the triangle, two of the special areas can be negative—but the sum is still correct. The real problem is to explain the special area $\frac{1}{2}(x_1y_2 - x_2y_1)$.

Why is this the area of a triangle? We can remove the factor $\frac{1}{2}$ and change to a parallelogram (twice as big, because the parallelogram contains two equal triangles). We now prove that the parallelogram area is the determinant $x_1y_2 - x_2y_1$. This area in Figure 5.2 is 11, and therefore the triangle has area $\frac{11}{2}$.

Proof that a parallelogram starting from $(0,0)$ has area = 2 by 2 determinant.



Parallelogram

$$\text{Area} = \begin{vmatrix} 4 & 1 \\ 1 & 3 \end{vmatrix} = 11$$

$$\text{Triangle: Area} = \tfrac{11}{2}$$

Figure 5.2: A triangle is half of a parallelogram. Area is half of a determinant.

There are many proofs but this one fits with the book. We show that the area has the same properties 1-2-3 as the determinant. Then area = determinant! Remember that those three rules defined the determinant and led to all its other properties.

1 When $A = I$, the parallelogram becomes the unit square. Its area is $\det I = 1$.

2 When rows are exchanged, the determinant reverses sign. The absolute value (positive area) stays the same—it is the same parallelogram.

3 If row 1 is multiplied by t, Figure 5.3a shows that the area is also multiplied by t. Suppose a new row (x_1', y_1') is added to (x_1, y_1) (keeping row 2 fixed). Figure 5.3b shows that the solid parallelogram areas add to the dotted parallelogram area (because the two triangles completed by dotted lines are the same).

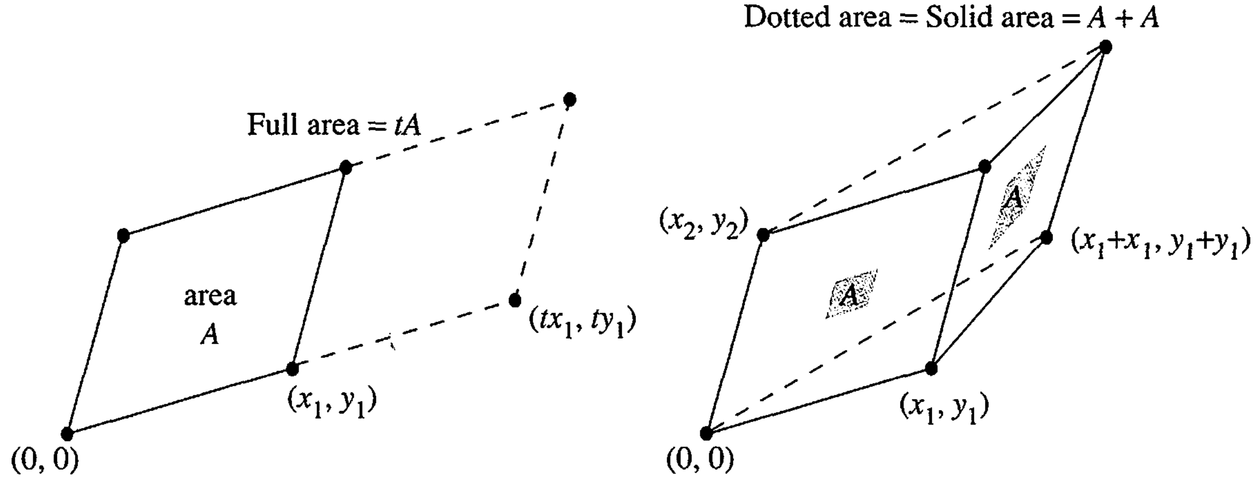


Figure 5.3: Areas obey the rule of linearity (keeping the side (x_2, y_2) constant).

That is an exotic proof, when we could use plane geometry. But the proof has a major attraction—it applies in n dimensions. The n edges going out from the origin are given by the *rows of an n by n matrix*. The box is completed by more edges, just like the parallelogram.

Figure 5.4 shows a three-dimensional box—whose edges are not at right angles. ***The volume equals the absolute value of*** $\det A$. Our proof checks again that rules 1–3 for determinants are also obeyed by volumes. When an edge is stretched by a factor t, the volume is multiplied by t. When edge 1 is added to edge 1′, the new box has edge 1 + 1′. Its volume is the sum of the two original volumes. This is Figure 5.3b lifted into three dimensions or n dimensions. I would draw the boxes but this paper is only two-dimensional.

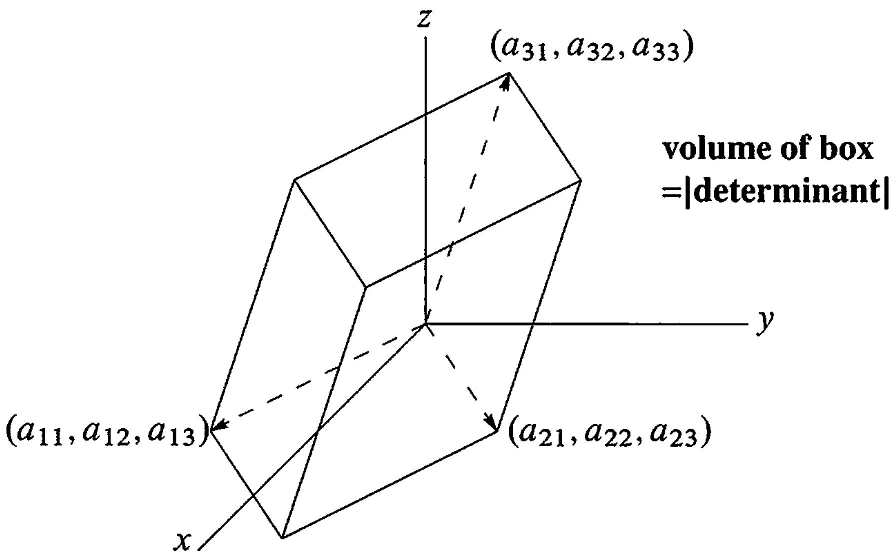


Figure 5.4: Three-dimensional box formed from the three rows of A.

The unit cube has volume = 1, which is $\det I$. Row exchanges or edge exchanges leave the same box and the same absolute volume. The determinant changes sign, to indicate whether the edges are a *right-handed triple* ($\det A > 0$) or a *left-handed triple* ($\det A < 0$). The box volume follows the rules for determinants, so volume of the box = absolute value of the determinant.

Example 5 Suppose a rectangular box (90° angles) has side lengths r, s, and t. Its volume is r times s times t. The diagonal matrix with entries r, s, and t produces those three sides. Then $\det A$ also equals $r\,s\,t$.

Example 6 In calculus, the box is infinitesimally small! To integrate over a circle, we might change x and y to r and θ. Those are polar coordinates: $x = r\cos\theta$ and $y = r\sin\theta$. The area of a "polar box" is a determinant J times $dr\,d\theta$:

$$J = \begin{vmatrix} \partial x/\partial r & \partial x/\partial\theta \\ \partial y/\partial r & \partial y/\partial\theta \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r.$$

This determinant is the r in the small area $dA = r\,dr\,d\theta$. The stretching factor J goes into double integrals just as dx/du goes into an ordinary integral $\int dx = \int (dx/du)\,du$. For triple integrals the Jacobian matrix J with nine derivatives will be 3 by 3.

The Cross Product

The *cross product* is an extra (and optional) application, special for three dimensions. Start with vectors $\boldsymbol{u} = (u_1, u_2, u_3)$ and $\boldsymbol{v} = (v_1, v_2, v_3)$. Unlike the dot product, which is a number, the cross product is a vector—also in three dimensions. It is written $\boldsymbol{u} \times \boldsymbol{v}$ and pronounced "$\boldsymbol{u}$ cross $\boldsymbol{v}$." *The components of this cross product are just* 2 *by* 2 *cofactors.* We will explain the properties that make $\boldsymbol{u} \times \boldsymbol{v}$ useful in geometry and physics.

This time we bite the bullet, and write down the formula before the properties.

DEFINITION The *cross product* of $\boldsymbol{u} = (u_1, u_2, u_3)$ and $\boldsymbol{v} = (v_1, v_2, v_3)$ is a vector

$$\boldsymbol{u} \times \boldsymbol{v} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = (u_2v_3 - u_3v_2)\boldsymbol{i} + (u_3v_1 - u_1v_3)\boldsymbol{j} + (u_1v_2 - u_2v_1)\boldsymbol{k}. \quad (10)$$

This vector is perpendicular to u and v. The cross product $\boldsymbol{v} \times \boldsymbol{u}$ is $-(\boldsymbol{u} \times \boldsymbol{v})$.

Comment The 3 by 3 determinant is the easiest way to remember $\boldsymbol{u} \times \boldsymbol{v}$. It is not especially legal, because the first row contains vectors $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ and the other rows contain numbers. In the determinant, the vector $\boldsymbol{i} = (1, 0, 0)$ multiplies u_2v_3 and $-u_3v_2$. The result is $(u_2v_3 - u_3v_2, 0, 0)$, which displays the first component of the cross product.

Notice the cyclic pattern of the subscripts: 2 and 3 give component 1 of $\boldsymbol{u} \times \boldsymbol{v}$, then 3 and 1 give component 2, then 1 and 2 give component 3. This completes the definition of $\boldsymbol{u} \times \boldsymbol{v}$. Now we list the properties of the cross product:

Property 1 $\boldsymbol{v} \times \boldsymbol{u}$ reverses rows 2 and 3 in the determinant so it equals $-(\boldsymbol{u} \times \boldsymbol{v})$.

Property 2 The cross product $\boldsymbol{u} \times \boldsymbol{v}$ is perpendicular to $\boldsymbol{u}$ (and also to $\boldsymbol{v}$). The direct proof is to watch terms cancel. Perpendicularity is a zero dot product:

$$\boldsymbol{u} \cdot (\boldsymbol{u} \times \boldsymbol{v}) = u_1(u_2v_3 - u_3v_2) + u_2(u_3v_1 - u_1v_3) + u_3(u_1v_2 - u_2v_1) = 0. \quad (11)$$

The determinant now has rows $\boldsymbol{u}, \boldsymbol{u}$ and $\boldsymbol{v}$ so it is zero.

Property 3 The cross product of any vector with itself (two equal rows) is $\boldsymbol{u} \times \boldsymbol{u} = \boldsymbol{0}$.

When $\boldsymbol{u}$ and $\boldsymbol{v}$ are parallel, the cross product is zero. When $\boldsymbol{u}$ and $\boldsymbol{v}$ are perpendicular, the dot product is zero. One involves $\sin\theta$ and the other involves $\cos\theta$:

$$\|\boldsymbol{u} \times \boldsymbol{v}\| = \|\boldsymbol{u}\|\,\|\boldsymbol{v}\|\,|\sin\theta| \quad \text{and} \quad |\boldsymbol{u} \cdot \boldsymbol{v}| = \|\boldsymbol{u}\|\,\|\boldsymbol{v}\|\,|\cos\theta|. \quad (12)$$

Example 7 Since $\boldsymbol{u} = (3, 2, 0)$ and $\boldsymbol{v} = (1, 4, 0)$ are in the xy plane, $\boldsymbol{u} \times \boldsymbol{v}$ goes up the z axis:

$$\boldsymbol{u} \times \boldsymbol{v} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 3 & 2 & 0 \\ 1 & 4 & 0 \end{vmatrix} = 10\boldsymbol{k}. \quad \text{The cross product is } \boldsymbol{u} \times \boldsymbol{v} = (0, 0, 10).$$

The length of $u \times v$ equals the area of the parallelogram with sides u and v. This will be important: In this example the area is 10.

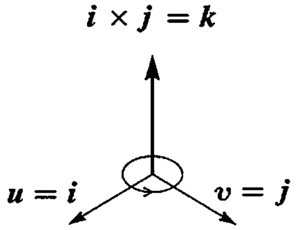
Example 8 The cross product of $\boldsymbol{u} = (1,1,1)$ and $\boldsymbol{v} = (1,1,2)$ is $(1,-1,0)$:

$$\begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{vmatrix} = \boldsymbol{i}\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} - \boldsymbol{j}\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} + \boldsymbol{k}\begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = \boldsymbol{i} - \boldsymbol{j}.$$

This vector $(1,-1,0)$ is perpendicular to $(1,1,1)$ and $(1,1,2)$ as predicted. Area $= \sqrt{2}$.

Example 9 The cross product of $(1,0,0)$ and $(0,1,0)$ obeys the *right hand rule*. It goes up not down:

$$\begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = \boldsymbol{k}$$



Rule $\boldsymbol{u} \times \boldsymbol{v}$ points along your right thumb when the fingers curl from $\boldsymbol{u}$ to $\boldsymbol{v}$.

Thus $\boldsymbol{i} \times \boldsymbol{j} = \boldsymbol{k}$. The right hand rule also gives $\boldsymbol{j} \times \boldsymbol{k} = \boldsymbol{i}$ and $\boldsymbol{k} \times \boldsymbol{i} = \boldsymbol{j}$. Note the cyclic order. In the opposite order (anti-cyclic) the thumb is reversed and the cross product goes the other way: $\boldsymbol{k} \times \boldsymbol{j} = -\boldsymbol{i}$ and $\boldsymbol{i} \times \boldsymbol{k} = -\boldsymbol{j}$ and $\boldsymbol{j} \times \boldsymbol{i} = -\boldsymbol{k}$. You see the three plus signs and three minus signs from a 3 by 3 determinant.

The definition of $\boldsymbol{u} \times \boldsymbol{v}$ can be based on vectors instead of their components:

DEFINITION The ***cross product*** is a vector with length $\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|\,|\sin\theta|$. Its direction is perpendicular to $\boldsymbol{u}$ and $\boldsymbol{v}$. It points "up" or "down" by the right hand rule.

This definition appeals to physicists, who hate to choose axes and coordinates. They see (u_1, u_2, u_3) as the position of a mass and (F_x, F_y, F_z) as a force acting on it. If $\boldsymbol{F}$ is parallel to $\boldsymbol{u}$, then $\boldsymbol{u} \times \boldsymbol{F} = \boldsymbol{0}$—there is no turning. The cross product $\boldsymbol{u} \times \boldsymbol{F}$ is the turning force or *torque*. It points along the turning axis (perpendicular to $\boldsymbol{u}$ and $\boldsymbol{F}$). Its length $\|\boldsymbol{u}\|\,\|\boldsymbol{F}\|\sin\theta$ measures the "moment" that produces turning.

Triple Product = Determinant = Volume

Since $\boldsymbol{u} \times \boldsymbol{v}$ is a vector, we can take its dot product with a third vector $\boldsymbol{w}$. That produces the ***triple product*** $(\boldsymbol{u} \times \boldsymbol{v}) \cdot \boldsymbol{w}$. It is called a "scalar" triple product, because it is a number. In fact it is a determinant—it gives the volume of the $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ box:

Triple product

$$(\boldsymbol{u} \times \boldsymbol{v}) \cdot \boldsymbol{w} = \begin{vmatrix} w_1 & w_2 & w_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}. \tag{13}$$

We can put w in the top or bottom row. The two determinants are the same because _____ row exchanges go from one to the other. Notice when this determinant is zero:

$$(u \times v) \cdot w = 0 \quad \text{exactly when the vectors } u, v, w \text{ lie in the } \textit{same plane.}$$

First reason $u \times v$ is perpendicular to that plane so its dot product with w is zero.

Second reason Three vectors in a plane are dependent. The matrix is singular (det $= 0$).

Third reason Zero volume when the u, v, w box is squashed onto a plane.

It is remarkable that $(u \times v) \cdot w$ equals the volume of the box with sides u, v, w. This 3 by 3 determinant carries tremendous information. Like $ad - bc$ for a 2 by 2 matrix, it separates invertible from singular. Chapter 6 will be looking for singular.

■ REVIEW OF THE KEY IDEAS ■

1. Cramer's Rule solves $Ax = b$ by ratios like $x_1 = |B_1|/|A| = |b\, a_2 \cdots a_n|/|A|$.
2. When C is the cofactor matrix for A, the inverse is $A^{-1} = C^{\mathrm{T}}/\det A$.
3. The volume of a box is $|\det A|$, when the box edges are the rows of A.
4. Area and volume are needed to change variables in double and triple integrals.
5. In $\mathbf{R}^3$, the cross product $u \times v$ is perpendicular to u and v.

■ WORKED EXAMPLES ■

5.3 A If A is singular, the equation $AC^{\mathrm{T}} = (\det A)I$ becomes $AC^{\mathrm{T}} =$ **zero matrix.** *Then each column of C^{T} is in the nullspace of A.* Those columns contain cofactors along rows of A. So the cofactors quickly find the nullspace of a 3 by 3 matrix—my apologies that this comes so late!

Solve $Ax = \mathbf{0}$ by $x =$ cofactors along a row, for these singular matrices of rank 2:

Cofactors give Nullspace
$$A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 3 & 9 \\ 2 & 2 & 8 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Any nonzero column of C^{T} will give the desired solution to $Ax = \mathbf{0}$. With rank 2, A has at least one nonzero cofactor. If A has rank 1 we get $x = \mathbf{0}$ and the idea fails.

Solution The first matrix has these cofactors along its top row (note each minus sign):

$$\begin{vmatrix} 3 & 9 \\ 2 & 8 \end{vmatrix} = 6 \qquad -\begin{vmatrix} 2 & 9 \\ 2 & 8 \end{vmatrix} = 2 \qquad \begin{vmatrix} 2 & 3 \\ 2 & 2 \end{vmatrix} = -2$$

Then $\boldsymbol{x} = (6, 2, -2)$ solves $A\boldsymbol{x} = \mathbf{0}$. The cofactors along the second row are $(-18, -6, 6)$ which is just $-3\boldsymbol{x}$. This is also in the one-dimensional nullspace of A.

The second matrix has *zero cofactors* along its first row. The nullvector $\boldsymbol{x} = (0, 0, 0)$ is not interesting. The cofactors of row 2 give $\boldsymbol{x} = (1, -1, 0)$ which solves $A\boldsymbol{x} = \mathbf{0}$.

Every n by n matrix of rank $n - 1$ has at least one nonzero cofactor by Problem 3.3.12. But for rank $n - 2$, all cofactors are zero and we only find $\boldsymbol{x} = \mathbf{0}$.

5.3 B Use Cramer's Rule with ratios $\det B_j / \det A$ to solve $A\boldsymbol{x} = \boldsymbol{b}$. Also find the inverse matrix $A^{-1} = C^{\mathrm{T}} / \det A$. Why is the solution $\boldsymbol{x}$ for this $\boldsymbol{b}$ the same as column 3 of A^{-1}? Which cofactors are involved in computing that column $\boldsymbol{x}$?

$$A\boldsymbol{x} = \boldsymbol{b} \quad \text{is} \quad \begin{bmatrix} 2 & 6 & 2 \\ 1 & 4 & 2 \\ 5 & 9 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Find the volumes of the boxes whose edges are *columns* of A and then rows of A^{-1}.

Solution The determinants of the B_j (with right side $\boldsymbol{b}$ placed in column j) are

$$|B_1| = \begin{vmatrix} \mathbf{0} & 6 & 2 \\ \mathbf{0} & 4 & 2 \\ \mathbf{1} & 9 & 0 \end{vmatrix} = 4 \qquad |B_2| = \begin{vmatrix} 2 & \mathbf{0} & 2 \\ 1 & \mathbf{0} & 2 \\ 5 & \mathbf{1} & 0 \end{vmatrix} = -2 \qquad |B_3| = \begin{vmatrix} 2 & 6 & \mathbf{0} \\ 1 & 4 & \mathbf{0} \\ 5 & 9 & \mathbf{1} \end{vmatrix} = 2.$$

Those are cofactors C_{31}, C_{32}, C_{33} of row 3. Their dot product with row 3 is $\det A$:

$$\det A = a_{31}C_{31} + a_{32}C_{32} + a_{33}C_{33} = (5, 9, 0) \cdot (4, -2, 2) = 2.$$

The three ratios $\det B_j / \det A$ give the three components of $\boldsymbol{x} = (2, -1, 1)$. This $\boldsymbol{x}$ is the third column of A^{-1} because $\boldsymbol{b} = (0, 0, 1)$ is the third column of I. The cofactors along the other rows of A, divided by $\det A = 2$, give the other *columns* of A^{-1}:

$$A^{-1} = \frac{C^{\mathrm{T}}}{\det A} = \frac{1}{2} \begin{bmatrix} -18 & 18 & 4 \\ 10 & -10 & -2 \\ -11 & 12 & 2 \end{bmatrix}. \quad \text{Multiply to check} \quad AA^{-1} = I$$

The box from the columns of A has volume $= \det A = 2$ (the same as the box from the rows, since $|A^{\mathrm{T}}| = |A|$). The box from A^{-1} has volume $1/|A| = \frac{1}{2}$.

Problem Set 5.3

Problems 1–5 are about Cramer's Rule for $x = A^{-1}b$.

1 Solve these linear equations by Cramer's Rule $x_j = \det B_j / \det A$:

$$\text{(a)} \quad \begin{aligned} 2x_1 + 5x_2 &= 1 \\ x_1 + 4x_2 &= 2 \end{aligned} \qquad \text{(b)} \quad \begin{aligned} 2x_1 + \ x_2 \qquad &= 1 \\ x_1 + 2x_2 + \ x_3 &= 0 \\ x_2 + 2x_3 &= 0. \end{aligned}$$

2 Use Cramer's Rule to solve for y (only). Call the 3 by 3 determinant D:

$$\text{(a)} \quad \begin{aligned} ax + by &= 1 \\ cx + dy &= 0 \end{aligned} \qquad \text{(b)} \quad \begin{aligned} ax + by + \ cz &= 1 \\ dx + ey + fz &= 0 \\ gx + hy + \ iz &= 0. \end{aligned}$$

3 Cramer's Rule breaks down when $\det A = 0$. Example (a) has no solution while (b) has infinitely many. What are the ratios $x_j = \det B_j / \det A$ in these two cases?

$$\text{(a)} \quad \begin{aligned} 2x_1 + 3x_2 &= 1 \\ 4x_1 + 6x_2 &= 1 \end{aligned} \quad \text{(parallel lines)} \qquad \text{(b)} \quad \begin{aligned} 2x_1 + 3x_2 &= 1 \\ 4x_1 + 6x_2 &= 2 \end{aligned} \quad \text{(same line)}$$

4 *Quick proof of Cramer's rule.* The determinant is a linear function of column 1. It is zero if two columns are equal. When $\boldsymbol{b} = A\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + x_3\boldsymbol{a}_3$ goes into the first column of A, the determinant of this matrix B_1 is

$$|\boldsymbol{b} \quad \boldsymbol{a}_2 \quad \boldsymbol{a}_3| = |x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + x_3\boldsymbol{a}_3 \quad \boldsymbol{a}_2 \quad \boldsymbol{a}_3| = x_1|\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \boldsymbol{a}_3| = x_1 \det A.$$

(a) What formula for x_1 comes from left side = right side?

(b) What steps lead to the middle equation?

5 If the right side $\boldsymbol{b}$ is the first column of A, solve the 3 by 3 system $A\boldsymbol{x} = \boldsymbol{b}$. How does each determinant in Cramer's Rule lead to this solution $\boldsymbol{x}$?

Problems 6–15 are about $A^{-1} = C^{\mathrm{T}} / \det A$. Remember to transpose C.

6 Find A^{-1} from the cofactor formula $C^{\mathrm{T}} / \det A$. Use symmetry in part (b).

$$\text{(a)}\ A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 0 \\ 0 & 7 & 1 \end{bmatrix} \qquad \text{(b)} \quad A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

7 If all the cofactors are zero, how do you know that A has no inverse? If none of the cofactors are zero, is A sure to be invertible?

8 Find the cofactors of A and multiply AC^{T} to find $\det A$:

$$A = \begin{bmatrix} 1 & 1 & 4 \\ 1 & 2 & 2 \\ 1 & 2 & 5 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 6 & -3 & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} \quad \text{and } AC^{\mathrm{T}} = \underline{\qquad}.$$

If you change that 4 to 100, why is $\det A$ unchanged?

9 Suppose $\det A = 1$ and you know all the cofactors in C. How can you find A?

10 From the formula $AC^{\mathrm{T}} = (\det A)I$ show that $\det C = (\det A)^{n-1}$.

11 If all entries of A are integers, and $\det A = 1$ or -1, prove that all entries of A^{-1} are integers. Give a 2 by 2 example with no zero entries.

12 If all entries of A and A^{-1} are integers, prove that $\det A = 1$ or -1. Hint: What is $\det A$ times $\det A^{-1}$?

13 Complete the calculation of A^{-1} by cofactors that was started in Example 5.

14 L is lower triangular and S is symmetric. Assume they are invertible:

To invert
triangular L
symmetric S

$$L = \begin{bmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{bmatrix} \qquad S = \begin{bmatrix} a & b & d \\ b & c & e \\ d & e & f \end{bmatrix}.$$

(a) Which three cofactors of L are zero? Then L^{-1} is also lower triangular.

(b) Which three pairs of cofactors of S are equal? Then S^{-1} is also symmetric.

(c) The cofactor matrix C of an orthogonal Q will be _____. *Why*?

15 For $n = 5$ the matrix C contains _____ cofactors. Each 4 by 4 cofactor contains _____ terms and each term needs _____ multiplications. Compare with $5^3 = 125$ for the Gauss-Jordan computation of A^{-1} in Section 2.4.

Problems 16–26 are about area and volume by determinants.

16 (a) Find the area of the parallelogram with edges $\boldsymbol{v} = (3,2)$ and $\boldsymbol{w} = (1,4)$.

(b) Find the area of the triangle with sides $\boldsymbol{v}$, $\boldsymbol{w}$, and $\boldsymbol{v} + \boldsymbol{w}$. Draw it.

(c) Find the area of the triangle with sides $\boldsymbol{v}$, $\boldsymbol{w}$, and $\boldsymbol{w} - \boldsymbol{v}$. Draw it.

17 A box has edges from $(0,0,0)$ to $(3,1,1)$ and $(1,3,1)$ and $(1,1,3)$. Find its volume. Also find the area of each parallelogram face using $\|\boldsymbol{u} \times \boldsymbol{v}\|$.

18 (a) The corners of a triangle are $(2,1)$ and $(3,4)$ and $(0,5)$. What is the area?

(b) Add a corner at $(-1,0)$ to make a lopsided region (four sides). Find the area.

19 The parallelogram with sides $(2,1)$ and $(2,3)$ has the same area as the parallelogram with sides $(2,2)$ and $(1,3)$. Find those areas from 2 by 2 determinants and say why they must be equal. (I can't see why from a picture. Please write to me if you do.)

20 The Hadamard matrix H has orthogonal rows. The box is a hypercube!

$$\text{What is} \quad |H| = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{vmatrix} = \text{volume of a hypercube in } \mathbf{R}^4?$$

21 If the columns of a 4 by 4 matrix have lengths L_1, L_2, L_3, L_4, what is the largest possible value for the determinant (based on volume)? If all entries of the matrix are 1 or -1, what are those lengths and the maximum determinant?

22 Show by a picture how a rectangle with area $x_1 y_2$ minus a rectangle with area $x_2 y_1$ produces the same area as our parallelogram.

23 When the edge vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ are perpendicular, the volume of the box is $\|\boldsymbol{a}\|$ times $\|\boldsymbol{b}\|$ times $\|\boldsymbol{c}\|$. The matrix $A^{\mathrm{T}}A$ is _____. Find det $A^{\mathrm{T}}A$ and det A.

24 The box with edges $\boldsymbol{i}$ and $\boldsymbol{j}$ and $\boldsymbol{w} = 2\boldsymbol{i} + 3\boldsymbol{j} + 4\boldsymbol{k}$ has height _____. What is the volume? What is the matrix with this determinant? What is $\boldsymbol{i} \times \boldsymbol{j}$ and what is its dot product with $\boldsymbol{w}$?

25 An n-dimensional cube has how many corners? How many edges? How many $(n-1)$-dimensional faces? The cube in $\mathbf{R}^n$ whose edges are the rows of $2I$ has volume _____. A hypercube computer has parallel processors at the corners with connections along the edges.

26 The triangle with corners $(0,0), (1,0), (0,1)$ has area $\frac{1}{2}$. The pyramid in $\mathbf{R}^3$ with four corners $(0,0,0), (1,0,0), (0,1,0), (0,0,1)$ has volume _____. What is the volume of a pyramid in $\mathbf{R}^4$ with five corners at $(0,0,0,0)$ and the rows of I?

Problems 27–30 are about areas dA and volumes dV in calculus.

27 Polar coordinates satisfy $x = r\cos\theta$ and $y = r\sin\theta$. Polar area is $J\,dr\,d\theta$:

$$J = \begin{vmatrix} \partial x/\partial r & \partial x/\partial \theta \\ \partial y/\partial r & \partial y/\partial \theta \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix}.$$

The two columns are orthogonal. Their lengths are _____. Thus $J =$ _____.

28 Spherical coordinates ρ, ϕ, θ satisfy $x = \rho\sin\phi\cos\theta$ and $y = \rho\sin\phi\sin\theta$ and $z = \rho\cos\phi$. Find the 3 by 3 matrix of partial derivatives: $\partial x/\partial\rho, \partial x/\partial\phi, \partial x/\partial\theta$ in row 1. Simplify its determinant to $J = \rho^2\sin\phi$. Then dV in spherical coordinates is $\rho^2 \sin\phi\, d\rho\, d\phi d\theta$, the volume of an infinitesimal "coordinate box".

29 The matrix that connects r, θ to x, y is in Problem 27. Invert that 2 by 2 matrix:

$$J^{-1} = \begin{vmatrix} \partial r/\partial x & \partial r/\partial y \\ \partial\theta/\partial x & \partial\theta/\partial y \end{vmatrix} = \begin{vmatrix} \cos\theta & ? \\ ? & ? \end{vmatrix} = ?$$

It is surprising that $\partial r/\partial x = \partial x/\partial r$ (***Calculus***, Gilbert Strang, p. 501). Multiplying the matrices J and J^{-1} gives the chain rule $\frac{\partial x}{\partial x} = \frac{\partial x}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial x}{\partial\theta}\frac{\partial\theta}{\partial x} = 1$.

30 The triangle with corners $(0,0)$, $(6,0)$, and $(1,4)$ has area ______. When you rotate it by $\theta = 60°$ the area is ______. The determinant of the rotation matrix is

$$J = \begin{vmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & ? \\ ? & ? \end{vmatrix} = ?$$

Problems 31–38 are about the triple product $(\boldsymbol{u} \times \boldsymbol{v}) \cdot \boldsymbol{w}$ in three dimensions.

31 A box has base area $\|\boldsymbol{u} \times \boldsymbol{v}\|$. Its perpendicular height is $\|\boldsymbol{w}\| \cos\theta$. Base area times height = volume = $\|\boldsymbol{u} \times \boldsymbol{v}\| \, \|\boldsymbol{w}\| \cos\theta$ which is $(\boldsymbol{u} \times \boldsymbol{v}) \cdot \boldsymbol{w}$. Compute base area, height, and volume for $\boldsymbol{u} = (2,4,0)$, $\boldsymbol{v} = (-1,3,0)$, $\boldsymbol{w} = (1,2,2)$.

32 The volume of the same box is given more directly by a 3 by 3 determinant. Evaluate that determinant.

33 Expand the 3 by 3 determinant in equation (13) in cofactors of its row u_1, u_2, u_3. This expansion is the dot product of $\boldsymbol{u}$ with the vector _____.

34 Which of the triple products $(\boldsymbol{u} \times \boldsymbol{w}) \cdot \boldsymbol{v}$ and $(\boldsymbol{w} \times \boldsymbol{u}) \cdot \boldsymbol{v}$ and $(\boldsymbol{v} \times \boldsymbol{w}) \cdot \boldsymbol{u}$ are the same as $(\boldsymbol{u} \times \boldsymbol{v}) \cdot \boldsymbol{w}$? Which orders of the rows $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ give the correct determinant?

35 Let $P = (1,0,-1)$ and $Q = (1,1,1)$ and $R = (2,2,1)$. Choose S so that $PQRS$ is a parallelogram and compute its area. Choose T, U, V so that $OPQRSTUV$ is a tilted box and compute its volume.

36 Suppose (x,y,z) and $(1,1,0)$ and $(1,2,1)$ lie on a plane through the origin. What determinant is zero? What equation does this give for the plane?

37 Suppose (x,y,z) is a linear combination of $(2,3,1)$ and $(1,2,3)$. What determinant is zero? What equation does this give for the plane of all combinations?

38 (a) Explain from volumes why $\det 2A = 2^n \det A$ for n by n matrices.

(b) For what size matrix is the false statement $\det A + \det A = \det(A + A)$ true?

Challenge Problems

39 If you know all 16 cofactors of a 4 by 4 invertible matrix A, how would you find A?

40 Suppose A is a 5 by 5 matrix. Its entries in row 1 multiply determinants (cofactors) in rows 2–5 to give the determinant. Can you guess a "Jacobi formula" for $\det A$ using 2 by 2 determinants from rows 1–2 *times* 3 by 3 determinants from rows 3–5?

Test your formula on the $-1, 2, -1$ tridiagonal matrix that has determinant = 6.

41 The 2 by 2 matrix AB =(2 by 3)(3 by 2) has a "Cauchy-Binet formula" for $\det AB$:

$$\det AB = \text{sum of (2 by 2 determinants in } A\text{) (2 by 2 determinants in } B\text{)}$$

(a) Guess which 2 by 2 determinants to use from A and B.

(b) Test your formula when the rows of A are $1,2,3$ and $1,4,7$ with $B = A^{\mathrm{T}}$.

Chapter 6

Eigenvalues and Eigenvectors

6.1 Introduction to Eigenvalues

Linear equations $Ax = b$ come from steady state problems. Eigenvalues have their greatest importance in *dynamic problems*. The solution of $du/dt = Au$ is changing with time—growing or decaying or oscillating. We can't find it by elimination. This chapter enters a new part of linear algebra, based on $Ax = \lambda x$. All matrices in this chapter are square.

A good model comes from the powers $A, A^2, A^3, \ldots$ of a matrix. Suppose you need the hundredth power A^{100}. The starting matrix A becomes unrecognizable after a few steps, and A^{100} is very close to $[\,.6\;\;.6;\;\;.4\;\;.4\,]$:

$$\underset{A}{\begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}} \quad \underset{A^2}{\begin{bmatrix} .70 & .45 \\ .30 & .55 \end{bmatrix}} \quad \underset{A^3}{\begin{bmatrix} .650 & .525 \\ .350 & .475 \end{bmatrix}} \quad \cdots \quad \underset{A^{100}}{\begin{bmatrix} .6000 & .6000 \\ .4000 & .4000 \end{bmatrix}}$$

A^{100} was found by using the *eigenvalues* of A, not by multiplying 100 matrices. Those eigenvalues (here they are 1 and $1/2$) are a new way to see into the heart of a matrix.

To explain eigenvalues, we first explain eigenvectors. Almost all vectors change direction, when they are multiplied by A. ***Certain exceptional vectors x are in the same direction as Ax. Those are the "eigenvectors".*** Multiply an eigenvector by A, and the vector Ax is a number λ times the original x.

The basic equation is $Ax = \lambda x$. The number λ is an eigenvalue of A.

The eigenvalue λ tells whether the special vector x is stretched or shrunk or reversed or left unchanged—when it is multiplied by A. We may find $\lambda = 2$ or $\frac{1}{2}$ or -1 or 1. The eigenvalue λ could be zero! Then $Ax = 0x$ means that this eigenvector x is in the nullspace.

If A is the identity matrix, every vector has $Ax = x$. All vectors are eigenvectors of I. All eigenvalues "lambda" are $\lambda = 1$. This is unusual to say the least. Most 2 by 2 matrices have *two* eigenvector directions and *two* eigenvalues. We will show that $\det(A - \lambda I) = 0$.

This section will explain how to compute the x's and λ's. It can come early in the course because we only need the determinant of a 2 by 2 matrix. Let me use $\det(A - \lambda I) = 0$ to find the eigenvalues for this first example, and then derive it properly in equation (3).

Example 1 The matrix A has two eigenvalues $\lambda = 1$ and $\lambda = 1/2$. Look at $\det(A - \lambda I)$:

$$A = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix} \quad \det \begin{bmatrix} .8 - \lambda & .3 \\ .2 & .7 - \lambda \end{bmatrix} = \lambda^2 - \frac{3}{2}\lambda + \frac{1}{2} = (\lambda - 1)\left(\lambda - \frac{1}{2}\right).$$

I factored the quadratic into $\lambda - 1$ times $\lambda - \frac{1}{2}$, to see the two eigenvalues $\boldsymbol{\lambda} = \mathbf{1}$ and $\boldsymbol{\lambda} = \frac{1}{2}$. For those numbers, the matrix $A - \lambda I$ becomes *singular* (zero determinant). The eigenvectors x_1 and x_2 are in the nullspaces of $A - I$ and $A - \frac{1}{2}I$.

$(A - I)x_1 = 0$ is $Ax_1 = x_1$ and the first eigenvector is $(.6, .4)$.

$(A - \frac{1}{2}I)x_2 = 0$ is $Ax_2 = \frac{1}{2}x_2$ and the second eigenvector is $(1, -1)$:

$$x_1 = \begin{bmatrix} .6 \\ .4 \end{bmatrix} \quad \text{and} \quad Ax_1 = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}\begin{bmatrix} .6 \\ .4 \end{bmatrix} = x_1 \quad (Ax = x \text{ means that } \lambda_1 = 1)$$

$$x_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \text{and} \quad Ax_2 = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} .5 \\ -.5 \end{bmatrix} \quad (\text{this is } \tfrac{1}{2}x_2 \text{ so } \lambda_2 = \tfrac{1}{2}).$$

If x_1 is multiplied again by A, we still get x_1. Every power of A will give $A^n x_1 = x_1$. Multiplying x_2 by A gave $\frac{1}{2}x_2$, and if we multiply again we get $(\frac{1}{2})^2$ times x_2.

When A is squared, the eigenvectors stay the same. The eigenvalues are squared.

This pattern keeps going, because the eigenvectors stay in their own directions (Figure 6.1) and never get mixed. The eigenvectors of A^{100} are the same x_1 and x_2. The eigenvalues of A^{100} are $1^{100} = 1$ and $(\frac{1}{2})^{100}$ = very small number.

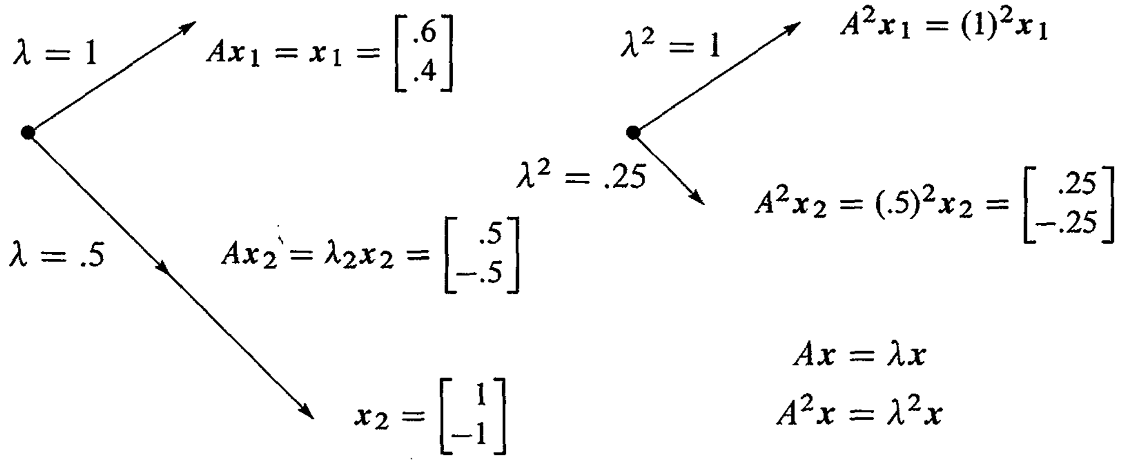


Figure 6.1: The eigenvectors keep their directions. A^2 has eigenvalues 1^2 and $(.5)^2$.

Other vectors do change direction. But all other vectors are combinations of the two eigenvectors. The first column of A is the combination $x_1 + (.2)x_2$:

$$\textbf{Separate into eigenvectors} \qquad \begin{bmatrix} .8 \\ .2 \end{bmatrix} = x_1 + (.2)x_2 = \begin{bmatrix} .6 \\ .4 \end{bmatrix} + \begin{bmatrix} .2 \\ -.2 \end{bmatrix}. \qquad (1)$$

Multiplying by A gives $(.7, .3)$, the first column of A^2. Do it separately for $\boldsymbol{x}_1$ and $(.2)\boldsymbol{x}_2$. Of course $A\boldsymbol{x}_1 = \boldsymbol{x}_1$. And A multiplies $\boldsymbol{x}_2$ by its eigenvalue $\frac{1}{2}$:

$$\textbf{Multiply each } \boldsymbol{x}_i \textbf{ by } \boldsymbol{\lambda}_i \qquad A\begin{bmatrix} .8 \\ .2 \end{bmatrix} = \begin{bmatrix} .7 \\ .3 \end{bmatrix} \quad \text{is} \quad \boldsymbol{x}_1 + \frac{1}{2}(.2)\boldsymbol{x}_2 = \begin{bmatrix} .6 \\ .4 \end{bmatrix} + \begin{bmatrix} .1 \\ -.1 \end{bmatrix}.$$

Each eigenvector is multiplied by its eigenvalue, when we multiply by A. We didn't need these eigenvectors to find A^2. But it is the good way to do 99 multiplications. At every step $\boldsymbol{x}_1$ is unchanged and $\boldsymbol{x}_2$ is multiplied by $(\frac{1}{2})$, so we have $(\frac{1}{2})^{99}$:

$$A^{99}\begin{bmatrix} .8 \\ .2 \end{bmatrix} \quad \text{is really} \quad \boldsymbol{x}_1 + (.2)(\tfrac{1}{2})^{99}\boldsymbol{x}_2 = \begin{bmatrix} .6 \\ .4 \end{bmatrix} + \begin{bmatrix} \text{very} \\ \text{small} \\ \text{vector} \end{bmatrix}.$$

This is the first column of A^{100}. The number we originally wrote as $.6000$ was not exact. We left out $(.2)(\frac{1}{2})^{99}$ which wouldn't show up for 30 decimal places.

The eigenvector $\boldsymbol{x}_1$ is a "steady state" that doesn't change (because $\lambda_1 = 1$). The eigenvector $\boldsymbol{x}_2$ is a "decaying mode" that virtually disappears (because $\lambda_2 = .5$). The higher the power of A, the closer its columns approach the steady state.

We mention that this particular A is a ***Markov matrix***. Its entries are positive and every column adds to 1. Those facts guarantee that the largest eigenvalue is $\lambda = 1$ (as we found). Its eigenvector $\boldsymbol{x}_1 = (.6, .4)$ is the *steady state*—which all columns of A^k will approach. Section 8.3 shows how Markov matrices appear in applications like Google.

For projections we can spot the steady state ($\lambda = 1$) and the nullspace ($\lambda = 0$).

Example 2 **The projection matrix $P = \begin{bmatrix} .5 & .5 \\ .5 & .5 \end{bmatrix}$ has eigenvalues $\lambda = 1$ and $\lambda = 0$.**

Its eigenvectors are $\boldsymbol{x}_1 = (1, 1)$ and $\boldsymbol{x}_2 = (1, -1)$. For those vectors, $P\boldsymbol{x}_1 = \boldsymbol{x}_1$ (steady state) and $P\boldsymbol{x}_2 = \boldsymbol{0}$ (nullspace). This example illustrates Markov matrices and singular matrices and (most important) symmetric matrices. All have special λ's and $\boldsymbol{x}$'s:

1. Each column of $P = \begin{bmatrix} .5 & .5 \\ .5 & .5 \end{bmatrix}$ adds to 1, so $\lambda = 1$ is an eigenvalue.
2. P is **singular**, so $\lambda = 0$ is an eigenvalue.
3. P is **symmetric**, so its eigenvectors $(1, 1)$ and $(1, -1)$ are perpendicular.

The only eigenvalues of a projection matrix are 0 and 1. The eigenvectors for $\lambda = 0$ (which means $P\boldsymbol{x} = 0\boldsymbol{x}$) fill up the nullspace. The eigenvectors for $\lambda = 1$ (which means $P\boldsymbol{x} = \boldsymbol{x}$) fill up the column space. The nullspace is projected to zero. The column space projects onto itself. The projection keeps the column space and destroys the nullspace:

$$\textbf{Project each part} \quad \boldsymbol{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \textbf{projects onto} \quad P\boldsymbol{v} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 \\ 2 \end{bmatrix}.$$

Special properties of a matrix lead to special eigenvalues and eigenvectors. That is a major theme of this chapter (it is captured in a table at the very end).

Projections have $\lambda = 0$ and 1. Permutations have all $|\lambda| = 1$. The next matrix R (a reflection and at the same time a permutation) is also special.

Example 3 **The reflection matrix $R = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ has eigenvalues 1 and -1.**

The eigenvector $(1, 1)$ is unchanged by R. The second eigenvector is $(1, -1)$—its signs are reversed by R. A matrix with no negative entries can still have a negative eigenvalue! The eigenvectors for R are the same as for P, because *reflection* $= 2$(*projection*) $- I$:

$$\boldsymbol{R = 2P - I} \qquad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = 2\begin{bmatrix} .5 & .5 \\ .5 & .5 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{2}$$

Here is the point. If $Px = \lambda x$ then $2Px = 2\lambda x$. The eigenvalues are doubled when the matrix is doubled. Now subtract $Ix = x$. The result is $(2P - I)x = (2\lambda - 1)x$. ***When a matrix is shifted by I, each λ is shifted by*** 1. No change in eigenvectors.

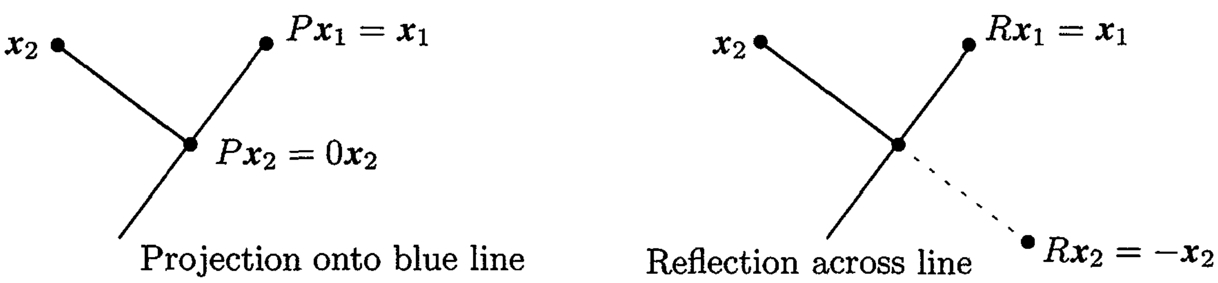


Figure 6.2: Projections P have eigenvalues 1 and 0. Reflections R have $\lambda = 1$ and -1. A typical x changes direction, but not the eigenvectors x_1 and x_2.

Key idea: The eigenvalues of R and P are related exactly as the matrices are related:

The eigenvalues of $R = 2P - I$ are $2(1) - 1 = 1$ and $2(0) - 1 = -1$.

The eigenvalues of R^2 are λ^2. In this case $R^2 = I$. Check $(1)^2 = 1$ and $(-1)^2 = 1$.

The Equation for the Eigenvalues

For projections and reflections we found λ's and x's by geometry: $Px = x, Px = \mathbf{0}$, $Rx = -x$. Now we use determinants and linear algebra. *This is the key calculation in the chapter*—almost every application starts by solving $Ax = \lambda x$.

First move λx to the left side. Write the equation $Ax = \lambda x$ as $(A - \lambda I)x = \mathbf{0}$. The matrix $A - \lambda I$ times the eigenvector x is the zero vector. ***The eigenvectors make up the nullspace of*** $A - \lambda I$. When we know an eigenvalue λ, we find an eigenvector by solving $(A - \lambda I)x = \mathbf{0}$.

Eigenvalues first. If $(A - \lambda I)x = \mathbf{0}$ has a nonzero solution, $A - \lambda I$ is not invertible. ***The determinant of $A - \lambda I$ must be zero.*** This is how to recognize an eigenvalue λ:

Eigenvalues The number λ is an eigenvalue of A if and only if $A - \lambda I$ is singular:

Equation for the eigenvalues $\qquad \det(A - \lambda I) = 0. \qquad (3)$

This "*characteristic polynomial*" $\det(A - \lambda I)$ involves only λ, not $\boldsymbol{x}$. When A is n by n, equation (3) has degree n. Then A has n eigenvalues (repeats possible!) Each λ leads to $\boldsymbol{x}$:

For each eigenvalue λ solve $(A - \lambda I)\boldsymbol{x} = \mathbf{0}$ or $A\boldsymbol{x} = \lambda\boldsymbol{x}$ to find an eigenvector $\boldsymbol{x}$.

Example 4 $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ is already singular (zero determinant). Find its λ's and $\boldsymbol{x}$'s.

When A is singular, $\lambda = 0$ is one of the eigenvalues. The equation $A\boldsymbol{x} = 0\boldsymbol{x}$ has solutions. They are the eigenvectors for $\lambda = 0$. But $\det(A - \lambda I) = 0$ is the way to find *all* λ's and $\boldsymbol{x}$'s. Always subtract λI from A:

$$\textit{Subtract } \lambda \textit{ from the diagonal to find} \quad A - \lambda I = \begin{bmatrix} 1-\lambda & 2 \\ 2 & 4-\lambda \end{bmatrix}. \tag{4}$$

Take the determinant "$ad - bc$" of this 2 by 2 matrix. From $1 - \lambda$ times $4 - \lambda$, the "ad" part is $\lambda^2 - 5\lambda + 4$. The "bc" part, not containing λ, is 2 times 2.

$$\det \begin{bmatrix} 1-\lambda & 2 \\ 2 & 4-\lambda \end{bmatrix} = (1-\lambda)(4-\lambda) - (2)(2) = \lambda^2 - 5\lambda. \tag{5}$$

Set this determinant $\lambda^2 - 5\lambda$ to zero. One solution is $\lambda = 0$ (as expected, since A is singular). Factoring into λ times $\lambda - 5$, the other root is $\lambda = 5$:

$$\det(A - \lambda I) = \lambda^2 - 5\lambda = 0 \quad \text{yields the eigenvalues} \quad \lambda_1 = 0 \quad \text{and} \quad \lambda_2 = 5\,.$$

Now find the eigenvectors. Solve $(A - \lambda I)\boldsymbol{x} = \mathbf{0}$ separately for $\lambda_1 = 0$ and $\lambda_2 = 5$:

$$(A - 0I)\boldsymbol{x} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ yields an eigenvector } \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \text{ for } \lambda_1 = 0$$

$$(A - 5I)\boldsymbol{x} = \begin{bmatrix} -4 & 2 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ yields an eigenvector } \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ for } \lambda_2 = 5.$$

The matrices $A - 0I$ and $A - 5I$ are singular (because 0 and 5 are eigenvalues). The eigenvectors $(2, -1)$ and $(1, 2)$ are in the nullspaces: $(A - \lambda I)\boldsymbol{x} = \mathbf{0}$ is $A\boldsymbol{x} = \lambda\boldsymbol{x}$.

We need to emphasize: *There is nothing exceptional about* $\lambda = 0$. Like every other number, zero might be an eigenvalue and it might not. If A is singular, it is. The eigenvectors fill the nullspace: $A\boldsymbol{x} = 0\boldsymbol{x} = \mathbf{0}$. If A is invertible, zero is not an eigenvalue. We shift A by a multiple of I to *make it singular.*

In the example, the shifted matrix $A - 5I$ is singular and 5 is the other eigenvalue.

Summary To solve the eigenvalue problem for an n by n matrix, follow these steps:

1. ***Compute the determinant of*** $A - \lambda I$. With λ subtracted along the diagonal, this determinant starts with λ^n or $-\lambda^n$. It is a polynomial in λ of degree n.

2. ***Find the roots of this polynomial,*** by solving $\det(A - \lambda I) = 0$. The n roots are the n eigenvalues of A. They make $A - \lambda I$ singular.

3. For each eigenvalue λ, ***solve*** $(A - \lambda I)x = \mathbf{0}$ ***to find an eigenvector*** x.

A note on the eigenvectors of 2 by 2 matrices. When $A - \lambda I$ is singular, both rows are multiples of a vector (a, b). *The eigenvector is any multiple of* $(b, -a)$. The example had $\lambda = 0$ and $\lambda = 5$:

$\lambda = 0$: rows of $A - 0I$ in the direction $(1, 2)$; eigenvector in the direction $(2, -1)$

$\lambda = 5$: rows of $A - 5I$ in the direction $(-4, 2)$; eigenvector in the direction $(2, 4)$.

Previously we wrote that last eigenvector as $(1, 2)$. Both $(1, 2)$ and $(2, 4)$ are correct. There is a whole *line of eigenvectors*—any nonzero multiple of x is as good as x. MATLAB's **eig**(A) divides by the length, to make the eigenvector into a unit vector.

We end with a warning. Some 2 by 2 matrices have only *one* line of eigenvectors. This can only happen when two eigenvalues are equal. (On the other hand $A = I$ has equal eigenvalues and plenty of eigenvectors.) Similarly some n by n matrices don't have n independent eigenvectors. Without n eigenvectors, we don't have a basis. We can't write every v as a combination of eigenvectors. In the language of the next section, we can't diagonalize a matrix without n independent eigenvectors.

Good News, Bad News

Bad news first: If you add a row of A to another row, or exchange rows, the eigenvalues usually change. *Elimination does not preserve the* λ's. The triangular U has *its* eigenvalues sitting along the diagonal—they are the pivots. But they are not the eigenvalues of A! Eigenvalues are changed when row 1 is added to row 2:

$$U = \begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix} \quad \text{has } \lambda = 0 \text{ and } \lambda = 1; \quad A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} \quad \text{has } \lambda = 0 \text{ and } \lambda = 7.$$

Good news second: The *product* λ_1 *times* λ_2 *and the sum* $\lambda_1 + \lambda_2$ *can be found quickly from the matrix.* For this A, the product is 0 times 7. That agrees with the determinant (which is 0). The sum of eigenvalues is $0 + 7$. That agrees with the sum down the main diagonal (the **trace** is $1 + 6$). These quick checks always work:

The product of the n eigenvalues equals the determinant.
The sum of the n eigenvalues equals the sum of the n diagonal entries.

The sum of the entries on the main diagonal is called the ***trace*** of A:

$$\lambda_1 + \lambda_2 + \cdots + \lambda_n = \textbf{\textit{trace}} = a_{11} + a_{22} + \cdots + a_{nn}. \tag{6}$$

Those checks are very useful. They are proved in Problems 16–17 and again in the next section. They don't remove the pain of computing λ's. But when the computation is wrong, they generally tell us so. To compute the correct λ's, go back to $\det(A - \lambda I) = 0$.

The determinant test makes the *product* of the λ's equal to the *product* of the pivots (assuming no row exchanges). But the sum of the λ's is not the sum of the pivots—as the example showed. The individual λ's have almost nothing to do with the pivots. In this new part of linear algebra, the key equation is really *nonlinear*: λ multiplies $\boldsymbol{x}$.

Why do the eigenvalues of a triangular matrix lie on its diagonal?

Imaginary Eigenvalues

One more bit of news (not too terrible). The eigenvalues might not be real numbers.

Example 5 *The* $90°$ *rotation* $Q = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ *has no real eigenvectors. Its eigenvalues are* $\lambda = i$ *and* $\lambda = -i$. *Sum of* λ*'s* $=$ trace $= 0$. *Product* $=$ *determinant* $= 1$.

After a rotation, *no vector* $Q\boldsymbol{x}$ *stays in the same direction as* $\boldsymbol{x}$ (except $\boldsymbol{x} = \boldsymbol{0}$ which is useless). There cannot be an eigenvector, unless we go to ***imaginary numbers***. Which we do.

To see how i can help, look at Q^2 which is $-I$. If Q is rotation through $90°$, then Q^2 is rotation through $180°$. Its eigenvalues are -1 and -1. (Certainly $-I\boldsymbol{x} = -1\boldsymbol{x}$.) Squaring Q will square each λ, so we must have $\lambda^2 = -1$. *The eigenvalues of the* $90°$ *rotation matrix* Q *are* $+i$ *and* $-i$, because $i^2 = -1$.

Those λ's come as usual from $\det(Q - \lambda I) = 0$. This equation gives $\lambda^2 + 1 = 0$. Its roots are i and $-i$. We meet the imaginary number i also in the eigenvectors:

Complex eigenvectors
$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ i \end{bmatrix} = -i\begin{bmatrix} 1 \\ i \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} i \\ 1 \end{bmatrix} = i\begin{bmatrix} i \\ 1 \end{bmatrix}.$$

Somehow these complex vectors $\boldsymbol{x}_1 = (1, i)$ and $\boldsymbol{x}_2 = (i, 1)$ keep their direction as they are rotated. Don't ask me how. This example makes the all-important point that real matrices can easily have complex eigenvalues and eigenvectors. The particular eigenvalues i and $-i$ also illustrate two special properties of Q:

1. Q is an orthogonal matrix so the absolute value of each λ is $|\lambda| = 1$.

2. Q is a skew-symmetric matrix so each λ is pure imaginary.