INTRODUCTION TO LINEAR ALGEBRA

Fifth Edition

GILBERT STRANG
Massachusetts Institute of Technology

WELLESLEY - CAMBRIDGE PRESS
Box 812060 Wellesley MA 02482

Introduction to Linear Algebra, 5th Edition

ISBN 978-0-9802327-7-6

LaTeX typesetting by Ashley C. Fernandes (info@problemsolvingpathway.com)
Printed in the United States of America

9 8 7 6 5 4 3

QA184.S78 2016 512'.5 93-14092

Other texts from Wellesley - Cambridge Press

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

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

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

Calculus Third Edition (2017), Gilbert Strang ISBN 978-0-9802327-5-2

Algorithms for Global Positioning, Kai Borre & Gilbert Strang ISBN 978-0-9802327-3-8

Essays in Linear Algebra, Gilbert Strang ISBN 978-0-9802327-6-9

Differential Equations and Linear Algebra, Gilbert Strang ISBN 978-0-9802327-9-0

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

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

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

The website for this book is **math.mit.edu/linearalgebra**.
The Solution Manual can be printed from that website.

Course material including syllabus and exams and also videotaped lectures are available on the book website and 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 and 18.06 SC.
MATLAB® is a registered trademark of The MathWorks, Inc.

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

The cover design was the inspiration of Lois Sellers and Gail Corbett.

Table of Contents

Preface

I am happy for you to see this Fifth Edition of Introduction to Linear Algebra. This is the text for my video lectures on MIT's OpenCourseWare (**ocw.mit.edu** and also **YouTube**). I hope those lectures will be useful to you (maybe even enjoyable !).

Hundreds of colleges and universities have chosen this textbook for their basic linear algebra course. A sabbatical gave me a chance to prepare two new chapters about probability and statistics and understanding data. Thousands of other improvements too—probably only noticed by the author... Here is a new addition for students and all readers:

> Every section opens with a brief summary to explain its contents. When you read a new section, and when you revisit a section to review and organize it in your mind, those lines are a quick guide and an aid to memory.

Another big change comes on this book's website **math.mit.edu/linearalgebra**. That site now contains solutions to the Problem Sets in the book. With unlimited space, this is much more flexible than printing short solutions. There are three key websites :

ocw.mit.edu Messages come from thousands of students and faculty about linear algebra on this OpenCourseWare site. The 18.06 and 18.06 SC courses include 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 2 a.m. (The reader doesn't have to be in a class at all.) Six 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 to you as possible, with all the course material we can provide.

math.mit.edu/linearalgebra This has become an active website. It now has Solutions to Exercises—with space to explain ideas. There are also new exercises from many different sources—practice problems, development of textbook examples, codes in MATLAB and *Julia* and *Python*, plus whole collections of exams (18.06 and others) for review.

Please visit this linear algebra site. *Send suggestions to* **linearalgebrabook@gmail.com**

The Fifth Edition

The 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! When you meet those four spaces in Chapter 3, 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 to take ***linear combinations of column vectors***. This is exactly the result of a matrix-vector multiplication. *Ax is a combination of the columns of A.*

When we take *all* combinations Ax of the column vectors, we get the *column space.* If this space includes the vector b, we can solve the equation $Ax = b$.

May I call special attention to Section 1.3, where 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. There is even an *inverse matrix* and its connection to calculus. 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 ends with a large collection of review problems. They 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 *Challenge Problems* go a step further, and sometimes deeper. Let me give four examples:

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

Section 2.7: If P is a permutation matrix, why is some power P^k equal to I ?

Section 3.4: If $Ax = b$ and $Cx = b$ have the same solutions for every b, does A equal 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 matrix: $x = A^{-1}b$ (if A has an inverse).

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

That vector space solution $x = y + z$ is shown on the cover of the book.

Direct elimination is the most frequently used algorithm in scientific computing. The matrix A becomes triangular—then solutions come quickly. We also see bases for the four subspaces. But don't spend forever on practicing elimination ... good ideas are coming.

The speed of every new supercomputer is tested on $Ax = b$: pure linear algebra. But 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. This book moves gradually and steadily from *numbers* to *vectors* to *subspaces*—each level comes naturally and everyone can get it.

Here are 12 points about learning and teaching from this book:

1. Chapter 1 starts with vectors and dot products. If the class has met them before, focus quickly on linear combinations. 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 of A: 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 is to capture the whole process—start with A, multiply by E's, end with U.

   Elimination is seen in the beautiful form $A = LU$. The ***lower triangular*** L holds the forward elimination steps, and U is ***upper triangular*** 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. With more equations than unknowns, it is almost sure that $Ax = b$ has no solution. We cannot throw out every measurement that is close but not perfectly exact! When we solve by ***least squares***, the key will be the matrix $A^{\mathrm{T}}A$. This wonderful matrix appears everywhere in applied mathematics, when A is rectangular.

5. ***Determinants*** give formulas for all that has come before—Cramer's Rule, inverse matrices, volumes in n dimensions. We don't need those formulas to compute. They slow us down. But $\det A = 0$ tells when a matrix is singular: this is the key to eigenvalues.

6. ***Section 6.1 explains 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.

   An essential highlight of Chapter 6 is ***diagonalizing a symmetric matrix***. When all the eigenvalues are positive, the matrix is "positive definite". This key idea connects the whole course—positive pivots and determinants and eigenvalues and energy. I work hard to reach this point in the book and to explain it by examples.

7. Chapter 7 is new. It introduces ***singular values*** and ***singular vectors***. They separate all martices into simple pieces, ranked in order of their importance. You will see one way to compress an image. Especially you can analyze a matrix full of data.

8. Chapter 8 explains ***linear transformations***. This is geometry without axes, algebra with no coordinates. When we choose a basis, we reach the best possible matrix.

9. Chapter 9 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 revolutionary.

10. Chapter 10 is full of applications, more than any single course could need:

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

    10.2 *Matrices in Engineering*—differential equations parallel to matrix equations

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

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

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

    10.6 *Computer Graphics*—matrices move and rotate and compress images

    10.7 *Linear Algebra in Cryptography*—this new section was fun to write. The Hill Cipher is not too secure. It uses modular arithmetic: integers from 0 to $p-1$. Multiplication gives $4 \times 5 \equiv 1\,(mod\ 19)$. For decoding this gives $4^{-1} \equiv 5$.

11. How should computing be included in a linear algebra course? It can open a new understanding of matrices—every class will find a balance. MATLAB and *Maple* and *Mathematica* are powerful in different ways. *Julia* and *Python* are free and directly accessible on the Web. Those newer languages are powerful too !

    Basic commands begin in Chapter 2. Then Chapter 11 moves toward professional algorithms.You can upload and download codes for this course on the website.

12. Chapter 12 on Probability and Statistics is new, with truly important applications. When random variables are not independent we get covariance matrices. Fortunately they are symmetric positive definite. The linear algebra in Chapter 6 is needed now.

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.

I have used *italics* and ***boldface*** to pick out the key words on each page. I know there are times when you want to read quickly, looking for the important lines.

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

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.*

One person has helped with every word in this book. He is Ashley C. Fernandes, who prepared the LaTeX files. It is now six books that he has allowed me to write and rewrite, aiming for accuracy and also for life. Working with friends is a happy way to live.

Friends inside and outside the MIT math department have been wonderful. Alan Edelman for *Julia* and much more, Alex Townsend for the flag examples in 7.1, and Peter Kempthorne for the finance example in 7.3 : those stand out. Don Spickler's website on cryptography is simply excellent. I thank Jon Bloom, Jack Dongarra, Hilary Finucane, Pavel Grinfeld, Randy LeVeque, David Vogan, Liang Wang, and Karen Willcox. The "eigenfaces" in 7.3 came from Matthew Turk and Jeff Jauregui. And the big step to singular values was accelerated by Raj Rao's great course at Michigan.

This book owes so much to my happy sabbatical in Oxford. Thank you, Nick Trefethen and everyone. Especially you the reader! Best wishes in your work.

Background of the Author

This is my 9th textbook on linear algebra, and I hesitate to write about myself. It is the mathematics that is important, and the reader. The next paragraphs add something brief and 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 6 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.**

I am so grateful for a life of teaching mathematics, more than I could possibly tell you.

Gilbert Strang

PS I hope the next book (2018 ?) will include *Learning from Data*. This subject is growing quickly, especially "deep learning". By knowing a function on a training set of old data, we approximate the function on new data. The approximation only uses one simple nonlinear function $f(x) = \max(0, x)$. It is n matrix multiplications that we optimize to make the learning deep: $\boldsymbol{x}_1 = f(A_1\boldsymbol{x} + \boldsymbol{b}_1), \boldsymbol{x}_2 = f(A_2\boldsymbol{x}_1 + \boldsymbol{b}_2), \ldots, \boldsymbol{x}_n = f(A_n\boldsymbol{x}_{n-1} + \boldsymbol{b}_n)$. Those are $n-1$ hidden layers between the input $\boldsymbol{x}$ and the output $\boldsymbol{x}_n$—which approximates $F(\boldsymbol{x})$ on the training set.

THE MATRIX ALPHABET

$\boldsymbol{A}$	Any Matrix	$\boldsymbol{P}$	Permutation Matrix
$\boldsymbol{B}$	Basis Matrix	$\boldsymbol{P}$	Projection Matrix
$\boldsymbol{C}$	Cofactor Matrix	$\boldsymbol{Q}$	Orthogonal Matrix
$\boldsymbol{D}$	Diagonal Matrix	$\boldsymbol{R}$	Upper Triangular Matrix
$\boldsymbol{E}$	Elimination Matrix	$\boldsymbol{R}$	Reduced Echelon Matrix
$\boldsymbol{F}$	Fourier Matrix	$\boldsymbol{S}$	Symmetric Matrix
$\boldsymbol{H}$	Hadamard Matrix	$\boldsymbol{T}$	Linear Transformation
$\boldsymbol{I}$	Identity Matrix	$\boldsymbol{U}$	Upper Triangular Matrix
$\boldsymbol{J}$	Jordan Matrix	$\boldsymbol{U}$	Left Singular Vectors
$\boldsymbol{K}$	Stiffness Matrix	$\boldsymbol{V}$	Right Singular Vectors
$\boldsymbol{L}$	Lower Triangular Matrix	$\boldsymbol{X}$	Eigenvector Matrix
$\boldsymbol{M}$	Markov Matrix	$\boldsymbol{\Lambda}$	Eigenvalue Matrix
$\boldsymbol{N}$	Nullspace Matrix	$\boldsymbol{\Sigma}$	Singular Value Matrix

Chapter 1

Introduction to Vectors

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

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

Example $\boldsymbol{v} + \boldsymbol{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ 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 $c\boldsymbol{v} + d\boldsymbol{w} = (4, 5)$. Other times we want *all the combinations* of $\boldsymbol{v}$ and $\boldsymbol{w}$ (coming from all c and d).

The vectors $c\boldsymbol{v}$ lie along a line. When $\boldsymbol{w}$ is not on that line, **the combinations** $c\boldsymbol{v} + d\boldsymbol{w}$ **fill the whole two-dimensional plane**. Starting from four vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{z}$ in four-dimensional space, their combinations $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w} + f\boldsymbol{z}$ are likely to fill the space—but not always. The vectors and their combinations could lie in a plane or on a 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*** $\boldsymbol{v} + \boldsymbol{w}$ ***and linear combinations*** $c\boldsymbol{v} + d\boldsymbol{w}$.

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

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

1.1 Vectors and Linear Combinations

1 $3\boldsymbol{v} + 5\boldsymbol{w}$ is a typical **linear combination** $c\boldsymbol{v} + d\boldsymbol{w}$ of the vectors $\boldsymbol{v}$ and $\boldsymbol{w}$.

2 For $\boldsymbol{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ that combination is $3\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 5\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3+10 \\ 3+15 \end{bmatrix} = \begin{bmatrix} 13 \\ 18 \end{bmatrix}$.

3 The vector $\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \end{bmatrix}$ goes across to $x = 2$ and up to $y = 3$ in the xy plane.

4 The combinations $c\begin{bmatrix} 1 \\ 1 \end{bmatrix} + d\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ fill the whole xy plane. They produce every $\begin{bmatrix} x \\ y \end{bmatrix}$.

5 The combinations $c\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + d\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$ fill a **plane** in xyz space. Same plane for $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$.

6 But $\begin{matrix} c + 2d = 1 \\ c + 3d = 0 \\ c + 4d = 0 \end{matrix}$ has no solution because its right side $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ is not on that plane.

"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*** $\boldsymbol{v}$:

Column vector $\boldsymbol{v}$ $\qquad \boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \qquad \begin{matrix} v_1 = \text{first component of } \boldsymbol{v} \\ v_2 = \text{second component of } \boldsymbol{v} \end{matrix}$

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:

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

Subtraction 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. To find $2\boldsymbol{v}$, multiply each component of $\boldsymbol{v}$ by 2:

SCALAR MULTIPLICATION $\qquad 2\boldsymbol{v} = \begin{bmatrix} 2v_1 \\ 2v_2 \end{bmatrix} = \boldsymbol{v} + \boldsymbol{v} \qquad -\boldsymbol{v} = \begin{bmatrix} -v_1 \\ -v_2 \end{bmatrix}$.

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 $c\boldsymbol{v}$ and $d\boldsymbol{w}$—***adding vectors and multiplying by scalars***.

Linear Combinations

Now we combine addition with scalar multiplication to produce a "**linear combination**" of $\boldsymbol{v}$ and $\boldsymbol{w}$. Multiply $\boldsymbol{v}$ by c and multiply $\boldsymbol{w}$ by d. Then add $c\boldsymbol{v} + d\boldsymbol{w}$.

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

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

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

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*** $\boldsymbol{w}$.

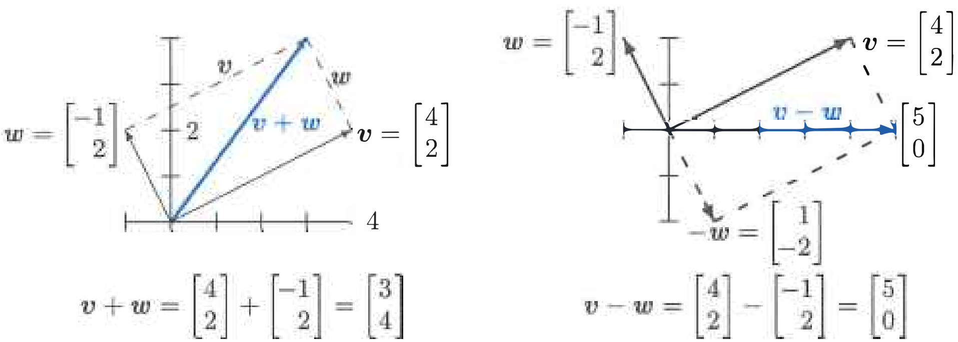


Figure 1.1: Vector addition $\boldsymbol{v} + \boldsymbol{w} = (3,\ 4)$ produces the diagonal of a parallelogram. The reverse of $\boldsymbol{w}$ is $-\boldsymbol{w}$. The linear combination on the right is $\boldsymbol{v} - \boldsymbol{w} = (5,\ 0)$.

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).

Vectors in Three Dimensions

A vector with two components corresponds to a point in the xy plane. The components of $\boldsymbol{v}$ 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 xyz space. Here are typical vectors (still column vectors but with three components):

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

The vector $\boldsymbol{v}$ 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***.

The vector (x, y) in the plane is different from $(x, y, 0)$ in 3-space!

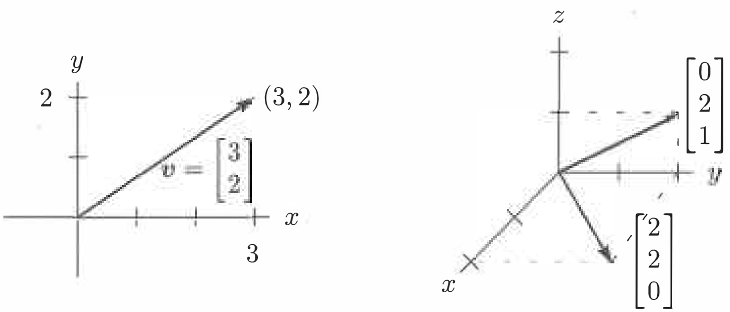


Figure 1.2: Vectors $\begin{bmatrix} x \\ y \end{bmatrix}$ and $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ correspond to points (x, y) and (x, y, z).

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

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 1 by 3 row vector is the "transpose" of the 3 by 1 column vector $\boldsymbol{v}$.

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 through*** $(\mathbf{0}, \mathbf{0}, \mathbf{0})$.

2. The combinations $c\boldsymbol{u} + d\boldsymbol{v}$ fill a ***plane through*** $(\mathbf{0}, \mathbf{0}, \mathbf{0})$.

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 could both 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.

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

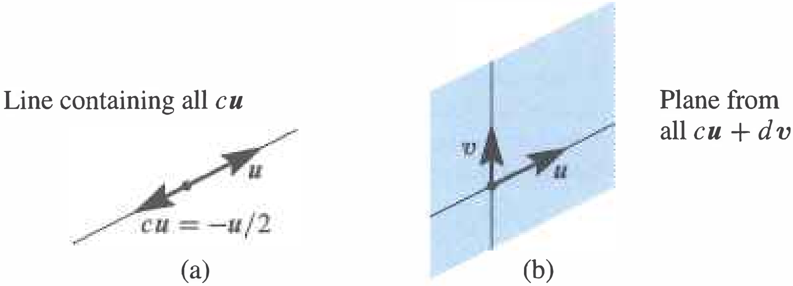


Figure 1.3: (a) Line through $\boldsymbol{u}$. (b) The plane containing the lines through $\boldsymbol{u}$ and $\boldsymbol{v}$.

■ REVIEW OF THE KEY IDEAS ■

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

■ WORKED EXAMPLES ■

1.1 A The linear combinations of $\boldsymbol{v} = (1,1,0)$ and $\boldsymbol{w} = (0,1,1)$ fill a plane in $\mathbf{R}^3$. *Describe that plane*. Find a vector that is *not* a combination of $\boldsymbol{v}$ and $\boldsymbol{w}$—not on the plane.

Solution The plane of $\boldsymbol{v}$ and $\boldsymbol{w}$ contains all combinations $c\boldsymbol{v} + d\boldsymbol{w}$. The vectors in that plane allow any c and d. The plane of Figure 1.3 fills in between the two lines.

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

Four 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. Like most vectors, $(1,2,3)$ *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 $\boldsymbol{w} \cdot \boldsymbol{n} = 0$. Perpendicular vectors have zero dot products.

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

Solution

$(\mathbf{1})$ 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 $(-2, 0), (-1, 0), (0, 0), (1, 0), (2, 0)$.

$(\mathbf{2})$ 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 $(100, 0)$ and $(\pi, 0)$ but not $(-100, 0)$.

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

$(\mathbf{2'})$ 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***. The right half of the xy plane has any $x \geq 0$ and any y.

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

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

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 solution). Our example fits into a fundamental model for linear algebra:

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

For $n = 2$ we will find a formula for the $\boldsymbol{c}$'s. The "elimination method" in Chapter 2 succeeds far beyond $n = 1000$. For n greater than 1 billion, see Chapter 11. Here $n = 2$:

Vector equation
$c\boldsymbol{v} + d\boldsymbol{w} = \boldsymbol{b}$

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

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

Two ordinary equations

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

Each equation produces a line. The two lines cross at the solution $c = \dfrac{2}{3}, d = \dfrac{1}{3}$. Why not see this also as a **matrix equation**, since that is where we are going:

2 by 2 matrix
$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

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 the vectors $\boldsymbol{v}$ 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?

These lie in a plane because $\boldsymbol{w} = c\boldsymbol{u} + d\boldsymbol{v}$. Find c and d

$$\boldsymbol{u} = \begin{bmatrix}1\\2\\3\end{bmatrix}, \quad \boldsymbol{v} = \begin{bmatrix}-3\\1\\-2\end{bmatrix}, \quad \boldsymbol{w} = \begin{bmatrix}2\\-3\\-1\end{bmatrix}.$$

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 $c\boldsymbol{v}+d\boldsymbol{w} = (3, 3, -6)$. Why is $(3, 3, 6)$ impossible?

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 this unit 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 _____. The cube has how many edges?

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

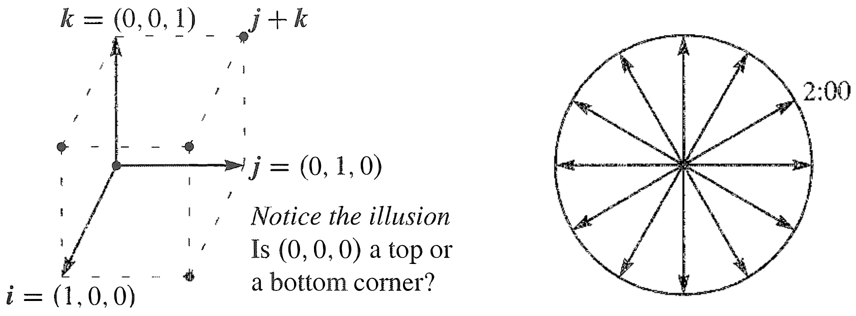


Figure 1.4: Unit cube from $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ and twelve clock vectors.

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 x, y 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 $(0, 2)$. The new twelve vectors add to ____.

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 $c\boldsymbol{v} + d\boldsymbol{w}$.

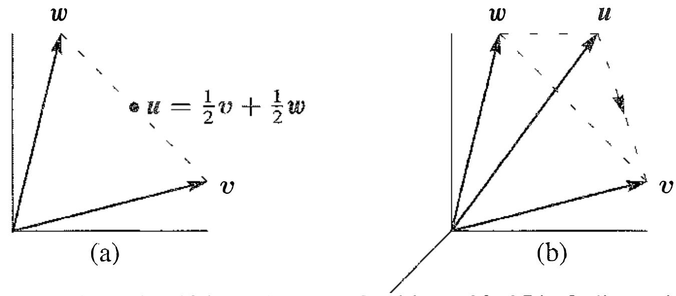


Figure 1.5: Problems **15–19** in a plane   Problems **20–25** in 3-dimensional space

Problems 20–25 deal with u, v, 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.

Challenge Problems

27 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 $(0, 1, 0, 0)$.

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, e for this $\boldsymbol{b}$?

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

1.2 Lengths and Dot Products

1 The "dot product" of $\boldsymbol{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$ is $\boldsymbol{v} \cdot \boldsymbol{w} = (1)(4) + (2)(5) = 4 + 10 = \mathbf{14}$.

2 $\boldsymbol{v} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 4 \\ -4 \\ 4 \end{bmatrix}$ are perpendicular because $\boldsymbol{v} \cdot \boldsymbol{w}$ is zero: $(1)(4) + (3)(-4) + (2)(4) = \mathbf{0}$.

3 The length squared of $\boldsymbol{v} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ is $\boldsymbol{v} \cdot \boldsymbol{v} = 1 + 9 + 4 = 14$. **The length is** $||\boldsymbol{v}|| = \sqrt{\mathbf{14}}$.

4 Then $\boldsymbol{u} = \dfrac{\boldsymbol{v}}{||\boldsymbol{v}||} = \dfrac{\boldsymbol{v}}{\sqrt{14}} = \dfrac{1}{\sqrt{14}} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ has length $||\boldsymbol{u}|| = \mathbf{1}$. Check $\dfrac{1}{14} + \dfrac{9}{14} + \dfrac{4}{14} = 1$.

5 The angle θ between $\boldsymbol{v}$ and $\boldsymbol{w}$ has $\cos\theta = \dfrac{\boldsymbol{v} \cdot \boldsymbol{w}}{||\boldsymbol{v}||\ ||\boldsymbol{w}||}$.

6 The angle between $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ has $\cos\theta = \dfrac{1}{(1)(\sqrt{2})}$. That angle is $\theta = 45\,°$.

7 All angles have $|\cos\theta| \leq 1$. So all vectors have $\boxed{|\boldsymbol{v} \cdot \boldsymbol{w}| \leq ||\boldsymbol{v}||\ ||\boldsymbol{w}||.}$

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 $v_1 w_1$ and $v_2 w_2$, but it doesn't stop there. Those two numbers are added to produce one number $\boldsymbol{v} \cdot \boldsymbol{w}$.

This is the geometry section (lengths of vectors and cosines of angles between them).

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

$$\boldsymbol{v} \cdot \boldsymbol{w} = v_1 w_1 + v_2 w_2. \tag{1}$$

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 For $\boldsymbol{v}\cdot\boldsymbol{w}$, multiply each v_i times w_i. Then $\boldsymbol{v}\cdot\boldsymbol{w} = v_1w_1 + \cdots + v_nw_n$.

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 $\boldsymbol{v}\cdot\boldsymbol{v}$:

$$\textbf{length} = \|\boldsymbol{v}\| = \sqrt{\boldsymbol{v}\cdot\boldsymbol{v}} = \left(v_1^2 + v_2^2 + \cdots + v_n^2\right)^{1/2}.$$

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. If the components are 1 and 2, the arrow is the third side of a right triangle (Figure 1.6). The Pythagoras formula $a^2 + b^2 = c^2$ connects the three sides: $1^2 + 2^2 = \|\boldsymbol{v}\|^2$.

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. That diagonal in n dimensions has length $\sqrt{n}$.

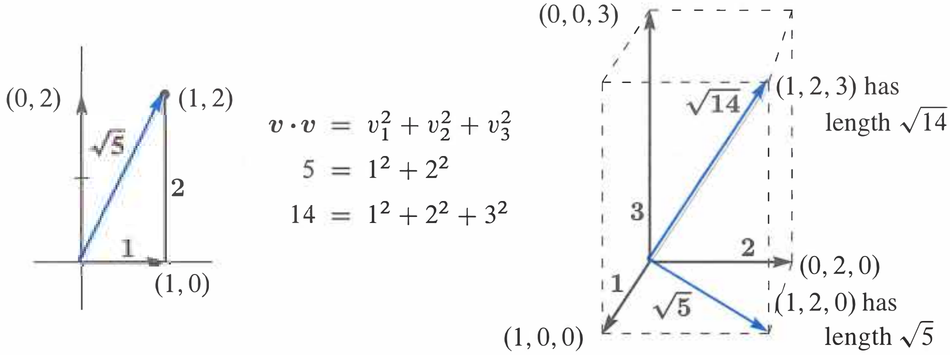


Figure 1.6: The length $\sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}$ of two-dimensional and three-dimensional vectors.

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 see the meaning 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 $\boldsymbol{u} = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$. 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.

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 $\theta = 90°$ (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 $\left(\frac{2}{3},\frac{2}{3},\frac{1}{3}\right)$ 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 vector $\boldsymbol{v}$ by its length** $\|\boldsymbol{v}\|$.

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

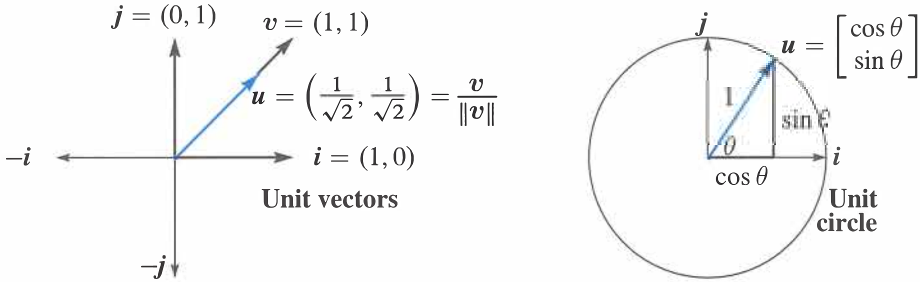


Figure 1.7: The coordinate vectors $\boldsymbol{i}$ and $\boldsymbol{j}$. The unit vector $\boldsymbol{u}$ at angle $45°$ (left) divides $\boldsymbol{v} = (1,1)$ by its length $\|\boldsymbol{v}\| = \sqrt{2}$. The unit vector $\boldsymbol{u} = (\cos\theta, \sin\theta)$ is at angle θ.

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 $a^2+b^2=c^2$:

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

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 to see $\boldsymbol{v}-\boldsymbol{w} = 0$:

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

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.

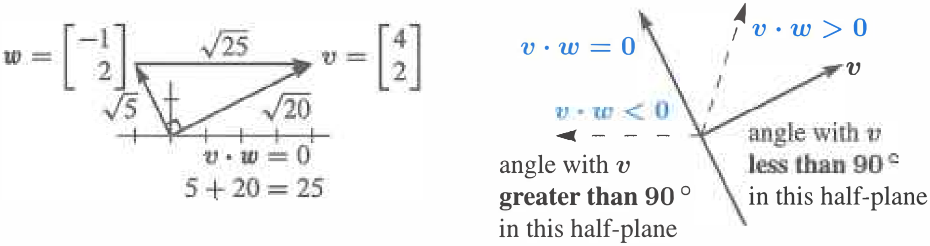


Figure 1.8: Perpendicular vectors have $\boldsymbol{v} \cdot \boldsymbol{w} = 0$. Then $\|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2 = \|\boldsymbol{v} - \boldsymbol{w}\|^2$.

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 $\boldsymbol{\theta}$. For unit vectors $\boldsymbol{u}$ and $\boldsymbol{U}$, the sign of $\boldsymbol{u} \cdot \boldsymbol{U}$ tells whether $\theta < 90^\circ$ or $\theta > 90^\circ$. More than that, ***the dot product $\boldsymbol{u} \cdot \boldsymbol{U}$ is the cosine of θ.*** This remains true in n 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. **The cosine of θ is revealed by $\boldsymbol{u} \cdot \boldsymbol{U}$.**

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 $\cos(\beta - \alpha) = \cos\theta$.

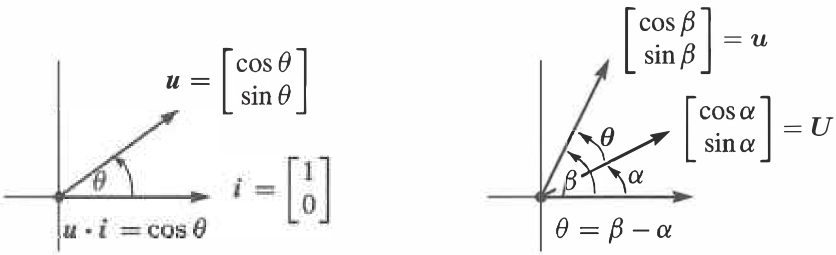


Figure 1.9: Unit vectors : $\boldsymbol{u} \cdot \boldsymbol{U}$ is the cosine of θ (the angle between).

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 $\dfrac{\boldsymbol{v}\cdot\boldsymbol{w}}{\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|} = \cos\theta.$ (5)

Whatever the angle, this dot product of $\boldsymbol{v}/\|\boldsymbol{v}\|$ with $\boldsymbol{w}/\|\boldsymbol{w}\|$ never exceeds one. That is the ***"Schwarz inequality"*** $|\boldsymbol{v}\cdot\boldsymbol{w}| \le \|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$ 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	$\|\boldsymbol{v}\cdot\boldsymbol{w}\| \le \\|\boldsymbol{v}\\|\,\\|\boldsymbol{w}\\|$
TRIANGLE INEQUALITY	$\\|\boldsymbol{v}+\boldsymbol{w}\\| \le \\|\boldsymbol{v}\\| + \\|\boldsymbol{w}\\|$

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

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

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

By the Schwarz inequality, $\boldsymbol{v}\cdot\boldsymbol{w} = 4$ is less than $\|\boldsymbol{v}\|\,\|\boldsymbol{w}\| = 5$. By the triangle inequality, side $3 = \|\boldsymbol{v}+\boldsymbol{w}\|$ is less than side 1 + side 2. For $\boldsymbol{v}+\boldsymbol{w} = (3,3)$ the three sides are $\sqrt{18} < \sqrt{5}+\sqrt{5}$. Square this triangle inequality to get $18 < 20$.

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

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

Geometric mean $\le$ **Arithmetic mean** $\quad ab \le \dfrac{a^2+b^2}{2}$ becomes $\sqrt{xy} \le \dfrac{x+y}{2}$.

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

MATLAB, Python *and* Julia *work directly with whole vectors, not their components.* When $\boldsymbol{v}$ and $\boldsymbol{w}$ have been defined, $\boldsymbol{v}+\boldsymbol{w}$ is immediately understood. Input $\boldsymbol{v}$ and $\boldsymbol{w}$ as rows—the prime $'$ transposes them to columns. $2\boldsymbol{v}+3\boldsymbol{w}$ becomes $2*\boldsymbol{v}+3*\boldsymbol{w}$. The result will be printed unless the line ends in a semicolon.

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

The dot product $v \cdot w$ is ***a row vector times a column vector (use * instead of $\cdot$)***:

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

The length of v is known to MATLAB as norm (v). This is sqrt $(v' * v)$. Then find the cosine from the dot product $v' * w$ and the angle (in radians) that has that cosine:

Cosine formula	cosine $= v' * w / ($norm $(v) *$ norm $(w))$
The arc cosine	angle $=$ acos (cosine)

An M-file would create a new function **cosine** (v, w). Python and Julia are open source.

■ 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 $\|v\|$ is the square root of $v \cdot v$. Then $u = v / \|v\|$ is a ***unit vector***: length 1.
3. The dot product is $v \cdot w = 0$ when vectors v and w are perpendicular.
4. The cosine of θ (the angle between any nonzero v and w) never exceeds 1:

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

■ 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. Which v and w give *equality* $|v \cdot w| = \|v\| \, \|w\|$ and $\|v + w\| = \|v\| + \|w\|$?

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

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

Equality: One vector is a multiple of the other as in $w = cv$. Then the angle is $0°$ or $180°$. In this case $|\cos\theta| = 1$ and $|v \cdot w|$ *equals* $\|v\| \, \|w\|$. If the angle is $0°$, as in $w = 2v$, then $\|v + w\| = \|v\| + \|w\|$ (both sides give $3\|v\|$). This $v, 2v, 3v$ triangle is 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 perpendicular to $\boldsymbol{u}$, 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 two given vectors $\boldsymbol{r} = (2, -1)$ and $\boldsymbol{s} = (-1, 2)$.

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

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

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

Section 1.1 would start with columns $\boldsymbol{v}_j$. The goal is to produce $x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n = \boldsymbol{b}$. This section would start from rows $\boldsymbol{r}_i$. Now the goal is to find $\boldsymbol{x}$ with $\boldsymbol{x} \cdot \boldsymbol{r}_i = b_i$.

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 $\boldsymbol{w} \cdot \boldsymbol{v}$:

$$\boldsymbol{u} = \begin{bmatrix} -.6 \\ .8 \end{bmatrix} \qquad \boldsymbol{v} = \begin{bmatrix} 4 \\ 3 \end{bmatrix} \qquad \boldsymbol{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

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 $\boldsymbol{v} = (1, 3)$ 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) All vectors perpendicular to $V = (1, 1, 1)$ lie on a _____ in 3 dimensions.

(c) The vectors perpendicular to both $(1, 1, 1)$ 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 find a counterexample if false):

(a) If $\boldsymbol{u} = (1, 1, 1)$ is perpendicular to $\boldsymbol{v}$ and $\boldsymbol{w}$, then $\boldsymbol{v}$ is parallel to $\boldsymbol{w}$.

(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 $\|\boldsymbol{u} - \boldsymbol{v}\| = \sqrt{2}$. *Yes*!

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** $v_2 w_2 / v_1 w_1$ **of those slopes is** -1. Show that $\boldsymbol{v} \cdot \boldsymbol{w} = 0$ and the vectors are perpendicular. (The line $y = 4x$ is perpendicular to $y = -\frac{1}{4}x$.)

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 for c starting from *any* nonzero $\boldsymbol{v}$ and $\boldsymbol{w}$.

13 Find nonzero 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 18–28 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 $(\boldsymbol{v} - \boldsymbol{w}) \cdot (\boldsymbol{v} - \boldsymbol{w}) = \boldsymbol{v} \cdot \boldsymbol{v} - 2\boldsymbol{v} \cdot \boldsymbol{w} + \boldsymbol{w} \cdot \boldsymbol{w}$:

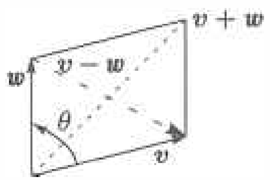
Cosine Law $$\|\boldsymbol{v} - \boldsymbol{w}\|^2 = \|\boldsymbol{v}\|^2 - 2\|\boldsymbol{v}\|\,\|\boldsymbol{w}\| \cos\theta + \|\boldsymbol{w}\|^2.$$

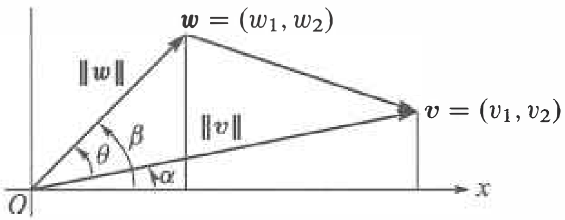
Draw a triangle with sides $\boldsymbol{v}$ and $\boldsymbol{w}$ and $\boldsymbol{v} - \boldsymbol{w}$. Which of the angles is θ ?

21 The ***triangle inequality*** says: (length of $\boldsymbol{v} + \boldsymbol{w}$) $\leq$ (length of $\boldsymbol{v}$) + (length of $\boldsymbol{w}$).

Problem 19 found $\|\boldsymbol{v} + \boldsymbol{w}\|^2 = \|\boldsymbol{v}\|^2 + 2\boldsymbol{v} \cdot \boldsymbol{w} + \|\boldsymbol{w}\|^2$. Increase that $\boldsymbol{v} \cdot \boldsymbol{w}$ to $\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$ to show that $\|$**side 3**$\|$ can not exceed $\|$**side 1**$\|$ + $\|$**side 2**$\|$:

Triangle inequality $\quad \|\boldsymbol{v} + \boldsymbol{w}\|^2 \leq (\|\boldsymbol{v}\| + \|\boldsymbol{w}\|)^2$ **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 $(v_1w_1 + v_2w_2)^2 \leq \left(v_1^2 + v_2^2\right)\left(w_1^2 + w_2^2\right)$.

(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 $\cos\alpha = v_1/\|\boldsymbol{v}\|$ and $\sin\alpha = v_2/\|\boldsymbol{v}\|$. 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 inequality $|\boldsymbol{u}\cdot\boldsymbol{U}|\le 1$ for unit vectors (u_1,u_2) and (U_1,U_2):

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

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 (*Recommended*) Draw a parallelogram

27 Parallelogram with two sides $\boldsymbol{v}$ and $\boldsymbol{w}$. Show that the squared diagonal lengths $\|\boldsymbol{v}+\boldsymbol{w}\|^2+\|\boldsymbol{v}-\boldsymbol{w}\|^2$ add to the sum of four squared side lengths $2\|\boldsymbol{v}\|^2+2\|\boldsymbol{w}\|^2$.

28 If $\boldsymbol{v}=(1,2)$ draw all vectors $\boldsymbol{w}=(x,y)$ in the xy plane with $\boldsymbol{v}\cdot\boldsymbol{w}=x+2y=5$. Why do those $\boldsymbol{w}$'s lie along a line? Which is the shortest $\boldsymbol{w}$?

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

Challenge Problems

30 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 ...).

31 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 $-\frac{1}{2}$.

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

33 Find 4 perpendicular unit vectors of the form $(\pm\frac{1}{2},\pm\frac{1}{2},\pm\frac{1}{2},\pm\frac{1}{2})$: Choose $+$ or $-$.

34 Using $\boldsymbol{v}=\mathsf{randn}(3,1)$ in MATLAB, create a random unit vector $\boldsymbol{u}=\boldsymbol{v}/\|\boldsymbol{v}\|$. Using $V=\mathsf{randn}(3,30)$ 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 is $\int_0^\pi|\cos\theta|d\theta/\pi=2/\pi$.

1.3 Matrices

1 $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$ is a 3 by 2 matrix : $m = 3$ rows and $n = 2$ columns.

2 $A\boldsymbol{x} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is a **combination of the columns** $\quad A\boldsymbol{x} = x_1 \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$.

3 The 3 components of $A\boldsymbol{x}$ are dot products of the 3 rows of A with the vector $\boldsymbol{x}$:

Row at a time $\quad \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 7 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 \cdot 7 + 2 \cdot 8 \\ 3 \cdot 7 + 4 \cdot 8 \\ 5 \cdot 7 + 6 \cdot 8 \end{bmatrix} = \begin{bmatrix} 23 \\ 53 \\ 83 \end{bmatrix}$.

4 Equations in matrix form $A\boldsymbol{x} = \boldsymbol{b}$: $\begin{bmatrix} 2 & 5 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ replaces $\begin{matrix} 2x_1 + 5x_2 = b_1 \\ 3x_1 + 7x_2 = b_2 \end{matrix}$.

5 The solution to $A\boldsymbol{x} = \boldsymbol{b}$ can be written as $\boldsymbol{x} = A^{-1}\boldsymbol{b}$. But some matrices don't allow A^{-1}.

This section starts with three vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$. I will combine them using *matrices*.

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

Their linear combinations in three-dimensional space are $x_1\boldsymbol{u} + x_2\boldsymbol{v} + x_3\boldsymbol{w}$:

Combination of the vectors
$$x_1 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix}. \qquad (1)$$

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*" the vector (x_1, x_2, x_3) :

Matrix times vector
Combination of columns
$$A\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix}. \qquad (2)$$

The numbers x_1, x_2, x_3 are the components of a vector $\boldsymbol{x}$. The matrix A times the vector $\boldsymbol{x}$ is the **same** as the combination $x_1\boldsymbol{u} + x_2\boldsymbol{v} + x_3\boldsymbol{w}$ of the three columns in equation (1).

This is more than a definition of $A\boldsymbol{x}$, because the rewriting brings a crucial change in viewpoint. At first, the numbers x_1, x_2, x_3 were multiplying the vectors. Now the matrix is multiplying those numbers. **The matrix A acts on the vector x.** The output Ax is a **combination b of the columns of A.**

To see that action, I will write b_1, b_2, b_3 for the components of Ax:

$$Ax = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{x_1} \\ \boldsymbol{x_2 - x_1} \\ \boldsymbol{x_3 - x_2} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \boldsymbol{b}\,. \tag{3}$$

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 $\boldsymbol{x} = (1, 4, 9)$: squares in $\boldsymbol{x}$, odd numbers in $\boldsymbol{b}$.

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

That pattern would continue for a 4 by 4 difference matrix. The next square would be $x_4 = 16$. The next difference would be $x_4 - x_3 = 16 - 9 = 7$ (the next odd number). The matrix finds all the differences $1, 3, 5, 7$ at once.

Important Note: Multiplication a row at a time. 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 $\boldsymbol{x}$:

$A\boldsymbol{x}$ is also dot products with rows
$$A\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} (1,\ 0,\ 0) \cdot (x_1, x_2, x_3) \\ (-1, 1, 0) \cdot (x_1, x_2, x_3) \\ (0, -1, 1) \cdot (x_1, x_2, x_3) \end{bmatrix}. \tag{5}$$

Those dot products are the same x_1 and $x_2 - x_1$ and $x_3 - x_2$ that we wrote in equation (3). 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 $A\boldsymbol{x}$ by rows. With letters, columns are the good way. Chapter 2 will repeat these rules of matrix multiplication, and explain the ideas.

Linear Equations

One more change in viewpoint is crucial. Up to now, the numbers x_1, x_2, x_3 were known. The right hand side $\boldsymbol{b}$ was not known. We found that vector of differences by multiplying A times $\boldsymbol{x}$. **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. I will now solve that system $A\boldsymbol{x} = \boldsymbol{b}$ to find x_1, x_2, x_3:

Equations $Ax = b$	$\begin{aligned} x_1 \quad\quad\quad &= b_1 \\ -x_1 + x_2 \quad\quad &= b_2 \\ -x_2 + x_3 &= b_3 \end{aligned}$	**Solution** $x = A^{-1}b$	$\begin{aligned} x_1 &= b_1 \\ x_2 &= b_1 + b_2 \\ x_3 &= b_1 + b_2 + b_3. \end{aligned}$	(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 can be solved in order* (top to bottom) *because A is a triangular matrix.*

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 $\boldsymbol{x}$. We write $\boldsymbol{x}$ as $A^{-1}\boldsymbol{b}$.

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 in equation** (7) **is the inverse** A^{-1} **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 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$:

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad A^{-1}\boldsymbol{b} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

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.** The matrix A^{-1} produces $\boldsymbol{x} = A^{-1}\boldsymbol{b}$.

The next chapters ask about other equations $A\boldsymbol{x} = \boldsymbol{b}$. Is there a solution? How to find it?

Note on calculus. Let me connect these special matrices 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 sums $A^{-1}\boldsymbol{b}$ become the *integral* of $b(t)$. **Sums of differences are like integrals of derivatives.**

The Fundamental Theorem of Calculus says: ***integration is the inverse of differentiation***.

$$\boldsymbol{Ax = b \text{ and } x = A^{-1}b} \qquad \frac{dx}{dt} = b \text{ and } x(t) = \int_0^t b\,dt. \tag{8}$$

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 $2t - 1$:

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

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

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

Difference matrices are great. Centered is the 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 $\boldsymbol{w}^*$:

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

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

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

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** (sometimes) or else **no solution** (usually):

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

Every constant vector like $\boldsymbol{x} = (3, 3, 3)$ has zero differences when we go cyclically. The undetermined constant c is exactly like the $+\,C$ that we add to integrals. The cyclic differences cycle around to $x_1 - x_3$ in the first component, instead of starting from $x_0 = 0$.

The more likely possibility for $C\boldsymbol{x} = \boldsymbol{b}$ is **no solution** $\boldsymbol{x}$ at all:

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

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 $C\boldsymbol{x} = \boldsymbol{b}$, 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 $x_1\boldsymbol{u} + x_2\boldsymbol{v} + x_3\boldsymbol{w}^*$ **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 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}^*$ (all in the same plane).

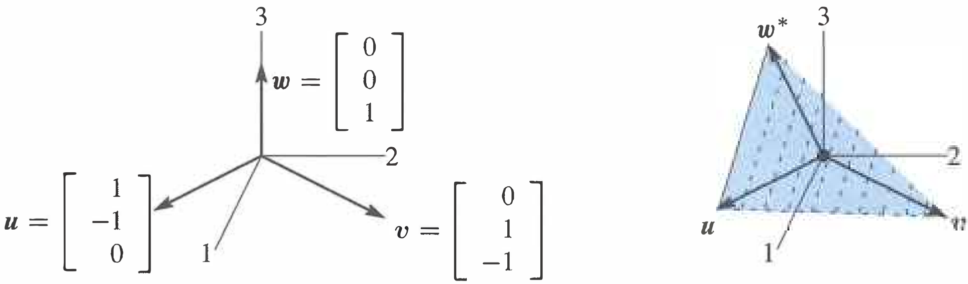


Figure 1.10: Independent vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$. Dependent vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}^*$ in a plane.

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 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ 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 like $\boldsymbol{u} + \boldsymbol{v} + \boldsymbol{w}^*$ 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: $C\boldsymbol{x} = \boldsymbol{0}$ has many solutions. C 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 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.
4. 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 $a_{31} = \mathbf{1}$:

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

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

Solution Solve the (linear triangular) system $A\boldsymbol{x} = \boldsymbol{b}$ from top to bottom:

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

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 $\boldsymbol{b} = (1,0,0)$. The second column is the solution for $\boldsymbol{b} = (0,1,0)$. The third column $\boldsymbol{x}$ of A^{-1} is the solution for $A\boldsymbol{x} = \boldsymbol{b} = (0,0,1)$.

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 $\boldsymbol{b} = (b_1, b_2, b_3)$. Those weights come from $\boldsymbol{x} = A^{-1}\boldsymbol{b}$.

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

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

The first equation is $x_1 = b_1$. The second equation is $x_2 - \ell x_1 = b_2$. The inverse will *add* ℓb_1 to b_2, because the elimination matrix *subtracted* :

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

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

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

$C\boldsymbol{x} = \boldsymbol{b}$ can only be solved when $b_1 + b_3 = x_2 - x_2 = 0$. That is a plane of vectors $\boldsymbol{b}$ 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 $C\boldsymbol{x}$).

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

$$\begin{array}{l} C\boldsymbol{x} = \boldsymbol{b} \\ \textbf{Centered} \\ \textbf{differences} \end{array} \qquad \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_2 - 0 \\ x_3 - x_1 \\ x_4 - x_2 \\ 0 - x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \tag{16}$$

Surprisingly this matrix is now invertible! The first and last rows tell you 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 $3s_1 + 4s_2 + 5s_3 = b$. Then write b as a matrix-vector multiplication Sx, with $3, 4, 5$ in x. Compute the three dot products (row of S) $\cdot\, x$:

$$s_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad s_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \quad s_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \text{ go into the columns of } S.$$

2 Solve these equations $Sy = b$ with s_1, s_2, s_3 in the columns of S:

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

S is a sum matrix. The sum of the first 5 odd numbers is _____.

3 Solve these three equations for y_1, y_2, y_3 in terms of c_1, c_2, c_3:

$$Sy = c \qquad \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}.$$

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

4 Find a combination $x_1 w_1 + x_2 w_2 + x_3 w_3$ that gives the zero vector with $x_1 = 1$:

$$w_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad w_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \quad w_3 = \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}.$$

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

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

$$r_1 = \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix} \quad r_2 = \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} \quad r_3 = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}.$$

Linear algebra says that these vectors must also lie in a plane. There must be many combinations with $y_1 r_1 + y_2 r_2 + y_3 r_3 = \mathbf{0}$. Find two sets of y's.

6 Which numbers c give dependent columns so a combination of columns equals zero ?

$$\begin{bmatrix} 1 & 1 & 0 \\ 3 & 2 & 1 \\ 7 & 4 & c \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & c \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \qquad \begin{bmatrix} c & c & c \\ 2 & 1 & 5 \\ 3 & 3 & 6 \end{bmatrix} \begin{matrix} \text{maybe} \\ \text{always} \\ \text{independent for } c \neq 0\,? \end{matrix}$$

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

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

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 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ to find the inverse matrix:

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

9 What is the *cyclic* 4 by 4 difference matrix C? It will have 1 and -1 in each row and each column. 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. *Hard to visualize.*)

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:

If $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ has dependent rows, then it also has dependent columns.

Chapter 2

Solving Linear Equations

2.1 Vectors and Linear Equations

1 The **column picture of** $A\boldsymbol{x} = \boldsymbol{b}$: a combination of n columns of A produces the vector $\boldsymbol{b}$.

2 This is a vector equation $A\boldsymbol{x} = x_1\boldsymbol{a}_1 + \cdots + x_n\boldsymbol{a}_n = \boldsymbol{b}$: the columns of A are $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$.

3 When $\boldsymbol{b} = \mathbf{0}$, a combination $A\boldsymbol{x}$ of the columns is *zero*: one possibility is $\boldsymbol{x} = (0, \ldots, 0)$.

4 The **row picture of** $A\boldsymbol{x} = \boldsymbol{b}$: m equations from m rows give m planes meeting at $\boldsymbol{x}$.

5 A dot product gives the equation of each plane : $(\textbf{row } \mathbf{1}) \cdot \boldsymbol{x} = b_1, \ \ldots, (\textbf{row } \boldsymbol{m}) \cdot \boldsymbol{x} = b_m$.

6 When $\boldsymbol{b} = \mathbf{0}$, all the planes $(\textbf{row } \boldsymbol{i}) \cdot \boldsymbol{x} = 0$ go through the center point $\boldsymbol{x} = (0, 0, \ \ldots, \ 0)$.

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 small. But you will see how far it leads:

Two equations
Two unknowns

$$\begin{aligned} x - 2y &= 1 \\ 3x + 2y &= 11 \end{aligned} \qquad (1)$$

We begin *a row at a time*. The first equation $\boldsymbol{x} - \mathbf{2}\boldsymbol{y} = \mathbf{1}$ 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 will show that first line $x - 2y = 1$. The second line in this "row picture" comes from the second equation $3x + 2y = 11$. You can't miss the point $x = 3, y = 1$ where the two lines meet. *That point* $(3, 1)$ *lies on both lines and solves both equations.*

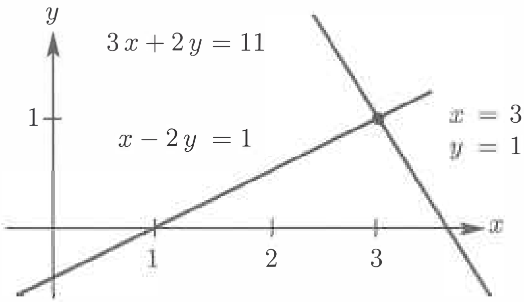


Figure 2.1: *Row picture*: The point $(3, 1)$ where the lines meet solves both equations.

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 $3\,(\textbf{\textit{column 1}}) + 1\,(\textbf{\textit{column 2}}) = \boldsymbol{b}$.

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

Figure 2.2 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 $(1, 11)$, the desired vector $\boldsymbol{b}$.

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

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

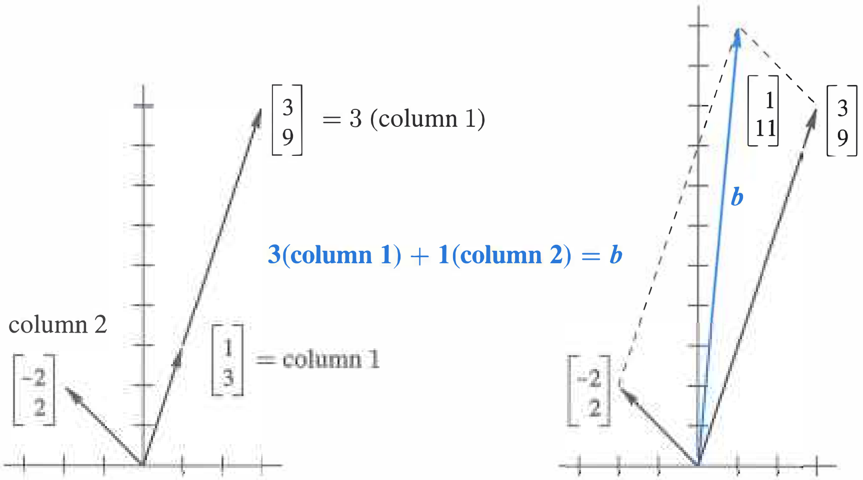


Figure 2.2: *Column picture*: A combination of columns produces the right side (1, 11).

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: ***Multiply by* 3 *and* 1*, then add.***

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

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 combine those equations into a matrix problem $A\boldsymbol{x} = \boldsymbol{b}$:

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

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:

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

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 triangular system is solved by ***back substitution***: working bottom to top.

3. In matrix language A is factored into $LU =$ (lower triangular) (upper triangular).

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

The most-used algorithm in computational science takes those steps (MATLAB calls it **lu**). Its quickest form is *backslash*: $\boldsymbol{x} = A \backslash \boldsymbol{b}$. But linear algebra goes beyond square invertible matrices! For m by n matrices, $A\boldsymbol{x} = \mathbf{0}$ 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:

$$A\boldsymbol{x} = \boldsymbol{b} \qquad \begin{aligned} x + 2y + 3z &= 6 \\ 2x + 5y + 2z &= 4 \\ 6x - 3y + z &= 2 \end{aligned} \tag{3}$$

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, in this case $3 = 3$, 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*** $b = (6, 4, 2)$.

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** $z = 2$.

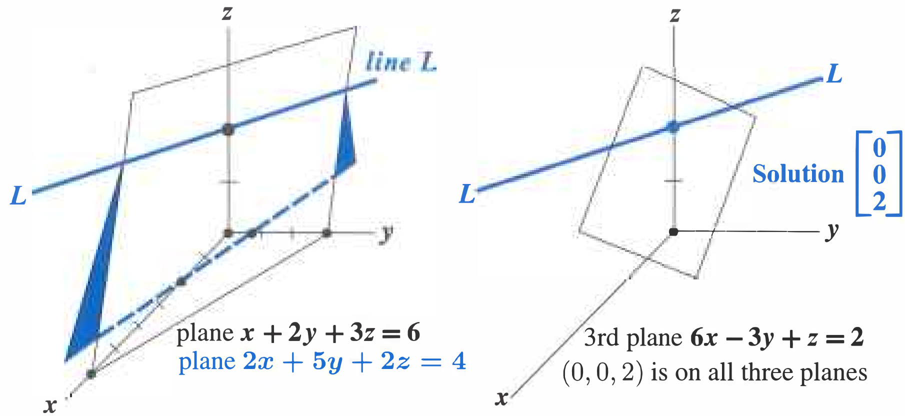


Figure 2.3: *Row picture*: Two planes meet at a line L. Three planes meet at a point.

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

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

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 $\boldsymbol{b} = (6, 4, 2)$.

Figure 2.4 shows this column picture. Linear combinations of those columns can produce any vector $\boldsymbol{b}$! The combination that produces $\boldsymbol{b} = (6, 4, 2)$ 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
$(x, y, z) = (\mathbf{0}, \mathbf{0}, \mathbf{2})$
$$\mathbf{0} \begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix} + \mathbf{0} \begin{bmatrix} 2 \\ 5 \\ -3 \end{bmatrix} + \mathbf{2} \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix}.$$

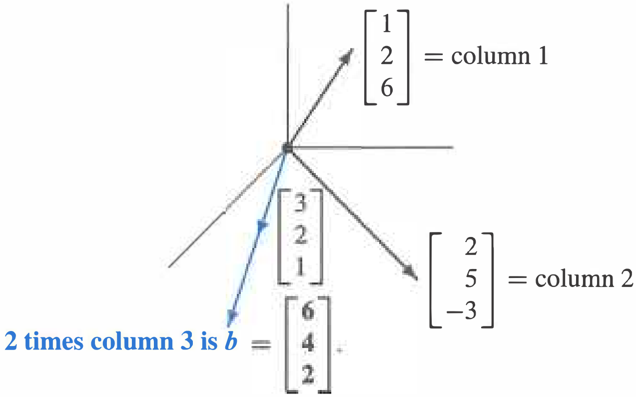


Figure 2.4: *Column picture: Combine the columns with weights* $(x, y, z) = (0, 0, 2)$.

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:

The "coefficient matrix" in $A\boldsymbol{x} = \boldsymbol{b}$ ***is*** $\quad A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 2 \\ 6 & -3 & 1 \end{bmatrix}.$

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):

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

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 statement of the three equations. You do the same nine multiplications either way.

Multiplication by rows

$A\boldsymbol{x}$ 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

Ax is a ***combination of column vectors***:
$Ax = x$ (*column 1*)$+ y$ (*column 2*)$+z$ (*column 3*). (7)

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

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

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*** Ax ***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 :

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix} \qquad I\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

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 $A\boldsymbol{x}$ 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*** :

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

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 $\boldsymbol{x}$ and their product $A\boldsymbol{x}$ using MATLAB commands. This is a first step in learning that language (and others). I begin by defining A and $\boldsymbol{x}$. A vector $\boldsymbol{x}$ in $\mathbf{R}^n$ is an n by 1 matrix (as in this book). Enter matrices *a row at a time*, and use a semicolon to signal the end of a row. Or enter by columns and transpose by $'$:

$$A = [1 \quad 2 \quad 3\,; \quad 2 \quad 5 \quad 2\,; \quad 6 \,-3 \quad 1]$$
$$\boldsymbol{x} = [0 \quad 0 \quad 2]' \quad \textbf{or} \quad \boldsymbol{x} = [0\,;0\,;2]$$

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

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

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

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

Each entry of $\boldsymbol{b}$ 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 : comes first, *to keep all rows of column* 1. This column multiplies $x(1)$ and the other columns multiply $x(2)$ and $x(3)$:

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

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.

Programming Languages for Mathematics and Statistics

Here are five more important languages and their commands for the multiplication $A\boldsymbol{x}$:

Julia	$A * \boldsymbol{x}$	**julialang.org**
Python	dot(A, $\boldsymbol{x}$)	**python.org**
R	$A \% * \% \boldsymbol{x}$	**r-project.org**
Mathematica	$A\,.\,\boldsymbol{x}$	**wolfram.com/mathematica**
Maple	$A * \boldsymbol{x}$	**maplesoft.com**

Julia, Python, and **R** are free and open source languages. R is developed particularly for applications in statistics. Other software for statistics (SAS, JMP, and many more) is described on Wikipedia's Comparison of Statistical Packages.

Mathematica and **Maple** allow symbolic entries $a, b, x, \ldots$ and not only real numbers. As in MATLAB's Symbolic Toolbox, they work with symbolic expressions like $x^2 - x$. The power of Mathematica is seen in Wolfram Alpha.

Julia combines the high productivity of SciPy or R for technical computing with performance comparable to C or Fortran. It can call Python and C/Fortran libraries. But it doesn't rely on "vectorized" library functions for speed; Julia is designed to be fast.

I entered **juliabox.org**. I clicked *Sign in via Google* to access my gmail space. Then I clicked *new* at the right and chose a Julia notebook. I chose 0.4.5 and not one under development. The Julia command line came up immediately.

As a novice, I computed $1 + 1$. To see the answer I pressed *Shift+Enter*. I also learned that $1.0 + 1.0$ uses floating point, much faster for a large problem. The website **math.mit.edu/linearalgebra** will show part of the power of Julia and Python and R.

Python is a popular general-purpose programming language. When combined with packages like NumPy and the SciPy library, it provides a full-featured environment for technical computing. NumPy has the basic linear algebra commands. Download the Anaconda Python distribution from **https://www.continuum.io** (a prepackaged collection of Python and most important mathematical libraries, with a graphical installer).

R is free software for statistical computing and graphics. To download and install R, go to **r-project.org** (prefix **https://www.**). Commands are prompted by $>$ and R is a scripted language. It works with lists that can be shaped into vectors and matrices.

It is important to recommend RStudio for editing and graphing (and help resources). When you download from **www.RStudio.com**, a window opens for R commands—plus windows for editing and managing files and plots. Tell R the form of the matrix as well as the list of numerical entries:

$> A =$ matrix $(c\,(1, 2, 3, 2, 5, 2, 6, -3, 1),$ nrow $= 3,$ byrow $=$ TRUE$)$

$> x =$ matrix $(c\,(0, 0, 2),$ nrow $= 3)$

To see A and x, type their names at the new prompt $>$. To multiply type $b = A\,\% * \%\,x$. Transpose by $t(A)$ and use as.matrix to turn a vector into a matrix.

MATLAB and Julia have a cleaner syntax for matrix computations than R. But R has become very familiar and widely used. The website for this book has space for proper demos (including the *Manipulate* command) of **MATLAB** and **Julia** and **Python** and **R**.

■ 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 $A\boldsymbol{x}$ must 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 $A\boldsymbol{x} = \boldsymbol{b}$. Solve by careful inspection of the columns (instead of elimination):

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

Solution The column picture asks for a linear combination that produces $\boldsymbol{b}$ from the three columns of A. In this example $\boldsymbol{b}$ 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 $\boldsymbol{b} = (4, 4, 8) =$ sum of the first two columns. Then the good combination has $x = 1, y = 1, z = 0$. The solution becomes $\boldsymbol{x} = (1, 1, 0)$.

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 ?*

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

Idea Add (equation 1) + (equation 2) − (equation 3). The result is $\mathbf{0} = \mathbf{1}$. This system cannot have a solution. We could say: The vector $(1, 1, -1)$ is orthogonal to all three columns of A but *not* orthogonal to $\boldsymbol{b}$.

(1) Are any two of the three 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 $\boldsymbol{b}$) with $\boldsymbol{y} = (1, 1, -1)$. How do those dot products show that no combination of columns equals $\boldsymbol{b}$?

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

Solution

(1) 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, $\boldsymbol{y} \cdot \boldsymbol{b} = (1, 1, -1) \cdot (4, 5, 8) = \mathbf{1}$ is *not zero*. $A\boldsymbol{x} = \boldsymbol{b}$ led to $0 = 1$: **no solution**.

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

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

Problem Set 2.1

Problems 1–8 are about the row and column pictures of $Ax = b$.

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

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

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 $DX = B$:

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

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:

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

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

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

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 $A\boldsymbol{x} = \mathbf{0}$ 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 $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, $\boldsymbol{u}_3$:

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

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 **Multiplication by A is a "linear transformation".** Those words mean:

If $\boldsymbol{w}$ is a combination of $\boldsymbol{u}$ and $\boldsymbol{v}$, then $A\boldsymbol{w}$ is the same combination of $A\boldsymbol{u}$ and $A\boldsymbol{v}$.

It is this "***linearity***" $A\boldsymbol{w} = cA\boldsymbol{u} + dA\boldsymbol{v}$ that gives us the name *"linear algebra"*.

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

Combine $\boldsymbol{w} = c\boldsymbol{u} + d\boldsymbol{v}$. **If** $\boldsymbol{w} = \begin{bmatrix} \mathbf{5} \\ \mathbf{7} \end{bmatrix}$ **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 $A\boldsymbol{x} = \boldsymbol{b}$. 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 every column, and in every 3 by 3 block. For the all-ones vector $\boldsymbol{x} = (1, \ldots, 1)$, what is $S\boldsymbol{x}$?

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

1 For $\boldsymbol{m} = \boldsymbol{n} = 3$, there are three equations $A\boldsymbol{x} = \boldsymbol{b}$ and three unknowns x_1, x_2, x_3.

2 The first two equations are $a_{11}x_1 + \cdots = b_1$ and $a_{21}x_1 + \cdots = b_2$.

3 Multiply the first equation by a_{21}/a_{11} and subtract from the second : then x_1 **is eliminated**.

4 The corner entry a_{11} is the first "pivot" and the ratio a_{21}/a_{11} is the first "multiplier."

5 Eliminate x_1 from every remaining equation i by subtracting a_{i1}/a_{11} times the first equation.

6 Now the last $n-1$ equations contain $n-1$ unknowns $x_2, \ldots, x_n$. *Repeat to eliminate* x_2.

7 Elimination breaks down if zero appears in the pivot. Exchanging two equations may save it.

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.

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** ":*

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

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.

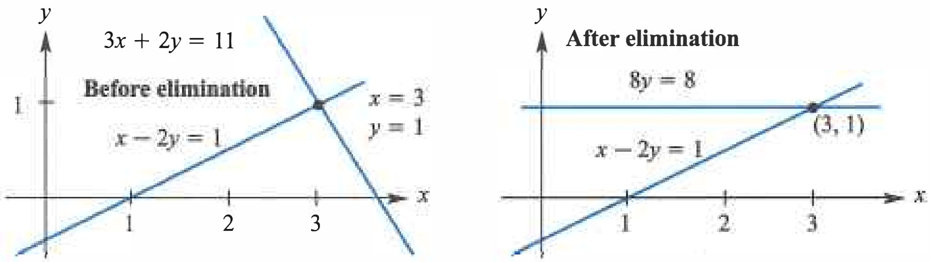


Figure 2.5: Eliminating x makes the second line horizontal. Then $8y = 8$ gives $y = 1$.

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.

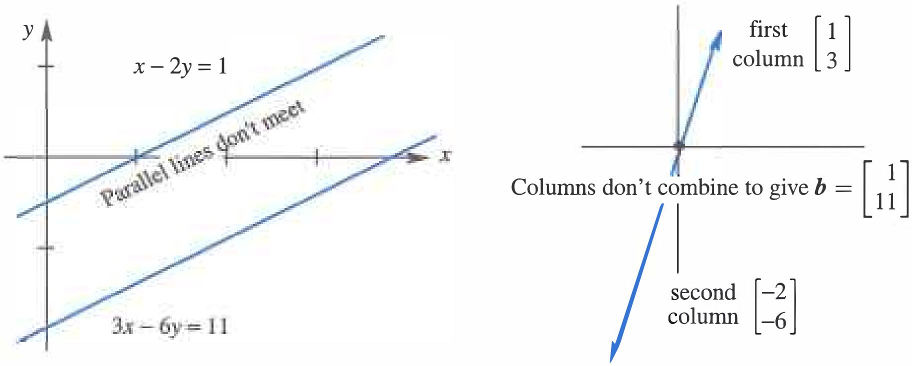


Figure 2.6: Row picture and column picture for Example 1: ***no solution***.

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

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

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.

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

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

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.

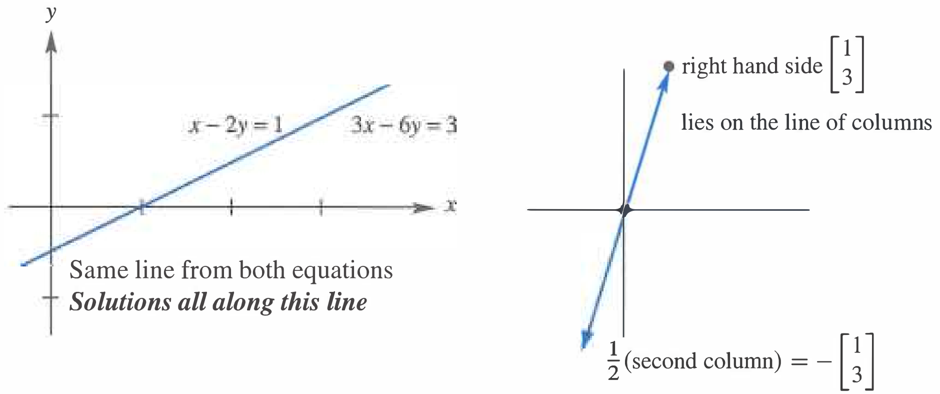


Figure 2.7: Row and column pictures for Example 2: *infinitely many solutions.*

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:***

Permutation $\begin{aligned} 0x + 2y &= 4 \\ 3x - 2y &= 5 \end{aligned}$ Exchange the two equations $\begin{aligned} 3x - 2y &= 5 \\ 2y &= 4. \end{aligned}$

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.

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 elimination 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 $U\boldsymbol{x} = \boldsymbol{c}$:

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

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):

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

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

Elimination from A to U

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

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

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 won't be changed in elimination 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:

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

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. When zero is in the pivot position, **exchange rows** if there is a nonzero below it.

5. The upper triangular $U\boldsymbol{x} = \boldsymbol{c}$ is solved by **back substitution** (starting at the bottom).

6. When **breakdown** is permanent, $A\boldsymbol{x} = \boldsymbol{b}$ 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)?

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

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 $1, 1$. When 1 times row 1 is subtracted from row 2, the second pivot is revealed as another 1. If the original middle entry had been 1 instead of 2, that would have forced a row exchange.

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*". Everything stays zero outside the band.

The last pivot is also 1. So if the original corner entry $a_{33} = 2$ reduced by 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.

$$\textbf{Success} \quad \begin{aligned} x + y + z &= 7 \\ x + y - z &= 5 \\ x - y + z &= 3 \end{aligned} \qquad\qquad \textbf{Failure} \quad \begin{aligned} x + y + z &= 7 \\ x + y - z &= 5 \\ -x - y + z &= 3 \end{aligned}$$

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 2 and 3:

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

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*:

	$x + y + z = 7$	There is **no pivot in column 2** (it was – column 1)
Failure	$0y - 2z = -2$	A further elimination step gives $\mathbf{0z = 8}$
	$0y + 2z = 10$	The three planes **don't meet**

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** (free variables 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 elimination, write down the upper triangular system and circle the two pivots. The numbers 1 and 11 don't affect the pivots—use them now in back substitution.

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

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

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.

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

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?

$$\begin{array}{llll} \textbf{Equal} & 2x - y + z = 0 & 2x + 2y + z = 0 & \textbf{Equal} \\ \textbf{rows} & 2x - y + z = 0 & 4x + 4y + z = 0 & \textbf{columns} \\ & 4x + y + z = 2 & 6x + 6y + z = 2. & \end{array}$$

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 $K\boldsymbol{x} = \boldsymbol{b}$):

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

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:

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

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 $[L, U] = \mathbf{lu}\,(\mathbf{rand}\,(3))$. The average size $\mathbf{abs}\,(U(1,1))$ is above $\frac{1}{2}$ because **lu** picks the largest available pivot in column 1. Here $A = \mathbf{rand}\,(3)$ 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

1 The first step multiplies the equations $A\boldsymbol{x} = \boldsymbol{b}$ by a matrix E_{21} to produce $E_{21}A\boldsymbol{x} = E_{21}\boldsymbol{b}$.

2 That matrix $E_{21}A$ has a zero in row 2, column 1 because x_1 is eliminated from equation 2.

3 E_{21} is the **identity matrix** (diagonal of 1's) minus the multiplier a_{21}/a_{11} in row 2, column 1.

4 Matrix-matrix multiplication is n matrix-vector multiplications: $\boldsymbol{EA} = [\boldsymbol{Ea_1} \;\ldots\; \boldsymbol{Ea_n}]$.

5 We must also multiply $E\boldsymbol{b}$! So E is multiplying the **augmented matrix** $[A\,\boldsymbol{b}] = [\boldsymbol{a}_1 \;\ldots\; \boldsymbol{a}_n\,\boldsymbol{b}]$.

6 Elimination multiplies $A\boldsymbol{x} = \boldsymbol{b}$ by $E_{21}, E_{31}, \ldots, E_{n1}$, then $E_{32}, E_{42}, \ldots, E_{n2}$, and onward.

7 The **row exchange matrix** is not E_{ij} but P_{ij}. To find P_{ij}, exchange rows i and j of I.

This section gives our first examples of **matrix multiplication**. Naturally we start with matrices that contain many zeros. Our goal is to see that matrices *do something*. E acts on a vector $\boldsymbol{b}$ or a matrix A to produce a new vector $E\boldsymbol{b}$ or a new matrix EA.

Our first examples will be "**elimination matrices**." They execute the elimination steps. Multiply the j^{th} equation by ℓ_{ij} and subtract from the i^{th} equation. (This eliminates x_j from equation i.) We need a lot of these simple matrices E_{ij}, one for every nonzero to be eliminated below the main diagonal.

Fortunately we won't see all these matrices E_{ij} in later chapters. They are good examples to start with, but there are too many. They can combine into one overall matrix E that takes all steps at once. The neatest way is to combine all their inverses $(E_{ij})^{-1}$ into one overall matrix $L = E^{-1}$. Here is the purpose of the next pages.

1. To see how each step is a matrix multiplication.
2. To assemble all those steps E_{ij} into one elimination matrix E.
3. To see how each E_{ij} is inverted by its inverse matrix E_{ij}^{-1}.
4. To assemble all those inverses E_{ij}^{-1} (in the right order) into L.

The special property of L is that all the multipliers ℓ_{ij} fall into place. Those numbers are mixed up in E (forward elimination from A to U). They are perfect in L (undoing elimination, returning from U to A). Inverting puts the steps and their matrices E_{ij}^{-1} in the opposite order and that prevents the mixup.

This section finds the matrices E_{ij}. Section 2.4 presents four ways to multiply matrices. Section 2.5 inverts every step. (For elimination matrices we can already see E_{ij}^{-1} here.) Then those inverses go into L.

Matrices times Vectors and $Ax = b$

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

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

The nine numbers on the left go into the matrix A. That matrix not only sits beside $\boldsymbol{x}$. *A 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.

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

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

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

$\boldsymbol{Ax}$ is a combination of the columns of A. To compute each component of $\boldsymbol{Ax}$, we use the **row form** of matrix multiplication. ***Components of $\boldsymbol{Ax}$ are dot products with rows of A.*** The short formula for that dot product with $\boldsymbol{x}$ uses "sigma notation".

The first component of $\boldsymbol{Ax}$ above is $\quad (-1)(2) + (2)(4) + (2)(-2)$.

The ith component of $\boldsymbol{Ax}$ is $\quad (\text{row } i) \cdot \boldsymbol{x} = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n$.

This is sometimes written with the sigma symbol as $\sum_{j=1}^{n} \boldsymbol{a_{ij}x_j}$.

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

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 $\boldsymbol{Ax}$ by rows with numbers and letters:

$$\begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \cdot 2 + 4 \cdot 1 \\ 5 \cdot 2 + 6 \cdot 1 \end{bmatrix} \qquad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}.$$

A row times a column gives a dot product.

[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 $\sum$.

The Matrix Form of One Elimination Step

$A\boldsymbol{x} = \boldsymbol{b}$ is a convenient form for the original equation. What about the elimination steps? In this example, 2 times the first equation is subtracted from the second equation. On the right side, 2 times the first component of $\boldsymbol{b}$ is subtracted from the second component.

$$\textbf{First step} \qquad \boldsymbol{b} = \begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix} \quad \text{changes to} \quad \boldsymbol{b}_{\text{new}} = \begin{bmatrix} 2 \\ 4 \\ 10 \end{bmatrix}.$$

We want to do that subtraction with a matrix! The same result $\boldsymbol{b}_{\text{new}} = E\boldsymbol{b}$ is achieved when we multiply an "elimination matrix" E times $\boldsymbol{b}$. It subtracts $2b_1$ from b_2:

$$\textbf{\textit{The elimination matrix is}} \qquad E = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Multiplication by E subtracts 2 times row 1 from row 2. Rows 1 and 3 stay the same:

$$\begin{bmatrix} 1 & 0 & 0 \\ \mathbf{-2} & \mathbf{1} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{2} \\ \mathbf{8} \\ 10 \end{bmatrix} = \begin{bmatrix} 2 \\ \mathbf{4} \\ 10 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 \\ \mathbf{-2} & \mathbf{1} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{b_1} \\ \boldsymbol{b_2} \\ b_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ \boldsymbol{b_2 - 2b_1} \\ b_3 \end{bmatrix}$$

The first and third rows of E come from the identity matrix I. They don't change the first and third numbers (2 and 10). 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 $I\boldsymbol{b} = \boldsymbol{b}$ for all $\boldsymbol{b}$. The ***elementary matrix or elimination matrix*** E_{ij} has the extra nonzero entry $-\ell$ in the i, j position. Then E_{ij} subtracts a multiple ℓ of row j from row i.

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 $9 - 4 = 5$:

$$I\boldsymbol{b} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix} \quad \text{and} \quad E\boldsymbol{b} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}.$$

What about the left side of $A\boldsymbol{x} = \boldsymbol{b}$? Both sides will be multiplied by this 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 $\boldsymbol{x}$ 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 use.

Matrix Multiplication

The big question is: ***How do we multiply two matrices?*** When the first matrix is E, we know what to expect for EA. This particular E subtracts 2 times row 1 from row 2. The multiplier is $\ell = 2$:

$$EA = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{2} & \mathbf{4} & \mathbf{-2} \\ \mathbf{4} & \mathbf{9} & \mathbf{-3} \\ -2 & -3 & 7 \end{bmatrix} = \begin{bmatrix} 2 & 4 & -2 \\ \mathbf{0} & \mathbf{1} & \mathbf{1} \\ -2 & -3 & 7 \end{bmatrix} \quad \textit{(with the zero).} \tag{3}$$

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 Ax, the second is EA times 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. 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—not the rows. AE actually subtracts 2 times column 2 from column 1. So $EA \neq AE$.

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 $$AB = A\,[b_1\ b_2\ b_3] = [Ab_1\ Ab_2\ Ab_3]. \tag{4}$$

This holds true for the matrix multiplication in (3). 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:

Permutation matrix $$P_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

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.

To exchange equations 1 and 3 multiply by $P_{13} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$.

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 $\boldsymbol{b}$:

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 $\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix}$ to give a row of $\begin{bmatrix} EA & E\boldsymbol{b} \end{bmatrix}$.

COLUMNS E acts on each column of $\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix}$ to give a column of $\begin{bmatrix} EA & E\boldsymbol{b} \end{bmatrix}$.

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. $A\boldsymbol{x} = x_1$ times column $1 + \cdots + x_n$ times column n. And $(A\boldsymbol{x})_i = \sum_{j=1}^{n} a_{ij}x_j$.
2. Identity matrix $= I$, elimination matrix $= E_{ij}$ using ℓ_{ij}, exchange matrix $= P_{ij}$.
3. Multiplying $A\boldsymbol{x} = \boldsymbol{b}$ 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 $\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix}$, that elimination step gives $\begin{bmatrix} E_{21}A & E_{21}\boldsymbol{b} \end{bmatrix}$.
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 also appears in column 2:

$$[A \quad b] = \begin{bmatrix} 1 & 2 & 2 & 1 \\ 4 & 8 & 9 & 3 \\ 0 & 3 & 2 & 1 \end{bmatrix} \quad \text{and} \quad E_{21}[A \quad b] = \begin{bmatrix} 1 & 2 & 2 & 1 \\ \mathbf{0} & \mathbf{0} & 1 & -1 \\ 0 & 3 & 2 & 1 \end{bmatrix}$$

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:

$$(\text{row } 1) \cdot (\text{column } 1) = \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \mathbf{10} \quad \text{is the } (1,1) \text{ entry of } AB$$

$$(\text{row } 2) \cdot (\text{column } 1) = \begin{bmatrix} 1 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \mathbf{7} \quad \text{is the } (2,1) \text{ entry of } AB$$

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 $E = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{-2} & \mathbf{1} \end{bmatrix}$ 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.

$$\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & a \\ 0 & 4 & 5 & b \\ 0 & 0 & d & c \end{bmatrix}$$

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:

Elimination on 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 $M = \left[\begin{smallmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{c} & \mathbf{d} \end{smallmatrix}\right] > 0$ that have $ad - bc = 1$.

31 Find elimination matrices E_{21} then E_{32} then E_{43} to change K into U:

$$E_{43}\, E_{32}\, E_{21} \begin{bmatrix} 1 & 0 & 0 & 0 \\ -a & 1 & 0 & 0 \\ 0 & -b & 1 & 0 \\ 0 & 0 & -c & 1 \end{bmatrix} = I.$$

Apply those three steps to the identity matrix I, to multiply $E_{43}E_{32}E_{21}$.

2.4 Rules for Matrix Operations

1 Matrices A with n columns multiply matrices B with n rows : $\boxed{A_{m\times n}\, B_{n\times p} = C_{m\times p}.}$

2 Each entry in $AB = C$ is a dot product : $C_{ij} =$ (row i of A) $\cdot$ (column j of B).

3 This rule is chosen so that ***AB* times *C* equals *A* times *BC***. And $(AB)\boldsymbol{x} = A(B\boldsymbol{x})$.

4 More ways to compute AB : (A times columns of B) (rows of A times B) (*columns times rows*).

5 It is not usually true that $AB = BA$. In most cases A *doesn't commute with* B.

6 Matrices can be multiplied by *blocks* : $A = [A_1\ \ A_2]$ times $B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$ is $A_1B_1 + A_2B_2$.

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? *This section gives* 4 *ways to find* AB. But 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 return to this *column way*: A times columns of B. Both ways follow this rule:

Fundamental Law of Matrix Multiplication ***AB times C equals A times BC*** (1)

The parentheses can move safely in $(AB)C = A(BC)$. Linear algebra depends on this law.

Suppose A is m by n and B is n by p. We can multiply. The product AB is m by p.

$$(\boldsymbol{m}\times n)(n\times \boldsymbol{p}) = (\boldsymbol{m}\times \boldsymbol{p}) \qquad \begin{bmatrix} m \textbf{ rows} \\ n \textit{ columns} \end{bmatrix} \begin{bmatrix} n \textit{ rows} \\ p \textbf{ columns} \end{bmatrix} = \begin{bmatrix} m \textbf{ rows} \\ p \textbf{ columns} \end{bmatrix}.$$

A row times a column is an extreme case. Then 1 by n multiplies n by 1. The result will be 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). This is the first way, and the usual way, to multiply matrices. ***Take the dot product of each row of A with each column of B.***

1. ***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.

$$\begin{bmatrix} * & & & \\ a_{i1} & a_{i2} & \cdots & a_{i5} \\ * & & & \\ * & & & \end{bmatrix} \begin{bmatrix} * & * & b_{1j} & * & * & * \\ & & b_{2j} & & & \\ & & \vdots & & & \\ & & b_{5j} & & & \end{bmatrix} = \begin{bmatrix} & & * & & & \\ * & * & (AB)_{ij} & * & * & * \\ & & * & & & \\ & & * & & & \end{bmatrix}$$

A is 4 by 5 $\qquad$ B is 5 by 6 $\qquad$ AB is $(4\times 5)(5\times 6) = 4$ by 6

Figure 2.8: Here $i=2$ and $j=3$. Then $(AB)_{23}$ is (**row 2**) $\cdot$ (**column 3**) $=$ sum of $a_{2k}b_{k3}$.

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*** $\boldsymbol{n^3}$ ***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 to multiply large matrices. Instead of n^3 multiplications the count has now dropped to $n^{2.376}$. Maybe n^2 is possible? But the algorithms are so awkward that scientific computing is done the regular n^3 way.

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
$(\boldsymbol{n} \times \mathbf{1})(\mathbf{1} \times \boldsymbol{n}) = (\boldsymbol{n} \times \boldsymbol{n})$

$$\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix}.$$

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.

The Second and Third Ways: Rows and Columns

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 the columns of A.*** That is the column picture of matrix multiplication:

2. Matrix A times every column of B $\qquad A\left[\boldsymbol{b}_1 \cdots \boldsymbol{b}_p\right] = \left[A\boldsymbol{b}_1 \cdots A\boldsymbol{b}_p\right].$

The row picture is reversed. Each row of A multiplies the whole matrix B. The result is a row of AB. **Every row of AB is a combination of the rows of B**:

3. Every row of A times matrix B $\qquad \left[\textbf{row } i \textbf{ of } A\right] \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \left[\textbf{row } i \textbf{ of } AB\right].$

We see row operations in elimination (E times A). Soon we see columns in $AA^{-1} = I$. The "row-column picture" has the dot products of rows with columns. Dot products are the usual way to multiply matrices by hand: mnp separate steps of multiply/add.

$$\boldsymbol{AB} = (\boldsymbol{m} \times \boldsymbol{n})(\boldsymbol{n} \times \boldsymbol{p}) = (\boldsymbol{m} \times \boldsymbol{p}) \quad \boldsymbol{mp} \textbf{ dot products with } \boldsymbol{n} \textbf{ steps each} \tag{2}$$

The Fourth Way: Columns Multiply Rows

There is a fourth way to multiply matrices. Not many people realize how important this is. I feel like a magician explaining a trick. Magicians won't do it but mathematicians try. The fourth way was in previous editions of this book, but I didn't emphasize it enough.

4. Multiply columns 1 to n of A times rows 1 to n of B. Add those matrices.

Column 1 of A multiplies row 1 of B. Columns 2 and 3 multiply rows 2 and 3. Then add:

$$\begin{bmatrix} \textbf{col 1} & \textbf{col 2} & \textbf{col 3} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \textbf{row 1} & \cdot & \cdot & \cdot \\ \textbf{row 2} & \cdot & \cdot & \cdot \\ \textbf{row 3} & \cdot & \cdot & \cdot \end{bmatrix} = (\textbf{col 1})\,(\textbf{row 1}) + (\textbf{col 2})\,(\textbf{row 2}) + (\textbf{col 3})\,(\textbf{row 3}).$$

If I multiply 2 by 2 matrices this column–row way, you will see that AB is correct.

$$AB = \begin{bmatrix} \boldsymbol{a} & b \\ \boldsymbol{c} & d \end{bmatrix} \begin{bmatrix} \boldsymbol{E} & \boldsymbol{F} \\ G & H \end{bmatrix} = \begin{bmatrix} \boldsymbol{aE} + bG & \boldsymbol{aF} + bH \\ \boldsymbol{cE} + dG & \boldsymbol{cF} + dH \end{bmatrix}$$

Add columns of A times rows of B

$$AB = \begin{bmatrix} a \\ c \end{bmatrix} \begin{bmatrix} E & F \end{bmatrix} + \begin{bmatrix} b \\ d \end{bmatrix} \begin{bmatrix} G & H \end{bmatrix} \tag{3}$$

Column k of A multiplies row k of B. That gives a matrix (not just a number). Then you add those matrices for $k = 1, 2, \ldots, n$ to produce AB.

If AB is (m by n) (n by p) then n matrices will be (*column*) (*row*). They are all m by p. This uses the same mnp steps as in the dot products—but in a new order.

The Laws for Matrix Operations

May I put on record six laws that matrices do obey, while emphasizing a rule 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:

$$\begin{aligned} A + B &= B + A && \text{(commutative law)} \\ c\,(A + B) &= cA + cB && \text{(distributive law)} \\ A + (B + C) &= (A + B) + C && \text{(associative law).} \end{aligned}$$

Three more laws hold for multiplication, but $AB = BA$ is not one of them:

$$\begin{aligned} \boxed{AB \neq BA} & \quad \text{(the commutative "law" is \textit{usually broken})} \\ A\,(B + C) = AB + AC & \quad \text{(distributive law from the left)} \\ (A + B)\,C = AC + BC & \quad \text{(distributive law from the right)} \\ \boxed{A\,(BC) = (AB)C} & \quad \text{(associative law for } ABC\text{) (\textbf{\textit{parentheses not needed}}).} \end{aligned}$$

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:

$$\boldsymbol{A^p = AAA \cdots A \ (p \text{ factors})} \qquad (A^p)(A^q) = A^{p+q} \qquad (A^p)^q = A^{pq}.$$

Those are the ordinary laws for exponents. A^3 times A^4 is A^7 (seven factors). But the fourth power of A^3 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 in analogy with $2^0 = 1$.

For a number, a^{-1} is $1/a$. For a matrix, the inverse is written A^{-1}. (It is *not* I/A, except 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 give
2 by 3 block matrix

$$A = \left[\begin{array}{cc|cc|cc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right] = \begin{bmatrix} I & I & I \\ I & I & I \end{bmatrix}.$$

If B is also 4 by 6 and the block sizes match, you can add $A + B$ *a block at a time*.

You have seen block matrices before. The right side vector $\boldsymbol{b}$ was placed next to A in the "augmented matrix". Then $[\,A \quad \boldsymbol{b}\,]$ has two blocks of different sizes. Multiplying by an elimination matrix gave $[\,EA \quad E\boldsymbol{b}\,]$. No problem to multiply blocks times blocks, when their shapes permit.

Block multiplication If blocks of A can multiply blocks of B, then block multiplication of AB is allowed. Cuts between columns of A match cuts between rows of B.

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} \\ A_{21}B_{11} + A_{22}B_{21} \end{bmatrix}. \tag{4}$$

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***:

$$\textbf{Columns times rows} \qquad \begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix} \begin{bmatrix} - & b_1 & - \\ & \vdots & \\ - & b_n & - \end{bmatrix} = \begin{bmatrix} a_1b_1 + \cdots + a_nb_n \end{bmatrix}. \qquad (5)$$

This is Rule 4 to multiply matrices. Here is a numerical example:

$$\begin{bmatrix} 1 & 4 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 3 & 2 \end{bmatrix} + \begin{bmatrix} 4 \\ 5 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 3 & 2 \end{bmatrix} + \begin{bmatrix} 4 & 0 \\ 5 & 0 \end{bmatrix} = \begin{bmatrix} 7 & 2 \\ 8 & 2 \end{bmatrix}.$$

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).

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} :

$$\textbf{One at a time} \qquad E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad E_{31} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}.$$

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 $a = 1$:

$$E = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix} \quad \text{multiplies} \quad \begin{bmatrix} \mathbf{1} & x & x \\ \mathbf{3} & x & x \\ \mathbf{4} & x & x \end{bmatrix} \quad \text{to give} \quad EA = \begin{bmatrix} \mathbf{1} & x & x \\ \mathbf{0} & y & y \\ \mathbf{0} & z & z \end{bmatrix}.$$

Using inverse matrices, a block matrix E can do elimination on a whole (block) column. Suppose a matrix has four blocks A, B, C, D. Watch how E eliminates C by blocks :

$$\textbf{Block elimination} \qquad \left[\begin{array}{c|c} I & \mathbf{0} \\ \hline -CA^{-1} & I \end{array}\right] \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = \left[\begin{array}{c|c} A & B \\ \hline \mathbf{0} & \boldsymbol{D - CA^{-1}B} \end{array}\right]. \qquad (6)$$

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 $S = D - CA^{-1}B$.

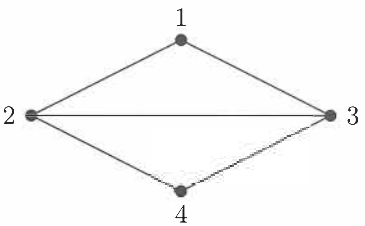
This is ordinary elimination, a column at a time—using blocks. The pivot block is A. That final block is $D - CA^{-1}B$, just like $d - cb/a$. This is called the ***Schur complement***.

■ REVIEW OF THE KEY IDEAS ■

1. The (i, j) entry of AB is (row i of A) · (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 n 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 A graph or a network has n nodes. Its **adjacency matrix** S is n by n. This is a 0–1 matrix with $s_{ij} = 1$ when nodes i and j are connected by an edge.



Adjacency matrix
Square and symmetric
for undirected graphs
Edges go both ways

$$S = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

The matrix S^2 has a useful interpretation. $(\boldsymbol{S^2})_{ij}$ **counts the walks of length 2** between node i and node j. Between nodes 2 and 3 the graph has two walks: go via 1 or go via 4. From node 1 to node 1, there are also two walks: 1–2–1 and 1–3–1.

$$S^2 = \begin{bmatrix} 2 & 1 & 1 & 2 \\ 1 & 3 & 2 & 1 \\ 1 & 2 & 3 & 1 \\ 2 & 1 & 1 & 2 \end{bmatrix} \qquad S^3 = \begin{bmatrix} 2 & \mathbf{5} & 5 & 2 \\ 5 & 4 & 5 & 5 \\ 5 & 5 & 4 & 5 \\ 2 & 5 & 5 & 2 \end{bmatrix}$$

Can you find 5 walks of length 3 between nodes 1 and 2 ?

The real question is why S^N counts all the N-step paths between pairs of nodes. Start with S^2 and look at matrix multiplication by dot products:

$$(S^2)_{ij} = (\text{row } i \text{ of } S) \cdot (\text{column } j \text{ of } S) = s_{i1}s_{1j} + s_{i2}s_{2j} + s_{i3}s_{3j} + s_{i4}s_{4j}. \qquad (7)$$

If there is a 2-step path $i \to 1 \to j$, the first multiplication gives $s_{i1}s_{1j} = (1)(1) = 1$. If $i \to 1 \to j$ is *not* a path, then either $i \to 1$ is missing or $1 \to j$ is missing. So the multiplication gives $s_{i1}s_{1j} = 0$ in that case.

$(S^2)_{ij}$ is adding up 1's for all the 2-step paths $i \to k \to j$. So it counts those paths. In the same way $S^{N-1}S$ will count N-step paths, because those are $(N-1)$-step paths from i to k followed by one step from k to j. Matrix multiplication is exactly suited to counting paths on a graph—channels of communication between employees in a company.

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.

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?

$$BA \qquad AB \qquad ABD \qquad DC \qquad A(B+C).$$

2 What rows or columns or matrices do you multiply to find

(a) the second column of AB?

(b) the first row of AB?

(c) the entry in row 3, column 5 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 $(A+B)(A+B) = A^2 + \_\_\_\_ + B^2$.

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 This fact still amazes me. If you do a row operation on A and then a column operation, the result is the same as if you did the column operation first. (Try it.) Why is this true?

12 (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.

13 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.

14 Which of the following matrices are guaranteed to equal $(A - B)^2$: $A^2 - B^2$, $(B - A)^2$, $A^2 - 2AB + B^2$, $A(A - B) - B(A - B)$, $A^2 - AB - BA + B^2$?

15 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 $A = I$.

16 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 $m = n$.

17 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 18–20 use a_{ij} for the entry in row i, column j of A.

18 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) $a_{ij} = i/j$.

19 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) $a_{ij} = a_{1j}$.

20 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 21–24 involve powers of A.

21 Compute A^2, A^3, A^4 and also $A\boldsymbol{v}, A^2\boldsymbol{v}, A^3\boldsymbol{v}, A^4\boldsymbol{v}$ for

$$A = \begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{v} = \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix}.$$

22 By trial and error find real nonzero 2 by 2 matrices such that

$$A^2 = -I \qquad BC = 0 \qquad DE = -ED \text{ (not allowing } DE = 0).$$

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:

$$AB = \begin{bmatrix} 1 & 0 \\ 2 & 4 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 3 & 3 & 0 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \begin{bmatrix} 3 & 3 & 0 \end{bmatrix} + \underline{\qquad} = \underline{\qquad}.$$

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

$$EA = \begin{bmatrix} 1 & \mathbf{0} \\ -\boldsymbol{c}/a & I \end{bmatrix} \begin{bmatrix} a & \boldsymbol{b} \\ \boldsymbol{c} & D \end{bmatrix} = \begin{bmatrix} a & \boldsymbol{b} \\ \mathbf{0} & D - \boldsymbol{cb}/a \end{bmatrix}.$$

In Problem 29, what numbers go into $\boldsymbol{c}$ 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 $Ax = b$ for three special right sides b:

$$Ax_1 = \begin{bmatrix}1\\0\\0\end{bmatrix} \quad \text{and} \quad Ax_2 = \begin{bmatrix}0\\1\\0\end{bmatrix} \quad \text{and} \quad Ax_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}.$$

If the three solutions x_1, x_2, x_3 are the columns of a matrix X, what is A times X?

33 If the three solutions in Question 32 are $x_1 = (1,1,1)$ and $x_2 = (0,1,1)$ and $x_3 = (0,0,1)$, solve $Ax = b$ when $b = (3,5,8)$. Challenge problem: What is A?

34 Find all matrices $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ that satisfy $A\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}A$.

35 Suppose a "circle graph" has 4 nodes connected (in both directions) by edges around a circle. What is its adjacency matrix S from Worked Example **2.4 A**? What is S^2? Find all the 2-step paths predicted by S^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 is $\boldsymbol{mnp}$ for $AB + \boldsymbol{mpq}$ for $(AB)\,C$. The same matrix comes from A times BC with $\boldsymbol{mnq} + \boldsymbol{npq}$ 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 $\boldsymbol{(AB)\,C = A\,(BC)}$, 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 $c_1A\boldsymbol{b}_1 + \cdots + c_nA\boldsymbol{b}_n$.

$B\boldsymbol{c}$ 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.

38 (a) Suppose A has rows $\boldsymbol{a}_1^{\mathrm{T}}, \ldots, \boldsymbol{a}_m^{\mathrm{T}}$. Why does $A^{\mathrm{T}}A$ equal $\boldsymbol{a}_1\boldsymbol{a}_1^{\mathrm{T}} + \cdots + \boldsymbol{a}_m\boldsymbol{a}_m^{\mathrm{T}}$?

(b) If C is a diagonal matrix with $c_1, \ldots, c_m$ on its diagonal, find a similar sum of columns times rows for $A^{\mathrm{T}}CA$. First do an example with $m = n = 2$.

2.5 Inverse Matrices

1 If the square matrix A has an inverse, then both $A^{-1}A = I$ and $AA^{-1} = I$.

2 The *algorithm* to test invertibility is elimination: A must have n (nonzero) pivots.

3 The *algebra* test for invertibility is the determinant of A: $\det A$ must not be zero.

4 The *equation* that tests for invertibility is $A\boldsymbol{x} = \mathbf{0}$: $\boldsymbol{x} = \mathbf{0}$ **must be the only solution.**

5 If A and B (same size) are invertible then so is AB: $\boxed{(AB)^{-1} = B^{-1}A^{-1}.}$

6 $AA^{-1} = I$ is n equations for n columns of A^{-1}. Gauss-Jordan eliminates $[A\ I]$ to $[I\ A^{-1}]$.

7 The last page of the book gives 14 equivalent conditions for a square A to be invertible.

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} that "inverts" A:

Two-sided inverse $\qquad A^{-1}A = I \quad \text{and} \quad AA^{-1} = I. \qquad (1)$

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":

$$B(AC) = (BA)C \quad \text{gives} \quad BI = IC \quad \text{or} \quad B = C. \qquad (2)$$

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 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$:

Multiply $A\boldsymbol{x} = \boldsymbol{b}$ ***by*** A^{-1}. ***Then*** $\boldsymbol{x} = A^{-1}A\boldsymbol{x} = A^{-1}\boldsymbol{b}$.

Note 4 (Important) ***Suppose there is a nonzero vector x such that*** $Ax = \mathbf{0}$. *Then A cannot have an inverse.* No matrix can bring $\mathbf{0}$ back to x.

If A is invertible, then $Ax = \mathbf{0}$ can only have the zero solution $x = A^{-1}\mathbf{0} = \mathbf{0}$.

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 $A = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$ 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 $BB^{-1} = I$:

$$\textbf{Inverse of } \boldsymbol{AB} \qquad (AB)(B^{-1}A^{-1}) = AIA^{-1} = AA^{-1} = I.$$

We moved parentheses to multiply BB^{-1} first. Similarly $B^{-1}A^{-1}$ times AB equals I.

$B^{-1}A^{-1}$ 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:

Reverse order $$(ABC)^{-1} = C^{-1}B^{-1}A^{-1}. \tag{5}$$

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 subtracts
E^{-1} adds
$$E = \begin{bmatrix} 1 & 0 & 0 \\ \mathbf{-5} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ \mathbf{5} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

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. If $AC = I$ then automatically $CA = I$.

For square matrices, an inverse on one side is automatically an inverse on the other side.

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 $E^{-1}F^{-1}$!

$$FE = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ \mathbf{20} & -4 & 1 \end{bmatrix} \quad \text{is inverted by} \quad E^{-1}F^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ \mathbf{5} & 1 & 0 \\ \mathbf{0} & 4 & 1 \end{bmatrix}. \tag{6}$$

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.***

This is why the next section chooses $A = LU$, to go back from the triangular U to A. The multipliers fall into place perfectly in the lower triangular L.

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.

Calculating A^{-1} by Gauss-Jordan Elimination

I hinted that A^{-1} might not be explicitly needed. The equation $A\boldsymbol{x} = \boldsymbol{b}$ is solved by $\boldsymbol{x} = A^{-1}\boldsymbol{b}$. But it is not necessary or efficient to compute A^{-1} and multiply it times $\boldsymbol{b}$. *Elimination goes directly to* $\boldsymbol{x}$. And 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 $\boldsymbol{x}_1$) to give the first column of I (call that $\boldsymbol{e}_1$). This is our equation $A\boldsymbol{x}_1 = \boldsymbol{e}_1 = (1,0,0)$. There will be two more equations. ***Each of the columns*** $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ ***of*** A^{-1} ***is multiplied by*** A ***to produce a column of*** I***:***

3 columns of A^{-1}
$$AA^{-1} = A\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3\end{bmatrix} = \begin{bmatrix}\boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3\end{bmatrix} = I. \tag{7}$$

To invert a 3 by 3 matrix A, we have to solve three systems of equations: $A\boldsymbol{x}_1 = \boldsymbol{e}_1$ and $A\boldsymbol{x}_2 = \boldsymbol{e}_2 = (0,1,0)$ and $A\boldsymbol{x}_3 = \boldsymbol{e}_3 = (0,0,1)$. 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:

$$\begin{bmatrix}K & \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3\end{bmatrix} = \begin{bmatrix} \mathbf{2} & \mathbf{-1} & \mathbf{0} & 1 & 0 & 0 \\ \mathbf{-1} & \mathbf{2} & \mathbf{-1} & 0 & 1 & 0 \\ \mathbf{0} & \mathbf{-1} & \mathbf{2} & 0 & 0 & 1 \end{bmatrix} \quad \textbf{Start Gauss-Jordan on } K$$

$$\to \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ \mathbf{0} & \mathbf{\tfrac{3}{2}} & \mathbf{-1} & \mathbf{\tfrac{1}{2}} & \mathbf{1} & \mathbf{0} \\ 0 & -1 & 2 & 0 & 0 & 1 \end{bmatrix} \quad (\tfrac{1}{2}\; \textbf{\textit{row}}\; \mathbf{1} + \textbf{\textit{row}}\; \mathbf{2})$$

$$\to \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \tfrac{3}{2} & -1 & \tfrac{1}{2} & 1 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{\tfrac{4}{3}} & \mathbf{\tfrac{1}{3}} & \mathbf{\tfrac{2}{3}} & \mathbf{1} \end{bmatrix} \quad (\tfrac{2}{3}\; \textbf{\textit{row}}\; \mathbf{2} + \textbf{\textit{row}}\; \mathbf{3})$$

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*** $\boldsymbol{R} = \boldsymbol{I}$. Rows are added to rows above them, to produce ***zeros above the pivots*** :

$$\begin{pmatrix}\text{Zero above} \\ \text{third pivot}\end{pmatrix} \to \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ \mathbf{0} & \mathbf{\tfrac{3}{2}} & \mathbf{0} & \mathbf{\tfrac{3}{4}} & \mathbf{\tfrac{3}{2}} & \mathbf{\tfrac{3}{4}} \\ 0 & 0 & \tfrac{4}{3} & \tfrac{1}{3} & \tfrac{2}{3} & 1 \end{bmatrix} \quad (\tfrac{3}{4}\; \textbf{\textit{row}}\; \mathbf{3} + \textbf{\textit{row}}\; \mathbf{2})$$

$$\begin{pmatrix}\text{Zero above} \\ \text{second pivot}\end{pmatrix} \to \begin{bmatrix} \mathbf{2} & \mathbf{0} & \mathbf{0} & \mathbf{\tfrac{3}{2}} & \mathbf{1} & \mathbf{\tfrac{1}{2}} \\ 0 & \tfrac{3}{2} & 0 & \tfrac{3}{4} & \tfrac{3}{2} & \tfrac{3}{4} \\ 0 & 0 & \tfrac{4}{3} & \tfrac{1}{3} & \tfrac{2}{3} & 1 \end{bmatrix} \quad (\tfrac{2}{3}\; \textbf{\textit{row}}\; \mathbf{2} + \textbf{\textit{row}}\; \mathbf{1})$$

The final Gauss-Jordan step is to divide each row by its pivot. The new pivots are all 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*** $[\,I\;K^{-1}\,]$:

$$\begin{matrix}\text{(divide by 2)}\\ \text{(divide by }\frac{3}{2})\\ \text{(divide by }\frac{4}{3})\end{matrix}\quad \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{3}{4} & \frac{1}{2} & \frac{1}{4}\\ 0 & 1 & 0 & \frac{1}{2} & 1 & \frac{1}{2}\\ 0 & 0 & 1 & \frac{1}{4} & \frac{1}{2} & \frac{3}{4}\end{array}\right] = \left[\,I\;\; \boldsymbol{x}_1\;\; \boldsymbol{x}_2\;\; \boldsymbol{x}_3\,\right] = \left[\,I\;\; \boldsymbol{K^{-1}}\,\right].$$

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 K^{-1}: an important example.

1. K is ***symmetric*** across its main diagonal. Then K^{-1} is also symmetric.

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 $2\left(\frac{3}{2}\right)\left(\frac{4}{3}\right)=4$. This number 4 is the ***determinant*** of K.

K^{-1} involves division by the determinant of K
$$K^{-1}=\frac{1}{4}\begin{bmatrix}3&2&1\\2&4&2\\1&2&3\end{bmatrix}. \tag{8}$$

This is why an invertible matrix cannot have a zero determinant: we need to divide.

Example 4 Find A^{-1} by Gauss-Jordan elimination starting from $A=\left[\begin{smallmatrix}2&3\\4&7\end{smallmatrix}\right]$.

$$\begin{aligned}[\,A\;\;I\,]=\begin{bmatrix}2&3&1&0\\4&7&0&1\end{bmatrix}&\rightarrow\begin{bmatrix}2&3&1&0\\0&1&-2&1\end{bmatrix} &&(\text{this is } [\,U\;\;L^{-1}\,])\\ \rightarrow\begin{bmatrix}2&0&7&-3\\0&1&-2&1\end{bmatrix}&\rightarrow\begin{bmatrix}1&0&\frac{7}{2}&-\frac{3}{2}\\0&1&-2&1\end{bmatrix} &&(\text{this is } [\,I\;\;\boldsymbol{A^{-1}}\,]).\end{aligned}$$

Example 5 ***If A is invertible and upper triangular, so is A^{-1}.*** Start with $\boldsymbol{AA^{-1}=I}$.

1 A times *column j of A^{-1}* equals *column j of I*, ending with $n-j$ zeros.

2 Back substitution keeps those $n-j$ zeros at the end of column j of A^{-1}.

3 Put those columns $[*\ldots*0\ldots0]^{\mathrm{T}}$ into A^{-1} and that matrix is upper triangular!

$$A^{-1} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$ Columns $j = 1$ and 2 end with $3 - j = 2$ and 1 zeros.

The code for $X = \textbf{inv}(A)$ can use **rref**, the reduced row echelon form from Chapter 3:

$I =$ eye (n);	% Define the n by n identity matrix
$R =$ rref $([A \quad I])$;	% Eliminate on the augmented matrix $[A \quad I]$
$X = R(:, n+1 : n+n)$	% Pick $X = A^{-1}$ 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 solve n equations for its n columns. But all those equations involve the same matrix A on the left side (where most of the work is done). The total cost for A^{-1} is n^3 multiplications and subtractions. To solve a single $A\boldsymbol{x} = \boldsymbol{b}$ that cost (see the next section) is $n^3/3$.

To solve $A\boldsymbol{x} = \boldsymbol{b}$ without A^{-1}, we deal with *one* column $\boldsymbol{b}$ to find one column $\boldsymbol{x}$.

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 $A\boldsymbol{x}_i = \boldsymbol{e}_i$. The columns $\boldsymbol{x}_i$ 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 C $$CA = (D^{-1} \cdots E \cdots P \cdots E)A = I. \quad (9)$$

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 of A***. 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*** $AC = I$.

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. C is A^{-1}.

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:

$$\text{If } AC = I \text{ then } CA = I \text{ and } C = A^{-1} \tag{10}$$

Example 6 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} from E_{32}, E_{31}, E_{21}. Notice how L^{-1} contains the strange entry 11, from 3 times 5 minus 4.

Gauss-Jordan on triangular L

$$\begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & 1 & 0 & 0 \\ \mathbf{3} & \mathbf{1} & \mathbf{0} & 0 & 1 & 0 \\ \mathbf{4} & \mathbf{5} & \mathbf{1} & 0 & 0 & 1 \end{bmatrix} = [\, L \quad I \,]$$

$$\begin{array}{c} \\ \to \\ \to \end{array} \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -3 & 1 & 0 \\ 0 & 5 & 1 & -4 & 0 & 1 \end{bmatrix} \quad \begin{array}{l} \text{(3 times row 1 from row 2)} \\ \text{(4 times row 1 from row 3)} \\ \text{(then 5 times row 2 from row 3)} \end{array}$$

The inverse is still triangular

$$\to \begin{bmatrix} 1 & 0 & 0 & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ 0 & 1 & 0 & \mathbf{-3} & \mathbf{1} & \mathbf{0} \\ 0 & 0 & 1 & \mathbf{11} & \mathbf{-5} & \mathbf{1} \end{bmatrix} = [\, I \quad \boldsymbol{L^{-1}} \,].$$

Recognizing an Invertible Matrix

Normally, it takes work to decide if a matrix is invertible. The usual way is to find a full set of nonzero pivots in elimination. (Then the nonzero determinant comes from multiplying those pivots.) But for some matrices you can see quickly that they are invertible because every number a_{ii} on their main diagonal dominates the off-diagonal part of that row i.

Diagonally dominant matrices are invertible. Each a_{ii} on the diagonal is larger than the total sum along the rest of row i. On every row,

$$|\boldsymbol{a_{ii}}| > \sum_{j \neq i} |\boldsymbol{a_{ij}}| \quad \text{means that} \quad |a_{ii}| > |a_{i1}| + \cdots (\text{skip } |a_{ii}|) \cdots + |a_{in}|. \tag{11}$$

Examples. A is diagonally dominant $(3 > 2)$. B is not (but still invertible). C is singular.

$$A = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 3 \end{bmatrix}$$

Reasoning. Take any nonzero vector $\boldsymbol{x}$. *Suppose its largest component is* $|x_i|$. Then $A\boldsymbol{x} = \mathbf{0}$ is impossible, because row i of $A\boldsymbol{x} = \mathbf{0}$ would need

$$a_{i1}x_1 + \cdots + a_{ii}x_i + \cdots + a_{in}x_n = 0.$$

Those can't add to zero when A is diagonally dominant! The size of $a_{ii}x_i$ (that one particular term) is greater than all the other terms combined:

$$\textbf{All } |x_j| \le |x_i| \qquad \sum_{j \neq i} |a_{ij}x_j| \le \sum_{j \neq i} |a_{ij}|\,|x_i| < |a_{ii}|\,|x_i| \quad \textbf{because } a_{ii} \textbf{ dominates}$$

This shows that $A\boldsymbol{x} = \mathbf{0}$ is only possible when $\boldsymbol{x} = \mathbf{0}$. *So A is invertible*. The example B was also invertible but not quite diagonally dominant: 2 is not larger than $1+1$.

■ 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. *Important*. 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 $\begin{bmatrix} I & A^{-1} \end{bmatrix}$.
6. Diagonally dominant matrices are invertible. Each $|a_{ii}|$ dominates its row.

■ WORKED EXAMPLES ■

2.5 A The inverse of a triangular **difference matrix** A is a triangular **sum matrix** S:

$$\begin{bmatrix} A & I \end{bmatrix} = \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$\to \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right] = \begin{bmatrix} I & A^{-1} \end{bmatrix} = \begin{bmatrix} I & \textit{sum matrix} \end{bmatrix}.$$

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 $A\boldsymbol{x} = \mathbf{0}$) 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 has two equal rows (and the second column minus the first column is zero). In other words $E\boldsymbol{x} = \mathbf{0}$ has the solution $\boldsymbol{x} = (-1, 1, 0)$.

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 $1, 4, 6, 4, 1$.

$$\textbf{Triangular Pascal matrix} \quad L = \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 \\ \mathbf{1} & \mathbf{1} & 0 & 0 \\ \mathbf{1} & \mathbf{2} & \mathbf{1} & 0 \\ \mathbf{1} & \mathbf{3} & \mathbf{3} & \mathbf{1} \end{bmatrix} = \text{abs(pascal (4,1))}$$

Solution Gauss-Jordan starts with $[\,L \;\; I\,]$ and produces zeros by subtracting row 1:

$$[\,\boldsymbol{L} \;\; \boldsymbol{I}\,] = \left[\begin{array}{cccc|cccc} \mathbf{1} & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \mathbf{1} & \mathbf{1} & 0 & 0 & 0 & 1 & 0 & 0 \\ \mathbf{1} & \mathbf{2} & \mathbf{1} & 0 & 0 & 0 & 1 & 0 \\ \mathbf{1} & \mathbf{3} & \mathbf{3} & \mathbf{1} & 0 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \mathbf{0} & 1 & 0 & 0 & \mathbf{-1} & 1 & 0 & 0 \\ \mathbf{0} & 2 & 1 & 0 & \mathbf{-1} & 0 & 1 & 0 \\ \mathbf{0} & 3 & 3 & 1 & \mathbf{-1} & 0 & 0 & 1 \end{array}\right].$$

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:

$$\to \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & \mathbf{0} & 1 & 0 & \mathbf{1} & \mathbf{-2} & 1 & 0 \\ 0 & \mathbf{0} & 3 & 1 & \mathbf{2} & \mathbf{-3} & 0 & 1 \end{array}\right] \to \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & 1 & 0 & 0 & \mathbf{-1} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ 0 & 0 & 1 & 0 & \mathbf{1} & \mathbf{-2} & \mathbf{1} & \mathbf{0} \\ 0 & 0 & 0 & 1 & \mathbf{-1} & \mathbf{3} & \mathbf{-3} & \mathbf{1} \end{array}\right] = [\,\boldsymbol{I} \;\; \boldsymbol{L}^{-1}\,].$$

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 $A\boldsymbol{x} = (0, 0, 1)$ cannot have a solution. Add eqn 1 + eqn 2.

(b) Which right sides (b_1, b_2, b_3) might allow a solution to $A\boldsymbol{x} = \boldsymbol{b}$?

(c) In elimination, what happens to equation 3?

8 If A has column 1 + column 2 = column 3, show that A is not invertible:

(a) Find a nonzero solution $\boldsymbol{x}$ to $A\boldsymbol{x} = \mathbf{0}$. 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? 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 $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ times $\left[\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right]$. 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 $5 * \mathsf{eye}(4) - \mathsf{ones}(4,4)$:

$$\begin{bmatrix} 4 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 \\ -1 & -1 & 4 & -1 \\ -1 & -1 & -1 & 4 \end{bmatrix}^{-1} = \begin{bmatrix} a & b & b & b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \end{bmatrix}.$$

What are a and b in the inverse of $6 * \mathsf{eye}(5) - \mathsf{ones}(5,5)$?

20 Show that $\mathsf{A} = 4 * \mathsf{eye}(4) - \mathsf{ones}(4,4)$ is *not* invertible: Multiply $\mathsf{A} * \mathsf{ones}(4,1)$.

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 $[\,A \;\; I\,]$ to $[\,I \;\; A^{-1}\,]$:

$$\begin{bmatrix} A & I \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{bmatrix}.$$

24 Use Gauss-Jordan elimination on $[\,U \;\; I\,]$ 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 $[\,A \;\; I\,]$ and $[\,B \;\; I\,]$:

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

26 What three matrices E_{21} and E_{12} and D^{-1} reduce $A = \begin{bmatrix} 1 & 2 \\ 2 & 6 \end{bmatrix}$ 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 $[\,A \;\; I\,]$:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

28 Exchange rows and continue with Gauss-Jordan to find A^{-1}:

$$\begin{bmatrix} A & I \end{bmatrix} = \begin{bmatrix} 0 & 2 & 1 & 0 \\ 2 & 2 & 0 & 1 \end{bmatrix}.$$

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 (Recommended) Prove that A is invertible if $a \neq 0$ and $a \neq b$ (find the pivots or A^{-1}). Then find three numbers c so that C is not invertible:

$$A = \begin{bmatrix} a & b & b \\ a & a & b \\ a & a & a \end{bmatrix} \qquad C = \begin{bmatrix} 2 & c & c \\ c & c & c \\ 8 & 7 & c \end{bmatrix}.$$

31 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.

$$\text{Invert } A = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and solve } A\boldsymbol{x} = (1,1,1,1).$$

32 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 $(P - Q)\,\boldsymbol{x} = \mathbf{0}$.

33 Find and check the inverses (assuming they exist) of these block matrices:

$$\begin{bmatrix} I & 0 \\ C & I \end{bmatrix} \quad \begin{bmatrix} A & 0 \\ C & D \end{bmatrix} \quad \begin{bmatrix} 0 & I \\ I & D \end{bmatrix}.$$

34 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?

35 In the Worked Example **2.5 C**, the triangular Pascal matrix L has $L^{-1} = DLD$, where the diagonal matrix D has alternating entries $1, -1, 1, -1$. Then $LDLD = I$, so what is the inverse of $LD = \texttt{pascal}\,(4, 1)$?

36 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?

37 (a) Use inv(P) to invert MATLAB's 4 by 4 symmetric matrix $P = \texttt{pascal}\,(4)$.

(b) Create Pascal's lower triangular $L = \texttt{abs}\,(\texttt{pascal}\,(4, 1))$ and test $P = LL^{\mathrm{T}}$.

38 If $A = \texttt{ones}\,(4)$ and $\boldsymbol{b} = \texttt{rand}\,(4, 1)$, how does MATLAB tell you that $A\boldsymbol{x} = \boldsymbol{b}$ has no solution? For the special $\boldsymbol{b} = \texttt{ones}\,(4, 1)$, which solution to $A\boldsymbol{x} = \boldsymbol{b}$ is found by $A\backslash\boldsymbol{b}$?

Challenge Problems

39 (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.

40 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 $L = E_1 E_2 E_3$ to show that all these nonzeros are copied into L.

$E_1 E_2 E_3$ is in the *opposite* order from elimination (because E_3 is acting first). But $E_1 E_2 E_3 = L$ is in the *correct* order to invert elimination and recover A.

41 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:

$$T = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \qquad T^{-1} = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

One approach is Gauss-Jordan elimination on $[\,T \quad I\,]$. 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 $T = LU$ and $T^{-1} = U^{-1}L^{-1}$:

$$T = \underset{\textbf{difference}}{\begin{bmatrix} 1 & & \\ -1 & 1 & \\ 0 & -1 & 1 \end{bmatrix}} \underset{\textbf{difference}}{\begin{bmatrix} 1 & -1 & 0 \\ & 1 & -1 \\ & & 1 \end{bmatrix}} \qquad T^{-1} = \underset{\textbf{sum}}{\begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{bmatrix}} \underset{\textbf{sum}}{\begin{bmatrix} 1 & & \\ 1 & 1 & \\ 1 & 1 & 1 \end{bmatrix}}$$

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^*?

42 Here are two more difference matrices, both important. ***But are they invertible***?

$$\textbf{Cyclic } C = \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix} \qquad \textbf{Free ends } F = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}.$$

43 *Elimination for a block matrix*: When you multiply the first block row $[A \;\; B]$ by CA^{-1} and subtract from the second row $[C \;\; D]$, 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*** $[\,A \quad B\,;\;\; C \quad D\,]$.

44 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: $A = LU$

1 Each elimination step E_{ij} is inverted by L_{ij}. Off the main diagonal change $-\ell_{ij}$ to $+\ell_{ij}$.

2 The whole forward elimination process (with no row exchanges) is inverted by $\boldsymbol{L}$:
$$\boldsymbol{L} = (L_{21}L_{31}\ldots L_{n1})(L_{32}\ldots L_{n2})(L_{43}\ldots L_{n3})\ldots(L_{n\,n-1}).$$

3 That product matrix $\boldsymbol{L}$ is still lower triangular. **Every multiplier $\boldsymbol{\ell_{ij}}$ is in row $\boldsymbol{i}$, column $\boldsymbol{j}$.**

4 The original A is recovered from U by $\boldsymbol{A = LU} =$ (lower triangular)(upper triangular).

5 Elimination on $A\boldsymbol{x} = \boldsymbol{b}$ reaches $U\boldsymbol{x} = \boldsymbol{c}$. Then back-substitution solves $U\boldsymbol{x} = \boldsymbol{c}$.

6 Solving a triangular system takes $n^2/2$ multiply-subtracts. Elimination to find U takes $n^3/3$.

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:

$$(E_{32}E_{31}E_{21})A = U \quad \textbf{becomes} \quad \boxed{A = (E_{21}^{-1}E_{31}^{-1}E_{32}^{-1})\,U} \quad \textbf{which is} \quad \boxed{A = LU.}$$

The inverses go in opposite order, as they must. That product of three inverses is L. ***We have reached*** $\boldsymbol{A = LU}$. 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 $E = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}$ and $L = E^{-1} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$.

Second point: Equation (2) 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 (2).

Since each E^{-1} has 1's down its diagonal, the final good point is that L does too.

$\boldsymbol{A = LU}$	***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*** $\boldsymbol{\ell_{ij}}$ ***are below the diagonal of*** $\boldsymbol{L}$.

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 in A to 1 in B. The pivots all become 1. The multipliers are all 1. That pattern continues when B is 4 by 4:

$$\textbf{Special pattern} \qquad B = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ 0 & 1 & 1 & \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ & 1 & 1 & 0 \\ & & 1 & 1 \\ & & & 1 \end{bmatrix}.$$

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!):

$$\text{Row 3 of } U = (\text{Row 3 of } A) - \ell_{31}(\text{Row 1 of } U) - \ell_{32}(\text{Row 2 of } U). \tag{1}$$

Rewrite this equation to see that the row $[\,\ell_{31} \quad \ell_{32} \quad 1\,]$ is multiplying the matrix U:

$$(\text{Row 3 of } A) = \ell_{31}(\text{Row 1 of } U) + \ell_{32}(\text{Row 2 of } U) + 1(\text{Row 3 of } U). \tag{2}$$

This is exactly row 3 ***of*** $A = LU$. That row of L holds ℓ_{31}, ℓ_{32}, 1. All rows look like this, whatever the size of A. With no row exchanges, we have $A = LU$.

Better balance from LDU $A = LU$ 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 triangular 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 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:

$$\begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 8 \\ 0 & 5 \end{bmatrix} \quad \text{splits further into} \quad \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & \\ & 5 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}. \tag{3}$$

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. I go forward to 2.7. 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***:

$$\textbf{Forward and backward} \qquad \textit{Solve} \quad L\boldsymbol{c} = \boldsymbol{b} \quad \textit{and then solve} \quad U\boldsymbol{x} = \boldsymbol{c}\,. \tag{4}$$

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 $U\boldsymbol{x} = \boldsymbol{c}$:

$$A\boldsymbol{x} = \boldsymbol{b} \qquad \begin{aligned} u + 2v &= 5 \\ 4u + 9v &= 21 \end{aligned} \qquad \text{becomes} \qquad \begin{aligned} u + 2v &= 5 \\ v &= 1 \end{aligned} \qquad U\boldsymbol{x} = \boldsymbol{c}$$

The multiplier was 4, which is saved in L. The right side used that 4 to change 21 to 1:

$Lc = b$ The lower triangular system $\begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} c \end{bmatrix} = \begin{bmatrix} 5 \\ 21 \end{bmatrix}$ gave $c = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$.

$Ux = c$ The upper triangular system $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ gives $x = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$.

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 $n = 100,000$? (*Is A dense or sparse*?) 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 produces zeros below the first pivot in column 1. 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 $\frac{1}{3}n\left(n + \frac{1}{2}\right)(n+1)$ 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:

$$[(n-1) + (n-2) + \cdots + 1] \; + \; [1 + 2 + \cdots + (n-1) + n] = n^2. \tag{5}$$

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 above its main diagonal. The zero entries outside the band stay zero in elimination (they are zero 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 matrix	$\boldsymbol{A}$ **to** $\boldsymbol{U}$ $\quad \frac{1}{3}\boldsymbol{n}^3$ reduces to $\boldsymbol{nw}^2$	**Solve** $\quad \boldsymbol{n}^2$ reduces to $\mathbf{2}\,\boldsymbol{nw}$

A tridiagonal matrix (bandwidth $w = 1$) allows very fast computation. Don't store zeros!

The book's website has Teaching Codes to factor A into LU and to solve $Ax = b$. Professional codes will look down each column for the *largest available pivot*, to exchange rows and reduce roundoff error.

MATLAB's backslash command $\boldsymbol{x} = A\backslash \boldsymbol{b}$ combines **Factor** and **Solve** to reach $\boldsymbol{x}$.

How long does it take to solve $\boldsymbol{Ax = b}$***?*** For a random matrix of order $n = 1000$, a typical time on a PC is 1 second. 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.

■ 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 inverted L in the worked example **2.5 C**. Here we factor Pascal.

The symmetric Pascal matrix P is a product of triangular Pascal matrices L and 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* $P = LU$.

$$\text{pascal}(4) = \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{2} & \mathbf{3} & 4 \\ \mathbf{1} & \mathbf{3} & 6 & 10 \\ \mathbf{1} & 4 & 10 & 20 \end{bmatrix} = \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 \\ \mathbf{1} & \mathbf{1} & 0 & 0 \\ \mathbf{1} & \mathbf{2} & \mathbf{1} & 0 \\ \mathbf{1} & \mathbf{3} & \mathbf{3} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ 0 & \mathbf{1} & \mathbf{2} & \mathbf{3} \\ 0 & 0 & \mathbf{1} & \mathbf{3} \\ 0 & 0 & 0 & \mathbf{1} \end{bmatrix} = LU.$$

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:

$$P = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 2 & 5 & 9 \\ 0 & 3 & 9 & 19 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 3 & 10 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix} = U.$$

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* $P\boldsymbol{x} = \boldsymbol{b} = (1,0,0,0)$. This right side = column of I means that $\boldsymbol{x}$ 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:

Two triangular systems $\quad L\boldsymbol{c} = \boldsymbol{b}$ (forward) $\quad U\boldsymbol{x} = \boldsymbol{c}$ (back).

Solution The lower triangular system $L\boldsymbol{c} = \boldsymbol{b}$ is solved *top to bottom*:

$$\begin{aligned} c_1 \qquad\qquad\qquad &= 1 \\ c_1 + \ c_2 \qquad\qquad &= 0 \\ c_1 + 2c_2 + \ c_3 \qquad &= 0 \\ c_1 + 3c_2 + 3c_3 + c_4 &= 0 \end{aligned} \quad \text{gives} \quad \begin{aligned} c_1 &= +1 \\ c_2 &= -1 \\ c_3 &= +1 \\ c_4 &= -1 \end{aligned}$$

Forward elimination is multiplication by L^{-1}. It produces the upper triangular system $U\boldsymbol{x} = \boldsymbol{c}$. The solution $\boldsymbol{x}$ comes as always by back substitution, *bottom to top*:

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 1 \\ x_2 + 2x_3 + 3x_4 &= -1 \\ x_3 + 3x_4 &= 1 \\ x_4 &= -1 \end{aligned} \qquad \text{gives} \qquad \begin{aligned} x_1 &= \mathbf{+4} \\ x_2 &= \mathbf{-6} \\ x_3 &= \mathbf{+4} \\ x_4 &= \mathbf{-1} \end{aligned}$$

I see a pattern in that $\boldsymbol{x}$, 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 $\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\boldsymbol{x} = \boldsymbol{b}$ to a triangular $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\boldsymbol{x} = \boldsymbol{c}$:

$$\begin{aligned} x + y &= \mathbf{5} \\ x + 2y &= \mathbf{7} \end{aligned} \quad \longrightarrow \quad \begin{aligned} x + y &= \mathbf{5} \\ y &= \mathbf{2} \end{aligned} \qquad \begin{bmatrix} 1 & 1 & \mathbf{5} \\ 1 & 2 & \mathbf{7} \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & \mathbf{5} \\ 0 & 1 & \mathbf{2} \end{bmatrix}$$

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 $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\boldsymbol{x}_1 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$ to get _____ $=$ _____. In letters, L multiplies $U\boldsymbol{x} = \boldsymbol{c}$ to give _____.

2 Write down the 2 by 2 triangular systems $L\boldsymbol{c} = \boldsymbol{b}$ and $U\boldsymbol{x} = \boldsymbol{c}$ from Problem 1. Check that $\boldsymbol{c} = (5, 2)$ solves the first one. Find $\boldsymbol{x}$ 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 $\begin{bmatrix} 1 & 3 & 6 & 11 \end{bmatrix}$ in the last row of A and $\boldsymbol{b}$ from the final $\begin{bmatrix} 1 & 1 & 1 & 5 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 & 2 & 2 \end{bmatrix}$ and $\begin{bmatrix} 0 & 0 & 1 & 2 \end{bmatrix}$ 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 **This is the problem that shows how the inverses E_{ij}^{-1} multiply to give L.** You see this best when 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.
The multipliers a, b, c are mixed up in E but perfect in L.

9 *When zero appears in a pivot position, $A = LU$ is not possible*! (We are requiring nonzero pivots in U.) Show directly why these equations are both impossible:

$$\begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \ell & 1 \end{bmatrix}\begin{bmatrix} d & e \\ 0 & f \end{bmatrix} \qquad \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ \ell & 1 & \\ m & n & 1 \end{bmatrix}\begin{bmatrix} d & e & g \\ & f & h \\ & & i \end{bmatrix}.$$

These matrices need a row exchange. That uses a "*permutation matrix*" 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 $A = LDU$?

$$\textbf{Already triangular} \qquad A = \begin{bmatrix} 2 & 4 & 8 \\ 0 & 3 & 9 \\ 0 & 0 & 7 \end{bmatrix}.$$

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:

$$\textbf{Symmetric} \qquad A = \begin{bmatrix} 2 & 4 \\ 4 & 11 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 4 & 0 \\ 4 & 12 & 4 \\ 0 & 4 & 0 \end{bmatrix}.$$

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**:

$$\textbf{Find } L \textbf{ and } U \textbf{ for} \qquad A = \begin{bmatrix} a & r & r & r \\ a & b & s & s \\ a & b & c & t \\ a & b & c & d \end{bmatrix}.$$

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 $A\boldsymbol{x} = \boldsymbol{b}$.

15 Solve the triangular system $L\boldsymbol{c} = \boldsymbol{b}$ to find $\boldsymbol{c}$. Then solve $U\boldsymbol{x} = \boldsymbol{c}$ to find $\boldsymbol{x}$:

$$L = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} 2 & 4 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 2 \\ 11 \end{bmatrix}.$$

For safety multiply LU and solve $A\boldsymbol{x} = \boldsymbol{b}$ as usual. Circle $\boldsymbol{c}$ when you see it.

16 Solve $L\boldsymbol{c} = \boldsymbol{b}$ to find $\boldsymbol{c}$. Then solve $U\boldsymbol{x} = \boldsymbol{c}$ to find $\boldsymbol{x}$. **What was** A?

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}.$$

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 A_k 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.

$$\textbf{−1, 2, −1 matrix} \quad K = \begin{bmatrix} 2 & -1 & & & & \\ -1 & \cdot & \cdot & & & \\ & \cdot & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & -1 \\ & & & & -1 & 2 \end{bmatrix} = \text{toeplitz}\,([2 \quad -1 \quad 0 \quad 0 \quad 0 \quad 0])$$

26 If you print K^{-1}, it doesn't look so good (6 by 6). But if you print $7K^{-1}$, 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 $4K^{-1}$ for $n = 3$:

$$\begin{array}{l} \textbf{3 by 3 case} \\ \textbf{The determinant} \\ \textbf{of this } \boldsymbol{K} \textbf{ is 4} \end{array} \quad (K)(4K^{-1}) = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} \mathbf{3} & \mathbf{2} & \mathbf{1} \\ \mathbf{2} & \mathbf{4} & \mathbf{2} \\ \mathbf{1} & \mathbf{2} & \mathbf{3} \end{bmatrix} = \begin{bmatrix} 4 & & \\ & 4 & \\ & & 4 \end{bmatrix}.$$

2.7 Transposes and Permutations

1 The transposes of $A\boldsymbol{x}$ and AB and A^{-1} are $\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}$ and $B^{\mathrm{T}}A^{\mathrm{T}}$ and $(A^{\mathrm{T}})^{-1}$.

2 The dot product (inner product) is $\boldsymbol{x}\cdot\boldsymbol{y}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$. This is $(1\times n)(n\times 1)=(1\times 1)$.
The outer product is $\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}$ = column times row = $(n\times 1)(1\times n)=n\times n$ matrix.

3 The idea behind A^{T} is that $A\boldsymbol{x}\cdot\boldsymbol{y}$ equals $\boldsymbol{x}\cdot A^{\mathrm{T}}\boldsymbol{y}$ because $(A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{y})$.

4 A **symmetric matrix** has $\boldsymbol{S^{\mathrm{T}}=S}$ (and the product $A^{\mathrm{T}}A$ is always symmetric).

5 An **orthogonal matrix** has $\boldsymbol{Q^{\mathrm{T}}=Q^{-1}}$. The columns of Q are orthogonal unit vectors.

6 A **permutation matrix** $\boldsymbol{P}$ has the same rows as I (in any order). There are $\boldsymbol{n}\,!$ different orders.

7 Then $P\boldsymbol{x}$ puts the components $x_1, x_2, \ldots, x_n$ in that new order. And P^{T} equals P^{-1}.

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:

Transpose $\qquad$ If $\quad A=\begin{bmatrix}1&2&3\\0&0&4\end{bmatrix}$ $\quad$ 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:

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}:

Sum	The transpose of $A+B$ is $A^{\mathrm{T}}+B^{\mathrm{T}}$.	(1)
Product	The transpose of AB is $(AB)^{\mathrm{T}}=B^{\mathrm{T}}A^{\mathrm{T}}$.	(2)
Inverse	The transpose of A^{-1} is $(A^{-1})^{\mathrm{T}}=(A^{\mathrm{T}})^{-1}$.	(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 $(A\boldsymbol{x})^{\mathrm{T}}=\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}$ when B is just a vector:

$A\boldsymbol{x}$ combines the columns of A while $\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}$ combines the rows of A^{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 appear correctly in the rows of $B^{\mathrm{T}}A^{\mathrm{T}}$:

$$\text{Transposing } AB=\begin{bmatrix} A\boldsymbol{x}_1 & A\boldsymbol{x}_2 & \cdots \end{bmatrix} \text{ gives } \begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}}A^{\mathrm{T}} \\ \boldsymbol{x}_2^{\mathrm{T}}A^{\mathrm{T}} \\ \vdots \end{bmatrix} \text{ which is } B^{\mathrm{T}}A^{\mathrm{T}}. \tag{4}$$

The right answer $B^{\mathrm{T}}A^{\mathrm{T}}$ comes out a row at a time. Here are numbers in $(AB)^{\mathrm{T}}=B^{\mathrm{T}}A^{\mathrm{T}}$:

$$AB=\begin{bmatrix}1&0\\1&1\end{bmatrix}\begin{bmatrix}5&0\\4&1\end{bmatrix}=\begin{bmatrix}5&0\\9&1\end{bmatrix} \quad\text{and}\quad B^{\mathrm{T}}A^{\mathrm{T}}=\begin{bmatrix}5&4\\0&1\end{bmatrix}\begin{bmatrix}1&1\\0&1\end{bmatrix}=\begin{bmatrix}5&9\\0&1\end{bmatrix}.$$

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 by transposing 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}. Their product is I:

$$\textbf{Transpose of inverse} \qquad A^{-1}A=I \ \text{ is transposed to } \ A^{\mathrm{T}}(A^{-1})^{\mathrm{T}}=I. \tag{5}$$

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). Maybe the universe is governed by linear algebra. Here are three more 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 emphasize. 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 $\boldsymbol{y}$*:

$$(A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{y}) \quad \textbf{Inner product of } A\boldsymbol{x} \textbf{ with } \boldsymbol{y} = \textbf{Inner product of } \boldsymbol{x} \textbf{ with } A^{\mathrm{T}}\boldsymbol{y}$$

$$\text{Start with } A = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

On one side we have $A\boldsymbol{x}$ multiplying $\boldsymbol{y}$: $(x_2 - x_1)\,y_1 + (x_3 - x_2)\,y_2$
That is the same as $x_1\,(-y_1) + x_2\,(y_1 - y_2) + x_3\,(y_2)$. Now $\boldsymbol{x}$ is multiplying $A^{\mathrm{T}}\boldsymbol{y}$.

$$A^{\mathrm{T}}\boldsymbol{y} \text{ must be } \begin{bmatrix} -y_1 \\ y_1 - y_2 \\ y_2 \end{bmatrix} \text{ which produces } \quad A^{\mathrm{T}} = \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \text{ as expected.}$$

Symmetric Matrices

For a *symmetric matrix*, transposing A to A^{T} produces no change. Then A^{T} equals 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. We give symmetric matrices the special letter S.

DEFINITION A ***symmetric matrix*** has $S^{\mathrm{T}} = S$. This means that $s_{ji} = s_{ij}$.

Symmetric matrices $\quad S = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} = S^{\mathrm{T}}$ and $D = \begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix} = D^{\mathrm{T}}$.

The inverse of a symmetric matrix is also symmetric. The transpose of S^{-1} is $(S^{-1})^{\mathrm{T}} = (S^{\mathrm{T}})^{-1} = S^{-1}$. That says S^{-1} is symmetric (when S is invertible):

Symmetric inverses $\quad S^{-1} = \begin{bmatrix} 5 & -2 \\ -2 & 1 \end{bmatrix}$ and $D^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 0.1 \end{bmatrix}$.

Now we produce a symmetric matrix S by ***multiplying any matrix A by*** $\boldsymbol{A^{\mathrm{T}}}$.

Symmetric Products $A^{\mathrm{T}}A$ and AA^{T} and LDL^{T}

Choose any matrix A, probably rectangular. Multiply A^{T} times A. Then the product $S = A^{\mathrm{T}}A$ is automatically a square symmetric matrix:

$$\textit{The transpose of} \quad A^{\mathrm{T}}A \quad \textit{is} \quad A^{\mathrm{T}}(A^{\mathrm{T}})^{\mathrm{T}} \quad \textit{which is} \quad A^{\mathrm{T}}A \quad \textit{again.} \tag{6}$$

That is a quick proof of symmetry for $A^{\mathrm{T}}A$. We could look at the (i,j) entry of $A^{\mathrm{T}}A$. It is the dot product of row i of A^{T} (column i of A) with column j of A. The (j,i) entry is the same dot product, column j with column i. So $A^{\mathrm{T}}A$ is symmetric.

The matrix AA^{T} is also symmetric. (The shapes of A and A^{T} allow multiplication.) But AA^{T} is a different matrix from $A^{\mathrm{T}}A$. In our experience, most scientific problems that start with a rectangular matrix A end up with $A^{\mathrm{T}}A$ or AA^{T} or both. As in least squares.

Example 2 Multiply $A = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$ and $A^{\mathrm{T}} = \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}$ in both orders.

$$AA^{\mathrm{T}} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \text{ and } A^{\mathrm{T}}A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \text{ are both symmetric matrices.}$$

The product $A^{\mathrm{T}}A$ is n by n. In the opposite order, AA^{T} is m by m. Both are symmetric, with positive diagonal (*why*?). But even if $m = n$, it is very likely that $A^{\mathrm{T}}A \neq AA^{\mathrm{T}}$. Equality can happen, but it is abnormal.

Symmetric matrices in elimination $\quad S^{\mathrm{T}} = S$ 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*** $S = LDU$. Remember how the diagonal matrix D of pivots can be divided out, to leave 1's on the diagonal of both L and U:

$$\begin{bmatrix} 1 & 2 \\ 2 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \quad LU \text{ misses the symmetry of } S$$

$$\begin{bmatrix} 1 & 2 \\ 2 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \quad \begin{array}{l} LDL^{\mathrm{T}} \text{ captures the symmetry} \\ \textit{Now } U \textit{ is the transpose of } L. \end{array}$$

When S is symmetric, the usual form $A = LDU$ becomes $\boldsymbol{S = LDL^{\mathrm{T}}}$. 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 $S = S^{\mathrm{T}}$ is factored into LDU with no row exchanges, then U is exactly L^{T}.

The symmetric factorization of a symmetric matrix is $S = LDL^{\mathrm{T}}$.

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 as P or maybe different. Any product P_1P_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 3 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.

Permutations (row exchanges before elimination) 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. 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*** $PA = LU$.

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*** $A = L_1P_1U_1$.

$PA = LU$ is constantly used in all computing. ***We will concentrate on this form***.

The factorization $A = L_1P_1U_1$ 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.

This matrix A starts with $a_{11} = 0$. Exchange rows 1 and 2 to bring the first pivot into 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:

$$\boldsymbol{P} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \boldsymbol{PA} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{bmatrix} = \boldsymbol{LU}. \tag{7}$$

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 $[\,L, U, P\,] = \textbf{lu}\,(A)$

$$\begin{aligned}
&A\,([r\ k], :) = A\,([k\ r], :);\\
&L\,([r\ k], 1 : k-1) = L\,([k\ r], 1 : k-1);\\
&P\,([r\ k], :) = P\,([k\ r], :);\\
&\text{sign} \ = -\text{sign}
\end{aligned}$$

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 $\boldsymbol{P}$** and it does not depend on the order of the row exchanges.

For PA we get back to the familiar LU. In reality, a code like **lu** (A) often does not use the first available pivot. Mathematically we can accept a small pivot— anything but zero. **All good codes look down the column for the largest pivot**.

Section 11.1 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 $A\boldsymbol{x} = \boldsymbol{b}$ for any invertible A. The program splu on the website 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 S is symmetric ($S^{\mathrm{T}} = S$), its LDU factorization is symmetric: $S = LDL^{\mathrm{T}}$.

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 S destroys its symmetry:

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \qquad S = \begin{bmatrix} 1 & 4 & 5 \\ 4 & \mathbf{2} & 6 \\ 5 & 6 & \mathbf{3} \end{bmatrix} \qquad PS = \begin{bmatrix} 4 & \mathbf{2} & 6 \\ 5 & 6 & \mathbf{3} \\ 1 & 4 & 5 \end{bmatrix}$$

What permutation Q applied to the *columns* of PS will recover symmetry in PSQ? 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** PSP^{T}.

Solution To recover symmetry and put "2" back on the diagonal, column 2 of PS must move to column 1. Column 3 of PS (containing "3") must move to column 2. Then the "1" moves to the $3, 3$ position. The matrix that permutes columns is Q:

$$PS = \begin{bmatrix} 4 & \mathbf{2} & 6 \\ 5 & 6 & \mathbf{3} \\ 1 & 4 & 5 \end{bmatrix} \qquad Q = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \qquad PSQ = \begin{bmatrix} \mathbf{2} & 6 & 4 \\ 6 & \mathbf{3} & 5 \\ 4 & 5 & 1 \end{bmatrix} \text{ is symmetric.}$$

The matrix Q is P^{T}. This choice always recovers symmetry, because PSP^{T} is guaranteed to be symmetric. (Its transpose is again PSP^{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 $S = LDL^{\mathrm{T}}$ for the matrix S above.

Solution To factor S into LDL^{T} we eliminate as usual to reach U:

$$S = \begin{bmatrix} 1 & 4 & 5 \\ 4 & 2 & 6 \\ 5 & 6 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 4 & 5 \\ 0 & -14 & -14 \\ 0 & -14 & -22 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 4 & 5 \\ 0 & -14 & -14 \\ 0 & 0 & -8 \end{bmatrix} = U.$$

The multipliers were $\ell_{21} = 4$ and $\ell_{31} = 5$ and $\ell_{32} = 1$. **The pivots $\mathbf{1, -14, -8}$ go into $\boldsymbol{D}$.** When we divide the rows of U by those pivots, L^{T} should appear:

Symmetric factorization when $\boldsymbol{S = S^{\mathrm{T}}}$

$$\boldsymbol{S = LDL^{\mathrm{T}}} = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 5 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & -14 & \\ & & -8 \end{bmatrix} \begin{bmatrix} 1 & 4 & 5 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

This matrix S is invertible because *it has three pivots*. Its inverse is $(L^{\mathrm{T}})^{-1}D^{-1}L^{-1}$ and S^{-1} is also symmetric. The numbers 14 and 8 will turn up in the denominators of S^{-1}. The "determinant" of S is the product of the pivots $(1)(-14)(-8) = 112$.

2.7 C For a rectangular A, this ***saddle-point matrix*** S is symmetric and important:

Block matrix from least squares $\quad S = \begin{bmatrix} I & A \\ A^{\mathrm{T}} & 0 \end{bmatrix} = S^{\mathrm{T}}$ has size $m+n$.

Apply block elimination to find a **block factorization** $S = LDL^{\mathrm{T}}$. Then test invertibility:

S is invertible $\iff$ ***$A^{\mathrm{T}}A$ is invertible*** $\iff$ ***$Ax \neq 0$ whenever $x \neq 0$***

Solution The first block pivot is I. Subtract A^{T} times row 1 from row 2:

Block elimination $\quad S = \begin{bmatrix} I & A \\ A^{\mathrm{T}} & 0 \end{bmatrix}$ goes to $\begin{bmatrix} I & A \\ 0 & -\boldsymbol{A^{\mathrm{T}}A} \end{bmatrix}$. This is U.

The block pivot matrix D contains I and $-A^{\mathrm{T}}A$. Then L and L^{T} contain A^{T} and A:

$$\boxed{\textbf{Block factorization} \quad S = LDL^{\mathrm{T}} = \begin{bmatrix} I & 0 \\ A^{\mathrm{T}} & I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & -A^{\mathrm{T}}A \end{bmatrix} \begin{bmatrix} I & A \\ 0 & I \end{bmatrix}.}$$

L is certainly invertible, with diagonal 1's. The inverse of the middle matrix involves $(A^{\mathrm{T}}A)^{-1}$. Section 4.2 answers a key question about the matrix $A^{\mathrm{T}}A$:

When is $A^{\mathrm{T}}A$ invertible? *Answer: A must have independent columns.*
Then $Ax=0$ only if $x=0$. Otherwise $Ax = 0$ will lead to $A^{\mathrm{T}}Ax=0$.

Problem Set 2.7

Questions 1–7 are about the rules for transpose matrices.

1 Find A^{T} and A^{-1} and $(A^{-1})^{\mathrm{T}}$ and $(A^{\mathrm{T}})^{-1}$ 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 $(AB)^{\mathrm{T}}$ equals $B^{\mathrm{T}}A^{\mathrm{T}}$ but those are different from $A^{\mathrm{T}}B^{\mathrm{T}}$:

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix} \qquad AB = \begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix}.$$

Show also that AA^{T} is different from $A^{\mathrm{T}}A$. But both of those matrices are ____.

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?

$$\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{y} = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \_\_\_\_ .$$

(b) This is the row $\boldsymbol{x}^{\mathrm{T}}A =$ ____ times the column $\boldsymbol{y} = (0, 1, 0)$.

(c) This is the row $\boldsymbol{x}^{\mathrm{T}} = [\,0 \;\; 1\,]$ times the column $A\boldsymbol{y} =$ ____ .

6 The transpose of a block matrix $M = \left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right]$ 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 $\left[\begin{smallmatrix} 0 & A \\ A & 0 \end{smallmatrix}\right]$ 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 $(P\boldsymbol{x})^{\mathrm{T}}(P\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ tells us 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 $S = S^{\mathrm{T}}$ with these properties:

(a) S is not invertible.

(b) S is invertible but cannot be factored into LU (row exchanges needed).

(c) S can be factored into LDL^{T} but not into LL^{T} (because of negative D).

18 (a) How many entries of S can be chosen independently, if $S = S^{\mathrm{T}}$ 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 A is rectangular (m by n) and S is symmetric (m by m).

(a) Transpose $A^{\mathrm{T}}SA$ to show its symmetry. What shape is this matrix?

(b) Show why $A^{\mathrm{T}}A$ has no negative numbers on its diagonal.

20 Factor these symmetric matrices into $S = LDL^{\mathrm{T}}$. The pivot matrix D is diagonal:

$$S = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix} \quad \text{and} \quad S = \begin{bmatrix} 1 & b \\ b & c \end{bmatrix} \quad \text{and} \quad S = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

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:

$$\text{Start from } S = \begin{bmatrix} 2 & 4 & 8 \\ 4 & 3 & 9 \\ 8 & 9 & 0 \end{bmatrix} \quad \text{and} \quad S = \begin{bmatrix} 1 & b & c \\ b & d & e \\ c & e & f \end{bmatrix}.$$

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 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).

26 (a) Choose E_{21} to remove the 3 below the first pivot. Then multiply $E_{21}SE_{21}^{\mathrm{T}}$ to remove both 3's:

$$S = \begin{bmatrix} 1 & 3 & 0 \\ 3 & 11 & 4 \\ 0 & 4 & 9 \end{bmatrix} \quad \text{is going toward} \quad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

(b) Choose E_{32} to remove the 4 below the second pivot. Then S is reduced to D by $E_{32}E_{21}SE_{21}^{\mathrm{T}}E_{32}^{\mathrm{T}} = D$. Invert the E's to find L in $S = LDL^{\mathrm{T}}$.

27 If every row of a 4 by 4 matrix contains the numbers $0, 1, 2, 3$ in some order, can the matrix be symmetric?

28 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 $(A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{y})$.

29 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 $\boldsymbol{y}$:

$$\boldsymbol{y} = A\boldsymbol{x} \quad \text{is} \quad \begin{bmatrix} y_{BC} \\ y_{CS} \\ y_{BS} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_B \\ x_C \\ x_S \end{bmatrix}.$$

(a) Find the total currents $A^{\mathrm{T}}\boldsymbol{y}$ out of the three cities.

(b) Verify that $(A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y}$ agrees with $\boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{y})$—six terms in both.

30 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 $A^{\mathrm{T}}\boldsymbol{y}$.

31 $A\boldsymbol{x}$ 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 _____.

32 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 $P\boldsymbol{v} = (-5, 2, 3)$?

33 Write $A = \begin{bmatrix} 1 & 2 \\ 4 & 9 \end{bmatrix}$ as the product ES of an elementary row operation matrix E and a symmetric matrix S.

34 Here is a new factorization of A into LS: *triangular* (with 1's) *times symmetric*:

$$\text{Start from } A = LDU. \text{ Then } A \text{ equals } L\,(U^{\mathrm{T}})^{-1} \text{ times } S = U^{\mathrm{T}}DU.$$

Why is $L\,(U^{\mathrm{T}})^{-1}$ triangular? Its diagonal is all 1's. Why is $U^{\mathrm{T}}DU$ symmetric?

35 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

36 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***?

37 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.

38 (a) Write down any 3 by 3 matrix M. Split M into $S + A$ where $S = S^{\mathrm{T}}$ is symmetric and $A = -A^{\mathrm{T}}$ is anti-symmetric.

(b) Find formulas for S and A involving M and M^{T}. We want $M = S + A$.

39 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 $q_{11} = \cos\theta$.

The Transpose of a Derivative

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 matrix changes to a derivative so $\boldsymbol{A = d/dt}$.** To find the transpose of this unusual A we need to define the inner product between two functions $x(t)$ and $y(t)$.

The inner product changes from the sum of $x_k\, y_k$ to the *integral* of $x(t)\, y(t)$.

Inner product of functions $$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = (\boldsymbol{x}, \boldsymbol{y}) = \int_{-\infty}^{\infty} \boldsymbol{x(t)\, y(t)\, dt}$$

From this inner product we know the requirement on A^{T}. The word "adjoint" is more correct than "transpose" when we are working with derivatives.

The transpose of a matrix has $(A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{y})$. The adjoint of $A = \dfrac{d}{dt}$ has

$$(Ax, y) = \int_{-\infty}^{\infty} \frac{dx}{dt}\, y(t)\, dt = \int_{-\infty}^{\infty} x(t)\left(-\frac{dy}{dt}\right) dt = (x, A^{\mathrm{T}}y)$$

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 *antisymmetric*: $\boldsymbol{A = d/dt}$ ***and*** $\boldsymbol{A^{\mathrm{T}} = -d/dt}$. Symmetric matrices have $S^{\mathrm{T}} = S$, antisymmetric matrices have $A^{\mathrm{T}} = -A$. Linear algebra includes derivatives and integrals in Chapter 8, *because those are both linear*.

This antisymmetry of the derivative applies also to centered difference matrices.

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix} \quad \text{transposes to} \quad A^{\mathrm{T}} = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} = -A.$$

And a forward difference matrix transposes to a backward difference matrix, *multiplied by* -1. In differential equations, the second derivative (acceleration) is symmetric. The first derivative (damping proportional to velocity) is *anti*symmetric.

Chapter 3

Vector Spaces and Subspaces

3.1 Spaces of Vectors

1 The standard n-dimensional space $\mathbf{R}^n$ contains all real column vectors with n components.

2 If $\boldsymbol{v}$ and $\boldsymbol{w}$ are in a **vector space** $\boldsymbol{S}$, every combination $c\boldsymbol{v} + d\boldsymbol{w}$ must be in $\boldsymbol{S}$.

3 The "vectors" in $\boldsymbol{S}$ can be matrices or functions of x. The 1-point space $\boldsymbol{Z}$ consists of $\boldsymbol{x} = \mathbf{0}$.

4 A **subspace** of $\mathbf{R}^n$ is a vector space inside $\mathbf{R}^n$. *Example*: The line $y = 3x$ inside $\mathbf{R}^2$.

5 The **column space** of A contains all combinations of the columns of A : a subspace of $\mathbf{R}^m$.

6 The column space contains all the vectors $A\boldsymbol{x}$. So $A\boldsymbol{x} = \boldsymbol{b}$ is solvable when $\boldsymbol{b}$ is in $\boldsymbol{C}(A)$.

To a newcomer, matrix calculations involve a lot of numbers. To you, they involve vectors. The columns of $A\boldsymbol{x}$ 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 $A\boldsymbol{x} = \boldsymbol{b}$.

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 $\boldsymbol{v}$ 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 $\boldsymbol{v}$ 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: $\boldsymbol{v} = (x, y)$.

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 $\boldsymbol{v}$ 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 $\boldsymbol{v}$ by any scalar c.

"Inside the vector space" means that ***the result stays in the space.*** If $\boldsymbol{v}$ is the vector in $\mathbf{R}^4$ with components $1, 0, 0, 1$, then $2\boldsymbol{v}$ 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 $\boldsymbol{v} + \boldsymbol{w} = \boldsymbol{w} + \boldsymbol{v}$; the distributive law is $c(\boldsymbol{v} + \boldsymbol{w}) = c\boldsymbol{v} + c\boldsymbol{w}$. There is a unique "zero vector" satisfying $\mathbf{0} + \boldsymbol{v} = \boldsymbol{v}$. 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 there are 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 $\mathbf{M}$ the "vectors" are really matrices. In $\mathbf{F}$ the vectors are functions. In $\mathbf{Z}$ 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 $\mathbf{M}$ or $\mathbf{F}$ or $\mathbf{Z}$. The eight conditions are all easily checked.

The function space $\mathbf{F}$ is infinite-dimensional. A smaller function space is $\mathbf{P}$, or $\mathbf{P}_n$, containing all polynomials $a_0 + a_1x + \cdots + a_nx^n$ of degree n.

The space $\mathbf{Z}$ is zero-dimensional (by any reasonable definition of dimension). $\mathbf{Z}$ 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 $\mathbf{Z}$ 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 $\mathbf{R}^3$.

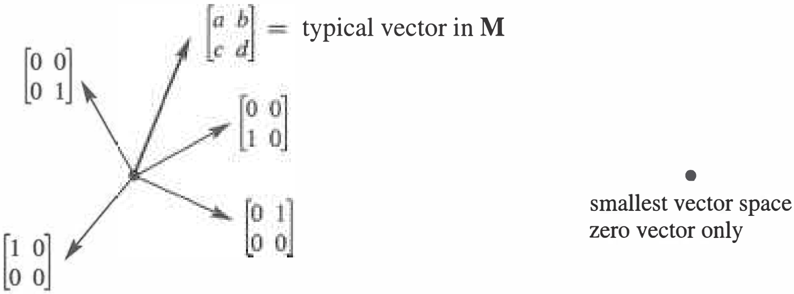


Figure 3.1: "Four-dimensional" matrix space $\mathbf{M}$. The "zero-dimensional" space $\mathbf{Z}$.

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. 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) $\boldsymbol{v} + \boldsymbol{w}$ is in the subspace

(ii) $c\boldsymbol{v}$ is in the subspace.

In other words, the set of vectors is "closed" under addition $\boldsymbol{v}+\boldsymbol{w}$ and multiplication $c\boldsymbol{v}$ (and $d\boldsymbol{w}$). Those operations leave us in the subspace. We can also subtract, because $-\boldsymbol{w}$ is in the subspace and its sum with $\boldsymbol{v}$ is $\boldsymbol{v}-\boldsymbol{w}$. 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 linear combinations requirement for a subspace.

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 $0\boldsymbol{v}$ to be in the subspace.

Planes that don't contain the origin fail those tests. Those planes are not subspaces.

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 $\mathbf{R}^3$:

$(\mathbf{L})$ Any line through $(0,0,0)$	$(\mathbf{R}^3)$ The whole space
$(\mathbf{P})$ Any plane through $(0,0,0)$	$(\mathbf{Z})$ 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 $\mathbf{R}^2$—they are not subspaces.

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 c and d. The rules can be combined into a single requirement—*the rule for subspaces*:

A subspace containing v and w must contain all linear combinations $c\boldsymbol{v}+d\boldsymbol{w}$.

Example 3 Inside the vector space $\mathbf{M}$ 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 $\mathbf{U}$, and the sum is in $\mathbf{U}$. Add diagonal matrices, and the sum is diagonal. In this case $\mathbf{D}$ is also a subspace of $\mathbf{U}$! Of course the zero matrix is in these subspaces, when a, b, and d all equal zero. $\mathbf{Z}$ is always a subspace.

Multiples of the identity matrix also form a subspace. $2I+3I$ is in this subspace, and so is 3 times $4I$. The matrices cI form a "line of matrices" inside $\mathbf{M}$ and $\mathbf{U}$ and $\mathbf{D}$.

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 $\boldsymbol{x}$. 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 $A\boldsymbol{x}$. They fill the column space $\boldsymbol{C}(A)$.

This column space is crucial to the whole book, and here is why. ***To solve*** $A\boldsymbol{x} = \boldsymbol{b}$ ***is to express b as a combination of the columns***. The right side $\boldsymbol{b}$ has to be *in the column space* produced by A on the left side, or no solution!

The system $A\boldsymbol{x} = \boldsymbol{b}$ ***is solvable if and only if b is in the column space of*** A.

When $\boldsymbol{b}$ is in the column space, it is a combination of the columns. The coefficients in that combination give us a solution $\boldsymbol{x}$ to the system $A\boldsymbol{x} = \boldsymbol{b}$.

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 $A\boldsymbol{x}$ 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. With only 2 columns, $\boldsymbol{C}(A)$ can't be all of $\mathbf{R}^3$.

Example 4

$$A\boldsymbol{x} \quad \text{is} \quad \begin{bmatrix} 1 & 0 \\ 4 & 3 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{which is} \quad x_1 \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 3 \\ 3 \end{bmatrix}.$$

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.

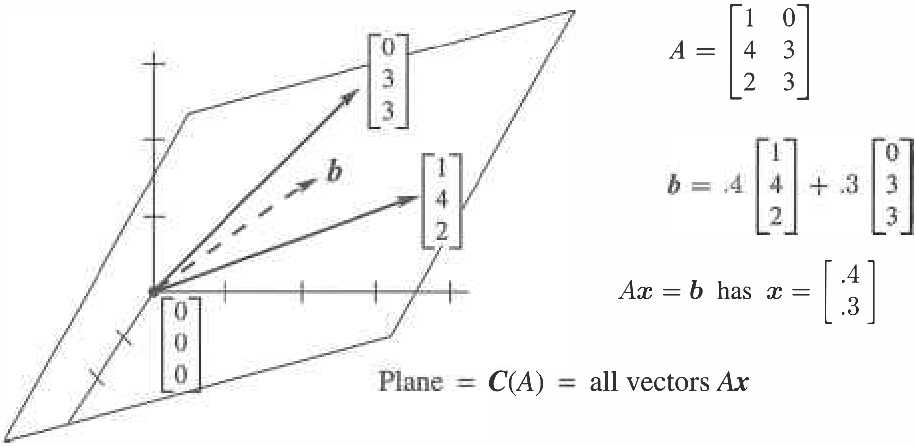


Figure 3.2: The column space $C(A)$ is a plane containing the two columns. $Ax = b$ is solvable when b is on that plane. Then b is a combination of the columns.

Of course $(0,0,0)$ is in the column space. The plane passes through the origin. There is certainly a solution to $Ax = \mathbf{0}$. That solution, always available, is $x =$ _____.

To repeat, the attainable right sides 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 $C(A)$. 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:

S = set of vectors in $\mathbf{V}$ (probably *not* a subspace)
SS = all combinations of vectors in S (definitely a subspace)

SS = all $c_1 v_1 + \cdots + c_N v_N$ = **the subspace of V "spanned" by S**

When S is the set of columns, SS is the column space. When there is only one nonzero vector v in S, the subspace SS is the line through v. *Always* SS *is the smallest subspace containing* S. This is a fundamental way to create subspaces and we will come back to it.

To repeat: The columns "span" the column space.

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.

$$\boldsymbol{C}(A) = \textbf{all combinations of the columns} = \textbf{all vectors } A\boldsymbol{x}.$$

■ 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 $\boldsymbol{b}_1$ 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* ....:

$$\begin{aligned}
&\mathbf{V}_1 = \text{all combinations of the 3 vectors} && \mathbf{V}_1 = \text{all solutions of } v_1 - v_2 = 0 \\
&\mathbf{V}_2 = \text{all combinations of } (1,0,-1) \text{ and } (1,-1,1) && \mathbf{V}_2 = \text{all solutions of } \boldsymbol{u} \cdot \boldsymbol{v} = 0. \\
&\mathbf{V}_3 = \text{all combinations of } \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right]. && \mathbf{V}_3 = \text{ all solutions } \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \text{ of } b = c \\
&\mathbf{V}_4 = \text{all combinations of } 1, x, x^2, x^3 && \mathbf{V}_4 = \text{all solutions to } d^4y/dx^4 = 0.
\end{aligned}$$

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) $(c_1c_2)\boldsymbol{x}=c_1(c_2\boldsymbol{x})$ — (1) to (4) about $\boldsymbol{x}+\boldsymbol{y}$

(7) $c(\boldsymbol{x}+\boldsymbol{y})=c\boldsymbol{x}+c\boldsymbol{y}$ — (5) to (6) about $c\boldsymbol{x}$

(8) $(c_1+c_2)\boldsymbol{x}=c_1\boldsymbol{x}+c_2\boldsymbol{x}$. — (7) to (8) connects them

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 $\mathbf{M}$ does contain A and B, must it contain I?

(c) Describe a subspace of $\mathbf{M}$ that contains no nonzero diagonal matrices.

6 The functions $\boldsymbol{f}(x)=x^2$ and $\boldsymbol{g}(x)=5x$ are "vectors" in $\mathbf{F}$. This is the vector space of all real functions. (The functions are defined for $-\infty<x<\infty$.) The combination $3\boldsymbol{f}(x)-4\boldsymbol{g}(x)$ is the function $\boldsymbol{h}(x)=$ ____ .

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 $cf(x)$ and find two rules that are broken.

Questions 9–18 are about the "subspace requirements": $x + y$ and cx (and then all linear combinations $cx + dy$) 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 $x + y$ stays in the set but $\frac{1}{2}x$ may be outside.

(b) A set of vectors in $\mathbf{R}^2$ (other than two quarter-planes) for which every cx stays in the set but $x + y$ 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 $v = (1, 4, 0)$ and $w = (2, 2, 2)$.

(e) All vectors that satisfy $b_1 + b_2 + b_3 = 0$.

(f) All vectors with $b_1 \leq b_2 \leq b_3$.

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 _____ in $\mathbf{R}^3$ 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 that $\boldsymbol{x}+\boldsymbol{y}$ and $c\boldsymbol{x}$ are in $\mathbf{S} \cap \mathbf{T}$ if $\boldsymbol{x}$ and $\boldsymbol{y}$ are in both spaces.*

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 $\mathbf{S}$ and $\mathbf{T}$ are two subspaces of a vector space $\mathbf{V}$.

(a) **Definition**: The **sum** $\mathbf{S} + \mathbf{T}$ contains all sums $\boldsymbol{s} + \boldsymbol{t}$ of a vector $\boldsymbol{s}$ in $\mathbf{S}$ and a vector $\boldsymbol{t}$ in $\mathbf{T}$. Show that $\mathbf{S} + \mathbf{T}$ satisfies the requirements (addition and scalar multiplication) for a vector space.

(b) If $\mathbf{S}$ and $\mathbf{T}$ are lines in $\mathbf{R}^m$, what is the difference between $\mathbf{S} + \mathbf{T}$ and $\mathbf{S} \cup \mathbf{T}$? That union contains all vectors from $\mathbf{S}$ or $\mathbf{T}$ or both. Explain this statement: *The span of* $\mathbf{S} \cup \mathbf{T}$ *is* $\mathbf{S} + \mathbf{T}$. (Section 3.5 returns to this word "span".)

31 If $\mathbf{S}$ is the column space of A and $\mathbf{T}$ is $\boldsymbol{C}(B)$, then $\mathbf{S} + \mathbf{T}$ 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 $Ax = 0$ and $Rx = 0$

1 The **nullspace** $\boldsymbol{N}(A)$ in $\mathbf{R}^n$ contains all solutions $\boldsymbol{x}$ to $A\boldsymbol{x} = \mathbf{0}$. This includes $\boldsymbol{x} = \mathbf{0}$.

2 Elimination (from A to U to R) does not change the nullspace: $\boldsymbol{N}(A) = \boldsymbol{N}(U) = \boldsymbol{N}(R)$.

3 The **reduced row echelon form $\boldsymbol{R}$ = rref$(\boldsymbol{A})$** has all pivots $= 1$, with zeros above and below.

4 If column j of R is free (no pivot), there is a "*special solution*" to $A\boldsymbol{x} = \mathbf{0}$ with $x_j = 1$.

5 Number of pivots = number of nonzero rows in R = **rank $\boldsymbol{r}$**. There are $n - r$ free columns.

6 Every matrix with $m < n$ has nonzero solutions to $A\boldsymbol{x} = \mathbf{0}$ in its nullspace.

This section is about the subspace containing all solutions to $A\boldsymbol{x} = \mathbf{0}$. The m by n matrix A can be square or rectangular. The right hand side is $\boldsymbol{b} = \mathbf{0}$. *One immediate solution is* $\boldsymbol{x} = \mathbf{0}$. For invertible matrices this is the only solution. For other matrices, not invertible, there are nonzero solutions to $A\boldsymbol{x} = \mathbf{0}$. *Each solution $\boldsymbol{x}$ belongs to the nullspace of A.*

Elimination will find all solutions and identify this very important subspace.

The nullspace $\boldsymbol{N}(A)$ consists of all solutions to $A\boldsymbol{x} = \mathbf{0}$. These vectors $\boldsymbol{x}$ are in $\mathbf{R}^n$.

Check that the solution vectors form a subspace. Suppose $\boldsymbol{x}$ and $\boldsymbol{y}$ are in the nullspace (this means $A\boldsymbol{x} = \mathbf{0}$ and $A\boldsymbol{y} = \mathbf{0}$). The rules of matrix multiplication give $A(\boldsymbol{x} + \boldsymbol{y}) = \mathbf{0} + \mathbf{0}$. The rules also give $A(c\boldsymbol{x}) = c\mathbf{0}$. 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 $\mathbf{R}^m$.

Example 1 Describe the nullspace of $A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$. This matrix is singular !

Solution Apply elimination to the linear equations $A\boldsymbol{x} = \mathbf{0}$:

$$\begin{aligned} x_1 + 2x_2 &= 0 \\ 3x_1 + 6x_2 &= 0 \end{aligned} \quad \rightarrow \quad \begin{aligned} x_1 + 2x_2 &= 0 \\ 0 &= 0 \end{aligned}$$

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 the solutions to $A\boldsymbol{x} = \mathbf{0}$, 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 is $\boldsymbol{s} = (-2, 1)$.**

Special solution	$As = \mathbf{0}$	The nullspace of $A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$ contains all multiples of $s = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$.

This is the best way to describe the nullspace, by computing special solutions to $A\boldsymbol{x} = \mathbf{0}$. **The solution is special because we set the free variable to $x_2 = 1$.**

The nullspace of A consists of all combinations of the special solutions to $A\boldsymbol{x} = \mathbf{0}$.

Example 2 $x + 2y + 3z = 0$ comes from the 1 by 3 matrix $A = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$. Then $A\boldsymbol{x} = \mathbf{0}$ produces a plane. All vectors on the plane are perpendicular to $(1, 2, 3)$. *The plane is the nullspace of A.* There are two free variables y and z: Set to 0 and 1.

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0 \text{ has two special solutions } \boldsymbol{s}_1 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \text{ and } \boldsymbol{s}_2 = \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}.$$

Those vectors $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ lie on the plane $x + 2y + 3z = 0$. All vectors on the plane are combinations of $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$.

Notice what is special about $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$. *The last two components are "free" and we choose them specially as* $1, 0$ *and* $0, 1$. Then the first components -2 and -3 are determined by the equation $A\boldsymbol{x} = \mathbf{0}$.

The solutions to $x + 2y + 3z = \mathbf{6}$ also lie on a plane, but that plane is not a subspace. The vector $\boldsymbol{x} = \mathbf{0}$ is only a solution if $\boldsymbol{b} = \mathbf{0}$. Section 3.3 will show how the solutions to $A\boldsymbol{x} = \boldsymbol{b}$ (if there are any solutions) are shifted away from zero by one particular solution.

The two key steps of this section are
(**1**) reducing A to its **row echelon form R**
(**2**) finding the **special solutions to $A\boldsymbol{x} = \mathbf{0}$**

The display on page 138 shows 4 by 5 matrices A and R, with 3 pivots.
The equations $A\boldsymbol{x} = \mathbf{0}$ and also $R\boldsymbol{x} = \mathbf{0}$ have $5 - 3 = 2$ special solutions $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$.

Pivot Columns and Free Columns

The first column of $A = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$ contains the only pivot, so the first component of $\boldsymbol{x}$ is *not free*. **The free components correspond to columns with no pivots**. The special choice (one or zero) is only for the free variables in the special solutions.

Example 3 Find the nullspaces of A, B, C and the two special solutions to $C\boldsymbol{x} = \mathbf{0}$.

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 3 & 8 \end{bmatrix} \quad \boldsymbol{B} = \begin{bmatrix} A \\ 2A \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 8 \\ 2 & 4 \\ 6 & 16 \end{bmatrix} \quad \boldsymbol{C} = \begin{bmatrix} A & 2A \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 3 & 8 & 6 & 16 \end{bmatrix}.$$

Solution The equation $Ax = \mathbf{0}$ has only the zero solution $\boldsymbol{x} = \mathbf{0}$. *The nullspace is* $\mathbf{Z}$. It contains only the single point $\boldsymbol{x} = \mathbf{0}$ in $\mathbf{R}^2$. This fact comes from elimination:

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 2 \\ 3 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ yields } \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} \boldsymbol{x_1} = \mathbf{0} \\ \boldsymbol{x_2} = \mathbf{0} \end{bmatrix}.$$

A is invertible. There are no special solutions. Both columns of this matrix have pivots.

The rectangular matrix B has the same nullspace $\mathbf{Z}$. The first two equations in $B\boldsymbol{x} = \mathbf{0}$ again require $\boldsymbol{x} = \mathbf{0}$. The last two equations would also force $\boldsymbol{x} = \mathbf{0}$. When we add extra equations (giving extra rows), the nullspace certainly cannot become larger. The extra rows impose more conditions on the vectors $\boldsymbol{x}$ in the nullspace.

The rectangular matrix C is different. It has extra columns instead of extra rows. The solution vector $\boldsymbol{x}$ has *four* components. Elimination will produce pivots in the first two columns of C, but **the last two columns of C and U are "free". They don't have pivots:**

Subtract 3 (row 1) from row 2 of C

$$C = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 3 & 8 & 6 & 16 \end{bmatrix} \text{ becomes } U = \begin{bmatrix} \mathbf{1} & 2 & 2 & 4 \\ 0 & \mathbf{2} & 0 & 4 \end{bmatrix}$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \uparrow \quad \uparrow \quad \uparrow \quad \uparrow$$

pivot columns **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 $U\boldsymbol{x} = \mathbf{0}$ (or $C\boldsymbol{x} = \mathbf{0}$ or eventually $R\boldsymbol{x} = \mathbf{0}$). We get two special solutions in the nullspace of C. This is also the nullspace of U: elimination doesn't change solutions.

Special solutions
$\boldsymbol{Cs = 0}$
$\boldsymbol{Us = 0}$

$$s_1 = \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \end{bmatrix} \text{ and } s_2 = \begin{bmatrix} 0 \\ -2 \\ 0 \\ 1 \end{bmatrix} \begin{array}{l} \leftarrow \textbf{pivot} \\ \leftarrow \quad \textbf{variables} \\ \leftarrow \textbf{free} \\ \leftarrow \quad \textbf{variables} \end{array}$$

The Reduced Row Echelon Form R

When A is rectangular, elimination will not stop at the upper triangular U. We can continue to make this matrix simpler, in two ways. These steps bring us to the best matrix R:

1. *Produce zeros above the pivots.* **Use pivot rows to eliminate upward in R.**

2. *Produce ones in the pivots.* **Divide the whole pivot row by its pivot.**

Those steps don't change the zero vector on the right side of the equation. The nullspace stays the same: $\boldsymbol{N}(A) = \boldsymbol{N}(U) = \boldsymbol{N}(R)$. This nullspace becomes easiest to see when we reach the ***reduced row echelon form*** $\boldsymbol{R} = \mathbf{rref}\,(\boldsymbol{A})$. ***The pivot columns of R contain I.***

Reduced form R $\quad U = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 2 & 0 & 4 \end{bmatrix}$ becomes $R = \begin{bmatrix} \mathbf{1} & \mathbf{0} & 2 & 0 \\ \mathbf{0} & \mathbf{1} & 0 & 2 \end{bmatrix}$.
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\uparrow\ \uparrow$

I subtracted row 2 of U from row 1. Then I multiplied row 2 by $\frac{1}{2}$ to get pivot $= 1$.

Now (**free column 3**) $=$ **2** (**pivot column 1**), so -2 appears in $\boldsymbol{s}_1 = (-2, 0, 1, 0)$. The special solutions are much easier to find from the reduced system $R\boldsymbol{x} = \mathbf{0}$. In each free column of R, I change all the signs to find $\boldsymbol{s}$. Second special solution $\boldsymbol{s}_2 = (0, -2, 0, 1)$.

Before moving to m by n matrices A and their nullspaces $\boldsymbol{N}(A)$ and special solutions, allow me to repeat one comment. For many matrices, the only solution to $A\boldsymbol{x} = \mathbf{0}$ is $\boldsymbol{x} = \mathbf{0}$. Their nullspaces $\boldsymbol{N}(A) = \mathbf{Z}$ contain only that zero vector: *no* special solutions. The only combination of the columns that produces $\boldsymbol{b} = \mathbf{0}$ is then the "zero combination". The solution to $A\boldsymbol{x} = \mathbf{0}$ is trivial (just $\boldsymbol{x} = \mathbf{0}$) 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. . .

Pivot Variables and Free Variables in the Echelon Matrix R

$$\boldsymbol{A} = \begin{bmatrix} p & p & f & p & f \\ | & | & | & | & | \\ | & | & | & | & | \end{bmatrix} \qquad \boldsymbol{R} = \begin{bmatrix} \mathbf{1} & 0 & a & 0 & c \\ 0 & \mathbf{1} & b & 0 & d \\ 0 & 0 & 0 & \mathbf{1} & e \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \boldsymbol{s}_1 = \begin{bmatrix} -a \\ -b \\ \mathbf{1} \\ 0 \\ \mathbf{0} \end{bmatrix} \qquad \boldsymbol{s}_2 = \begin{bmatrix} -c \\ -d \\ \mathbf{0} \\ -e \\ \mathbf{1} \end{bmatrix}$$

3 pivot columns p	I in pivot columns	special $R\boldsymbol{s}_1 = \mathbf{0}$ and $R\boldsymbol{s}_2 = \mathbf{0}$
2 free columns f	F in free columns	take $-a$ to $-e$ from R
to be revealed by R	3 pivots: rank $r = 3$	$R\boldsymbol{s} = \mathbf{0}$ means $A\boldsymbol{s} = \mathbf{0}$

R shows clearly: *column* 3 $= a\,$(*column* 1) $+ b\,$(*column* 2). The same must be true for A.
The special solution $\boldsymbol{s}_1$ repeats that combination so $(-\boldsymbol{a}, -\boldsymbol{b}, \mathbf{1}, 0, 0)$ has $R\boldsymbol{s}_1 = \mathbf{0}$.
Nullspace of $A =$ Nullspace of $R =$ all combinations of $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$.

Here are those steps for a 4 by 7 ***reduced row echelon matrix R*** with three pivots:

$$\boldsymbol{R} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & x & x & x & \mathbf{0} & x \\ 0 & \mathbf{1} & x & x & x & 0 & x \\ \mathbf{0} & \mathbf{0} & 0 & 0 & 0 & \mathbf{1} & x \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Three pivot variables x_1, x_2, x_6
Four free variables x_3, x_4, x_5, x_7
Four special solutions $\boldsymbol{s}$ **in** $\boldsymbol{N}(\boldsymbol{R})$
The pivot rows and columns contain $\boldsymbol{I}$

Question What are the column space and the nullspace for this matrix R?

Answer The columns of R 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* $\boldsymbol{C}(R)$ *consists of all vectors of the form* $(b_1, b_2, b_3, 0)$. For those vectors we can solve $R\boldsymbol{x} = \boldsymbol{b}$.

The nullspace $\boldsymbol{N}(R)$ is a subspace of $\mathbf{R}^7$. The solutions to $R\boldsymbol{x} = \mathbf{0}$ 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 four free variables are x_3, x_4, x_5, x_7.

2. Set one free variable to 1 and set the other three free variables to zero.

3. To find $\boldsymbol{s}$, solve $R\boldsymbol{x} = \mathbf{0}$ for the pivot variables x_1, x_2, x_6.

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 $A\boldsymbol{x} = \mathbf{0}$ has at least one special solution. This solution is *not zero*!

Suppose $A\boldsymbol{x} = \mathbf{0}$ has more unknowns than equations ($\boldsymbol{n} > \boldsymbol{m}$, more columns than rows). There must be at least one free column. **Then $A\boldsymbol{x} = \mathbf{0}$ has nonzero solutions.**

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 $A\boldsymbol{x} = \mathbf{0}$ has at least a line of nonzero solutions.

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 Rank of a Matrix

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 OF RANK ***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 final matrix R will have r nonzero rows. Start with a 3 by 4 example of rank $r = 2$:

Four columns
Two pivots

$$A = \begin{bmatrix} 1 & 1 & 2 & 4 \\ 1 & 2 & 2 & 5 \\ 1 & 3 & 2 & 6 \end{bmatrix} \qquad R = \begin{bmatrix} \mathbf{1} & 0 & 2 & 3 \\ 0 & \mathbf{1} & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The first two columns of A are $(1, 1, 1)$ and $(1, 2, 3)$, going in different directions. Those will be pivot columns (revealed by R). 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 the sum of the first three. That fourth column will also have no pivot. The rank of A and R is **2**.

Every "free column" is a combination of earlier pivot columns. It is the special solutions s that tell us those combinations:

$$\begin{array}{ll} \text{Column } 3 = \mathbf{2}\ (\text{column } 1) + \mathbf{0}\ (\text{column } 2) & \boldsymbol{s}_1 = (\mathbf{-2}, \mathbf{-0}, 1, 0) \\ \text{Column } 4 = \mathbf{3}\ (\text{column } 1) + \mathbf{1}\ (\text{column } 2) & \boldsymbol{s}_2 = (\mathbf{-3}, \mathbf{-1}, 0, 1) \end{array}$$

The numbers $2, 0$ in column 3 of R show up in $\boldsymbol{s}_1$ (with signs reversed). And the numbers $3, 1$ in column 4 of R show up in $\boldsymbol{s}_2$ (with signs reversed to $-3, -1$).

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!

Rank one matrix $$A = \begin{bmatrix} \mathbf{1} & 3 & 10 \\ \mathbf{2} & 6 & 20 \\ \mathbf{3} & 9 & 30 \end{bmatrix} \longrightarrow R = \begin{bmatrix} \mathbf{1} & 3 & 10 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

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 $A = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$:

$$\boldsymbol{A} = \textbf{column times row} = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} \qquad \begin{bmatrix} 1 & 3 & 10 \\ 2 & 6 & 20 \\ 3 & 9 & 30 \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \end{bmatrix} [\mathbf{1} \quad \mathbf{3} \quad \mathbf{10}]$$

With rank one, $A\boldsymbol{x} = \mathbf{0}$ is 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 when $r = 1$: *row space = line, nullspace = perpendicular plane*.

Example 4 When all rows are multiples of one pivot row, the rank is $r = 1$:

$$\begin{bmatrix} 1 & 3 & 4 \\ 2 & 6 & 8 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 3 \\ 0 & \mathbf{5} \end{bmatrix} \text{ and } \begin{bmatrix} 5 \\ 2 \end{bmatrix} \text{ and } \begin{bmatrix} 6 \end{bmatrix} \text{ all have rank 1.}$$

For those matrices, the reduced row echelon $R =$ **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. All three matrices A and U and R have r **independent rows**.

A and U and R also have $\boldsymbol{r}$ **independent columns** (the pivot columns). Section 3.4 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 $\boldsymbol{n} - \boldsymbol{r}$ **is the dimension of the nullspace.**

■ REVIEW OF THE KEY IDEAS ■

1. The nullspace $\boldsymbol{N}(A)$ is a subspace of $\mathbf{R}^n$. It contains all solutions to $A\boldsymbol{x} = \mathbf{0}$.
2. Elimination on A produces a row reduced R with pivot columns and free columns.
3. Every free column leads to a special solution. That free variable is 1, the others are 0.
4. The *rank* r of A is the number of pivots. All pivots are 1's in $R = \text{rref(A)}$.
5. The complete solution to $A\boldsymbol{x} = \mathbf{0}$ is a combination of the $n - r$ special solutions.
6. A always has a free column if $n > m$, giving a *nonzero solution* to $A\boldsymbol{x} = \mathbf{0}$.

■ WORKED EXAMPLES ■

3.2 A Why do A and R have the same nullspace if $EA = R$ and E is invertible?

Solution If $A\boldsymbol{x} = \mathbf{0}$ then $R\boldsymbol{x} = EA\boldsymbol{x} = E\mathbf{0} = \mathbf{0}$
If $R\boldsymbol{x} = \mathbf{0}$ then $A\boldsymbol{x} = E^{-1}R\boldsymbol{x} = E^{-1}\mathbf{0} = \mathbf{0}$
A and R also have the same row space and the same rank.

3.2 B Create a 3 by 4 matrix R whose special solutions to $R\boldsymbol{x} = \mathbf{0}$ are $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$:

$$\boldsymbol{s}_1 = \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{s}_2 = \begin{bmatrix} -2 \\ 0 \\ -6 \\ 1 \end{bmatrix} \qquad \begin{array}{l} \text{pivot columns 1 and 3} \\ \text{free variables } x_2 \text{ and } x_4 \end{array}$$

Describe all possible matrices A with this nullspace $\boldsymbol{N}(A) =$ all combinations of $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$.

Solution The reduced matrix R has pivots $= 1$ in columns 1 and 3. There is no third pivot, so row 3 of R is all zeros. The free columns 2 and 4 will be combinations of the pivot columns: $3, 0, 2, 6$ in R come from $-3, -0, -2, -6$ in $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$. **Every** $\boldsymbol{A} = \boldsymbol{ER}$.

Every 3 by 4 matrix has at least one special solution. *These matrices have two.*

$$R = \begin{bmatrix} \mathbf{1} & \mathbf{3} & 0 & \mathbf{2} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{6} \\ \mathbf{0} & \mathbf{0} & 0 & 0 \end{bmatrix} \quad \text{has} \quad R\boldsymbol{s}_1 = \mathbf{0} \quad \text{and} \quad R\boldsymbol{s}_2 = \mathbf{0}.$$

3.2 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?

Find special solutions $A = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 6 & 3 \\ 4 & 8 & c \end{bmatrix}$ and $B = \begin{bmatrix} c & c \\ c & c \end{bmatrix}$.

Solution The matrix A has row $2 = 3$ (row 1). The rank of A is $r = 2$ ***except if*** $\boldsymbol{c = 4}$. Row $4 - 4$ (row 1) ends in $c - 4$. The pivots are in columns 1 and 3. The second variable x_2 is free. Notice the form of R: Row 3 has moved up into row 2.

$$\boldsymbol{c \neq 4} \quad R = \begin{bmatrix} \mathbf{1} & 2 & 0 \\ 0 & 0 & \mathbf{1} \\ 0 & 0 & 0 \end{bmatrix} \qquad \boldsymbol{c = 4} \quad R = \begin{bmatrix} \mathbf{1} & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

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:

$\boldsymbol{c \neq 4}$ Special solution $(-2, 1, 0)$ $\quad\boldsymbol{c = 4}$ Another special solution $(-1, 0, 1)$.

The 2 by 2 matrix $B = \begin{bmatrix} c & c \\ c & c \end{bmatrix}$ has rank $r = 1$ ***except if*** $\boldsymbol{c = 0}$, when the rank is zero!

$$\boldsymbol{c \neq 0} \quad R = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \qquad \boldsymbol{c = 0} \quad R = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \text{ and nullspace} = \mathbf{R}^2.$$

Problem Set 3.2

1 Reduce A and B to their triangular echelon forms U. Which variables are free?

$$\text{(a)}\ A = \begin{bmatrix} 1 & 2 & 2 & 4 & 6 \\ 1 & 2 & 3 & 6 & 9 \\ 0 & 0 & 1 & 2 & 3 \end{bmatrix} \qquad \text{(b)}\ B = \begin{bmatrix} 2 & 4 & 2 \\ 0 & 4 & 4 \\ 0 & 8 & 8 \end{bmatrix}.$$

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 further row operations on each U in Problem 1, find the reduced echelon form R. *True or false with a reason*: The nullspace of R equals the nullspace of U.

4 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 ______. This is the **Counting Theorem**: $\boldsymbol{r + (n - r) = n}$.

$$\text{(a)}\ A = \begin{bmatrix} -1 & 3 & 5 \\ -2 & 6 & 10 \end{bmatrix} \qquad \text{(b)}\ B = \begin{bmatrix} -1 & 3 & 5 \\ -2 & 6 & 7 \end{bmatrix}.$$

Questions 5–14 are about free variables and pivot variables.

5 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.

6 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.

7 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.

8 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 $\boldsymbol{x} =$ _____.

9 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.

10 Suppose an m by n matrix has r pivots. The number of special solutions is _____. The nullspace contains only $\boldsymbol{x} = \boldsymbol{0}$ when $r =$ _____. The column space is all of $\mathbf{R}^m$ when $r =$ _____.

11 The nullspace of a 5 by 5 matrix contains only $\boldsymbol{x} = \boldsymbol{0}$ when the matrix has _____ pivots. The column space is $\mathbf{R}^5$ when there are _____ pivots. Explain why.

12 The equation $x - 3y - z = 0$ determines a plane in $\mathbf{R}^3$. What is the matrix A in this equation? Which variables are free? The special solutions are _____ and _____.

13 (Recommended) The plane $x - 3y - z = 12$ is parallel to $x - 3y - z = 0$. One particular point on this plane is $(12, 0, 0)$. All points on the plane have the form

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \\ 0 \\ 0 \end{bmatrix} + y \begin{bmatrix} \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} \\ 0 \\ 1 \end{bmatrix}.$$

14 Suppose column 1 + column 3 + column 5 = $\boldsymbol{0}$ in a 4 by 5 matrix with four pivots. Which column has no pivot? What is the special solution? Describe $\boldsymbol{N}(A)$.

Questions 15–22 ask for matrices (if possible) with specific properties.

15 Construct a matrix for which $\boldsymbol{N}(A) =$ all combinations of $(2, 2, 1, 0)$ and $(3, 1, 0, 1)$.

16 Construct A so that $\boldsymbol{N}(A) =$ all multiples of $(4, 3, 2, 1)$. Its rank is _____.

17 Construct a matrix whose column space contains $(1,1,5)$ and $(0,3,1)$ and whose nullspace contains $(1,1,2)$.

18 Construct a matrix whose column space contains $(1,1,0)$ and $(0,1,1)$ and whose nullspace contains $(1,0,1)$ and $(0,0,1)$.

19 Construct a matrix whose column space contains $(1,1,1)$ and whose nullspace is the line of multiples of $(1,1,1,1)$.

20 Construct a 2 by 2 matrix whose nullspace equals its column space. This is possible.

21 Why does no 3 by 3 matrix have a nullspace that equals its column space?

22 If $AB = 0$ then the column space of B is contained in the _____ of A. Why?

23 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?

24 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 $\mathbf{rref}(A^{\mathrm{T}})$.

25 If $\boldsymbol{N}(A) =$ all multiples of $\boldsymbol{x} = (2,1,0,1)$, what is R and what is its rank?

26 If the special solutions to $R\boldsymbol{x} = \mathbf{0}$ are in the columns of these nullspace matrices N, go backward to find the nonzero rows of the reduced matrices R:

$$N = \begin{bmatrix} 2 & 3 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad N = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad N = \begin{bmatrix} \\ \\ \\ \end{bmatrix} \quad \text{(empty 3 by 1).}$$

27 (a) What are the five 2 by 2 reduced 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?

28 Explain why A and $-A$ always have the same reduced echelon form R.

29 If A is 4 by 4 and invertible, describe the nullspace of the 4 by 8 matrix $B = [A \ \ A]$.

30 How is the nullspace $\boldsymbol{N}(C)$ related to the spaces $\boldsymbol{N}(A)$ and $\boldsymbol{N}(B)$, if $C = \begin{bmatrix} A \\ B \end{bmatrix}$?

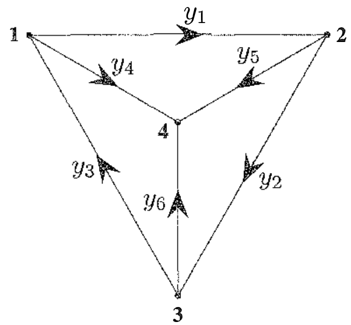
31 Find the reduced row echelon forms R and the rank of these matrices:

(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 $a_{ij} = (-1)^j$.

32 Kirchhoff's Current Law $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$ says that *current in* = *current out* at every node. At node 1 this is $y_3 = y_1 + y_4$. Write the four equations for Kirchhoff's Law at the four nodes (arrows show the positive direction of each y). Reduce A^{T} to R and find three special solutions in the nullspace of A^{T} (4 by 6 matrix).



$$A = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

33 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.

34 Find the reduced R for each of these (block) matrices:

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 3 \\ 2 & 4 & 6 \end{bmatrix} \qquad B = \begin{bmatrix} A & A \end{bmatrix} \qquad C = \begin{bmatrix} A & A \\ A & 0 \end{bmatrix}$$

35 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?

36 (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 $\mathbf{rref}(A_1 - A_2)$?

37 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}.

38 What are the special solutions to $R\boldsymbol{x} = \mathbf{0}$ and $\boldsymbol{y}^{\mathrm{T}}R = \mathbf{0}$ for these R?

$$R = \begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & 4 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad R = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

39 Fill out these matrices so that they have rank 1:

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 2 & & \\ 4 & & \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} & 9 & \\ 1 & & \\ 2 & 6 & -3 \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} a & b \\ c & \end{bmatrix}.$$

40 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 _____.

41 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* $r = 1$.

42 If A is a rank one matrix, the second row of R is _____. Do an example.

Problems 43–45 are about r by r invertible matrices inside A.

43 *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:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix} \qquad C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

44 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.

45 Transpose P in Problem 44. Find the r pivot columns of P^{T} (which is r by m). 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 46–51 show that $\operatorname{rank}(AB)$ is not greater than $\operatorname{rank}(A)$ or $\operatorname{rank}(B)$.

46 Find the ranks of AB and AC (rank one matrix times rank one matrix):

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & 1 & 4 \\ 3 & 1.5 & 6 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 1 & b \\ c & bc \end{bmatrix}.$$

47 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 _____ $= 0$.

48 (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 $\mathbf{rank}(\boldsymbol{AB}) \leq \mathbf{rank}(\boldsymbol{B})$.

(b) Find A_1 and A_2 so that $\operatorname{rank}(A_1B) = 1$ and $\operatorname{rank}(A_2B) = 0$ for $B = \left[\begin{smallmatrix}1&1\\1&1\end{smallmatrix}\right]$.

49 Problem 48 proved that $\operatorname{rank}(AB) \leq \operatorname{rank}(B)$. Then the same reasoning gives $\mathbf{rank}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}) \leq \operatorname{rank}(A^{\mathrm{T}})$. How do you deduce that $\mathbf{rank}(\boldsymbol{AB}) \leq \mathbf{rank}\ \boldsymbol{A}$?

50 *(Important)* Suppose A and B are n by n matrices, and $AB = I$. Prove from $\operatorname{rank}(AB) \leq \operatorname{rank}(A)$ 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!)*.

51 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.

52 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.

53 Express A and then B as the sum of two rank one matrices:

$$\mathbf{rank} = \mathbf{2} \qquad A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 4 \\ 1 & 1 & 8 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}.$$

54 Answer the same questions as in Worked Example **3.2 C** for

$$A = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 2 & 2 & 4 & 4 \\ 1 & c & 2 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1-c & 2 \\ 0 & 2-c \end{bmatrix}.$$

55 What is the nullspace matrix N (containing the special solutions) for A, B, C?

$$\text{Block matrices} \quad A = \begin{bmatrix} I & I \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} I & I \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} I & I & I \end{bmatrix}.$$

56 *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 of all ones as the product of the 3 by 1 matrix from the pivot columns and the 1 by 4 matrix from R.

Challenge Problems

57 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^{T} :

$$R = \text{rref}(A) \quad \text{and} \quad Z = (\text{rref}(A^{\mathrm{T}}))^{\mathrm{T}}.$$

58 (Recommended) 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 $r = m$. The zero blocks are gone.

(c) Find a *left-inverse* C with $CR = I$ if $r = n$. The F and 0 column is gone.

(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)?

59 I think that the reduced echelon form of $R^{\mathrm{T}}R$ is always R (except for extra zero rows). Can you do an example when R is 2 by 3 ? Later we show that $A^{\mathrm{T}}A$ always has the same nullspace as A (a valuable fact).

60 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?

Elimination: The Big Picture

This page explains elimination at the vector level and subspace level, when A is reduced to R. You know the steps and I won't repeat them. Elimination starts with the first pivot. It moves a column at a time (left to right) and a row at a time (top to bottom). As it moves, elimination answers two questions:

Question 1 Is this column a combination of previous columns?

If the column contains a pivot, the answer is no. Pivot columns are "independent" of previous columns. If column 4 has no pivot, it is a combination of columns $1, 2, 3$.

Question 2 Is this row a combination of previous rows?

If the row contains a pivot, the answer is no. Pivot rows are "independent" of previous rows. If row 3 ends up with no pivot, it is a zero row and it is moved to the bottom of R.

It is amazing to me that one pass through the matrix answers both questions. Actually that pass reaches the triangular echelon matrix U, not the reduced echelon matrix R. Then the reduction from U to R goes bottom to top. U tells which columns are combinations of earlier columns (pivots are missing). Then *R tells us what those combinations are*.

In other words, **$\boldsymbol{R}$ tells us the special solutions to $\boldsymbol{Ax} = \mathbf{0}$**. We could reach R from A by different row exchanges and elimination steps, but it will always be the same R (because the special solutions are decided by A). In the language coming soon, R reveals a "basis" for three fundamental subspaces:

The **column space** of A—choose the pivot columns of A as a basis.

The **row space** of A—choose the nonzero rows of R as a basis.

The **nullspace** of A—choose the special solutions to $R\boldsymbol{x} = \mathbf{0}$ (and $A\boldsymbol{x} = \mathbf{0}$).

We learn from elimination the single most important number—**the rank $\boldsymbol{r}$**. That number counts the pivot columns and the pivot rows. Then $n - r$ counts the free columns and the special solutions.

I mention that reducing $[A \quad I]$ to $[R \quad E]$ will tell you even more about A—in fact virtually everything (including $EA = R$). The matrix E keeps a record, otherwise lost, of the elimination from A to R. When A is square and invertible, R is I and E is A^{-1}.

3.3 The Complete Solution to $Ax = b$

1 **Complete solution** to $A\boldsymbol{x} = \boldsymbol{b}$: $\boldsymbol{x} =$ (one particular solution $\boldsymbol{x}_p$) + (any $\boldsymbol{x}_n$ in the nullspace).

2 Elimination on $\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix}$ leads to $\begin{bmatrix} R & \boldsymbol{d} \end{bmatrix}$. Then $A\boldsymbol{x} = \boldsymbol{b}$ is equivalent to $R\boldsymbol{x} = \boldsymbol{d}$.

3 $A\boldsymbol{x} = \boldsymbol{b}$ and $R\boldsymbol{x} = \boldsymbol{d}$ are solvable only when all zero rows of R have zeros in $\boldsymbol{d}$.

4 When $R\boldsymbol{x} = \boldsymbol{d}$ is solvable, one very particular solution $\boldsymbol{x}_p$ has all free variables equal to zero.

5 A has **full column rank** $\boldsymbol{r = n}$ when its nullspace $\boldsymbol{N}(A) =$ zero vector: *no free variables*.

6 A has **full row rank** $\boldsymbol{r = m}$ when its column space $\boldsymbol{C}(A)$ is $\mathbf{R}^m$: $A\boldsymbol{x} = \boldsymbol{b}$ is always solvable.

7 The four cases are $r = m = n$ (A is invertible) and $r = m < n$ (every $A\boldsymbol{x} = \boldsymbol{b}$ is solvable) and $r = n < m$ ($A\boldsymbol{x} = \boldsymbol{b}$ has 1 or 0 solutions) and $r < m, r < n$ (0 or ∞ solutions).

The last section totally solved $A\boldsymbol{x} = \boldsymbol{0}$. Elimination converted the problem to $R\boldsymbol{x} = \boldsymbol{0}$. 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 stayed at 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. $A\boldsymbol{x} = \boldsymbol{b}$ is reduced to a simpler system $R\boldsymbol{x} = \boldsymbol{d}$ with the same solutions. One way to organize that is to ***add*** $\boldsymbol{b}$ ***as an extra column of the matrix***. I will *"augment"* A with the right side $(b_1, b_2, b_3) = (1, 6, 7)$ to produce the **augmented matrix** $\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix}$:

$$\begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{6} \\ \mathbf{7} \end{bmatrix} \quad \begin{matrix}\text{has the} \\ \text{augmented} \\ \text{matrix}\end{matrix} \quad \begin{bmatrix} 1 & 3 & 0 & 2 & \mathbf{1} \\ 0 & 0 & 1 & 4 & \mathbf{6} \\ 1 & 3 & 1 & 6 & \mathbf{7} \end{bmatrix} = \begin{bmatrix} A & \boldsymbol{b} \end{bmatrix}.$$

When we apply the usual elimination steps to A, reaching R, we also apply them to $\boldsymbol{b}$.

In this example we subtract row 1 from row 3. Then we subtract row 2 from row 3. This produces a *row of zeros in* R, and it changes $\boldsymbol{b}$ to a new right side $\boldsymbol{d} = (1, 6, 0)$:

$$\begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{6} \\ \mathbf{0} \end{bmatrix} \quad \begin{matrix}\text{has the} \\ \text{augmented} \\ \text{matrix}\end{matrix} \quad \begin{bmatrix} 1 & 3 & 0 & 2 & \mathbf{1} \\ 0 & 0 & 1 & 4 & \mathbf{6} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix} = \begin{bmatrix} R & \boldsymbol{d} \end{bmatrix}.$$

That very last zero is crucial. The third equation has become $0 = 0$. So 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 of the right side $\boldsymbol{b}$ was $\mathbf{1} + \mathbf{6} = \mathbf{7}$. That led to $0 = 0$.

Here are the same augmented matrices for a general $\boldsymbol{b} = (b_1, b_2, b_3)$:

$$\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} 1 & 3 & 0 & 2 & b_1 \\ 0 & 0 & 1 & 4 & b_2 \\ 1 & 3 & 1 & 6 & b_3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 3 & 0 & 2 & b_1 \\ 0 & 0 & 1 & 4 & b_2 \\ 0 & 0 & 0 & 0 & b_3 - b_1 - b_2 \end{bmatrix} = \begin{bmatrix} R & \boldsymbol{d} \end{bmatrix}$$

Now we get $0 = 0$ in the third equation only if $b_3 - b_1 - b_2 = 0$. This is $b_1 + b_2 = b_3$.

One Particular Solution $A\boldsymbol{x}_p = \boldsymbol{b}$

For an easy solution $\boldsymbol{x}_p$, *choose the free variables to be zero:* $x_2 = x_4 = 0$. Then the two nonzero equations give the two pivot variables $x_1 = 1$ and $x_3 = 6$. Our particular solution to $A\boldsymbol{x} = \boldsymbol{b}$ (and also $R\boldsymbol{x} = \boldsymbol{d}$) is $\boldsymbol{x}_p = (1, 0, 6, 0)$. This particular solution is my favorite: ***free variables* = *zero*, *pivot variables from* d.** The method always works.

For a solution to exist, zero rows in R must also be zero in $\boldsymbol{d}$. Since I is in the pivot rows and pivot columns of R, the pivot variables in $\boldsymbol{x}_{\text{particular}}$ come from $\boldsymbol{d}$:

$$R\boldsymbol{x}_p = \begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix}$$

Pivot variables 1, 6
Free variables $0, 0$
Solution $\boldsymbol{x}_p = (\mathbf{1}, \mathbf{0}, \mathbf{6}, \mathbf{0})$.

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 $\boldsymbol{x}_p$ are already seen in the right side vector $\boldsymbol{d}$.

$\boldsymbol{x}_{\textbf{particular}}$	***The particular solution solves***	$A\boldsymbol{x}_p = \boldsymbol{b}$
$\boldsymbol{x}_{\textbf{nullspace}}$	***The $n - r$ special solutions solve***	$A\boldsymbol{x}_n = \mathbf{0}$.

That particular solution is $(1, 0, 6, 0)$. The two special (nullspace) solutions to $R\boldsymbol{x} = \mathbf{0}$ come from the two free columns of R, by reversing signs of $3, 2$, and 4. ***Please notice how I write the complete solution $\boldsymbol{x}_p + \boldsymbol{x}_n$ to $A\boldsymbol{x} = \boldsymbol{b}$:***

Complete solution
one $\boldsymbol{x}_p$
many $\boldsymbol{x}_n$

$$\boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_n = \begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -2 \\ 0 \\ -4 \\ 1 \end{bmatrix}.$$

Question Suppose A is a square invertible matrix, $m = n = r$. What are $\boldsymbol{x}_p$ and $\boldsymbol{x}_n$?

Answer The particular solution is the one and *only* solution $\boldsymbol{x}_p = A^{-1}\boldsymbol{b}$. There are no special solutions or free variables. $R = I$ has no zero rows. The only vector in the nullspace is $\boldsymbol{x}_n = \mathbf{0}$. The complete solution is $\boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_n = A^{-1}\boldsymbol{b} + \mathbf{0}$.

We didn't mention the nullspace in Chapter 2, because A was invertible and $\boldsymbol{N}(A)$ contained only the zero vector. Reduction went from $\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix}$ to $\begin{bmatrix} I & A^{-1}\boldsymbol{b} \end{bmatrix}$. The matrix A was reduced all the way to I. Then $A\boldsymbol{x} = \boldsymbol{b}$ became $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ which is $\boldsymbol{d}$. 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 $\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix}$ to $\begin{bmatrix} R & \boldsymbol{d} \end{bmatrix}$. For a large matrix, MATLAB does it better. One particular solution (not necessarily ours) is $\boldsymbol{x} = A\backslash\boldsymbol{b}$ 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 $A\boldsymbol{x} = \boldsymbol{b}$ to be solvable, if

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ -2 & -3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

This condition puts $\boldsymbol{b}$ in the column space of A. Find the complete $\boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_n$.

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. Then 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 final column $\boldsymbol{d}$:

$$\textbf{Only solution} \text{ to } A\boldsymbol{x} = \boldsymbol{b} \qquad \boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_n = \begin{bmatrix} 2b_1 - b_2 \\ b_2 - b_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

If $b_3 + b_1 + b_2$ is not zero, there is no solution to $A\boldsymbol{x} = \boldsymbol{b}$ ($\boldsymbol{x}_p$ and $\boldsymbol{x}$ don'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:

$$\textbf{Full column rank} \qquad R = \begin{bmatrix} I \\ 0 \end{bmatrix} = \begin{bmatrix} n \text{ by } n \text{ identity matrix} \\ m - n \text{ rows of zeros} \end{bmatrix} \tag{1}$$

There are no free columns or free variables. The nullspace is $\mathbf{Z} = \{\text{zero vector}\}$.

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 $N(A)$ contains only the zero vector $x = 0$.
4. If $Ax = b$ 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***. $Ax = 0$ only happens when $x = 0$. 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 $Ax = b$ is *unique* (if it exists). There will be $m - n$ zero rows in R. So there are $m - n$ conditions on b in order to have $0 = 0$ in those rows, and b in the column space. With full column rank, $Ax = b$ 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* ($m \leq n$). ***A matrix has full row rank if $r = m$.*** "*The rows are independent.*" Every row has a pivot, and here is an example.

Example 2 This system $Ax = b$ has $n = 3$ unknowns but only $m = 2$ equations:

$$\textbf{Full row rank} \qquad \begin{array}{rcrcrcr} x & + & y & + & z & = & 3 \\ x & + & 2y & - & z & = & 4 \end{array} \qquad (\text{rank } r = m = 2)$$

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 in Figure 3.3***. Then $x = x_p + x_n$ gives the whole line of solutions.

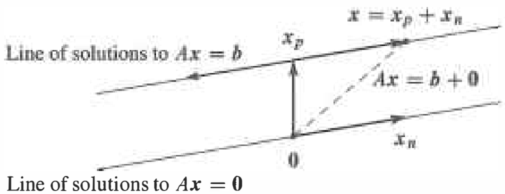


Figure 3.3: Complete solution = *one* particular solution + *all* nullspace solutions.

We find $\boldsymbol{x}_p$ and $\boldsymbol{x}_n$ by elimination on $\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix}$. Subtract row 1 from row 2 and then subtract row 2 from row 1:

$$\begin{bmatrix} 1 & 1 & 1 & \mathbf{3} \\ 1 & 2 & -1 & \mathbf{4} \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 1 & \mathbf{3} \\ 0 & 1 & -2 & \mathbf{1} \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 3 & \mathbf{2} \\ 0 & 1 & -2 & \mathbf{1} \end{bmatrix} = \begin{bmatrix} R & \boldsymbol{d} \end{bmatrix}.$$

The particular solution has free variable $x_3 = 0$. The special solution has $x_3 = 1$:

$\boldsymbol{x}_{\textbf{particular}}$ comes directly from $\boldsymbol{d}$ on the right side: $\boldsymbol{x}_p = (2, 1, 0)$
$\boldsymbol{x}_{\textbf{special}}$ comes from the third column (free column) of R: $\boldsymbol{s} = (-3, 2, 1)$

It is wise to check that $\boldsymbol{x}_p$ and $\boldsymbol{s}$ satisfy the original equations $A\boldsymbol{x}_p = \boldsymbol{b}$ and $A\boldsymbol{s} = \mathbf{0}$:

$$\begin{aligned} 2+1 &= 3 \qquad & -3+2+1 &= 0 \\ 2+2 &= 4 \qquad & -3+4-1 &= 0 \end{aligned}$$

The nullspace solution $\boldsymbol{x}_n$ is any multiple of $\boldsymbol{s}$. It moves along the line of solutions, starting at $\boldsymbol{x}_{\textbf{particular}}$. ***Please notice again how to write the answer***:

Complete solution $\qquad \boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_n = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}.$

This line of solutions 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 needs that constant, and you understand why—to produce all $\boldsymbol{x}_n$ in the nullspace.

Now we summarize this short wide case of *full row rank*. If $m < n$ the equation $A\boldsymbol{x} = \boldsymbol{b}$ is **underdetermined** (many solutions).

Every matrix A with ***full row rank*** $(\boldsymbol{r} = \boldsymbol{m})$ has all these properties:

1. All rows have pivots, and $\boldsymbol{R}$ **has no zero rows.**
2. $A\boldsymbol{x} = \boldsymbol{b}$ has a **solution for every right side** $\boldsymbol{b}$.
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. The nullspace of A^{T} is the zero vector.

We are 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*	$A\boldsymbol{x} = \boldsymbol{b}$	has 1 solution
$\boldsymbol{r = m}$	and	$r < n$	*Short and wide*	$A\boldsymbol{x} = \boldsymbol{b}$	has ∞ solutions
$r < m$	and	$\boldsymbol{r = n}$	*Tall and thin*	$A\boldsymbol{x} = \boldsymbol{b}$	has 0 or 1 solution
$r < m$	and	$r < n$	*Not full rank*	$A\boldsymbol{x} = \boldsymbol{b}$	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 $R\boldsymbol{x} = \boldsymbol{d}$ (and the original $A\boldsymbol{x} = \boldsymbol{b}$) to be solvable, $\boldsymbol{d}$ must end in $m - r$ zeros. F is the free part of R.

Four types for 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.

■ REVIEW OF THE KEY IDEAS ■

1. The rank r is the number of pivots. The matrix R has $m - r$ zero rows.

2. $A\boldsymbol{x} = \boldsymbol{b}$ is solvable if and only if the last $m - r$ equations reduce to $0 = 0$.

3. One particular solution $\boldsymbol{x}_p$ 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 ■

3.3 A This question connects elimination (**pivot columns and back substitution**) to **column space-nullspace-rank-solvability** (the higher level picture). A has rank 2:

$$A\boldsymbol{x} = \boldsymbol{b} \quad \text{is} \quad \begin{aligned} x_1 + 2x_2 + 3x_3 + 5x_4 &= b_1 \\ 2x_1 + 4x_2 + 8x_3 + 12x_4 &= b_2 \\ 3x_1 + 6x_2 + 7x_3 + 13x_4 &= b_3 \end{aligned}$$

1. Reduce $[A \;\; \boldsymbol{b}]$ to $[U \;\; \boldsymbol{c}]$, so that $A\boldsymbol{x} = \boldsymbol{b}$ becomes a triangular system $U\boldsymbol{x} = \boldsymbol{c}$.
2. Find the condition on b_1, b_2, b_3 for $A\boldsymbol{x} = \boldsymbol{b}$ 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. Reduce $[U \;\; c]$ to $[R \;\; \boldsymbol{d}]$: Special solutions from R, particular solution from $\boldsymbol{d}$.
6. Find a particular solution to $A\boldsymbol{x} = (0, 6, -6)$ and then the complete solution.

Solution

1. The multipliers in elimination are 2 and 3 and -1. They take $[A \;\; \boldsymbol{b}]$ into $[U \;\; \boldsymbol{c}]$.

$$\begin{bmatrix} 1 & 2 & 3 & 5 & b_1 \\ 2 & 4 & 8 & 12 & b_2 \\ 3 & 6 & 7 & 13 & b_3 \end{bmatrix} \rightarrow \left[\begin{array}{cccc|c} 1 & 2 & 3 & 5 & b_1 \\ 0 & 0 & 2 & 2 & b_2 - 2b_1 \\ 0 & 0 & -2 & -2 & b_3 - 3b_1 \end{array}\right] \rightarrow \left[\begin{array}{cccc|c} 1 & 2 & 3 & 5 & b_1 \\ 0 & 0 & 2 & 2 & b_2 - 2b_1 \\ 0 & 0 & 0 & 0 & b_3 + b_2 - 5b_1 \end{array}\right]$$

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 $A\boldsymbol{x} = \boldsymbol{b}$ 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 $U\boldsymbol{x} = \boldsymbol{0}$
or change signs of 2, 2, 1 in R

$$\boldsymbol{s}_1 = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad \boldsymbol{s}_2 = \begin{bmatrix} -2 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

The nullspace $\boldsymbol{N}(A)$ in $\mathbf{R}^4$ contains all $\boldsymbol{x}_n = c_1\boldsymbol{s}_1 + c_2\boldsymbol{s}_2$.

5. 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 $\boldsymbol{x}_p$:

$$[U \;\; \boldsymbol{c}] = \begin{bmatrix} 1 & 2 & 3 & 5 & \mathbf{0} \\ 0 & 0 & 2 & 2 & \mathbf{6} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix} \longrightarrow [R \;\; \boldsymbol{d}] = \begin{bmatrix} 1 & 2 & 0 & 2 & \mathbf{-9} \\ 0 & 0 & 1 & 1 & \mathbf{3} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix}$$

6. One particular solution $\boldsymbol{x}_p$ has free variables = zero. Back substitute in $U\boldsymbol{x} = \boldsymbol{c}$:

Particular solution to $A\boldsymbol{x}_p = \boldsymbol{b}$
Bring -9 and 3 from the vector $\boldsymbol{d}$
Free variables x_2 and x_4 are zero

$$\boldsymbol{x}_p = \begin{bmatrix} -9 \\ 0 \\ 3 \\ 0 \end{bmatrix}$$

The complete solution to $A\boldsymbol{x} = (0, 6, -6)$ is $\boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_n = \boldsymbol{x}_p + c_1\boldsymbol{s}_1 + c_2\boldsymbol{s}_2$.

3.3 B Suppose you have this information about the solutions to $Ax = \boldsymbol{b}$ for a specific $\boldsymbol{b}$. What does that tell you about m and n and r (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 $\boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.
3. There are no solutions.
4. All solutions to $A\boldsymbol{x} = \boldsymbol{b}$ have the form $\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + c\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$
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 $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ in the nullspace of A, column 2 is the *negative* of column 1. Also $A \neq 0$: the rank is 1. With $\boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ 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 $r = m$.

3.3 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 $y_1 b_1 + y_2 b_2 + y_3 b_3 = 0$. 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 or we continue to R.

$R\boldsymbol{x} = \boldsymbol{d}$ shows that the particular solution with free variables $= 0$ is $\boldsymbol{x_p} = (7, 0, -3, 0)$.

$$\begin{bmatrix} 1 & 2 & 1 & 0 & \mathbf{4} \\ 0 & 0 & 2 & 8 & \mathbf{-6} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 1 & 0 & \mathbf{4} \\ 0 & 0 & 1 & 4 & \mathbf{-3} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 0 & -4 & \mathbf{7} \\ 0 & 0 & 1 & 4 & \mathbf{-3} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix}.$$

For the nullspace part $\boldsymbol{x_n}$ with $\boldsymbol{b} = \mathbf{0}$, set the free variables x_2, x_4 to $1, 0$ and also $0, 1$:

Special solutions $\quad \boldsymbol{s}_1 = (-2, 1, 0, 0)$ **and** $\boldsymbol{s}_2 = (4, 0, -4, 1)$

Then the complete solution to $A\boldsymbol{x} = \boldsymbol{b}$ (and $R\boldsymbol{x} = \boldsymbol{d}$) is $\boldsymbol{x}_{\text{complete}} = \boldsymbol{x_p} + c_1\boldsymbol{s}_1 + c_2\boldsymbol{s}_2$.

The rows of A produced the zero row from $2(\text{row } 1) + (\text{row } 2) - (\text{row } 3) = (0, 0, 0, 0)$. 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 3.3

1 (Recommended) Execute the six steps of Worked Example **3.3 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 $\boldsymbol{x}_n$. Start from 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?

$$\text{(a)}\ A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 6 & 3 \\ 0 & 2 & 5 \end{bmatrix} \qquad \text{(b)}\quad A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 2 & 4 & 8 \end{bmatrix}.$$

9 (a) The Worked Example **3.3 A** reached $[\,U \;\; \boldsymbol{c}\,]$ from $[\,A \;\; \boldsymbol{b}\,]$. Put the multipliers into L and verify that LU equals A and $L\boldsymbol{c}$ 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 $Ax = b$ has two solutions x_1 and x_2, find two solutions to $Ax = 0$.

(b) Then find another solution to $Ax = 0$ and another solution to $Ax = b$.

13 Explain why these are all false:

(a) The complete solution is any linear combination of x_p and x_n.

(b) A system $Ax = b$ has at most one particular solution.

(c) The solution x_p with all free variables zero is the shortest solution (minimum length $\|x\|$). Find a 2 by 2 counterexample.

(d) If A is invertible there is no solution 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 $Ax = 0$. If $Ax = b$ has a solution, then it has _____ solutions.

15 Suppose row 3 of U has no pivot. Then that row is _____. The equation $Ux = c$ is only solvable provided _____. The equation $Ax = b$ (*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 $Ax = b$ (*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}:

$$A = \begin{bmatrix} 1 & 4 & 0 \\ 2 & 11 & 5 \\ -1 & 2 & 10 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 1 & 1 & q \end{bmatrix} \quad \text{(rank depends on } q\text{)}.$$

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 $\boldsymbol{x}_p + \boldsymbol{x}_n$ 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 $A\boldsymbol{x} = \boldsymbol{b}$ has infinitely many solutions, why is it impossible for $A\boldsymbol{x} = \boldsymbol{B}$ (new right side) to have only one solution? Could $A\boldsymbol{x} = \boldsymbol{B}$ 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 $A\boldsymbol{x} = \boldsymbol{b}$ is

(a) 0 or 1, depending on $\boldsymbol{b}$

(b) ∞, regardless of $\boldsymbol{b}$

(c) 0 or ∞, depending on $\boldsymbol{b}$

(d) 1, regardless of $\boldsymbol{b}$.

25 Write down all known relations between r and m and n if $A\boldsymbol{x} = \boldsymbol{b}$ has

(a) no solution for some $\boldsymbol{b}$

(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 If A is a triangular matrix, when is $R = \mathrm{rref}(A)$ equal to 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 $R\boldsymbol{x} = \boldsymbol{d}$:

$$\begin{bmatrix} U & \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & \boldsymbol{0} \\ 0 & 0 & 4 & \boldsymbol{0} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} U & \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & \boldsymbol{5} \\ 0 & 0 & 4 & \boldsymbol{8} \end{bmatrix}.$$

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 $R\boldsymbol{x} = \mathbf{0}$ and $R\boldsymbol{x} = \boldsymbol{d}$:

$$\begin{bmatrix} U & 0 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 6 & \mathbf{0} \\ 0 & 0 & 2 & \mathbf{0} \\ 0 & 0 & 0 & \mathbf{0} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} U & c \end{bmatrix} = \begin{bmatrix} 3 & 0 & 6 & \mathbf{9} \\ 0 & 0 & 2 & \mathbf{4} \\ 0 & 0 & 0 & \mathbf{5} \end{bmatrix}.$$

Solve $U\boldsymbol{x} = \mathbf{0}$ or $R\boldsymbol{x} = \mathbf{0}$ 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 $A\boldsymbol{x} = \boldsymbol{b}$:

$$A = \begin{bmatrix} 1 & 3 & 1 \\ 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 1 & 5 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 3 \\ 6 \\ 5 \end{bmatrix} \quad \text{and then} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

33 The complete solution to $A\boldsymbol{x} = \begin{bmatrix} \mathbf{1} \\ \mathbf{3} \end{bmatrix}$ is $\boldsymbol{x} = \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix} + c \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}$. Find A.

Challenge Problems

34 (Recommended!) Suppose you know that the 3 by 4 matrix A has the vector $\boldsymbol{s} = (2, 3, 1, 0)$ as the only special solution to $A\boldsymbol{x} = \boldsymbol{0}$.

(a) What is the *rank* of A and the complete solution to $A\boldsymbol{x} = \boldsymbol{0}$?

(b) What is the exact row reduced echelon form R of A?

(c) How do you know that $A\boldsymbol{x} = \boldsymbol{b}$ can be solved for all $\boldsymbol{b}$?

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 $K\boldsymbol{x} = \boldsymbol{b} = (10, \ldots, 10)$. 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 $A\boldsymbol{x} = \boldsymbol{b}$ and $C\boldsymbol{x} = \boldsymbol{b}$ have the same (complete) solutions for every $\boldsymbol{b}$. Is it true that A equals C?

37 Describe the column space of a reduced row echelon matrix R.

3.4 Independence, Basis and Dimension

1 **Independent columns** of A: The only solution to $A\boldsymbol{x} = \mathbf{0}$ is $\boldsymbol{x} = \mathbf{0}$. The nullspace is $\boldsymbol{Z}$.

2 Independent vectors: The only zero combination $c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k = 0$ has all c's $= 0$.

3 A matrix with $m < n$ has **dependent columns**: At least $n-m$ free variables / special solutions.

4 The vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ **span the space** $\boldsymbol{S}$ if $\boldsymbol{S}$ = all combinations of the $\boldsymbol{v}$'s.

5 The vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ are a **basis for** $\boldsymbol{S}$ if they are independent and they span $\boldsymbol{S}$.

6 The **dimension of a space** $\boldsymbol{S}$ is the number of vectors in every basis for $\boldsymbol{S}$.

7 If A is 4 by 4 and invertible, its columns are a basis for $\mathbf{R}^4$. The dimension of $\mathbf{R}^4$ is 4.

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 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ 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 $A\boldsymbol{x} = \mathbf{0}$ is $\boldsymbol{x} = \mathbf{0}$. ***No other combination*** $A\boldsymbol{x}$ ***of the columns gives the zero vector.***

The columns are independent when the nullspace $\boldsymbol{N}(A)$ 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 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ in Figure 3.4 gives zero except $0\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + 0\boldsymbol{v}_3$.

2. If three vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3$ are *in the same plane*, they are dependent.

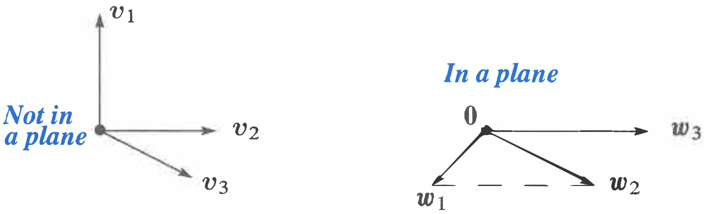


Figure 3.4: Independent vectors $\boldsymbol{v}_1,\ \boldsymbol{v}_2,\ \boldsymbol{v}_3$. Only $0\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + 0\boldsymbol{v}_3$ gives the vector $\mathbf{0}$. Dependent vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3$. The combination $\boldsymbol{w}_1 - \boldsymbol{w}_2 + \boldsymbol{w}_3$ is $(0,0,0)$.

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 $\boldsymbol{x} = \mathbf{0}$. 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.

DEFINITION The sequence of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ is ***linearly independent*** if the only combination that gives the zero vector is $0\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + \cdots + 0\boldsymbol{v}_n$.

Linear independence

$$x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_n\boldsymbol{v}_n = \mathbf{0} \quad \text{only happens when all } x\text{'s are zero.} \tag{1}$$

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 $A\boldsymbol{x} = \mathbf{0}$:

$$\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{for } x_1 = 1 \text{ and } x_2 = 1.$$

The columns are dependent exactly when *there is a nonzero vector in the nullspace*.

If one of the $\boldsymbol{v}$'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 $A\boldsymbol{x} = \mathbf{0}$. 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 $A\boldsymbol{x} = \mathbf{0}$.

Example 1 The columns of this A are dependent. $A\boldsymbol{x} = \mathbf{0}$ has a nonzero solution:

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 3 \\ 2 & 1 & 5 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \quad \text{is} \quad -3 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 3 \\ 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

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 $A\boldsymbol{x} = \mathbf{0}$? 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 $\boldsymbol{x} = (-3, 1, 1)$ 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 $\boldsymbol{x} = \mathbf{0}$ 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. $A\boldsymbol{x} = \mathbf{0}$ 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 $A\boldsymbol{x} = \mathbf{0}$ has a nonzero solution.

The columns might be dependent or might be independent if $n \le m$. 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 $\boldsymbol{v}$ in the space is a combination of the basis vectors, because they span the space. More than that, the combination that produces $\boldsymbol{v}$ is *unique*, because the basis vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are independent:

There is one and only one way to write $\boldsymbol{v}$ as a combination of the basis vectors.

Reason: Suppose $\boldsymbol{v} = a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n$ and also $\boldsymbol{v} = b_1\boldsymbol{v}_1 + \cdots + b_n\boldsymbol{v}_n$. By subtraction $(a_1 - b_1)\boldsymbol{v}_1 + \cdots + (a_n - b_n)\boldsymbol{v}_n$ is the zero vector. From the independence of the $\boldsymbol{v}$'s, each $a_i - b_i = 0$. Hence $a_i = b_i$, and there are not two ways to produce $\boldsymbol{v}$.

Example 6 The columns of $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ produce the "standard basis" for $\mathbf{R}^2$.

$$\text{The basis vectors} \quad \boldsymbol{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{are independent. They span } \mathbf{R}^2.$$

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 $\mathbf{R}^n$:

Invertible matrix
Independent columns
Column space is $\mathbf{R}^3$
$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Singular matrix
Dependent columns
Column space $\neq \mathbf{R}^3$
$$B = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 1 & 1 & 2 \end{bmatrix}.$$

The only solution to $A\boldsymbol{x} = \mathbf{0}$ is $\boldsymbol{x} = A^{-1}\mathbf{0} = \mathbf{0}$. The columns are independent. They span the whole space $\mathbf{R}^n$—because every vector $\boldsymbol{b}$ is a combination of the columns. $A\boldsymbol{x} = \boldsymbol{b}$ can always be solved by $\boldsymbol{x} = A^{-1}\boldsymbol{b}$. Do you see how everything comes together for invertible matrices? Here it is in one sentence:

The vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ 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!

One pivot column
One pivot row $(r = 1)$
$$A = \begin{bmatrix} 2 & 4 \\ 3 & 6 \end{bmatrix} \quad \text{reduces to} \quad R = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}.$$

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). Those pivot columns are a basis for the column space.

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 w is a combination of the v's

$$W = \begin{bmatrix} w_1 & w_2 & \ldots & w_n \end{bmatrix} = \begin{bmatrix} v_1 & \ldots & v_m \end{bmatrix} \begin{bmatrix} a_{11} & & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & & a_{mn} \end{bmatrix} = VA.$$

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 $A\boldsymbol{x} = \mathbf{0}$ ***has a nonzero solution***.

$A\boldsymbol{x} = 0$ gives $VA\boldsymbol{x} = \mathbf{0}$ which is $W\boldsymbol{x} = \mathbf{0}$. *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 $\boldsymbol{v} = (1, 5, 2)$ has dimension one. It is a subspace with this one vector $\boldsymbol{v}$ 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 $A\boldsymbol{x} = \mathbf{0}$. 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:

$$\begin{array}{l}\textbf{Every } \boldsymbol{A} \textbf{ combines} \\ \textbf{the basis matrices}\end{array} \qquad c_1A_1 + c_2A_2 + c_3A_3 + c_4A_4 = \begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix} = A.$$

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 $A\boldsymbol{x} = \boldsymbol{b}$. 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 ■

3.4 A Start with the vectors $\boldsymbol{v}_1 = (1, 2, 0)$ and $\boldsymbol{v}_2 = (2, 3, 0)$. **(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 $\boldsymbol{v}_3$ that complete a basis $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ for $\mathbf{R}^3$.

Solution

(a) $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are independent—the only combination to give $\mathbf{0}$ is $0\boldsymbol{v}_1 + 0\boldsymbol{v}_2$.

(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 $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. In particular A could just have columns $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$.

(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 $B\boldsymbol{v}_1 = \mathbf{0}$ and $B\boldsymbol{v}_2 = \mathbf{0}$.

(g) Any third vector $\boldsymbol{v}_3 = (a, b, c)$ will complete a basis for $\mathbf{R}^3$ provided $c \neq 0$.

3.4 B Start with three independent vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3$. Take combinations of those vectors to produce $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$. Write the combinations in matrix form as $V = WB$:

$$\begin{aligned} \boldsymbol{v}_1 &= \boldsymbol{w}_1 + \boldsymbol{w}_2 \\ \boldsymbol{v}_2 &= \boldsymbol{w}_1 + 2\boldsymbol{w}_2 + \boldsymbol{w}_3 \\ \boldsymbol{v}_3 &= \boldsymbol{w}_2 + c\boldsymbol{w}_3 \end{aligned} \quad \text{which is} \quad \begin{bmatrix} \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \boldsymbol{w}_1 & \boldsymbol{w}_2 & \boldsymbol{w}_3 \\ \\ \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & c \end{bmatrix}$$

What is the test on B to see if $V = WB$ has independent columns? If $c \neq 1$ show that $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ are linearly independent. If $c = 1$ show that the $\boldsymbol{v}$'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 $\boldsymbol{x} = (0, 0, 0)$ is the only combination of the columns that gives $V\boldsymbol{x} =$ zero vector.

If $c = 1$ in our problem, we can see *dependence* in two ways. First, $\boldsymbol{v}_1 + \boldsymbol{v}_3$ will be the same as $\boldsymbol{v}_2$. (If you add $\boldsymbol{w}_1 + \boldsymbol{w}_2$ to $\boldsymbol{w}_2 + \boldsymbol{w}_3$ you get $\boldsymbol{w}_1 + 2\boldsymbol{w}_2 + \boldsymbol{w}_3$ which is $\boldsymbol{v}_2$.) In other words $\boldsymbol{v}_1 - \boldsymbol{v}_2 + \boldsymbol{v}_3 = \mathbf{0}$—which says that the $\boldsymbol{v}$'s are not independent.

The other way is to look at the nullspace of B. If $c = 1$, the vector $\boldsymbol{x} = (1, -1, 1)$ is in that nullspace, and $B\boldsymbol{x} = \mathbf{0}$. Then certainly $WB\boldsymbol{x} = \mathbf{0}$ which is the same as $V\boldsymbol{x} = 0$. So the $\boldsymbol{v}$'s are dependent. This specific $\boldsymbol{x} = (1, -1, 1)$ from the nullspace tells us again that $\boldsymbol{v}_1 - \boldsymbol{v}_2 + \boldsymbol{v}_3 = \mathbf{0}$.

Now suppose $c \neq 1$. Then the matrix B is invertible. So if $\boldsymbol{x}$ is *any nonzero vector* we know that $B\boldsymbol{x}$ is nonzero. Since the $\boldsymbol{w}$'s are given as independent, we further know that $WB\boldsymbol{x}$ is nonzero. Since $V = WB$, 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 B 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 B is invertible."***

3.4 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 $c_1A\boldsymbol{v}_1 + \cdots + c_nA\boldsymbol{v}_n = \mathbf{0}$. 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 $c_1A\boldsymbol{v}_1 + \cdots + c_nA\boldsymbol{v}_n = \boldsymbol{b}$ 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 3.4

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 $U\boldsymbol{x} = \mathbf{0}$ is $\boldsymbol{x} = \mathbf{0}$. 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 11–14 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 $U = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}$. 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 $A\boldsymbol{x} = \mathbf{0}$ has only the solution $\boldsymbol{x} = \mathbf{0}$ 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 $\mathbf{S}$ is a 5-dimensional subspace of $\mathbf{R}^6$. True or false (example if false):

(a) Every basis for $\mathbf{S}$ 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 $\mathbf{S}$ 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 $c_1P_1 + \cdots + c_5P_5 =$ zero matrix, and check entries to prove that c_1 to c_5 must all be 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 Counting Theorem.

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 3.5, 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). So A multiplies each column of A to give the zero vector. 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 $r \leq 5$.

3.5 Dimensions of the Four Subspaces

1 The column space $\boldsymbol{C}(A)$ and the row space $\boldsymbol{C}(A^{\mathrm{T}})$ both have *dimension* r (the rank of A).

2 The nullspace $\boldsymbol{N}(A)$ has *dimension* $n-r$. The left nullspace $\boldsymbol{N}(A^{\mathrm{T}})$ has *dimension* $m-r$.

3 Elimination produces bases for the row space and nullspace of A: They are the same as for R.

4 Elimination often changes the column space and left nullspace (but dimensions don't change).

5 **Rank one matrices**: $A = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ = column times row : $\boldsymbol{C}(A)$ has basis $\boldsymbol{u}$, $\boldsymbol{C}(A^{\mathrm{T}})$ has basis $\boldsymbol{v}$.

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 $\boldsymbol{C}(A^{\mathrm{T}})$, a subspace of $\mathbf{R}^n$.
2. The ***column space*** is $\boldsymbol{C}(A)$, a subspace of $\mathbf{R}^m$.
3. The ***nullspace*** is $\boldsymbol{N}(A)$, a subspace of $\mathbf{R}^n$.
4. The ***left nullspace*** is $\boldsymbol{N}(A^{\mathrm{T}})$, a subspace of $\mathbf{R}^m$. This is our new space.

In this book the column space and nullspace came first. We know $\boldsymbol{C}(A)$ and $\boldsymbol{N}(A)$ pretty well. Now the other two subspaces come forward. The row space contains all combinations of the rows. *This row space of A 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 $\boldsymbol{y}^{\mathrm{T}}A = \mathbf{0}^{\mathrm{T}}$. 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. This number r is the **rank** of the matrix. The other important fact involves the two nullspaces:

$\boldsymbol{N}(A)$ ***and*** $\boldsymbol{N}(A^{\mathrm{T}})$ ***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 this echelon matrix R:

$$\begin{matrix} m=3 \\ n=5 \\ r=2 \end{matrix} \qquad R = \begin{bmatrix} \mathbf{1} & \mathbf{3} & \mathbf{5} & 0 & \mathbf{7} \\ 0 & 0 & 0 & \mathbf{1} & \mathbf{2} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{matrix} \textbf{pivot rows } 1 \textbf{ and } 2 \\ \\ \textbf{pivot columns } 1 \textbf{ and } 4 \end{matrix}$$

The rank of this matrix 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*** of R 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 $R\boldsymbol{x} = \mathbf{0}$. Set a free variable to 1, and solve for x_1 and x_4.

$$\boldsymbol{s}_2 = \begin{bmatrix} \mathbf{-3} \\ \mathbf{1} \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \boldsymbol{s}_3 = \begin{bmatrix} \mathbf{-5} \\ 0 \\ \mathbf{1} \\ 0 \\ 0 \end{bmatrix} \quad \boldsymbol{s}_5 = \begin{bmatrix} \mathbf{-7} \\ 0 \\ 0 \\ \mathbf{-2} \\ \mathbf{1} \end{bmatrix}$$

$R\boldsymbol{x} = \mathbf{0}$ has the complete solution
$\boldsymbol{x} = x_2\boldsymbol{s}_2 + x_3\boldsymbol{s}_3 + x_5\boldsymbol{s}_5$
The nullspace has dimension 3.

Reason: There is a special solution for each free variable. With n variables and r pivots, that leaves $n-r$ free variables and special solutions. The special solutions are independent, because they contain the identity matrix in rows 2, 3, 5. So $\boldsymbol{N}(R)$ has dimension $n - r$.

The nullspace has dimension $n - r$. The special solutions form a basis.

4. The ***nullspace of $\boldsymbol{R}^{\mathrm{T}}$*** (***left nullspace of $\boldsymbol{R}$***) has dimension $m - r = 3 - 2$.

Reason: The equation $R^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$ looks for combinations of the columns of R^{T} (*the rows of* R) that produce zero. This equation $R^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$ or $\boldsymbol{y}^{\mathrm{T}}R = \mathbf{0}^{\mathrm{T}}$ is

$$\begin{array}{ll} \textbf{Left nullspace} & \quad y_1\,[1,\ 3,\ 5,\ 0,\ 7] \\ \textbf{Combination} & +y_2\,[0,\ 0,\ 0,\ 1,\ 2] \\ \textbf{of rows is zero} & \underline{+y_3\,[0,\ 0,\ 0,\ 0,\ 0]} \\ & \qquad [0,\ 0,\ 0,\ 0,\ 0] \end{array} \tag{1}$$

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 $\boldsymbol{R}^{\mathrm{T}}$ contains all vectors $\boldsymbol{y}$** $= (0, 0, y_3)$.

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. So $\boldsymbol{y}$ in the left nullspace has $y_1 = 0, \ldots, y_r = 0$.

If A is m by n of rank r, its left nullspace has dimension $m - r$.

Why is this a "*left* nullspace"? The reason is that $R^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$ can be transposed to $\boldsymbol{y}^{\mathrm{T}}R = \mathbf{0}^{\mathrm{T}}$. Now $\boldsymbol{y}^{\mathrm{T}}$ 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).

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 $R = \text{rref}(A)$.

$$\textbf{This } A \textbf{ reduces to } R \qquad A = \begin{bmatrix} 1 & 3 & 5 & 0 & 7 \\ 0 & 0 & 0 & 1 & 2 \\ 1 & 3 & 5 & 1 & 9 \end{bmatrix} \qquad \text{Notice } \boldsymbol{C}(A) \neq \boldsymbol{C}(R)\,! \tag{2}$$

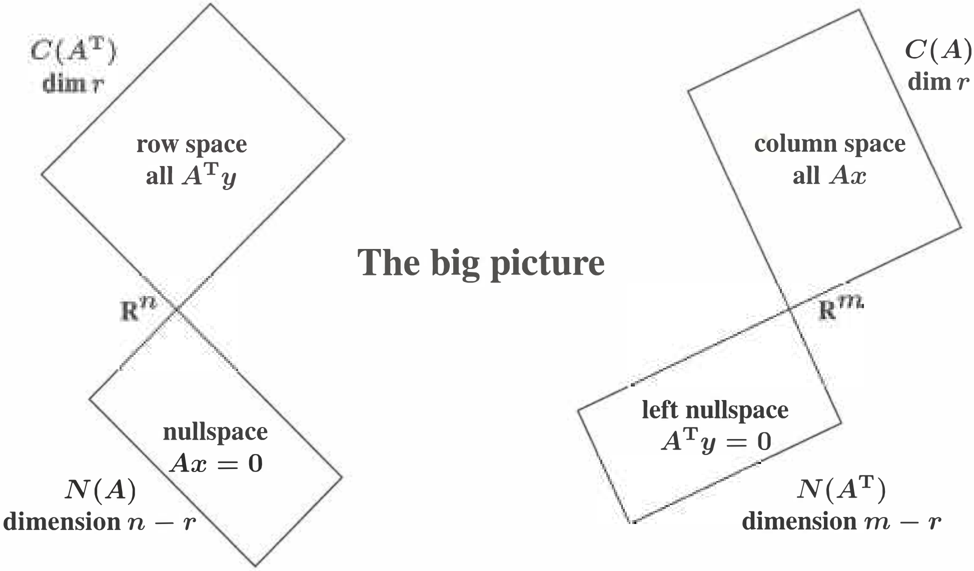


Figure 3.5: The dimensions of the Four Fundamental Subspaces (for R and for A).

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. The column rank equals the row rank.***

Rank Theorem: The number of independent columns = 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. Then $\boldsymbol{C}(A)$ is not $\boldsymbol{C}(R)$.

Right reason: The ***same*** combinations of the columns are zero (or nonzero) for A and R. Dependent in $A \Leftrightarrow$ dependent in R. Say that another way: $A\boldsymbol{x} = \mathbf{0}$ *exactly when* $R\boldsymbol{x} = \mathbf{0}$. The column spaces are different, but their *dimensions* are the same—equal to r.

Conclusion The r pivot columns of A are a basis for *its* column space $\boldsymbol{C}(A)$.

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 $n - r$. This is the **Counting Theorem**: $r + (n - r)$ 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 a big 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

$$\text{dimension of } \boldsymbol{C}(A) = \text{dimension of } \boldsymbol{C}(A^{\mathrm{T}}) = \text{rank of } A$$
$$\text{dimension of } \boldsymbol{C}(A) + \text{dimension of } \boldsymbol{N}(A) = 17.$$

Example 1 $A = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$ 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 $\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$ gives the zero row. Thus $\boldsymbol{N}(A^{\mathrm{T}})$ is $\mathbf{Z}$, the zero space with dimension 0 (which is $m - r$). In $\mathbf{R}^m$ the dimensions of $\boldsymbol{C}(A)$ and $\boldsymbol{N}(A^{\mathrm{T}})$ 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 $x_1 + 2x_2 + 3x_3 = 0$. The line and plane 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?

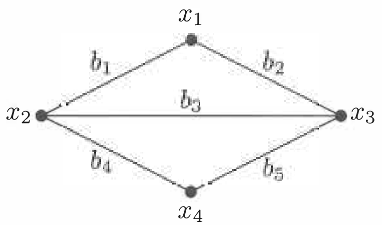
The y's in the left nullspace combine the rows to give the zero row.

Example 3 You have nearly finished three chapters with made-up equations, and this can't continue forever. Here is a better example of five equations (one for every edge in Figure 3.6). The five equations have four unknowns (one for every node). The matrix in $A\boldsymbol{x} = \boldsymbol{b}$ is an **incidence matrix**. This matrix A has 1 and -1 on every row.

Differences $A\boldsymbol{x} = \boldsymbol{b}$
across edges 1, 2, 3, 4, 5
between nodes 1, 2, 3, 4

$$\begin{array}{rrrrl} -x_1 & +x_2 & & & = b_1 \\ -x_1 & & +x_3 & & = b_2 \\ & -x_2 & +x_3 & & = b_3 \\ & -x_2 & & +x_4 & = b_4 \\ & & -x_3 & +x_4 & = b_5 \end{array} \qquad (3)$$

If you understand the four fundamental subspaces for this matrix (*the column spaces and the nullspaces for A and A^{T}*) you have captured the central ideas of linear algebra.



$$\boldsymbol{A} = \begin{bmatrix} -1 & 1 & & \\ -1 & & 1 & \\ & -1 & 1 & \\ & -1 & & 1 \\ & & -1 & 1 \end{bmatrix} \quad \begin{array}{c} \text{edges} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}$$

Figure 3.6: A "graph" with 5 edges and 4 nodes. A is its 5 by 4 incidence matrix.

The nullspace $\boldsymbol{N}(A)$ To find the nullspace we set $\boldsymbol{b} = \mathbf{0}$. Then the first equation says $x_1 = x_2$. The second equation is $x_3 = x_1$. Equation 4 is $x_2 = x_4$. *All four unknowns x_1, x_2, x_3, x_4 have the same value c.* The vectors $\boldsymbol{x} = (c,c,c,c)$ fill the nullspace of A.

That nullspace is a line in $\mathbf{R}^4$. The special solution $\boldsymbol{x} = (1,1,1,1)$ is a basis for $\boldsymbol{N}(A)$. The dimension of $\boldsymbol{N}(A)$ is 1 (one vector in the basis). *The rank of A must be* 3, *since* $n - r = 4 - 3 = 1$. We now know the dimensions of all four subspaces.

The column space $C(A)$ There must be $r = 3$ independent columns. The fast way is to look at the first 3 columns. The systematic way is to find $R = \text{rref}(A)$.

$$\begin{array}{c} \textbf{Columns} \\ \mathbf{1,2,3} \\ \textbf{of } \boldsymbol{A} \end{array} \quad \begin{matrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{matrix} \qquad R = \begin{array}{c} \textbf{reduced row} \\ \textbf{echelon form} \end{array} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

From R we see again the special solution $\boldsymbol{x} = (1,1,1,1)$. The first 3 columns are basic, the fourth column is free. To produce a basis for $\boldsymbol{C}(A)$ and not $\boldsymbol{C}(R)$, we go back to columns $1, 2, 3$ of A. The column space has dimension $r = 3$.

The row space $C(A^{\mathrm{T}})$ The dimension must again be $r = 3$. But the first 3 rows of A are *not independent*: row 3 $=$ row 2 $-$ row 1. So row 3 became zero in elimination, and row 3 was exchanged with row 4. *The first three independent rows are rows* $1, 2, 4$. Those three rows are a basis (one possible basis) for the row space.

I notice that edges $1, 2, 3$ form a **loop** in the picture: Dependent rows $1, 2, 3$. Edges $1, 2, 4$ form a **tree** in the picture. **Trees have no loops**! Independent rows $1, 2, 4$.

The left nullspace $N(A^{\mathrm{T}})$ Now we solve $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$. Combinations of the rows give zero. We already noticed that row 3 $=$ row 2 $-$ row 1, so one solution is $\boldsymbol{y} = (1, -1, 1, 0, 0)$. I would say: That $\boldsymbol{y}$ comes from following the upper loop in the picture. Another $\boldsymbol{y}$ comes from going around the lower loop and it is $\boldsymbol{y} = (0, 0, \mathbf{-1}, \mathbf{1}, \mathbf{-1})$: *row* 3 $=$ *row* 4 $-$ *row* 5. Those two $\boldsymbol{y}$'s are independent, they solve $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$, and the dimension of $\boldsymbol{N}(A^{\mathrm{T}})$ is $\boldsymbol{m} - \boldsymbol{r} = 5 - 3 = \mathbf{2}$. So we have a basis for the left nullspace.

You may ask how "loops" and "trees" got into this problem. That didn't have to happen. We could have used elimination to solve $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$. The 4 by 5 matrix A^{T} would have three pivot columns $1, 2, 4$ and two free columns $3, 5$. There are two special solutions and the nullspace of A^{T} has dimension two: $m - r = 5 - 3 = 2$. But *loops* and *trees* identify *dependent rows* and *independent rows* in a beautiful way. We use them in Section 10.1 for every incidence matrix like this A.

The equations $A\boldsymbol{x} = \boldsymbol{b}$ give "voltages" x_1, x_2, x_3, x_4 at the four nodes. The equations $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$ give "currents" y_1, y_2, y_3, y_4, y_5 on the five edges. These two equations are **Kirchhoff's Voltage Law** and **Kirchhoff's Current Law**. Those words apply to an electrical network. But the ideas behind the words apply all over engineering and science and economics and business.

Graphs are *the most important model in discrete applied mathematics*. You see graphs everywhere: roads, pipelines, blood flow, the brain, the Web, the economy of a country or the world. We can understand their matrices A and A^{T}.

Rank One Matrices (Review)

Suppose every row is a multiple of the first row. Here is a typical example:

$$\begin{bmatrix} 2 & 3 & 7 & 8 \\ 2a & 3a & 7a & 8a \\ 2b & 3b & 7b & 8b \end{bmatrix} = \begin{bmatrix} 1 \\ a \\ b \end{bmatrix} \begin{bmatrix} 2 & 3 & 7 & 8 \end{bmatrix} = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$$

On the left is a matrix with three rows. But its row *space* only has dimension $= 1$. The row vector $\boldsymbol{v}^{\mathrm{T}} = \begin{bmatrix} 2 & 3 & 7 & 8 \end{bmatrix}$ tells us a basis for that row space. *The row rank is* 1.

Now look at the columns. "The column rank equals the row rank which is 1." All columns of the matrix must be multiples of one column. Do you see that this key rule of linear algebra is true? The column vector $\boldsymbol{u} = (1, a, b)$ is multiplied by $2, 3, 7, 8$. That nonzero vector $\boldsymbol{u}$ is a basis for the column space. *The column rank is also* 1.

Every rank one matrix is one column times one row $\quad \boxed{\boldsymbol{A = uv^{\mathrm{T}}}}$

Rank Two Matrices = Rank One plus Rank One

Here is a matrix A of rank $r = 2$. We can't see r immediately from A. So we reduce the matrix by row operations to $R = \mathrm{rref}(A)$. Some elimination matrix E simplifies A to $\boldsymbol{EA = R}$. Then the inverse matrix $C = E^{-1}$ connects R back to $\boldsymbol{A = CR}$.

You know the main point already: $\boldsymbol{R}$ **has the same row space as** $\boldsymbol{A}$.

Rank two
$$A = \begin{bmatrix} 1 & 0 & 3 \\ 1 & 1 & 7 \\ 4 & 2 & 20 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 4 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix} = CR. \qquad (4)$$

The row space of R clearly has two basis vectors $\boldsymbol{v}_1^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 3 \end{bmatrix}$ and $\boldsymbol{v}_2^{\mathrm{T}} = \begin{bmatrix} 0 & 1 & 4 \end{bmatrix}$. So the (same!) row space of A also has this basis: *row rank* $= 2$. Multiplying C times R says that row 3 of A is $4\boldsymbol{v}_1^{\mathrm{T}} + 2\boldsymbol{v}_2^{\mathrm{T}}$.

Now look at columns. The pivot columns of R are clearly $(1, 0, 0)$ and $(0, 1, 0)$. Then the pivot columns of A are also in columns 1 and 2: $\boldsymbol{u}_1 = (1, 1, 4)$ and $\boldsymbol{u}_2 = (0, 1, 2)$. Notice that C has those same first two columns! That was guaranteed since multiplying by two columns of the identity matrix (in R) won't change the pivot columns $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$.

When you put in letters for the columns and rows, you see **rank 2 = rank 1 + rank 1**.

Matrix A Rank two
$$A = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \boldsymbol{u}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \\ \textbf{zero row} \end{bmatrix} = \boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}} = (\text{rank } 1) + (\text{rank } 1).$$

Did you see that last step? I multiplied the matrices using **columns times rows**. That was perfect for this problem. *Every rank r matrix is a sum of r rank one matrices*: Pivot columns of A times nonzero rows of R. The row $\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$ simply disappeared.

The pivot columns $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are a basis for the column space, which you knew.

■ 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 $\boldsymbol{C}(A)$.
3. The $n - r$ special solutions are a basis for the nullspaces of A and R (same space).
4. If $EA = R$, 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 10 (often I go to 10.1 for my next lecture). The equation for $\boldsymbol{y}$ in the left nullspace is $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$:

Flow into a node equals flow out. Kirchhoff's Current Law is the "balance equation".

This must be 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 ■

3.5 A 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 four different columns.

You can't do anything about the sum $r + (n - r) + r + (m - r) = \boldsymbol{n + m}$. 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?

3.5 B Fact: All the rows of AB are combinations of the rows of B. So the row space of AB is contained in (possibly equal to) the row space of B. **Rank** $(\boldsymbol{AB}) \leq$ **rank** $(\boldsymbol{B})$.

All columns of AB are combinations of the columns of A. So the column space of AB is contained in (possibly equal to) the column space of A. **Rank** $(\boldsymbol{AB}) \leq$ **rank** $(\boldsymbol{A})$.

If we multiply by an *invertible* matrix, the rank will not change. The rank can't drop, because when we multiply by the inverse matrix the rank can't jump back.

Problem Set 3.5

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) Nullspace contains $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$, column 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. Multiply AB.

6 Without using 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 $B = [\, A \;\; A \,]$. (The basis for $\boldsymbol{Z}$ is empty.)

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) $\begin{bmatrix} A \end{bmatrix}$ 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 random 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 $A^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{0}$ has solutions other than $\boldsymbol{y} = \boldsymbol{0}$?

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 after the row exchange.

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:

$$\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & b_1 \\ 4 & 5 & 6 & b_2 \\ 7 & 8 & 9 & b_3 \end{bmatrix} \quad \to \quad \begin{bmatrix} 1 & 2 & 3 & b_1 \\ 0 & -3 & -6 & b_2 - 4b_1 \\ 0 & 0 & 0 & b_3 - 2b_2 + b_1 \end{bmatrix}.$$

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 vectors 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:

$$\text{(a)} \quad \begin{bmatrix} 1 & 2 & b_1 \\ 3 & 4 & b_2 \\ 4 & 6 & b_3 \end{bmatrix} \qquad \text{(b)} \quad \begin{bmatrix} 1 & 2 & b_1 \\ 2 & 3 & b_2 \\ 2 & 4 & b_3 \\ 2 & 5 & b_4 \end{bmatrix}$$

From the $\boldsymbol{b}$ column after elimination, read off $m-r$ basis vectors in the left nullspace. Those $\boldsymbol{y}$'s are combinations of rows that give zero rows in the echelon form.

20 (a) Check that the solutions to $A\boldsymbol{x} = \boldsymbol{0}$ are perpendicular to the rows of A:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4 & 1 \end{bmatrix} \begin{bmatrix} 4 & 2 & 0 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} = ER.$$

(b) How many independent solutions to $A^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{0}$? Why does $\boldsymbol{y}^{\mathrm{T}} =$ row 3 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 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!

27 Find the ranks of the 8 by 8 checkerboard matrix B and the chess matrix C:

$$B = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} r & n & b & q & k & b & n & r \\ p & p & p & p & p & p & p & p \\ & & & \text{four zero rows} & & & & \\ p & p & p & p & p & p & p & p \\ r & n & b & q & k & b & n & r \end{bmatrix}$$

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.

28 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

29 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?

30 $\mathbf{M}$ 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 $(n - r) + r = 9$?

31 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

1 Orthogonal vectors have $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w} = 0$. Then $||\boldsymbol{v}||^2 + ||\boldsymbol{w}||^2 = ||\boldsymbol{v}+\boldsymbol{w}||^2 = ||\boldsymbol{v}-\boldsymbol{w}||^2$.

2 Subspaces $\boldsymbol{V}$ and $\boldsymbol{W}$ are orthogonal when $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w} = 0$ for every $\boldsymbol{v}$ in $\boldsymbol{V}$ and every $\boldsymbol{w}$ in $\boldsymbol{W}$.

3 The row space of A is orthogonal to the nullspace. The column space is orthogonal to $\boldsymbol{N}(A^{\mathrm{T}})$.

4 One pair of dimensions adds to $r + (n - r) = n$. The other pair has $r + (m - r) = m$.

5 Row space and nullspace are orthogonal *complements*: Every $\boldsymbol{x}$ in $\mathbf{R}^n$ splits into $\boldsymbol{x}_{\mathbf{row}} + \boldsymbol{x}_{\mathbf{null}}$.

6 Suppose a space $\boldsymbol{S}$ has dimension d. Then every basis for $\boldsymbol{S}$ consists of d vectors.

7 If d vectors in $\boldsymbol{S}$ are independent, they span $\boldsymbol{S}$. If d vectors span $\boldsymbol{S}$, they are independent.

Two vectors are orthogonal when their dot product is zero: $\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{w} = 0$. 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 column vectors in Q, all pairs will be orthogonal. Think of $a^2 + b^2 = c^2$ for a *right triangle* with sides $\boldsymbol{v}$ and $\boldsymbol{w}$.

Orthogonal vectors $\quad \boldsymbol{v}^{\mathrm{T}}\boldsymbol{w} = 0 \quad$ and $\quad ||\boldsymbol{v}||^2 + ||\boldsymbol{w}||^2 = ||\boldsymbol{v}+\boldsymbol{w}||^2.$

The right side is $(\boldsymbol{v}+\boldsymbol{w})^{\mathrm{T}}(\boldsymbol{v}+\boldsymbol{w})$. This equals $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}$ when $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w} = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{v} = 0$.

Subspaces entered Chapter 3 to throw light on $A\boldsymbol{x} = \boldsymbol{b}$. Right away we needed the column space and the nullspace. 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* $\boldsymbol{x}$. At the first level this is only numbers. At the second level $A\boldsymbol{x}$ 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.

The subspaces fit together to show the hidden reality of A times $\boldsymbol{x}$. The 90° angles between subspaces are new—and we can say now what those right angles mean.

The row space is perpendicular to the nullspace. Every row of A is perpendicular to every solution of $A\boldsymbol{x} = \boldsymbol{0}$. 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 $\boldsymbol{b}$ is outside the column space—when we want to solve $A\boldsymbol{x} = \boldsymbol{b}$ and can't do it—then this nullspace of A^{T} comes into its own. It contains the error $\boldsymbol{e} = \boldsymbol{b} - A\boldsymbol{x}$ 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*** $\mathbf{R}^n$.

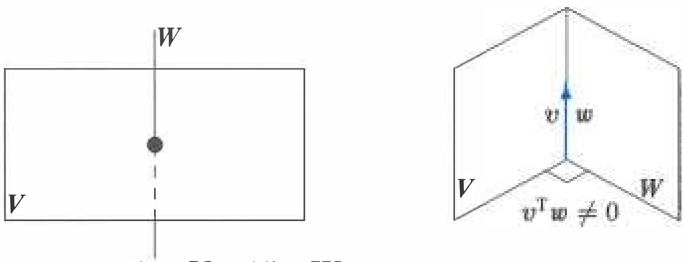
DEFINITION Two subspaces $\boldsymbol{V}$ and $\boldsymbol{W}$ of a vector space are ***orthogonal*** if every vector $\boldsymbol{v}$ in $\boldsymbol{V}$ is perpendicular to every vector $\boldsymbol{w}$ in $\boldsymbol{W}$:

Orthogonal subspaces	$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w} = 0$ ***for all*** $\boldsymbol{v}$ ***in*** $\boldsymbol{V}$ ***and all*** $\boldsymbol{w}$ ***in*** $\boldsymbol{W}$.

Example 1 The floor of your room (extended to infinity) is a subspace $\boldsymbol{V}$. The line where two walls meet is a subspace $\boldsymbol{W}$ (one-dimensional). Those subspaces are orthogonal. Every vector up the meeting line of the walls is perpendicular to every vector in the floor.

Example 2 Two walls look perpendicular but those two subspaces are not orthogonal! The meeting line is in both $\boldsymbol{V}$ and $\boldsymbol{W}$—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 $\boldsymbol{v}$ and it is $\boldsymbol{w}$, so $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} = 0$. This has to be the zero vector.



orthogonal plane $\boldsymbol{V}$ and line $\boldsymbol{W}$ non-orthogonal planes

Figure 4.1: Orthogonality is impossible when dim $\boldsymbol{V}$ + dim $\boldsymbol{W}$ > dim (whole space).

The crucial examples for linear algebra come from the four 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 $A\boldsymbol{x} = \boldsymbol{0}$:

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:

$$Ax = \begin{bmatrix} & \textbf{row 1} & \\ & \vdots & \\ & \textbf{row } m & \end{bmatrix}\begin{bmatrix} \\ x \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \quad \begin{matrix} \longleftarrow & (\textbf{row 1}) \cdot x \text{ is zero} \\ \\ \longleftarrow & (\textbf{row } m) \cdot x \text{ is zero} \end{matrix} \tag{1}$$

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 $A^{\mathrm{T}}y$ of the rows. Take the dot product of $A^{\mathrm{T}}y$ with any x in the nullspace. *These vectors are perpendicular*:

Nullspace orthogonal to row space $\quad x^{\mathrm{T}}(A^{\mathrm{T}}y) = (Ax)^{\mathrm{T}}y = \mathbf{0}^{\mathrm{T}}y = 0.$ (2)

We like the first proof. You can see those rows of A multiplying x to produce zeros in equation (1). The second proof shows why A and A^{T} are both in the Fundamental Theorem.

Example 3 The rows of A are perpendicular to $x = (1, 1, -1)$ in the nullspace:

$$Ax = \begin{bmatrix} 1 & 3 & 4 \\ 5 & 2 & 7 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{gives the dot products} \quad \begin{matrix} 1 + 3 - 4 = 0 \\ 5 + 2 - 7 = 0 \end{matrix}$$

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 column space of A and the nullspace of A^{T} are always orthogonal subspaces.

Every vector 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}. The nullspace of A^{T} is orthogonal to the row space of A^{T}—and the row space of A^{T} is the column space of A. Q.E.D.

For a visual proof, look at $A^{\mathrm{T}}y = \mathbf{0}$. Each column of A multiplies y to give 0:

$$C(A) \perp N(A^{\mathrm{T}}) \qquad A^{\mathrm{T}}y = \begin{bmatrix} (\textbf{column 1})^{\mathrm{T}} \\ \cdots \\ (\textbf{column } n)^{\mathrm{T}} \end{bmatrix}\begin{bmatrix} \\ y \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ \cdot \\ 0 \end{bmatrix}. \tag{3}$$

The dot product of y with every column of A is zero. Then y in the left nullspace is perpendicular to each column of A—and to the whole column space.

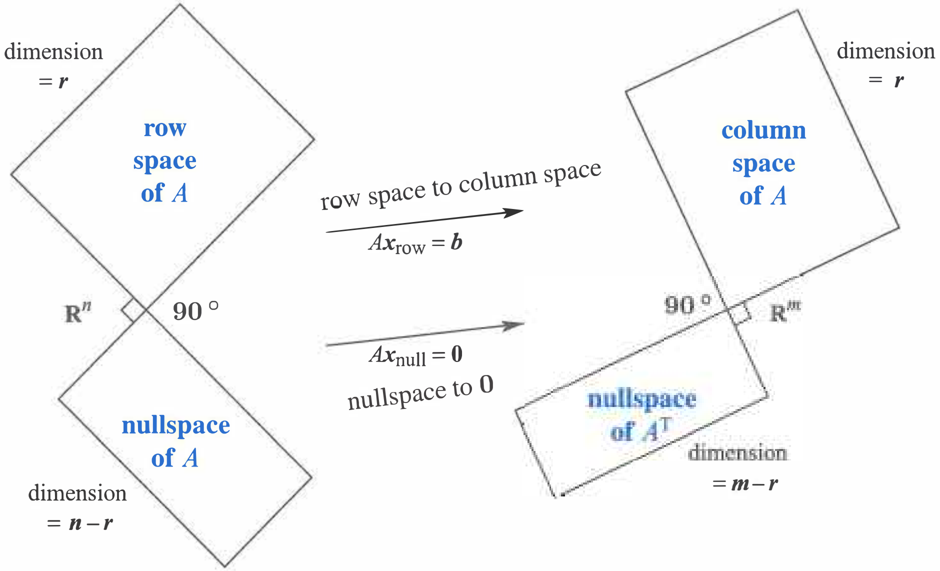


Figure 4.2: Two pairs of orthogonal subspaces. The dimensions add to n and add to m. **This is the Big Picture**—two subspaces in $\mathbf{R}^n$ and two subspaces in $\mathbf{R}^m$.

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. But the correct dimensions r and $n - r$ must add to $n = 3$.

The fundamental subspaces of a 3 by 3 matrix have dimensions 2 and 1, or 3 and 0. Those pairs of subspaces are not only orthogonal, they are *orthogonal complements*.

DEFINITION The ***orthogonal complement*** of a subspace $\boldsymbol{V}$ contains ***every*** vector that is perpendicular to $\boldsymbol{V}$. This orthogonal subspace is denoted by $\boldsymbol{V}^{\perp}$ (pronounced "$\boldsymbol{V}$ perp").

By this definition, the nullspace is the orthogonal complement of the row space. ***Every*** $\boldsymbol{x}$ that is perpendicular to the rows satisfies $A\boldsymbol{x} = \mathbf{0}$, and lies in the nullspace.

The reverse is also true. ***If*** $\boldsymbol{v}$ ***is orthogonal to the nullspace, it must be in the row space***. Otherwise we could add this $\boldsymbol{v}$ 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 $\boldsymbol{N}(A)^{\perp}$ is exactly the row space $\boldsymbol{C}(A^{\mathrm{T}})$.

In the same way, 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 $\boldsymbol{x}$ can be split into a *row space component* $\boldsymbol{x}_r$ and a *nullspace component* $\boldsymbol{x}_n$. When A multiplies $\boldsymbol{x} = \boldsymbol{x}_r + \boldsymbol{x}_n$, Figure 4.3 shows what happens to $A\boldsymbol{x} = A\boldsymbol{x}_r + A\boldsymbol{x}_n$:

The nullspace component goes to zero: $A\boldsymbol{x}_n = \mathbf{0}$.

The row space component goes to the column space: $A\boldsymbol{x}_r = A\boldsymbol{x}$.

Every vector goes to the column space! Multiplying by A cannot do anything else. More than that: *Every vector $\boldsymbol{b}$ in the column space comes from one and only one vector $\boldsymbol{x}_r$ 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 that part of A in Section 7.4.

Example 4 Every matrix of rank r 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 matrix can be diagonalized, when we choose the right bases for $\mathbf{R}^n$ and $\mathbf{R}^m$. This ***Singular Value Decomposition*** has become extremely important in applications.

Let me repeat one clear fact. A row of A can't be in the nullspace of A (except for a zero row). The only vector in two orthogonal subspaces is the zero vector.

If a vector v is orthogonal to itself then v is the zero vector.

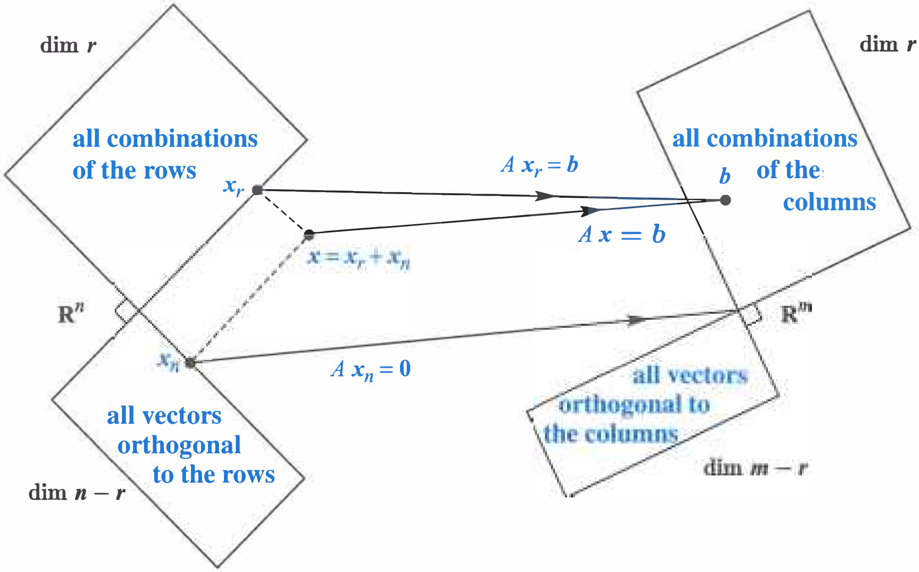


Figure 4.3: This update of Figure 4.2 shows the true action of A on $\boldsymbol{x} = \boldsymbol{x}_r + \boldsymbol{x}_n$. Row space vector $\boldsymbol{x}_r$ to column space, nullspace vector $\boldsymbol{x}_n$ to zero.

Drawing the Big Picture

I don't know the best way to draw the four subspaces in Figures 4.2 and 4.3. This big picture has to show the orthogonality of those subspaces. I can see a possible way to do it when a line meets a plane—maybe Figure 4.4 also shows that those spaces are infinite, more clearly than the rectangles in Figure 4.3. But how do I draw a pair of two-dimensional subspaces in $\mathbf{R}^4$, to show they are orthogonal to each other? Good ideas are welcome.

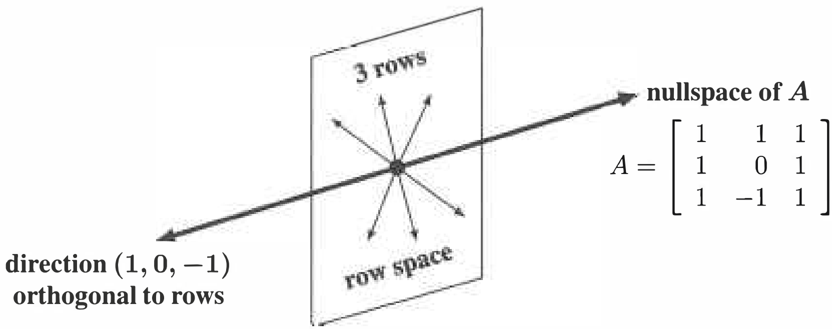


Figure 4.4: Row space of A = plane. Nullspace = orthogonal line. Dimensions $2 + 1 = 3$.

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 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 pivot columns. Then back substitution solves $A\boldsymbol{x} = \boldsymbol{b}$ (the solution exists).

Starting in the opposite direction, suppose that $A\boldsymbol{x} = \boldsymbol{b}$ 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 $\boldsymbol{x}_r = \boldsymbol{x}_n = \mathbf{0}$. All coefficients of the row space basis and of the nullspace basis must be zero. This 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, with dimensions $(n-r)+r=n$. Similarly $\boldsymbol{N}(A^{\mathrm{T}})$ and $\boldsymbol{C}(A)$ are orthogonal complements with $(m-r)+r=m$.

4. Any n independent vectors in $\mathbf{R}^n$ span $\mathbf{R}^n$. Any n spanning vectors are independent.

■ 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 smallest possible size of a matrix B that has nullspace $\boldsymbol{S}^{\perp}$?

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 will be a basis for $\boldsymbol{S}$).

(d) This is the same as question (c) !

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. The special solutions $\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3$ to $A\boldsymbol{x}=\mathbf{0}$. will be the same for $B\boldsymbol{x}=\mathbf{0}$. 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? *Ans*: $A=[1\ -3\ -4]$.

(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). *Answer*: $\boldsymbol{s}_1=(3,1,0)$ and $\boldsymbol{s}_2=(4,0,1)$.

(c) Find a basis for the line $\boldsymbol{P}^{\perp}$ that is perpendicular to $\boldsymbol{P}$. *Answer*: $(1,-3,-4)$**!**

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 (and put 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 $\boldsymbol{Z}$ (zero vector only)? The nullspace part of any vector $\boldsymbol{x}$ in $\mathbf{R}^2$ is $\boldsymbol{x}_n =$ ______.

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) $A\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ 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. With $AB = 0$, why can't A and B be 3 by 3 matrices of rank 2?

5 (a) If $A\boldsymbol{x} = \boldsymbol{b}$ has a solution and $A^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{0}$, *is* $(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} = 0)$ *or* $(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b} = 0)$?

(b) If $A^{\mathrm{T}}\boldsymbol{y} = (1, 1, 1)$ has a solution and $A\boldsymbol{x} = \boldsymbol{0}$, then ______.

6 This system of equations $A\boldsymbol{x} = \boldsymbol{b}$ 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 $\boldsymbol{y}$ in which subspace? Its dot product $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}$ is 1, so no solution $\boldsymbol{x}$.

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

$$A\boldsymbol{x} = \boldsymbol{b} \quad \textbf{OR} \quad A^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{0} \quad \text{with} \quad \boldsymbol{y}^{\mathrm{T}}\boldsymbol{b} = 1.$$

If $\boldsymbol{b}$ 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 $A\boldsymbol{x}_r$ is equal to $A\boldsymbol{x}$? How do we know that this vector is in the column space? If $A = \left[\begin{smallmatrix}1&1\\1&1\end{smallmatrix}\right]$ and $\boldsymbol{x} = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ what is $\boldsymbol{x}_r$?

9 If $A^{\mathrm{T}}A\boldsymbol{x} = \mathbf{0}$ then $A\boldsymbol{x} = \mathbf{0}$. Reason: $A\boldsymbol{x}$ 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 $A\boldsymbol{x} = \mathbf{0}$ and $A\boldsymbol{z} = 5\boldsymbol{z}$, which subspaces contain these "eigenvectors" $\boldsymbol{x}$ and $\boldsymbol{z}$? **Symmetric matrices have perpendicular eigenvectors** $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{z} = 0$.

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 $\boldsymbol{x}_r$ and $\boldsymbol{x}_n$ and draw Figure 4.3 properly if

$$A = \begin{bmatrix}1&-1\\0&0\\0&0\end{bmatrix} \quad \text{and} \quad \boldsymbol{x} = \begin{bmatrix}2\\0\end{bmatrix}.$$

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 S is the subspace of $\mathbf{R}^3$ containing only the zero vector, what is $S^\perp$? If S is spanned by $(1,1,1)$, what is $S^\perp$? If S is spanned by $(1,1,1)$ and $(1,1,-1)$, what is a basis for $S^\perp$?

18 Suppose S only contains two vectors $(1,5,1)$ and $(2,2,2)$ (not a subspace). Then $S^\perp$ is the nullspace of the matrix $A =$ _____. $S^\perp$ is a subspace even if 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 S is spanned by the vectors $(1,2,2,3)$ and $(1,3,3,2)$. Find two vectors that span $S^\perp$. 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 S is contained in a subspace $\boldsymbol{V}$, prove that $S^\perp$ 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 $\text{rank}(A) + \text{rank}(B) \leq 4$.

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

1 The projection of a vector $\boldsymbol{b}$ onto the line through $\boldsymbol{a}$ is the closest point $\boldsymbol{p} = \boldsymbol{a}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} \,/\, \boldsymbol{a}^{\mathrm{T}}\boldsymbol{a})$.

2 The error $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ is perpendicular to $\boldsymbol{a}$: Right triangle $\boldsymbol{b}\,\boldsymbol{p}\,\boldsymbol{e}$ has $||\boldsymbol{p}||^2 + ||\boldsymbol{e}||^2 = ||\boldsymbol{b}||^2$.

3 The **projection** of $\boldsymbol{b}$ onto a subspace $\boldsymbol{S}$ is the closest vector $\boldsymbol{p}$ in $\boldsymbol{S}$; $\boldsymbol{b} - \boldsymbol{p}$ is orthogonal to $\boldsymbol{S}$.

4 $A^{\mathrm{T}}A$ is invertible (and symmetric) only if A has independent columns: $\boldsymbol{N}(A^{\mathrm{T}}A) = \boldsymbol{N}(A)$.

5 Then the projection of $\boldsymbol{b}$ onto the column space of A is the vector $\boldsymbol{p} = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{b}$.

6 The **projection matrix** onto $\boldsymbol{C}(A)$ is $\boxed{P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}.}$ It has $\boldsymbol{p} = P\boldsymbol{b}$ and $P^2 = P = P^{\mathrm{T}}$.

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* $\boldsymbol{b}$ *is* $P\boldsymbol{b}$.

1 What are the projections of $\boldsymbol{b} = (2, 3, 4)$ onto the z axis and the xy plane?

2 What matrices P_1 and P_2 produce those projections onto a line and a plane?

When $\boldsymbol{b}$ is projected onto a line, ***its projection p is the part of b along that line***. If $\boldsymbol{b}$ is projected onto a plane, $\boldsymbol{p}$ is the part in that plane. ***The projection p is*** $P\boldsymbol{b}$.

The projection matrix P multiplies $\boldsymbol{b}$ to give $\boldsymbol{p}$. This section finds $\boldsymbol{p}$ and also P.

The projection onto the z axis we call $\boldsymbol{p}_1$. The second projection drops straight down to the xy plane. The picture in your mind should be Figure 4.5. Start with $\boldsymbol{b} = (2, 3, 4)$. The projection across gives $\boldsymbol{p}_1 = (0, 0, 4)$. The projection down gives $\boldsymbol{p}_2 = (2, 3, 0)$. Those are the parts of $\boldsymbol{b}$ along the z axis and in the xy plane.

The projection matrices P_1 and P_2 are 3 by 3. They multiply $\boldsymbol{b}$ with 3 components to produce $\boldsymbol{p}$ with 3 components. Projection onto a line comes from a rank one matrix. Projection onto a plane comes from a rank two matrix:

Projection matrix
Onto the z axis: $\quad P_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} \end{bmatrix}$ $\quad$ Onto the xy plane: $\quad P_2 = \begin{bmatrix} \mathbf{1} & 0 & 0 \\ 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

P_1 picks out the z component of every vector. P_2 picks out the x and y components. To find the projections $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ of $\boldsymbol{b}$, multiply $\boldsymbol{b}$ by P_1 and P_2 (small $\boldsymbol{p}$ for the vector, capital P for the matrix that produces it):

$$\boldsymbol{p}_1 = P_1\boldsymbol{b} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \boldsymbol{z} \end{bmatrix} \qquad \boldsymbol{p}_2 = P_2\boldsymbol{b} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ \mathbf{0} \end{bmatrix}.$$

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.

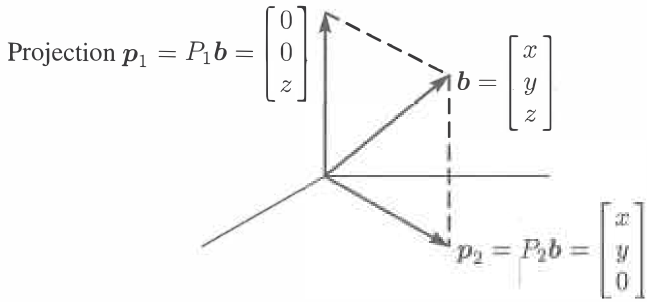


Figure 4.5: The projections $\boldsymbol{p}_1 = P_1\boldsymbol{b}$ and $\boldsymbol{p}_2 = P_2\boldsymbol{b}$ onto the z axis and the xy plane.

More than just orthogonal, 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 two parts of $\boldsymbol{b}$:

$$\text{The vectors give } \boldsymbol{p}_1 + \boldsymbol{p}_2 = \boldsymbol{b}. \qquad \text{The matrices give } P_1 + P_2 = I. \tag{1}$$

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). So $\boldsymbol{p}_2 = \boldsymbol{p}_3$ and $P_2 = P_3$.

$$A_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad A_3 = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 0 & 0 \end{bmatrix}.$$

Our problem is ***to project any $\boldsymbol{b}$ onto the column space of any m by n matrix***. Start with a line (dimension $n = 1$). The matrix A will have 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 $\boldsymbol{b}$ to $\boldsymbol{p}$ is perpendicular to the vector $\boldsymbol{a}$.*** This is the dotted line marked $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ for the error on the left side of Figure 4.6. We now compute $\boldsymbol{p}$ by algebra.

The projection $\boldsymbol{p}$ will be 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 will 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 the "error" $\boldsymbol{e} = \boldsymbol{b} - \widehat{x}\boldsymbol{a}$. It is perpendicular to $\boldsymbol{a}$—this will determine $\widehat{x}$. Use the fact that **$\boldsymbol{b} - \widehat{x}\boldsymbol{a}$ is perpendicular to $\boldsymbol{a}$** when their dot product is zero:

Projecting $\boldsymbol{b}$ onto $\boldsymbol{a}$ with error $\boldsymbol{e} = \boldsymbol{b} - \widehat{x}\boldsymbol{a}$

$\boldsymbol{a} \cdot (\boldsymbol{b} - \widehat{x}\boldsymbol{a}) = 0$ or $\boldsymbol{a} \cdot \boldsymbol{b} - \widehat{x}\boldsymbol{a} \cdot \boldsymbol{a} = 0$

$$\widehat{x} = \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{\boldsymbol{a} \cdot \boldsymbol{a}} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}. \tag{2}$$

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 $\boldsymbol{p} = \widehat{x}\boldsymbol{a}$.

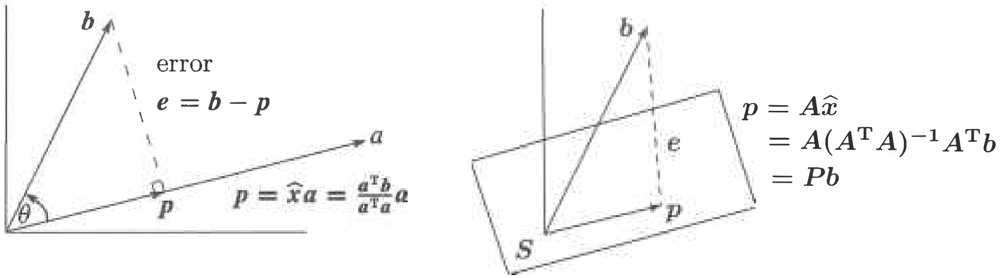


Figure 4.6: The projection $\boldsymbol{p}$ of $\boldsymbol{b}$ onto a line and onto $\boldsymbol{S}$ = column space of A.

The projection of b onto the line through a is the vector $\boldsymbol{p} = \widehat{x}\boldsymbol{a} = \dfrac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}\boldsymbol{a}$.

Special case 1: If $\boldsymbol{b} = \boldsymbol{a}$ then $\widehat{x} = 1$. The projection of $\boldsymbol{a}$ onto $\boldsymbol{a}$ is itself. $P\boldsymbol{a} = \boldsymbol{a}$.

Special case 2: If $\boldsymbol{b}$ is perpendicular to $\boldsymbol{a}$ then $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} = 0$. The projection is $\boldsymbol{p} = \boldsymbol{0}$.

Example 1 Project $\boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ onto $\boldsymbol{a} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ to find $\boldsymbol{p} = \widehat{x}\boldsymbol{a}$ in Figure 4.6.

Solution The number $\widehat{x}$ is the ratio of $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} = 5$ to $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a} = 9$. So the projection is $\boldsymbol{p} = \frac{5}{9}\boldsymbol{a}$.

The error vector between $\boldsymbol{b}$ and $\boldsymbol{p}$ is $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$. Those vectors $\boldsymbol{p}$ and $\boldsymbol{e}$ will add to $\boldsymbol{b} = (\mathbf{1}, \mathbf{1}, \mathbf{1})$:

$$p = \frac{5}{9}a = \left(\frac{5}{9}, \frac{10}{9}, \frac{10}{9}\right) \quad \text{and} \quad e = b - p = \left(\frac{4}{9}, -\frac{1}{9}, -\frac{1}{9}\right).$$

The error $\boldsymbol{e}$ should be perpendicular to $\boldsymbol{a} = (1, 2, 2)$ and it is: $\boldsymbol{e}^{\mathrm{T}}\boldsymbol{a} = \frac{4}{9} - \frac{2}{9} - \frac{2}{9} = 0$.

Look at the right triangle of $\boldsymbol{b}$, $\boldsymbol{p}$, and $\boldsymbol{e}$. The vector $\boldsymbol{b}$ is split into two parts—its component along the line is $\boldsymbol{p}$, its perpendicular part is $\boldsymbol{e}$. Those two sides $\boldsymbol{p}$ and $\boldsymbol{e}$ have length $||\boldsymbol{p}|| = ||\boldsymbol{b}|| \cos\theta$ and $||\boldsymbol{e}|| = ||\boldsymbol{b}|| \sin\theta$. Trigonometry matches the dot product:

$$\boldsymbol{p} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}\boldsymbol{a} \quad \text{has length} \quad \boxed{||\boldsymbol{p}||} = \frac{||\boldsymbol{a}||\,||\boldsymbol{b}|| \cos\theta}{||\boldsymbol{a}||^2}||\boldsymbol{a}|| = \boxed{||\boldsymbol{b}|| \cos\theta.} \tag{3}$$

The dot product is a lot simpler than getting involved with $\cos\theta$ and the length of $\boldsymbol{b}$. The example has square roots in $\cos\theta = 5/3\sqrt{3}$ and $||\boldsymbol{b}|| = \sqrt{3}$. There are no square roots in the projection $\boldsymbol{p} = 5\boldsymbol{a}/9$. The good way to $5/9$ is $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$.

Now comes the ***projection matrix***. In the formula for $\boldsymbol{p}$, what matrix is multiplying $\boldsymbol{b}$? You can see the matrix better if the number $\widehat{\boldsymbol{x}}$ is on the right side of $\boldsymbol{a}$:

Projection matrix P $\quad \boldsymbol{p} = \boldsymbol{a}\widehat{\boldsymbol{x}} = \boldsymbol{a}\dfrac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}} = P\boldsymbol{b}$ ***when the matrix is*** $\quad P = \dfrac{\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}.$

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 line 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:

$$\boldsymbol{p} = P\boldsymbol{b} = \frac{1}{9}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 5 \\ 10 \\ 10 \end{bmatrix} \quad \text{which is correct.}$$

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.

With $n=1$ (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 $\widehat{x} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ when $n=1$. For $n>1$, the best $\widehat{\boldsymbol{x}} = (\widehat{x}_1, \ldots, \widehat{x}_n)$ is to be found now.

We compute projections onto n-dimensional subspaces in three steps as before: ***Find the vector*** $\widehat{\boldsymbol{x}}$***, find the projection*** $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$***, find the projection matrix*** P.

The key is in the geometry! The dotted line in Figure 4.6 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 $\boldsymbol{b} - A\widehat{\boldsymbol{x}}$ makes a right angle with all the vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ in the base. The n right angles give the n equations for $\widehat{\boldsymbol{x}}$:

$$\begin{aligned} \boldsymbol{a}_1^{\mathrm{T}}(\boldsymbol{b} - A\widehat{\boldsymbol{x}}) &= 0 \\ &\vdots \\ \boldsymbol{a}_n^{\mathrm{T}}(\boldsymbol{b} - A\widehat{\boldsymbol{x}}) &= 0 \end{aligned} \qquad \text{or} \qquad \boxed{\begin{bmatrix} - & \boldsymbol{a}_1^{\mathrm{T}} & - \\ & \vdots & \\ - & \boldsymbol{a}_n^{\mathrm{T}} & - \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{b} - A\widehat{\boldsymbol{x}} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \mathbf{0} \\ \\ \end{bmatrix}.} \tag{4}$$

The matrix with those rows $\boldsymbol{a}_i^{\mathrm{T}}$ is A^{T}. The n equations are exactly $A^{\mathrm{T}}(\boldsymbol{b} - A\widehat{\boldsymbol{x}}) = \mathbf{0}$.

Rewrite $A^{\mathrm{T}}(\boldsymbol{b} - A\widehat{\boldsymbol{x}}) = \mathbf{0}$ in its famous form $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$. This is the equation for $\widehat{\boldsymbol{x}}$, and the coefficient matrix is $A^{\mathrm{T}}A$. Now we can find $\widehat{\boldsymbol{x}}$ and $\boldsymbol{p}$ 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 $\widehat{x}$:

$$\textbf{Find } \widehat{x}\ (n \times 1) \qquad \boxed{A^{\mathrm{T}}(b - A\widehat{x}) = 0 \quad \text{or} \quad A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b.} \tag{5}$$

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:

$$\textbf{Find } p\ (m \times 1) \qquad \boxed{p = A\widehat{x} = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b.} \tag{6}$$

The next formula picks out the ***projection matrix*** that is multiplying b in (6):

$$\textbf{Find } P\ (m \times m) \qquad \boxed{P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}.} \tag{7}$$

Compare with projection onto a line, when A has only one column: $A^{\mathrm{T}}A$ is $a^{\mathrm{T}}a$.

$$\textbf{For } n = 1 \qquad \boxed{\widehat{x} = \frac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a}} \quad \textbf{and} \quad \boxed{p = a\frac{a^{\mathrm{T}}b}{a^{\mathrm{T}}a}} \quad \textbf{and} \quad \boxed{P = \frac{aa^{\mathrm{T}}}{a^{\mathrm{T}}a}.}$$

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 $A^{\mathrm{T}}(b - A\widehat{x}) = 0$. We used geometry ($e$ is orthogonal to each a). *Linear algebra gives this "normal equation" too*, in a very quick and beautiful 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 $A^{\mathrm{T}}(b - A\widehat{x}) = 0$.

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 with sides p, e, and b.

Example 3 If $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ and $b = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$ find $\widehat{x}$ and p and P.

Solution Compute the square matrix $A^{\mathrm{T}}A$ and also the vector $A^{\mathrm{T}}b$:

$$A^{\mathrm{T}}A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \quad \text{and} \quad A^{\mathrm{T}}b = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}.$$

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:

$$\boldsymbol p = 5\begin{bmatrix}1\\1\\1\end{bmatrix}-3\begin{bmatrix}0\\1\\2\end{bmatrix}=\begin{bmatrix}5\\2\\-1\end{bmatrix}.\quad\text{The error is}\quad\boldsymbol e=\boldsymbol b-\boldsymbol p=\begin{bmatrix}1\\-2\\1\end{bmatrix}. \tag{9}$$

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 matrix P times $\boldsymbol b = (6,0,0)$ correctly gives $\boldsymbol p = (5,2,-1)$. That solves the problem for one particular $\boldsymbol b$, as soon as we find P.

The projection matrix is $P = A(A^{\mathrm T}A)^{-1}A^{\mathrm T}$. 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 $A^{\mathrm T}$ 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 $A\boldsymbol x = \mathbf 0$. Multiplying by $A^{\mathrm T}$ gives $A^{\mathrm T}A\boldsymbol x = \mathbf 0$. So $\boldsymbol x$ is also in the nullspace of $A^{\mathrm T}A$.

Now start with the nullspace of $A^{\mathrm T}A$. **From $A^{\mathrm T}A\boldsymbol x = \mathbf 0$ we must prove $A\boldsymbol x = \mathbf 0$.** We can't multiply by $(A^{\mathrm T})^{-1}$, which generally doesn't exist. Just multiply by $\boldsymbol x^{\mathrm T}$:

$$(\boldsymbol x^{\mathrm T})A^{\mathrm T}A\boldsymbol x = 0\quad\text{or}\quad(A\boldsymbol x)^{\mathrm T}(A\boldsymbol x)=0\quad\text{or}\quad\|A\boldsymbol x\|^2=0. \tag{11}$$

We have shown: If $A^{\mathrm T}A\boldsymbol x = \mathbf 0$ then $A\boldsymbol x$ has length zero. Therefore $A\boldsymbol x = \mathbf 0$. 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: $A^{\mathrm T}A$ is invertible.

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 $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$ 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 b to the subspace $C(A)$ is* $\|e\|$.

■ REVIEW OF THE KEY IDEAS ■

1. The projection of $\boldsymbol{b}$ onto the line through $\boldsymbol{a}$ is $\boldsymbol{p} = \boldsymbol{a}\widehat{\boldsymbol{x}} = \boldsymbol{a}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a})$.

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 $P^2 = P$ and $P\boldsymbol{b} = \boldsymbol{p}$.

■ 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*. Why is it exactly the error $\boldsymbol{e}^*$?

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 $$\boldsymbol{p} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}\boldsymbol{a} = \frac{18}{9}(2,2,1) = (4,4,2) = 2\boldsymbol{a}.$$

The error vector $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = (-1,0,2)$ is perpendicular to $\boldsymbol{a} = (2,2,1)$. So $\boldsymbol{p}$ is correct. The plane of $\boldsymbol{a} = (2,2,1)$ and $\boldsymbol{a}^* = (1,0,0)$ is the column space of $A = [\boldsymbol{a}\ \boldsymbol{a}^*]$:

$$A = \begin{bmatrix} 2 & 1 \\ 2 & 0 \\ 1 & 0 \end{bmatrix} \quad A^{\mathrm{T}}A = \begin{bmatrix} 9 & 2 \\ 2 & 1 \end{bmatrix} \quad (A^{\mathrm{T}}A)^{-1} = \frac{1}{5}\begin{bmatrix} 1 & -2 \\ -2 & 9 \end{bmatrix} \quad P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & .8 & .4 \\ 0 & .4 & .2 \end{bmatrix}$$

Now $\boldsymbol{p}^* = P\boldsymbol{b} = (3, 4.8, 2.4)$. 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 $P^2 = P = P^{\mathrm{T}}$.

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* $\widehat{x} = 90$:

$$\frac{dE}{dx} = 2(x-70) + 2(x-80) + 2(x-120) = 0 \quad \text{gives} \quad \widehat{x} = \frac{70+80+120}{3} = 90.$$

Also by projection : $$\widehat{x} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}} = \frac{(1,1,1)^{\mathrm{T}}(70,80,120)}{(1,1,1)^{\mathrm{T}}(1,1,1)} = \frac{70+80+120}{3} = 90.$$

In *recursive* least squares, a fourth measurement 130 changes the average $\widehat{x}_{\text{old}} = 90$ to $\widehat{x}_{\text{new}} = 100$. Verify the *update formula* $\widehat{x}_{\text{new}} = \widehat{x}_{\text{old}} + \frac{1}{4}(130 - \widehat{x}_{\text{old}})$. When a new measurement arrives, we don't have to average all the old measurements again!

Problem Set 4.2

Questions 1–9 ask for projections $\boldsymbol{p}$ onto lines. Also errors $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ 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 $\boldsymbol{p} = \widehat{x}\boldsymbol{a}$:

$$\text{(a)}\ \boldsymbol{b} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \quad \text{and} \quad \boldsymbol{a} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \text{(b)} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{a} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

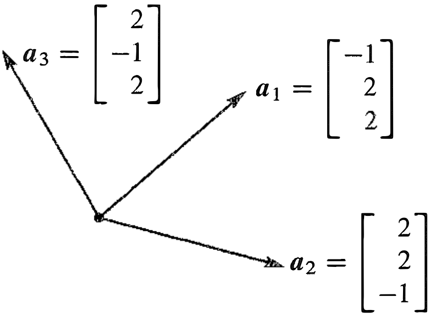
3 In Problem 1, find the projection matrix $P = \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ onto the line through each vector $\boldsymbol{a}$. Verify in both cases that $P^2 = P$. Multiply $P\boldsymbol{b}$ in each case to compute the projection $\boldsymbol{p}$.

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 $P_1P_2 = 0$.

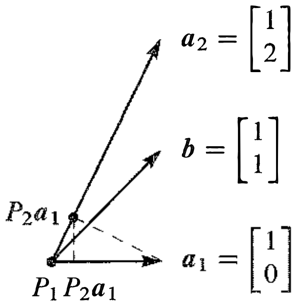
5 Compute the projection matrices $\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ onto the lines through $\boldsymbol{a}_1 = (-1, 2, 2)$ and $\boldsymbol{a}_2 = (2, 2, -1)$. Multiply those projection matrices and explain why their product P_1P_2 is what it is.

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 $P_1 + P_2 + P_3 = I$. This is because the basis $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ is orthogonal!



Questions 5–6–7: orthogonal



Questions 8–9–10: not 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.

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 $A = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} =$ invertible matrix.

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_1P_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 $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$:

$$\text{(a)}\ A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \qquad \text{(b)} \quad A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix}.$$

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 will be the column space of A! 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 of A so its projection onto that column space 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*** e?

24 The nullspace of A^{T} is _____ to the column space $\boldsymbol{C}(A)$. So if $A^{\mathrm{T}}\boldsymbol{b} = \mathbf{0}$, 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 of $\mathbf{R}^m$ 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*** $\boldsymbol{A^{\mathrm{T}}Ax} = \mathbf{0}$ ***then*** $\boldsymbol{Ax} = \mathbf{0}$. *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 also a combination $\boldsymbol{p}$ 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_2P_1.

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

33 Suppose you know the average $\widehat{x}_{\text{old}}$ of $b_1, b_2, \ldots, b_{999}$. When b_{1000} arrives, check that the new average is a combination of $\widehat{x}_{\text{old}}$ and the mismatch $b_{1000} - \widehat{x}_{\text{old}}$:

$$\widehat{\boldsymbol{x}}_{\textbf{new}} = \frac{b_1 + \cdots + b_{1000}}{1000} = \frac{b_1 + \cdots + b_{999}}{999} + \frac{1}{1000}\left(b_{1000} - \frac{b_1 + \cdots + b_{999}}{999}\right).$$

This is a "**Kalman filter**" $\widehat{x}_{\text{new}} = \widehat{x}_{\text{old}} + \frac{1}{1000}\,(b_{1000} - \widehat{x}_{\text{old}})$ with gain matrix $\frac{1}{1000}$. The last page of the book extends the Kalman filter to matrix updates.

34 (2017) Suppose P_1 and P_2 are projection matrices ($P_i^2 = P_i = P_i^{\mathrm{T}}$). Prove this fact:

$$P_1P_2 \text{ is a projection matrix if and only if } P_1P_2 = P_2P_1.$$

4.3 Least Squares Approximations

1 Solving $\boxed{A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}}$ gives the projection $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$ of $\boldsymbol{b}$ onto the column space of A.

2 When $A\boldsymbol{x} = \boldsymbol{b}$ has no solution, $\widehat{\boldsymbol{x}}$ is the "least-squares solution": $||\boldsymbol{b} - A\widehat{\boldsymbol{x}}||^2 =$ minimum.

3 Setting partial derivatives of $E = ||A\boldsymbol{x} - \boldsymbol{b}||^2$ to zero $\left(\frac{\partial E}{\partial x_i} = 0\right)$ also produces $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$.

4 To fit points $(t_1, b_1), \ldots, (t_m, b_m)$ by a straight line, A has columns $(1, \ldots, 1)$ and $(t_1, \ldots, t_m)$.

5 In that case $A^{\mathrm{T}}A$ is the 2 by 2 matrix $\begin{bmatrix} m & \Sigma\, t_i \\ \Sigma\, t_i & \Sigma\, t_i^2 \end{bmatrix}$ and $A^{\mathrm{T}}\boldsymbol{b}$ is the vector $\begin{bmatrix} \Sigma\, b_i \\ \Sigma\, t_i b_i \end{bmatrix}$.

It often happens that $A\boldsymbol{x} = \boldsymbol{b}$ has no solution. The usual reason is: *too many equations*. The matrix A 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, $\boldsymbol{b}$ is outside that column space of A. 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 $\boldsymbol{e} = \boldsymbol{b} - A\boldsymbol{x}$ down to zero. When $\boldsymbol{e}$ is zero, $\boldsymbol{x}$ is an exact solution to $A\boldsymbol{x} = \boldsymbol{b}$. *When the length of* $\boldsymbol{e}$ *is as small as possible,* $\widehat{\boldsymbol{x}}$ *is a least squares solution.* Our goal in this section is to compute $\widehat{\boldsymbol{x}}$ and use it. These are real problems and they need an answer.

The previous section emphasized $\boldsymbol{p}$ (the projection). This section emphasizes $\widehat{\boldsymbol{x}}$ (the least squares solution). They are connected by $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$. The fundamental equation is still $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$. Here is a short unofficial way to reach this "*normal equation*":

When $A\boldsymbol{x} = \boldsymbol{b}$ has no solution, multiply by A^{T} and solve $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$.

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: $n = 2$ and $m = 3$. Here are the three equations at $t = 0, 1, 2$ to match the given values $b = 6, 0, 0$:

$\boldsymbol{t = 0}$	The first point is on the line $b = C + Dt$ if	$C + D \cdot 0 = 6$
$\boldsymbol{t = 1}$	The second point is on the line $b = C + Dt$ if	$C + D \cdot 1 = 0$
$\boldsymbol{t = 2}$	The third point is on the line $b = C + Dt$ if	$C + D \cdot 2 = 0$.

This 3 by 2 system has *no solution*: $\boldsymbol{b} = (6,0,0)$ is not a combination of the columns $(1,1,1)$ and $(0,1,2)$. Read off A, $\boldsymbol{x}$, and $\boldsymbol{b}$ from those equations:

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad \boldsymbol{x} = \begin{bmatrix} C \\ D \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} \quad A\boldsymbol{x} = \boldsymbol{b} \text{ is } \textit{not} \text{ solvable.}$$

The same numbers were in Example 3 in the last section. We computed $\widehat{\boldsymbol{x}} = (5,-3)$. **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 $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$.

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 $\boldsymbol{e} = \boldsymbol{b} - A\boldsymbol{x}$ as small as possible? This is an important question with a beautiful answer. The best $\boldsymbol{x}$ (called $\widehat{\boldsymbol{x}}$) can be found by geometry (the error $\boldsymbol{e}$ meets the column space of A at 90°) and by algebra: $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$. Calculus gives the same $\widehat{\boldsymbol{x}}$: the derivative of the error $||A\boldsymbol{x} - \mathbf{b}||^2$ is zero at $\widehat{\boldsymbol{x}}$.

By geometry Every $A\boldsymbol{x}$ 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 $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$, perpendicular to the columns. *The three points at heights* (p_1, p_2, p_3) *do lie on a line*, because $\boldsymbol{p}$ is in the column space of A. In fitting a straight line, $\widehat{\boldsymbol{x}}$ is 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 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 can and do solve (by removing $\boldsymbol{e}$ and solving $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$):

$$A\boldsymbol{x} = \boldsymbol{b} = \boldsymbol{p} + \boldsymbol{e} \quad \text{is impossible} \qquad A\widehat{\boldsymbol{x}} = \boldsymbol{p} \quad \text{is solvable} \qquad \widehat{\boldsymbol{x}} \text{ is } (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{b}. \quad (1)$$

The solution to $A\widehat{\boldsymbol{x}} = \boldsymbol{p}$ leaves the least possible error (which is $\boldsymbol{e}$):

$$\textbf{Squared length for any } \boldsymbol{x} \qquad ||A\boldsymbol{x} - \mathbf{b}||^2 = ||A\boldsymbol{x} - \mathbf{p}||^2 + ||\mathbf{e}||^2. \quad (2)$$

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 $\boldsymbol{x} = \widehat{\boldsymbol{x}}$. That leaves the smallest possible error $\boldsymbol{e} = (e_1, e_2, e_3)$ which we can't reduce.

Notice what "smallest" means. The *squared length* of $A\boldsymbol{x} - \boldsymbol{b}$ is minimized:

The least squares solution $\widehat{\boldsymbol{x}}$ makes $\boldsymbol{E} = ||A\boldsymbol{x} - \mathbf{b}||^2$ as small as possible.

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 $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$.

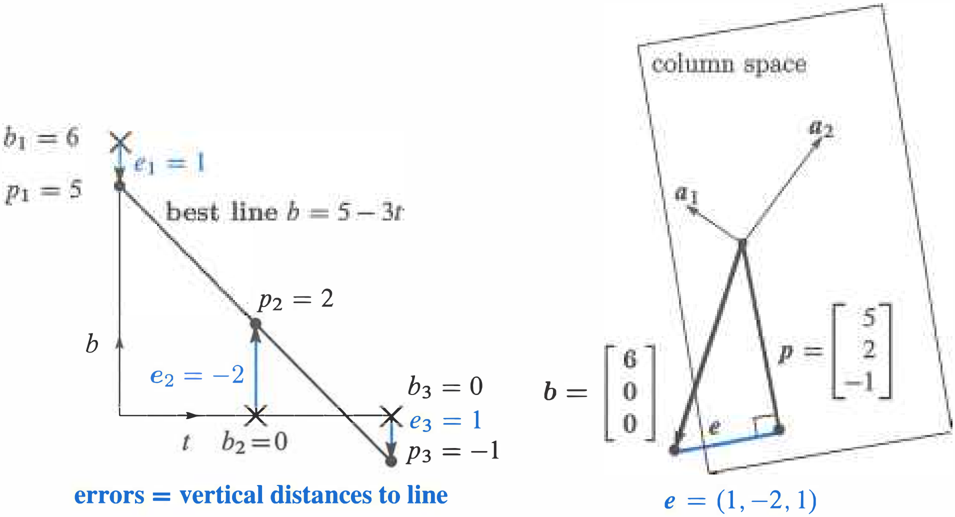


Figure 4.6: **Best line and projection: Two pictures, same problem.** The line has heights $\boldsymbol{p} = (5, 2, -1)$ with errors $\boldsymbol{e} = (1, -2, 1)$. The equations $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ give $\widehat{\boldsymbol{x}} = (5, -3)$. Same answer! The best line is $b = 5 - 3t$ and the closest point is $\boldsymbol{p} = 5\boldsymbol{a}_1 - 3\boldsymbol{a}_2$.

Notice that the errors $1, -2, 1$ add to zero. *Reason*: 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):

$$E = \|A\boldsymbol{x} - \boldsymbol{b}\|^2 = (C + D\cdot 0 - 6)^2 + (C + D\cdot 1)^2 + (C + D\cdot 2)^2. \qquad (3)$$

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 $(C + 2D)^2$ was 2 times $C + 2D$ times that extra 2.) Those factors are just $\mathbf{1}, \mathbf{1}, \mathbf{1}$ in $\partial E/\partial C$.

It is no accident that those factors 1, 1, 1 and 0, 1, 2 in the derivatives of $||Ax - b||^2$ are the columns of A. Now cancel 2 from every term and collect all C's and all D's:

$$\begin{array}{ll}\text{The } C \text{ derivative is zero:} & 3C + 3D = 6 \\ \text{The } D \text{ derivative is zero:} & 3C + 5D = 0\end{array} \qquad \textbf{This matrix} \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \textbf{ is } \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} \qquad (4)$$

These equations are identical with $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$. The best C and D are the components of $\widehat{x}$. 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 $||Ax - b||^2$ ***are zero when*** $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$.

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 $p = 5, 2, -1$. It could not go through $b = 6, 0, 0$. The errors are $1, -2, 1$. This is the vector e!

The Big Picture for Least Squares

The key figure of this book shows the four subspaces and the true action of a matrix. The vector x on the left side of Figure 4.3 went to $b = Ax$ on the right side. In that figure x was split into $x_r + x_n$. There were *many* solutions to $Ax = b$.

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* $b = p + e$. Figure 4.7 shows the big picture for least squares. Instead of $Ax = b$ we solve $A\widehat{x} = p$. The error $e = b - p$ is unavoidable.

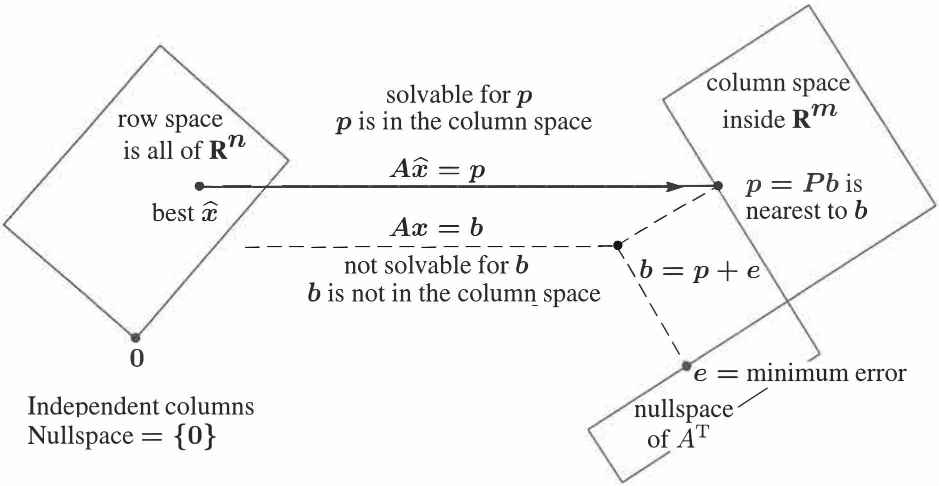


Figure 4.7: The projection $p = A\widehat{x}$ is closest to b, so $\widehat{x}$ minimizes $E = ||b - Ax||^2$.

Notice how the nullspace $N(A)$ is very small—just one point. With independent columns, the only solution to $Ax = 0$ is $x = 0$. 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 $A^{\mathrm{T}}e = 0$.

Chapter 7 will have the complete picture—all four subspaces included. Every $\boldsymbol{x}$ splits into $\boldsymbol{x}_r + \boldsymbol{x}_n$, and every $\boldsymbol{b}$ splits into $\boldsymbol{p} + \boldsymbol{e}$. The best solution is $\widehat{\boldsymbol{x}} = \widehat{\boldsymbol{x}}_r$ in the row space. We can't help $\boldsymbol{e}$ and we don't want $\boldsymbol{x}_n$ from the nullspace—this leaves $A\widehat{\boldsymbol{x}} = \boldsymbol{p}$.

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 $\widehat{\boldsymbol{x}}$ are still C and D.

A line goes through the m points when we exactly solve $A\boldsymbol{x} = \boldsymbol{b}$. 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 have two unknowns!).

$$A\boldsymbol{x} = \boldsymbol{b} \quad \text{is} \quad \begin{aligned} C + Dt_1 &= b_1 \\ C + Dt_2 &= b_2 \\ &\vdots \\ C + Dt_m &= b_m \end{aligned} \qquad \text{with} \qquad A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}. \tag{5}$$

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 $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$ for $\widehat{x} = (C, D)$. 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 $(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{b}$.

The line $C + Dt$ minimizes $e_1^2 + \cdots + e_m^2 = \|A\boldsymbol{x} - \boldsymbol{b}\|^2$ when $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$:

$$A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b} \qquad \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}. \tag{8}$$

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 (calculus):

$$E(\boldsymbol{x}) = \|A\boldsymbol{x} - \boldsymbol{b}\|^2 = (C + Dt_1 - b_1)^2 + \cdots + (C + Dt_m - b_m)^2.$$

Calculus sets the derivatives $\partial E/\partial C$ and $\partial E/\partial D$ to zero, and 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*: $(1, 1, 1)$ is orthogonal to $(-2, 0, 2)$:

$$\begin{aligned} C + D(-2) &= 1 \\ C + \quad D(0) &= 2 \\ C + \quad D(2) &= 4 \end{aligned} \qquad \text{or} \qquad A\boldsymbol{x} = \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}.$$

When the columns of A are orthogonal, $A^{\mathrm{T}}A$ will be a diagonal matrix:

$$A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b} \qquad \text{is} \qquad \begin{bmatrix} 3 & 0 \\ 0 & 8 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}. \tag{9}$$

Main point: Since $A^{\mathrm{T}}A$ *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 *shifting the times by subtracting the average time* $\widehat{t} = (t_1 + \cdots + t_m)/m$. If the original times were $1, 3, 5$ then their average is $\widehat{t} = 3$. The shifted times $T = t - \widehat{t} = t - 3$ add up to zero!

$$\begin{aligned} T_1 &= 1 - 3 = -2 \\ T_2 &= 3 - 3 = \ \ \, 0 \\ T_3 &= 5 - 3 = \ \ \, 2 \end{aligned} \qquad A_{\text{new}} = \begin{bmatrix} 1 & T_1 \\ 1 & T_2 \\ 1 & T_3 \end{bmatrix} \qquad A_{\text{new}}^{\mathrm{T}} A_{\text{new}} = \begin{bmatrix} 3 & \mathbf{0} \\ \mathbf{0} & 8 \end{bmatrix}.$$

Now C and D come from the easy equation (9). Then the best straight line uses $C + DT$ which is $C + D(t - \widehat{t}\,) = C + D(t - 3)$. Problem 30 even gives a formula for C and D.

That was a perfect example of the "Gram-Schmidt idea" coming in the next section: *Make the columns orthogonal in advance*. Then $A_{\text{new}}^{\mathrm{T}} A_{\text{new}}$ is diagonal and $\widehat{\boldsymbol{x}}_{\text{new}}$ is easy.

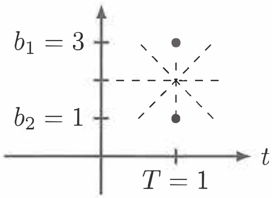
Dependent Columns in A: What is $\widehat{\boldsymbol{x}}$?

From the start, this chapter has assumed independent columns in A. Then $A^{\mathrm{T}}A$ is invertible. Then $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ produces the least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$.

Which $\widehat{\boldsymbol{x}}$ is best if A has *dependent columns*? Here is a specific example.

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \boldsymbol{b} \qquad \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \boldsymbol{p}$$

$$A\boldsymbol{x} = \boldsymbol{b} \qquad\qquad A\widehat{\boldsymbol{x}} = \boldsymbol{p}$$



The measurements $b_1 = 3$ and $b_2 = 1$ are at the same time T! A straight line $C + Dt$ cannot go through both points. I think we are right to project $\boldsymbol{b} = (3, 1)$ to $\boldsymbol{p} = (2, 2)$ in the column space of A. That changes the equation $A\boldsymbol{x} = \boldsymbol{b}$ to the equation $A\widehat{\boldsymbol{x}} = \boldsymbol{p}$. An equation with no solution has become an equation with infinitely many solutions. The problem is that A has dependent columns and $(1, -1)$ is in its nullspace.

Which solution $\widehat{\boldsymbol{x}}$ should we choose? All the dashed lines in the figure have the same two errors 1 and -1 at time T. Those errors $(1, -1) = \boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ are as small as possible. But this doesn't tell us which dashed line is best.

My instinct is to go for the horizontal line at height 2. If the equation for the best line is $b = C + Dt$, then my choice will have $\widehat{x}_1 = C = 2$ and $\widehat{x}_2 = D = 0$. But what if the line had been written as $b = ct + d$? This is equally correct (just reversing C and D). Now the horizontal line has $\widehat{x}_1 = c = 0$ and $\widehat{x}_2 = d = 2$. I don't see any way out.

In Section 7.4, the *"pseudoinverse" of* A will choose the **shortest solution to** $A\widehat{\boldsymbol{x}} = \boldsymbol{p}$. Here, that shortest solution will be $\boldsymbol{x}^+ = (1, 1)$. This is the particular solution in the row space of A, and $\boldsymbol{x}^+$ has length $\sqrt{2}$. (Both solutions $\widehat{\boldsymbol{x}} = (2, 0)$ and $(0, 2)$ have length 2.) We are arbitrarily choosing the nullspace component of the solution $\boldsymbol{x}^+$ to be zero.

When A has independent columns, the nullspace only contains the zero vector and the pseudoinverse is our usual left inverse $L = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$. When I write it that way, the pseudoinverse sounds like the best way to choose $\boldsymbol{x}$.

Comment MATLAB experiments with singular matrices produced either **Inf** or **NaN** (Not a Number) or $\mathbf{10^{16}}$ (a bad number). There is a warning in every case! I believe that **Inf** and **NaN** and $\mathbf{10^{16}}$ come from the possibilities $\mathbf{0}\boldsymbol{x} = \boldsymbol{b}$ and $\mathbf{0}\boldsymbol{x} = \mathbf{0}$ and $\mathbf{10^{-16}}\,\boldsymbol{x} = \mathbf{1}$.

Those are three small examples of three big difficulties: singular with no solution, singular with many solutions, and very very close to singular.

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 its orbit. But even with a nonlinear function like t^2, the unknowns C, D, E still appear linearly! Fitting points by 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:

$$\begin{aligned} C + Dt_1 + Et_1^2 &= b_1 \\ &\vdots \\ C + Dt_m + Et_m^2 &= b_m \end{aligned} \qquad \text{is } A\boldsymbol{x} = \boldsymbol{b} \text{ with the } m \text{ by 3 matrix} \qquad A = \begin{bmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 \end{bmatrix}. \tag{10}$$

Least squares The closest parabola $C + Dt + Et^2$ chooses $\widehat{\boldsymbol{x}} = (C, D, E)$ to satisfy the three normal equations $A^{\mathrm{T}} A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$.

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 $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$, 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 $\widehat{\boldsymbol{x}} = (C, D, E)$ solves the 3 by 3 system of equations $A^{\mathrm{T}} A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$.

Section 10.5 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 for C, D, E 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 $\boldsymbol{x} = (C, D, E) = (\mathbf{6}, \mathbf{-9}, \mathbf{3})$. The parabola through the three points in Figure 4.8a is $b = \mathbf{6 - 9t + 3t^2}$.

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 $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$, 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 $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$. Will the least squares parabola stay the same, with all the error at the fourth point? Not likely!

Least squares balances the four errors to get three equations for C, D, E.

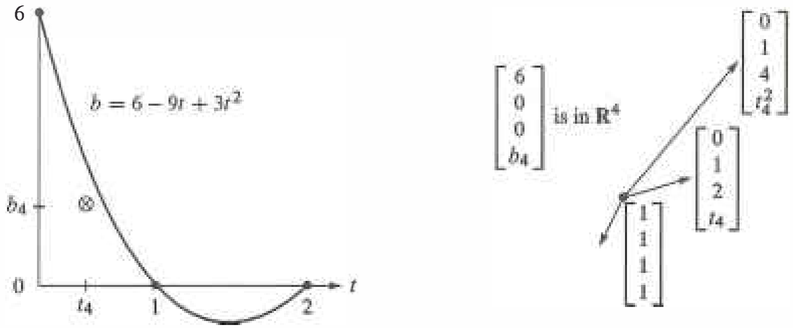


Figure 4.8: An exact fit of the parabola at $t = 0, 1, 2$ means that $\boldsymbol{p} = \boldsymbol{b}$ and $\boldsymbol{e} = \mathbf{0}$. The fourth point ⓧ off the parabola makes $m > n$ and we need least squares: project $\boldsymbol{b}$ on $\boldsymbol{C}(\boldsymbol{A})$. The figure on the right shows $\boldsymbol{b}$—not a combination of the three columns of A.

■ REVIEW OF THE KEY IDEAS ■

1. The least squares solution $\widehat{\boldsymbol{x}}$ minimizes $\|A\boldsymbol{x} - \boldsymbol{b}\|^2 = \boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{x} - 2\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}\boldsymbol{b} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}$. This is E, the sum of squares of the errors in the m equations $(m > n)$.

2. The best $\widehat{\boldsymbol{x}}$ comes from the normal equations $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$. E is a minimum.

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 $\boldsymbol{p} = (p_1, \ldots, p_m)$. The vertical distances to the data points are the errors $\boldsymbol{e} = (e_1, \ldots, e_m)$. A key equation is $A^{\mathrm{T}}\boldsymbol{e} = \mathbf{0}$.

5. If we try to fit m points by a combination of $n < m$ functions, the m equations $A\boldsymbol{x} = \boldsymbol{b}$ are generally unsolvable. The n equations $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ 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 options 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 $\boldsymbol{C = 4}$:

$$A = \text{column of 1's} \quad A^{\mathrm{T}}A = 10 \quad A^{\mathrm{T}}\boldsymbol{b} = \text{sum of } b_i = 40. \quad \text{So } 10\,C = 40.$$

(2) The least maximum error requires $\boldsymbol{C = 20}$, halfway between 0 and 40.

(3) The least sum requires $\boldsymbol{C = 0}$ (!!). 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 line $C + Dt$ through those ten points $(1,0)$ to $(10,40)$:

$$A^{\mathrm{T}}A = \begin{bmatrix} 10 & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} = \begin{bmatrix} 10 & 55 \\ 55 & 385 \end{bmatrix} \qquad A^{\mathrm{T}}\boldsymbol{b} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix} = \begin{bmatrix} 40 \\ 400 \end{bmatrix}$$

Those come from equation (8). Then $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ gives $C = -8$ and $D = 24/11$.

What happens to C and D if you multiply $\boldsymbol{b} = (0, 0, \ldots, 40)$ by 3 and then add 30 to get $\boldsymbol{b}_{\text{new}} = (30, 30, \ldots, 150)$? Linearity allows us to rescale $\boldsymbol{b}$. Multiplying $\boldsymbol{b}$ 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:

$$\begin{aligned} C + D\,(-2) + E\,(-2)^2 &= 0 \\ C + D\,(-1) + E\,(-1)^2 &= 0 \\ C + D\quad (0) + E\quad (0)^2 &= 1 \\ C + D\quad (1) + E\quad (1)^2 &= 0 \\ C + D\quad (2) + E\quad (2)^2 &= 0 \end{aligned} \qquad A = \begin{bmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \qquad A^{\mathrm{T}}A = \begin{bmatrix} 5 & \mathbf{0} & 10 \\ \mathbf{0} & 10 & \mathbf{0} \\ 10 & \mathbf{0} & 34 \end{bmatrix}$$

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 from C and E:

$$\begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix}\begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{leads to} \quad \begin{array}{l} C = 34/70 \\ D = 0 \ \text{ as predicted} \\ E = -10/70 \end{array}$$

Problem Set 4.3

Problems 1–11 use four data points $\boldsymbol{b} = (0,8,8,20)$ to bring out the key ideas.

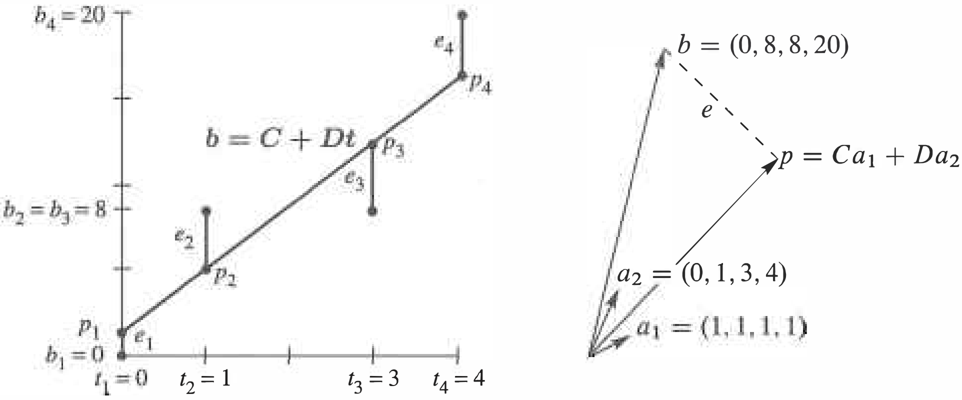


Figure 4.9: **Problems 1–11**: The closest line $C+Dt$ matches $C\boldsymbol{a}_1 + D\boldsymbol{a}_2$ in $\mathbf{R}^4$.

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 $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}\boldsymbol{b}$. 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 $A\boldsymbol{x} = \boldsymbol{b}$ in three unknowns $\boldsymbol{x} = (C, D, E)$. Set up the three normal equations $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ (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 $A\boldsymbol{x} = \boldsymbol{b}$. 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 $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$.

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*** W in Section 10.2.

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 $P\boldsymbol{e} = \mathbf{0}$?

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 $\boldsymbol{p} = \mathbf{0}$.

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 $\boldsymbol{e} = \mathbf{0}$ 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 $0, \boldsymbol{a}_1, \boldsymbol{a}_2$ unless _____.) What is the test to have exactly one plane in $\mathbf{R}^n$?

30 Example 2 shifted the times t_i to make them add to zero. We subtracted away the average time $\widehat{t} = (t_1 + \cdots + t_m)/m$ to get $T_i = t_i - \widehat{t}$. Those T_i add to zero.

With the columns $(1, \ldots, 1)$ and $(T_1, \ldots, T_m)$ now orthogonal, $A^{\mathrm{T}}A$ is diagonal. Its entries are m and $T_1^2 + \cdots + T_m^2$. Show that the best C and D have direct formulas:

$$\boldsymbol{T} \textbf{ is } \boldsymbol{t} - \widehat{\boldsymbol{t}} \qquad C = \frac{b_1 + \cdots + b_m}{m} \qquad \text{and} \qquad D = \frac{b_1T_1 + \cdots + b_mT_m}{T_1^2 + \cdots + T_m^2}.$$

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 of A in advance*.

4.4 Orthonormal Bases and Gram-Schmidt

1 The columns $q_1, \ldots, q_n$ are orthonormal if $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_j = \begin{cases} 0 \text{ for } i \neq j \\ 1 \text{ for } i = j \end{cases}$. Then $\boxed{Q^{\mathrm{T}}Q = I\,.}$

2 If Q is also square, then $QQ^{\mathrm{T}} = I$ and $\boxed{Q^{\mathrm{T}} = Q^{-1}}$. Q is an "orthogonal matrix".

3 The least squares solution to $Q\boldsymbol{x} = \boldsymbol{b}$ is $\widehat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b}$. Projection of $\boldsymbol{b}$: $\boldsymbol{p} = QQ^{\mathrm{T}}\boldsymbol{b} = P\boldsymbol{b}$.

4 The **Gram-Schmidt** process takes independent $\boldsymbol{a}_i$ to orthonormal $\boldsymbol{q}_i$. Start with $\boldsymbol{q}_1 = \boldsymbol{a}_1 / \,||\boldsymbol{a}_1||$.

5 $\boldsymbol{q}_i$ is $(\boldsymbol{a}_i -$ projection $\boldsymbol{p}_i) \,/\, ||\boldsymbol{a}_i - \boldsymbol{p}_i||$; projection $\boldsymbol{p}_i = (\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{q}_1)\boldsymbol{q}_1 + \cdots + (\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{q}_{i-1})\boldsymbol{q}_{i-1}$.

6 Each $\boldsymbol{a}_i$ will be a combination of $\boldsymbol{q}_1$ to $\boldsymbol{q}_i$. Then $\boldsymbol{A} = \boldsymbol{QR}$: orthogonal Q and triangular R.

This section has two goals, **why** and **how**. The first is to see why orthogonality is good. Dot products are zero, so $A^{\mathrm{T}}A$ will be diagonal. It becomes so easy to find $\widehat{\boldsymbol{x}}$ and $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$. ***The second goal is to construct orthogonal vectors.*** You will see how Gram-Schmidt chooses combinations of the original basis vectors to produce right angles. Those original vectors are the columns of A, probably *not* orthogonal. ***The orthonormal basis 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 (normal). 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 $\boxed{Q^{\mathrm{T}}Q = I}$:

$$\boxed{Q^{\mathrm{T}}Q} = \begin{bmatrix} - \boldsymbol{q}_1^{\mathrm{T}} - \\ - \boldsymbol{q}_2^{\mathrm{T}} - \\ \\ - \boldsymbol{q}_n^{\mathrm{T}} - \end{bmatrix} \begin{bmatrix} | & | & & | \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & & \boldsymbol{q}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \boxed{I}. \quad (1)$$

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). This diagonal matrix is almost as good as I. 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 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 $-\theta$ equals 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*: $Q^{-1} = Q^{\mathrm{T}}$. The inverse puts the components back into their original order:

[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.

$$\textbf{Inverse = transpose:} \quad \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ z \\ x \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

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:

$$\boxed{Q^{\mathrm{T}} = I - 2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}} = Q} \quad \text{and} \quad Q^{\mathrm{T}}Q = I - 4\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}} + 4\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}} = I. \qquad (2)$$

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, like $(-1)^2 = 1$. 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).

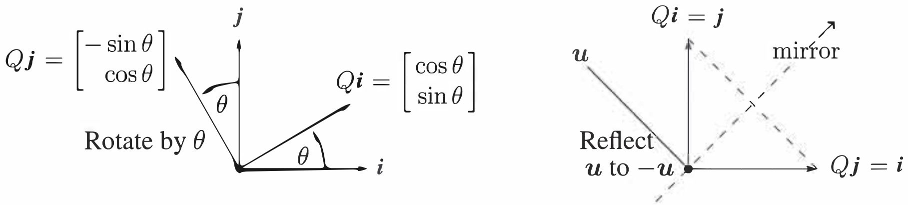


Figure 4.10: Rotation by $Q = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}$ and reflection across 45° by $Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

As example choose the direction $\boldsymbol{u} = (-1/\sqrt{2}, 1/\sqrt{2})$. Compute $2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ (column times row) and subtract from I to get the reflection matrix Q in the direction of $\boldsymbol{u}$:

$$\textbf{Reflection} \quad Q = I - 2\begin{bmatrix} .5 & -.5 \\ -.5 & .5 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}.$$

When (x, y) goes to (y, x), a vector like $(3, 3)$ doesn't move. It is on the mirror line.

Rotations preserve the length of every vector. So do reflections. So do permutations. So does multiplication by any orthogonal matrix Q—***lengths and angles don't change***.

Proof $\|Qx\|^2$ equals $\|x\|^2$ because $(Qx)^{\mathrm{T}}(Qx) = x^{\mathrm{T}}Q^{\mathrm{T}}Qx = x^{\mathrm{T}}Ix = x^{\mathrm{T}}x$.

If Q has orthonormal columns ($Q^{\mathrm{T}}Q = I$), it leaves lengths unchanged:

Same length for Qx $\quad$ $\|Qx\| = \|x\|$ **for every vector** x. $\qquad (3)$

Q also preserves dot products: $(Qx)^{\mathrm{T}}(Qy) = x^{\mathrm{T}}Q^{\mathrm{T}}Qy = x^{\mathrm{T}}y$. Just use $Q^{\mathrm{T}}Q = I$!

Projections Using Orthonormal Bases: Q Replaces A

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.

For projections onto subspaces, all formulas involve $A^{\mathrm{T}}A$. The entries of $A^{\mathrm{T}}A$ are the dot products $\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{a}_j$ of the basis vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$.

Suppose the basis vectors are actually orthonormal. The $\boldsymbol{a}$'s become the $\boldsymbol{q}$'s. Then $A^{\mathrm{T}}A$ *simplifies to* $Q^{\mathrm{T}}Q = I$. Look at the improvements in $\widehat{\boldsymbol{x}}$ and $\boldsymbol{p}$ and P. Instead of $Q^{\mathrm{T}}Q$ we print a blank for the identity matrix:

$$\underline{\qquad}\,\widehat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b} \quad \text{and} \quad \boldsymbol{p} = Q\widehat{\boldsymbol{x}} \quad \text{and} \quad P = Q\,\underline{\qquad}\,Q^{\mathrm{T}}. \tag{4}$$

The least squares solution of $Q\boldsymbol{x} = \boldsymbol{b}$ is $\widehat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b}$. The projection matrix is QQ^{T}.

There are no matrices to invert. This is the point of an orthonormal basis. The best $\widehat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b}$ just has dot products of $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ with $\boldsymbol{b}$. We have 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 $\boldsymbol{p} = Q\widehat{\boldsymbol{x}} = QQ^{\mathrm{T}}\boldsymbol{b}$:

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 $\widehat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b}$ 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 $\boldsymbol{p} = \boldsymbol{b}$ and $P = QQ^{\mathrm{T}} = I$.

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, then Q is square. Every $\boldsymbol{b} = QQ^{\mathrm{T}}\boldsymbol{b}$ *is the sum of its components along the* $\boldsymbol{q}$'s:

$$\boldsymbol{b} = \boldsymbol{q}_1(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b}) + \boldsymbol{q}_2(\boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b}) + \cdots + \boldsymbol{q}_n(\boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{b}). \tag{6}$$

Transforms $\;QQ^{\mathrm{T}} = I$ is the foundation of Fourier series and all the great "transforms" of applied mathematics. They break vectors $\boldsymbol{b}$ or functions $f(x)$ into perpendicular pieces. Then by adding the pieces in (6), the inverse transform puts $\boldsymbol{b}$ and $f(x)$ back together.

Example 4 The columns of this orthogonal Q are orthonormal vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$:

$$\boldsymbol{m = n = 3} \qquad Q = \frac{1}{3}\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix} \quad \text{has} \quad \boldsymbol{Q^{\mathrm{T}}Q = QQ^{\mathrm{T}} = I}.$$

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 $\boldsymbol{p}_1+\boldsymbol{p}_2+\boldsymbol{p}_3=\boldsymbol{b}$ itself:

Reconstruct $\boldsymbol{b}$
$\boldsymbol{b}=\boldsymbol{p}_1+\boldsymbol{p}_2+\boldsymbol{p}_3$

$$\frac{2}{3}\boldsymbol{q}_1+\frac{2}{3}\boldsymbol{q}_2-\frac{1}{3}\boldsymbol{q}_3=\frac{1}{9}\begin{bmatrix}-2+4-2\\4-2-2\\4+4+1\end{bmatrix}=\begin{bmatrix}0\\0\\1\end{bmatrix}=\boldsymbol{b}.$$

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 $\widehat{\boldsymbol{x}}$ 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 explain the "Gram-Schmidt 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 may be 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 as it comes. The next direction $\boldsymbol{B}$ must be perpendicular to $\boldsymbol{A}$. ***Start with $\boldsymbol{b}$ and subtract its projection along $\boldsymbol{A}$.*** 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. Multiply equation (7) by $\boldsymbol{A}^{\mathrm{T}}$ 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 a perpendicular direction $\boldsymbol{C}$:

Next Gram-Schmidt step
$$\boldsymbol{C}=\boldsymbol{c}-\frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c}}{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}}\boldsymbol{A}-\frac{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{c}}{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}}\boldsymbol{B}. \tag{8}$$

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 $\boldsymbol{D}$.

[2] I think Gram had the idea. I don't really know where Schmidt came in.

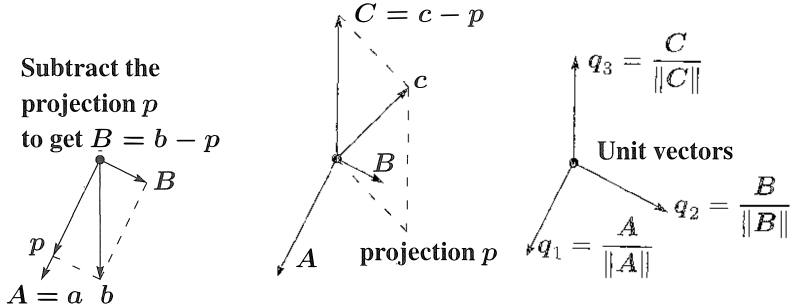


Figure 4.11: First project b onto the line through a and find the orthogonal B as $b - p$. Then project c onto the AB plane and find C as $c - p$. Divide by $\|A\|$, $\|B\|$, $\|C\|$.

At the end, *or immediately when each one is found*, divide the orthogonal vectors A, B, C, D by their lengths. The resulting vectors q_1, q_2, q_3, q_4 are orthonormal.

Example of **Gram-Schmidt** Suppose the independent non-orthogonal vectors a, b, c are

$$a = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} \quad \text{and} \quad c = \begin{bmatrix} 3 \\ -3 \\ 3 \end{bmatrix}.$$

Then $A = a$ has $A^{\mathrm{T}}A = 2$ and $A^{\mathrm{T}}b = 2$. Subtract from b its projection p along A:

First step $$B = b - \frac{A^{\mathrm{T}}b}{A^{\mathrm{T}}A}A = b - \frac{2}{2}A = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}.$$

Check: $A^{\mathrm{T}}B = 0$ as required. Now subtract the projections of c on A and B to get C:

Next step $$C = c - \frac{A^{\mathrm{T}}c}{A^{\mathrm{T}}A}A - \frac{B^{\mathrm{T}}c}{B^{\mathrm{T}}B}B = c - \frac{6}{2}A + \frac{6}{6}B = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Check: $C = (1, 1, 1)$ is perpendicular to both A and B. Finally convert A, B, C to unit vectors (length 1, orthonormal). The lengths of A, B, C are $\sqrt{2}$ and $\sqrt{6}$ and $\sqrt{3}$. Divide by those lengths, for an orthonormal basis:

$$q_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad q_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \quad \text{and} \quad q_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Usually A, B, C contain fractions. Almost always q_1, q_2, 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:

- The vectors $\boldsymbol{a}$ and $\boldsymbol{A}$ and $\boldsymbol{q}_1$ are all along a single line.
- The vectors $\boldsymbol{a}, \boldsymbol{b}$ and $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{q}_1, \boldsymbol{q}_2$ are all in the same plane.
- The vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ and $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ 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 $A = QR$:

$$\begin{bmatrix} & & \\ \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \\ & & \end{bmatrix} = \begin{bmatrix} & & \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \\ & & \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{c} \\ & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{c} \\ & & \boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{c} \end{bmatrix} \quad \text{or} \quad \boxed{\boldsymbol{A} = \boldsymbol{QR}.} \tag{9}$$

$A = QR$ is Gram-Schmidt in a nutshell. Multiply by Q^{T} to recognize $\boldsymbol{R} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}$ above.

(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 original $\boldsymbol{a}$'s and the final $\boldsymbol{q}$'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. The dot products $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{a}_j$ go into R. **Then $\boldsymbol{A} = \boldsymbol{QR}$:**

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & -3 \\ 0 & -2 & 3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & -2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \boldsymbol{\sqrt{2}} & \sqrt{2} & \sqrt{18} \\ \mathbf{0} & \boldsymbol{\sqrt{6}} & -\sqrt{6} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\sqrt{3}} \end{bmatrix} = QR.$$

Look closely at Q and R. The lengths of $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ are $\sqrt{2}, \sqrt{6}, \sqrt{3}$ on the diagonal of R. The columns of Q are orthonormal. Because of the square roots, QR might look harder 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 $A = QR$. 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: $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = (\boldsymbol{QR})^{\mathrm{T}}\boldsymbol{QR} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{QR} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$. The least squares equation $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ simplifies to $R^{\mathrm{T}}R\widehat{\boldsymbol{x}} = R^{\mathrm{T}}Q^{\mathrm{T}}\boldsymbol{b}$. Then finally we reach $R\widehat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b}$: good.

$$\textbf{Least squares} \qquad R^{\mathrm{T}}R\widehat{\boldsymbol{x}} = R^{\mathrm{T}}Q^{\mathrm{T}}\boldsymbol{b} \text{ or } R\widehat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b} \text{ or } \widehat{\boldsymbol{x}} = R^{-1}Q^{\mathrm{T}}\boldsymbol{b} \tag{10}$$

Instead of solving $A\boldsymbol{x} = \boldsymbol{b}$, which is impossible, we solve $R\widehat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b}$ 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 with $A = QR$.

Below is an informal code. It executes equations (11) for $j = 1$ then $j = 2$ and eventually $j = n$. The important lines 4-5 subtract from $\boldsymbol{v} = \boldsymbol{a}_j$ its projection onto each $\boldsymbol{q}_i, i < j$. The last line of that code normalizes $\boldsymbol{v}$ (divides by $r_{jj} = ||\boldsymbol{v}||$) to get the unit vector $\boldsymbol{q}_j$:

$$r_{kj} = \sum_{i=1}^{m} q_{ik}v_{ij} \text{ and } v_{ij} = v_{ij} - q_{ik}r_{kj} \text{ and } r_{jj} = \left(\sum_{i=1}^{m} v_{ij}^2\right)^{1/2} \text{ and } q_{ij} = \frac{v_{ij}}{r_{jj}}. \tag{11}$$

Starting from $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} = \boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ this code will construct $\boldsymbol{q}_1$, then $\boldsymbol{B}$, $\boldsymbol{q}_2$, then $\boldsymbol{C}$, $\boldsymbol{q}_3$:

$$\boldsymbol{q}_1 = \boldsymbol{a}_1/\|\boldsymbol{a}_1\| \qquad \boldsymbol{B} = \boldsymbol{a}_2 - (\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a}_2)\boldsymbol{q}_1 \qquad \boldsymbol{q}_2 = \boldsymbol{B}/\|\boldsymbol{B}\|$$

$$\boldsymbol{C}^* = \boldsymbol{a}_3 - (\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a}_3)\boldsymbol{q}_1 \qquad \boldsymbol{C} = \boldsymbol{C}^* - (\boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{C}^*)\boldsymbol{q}_2 \qquad \boldsymbol{q}_3 = \boldsymbol{C}/\|\boldsymbol{C}\|$$

Equation (11) subtracts **one projection at a time** as in $\boldsymbol{C}^*$ and $\boldsymbol{C}$. That change is called ***modified Gram-Schmidt***. This code 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 the original A
  for i = 1:j-1                    % columns q_1 to q_{j-1} are 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 v) q_i
  end                              % v is now perpendicular to all of q_1, ..., q_{j-1}
  R(j, j) = norm(v);               % the diagonal entries R_jj are lengths
  Q(:, j) = v/R(j, j);             % divide v by its length to get the next q_j
end                                % the "for j = 1 : n loop" produces all of the q_j
```

To recover column j of A, undo the last step and the middle steps of the code:

$$R(j,j)\boldsymbol{q}_j = (\boldsymbol{v} \textit{ minus its projections}) = (\text{column } j \text{ of } A) - \sum_{i=1}^{j-1} R(i,j)\boldsymbol{q}_i\,. \tag{12}$$

Moving the sum to the far left, this is column j in the multiplication $QR = A$.

Confession Good software like LAPACK, used in good systems like MATLAB and Julia and Python, will not use this Gram-Schmidt code. There is now a better way. "Householder reflections" act on A to produce the upper triangular R. This happens one column at a time in the same way that elimination produces the upper triangular U in LU.

Those reflection matrices $I - 2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ will be described in Chapter 11 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 to orthogonalize A 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 the matrix multiplication $Q^{\mathrm{T}}Q = I$.
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 of Q spanned by the $\boldsymbol{q}$'s is $P = QQ^{\mathrm{T}}$.
5. If Q is square then $P = QQ^{\mathrm{T}} = I$ 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. A 5 by 5 Hadamard matrix 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 $\sqrt{8}$.

$$H_8^{\mathrm{T}} H_8 = \begin{bmatrix} H^{\mathrm{T}} & H^{\mathrm{T}} \\ H^{\mathrm{T}} & -H^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} H & H \\ H & -H \end{bmatrix} = \begin{bmatrix} 2H^{\mathrm{T}}H & 0 \\ 0 & 2H^{\mathrm{T}}H \end{bmatrix} = \begin{bmatrix} 8I & 0 \\ 0 & 8I \end{bmatrix}. \quad Q_8 = \frac{H_8}{\sqrt{8}}$$

4.4 B What is the key point of orthogonal columns? Answer: $A^{\mathrm{T}}A$ is diagonal and easy to invert. **We can project onto lines and just add**. The axes are orthogonal.

Add p's Projection $\boldsymbol{p}_{1,2}$ onto a plane equals $\boldsymbol{p}_1 + \boldsymbol{p}_2$ onto orthogonal lines.

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 $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$. 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 $\boldsymbol{q}_1 = (1, 1, 1)/\sqrt{3}$.

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 $Q^{\mathrm{T}}Q = I$.)

7 If Q has orthonormal columns, what is the least squares solution $\widehat{x}$ to $Qx = b$?

8 If $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ are orthonormal vectors in $\mathbf{R}^5$, what combination ____ $q_1 +$ ____ q_2 is closest to a given vector $\boldsymbol{b}$?

9 (a) Compute $P = QQ^{\mathrm{T}}$ when $\boldsymbol{q}_1 = (.8, .6, 0)$ and $\boldsymbol{q}_2 = (-.6, .8, 0)$. Verify that $P^2 = P$.

(b) Prove that always $(QQ^{\mathrm{T}})^2 = QQ^{\mathrm{T}}$ by using $Q^{\mathrm{T}}Q = I$. Then $P = QQ^{\mathrm{T}}$ is the projection matrix onto the column space of Q.

10 Orthonormal vectors are automatically linearly independent.

(a) Vector proof: When $c_1\boldsymbol{q}_1 + c_2\boldsymbol{q}_2 + c_3\boldsymbol{q}_3 = \mathbf{0}$, what dot product leads to $c_1 = 0$? Similarly $c_2 = 0$ and $c_3 = 0$. Thus the $\boldsymbol{q}$'s are independent.

(b) Matrix proof: Show that $Q\boldsymbol{x} = \mathbf{0}$ leads to $\boldsymbol{x} = \mathbf{0}$. Since Q may be rectangular, you can use Q^{T} but not Q^{-1}.

11 (a) Gram-Schmidt: Find orthonormal vectors $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ in the plane spanned by $\boldsymbol{a} = (1, 3, 4, 5, 7)$ and $\boldsymbol{b} = (-6, 6, 8, 0, 8)$.

(b) Which vector in this plane is closest to $(1, 0, 0, 0, 0)$?

12 If $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ is a basis for $\mathbf{R}^3$, any vector $\boldsymbol{b}$ can be written as

$$\boldsymbol{b} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + x_3\boldsymbol{a}_3 \qquad \text{or} \qquad \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \boldsymbol{b}.$$

(a) Suppose the $\boldsymbol{a}$'s are orthonormal. Show that $x_1 = \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{b}$.

(b) Suppose the $\boldsymbol{a}$'s are orthogonal. Show that $x_1 = \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_1$.

(c) If the $\boldsymbol{a}$'s are independent, x_1 is the first component of ____ times $\boldsymbol{b}$.

Problems 13–25 are about the Gram-Schmidt process and $A = QR$.

13 What multiple of $\boldsymbol{a} = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ should be subtracted from $\boldsymbol{b} = \left[\begin{smallmatrix} 4 \\ 0 \end{smallmatrix}\right]$ to make the result $\boldsymbol{B}$ orthogonal to $\boldsymbol{a}$? Sketch a figure to show $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{B}$.

14 Complete the Gram-Schmidt process in Problem 13 by computing $\boldsymbol{q}_1 = \boldsymbol{a}/\|\boldsymbol{a}\|$ and $\boldsymbol{q}_2 = \boldsymbol{B}/\|\boldsymbol{B}\|$ and factoring into QR:

$$\begin{bmatrix} 1 & 4 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 \end{bmatrix} \begin{bmatrix} \|\boldsymbol{a}\| & ? \\ 0 & \|\boldsymbol{B}\| \end{bmatrix}.$$

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 equation (11). It executes Gram-Schmidt.

26 Show why $\boldsymbol{C}$ (found via $\boldsymbol{C}^*$ in the steps after (11)) 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 equation (11)?

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 $A = [Q \quad \boldsymbol{a}]$. Gram-Schmidt replaces $\boldsymbol{a}$ by what vector $\boldsymbol{q}$? Start with $\boldsymbol{a}$, subtract _____, divide by _____ to find $\boldsymbol{q}$.

Chapter 5

Determinants

1 The **determinant** of $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $\boldsymbol{ad - bc}$. Singular matrix $A = \begin{bmatrix} a & xa \\ c & xc \end{bmatrix}$ has $\textbf{det} = \mathbf{0}$.

2 **Row exchange reverses signs** $PA = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} c & d \\ a & b \end{bmatrix}$ has det $PA = bc - ad = \boldsymbol{-} \textbf{det}\, \boldsymbol{A}$.

3 The determinant of $\begin{bmatrix} xa + yA & xb + yB \\ c & d \end{bmatrix}$ is $x(ad - bc) + y(Ad - Bc)$. **Det is linear in row 1 by itself.**

4 **Elimination** $EA = \begin{bmatrix} a & b \\ 0 & d - \frac{c}{a} b \end{bmatrix}$ $\textbf{det}\, \boldsymbol{EA} = a\left(d - \frac{c}{a} b\right) =$ **product of pivots** $=$ **det** $\boldsymbol{A}$.

5 If A is n by n then $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}$ remain true: $\textbf{det} = \mathbf{0}$ when A is singular, **det reverses sign** when rows are exchanged, **det is linear in row 1 by itself**, det $=$ **product of the pivots**. Always $\textbf{det}\, \boldsymbol{BA} = (\textbf{det}\, \boldsymbol{B})(\textbf{det}\, \boldsymbol{A})$ and $\textbf{det}\, \boldsymbol{A}^{\mathrm{T}} = \textbf{det}\, \boldsymbol{A}$. This is an amazing number.

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. Notice the division by the determinant!

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{has inverse} \quad A^{-1} = \frac{\mathbf{1}}{\boldsymbol{ad - bc}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{1}$$

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 $\left[\begin{smallmatrix} \mathbf{c} & \mathbf{d} \\ \mathbf{a} & \mathbf{b} \end{smallmatrix}\right]$ 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 determinant 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 rules 1, 2, 3 lead to the answer we expect:

$$\text{The determinant of} \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{is} \quad \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

From rules 1–3 we will reach rules 4–10. The last two 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 A.

Rule 1 (the easiest) matches $\det I = 1$ with volume $= 1$ for a unit cube.

1 ***The determinant of the n by n identity matrix is* 1.**

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \quad \text{and} \quad \begin{vmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{vmatrix} = 1.$$

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 **det is multiplied by t**	$\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix}$
add row 1 of A to row 1 of A': **then determinants add**	$\begin{vmatrix} a+a' & b+b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}.$

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 **one row changes**.

$$A = \begin{vmatrix} 4 & 8 & 8 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix} = 4 \begin{vmatrix} 1 & 2 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} 4 & 8 & 8 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 4 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix} + \begin{vmatrix} 0 & 8 & 8 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix}.$$

By itself, rule 3 does not say what those determinants are ($\det A$ is 4).

Combining multiplication and addition, we get *any linear combination in one row*. Rule 2 for row exchanges can put that row 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:

$$\begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 2^2 = 4 \quad \text{and} \quad \begin{vmatrix} t & 0 \\ 0 & t \end{vmatrix} = t^2.$$

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, because 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 (like elimination) 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 $-\ell \begin{vmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{a} & \mathbf{b} \end{vmatrix}$. This extra term is zero by rule 4: equal rows. 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 $\begin{vmatrix} 0 & 0 \\ c & d \end{vmatrix} = 0 \quad \text{and} \quad \begin{vmatrix} a & b \\ 0 & 0 \end{vmatrix} = 0.$

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 are nonzero. Remove the off-diagonal entries by elimination! (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 $\det A = 0$. We reach the great test for **singular or invertible** matrices.

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 $d - (c/a)b$:

The determinant is $$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ 0 & d-(c/a)b \end{vmatrix} = ad - bc.$$

This is the first formula for the determinant. MATLAB multiplies the pivots to find $\det A$. The sign in $\pm \det U$ depends on whether the number of row exchanges is even or odd: $+1$ or -1 is the determinant of the permutation P that exchanges rows.

With no row exchanges, $P = I$ and $\det A = \det U =$ *product of pivots*. And $\det L = 1$:

$$\text{If} \quad PA = LU \quad \text{then} \quad \det P \det A = \det L \det U \quad \text{and} \quad \det A = \pm \det U. \tag{3}$$

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 $\mathbf{1/\det A}$:

A times A^{-1} $\quad$ $\boldsymbol{AA^{-1} = I}$ **so** $\quad$ $\boldsymbol{(\det A)(\det A^{-1}) = \det I = 1.}$

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 $\boldsymbol{D(A) = |AB|/|B|}$. *Check that this ratio $D(A)$ has properties* 1,2,3. Then $D(A)$ has to be the determinant and we have $|AB|/|B| = |A|$. Good.

Property 1 (*Determinant of I*) If $A = I$ then the ratio $D(A)$ 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 the determinant $|AB|$ by t. So the ratio $|AB|/|B|$ is multiplied by t.

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.***

Transpose $\quad \begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix}$ $\quad$ 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:

Compare $\quad \det P \det A = \det L \det U \quad$ 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?

$$\begin{bmatrix} a_{11} & -a_{12} & a_{13} \\ -a_{21} & a_{22} & -a_{23} \\ a_{31} & -a_{32} & a_{33} \end{bmatrix} \qquad \begin{bmatrix} \text{row } 1 - \text{row } 3 \\ \text{row } 2 - \text{row } 1 \\ \text{row } 3 - \text{row } 2 \end{bmatrix} \qquad \begin{bmatrix} \text{row } 1 + \text{row } 3 \\ \text{row } 2 + \text{row } 1 \\ \text{row } 3 + \text{row } 2 \end{bmatrix}$$

Solution The three determinants are $\det A$, 0, and $2\det A$. Here are reasons:

$$M_1 = \begin{bmatrix} 1 & & \\ & -1 & \\ & & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} 1 & & \\ & -1 & \\ & & 1 \end{bmatrix} \qquad \text{so } \det M_1 = (-1)(\det A)(-1).$$

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 separately):

$$\begin{vmatrix} \text{row } 1 + \text{row } 3 \\ \text{row } 2 + \text{row } 1 \\ \text{row } 3 + \text{row } 3 \end{vmatrix} = \begin{vmatrix} \text{row } 1 \\ \text{row } 2 \\ \text{row } 3 \end{vmatrix} + \begin{vmatrix} \text{row } 3 \\ \text{row } 2 \\ \text{row } 3 \end{vmatrix} + \begin{vmatrix} \text{row } 1 \\ \text{row } 1 \\ \text{row } 3 \end{vmatrix} + \cdots + \begin{vmatrix} \text{row } 3 \\ \text{row } 1 \\ \text{row } 2 \end{vmatrix}.$$

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:

$$\det\begin{bmatrix} 1-a & 1 & 1 \\ 1 & 1-a & 1 \\ 1 & 1 & 1-a \end{bmatrix} = a^2(3-a). \qquad (4)$$

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 $\det\left(\frac{1}{2}A\right)$ 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 $A\boldsymbol{x} = \boldsymbol{0}$ 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 :

$$A=\begin{bmatrix}1\\2\\3\end{bmatrix}\begin{bmatrix}1 & -4 & 5\end{bmatrix} \quad\text{and}\quad A=\begin{bmatrix}0 & 1 & 3\\ -1 & 0 & 4\\ -3 & -4 & 0\end{bmatrix}.$$

17 A skew-symmetric matrix has $A^{\mathrm{T}}=-A$. Insert a, b, c for $1, 3, 4$ in Question 16 and show that $|A|=0$. Write down a 4 by 4 example with $|A|=1$.

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?

$$\text{From } \det A = \begin{vmatrix} \text{row } 1 \\ \text{row } 2 \\ \text{row } 3 \end{vmatrix} = 6 \text{ find } \det B = \begin{vmatrix} \text{row } 3 + \text{row } 2 + \text{row } 1 \\ \text{row } 2 + \text{row } 1 \\ \text{row } 1 \end{vmatrix}.$$

5.2 Permutations and Cofactors

1 **2 by 2**: $ad - bc$ has $2\,!$ terms with $\pm$ signs. **n by n: det A adds $n!$ terms with $\pm$ signs.**

2 For $n = 3$, $\det A$ adds $3! = 6$ terms. Two terms are $+a_{12}a_{23}a_{31}$ and $-a_{13}a_{22}a_{31}$. **Rows $1, 2, 3$ and columns $1, 2, 3$ appear once in each term.**

3 That minus sign came because the column order $3, 2, 1$ needs one exchange to recover $1, 2, 3$.

4 The six terms include $+a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} = a_{11}(\boldsymbol{a_{22}a_{33} - a_{23}a_{32}}) = a_{11}(\textbf{cofactor } \boldsymbol{C_{11}})$.

5 Always $\det A = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}$. Cofactors are determinants of size $n - 1$.

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

When elimination leads to $A = LU$, the pivots $d_1, \ldots, d_n$ are 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 Section 5.1, with the further possibility of row exchanges. Then a permutation enters $PA = LU$. The determinant of P is -1 or $+1$.

$$(\det \boldsymbol{P})(\det \boldsymbol{A}) = (\det \boldsymbol{L})(\det \boldsymbol{U}) \quad \text{gives} \quad \boxed{\det A = \pm(d_1 d_2 \cdots d_n).} \tag{2}$$

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:

$$\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & \cdot & \\ & & \cdot & \cdot & -1 \\ & & & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ -\frac{1}{2} & 1 & & & \\ & -\frac{2}{3} & 1 & & \\ & & \cdot & \cdot & \\ & & & -\frac{n-1}{n} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{2} & -1 & & & \\ & \mathbf{\frac{3}{2}} & -1 & & \\ & & \mathbf{\frac{4}{3}} & -1 & \\ & & & \cdot & \cdot \\ & & & & \mathbf{\frac{n+1}{n}} \end{bmatrix}$$

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 $n+1$*:

$$\boldsymbol{-1, 2, -1} \textbf{ matrix} \qquad \det A = (\mathbf{2})\left(\tfrac{\mathbf{3}}{\mathbf{2}}\right)\left(\tfrac{\mathbf{4}}{\mathbf{3}}\right)\cdots\left(\tfrac{\mathbf{n+1}}{\mathbf{n}}\right) = \boldsymbol{n+1}.$$

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*** $d_1 d_2 \cdots d_k$ ***(first k pivots)***.

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 multiplies all n pivots.

Elimination deals with the matrix A_k in the upper left corner 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. ***Each pivot is a ratio of determinants***:

$$\textbf{Pivots from determinants} \qquad \textbf{The } k\textbf{th pivot is } \boldsymbol{d_k} = \frac{d_1 d_2 \cdots d_k}{d_1 d_2 \cdots d_{k-1}} = \boxed{\frac{\det A_k}{\det A_{k-1}}}. \tag{3}$$

We don't need row exchanges when all the upper left 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:

$$\text{3 by 3 determinant} \quad \begin{vmatrix} a_{11} & a_{12} & \boldsymbol{a_{13}} \\ \boldsymbol{a_{21}} & a_{22} & a_{23} \\ a_{31} & \boldsymbol{a_{32}} & a_{33} \end{vmatrix} = \begin{matrix} +a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} \boldsymbol{+ a_{13}a_{21}a_{32}} \\ -a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}. \end{matrix} \tag{4}$$

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 (boldface). 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:

$$\begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} a & 0 \end{bmatrix} + \begin{bmatrix} 0 & b \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} c & d \end{bmatrix} = \begin{bmatrix} c & 0 \end{bmatrix} + \begin{bmatrix} 0 & d \end{bmatrix}.$$

Now apply linearity, first in row 1 (with row 2 fixed) and then in row 2 (with row 1 fixed):

$$\begin{aligned} \begin{vmatrix} a & b \\ c & d \end{vmatrix} &= \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix} && \textit{(break up row 1)} \\ &= \begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix} + \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix} && \textit{(break up row 2).} \end{aligned} \tag{5}$$

The last line has $2^2 = 4$ determinants. The first and fourth are zero because 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 permutation tells the column sequence. In this case the column order is $(1, 2)$ or $(2, 1)$.

Now try $n = 3$. Each row splits into 3 simpler rows like $\begin{bmatrix} a_{11} & 0 & 0 \end{bmatrix}$. 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 the row $\begin{bmatrix} a_{21} & 0 & 0 \end{bmatrix}$—then the simple determinant is zero.

We pay attention only when ***the entries a_{ij} come from different columns***, like $(\mathbf{3}, \mathbf{1}, \mathbf{2})$:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & & \\ & a_{22} & \\ & & a_{33} \end{vmatrix} + \begin{vmatrix} & a_{12} & \\ & & a_{23} \\ a_{31} & & \end{vmatrix} + \begin{vmatrix} & & \mathbf{a_{13}} \\ \mathbf{a_{21}} & & \\ & \mathbf{a_{32}} & \end{vmatrix}$$

Six terms

$$+ \begin{vmatrix} a_{11} & & \\ & & a_{23} \\ & a_{32} & \end{vmatrix} + \begin{vmatrix} & a_{12} & \\ a_{21} & & \\ & & a_{33} \end{vmatrix} + \begin{vmatrix} & & a_{13} \\ & a_{22} & \\ a_{31} & & \end{vmatrix}.$$

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 $P = I$.

$$\textbf{Column numbers} = (1,2,3), (2,3,1), (\mathbf{3},\mathbf{1},\mathbf{2}), (1,3,2), (2,1,3), (3,2,1). \qquad (6)$$

The last three are *odd permutations* (one exchange). The first three are *even permutations* (0 or 2 exchanges). When the column sequence is $(\mathbf{3}, \mathbf{1}, \mathbf{2})$, we have chosen the entries $a_{1\mathbf{3}}a_{2\mathbf{1}}a_{3\mathbf{2}}$—that particular column sequence comes with a plus sign (2 exchanges). The determinant of A is now split into six simple terms. Factor out the a_{ij}:

$$\det A = a_{11}a_{22}a_{33}\begin{vmatrix} 1 & & \\ & 1 & \\ & & 1 \end{vmatrix} + a_{12}a_{23}a_{31}\begin{vmatrix} & 1 & \\ & & 1 \\ 1 & & \end{vmatrix} + a_{1\mathbf{3}}a_{2\mathbf{1}}a_{3\mathbf{2}}\begin{vmatrix} & & 1 \\ 1 & & \\ & 1 & \end{vmatrix}$$
$$+ a_{11}a_{23}a_{32}\begin{vmatrix} 1 & & \\ & & 1 \\ & 1 & \end{vmatrix} + a_{12}a_{21}a_{33}\begin{vmatrix} & 1 & \\ 1 & & \\ & & 1 \end{vmatrix} + a_{13}a_{22}a_{31}\begin{vmatrix} & & 1 \\ & 1 & \\ 1 & & \end{vmatrix}. \qquad (7)$$

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 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***. For 5 by 5, the term $a_{15}a_{22}a_{33}a_{44}a_{51}$ would have $\det P = -1$ from exchanging 5 and 1.

$$\det A = \text{sum over all } \mathbf{n}! \text{ column permutations } P = (\alpha, \beta, \ldots, \omega)$$
$$= \sum (\det P) a_{1\alpha}a_{2\beta}\cdots a_{n\omega} = \textbf{BIG FORMULA}. \qquad (8)$$

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 $u_{21} = 0$, 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

$$\text{The determinant of } Z = \begin{vmatrix} 1 & 0 & \boldsymbol{a} & 0 \\ 0 & 1 & \boldsymbol{b} & 0 \\ 0 & 0 & \boldsymbol{c} & 0 \\ 0 & 0 & \boldsymbol{d} & 1 \end{vmatrix} \quad \text{is } \boldsymbol{c}. \tag{9}$$

The term $(1)(1)(c)(1)$ comes from the main diagonal with a plus sign. There are $4! = 24$ products (choosing one factor from each row and column) but the other 23 products are 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 $n = 4$:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \text{have } \textbf{determinant 1}.$$

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 $\det P = \det A$. The determinant of P is $+1$ (two exchanges to reach $2, 1, 4, 3$). Therefore $\det A = +1$.

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 4-10 will follow). The easiest is $\det I = 1$, already checked.

When you separate out the factor a_{11} or a_{12} or $a_{1\alpha}$ that comes from the first row, you see linearity. For 3 by 3, separate the usual 6 terms of the determinant into 3 pairs:

$$\det A = \boldsymbol{a_{11}}\,(a_{22}a_{33} - a_{23}a_{32}) + \boldsymbol{a_{12}}\,(a_{23}a_{31} - a_{21}a_{33}) + \boldsymbol{a_{13}}\,(a_{21}a_{32} - a_{22}a_{31}). \tag{10}$$

Those three quantities in parentheses are called ***"cofactors"***. They are **2 by 2 determinants**, from 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 the sign $(-1)^{1+j}$ to get the cofactor:

$$\text{The cofactors along row 1 are } C_{1j} = (-1)^{1+j}\det M_{1j}.$$

$$\textbf{The cofactor expansion is } \det \boldsymbol{A} = \boldsymbol{a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}}. \tag{11}$$

In the big formula (8), the terms that multiply a_{11} combine to give $C_{11} = \det M_{11}$. 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 only 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:

$$A = \begin{bmatrix} \bullet & \bullet & & \bullet \\ \bullet & \bullet & & \bullet \\ \bullet & \bullet & & \bullet \\ & & a_{43} & \end{bmatrix} \qquad \textbf{signs } (-1)^{i+j} = \begin{bmatrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{bmatrix}.$$

The determinant is the dot product of any row i of A with its cofactors using other rows:

COFACTOR FORMULA $$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}. \qquad (12)$$

Each cofactor C_{ij} (order $n-1$, without row i and column j) includes its correct sign:

Cofactor $$C_{ij} = (-1)^{i+j} \det M_{ij}.$$

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 rules. 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:

Cofactors down column j $$\det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}. \qquad (13)$$

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} :

$$\begin{vmatrix} 2 & -1 & & \\ \mathbf{-1} & 2 & \mathbf{-1} & \\ & -1 & \mathbf{2} & \mathbf{-1} \\ & & \mathbf{-1} & \mathbf{2} \end{vmatrix} = 2\begin{vmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{vmatrix} - (-1)\begin{vmatrix} \mathbf{-1} & \mathbf{-1} & \\ & \mathbf{2} & \mathbf{-1} \\ & \mathbf{-1} & \mathbf{2} \end{vmatrix}. \qquad (14)$$

You see 2 times C_{11} first on the right, from crossing out row 1 and column 1. This cofactor C_{11} 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}:***

$$D_4 = 2D_3 - D_2 \quad \text{and generally} \quad D_n = 2D_{n-1} - D_{n-2}. \qquad (15)$$

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 $\boldsymbol{D_n = n+1}$. Then D_n equals $2n - (n-1)$. That "special tridiagonal answer" also came from the product of pivots in Example 2.

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 $n - (n-1) = 1$.

If you also change the last 2 into 1, why is det $= 0$?

■ REVIEW OF THE KEY IDEAS ■

1. With no row exchanges, det A = (*product of pivots*). In the upper left corner of A, 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 $|H_n|$?

$$H_2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \qquad H_3 = \begin{bmatrix} 2 & 1 & \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \qquad H_4 = \begin{bmatrix} 2 & 1 & & \\ 1 & 2 & 1 & \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}$$

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 $|H_n| = 2, 3, 5, 8, \ldots$ 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. I am assuming $n > 1$.)

In Question **2**, multiplying $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ 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 $n!$ 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 $D = \textbf{diag}(1 : n)$ 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: $|A| = \sum \pm a_{1\alpha}a_{2\beta}\cdots a_{n\omega}$. Every term uses each row and each column once.

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 :

$$E_1 = |1| \qquad E_2 = \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} \qquad E_3 = \begin{vmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{vmatrix} \qquad E_4 = \begin{vmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix}.$$

(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 $|B_4| = 2|B_3| - |B_2| = 1$.

$$B_4 = \begin{bmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{bmatrix} \quad B_3 = \begin{bmatrix} 1 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{bmatrix} \quad B_2 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}.$$

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?

$$G_2 = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} \quad G_3 = \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{vmatrix} \quad G_4 = \begin{vmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{vmatrix}.$$

21 Compute S_1, S_2, S_3 for these $1, 3, 1$ matrices. By Fibonacci guess and check S_4.

$$S_1 = |3| \quad S_2 = \begin{vmatrix} 3 & 1 \\ 1 & 3 \end{vmatrix} \quad S_3 = \begin{vmatrix} 3 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 3 \end{vmatrix}$$

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 $[\,A \;\; B\,]$ from the second row $[C \;\; D]$. 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:

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A|\,|D - CA^{-1}B| = |AD - CB| \quad \text{provided } AC = CA.$$

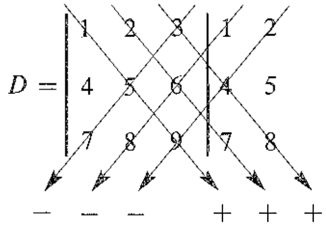
26 If A is m by n and B is n by m, block multiplication gives $\det M = \det AB$:

$$M = \begin{bmatrix} 0 & A \\ -B & I \end{bmatrix} = \begin{bmatrix} AB & A \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ -B & I \end{bmatrix}.$$

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}.

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

1 $\boldsymbol{A^{-1}}$ **equals** $\boldsymbol{C^{\mathrm{T}}/\det A}$. Then $(A^{-1})_{ij}$ = cofactor $\boldsymbol{C_{ji}}$ divided by the determinant of A.

2 **Cramer's Rule** computes $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ from $x_j = \det(A \text{ with column } j \text{ changed to } \boldsymbol{b}) \,/\, \det A$.

3 **Area of parallelogram** $= |ad-bc|$ if the four corners are $(0,0)$, (a,b), (c,d), and $(a+c, b+d)$.

4 **Volume of box** $= |\det \boldsymbol{A}|$ if the rows of A (or the columns of A) give the sides of the box.

5 The **cross product** $\boldsymbol{w} = \boldsymbol{u} \times \boldsymbol{v}$ is $\det \begin{bmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix}$. Notice $\boldsymbol{v} \times \boldsymbol{u} = -(\boldsymbol{u} \times \boldsymbol{v})$. w_1, w_2, w_3 are cofactors of row 1. Notice $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{u} = 0$ and $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{v} = 0$.

This section solves $A\boldsymbol{x} = \boldsymbol{b}$ and also finds A^{-1}—by algebra and not by elimination. In all formulas you will see a division by $\det A$. Each entry in A^{-1} and $A^{-1}\boldsymbol{b}$ is a determinant divided by the determinant of A. Let me start with Cramer's Rule.

Cramer's Rule solves $A\boldsymbol{x} = \boldsymbol{b}$. A neat idea gives the first component x_1. Replacing the first column of I by $\boldsymbol{x}$ gives a matrix with determinant x_1. When you multiply it by A, *the first column becomes $A\boldsymbol{x}$ which is $\boldsymbol{b}$.* The other columns of B_1 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 to find x_1:*

$$\boxed{\textbf{Product rule} \qquad (\det A)(x_1) = \det B_1 \qquad \text{or} \qquad x_1 = \frac{\det B_1}{\det A}.} \tag{2}$$

This is the first component of $\boldsymbol{x}$ in Cramer's Rule! Changing a column of A gave B_1. To find x_2 and B_2, put the vectors $\boldsymbol{x}$ and $\boldsymbol{b}$ into the *second* columns of I and A:

$$\textbf{Same idea} \qquad \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 \end{bmatrix} \begin{bmatrix} 1 & x_1 & 0 \\ 0 & x_2 & 0 \\ 0 & x_3 & 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{b} & \boldsymbol{a}_3 \end{bmatrix} = B_2. \tag{3}$$

Take determinants to find $(\det A)(x_2) = \det B_2$. This gives $\boldsymbol{x_2 = (\det B_2)/(\det A)}$.

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 of A, B_1, B_2 are -2 and -4 and 2. All ratios divide by $\det A = -2$:

$$\textbf{Find } \boldsymbol{x} = A^{-1}\boldsymbol{b} \qquad x_1 = \frac{-4}{-2} = 2 \qquad x_2 = \frac{2}{-2} = -1 \qquad \textbf{Check} \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

CRAMER's RULE If $\det A$ is not zero, $A\boldsymbol{x} = \boldsymbol{b}$ is solved by determinants:

$$x_1 = \frac{\det B_1}{\det A} \qquad x_2 = \frac{\det B_2}{\det A} \qquad \ldots \qquad x_n = \frac{\det B_n}{\det A} \tag{4}$$

The matrix B_j has the jth column of A replaced by the vector $\boldsymbol{b}$.

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 $\boldsymbol{x}$.

Example 2 Cramer's Rule is inefficient for numbers but it is well suited to letters. For $n = 2$, find the columns of $A^{-1} = [\boldsymbol{x} \;\; \boldsymbol{y}]$ by solving $AA^{-1} = I$:

$$\begin{array}{l}\textbf{Columns of } A^{-1} \\ \textbf{are } \boldsymbol{x} \textbf{ and } \boldsymbol{y}\end{array} \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Those share the same matrix A. We need $|A|$ and four determinants for x_1, x_2, y_1, y_2:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \text{ and } \begin{vmatrix} 1 & b \\ 0 & d \end{vmatrix} \quad \begin{vmatrix} a & 1 \\ c & 0 \end{vmatrix} \quad \begin{vmatrix} 0 & b \\ 1 & d \end{vmatrix} \quad \begin{vmatrix} a & 0 \\ c & 1 \end{vmatrix}$$

The last four determinants 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: $\boldsymbol{A^{-1}}$ **involves the cofactors**. When the right side is a column of the identity matrix I, as in $AA^{-1} = I$, **the determinant of each $\boldsymbol{B_j}$ in Cramer's Rule is a cofactor of $\boldsymbol{A}$.**

You can see those cofactors for $n = 3$. Solve $A\boldsymbol{x} = (1, 0, 0)$ to find column 1 of A^{-1}:

$$\begin{array}{l}\textbf{Determinants of } B\textbf{'s} \\ = \textbf{Cofactors of } A\end{array} \quad \begin{vmatrix} 1 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} \quad \begin{vmatrix} a_{11} & 1 & a_{13} \\ a_{21} & 0 & a_{23} \\ a_{31} & 0 & a_{33} \end{vmatrix} \quad \begin{vmatrix} a_{11} & a_{12} & 1 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 0 \end{vmatrix} \tag{5}$$

That first determinant $|B_1|$ is the cofactor $C_{11} = a_{22}a_{33} - a_{23}a_{32}$. Then $|B_2|$ 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 column 1 of A^{-1}. When we divide by $\det A$, we have the inverse matrix!

***The i, j entry of A^{-1} is the cofactor C_{ji} (not C_{ij}) divided by* det A:**

FORMULA FOR A^{-1} $$(A^{-1})_{ij} = \frac{C_{ji}}{\det A} \quad \text{and} \quad A^{-1} = \frac{C^{\mathrm{T}}}{\det A}. \tag{6}$$

The cofactors C_{ij} go into the *"cofactor matrix"* C. ***The transpose of C 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 C_{ji}, and divide by det A.

Check this rule for the $3, 1$ entry of A^{-1}. For column 1 we solve $Ax = (1, 0, 0)$. The third component x_3 needs the third determinant in equation (5), divided by det A. That determinant is exactly the cofactor $C_{13} = a_{21}a_{32} - a_{22}a_{31}$. So $(A^{-1})_{31} = C_{13}/\det A$.

Summary In solving $AA^{-1} = I$, each column of I leads to a column of A^{-1}. Every entry of A^{-1} is a ratio: determinant of size $n - 1$ / determinant of size n.

Direct proof of the formula $A^{-1} = C^{\mathrm{T}}/\det A$ This means $AC^{\mathrm{T}} = (\det A)I$:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix} = \begin{bmatrix} \det A & 0 & 0 \\ 0 & \det A & 0 \\ 0 & 0 & \det A \end{bmatrix}. \tag{7}$$

(Row 1 of A) times (column 1 of C^{T}) yields the first det A on the right:

$$a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} = \det A \quad \text{This is exactly the cofactor rule!}$$

Similarly row 2 of A times column 2 of C^{T} (*notice the transpose*) also 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)? The 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 to compute area.

Determinants are the best way to find area. ***The area of a triangle is half of a 3 by 3 determinant.*** The square roots in the base and height cancel out in the good formula. If one corner is at the origin, say $(x_3, y_3) = (0,0)$, the determinant is only 2 by 2.

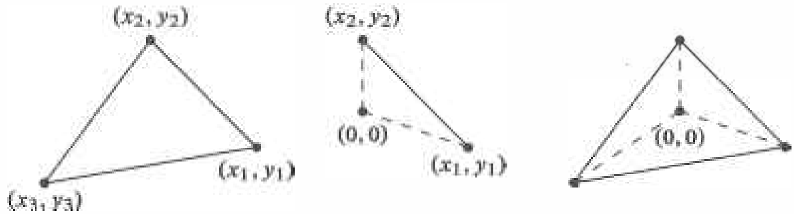


Figure 5.1: General triangle; special triangle from $(0,0)$; general from three specials.

The triangle with corners (x_1, y_1) and (x_2, y_2) and (x_3, y_3) has **area** $= \dfrac{\textbf{determinant}}{\textbf{2}}$:

Area of triangle $\quad \dfrac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$ $\qquad$ Area $= \dfrac{1}{2}\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}$ 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 (cofactors), just as the third triangle in Figure 5.1 breaks into three special triangles from $(0,0)$:

$$\textbf{Area} = \frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = \begin{array}{l} +\frac{1}{2}(x_1y_2 - x_2y_1) \\ +\frac{1}{2}(x_2y_3 - x_3y_2) \\ +\frac{1}{2}(x_3y_1 - x_1y_3). \end{array} \tag{9}$$

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 area of a triangle with corner $(0,0)$.

Why is $\frac{1}{2}|x_1y_2 - x_2y_1|$ the area of this triangle? We can remove the factor $\frac{1}{2}$ for 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 $\frac{11}{2}$.

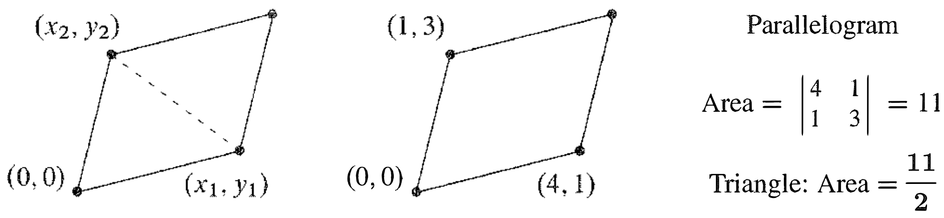


Figure 5.2: A triangle is half of a parallelogram. Area is half of a determinant.

Proof that a parallelogram starting from $(0,0)$ has area $=$ 2 by 2 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).

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, like the parallelogram.

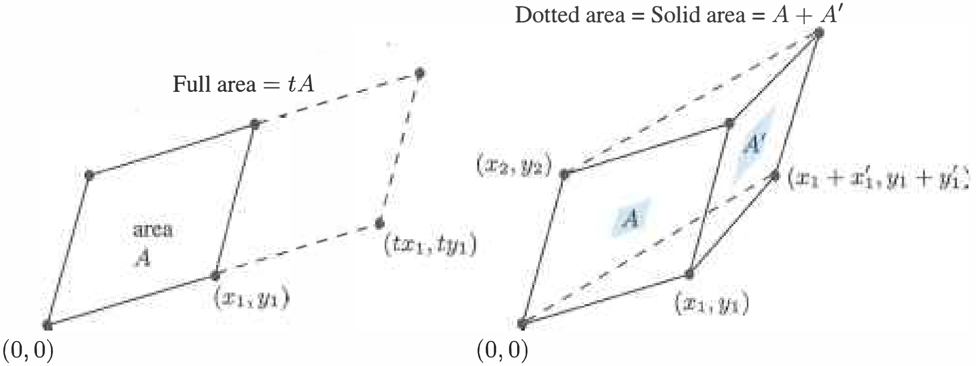


Figure 5.3: Areas obey the rule of linearity in side 1 (keeping the side (x_2, y_2) constant).

Figure 5.4 shows a three-dimensional box—whose edges are not at right angles. ***The volume equals the absolute value of*** **det** $\boldsymbol{A}$**.** 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 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.

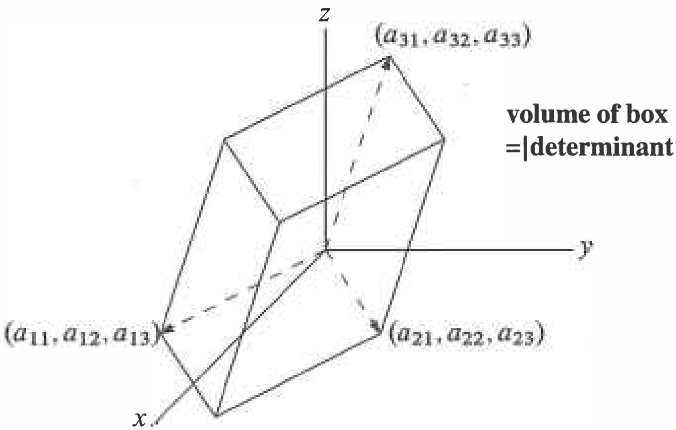


Figure 5.4: Three-dimensional box formed from the three rows of A.

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 $\det A =$ absolute value.

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 A with entries r, s, and t produces those three sides. Then $\det A$ also equals the volume $r\,s\,t$.

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 $dr\,d\theta$:

$$\textbf{Area } r\,dr\,d\theta \textbf{ in calculus} \qquad J = \begin{vmatrix} \partial x/\partial r & \partial x/\partial\theta \\ \partial y/\partial r & \partial y/\partial\theta \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = \boldsymbol{r}.$$

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* 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

$$\boldsymbol{u} \times \boldsymbol{v} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = (u_2v_3 - u_3v_2)\boldsymbol{i} + (u_3v_1 - u_1v_3)\boldsymbol{j} + (u_1v_2 - u_2v_1)\boldsymbol{k}. \tag{10}$$

This vector $\boldsymbol{u} \times \boldsymbol{v}$ ***is perpendicular to*** $\boldsymbol{u}$ ***and*** $\boldsymbol{v}$. 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, producing a zero dot product:

$$\boldsymbol{u} \cdot (\boldsymbol{u} \times \boldsymbol{v}) = u_1(u_2v_3 - u_3v_2) + u_2(u_3v_1 - u_1v_3) + u_3(u_1v_2 - u_2v_1) = 0. \tag{11}$$

The determinant for $\boldsymbol{u}\cdot(\boldsymbol{u} \times \boldsymbol{v})$ has rows $\boldsymbol{u}, \boldsymbol{u}$ and $\boldsymbol{v}$ (2 equal rows) so it is zero.

Property 3 The cross product of any vector with itself (two equal rows) is $\boldsymbol{u} \times \boldsymbol{u} = \mathbf{0}$.

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 $\cos\theta$:

$$\|\boldsymbol{u} \times \boldsymbol{v}\| = \|\boldsymbol{u}\|\,\|\boldsymbol{v}\|\,|\sin\theta| \qquad \text{and} \qquad |\boldsymbol{u} \cdot \boldsymbol{v}| = \|\boldsymbol{u}\|\,\|\boldsymbol{v}\|\,|\cos\theta|. \tag{12}$$

Example 7 $\boldsymbol{u} = (3, 2, 0)$ and $\boldsymbol{v} = (1, 4, 0)$ are in the xy plane, $\boldsymbol{u} \times \boldsymbol{v}$ goes up the z axis:

$$\boldsymbol{u} \times \boldsymbol{v} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 3 & 2 & 0 \\ 1 & 4 & 0 \end{vmatrix} = 10\boldsymbol{k}. \quad \text{The cross product is } \boldsymbol{u} \times \boldsymbol{v} = (\mathbf{0}, \mathbf{0}, \mathbf{10}).$$

The length of $\boldsymbol{u} \times \boldsymbol{v}$ equals the area of the parallelogram with sides $\boldsymbol{u}$ and $\boldsymbol{v}$. This will be important: In this example the area is 10.

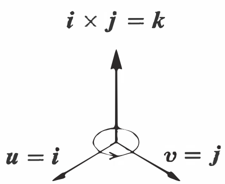
Example 8 The cross product of $\boldsymbol{u} = (1, 1, 1)$ and $\boldsymbol{v} = (1, 1, 2)$ is $(1, -1, 0)$:

$$\begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{vmatrix} = \boldsymbol{i}\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} - \boldsymbol{j}\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} + \boldsymbol{k}\begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = \boldsymbol{i} - \boldsymbol{j}.$$

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 $\boldsymbol{i} = (1, 0, 0)$ and $\boldsymbol{j} = (0, 1, 0)$ obeys the *right hand rule*. That cross product $\boldsymbol{k} = \boldsymbol{i} \times \boldsymbol{j}$ 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 $\boldsymbol{u} \times \boldsymbol{F} = \mathbf{0}$—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
$$\boxed{(\boldsymbol{u} \times \boldsymbol{v}) \cdot \boldsymbol{w}} = \begin{vmatrix} w_1 & w_2 & w_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}. \tag{13}$$

We can put $\boldsymbol{w}$ 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:

$$(\boldsymbol{u} \times \boldsymbol{v}) \cdot \boldsymbol{w} = 0 \quad \text{exactly when the vectors } \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \text{ lie in the } \textit{same plane}.$$

First reason $\boldsymbol{u} \times \boldsymbol{v}$ is perpendicular to that plane so its dot product with $\boldsymbol{w}$ is zero.

Second reason Three vectors in a plane are dependent. The matrix is singular (det $= 0$).

Third reason Zero volume when the $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ box is squashed onto a plane.

It is remarkable that $(\boldsymbol{u} \times \boldsymbol{v}) \cdot \boldsymbol{w}$ equals the volume of the box with sides $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$. 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 $A\boldsymbol{x} = \boldsymbol{b}$ by ratios like $x_1 = |B_1|/|A| = |\boldsymbol{b}\,\boldsymbol{a}_2 \cdots \boldsymbol{a}_n|/|A|$.

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 $\boldsymbol{u} \times \boldsymbol{v}$ is perpendicular to $\boldsymbol{u}$ and $\boldsymbol{v}$. Notice $\boldsymbol{i} \times \boldsymbol{j} = \boldsymbol{k}$.

■ WORKED EXAMPLES ■

5.3 A If A is singular, the equation $AC^{\mathrm{T}} = (\det A)I$ becomes $\boldsymbol{AC^{\mathrm{T}}} =$ **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 for a 3 by 3 matrix of rank 2. My apologies that this comes so late!

Solve $A\boldsymbol{x} = \boldsymbol{0}$ by $\boldsymbol{x} =$ cofactors along a row, for these singular matrices of rank 2:

Cofactors give nullspace $\quad A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 3 & 9 \\ 2 & 2 & 8 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

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 $A\boldsymbol{x} = \boldsymbol{0}$. 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 $A\boldsymbol{x} = \boldsymbol{0}$.

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 $\boldsymbol{x} = \boldsymbol{0}$.

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 $A^{-1} = C^{\mathrm{T}} / \det A$. For this $\boldsymbol{b} = (0, 0, 1)$ the solution $\boldsymbol{x}$ is column 3 of A^{-1}! Which cofactors are involved in computing that column $\boldsymbol{x} = (x, y, z)$?

Column 3 of $\boldsymbol{A^{-1}}$ $\quad \begin{bmatrix} 2 & 6 & 2 \\ 1 & 4 & 2 \\ 5 & 9 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$

Find the volumes of two boxes: edges are *columns* of A and edges are 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 $\det A = 2$:

$$\det A = a_{31}C_{31} + a_{32}C_{32} + a_{33}C_{33} = (5, 9, 0) \cdot (4, -2, 2) = 2.$$

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.