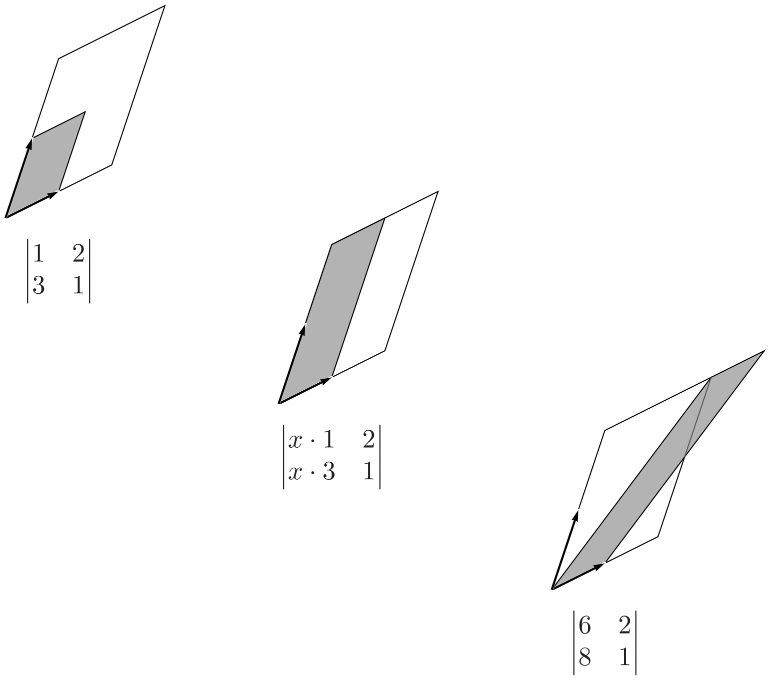
Linear Algebra



Jim Hefferon

Notation

$\mathbb{R}$	real numbers
$\mathbb{N}$	natural numbers: $\{0, 1, 2, \ldots\}$
$\mathbb{C}$	complex numbers
$\{\ldots \mid \ldots\}$	set of ... such that ...
$\langle \ldots \rangle$	sequence; like a set but order matters
V, W, U	vector spaces
$\vec{v}, \vec{w}$	vectors
$\vec{0}$, $\vec{0}_V$	zero vector, zero vector of V
B, D	bases
$\mathcal{E}_n = \langle \vec{e}_1, \ldots, \vec{e}_n \rangle$	standard basis for $\mathbb{R}^n$
$\vec{\beta}, \vec{\delta}$	basis vectors
$\mathrm{Rep}_B(\vec{v})$	matrix representing the vector
$\mathcal{P}_n$	set of n-th degree polynomials
$\mathcal{M}_{n\times m}$	set of $n\times m$ matrices
$[S]$	span of the set S
$M \oplus N$	direct sum of subspaces
$V \cong W$	isomorphic spaces
h, g	homomorphisms
H, G	matrices
t, s	transformations; maps from a space to itself
T, S	square matrices
$\mathrm{Rep}_{B,D}(h)$	matrix representing the map h
$h_{i,j}$	matrix entry from row i, column j
$\lvert T \rvert$	determinant of the matrix T
$\mathscr{R}(h), \mathscr{N}(h)$	rangespace and nullspace of the map h
$\mathscr{R}_\infty(h), \mathscr{N}_\infty(h)$	generalized rangespace and nullspace

Lower case Greek alphabet

name	symbol	name	symbol	name	symbol
alpha	α	iota	ι	rho	ρ
beta	β	kappa	κ	sigma	σ
gamma	γ	lambda	λ	tau	τ
delta	δ	mu	μ	upsilon	υ
epsilon	ϵ	nu	ν	phi	ϕ
zeta	ζ	xi	ξ	chi	χ
eta	η	omicron	o	psi	ψ
theta	θ	pi	π	omega	ω

Cover. This is Cramer's Rule applied to the system $x + 2y = 6$, $3x + y = 8$. The area of the first box is the determinant shown. The area of the second box is x times that, and equals the area of the final box. Hence, x is the final determinant divided by the first determinant.

Preface

In most mathematics programs linear algebra is taken in the first or second year, following or along with at least one course in calculus. While the location of this course is stable, lately the content has been under discussion. Some instructors have experimented with varying the traditional topics, trying courses focused on applications, or on the computer. Despite this (entirely healthy) debate, most instructors are still convinced, I think, that the right core material is vector spaces, linear maps, determinants, and eigenvalues and eigenvectors. Applications and computations certainly can have a part to play but most mathematicians agree that the themes of the course should remain unchanged.

Not that all is fine with the traditional course. Most of us do think that the standard text type for this course needs to be reexamined. Elementary texts have traditionally started with extensive computations of linear reduction, matrix multiplication, and determinants. These take up half of the course. Finally, when vector spaces and linear maps appear, and definitions and proofs start, the nature of the course takes a sudden turn. In the past, the computation drill was there because, as future practitioners, students needed to be fast and accurate with these. But that has changed. Being a whiz at 5×5 determinants just isn't important anymore. Instead, the availability of computers gives us an opportunity to move toward a focus on concepts.

This is an opportunity that we should seize. The courses at the start of most mathematics programs work at having students correctly apply formulas and algorithms, and imitate examples. Later courses require some mathematical maturity: reasoning skills that are developed enough to follow different types of proofs, a familiarity with the themes that underly many mathematical investigations like elementary set and function facts, and an ability to do some independent reading and thinking, Where do we work on the transition?

Linear algebra is an ideal spot. It comes early in a program so that progress made here pays off later. The material is straightforward, elegant, and accessible. The students are serious about mathematics, often majors and minors. There are a variety of argument styles—proofs by contradiction, if and only if statements, and proofs by induction, for instance—and examples are plentiful.

The goal of this text is, along with the development of undergraduate linear algebra, to help an instructor raise the students' level of mathematical sophistication. Most of the differences between this book and others follow straight from that goal.

One consequence of this goal of development is that, unlike in many computational texts, all of the results here are proved. On the other hand, in contrast with more abstract texts, many examples are given, and they are often quite detailed.

Another consequence of the goal is that while we start with a computational topic, linear reduction, from the first we do more than just compute. The solution of linear systems is done quickly but it is also done completely, proving everything (really these proofs are just verifications), all the way through the uniqueness of reduced echelon form. In particular, in this first chapter, the opportunity is taken to present a few induction proofs, where the arguments just go over bookkeeping details, so that when induction is needed later (e.g., to prove that all bases of a finite dimensional vector space have the same number of members), it will be familiar.

Still another consequence is that the second chapter immediately uses this background as motivation for the definition of a real vector space. This typically occurs by the end of the third week. We do not stop to introduce matrix multiplication and determinants as rote computations. Instead, those topics appear naturally in the development, after the definition of linear maps.

To help students make the transition from earlier courses, the presentation here stresses motivation and naturalness. An example is the third chapter, on linear maps. It does not start with the definition of homomorphism, as is the case in other books, but with the definition of isomorphism. That's because this definition is easily motivated by the observation that some spaces are just like each other. After that, the next section takes the reasonable step of defining homomorphisms by isolating the operation-preservation idea. A little mathematical slickness is lost, but it is in return for a large gain in sensibility to students.

Having extensive motivation in the text helps with time pressures. I ask students to, before each class, look ahead in the book, and they follow the classwork better because they have some prior exposure to the material. For example, I can start the linear independence class with the definition because I know students have some idea of what it is about. No book can take the place of an instructor, but a helpful book gives the instructor more class time for examples and questions.

Much of a student's progress takes place while doing the exercises; the exercises here work with the rest of the text. Besides computations, there are many proofs. These are spread over an approachability range, from simple checks to some much more involved arguments. There are even a few exercises that are reasonably challenging puzzles taken, with citation, from various journals, competitions, or problems collections (as part of the fun of these, the original wording has been retained as much as possible). In total, the questions are aimed to both build an ability at, and help students experience the pleasure of, *doing* mathematics.

Applications, and Computers. The point of view taken here, that linear algebra is about vector spaces and linear maps, is not taken to the exclusion of all other ideas. Applications, and the emerging role of the computer, are interesting, important, and vital aspects of the subject. Consequently, every chapter closes with a few application or computer-related topics. Some of the topics are: network flows, the speed and accuracy of computer linear reductions, Leontief Input/Output analysis, dimensional analysis, Markov chains, voting paradoxes, analytic projective geometry, and solving difference equations.

These are brief enough to be done in a day's class or to be given as independent projects for individuals or small groups. Most simply give a reader a feel for the subject, discuss how linear algebra comes in, point to some accessible further reading, and give a few exercises. I have kept the exposition lively and given an overall sense of breadth of application. In short, these topics invite readers to see for themselves that linear algebra is a tool that a professional must have.

For people reading this book on their own. The emphasis on motivation and development make this book a good choice for self-study. While a professional mathematician knows what pace and topics suit a class, perhaps an independent student would find some advice helpful. Here are two timetables for a semester. The first focuses on core material.

week	*Mon.*	*Wed.*	*Fri.*
1	1.I.1	1.I.1, 2	1.I.2, 3
2	1.I.3	1.II.1	1.II.2
3	1.III.1, 2	1.III.2	2.I.1
4	2.I.2	2.II	2.III.1
5	2.III.1, 2	2.III.2	exam
6	2.III.2, 3	2.III.3	3.I.1
7	3.I.2	3.II.1	3.II.2
8	3.II.2	3.II.2	3.III.1
9	3.III.1	3.III.2	3.IV.1, 2
10	3.IV.2, 3, 4	3.IV.4	exam
11	3.IV.4, 3.V.1	3.V.1, 2	4.I.1, 2
12	4.I.3	4.II	4.II
13	4.III.1	5.I	5.II.1
14	5.II.2	5.II.3	review

The second timetable is more ambitious (it presupposes 1.II, the elements of vectors, usually covered in third semester calculus).

week	*Mon.*	*Wed.*	*Fri.*
1	1.I.1	1.I.2	1.I.3
2	1.I.3	1.III.1, 2	1.III.2
3	2.I.1	2.I.2	2.II
4	2.III.1	2.III.2	2.III.3
5	2.III.4	3.I.1	exam
6	3.I.2	3.II.1	3.II.2
7	3.III.1	3.III.2	3.IV.1, 2
8	3.IV.2	3.IV.3	3.IV.4
9	3.V.1	3.V.2	3.VI.1
10	3.VI.2	4.I.1	exam
11	4.I.2	4.I.3	4.I.4
12	4.II	4.II, 4.III.1	4.III.2, 3
13	5.II.1, 2	5.II.3	5.III.1
14	5.III.2	5.IV.1, 2	5.IV.2

See the table of contents for the titles of these subsections.

For guidance, in the table of contents I have marked some subsections as optional if, in my opinion, some instructors will pass over them in favor of spending more time elsewhere. These subsections can be dropped or added, as desired. You might also adjust the length of your study by picking one or two Topics that appeal to you from the end of each chapter. You'll probably get more out of these if you have access to computer software that can do the big calculations.

Do many exercises. (The answers are available.) I have marked a good sample with ✓'s. Be warned about the exercises, however, that few inexperienced people can write correct proofs. Try to find a knowledgeable person to work with you on this aspect of the material.

Finally, if I may, a caution: I cannot overemphasize how much the statement (which I sometimes hear), "I understand the material, but it's only that I can't do any of the problems." reveals a lack of understanding of what we are up to. Being able to do particular things with the ideas is the entire point. The quote below expresses this sentiment admirably, and captures the essence of this book's approach. It states what I believe is the key to both the beauty and the power of mathematics and the sciences in general, and of linear algebra in particular.

I know of no better tactic
than the illustration of exciting principles
by well-chosen particulars.
–Stephen Jay Gould

Jim Hefferon
Saint Michael's College
Colchester, Vermont USA
jim@joshua.smcvt.edu
April 20, 2000

Author's Note. Inventing a good exercise, one that enlightens as well as tests, is a creative act, and hard work (at least half of the the effort on this text has gone into exercises and solutions). The inventor deserves recognition. But, somehow, the tradition in texts has been to not give attributions for questions. I have changed that here where I was sure of the source. I would greatly appreciate hearing from anyone who can help me to correctly attribute others of the questions. They will be incorporated into later versions of this book.

Contents

**Note:* starred subsections are optional.

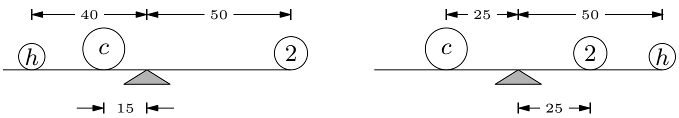
Chapter 1

Linear Systems

1.I Solving Linear Systems

Systems of linear equations are common in science and mathematics. These two examples from high school science [Onan] give a sense of how they arise.

The first example is from Physics. Suppose that we are given three objects, one with a mass of 2 kg, and are asked to find the unknown masses. Suppose further that experimentation with a meter stick produces these two balances.



Now, since the sum of moments on the left of each balance equals the sum of moments on the right (the moment of an object is its mass times its distance from the balance point), the two balances give this system of two equations.

$$\begin{aligned} 40h + 15c &= 100 \\ 25c &= 50 + 50h \end{aligned}$$

The second example of a linear system is from Chemistry. We can mix, under controlled conditions, toluene C_7H_8 and nitric acid HNO_3 to produce trinitrotoluene $C_7H_5O_6N_3$ along with the byproduct water (conditions have to be controlled very well, indeed — trinitrotoluene is better known as TNT). In what proportion should those components be mixed? The number of atoms of each element present before the reaction

$$x\,C_7H_8 \ + \ y\,HNO_3 \quad \longrightarrow \quad z\,C_7H_5O_6N_3 \ + \ w\,H_2O$$

must equal the number present afterward. Applying that principle to the elements C, H, N, and O in turn gives this system.

$$\begin{aligned} 7x &= 7z \\ 8x + 1y &= 5z + 2w \\ 1y &= 3z \\ 3y &= 6z + 1w \end{aligned}$$

To finish each of these examples requires solving a system of equations. In each, the equations involve only the first power of the variables. This chapter shows how to solve any such system.

1.I.1 Gauss' Method

1.1 Definition A *linear equation* in variables $x_1, x_2, \ldots, x_n$ has the form

$$a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n = d$$

where the numbers $a_1, \ldots, a_n \in \mathbb{R}$ are the equation's *coefficients* and $d \in \mathbb{R}$ is the *constant*. An n-tuple $(s_1, s_2, \ldots, s_n) \in \mathbb{R}^n$ is a *solution* of, or *satisfies*, that equation if substituting the numbers s_1, …, s_n for the variables gives a true statement: $a_1s_1 + a_2s_2 + \ldots + a_ns_n = d$.

A *system of linear equations*

$$\begin{aligned} a_{1,1}x_1 + \ a_{1,2}x_2 + \cdots + \ a_{1,n}x_n &= \ d_1 \\ a_{2,1}x_1 + \ a_{2,2}x_2 + \cdots + \ a_{2,n}x_n &= \ d_2 \\ &\ \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n &= d_m \end{aligned}$$

has the solution $(s_1, s_2, \ldots, s_n)$ if that n-tuple is a solution of all of the equations in the system.

1.2 Example The ordered pair $(-1, 5)$ is a solution of this system.

$$\begin{aligned} 3x_1 + 2x_2 &= 7 \\ -x_1 + \ \ x_2 &= 6 \end{aligned}$$

In contrast, $(5, -1)$ is not a solution.

Finding the set of all solutions is *solving* the system. No guesswork or good fortune is needed to solve a linear system. There is an algorithm that always works. The next example introduces that algorithm, called *Gauss' method*. It transforms the system, step by step, into one with a form that is easily solved.

1.3 Example To solve this system

$$\begin{aligned} 3x_3 &= 9 \\ x_1 + 5x_2 - 2x_3 &= 2 \\ \tfrac{1}{3}x_1 + 2x_2 \qquad\ &= 3 \end{aligned}$$

we repeatedly transform it until it is in a form that is easy to solve.

$$\xrightarrow{\text{swap row 1 with row 3}} \begin{aligned} \tfrac{1}{3}x_1 + 2x_2 &= 3 \\ x_1 + 5x_2 - 2x_3 &= 2 \\ 3x_3 &= 9 \end{aligned}$$

$$\xrightarrow{\text{multiply row 1 by 3}} \begin{aligned} x_1 + 6x_2 &= 9 \\ x_1 + 5x_2 - 2x_3 &= 2 \\ 3x_3 &= 9 \end{aligned}$$

$$\xrightarrow{\text{add } -1 \text{ times row 1 to row 2}} \begin{aligned} x_1 + 6x_2 &= 9 \\ -x_2 - 2x_3 &= -7 \\ 3x_3 &= 9 \end{aligned}$$

The third step is the only nontrivial one. We've mentally multiplied both sides of the first row by -1, mentally added that to the old second row, and written the result in as the new second row.

Now we can find the value of each variable. The bottom equation shows that $x_3 = 3$. Substituting 3 for x_3 in the middle equation shows that $x_2 = 1$. Substituting those two into the top equation gives that $x_1 = 3$ and so the system has a unique solution: the solution set is $\{ (3, 1, 3) \}$.

Most of this subsection and the next one consists of examples of solving linear systems by Gauss' method. We will use it throughout this book. It is fast and easy. But, before we get to those examples, we will first show that this method is also safe in that it never loses solutions or picks up extraneous solutions.

1.4 Theorem (Gauss' method) If a linear system is changed to another by one of these operations

(1) an equation is swapped with another

(2) an equation has both sides multiplied by a nonzero constant

(3) an equation is replaced by the sum of itself and a multiple of another

then the two systems have the same set of solutions.

Each of those three operations has a restriction. Multiplying a row by 0 is not allowed because obviously that can change the solution set of the system. Similarly, adding a multiple of a row to itself is not allowed because adding -1 times the row to itself has the effect of multiplying the row by 0. Finally, swapping a row with itself is disallowed to make some results in the fourth chapter easier to state and remember (and besides, self-swapping doesn't accomplish anything).

Proof. We will cover the equation swap operation here and save the other two cases for Exercise 29.

Consider this swap of row i with row j.

$$\begin{array}{rcl} a_{1,1}x_1 + a_{1,2}x_2 + \cdots a_{1,n}x_n = d_1 & & a_{1,1}x_1 + a_{1,2}x_2 + \cdots a_{1,n}x_n = d_1 \\ \vdots & & \vdots \\ a_{i,1}x_1 + a_{i,2}x_2 + \cdots a_{i,n}x_n = d_i & & a_{j,1}x_1 + a_{j,2}x_2 + \cdots a_{j,n}x_n = d_j \\ \vdots & \longrightarrow & \vdots \\ a_{j,1}x_1 + a_{j,2}x_2 + \cdots a_{j,n}x_n = d_j & & a_{i,1}x_1 + a_{i,2}x_2 + \cdots a_{i,n}x_n = d_i \\ \vdots & & \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots a_{m,n}x_n = d_m & & a_{m,1}x_1 + a_{m,2}x_2 + \cdots a_{m,n}x_n = d_m \end{array}$$

The n-tuple $(s_1, \ldots, s_n)$ satisfies the system before the swap if and only if substituting the values, the s's, for the variables, the x's, gives true statements: $a_{1,1}s_1 + a_{1,2}s_2 + \cdots + a_{1,n}s_n = d_1$ and ... $a_{i,1}s_1 + a_{i,2}s_2 + \cdots + a_{i,n}s_n = d_i$ and ... $a_{j,1}s_1 + a_{j,2}s_2 + \cdots + a_{j,n}s_n = d_j$ and ... $a_{m,1}s_1 + a_{m,2}s_2 + \cdots + a_{m,n}s_n = d_m$.

In a requirement consisting of statements and-ed together we can rearrange the order of the statements, so that this requirement is met if and only if $a_{1,1}s_1 + a_{1,2}s_2 + \cdots + a_{1,n}s_n = d_1$ and ... $a_{j,1}s_1 + a_{j,2}s_2 + \cdots + a_{j,n}s_n = d_j$ and ... $a_{i,1}s_1 + a_{i,2}s_2 + \cdots + a_{i,n}s_n = d_i$ and ... $a_{m,1}s_1 + a_{m,2}s_2 + \cdots + a_{m,n}s_n = d_m$. This is exactly the requirement that $(s_1, \ldots, s_n)$ solves the system after the row swap. QED

1.5 Definition The three operations from Theorem 1.4 are the *elementary reduction operations*, or *row operations*, or *Gaussian operations*. They are *swapping*, *multiplying by a scalar* or *rescaling*, and *pivoting*.

When writing out the calculations, we will abbreviate 'row i' by 'ρ_i'. For instance, we will denote a pivot operation by $k\rho_i + \rho_j$, with the row that is changed written second. We will also, to save writing, often list pivot steps together when they use the same ρ_i.

1.6 Example A typical use of Gauss' method is to solve this system.

$$\begin{array}{rcl} x + y &=& 0 \\ 2x - y + 3z &=& 3 \\ x - 2y - z &=& 3 \end{array}$$

The first transformation of the system involves using the first row to eliminate the x in the second row and the x in the third. To get rid of the second row's $2x$, we multiply the entire first row by -2, add that to the second row, and write the result in as the new second row. To get rid of the third row's x, we multiply the first row by -1, add that to the third row, and write the result in as the new third row.

$$\xrightarrow[-2\rho_1+\rho_2]{-\rho_1+\rho_3} \begin{array}{rcl} x + y &=& 0 \\ -3y + 3z &=& 3 \\ -3y - z &=& 3 \end{array}$$

(Note that the two ρ_1 steps $-2\rho_1 + \rho_2$ and $-\rho_1 + \rho_3$ are written as one operation.) In this second system, the last two equations involve only two unknowns.

To finish we transform the second system into a third system, where the last equation involves only one unknown. This transformation uses the second row to eliminate y from the third row.

$$\xrightarrow{-\rho_2+\rho_3}\quad \begin{aligned} x+\quad y\qquad\quad &=0\\ -3y+\quad 3z&=3\\ -4z&=0 \end{aligned}$$

Now we are set up for the solution. The third row shows that $z=0$. Substitute that back into the second row to get $y=-1$, and then substitute back into the first row to get $x=1$.

1.7 Example For the Physics problem from the start of this chapter, Gauss' method gives this.

$$\begin{aligned} 40h+15c&=100\\ -50h+25c&=\ \ 50 \end{aligned}\quad\xrightarrow{5/4\rho_1+\rho_2}\quad\begin{aligned} 40h+\qquad\quad 15c&=100\\ (175/4)c&=175 \end{aligned}$$

So $c=4$, and back-substitution gives that $h=1$. (The Chemistry problem is solved later.)

1.8 Example The reduction

$$\begin{aligned} x+\ \ y+\ \ z&=9\\ 2x+4y-3z&=1\\ 3x+6y-5z&=0 \end{aligned}\quad\xrightarrow[-3\rho_1+\rho_3]{-2\rho_1+\rho_2}\quad\begin{aligned} x+\ \ y+\ \ z&=\ \ \ 9\\ 2y-5z&=-17\\ 3y-8z&=-27 \end{aligned}$$

$$\xrightarrow{-(3/2)\rho_2+\rho_3}\quad\begin{aligned} x+\ \ y+\qquad z&=\ \ \ 9\\ 2y-\quad 5z&=-17\\ -\tfrac{1}{2}z&=\ -\tfrac{3}{2} \end{aligned}$$

shows that $z=3$, $y=-1$, and $x=7$.

As these examples illustrate, Gauss' method uses the elementary reduction operations to set up back-substitution.

1.9 Definition In each row, the first variable with a nonzero coefficient is the row's *leading variable*. A system is in *echelon form* if each leading variable is to the right of the leading variable in the row above it (except for the leading variable in the first row).

1.10 Example The only operation needed in the examples above is pivoting. Here is a linear system that requires the operation of swapping equations. After the first pivot

$$\begin{aligned} x-\ \ y\qquad\qquad\quad &=0\\ 2x-2y+\ \ z+2w&=4\\ y\qquad\ +\ \ w&=0\\ 2z+\ \ w&=5 \end{aligned}\quad\xrightarrow{-2\rho_1+\rho_2}\quad\begin{aligned} x-y\qquad\qquad\quad &=0\\ z+2w&=4\\ y\qquad+\ \ w&=0\\ 2z+\ \ w&=5 \end{aligned}$$

the second equation has no leading y. To get one, we look lower down in the system for a row that has a leading y and swap it in.

$$\xrightarrow{\rho_2\leftrightarrow\rho_3}\begin{array}{rcl} x-y & & =0 \\ y & +\ w & =0 \\ z+2w & & =4 \\ 2z+\ w & & =5 \end{array}$$

(Had there been more than one row below the second with a leading y then we could have swapped in any one.) The rest of Gauss' method goes as before.

$$\xrightarrow{-2\rho_3+\rho_4}\begin{array}{rcl} x-y & & =\ 0 \\ y & +\ w & =\ 0 \\ z+2w & & =\ 4 \\ -3w & & =-3 \end{array}$$

Back-substitution gives $w=1$, $z=2$, $y=-1$, and $x=-1$.

Strictly speaking, the operation of rescaling rows is not needed to solve linear systems. We have included it because we will use it later in this chapter as part of a variation on Gauss' method, the Gauss-Jordan method.

All of the systems seen so far have the same number of equations as unknowns. All of them have a solution, and for all of them there is only one solution. We finish this subsection by seeing for contrast some other things that can happen.

1.11 Example Linear systems need not have the same number of equations as unknowns. This system

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

has more equations than variables. Gauss' method helps us understand this system also, since this

$$\xrightarrow[-2\rho_1+\rho_3]{-2\rho_1+\rho_2}\begin{aligned} x+\ 3y&=\ \ 1 \\ -5y&=-5 \\ -4y&=-4 \end{aligned}$$

shows that one of the equations is redundant. Echelon form

$$\xrightarrow{-(4/5)\rho_2+\rho_3}\begin{aligned} x+\ 3y&=\ \ 1 \\ -5y&=-5 \\ 0&=\ \ 0 \end{aligned}$$

gives $y=1$ and $x=-2$. The '$0=0$' is derived from the redundancy.

That example's system has more equations than variables. Gauss' method is also useful on systems with more variables than equations. Many examples are in the next subsection.

Another way that linear systems can differ from the examples shown earlier is that some linear systems do not have a unique solution. This can happen in two ways.

The first is that it can fail to have any solution at all.

1.12 Example Contrast the system in the last example with this one.

$$\begin{array}{rcl} x+3y&=&1 \\ 2x+\ \ y&=&-3 \\ 2x+2y&=&0 \end{array} \xrightarrow[-2\rho_1+\rho_3]{-2\rho_1+\rho_2} \begin{array}{rcl} x+\ \ 3y&=&1 \\ -5y&=&-5 \\ -4y&=&-2 \end{array}$$

Here the system is inconsistent: no pair of numbers satisfies all of the equations simultaneously. Echelon form makes this inconsistency obvious.

$$\xrightarrow{-(4/5)\rho_2+\rho_3} \begin{array}{rcl} x+\ \ 3y&=&1 \\ -5y&=&-5 \\ 0&=&2 \end{array}$$

The solution set is empty.

1.13 Example The prior system has more equations than unknowns, but that is not what causes the inconsistency — Example 1.11 has more equations than unknowns and yet is consistent. Nor is having more equations than unknowns necessary for inconsistency, as is illustrated by this inconsistent system with the same number of equations as unknowns.

$$\begin{array}{rcl} x+2y&=&8 \\ 2x+4y&=&8 \end{array} \xrightarrow{-2\rho_1+\rho_2} \begin{array}{rcl} x+2y&=&8 \\ 0&=&-8 \end{array}$$

The other way that a linear system can fail to have a unique solution is to have many solutions.

1.14 Example In this system

$$\begin{array}{rcl} x+\ \ y&=&4 \\ 2x+2y&=&8 \end{array}$$

any pair of numbers satisfying the first equation automatically satisfies the second. The solution set $\{(x,y) \mid x+y=4\}$ is infinite — some of its members are $(0,4)$, $(-1,5)$, and $(2.5,1.5)$. The result of applying Gauss' method here contrasts with the prior example because we do not get a contradictory equation.

$$\xrightarrow{-2\rho_1+\rho_2} \begin{array}{rcl} x+y&=&4 \\ 0&=&0 \end{array}$$

Don't be fooled by the '$0 = 0$' equation in that example. It is not the signal that a system has many solutions.

1.15 Example The absence of a '$0 = 0$' does not keep a system from having many different solutions. This system is in echelon form

$$\begin{aligned} x + y + z &= 0 \\ y + z &= 0 \end{aligned}$$

has no '$0 = 0$', and yet has infinitely many solutions. (For instance, each of these is a solution: $(0, 1, -1)$, $(0, 1/2, -1/2)$, $(0, 0, 0)$, and $(0, -\pi, \pi)$. There are infinitely many solutions because any triple whose first component is 0 and whose second component is the negative of the third is a solution.)

Nor does the presence of a '$0 = 0$' mean that the system must have many solutions. Example 1.11 shows that. So does this system, which does not have many solutions — in fact it has none — despite that when it is brought to echelon form it has a '$0 = 0$' row.

$$\begin{array}{rcl} \begin{aligned} 2x \quad\;\; - 2z &= 6 \\ y + z &= 1 \\ 2x + y - z &= 7 \\ 3y + 3z &= 0 \end{aligned} & \xrightarrow{-\rho_1+\rho_3} & \begin{aligned} 2x \quad - 2z &= 6 \\ y + z &= 1 \\ y + z &= 1 \\ 3y + 3z &= 0 \end{aligned} \\ & \xrightarrow[-3\rho_2+\rho_4]{-\rho_2+\rho_3} & \begin{aligned} 2x \quad - 2z &= 6 \\ y + z &= 1 \\ 0 &= 0 \\ 0 &= -3 \end{aligned} \end{array}$$

We will finish this subsection with a summary of what we've seen so far about Gauss' method.

Gauss' method uses the three row operations to set a system up for back substitution. If any step shows a contradictory equation then we can stop with the conclusion that the system has no solutions. If we reach echelon form without a contradictory equation, and each variable is a leading variable in its row, then the system has a unique solution and we find it by back substitution. Finally, if we reach echelon form without a contradictory equation, and there is not a unique solution (at least one variable is not a leading variable) then the system has many solutions.

The next subsection deals with the third case — we will see how to describe the solution set of a system with many solutions.

Exercises

✓ **1.16** Use Gauss' method to find the unique solution for each system.

(a) $\begin{aligned} 2x + 3y &= 13 \\ x - y &= -1 \end{aligned}$ **(b)** $\begin{aligned} x \quad - z &= 0 \\ 3x + y \quad &= 1 \\ -x + y + z &= 4 \end{aligned}$

✓ **1.17** Use Gauss' method to solve each system or conclude 'many solutions' or 'no solutions'.

(a) $\begin{aligned} 2x + 2y &= 5 \\ x - 4y &= 0 \end{aligned}$ **(b)** $\begin{aligned} -x + y &= 1 \\ x + y &= 2 \end{aligned}$ **(c)** $\begin{aligned} x - 3y + z &= 1 \\ x + y + 2z &= 14 \end{aligned}$

(d) $\begin{aligned} -x - y &= 1 \\ -3x - 3y &= 2 \end{aligned}$ **(e)** $\begin{aligned} 4y + z &= 20 \\ 2x - 2y + z &= 0 \\ x + z &= 5 \\ x + y - z &= 10 \end{aligned}$ **(f)** $\begin{aligned} 2x + z + w &= 5 \\ y - w &= -1 \\ 3x - z - w &= 0 \\ 4x + y + 2z + w &= 9 \end{aligned}$

✓ **1.18** There are methods for solving linear systems other than Gauss' method. One often taught in high school is to solve one of the equations for a variable, then substitute the resulting expression into other equations. That step is repeated until there is an equation with only one variable. From that, the first number in the solution is derived, and then back-substitution can be done. This method both takes longer than Gauss' method, since it involves more arithmetic operations and is more likely to lead to errors. To illustrate how it can lead to wrong conclusions, we will use the system

$$\begin{aligned} x + 3y &= 1 \\ 2x + y &= -3 \\ 2x + 2y &= 0 \end{aligned}$$

from Example 1.12.

(a) Solve the first equation for x and substitute that expression into the second equation. Find the resulting y.

(b) Again solve the first equation for x, but this time substitute that expression into the third equation. Find this y.

What extra step must a user of this method take to avoid erroneously concluding a system has a solution?

✓ **1.19** For which values of k are there no solutions, many solutions, or a unique solution to this system?

$$\begin{aligned} x - y &= 1 \\ 3x - 3y &= k \end{aligned}$$

✓ **1.20** This system is not linear:

$$\begin{aligned} 2\sin\alpha - \cos\beta + 3\tan\gamma &= 3 \\ 4\sin\alpha + 2\cos\beta - 2\tan\gamma &= 10 \\ 6\sin\alpha - 3\cos\beta + \tan\gamma &= 9 \end{aligned}$$

but we can nonetheless apply Gauss' method. Do so. Does the system have a solution?

✓ **1.21** What conditions must the constants, the b's, satisfy so that each of these systems has a solution? *Hint.* Apply Gauss' method and see what happens to the right side.

(a) $\begin{aligned} x - 3y &= b_1 \\ 3x + y &= b_2 \\ x + 7y &= b_3 \\ 2x + 4y &= b_4 \end{aligned}$ **(b)** $\begin{aligned} x_1 + 2x_2 + 3x_3 &= b_1 \\ 2x_1 + 5x_2 + 3x_3 &= b_2 \\ x_1 + 8x_3 &= b_3 \end{aligned}$

1.22 True or false: a system with more unknowns than equations has at least one solution. (As always, to say 'true' you must prove it, while to say 'false' you must produce a counterexample.)

1.23 Must any Chemistry problem like the one that starts this subsection — a balance the reaction problem — have infinitely many solutions?

✓ **1.24** Find the coefficients a, b, and c so that the graph of $f(x) = ax^2 + bx + c$ passes through the points $(1, 2)$, $(-1, 6)$, and $(2, 3)$.

1.25 Gauss' method works by combining the equations in a system to make new equations.

(a) Can the equation $3x-2y=5$ be derived, by a sequence of Gaussian reduction steps, from the equations in this system?

$$\begin{aligned} x + y &= 1 \\ 4x - y &= 6 \end{aligned}$$

(b) Can the equation $5x-3y=2$ be derived, by a sequence of Gaussian reduction steps, from the equations in this system?

$$\begin{aligned} 2x + 2y &= 5 \\ 3x + y &= 4 \end{aligned}$$

(c) Can the equation $6x-9y+5z=-2$ be derived, by a sequence of Gaussian reduction steps, from the equations in the system?

$$\begin{aligned} 2x + y - z &= 4 \\ 6x - 3y + z &= 5 \end{aligned}$$

1.26 Prove that, where $a, b, \ldots, e$ are real numbers and $a \neq 0$, if

$$ax + by = c$$

has the same solution set as

$$ax + dy = e$$

then they are the same equation. What if $a = 0$?

✓ **1.27** Show that if $ad - bc \neq 0$ then

$$\begin{aligned} ax + by &= j \\ cx + dy &= k \end{aligned}$$

has a unique solution.

✓ **1.28** In the system

$$\begin{aligned} ax + by &= c \\ dx + ey &= f \end{aligned}$$

each of the equations describes a line in the xy-plane. By geometrical reasoning, show that there are three possibilities: there is a unique solution, there is no solution, and there are infinitely many solutions.

1.29 Finish the proof of Theorem 1.4.

1.30 Is there a two-unknowns linear system whose solution set is all of $\mathbb{R}^2$?

✓ **1.31** Are any of the operations used in Gauss' method redundant? That is, can any of the operations be synthesized from the others?

1.32 Prove that each operation of Gauss' method is reversible. That is, show that if two systems are related by a row operation $S_1 \leftrightarrow S_2$ then there is a row operation to go back $S_2 \leftrightarrow S_1$.

1.33 A box holding pennies, nickels and dimes contains thirteen coins with a total value of 83 cents. How many coins of each type are in the box?

1.34 [Con. Prob. 1955] Four positive integers are given. Select any three of the integers, find their arithmetic average, and add this result to the fourth integer. Thus the numbers 29, 23, 21, and 17 are obtained. One of the original integers is:

(a) 19 **(b)** 21 **(c)** 23 **(d)** 29 **(e)** 17

✓ **1.35** [Am. Math. Mon., Jan. 1935] Laugh at this: AHAHA + TEHE = TEHAW. It resulted from substituting a code letter for each digit of a simple example in addition, and it is required to identify the letters and prove the solution unique.

1.36 [Wohascum no. 2] The Wohascum County Board of Commissioners, which has 20 members, recently had to elect a President. There were three candidates (A, B, and C); on each ballot the three candidates were to be listed in order of preference, with no abstentions. It was found that 11 members, a majority, preferred A over B (thus the other 9 preferred B over A). Similarly, it was found that 12 members preferred C over A. Given these results, it was suggested that B should withdraw, to enable a runoff election between A and C. However, B protested, and it was then found that 14 members preferred B over C! The Board has not yet recovered from the resulting confusion. Given that every possible order of A, B, C appeared on at least one ballot, how many members voted for B as their first choice?

1.37 [Am. Math. Mon., Jan. 1963] "This system of n linear equations with n unknowns," said the Great Mathematician, "has a curious property."

"Good heavens!" said the Poor Nut, "What is it?"

"Note," said the Great Mathematician, "that the constants are in arithmetic progression."

"It's all so clear when you explain it!" said the Poor Nut. "Do you mean like $6x + 9y = 12$ and $15x + 18y = 21$?"

"Quite so," said the Great Mathematician, pulling out his bassoon. "Indeed, the system has a unique solution. Can you find it?"

"Good heavens!" cried the Poor Nut, "I am baffled."

Are you?

1.I.2 Describing the Solution Set

A linear system with a unique solution has a solution set with one element. A linear system with no solution has a solution set that is empty. In these cases the solution set is easy to describe. Solution sets are a challenge to describe only when they contain many elements.

2.1 Example This system has many solutions because in echelon form

$$\begin{array}{rcl} 2x \quad\;\; + z = 3 \\ x - y - z = 1 \\ 3x - y \quad\;\; = 4 \end{array} \xrightarrow[-(3/2)\rho_1+\rho_3]{-(1/2)\rho_1+\rho_2} \begin{array}{rcl} 2x \quad + \quad z = 3 \\ -y - (3/2)z = -1/2 \\ -y - (3/2)z = -1/2 \end{array}$$

$$\xrightarrow{-\rho_2+\rho_3} \begin{array}{rcl} 2x \quad + \quad z = 3 \\ -y - (3/2)z = -1/2 \\ 0 = 0 \end{array}$$

not all of the variables are leading variables. The Gauss' method theorem showed that a triple satisfies the first system if and only if it satisfies the third. Thus, the solution set $\{(x, y, z) \mid 2x + z = 3 \text{ and } x - y - z = 1 \text{ and } 3x - y = 4\}$

can also be described as $\{(x, y, z) \mid 2x + z = 3 \text{ and } -y - 3z/2 = -1/2\}$. However, this second description is not much of an improvement. It has two equations instead of three, but it still involves some hard-to-understand interaction among the variables.

To get a description that is free of any such interaction, we take the variable that does not lead any equation, z, and use it to describe the variables that do lead, x and y. The second equation gives $y = (1/2) - (3/2)z$ and the first equation gives $x = (3/2) - (1/2)z$. Thus, the solution set can be described as $\{(x, y, z) = ((3/2) - (1/2)z, (1/2) - (3/2)z, z) \mid z \in \mathbb{R}\}$. For instance, $(1/2, -5/2, 2)$ is a solution because taking $z = 2$ gives a first component of $1/2$ and a second component of $-5/2$.

The advantage of this description over the ones above is that the only variable appearing, z, is unrestricted — it can be any real number.

2.2 Definition The non-leading variables in an echelon-form linear system are *free variables.*

In the echelon form system derived in the above example, x and y are leading variables and z is free.

2.3 Example A linear system can end with more than one variable free. This row reduction

$$\begin{array}{rcl} \begin{aligned} x + y + z - w &= 1 \\ y - z + w &= -1 \\ 3x \quad + 6z - 6w &= 6 \\ -y + z - w &= 1 \end{aligned} & \xrightarrow{-3\rho_1+\rho_3} & \begin{aligned} x + y + z - w &= 1 \\ y - z + w &= -1 \\ -3y + 3z - 3w &= 3 \\ -y + z - w &= 1 \end{aligned} \\ & \xrightarrow[\rho_2+\rho_4]{3\rho_2+\rho_3} & \begin{aligned} x + y + z - w &= 1 \\ y - z + w &= -1 \\ 0 &= 0 \\ 0 &= 0 \end{aligned} \end{array}$$

ends with x and y leading, and with both z and w free. To get the description that we prefer we will start at the bottom. We first express y in terms of the free variables z and w with $y = -1 + z - w$. Next, moving up to the top equation, substituting for y in the first equation $x + (-1 + z - w) + z - w = 1$ and solving for x yields $x = 2 - 2z + 2w$. Thus, the solution set is $\{2 - 2z + 2w, -1 + z - w, z, w) \mid z, w \in \mathbb{R}\}$.

We prefer this description because the only variables that appear, z and w, are unrestricted. This makes the job of deciding which four-tuples are system solutions into an easy one. For instance, taking $z = 1$ and $w = 2$ gives the solution $(4, -2, 1, 2)$. In contrast, $(3, -2, 1, 2)$ is not a solution, since the first component of any solution must be 2 minus twice the third component plus twice the fourth.

2.4 Example After this reduction

$$\begin{array}{rcl} 2x-2y &=0 \\ z+3w &=2 \\ 3x-3y &=0 \\ x-y+2z+6w &=4 \end{array} \xrightarrow[-(1/2)\rho_1+\rho_4]{-(3/2)\rho_1+\rho_3} \begin{array}{rcl} 2x-2y &=0 \\ z+3w &=2 \\ 0 &=0 \\ 2z+6w &=4 \end{array}$$

$$\xrightarrow{-2\rho_2+\rho_4} \begin{array}{rcl} 2x-2y &=0 \\ z+3w &=2 \\ 0 &=0 \\ 0 &=0 \end{array}$$

x and z lead, y and w are free. The solution set is $\{(y, y, 2-3w, w) \mid y, w \in \mathbb{R}\}$. For instance, $(1, 1, 2, 0)$ satisfies the system — take $y = 1$ and $w = 0$. The four-tuple $(1, 0, 5, 4)$ is not a solution since its first coordinate does not equal its second.

We refer to a variable used to describe a family of solutions as a *parameter* and we say that the set above is *paramatrized* with y and w. (The terms 'parameter' and 'free variable' do not mean the same thing. Above, y and w are free because in the echelon form system they do not lead any row. They are parameters because they are used in the solution set description. We could have instead paramatrized with y and z by rewriting the second equation as $w = 2/3 - (1/3)z$. In that case, the free variables are still y and w, but the parameters are y and z. Notice that we could not have paramatrized with x and y, so there is sometimes a restriction on the choice of parameters. The terms 'parameter' and 'free' are related because, as we shall show later in this chapter, the solution set of a system can always be paramatrized with the free variables. Consequenlty, we shall paramatrize all of our descriptions in this way.)

2.5 Example This is another system with infinitely many solutions.

$$\begin{array}{rcl} x+2y &=1 \\ 2x +z &=2 \\ 3x+2y+z-w &=4 \end{array} \xrightarrow[-3\rho_1+\rho_3]{-2\rho_1+\rho_2} \begin{array}{rcl} x+2y &=1 \\ -4y+z &=0 \\ -4y+z-w &=1 \end{array}$$

$$\xrightarrow{-\rho_2+\rho_3} \begin{array}{rcl} x+2y &=1 \\ -4y+z &=0 \\ -w &=1 \end{array}$$

The leading variables are x, y, and w. The variable z is free. (Notice here that, although there are infinitely many solutions, the value of one of the variables is fixed — $w = -1$.) Write w in terms of z with $w = -1 + 0z$. Then $y = (1/4)z$. To express x in terms of z, substitute for y into the first equation to get $x = 1 - (1/2)z$. The solution set is $\{(1 - (1/2)z, (1/4)z, z, -1) \mid z \in \mathbb{R}\}$.

We finish this subsection by developing the notation for linear systems and their solution sets that we shall use in the rest of this book.

2.6 Definition An $m \times n$ *matrix* is a rectangular array of numbers with m *rows* and n *columns*. Each number in the matrix is an *entry*, Matrices are usually named by upper case roman letters, e.g. A. Each entry is denoted by the corresponding lower-case letter, e.g. $a_{i,j}$ is the number in row i and column j of the array. For instance,

$$A = \begin{pmatrix} 1 & 2.2 & 5 \\ 3 & 4 & -7 \end{pmatrix}$$

has two rows and three columns, and so is a 2×3 matrix. (Read that "two-by-three"; the number of rows is always stated first.) The entry in the second row and first column is $a_{2,1} = 3$. Note that the order of the subscripts matters: $a_{1,2} \neq a_{2,1}$ since $a_{1,2} = 2.2$. (The parentheses around the array are a typographic device so that when two matrices are side by side we can tell where one ends and the other starts.)

2.7 Example We can abbreviate this linear system

$$\begin{aligned} x_1 + 2x_2 \quad\quad &= 4 \\ x_2 - \ x_3 &= 0 \\ x_1 \quad\quad + 2x_3 &= 4 \end{aligned}$$

with this matrix.

$$\left(\begin{array}{ccc|c} 1 & 2 & 0 & 4 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 2 & 4 \end{array}\right)$$

The vertical bar just reminds a reader of the difference between the coefficients on the systems's left hand side and the constants on the right. When a bar is used to divide a matrix into parts, we call it an *augmented* matrix. In this notation, Gauss' method goes this way.

$$\left(\begin{array}{ccc|c} 1 & 2 & 0 & 4 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 2 & 4 \end{array}\right) \xrightarrow{-\rho_1+\rho_3} \left(\begin{array}{ccc|c} 1 & 2 & 0 & 4 \\ 0 & 1 & -1 & 0 \\ 0 & -2 & 2 & 0 \end{array}\right) \xrightarrow{2\rho_2+\rho_3} \left(\begin{array}{ccc|c} 1 & 2 & 0 & 4 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

The second row stands for $y - z = 0$ and the first row stands for $x + 2y = 4$ so the solution set is $\{(4-2z, z, z) \mid z \in \mathbb{R}\}$. One advantage of the new notation is that the clerical load of Gauss' method — the copying of variables, the writing of $+$'s and $=$'s, etc. — is lighter.

We will also use the array notation to clarify the descriptions of solution sets. A description like $\{(2-2z+2w, -1+z-w, z, w) \mid z, w \in \mathbb{R}\}$ from Example 2.3 is hard to read. We will rewrite it to group all the constants together, all the coefficients of z together, and all the coefficients of w together. We will write them vertically, in one-column wide matrices.

$$\{\begin{pmatrix} 2 \\ -1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \end{pmatrix} \cdot z + \begin{pmatrix} 2 \\ -1 \\ 0 \\ 1 \end{pmatrix} \cdot w \mid z, w \in \mathbb{R}\}$$