# 2

# Solving equations

*In 1985, John Hubbard was asked to testify before the Committee on Science and Technology of the U.S. House of Representatives. He was preceded by a chemist from DuPont, who spoke of modeling molecules, and by an official from the geophysics institute of California, who spoke of exploring for oil and attempting to predict tsunamis.*

*When it was his turn, he explained that when chemists model molecules, they are solving Schrödinger's equation, that exploring for oil requires solving the Gelfand-Levitan equation, and that predicting tsunamis means solving the Navier-Stokes equation. Astounded, the chairman of the committee interrupted him and turned to the previous speakers. "Is that true, what Professor Hubbard says?" he demanded. "Is it true that what you do is solve equations?"*

## 2.0 INTRODUCTION

In every subject, language is intimately related to understanding.

"It is impossible to dissociate language from science or science from language, because every natural science always involves three things: the sequence of phenomena on which the science is based; the abstract concepts which call these phenomena to mind; and the words in which the concepts are expressed. To call forth a concept, a word is needed; to portray a phenomenon, a concept is needed. All three mirror one and the same reality."—Antoine Lavoisier, 1789.

"Professor Hubbard, you always underestimate the difficulty of vocabulary."—Helen Chigirinskaya, Cornell University, 1997.

All readers of this book will have solved systems of simultaneous *linear* equations. Such problems arise throughout mathematics and its applications, so a thorough understanding of the problem is essential.

What most students encounter in high school is systems of $n$ equations in $n$ unknowns, where $n$ might be general or might be restricted to $n = 2$ and $n = 3$. Such a system usually has a unique solution, but sometimes something goes wrong: some equations are "consequences of others" and have infinitely many solutions; other systems of equations are "incompatible" and have no solutions. The first half of this chapter is largely concerned with making these notions systematic.

A language has evolved to deal with these concepts: "linear transformation", "linear combination", "linear independence", "kernel", "span", "basis", and "dimension". These words may sound unfriendly, but they are actually quite transparent if thought of in terms of linear equations. They are needed to answer questions like, "How many equations are consequences of the others?" The relationship of these words to linear equations goes further. Theorems in linear algebra can be proved with abstract induction proofs, but students generally prefer the following method, which we discuss in this chapter:

> *Reduce the statement to a statement about linear equations, row reduce the resulting matrix, and see whether the statement becomes obvious.*

If so, the statement is true; otherwise it is likely to be false.

In Section 2.6 we discuss abstract vector spaces and the change of basis. In Section 2.7 we explore the advantages of expressing a linear transformation in an *eigenbasis*, when such a basis exists, and how to find it when it does. The Perron-Frobenius theorem (Theorem 2.7.10) introduces the notion of a *leading eigenvalue*; Example 2.7.12 shows how Google uses this notion in its PageRank algorithm.

We added Example 2.7.12 on the PageRank algorithm and Example 3.8.10 on the use of *eigenfaces* to identify photographs in response to a request from computer scientists for more "big matrices" reflecting current applications of linear algebra.

Solving nonlinear equations is much harder than solving linear equations. In the days before computers, finding solutions was virtually impossible, and mathematicians had to be satisfied with proving that solutions existed. Today, knowing that a system of equations has solutions is no longer enough; we want a practical algorithm that will enable us to find them. The algorithm most often used is *Newton's method*. In Section 2.8 we show Newton's method in action and state Kantorovich's theorem, which guarantees that under appropriate circumstances Newton's method converges to a solution; in Section 2.9 we see when Newton's method superconverges and state a stronger version of Kantorovich's theorem.

In Section 2.10 we use Kantorovich's theorem to see when a function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ has a local inverse function, and when a function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$, with $n > m$, gives rise to an *implicit function* locally expressing some variables in terms of others.

## 2.1  The main algorithm: row reduction

Suppose we want to solve the system of linear equations

$$
\begin{aligned}
2x + y + 3z &= 1 \\
x - y \quad\;\; &= 1 \\
2x \quad\;\; + z &= 1.
\end{aligned}
\qquad 2.1.1
$$

We could add together the first and second equations to get $3x + 3z = 2$. Substituting $(2 - 3z)/3$ for $x$ in the third equation gives $z = 1/3$, so $x = 1/3$; putting this value for $x$ into the second equation then gives $y = -2/3$.

The big advantage of row reduction is that it requires no cleverness.

In this section we show how to make this approach systematic, using *row reduction*. The first step is to write the system of equations in matrix form:

$$
\underbrace{\begin{bmatrix} 2 & 1 & 3 \\ 1 & -1 & 0 \\ 2 & 0 & 1 \end{bmatrix}}_{\text{coefficient matrix } A}
\underbrace{\begin{bmatrix} x \\ y \\ z \end{bmatrix}}_{\text{vector of unknowns } \vec{\mathbf{x}}}
= \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{\text{constants } \vec{\mathbf{b}}},
\qquad 2.1.2
$$

which can be written as the matrix multiplication $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$:

$$
\overbrace{\begin{bmatrix} x \\ y \\ z \end{bmatrix}}^{\vec{\mathbf{x}}}
$$

$$
\underbrace{\begin{bmatrix} 2 & 1 & 3 \\ 1 & -1 & 0 \\ 2 & 0 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{\vec{\mathbf{b}}}.
\qquad 2.1.3
$$

We now use a shorthand notation, omitting the vector $\vec{\mathbf{x}}$ and writing $A$ and $\vec{\mathbf{b}}$ as a single matrix, with $\vec{\mathbf{b}}$ the last column of the new matrix:

$$\underbrace{\begin{bmatrix} 2 & 1 & 3 \\ 1 & -1 & 0 \\ 2 & 0 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{\vec{\mathbf{b}}}. \qquad 2.1.4$$

More generally, the system of equations

$$\begin{array}{ccccc} a_{1,1}x_1 & +\cdots+ & a_{1,n}x_n & = b_1 \\ \vdots & \cdots & \vdots & \vdots \\ a_{m,1}x_1 & +\cdots+ & a_{m,n}x_n & = b_m \end{array} \qquad 2.1.5$$

is the same as $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$:

$$\underbrace{\begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}_{\vec{\mathbf{x}}} = \underbrace{\begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}}_{\vec{\mathbf{b}}}, \qquad 2.1.6$$

represented by

$$\underbrace{\begin{bmatrix} a_{1,1} & \cdots & a_{1,n} & b_1 \\ \vdots & \cdots & \vdots & \vdots \\ a_{m,1} & \cdots & a_{m,n} & b_m \end{bmatrix}}_{[A|\vec{\mathbf{b}}]}. \qquad 2.1.7$$

We denote by $[A\,|\,\vec{\mathbf{b}}]$ the matrix obtained by putting $\vec{\mathbf{b}}$ next to the columns of $A$. The $i$th column of the matrix $A$ contains the coefficients of $x_i$; the rows of $[A\,|\,\vec{\mathbf{b}}]$ represent equations. The vertical line in $[A\,|\,\vec{\mathbf{b}}]$ is intended to avoid confusion with multiplication; we are not multiplying $A$ and $\vec{\mathbf{b}}$.

## Row operations

We can solve a system of linear equations $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$ by *row reducing* the matrix $[A\,|\,\vec{\mathbf{b}}]$, using *row operations*.

**Definition 2.1.1 (Row operations).** A *row operation* on a matrix is one of three operations:

1. Multiplying a row by a nonzero number
2. Adding a multiple of a row onto another row
3. Exchanging two rows

Row operations are important for two reasons. First, they require only arithmetic: addition, subtraction, multiplication, and division. This is what computers do well; in some sense it is all they can do. They spend a lot of time doing it: row operations are fundamental to most other mathematical

Note that writing the system of equations 2.1.5 in the form 2.1.4 uses position to impart information. In equations 2.1.5 we could write the terms of

$$a_{i,1}x_1 + \cdots + a_{i,n}x_n$$

in any order; in 2.1.4 the coefficient of $x_1$ must be in the first column, the coefficient of $x_2$ in the second column, and so on.

Using position to impart information allows for concision; in Roman numerals, $4\,084$ is
MMMMLXXXIIII.
(When we write IV = 4 and VI = 6 we are using position, but the Romans themselves were quite happy writing their numbers in any order, MMXXM for $3\,020$, for example.)

Recall that the first subscript in a pair of subscripts refers to vertical position, and the second to horizontal position: $a_{i,j}$ is the entry in the $i$th row, $j$th column: *first take the elevator, then walk down the hall.*

Exercise 2.1.4 asks you to show that the third operation is not necessary; one can exchange rows using operations 1 and 2.

*Column operations* are defined by replacing *row* by *column*. We will use column operations in Section 4.8.

algorithms. The other reason is that row operations enable us to solve systems of linear equations:

> **Theorem 2.1.2 (Solutions of $A\vec{x} = \vec{b}$ unchanged by row operations).** *If the matrix $[A|\vec{b}]$ representing a system of linear equations $A\vec{x} = \vec{b}$ can be turned into $[A'|\vec{b}']$ by a sequence of row operations, then the set of solutions of $A\vec{x} = \vec{b}$ and the set of solutions of $A'\vec{x} = \vec{b}'$ coincide.*

**Proof.** Row operations consist of multiplying one equation by a nonzero number, adding a multiple of one equation to another, and exchanging two equations. Any solution of $A\vec{x} = \vec{b}$ is thus a solution of $A'\vec{x} = \vec{b}'$. In the other direction, any row operation can be undone by another row operation (Exercise 2.1.5), so any solution $A'\vec{x} = \vec{b}'$ is also a solution of $A\vec{x} = \vec{b}$.  $\square$

Theorem 2.1.2 suggests that we solve $A\vec{x} = \vec{b}$ by using row operations to bring the system of equations to the most convenient form. In Example 2.1.3 we apply this technique to equation 2.1.1. For now, don't worry about how the row reduction was achieved. Concentrate instead on what the row-reduced matrix tells us about solutions to the system of equations.

**Example 2.1.3 (Solving a system of equations with row operations).** To solve the system of equations 2.1.1 we can use row operations to bring the matrix

$$\begin{bmatrix} 2 & 1 & 3 & 1 \\ 1 & -1 & 0 & 1 \\ 2 & 0 & 1 & 1 \end{bmatrix} \quad \text{to the form} \quad \begin{bmatrix} 1 & 0 & 0 & 1/3 \\ 0 & 1 & 0 & -2/3 \\ 0 & 0 & 1 & 1/3 \end{bmatrix}. \qquad 2.1.8$$

$$\underbrace{\phantom{xxxxx}}_{\widetilde{A}} \underbrace{\phantom{xx}}_{\widetilde{\mathbf{b}}}$$

(To distinguish the new $A$ and $\vec{b}$ from the old, we put a "tilde" on top: $\widetilde{A}, \widetilde{\mathbf{b}}$; to lighten notation, we drop the arrow on the $\mathbf{b}$.) In this case, the solution can just be read off the matrix. If we put the unknowns back in the matrix, we get

$$\begin{bmatrix} x & 0 & 0 & 1/3 \\ 0 & y & 0 & -2/3 \\ 0 & 0 & z & 1/3 \end{bmatrix} \quad \text{or} \quad \begin{matrix} x = & 1/3 \\ y = & -2/3 \\ z = & 1/3 \end{matrix} \quad \triangle \qquad 2.1.9$$

## Echelon form

Some systems of linear equations may have no solutions, and others may have infinitely many. But if a system has solutions, they can be found by an appropriate sequence of row operations, called *row reduction*, bringing the matrix to *echelon form*, as in the second matrix of formula 2.1.8.

---

Formula 2.1.8: We said not to worry about how we did this row reduction. But if you do worry, here are the steps: To get (1), divide row 1 of the original matrix by 2, and add $-1/2$ row 1 to row 2, and subtract row 1 from row 3. To get from (1) to (2), multiply row 2 of the matrix (1) by $-2/3$, and then add that result to row 3. From (2) to (3), subtract half of row 2 from row 1. For (4), subtract row 3 from row 1. For (5), subtract row 3 from row 2.

$$1. \begin{bmatrix} 1 & 1/2 & 3/2 & 1/2 \\ 0 & -3/2 & -3/2 & 1/2 \\ 0 & -1 & -2 & 0 \end{bmatrix}$$

$$2. \begin{bmatrix} 1 & 1/2 & 3/2 & 1/2 \\ 0 & 1 & 1 & -1/3 \\ 0 & 0 & -1 & -1/3 \end{bmatrix}$$

$$3. \begin{bmatrix} 1 & 0 & 1 & 2/3 \\ 0 & 1 & 1 & -1/3 \\ 0 & 0 & 1 & 1/3 \end{bmatrix}$$

$$4. \begin{bmatrix} 1 & 0 & 0 & 1/3 \\ 0 & 1 & 1 & -1/3 \\ 0 & 0 & 1 & 1/3 \end{bmatrix}$$

$$5. \begin{bmatrix} 1 & 0 & 0 & 1/3 \\ 0 & 1 & 0 & -2/3 \\ 0 & 0 & 1 & 1/3 \end{bmatrix}$$