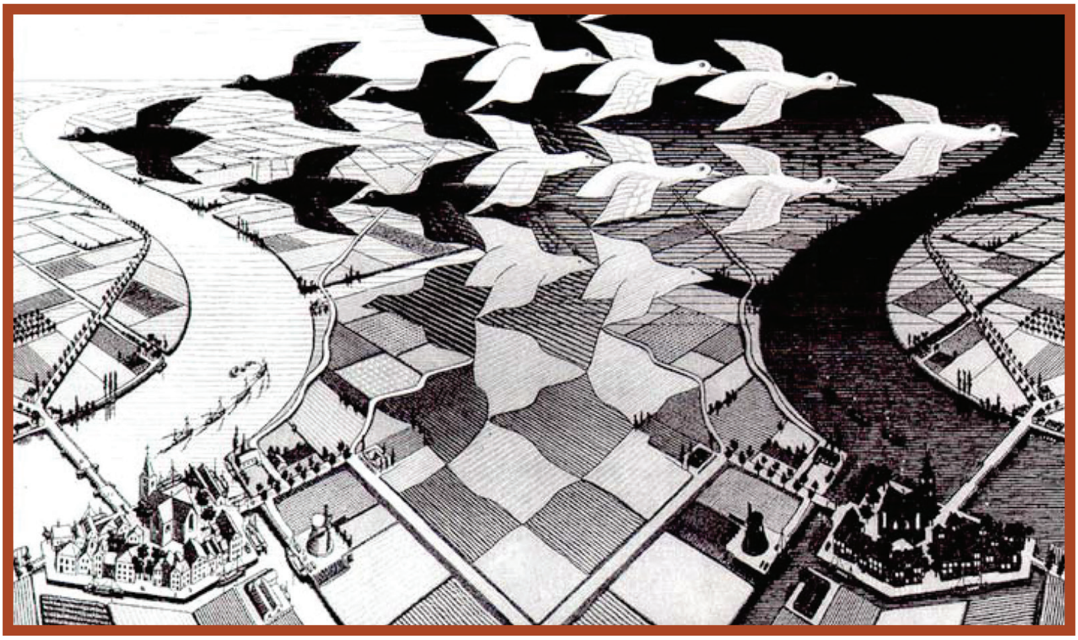
TEXTBOOKS in MATHEMATICS

LINEAR ALGEBRA
GEOMETRY AND TRANSFORMATION



BRUCE SOLOMON

WITH VITALSOURCE® EBOOK

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

LINEAR ALGEBRA
GEOMETRY AND TRANSFORMATION

TEXTBOOKS in MATHEMATICS

Series Editors: Al Boggess and Ken Rosen

PUBLISHED TITLES

ABSTRACT ALGEBRA: AN INQUIRY-BASED APPROACH
Jonathan K. Hodge, Steven Schlicker, and Ted Sundstrom

ABSTRACT ALGEBRA: AN INTERACTIVE APPROACH
William Paulsen

ADVANCED CALCULUS: THEORY AND PRACTICE
John Srdjan Petrovic

ADVANCED LINEAR ALGEBRA
Nicholas Loehr

ANALYSIS WITH ULTRASMALL NUMBERS
Karel Hrbacek, Olivier Lessmann, and Richard O'Donovan

APPLYING ANALYTICS: A PRACTICAL APPROACH
Evan S. Levine

COMPUTATIONS OF IMPROPER REIMANN INTEGRALS
Ioannis Roussos

CONVEX ANALYSIS
Steven G. Krantz

COUNTEREXAMPLES: FROM ELEMENTARY CALCULUS TO THE BEGINNINGS OF ANALYSIS
Andrei Bourchtein and Ludmila Bourchtein

DIFFERENTIAL EQUATIONS: THEORY, TECHNIQUE, AND PRACTICE, SECOND EDITION
Steven G. Krantz

DIFFERENTIAL EQUATIONS WITH MATLAB®: EXPLORATION, APPLICATIONS, AND THEORY
Mark A. McKibben and Micah D. Webster

ELEMENTARY NUMBER THEORY
James Kraft and Larry Washington

ELEMENTS OF ADVANCED MATHEMATICS, THIRD EDITION
Steven G. Krantz

EXPLORING LINEAR ALGEBRA: LABS AND PROJECTS WITH MATHEMATICA®
Crista Arangala

PUBLISHED TITLES CONTINUED

AN INTRODUCTION TO NUMBER THEORY WITH CRYPTOGRAPHY
James Kraft and Larry Washington

AN INTRODUCTION TO PARTIAL DIFFERENTIAL EQUATIONS WITH MATLAB®, SECOND EDITION
Mathew Coleman

INTRODUCTION TO THE CALCULUS OF VARIATIONS AND CONTROL WITH MODERN APPLICATIONS
John T. Burns

LINEAR ALGEBRA, GEOMETRY AND TRANSFORMATION
Bruce Solomon

THE MATHEMATICS OF GAMES: AN INTRODUCTION TO PROBABILITY
David G. Taylor

QUADRACTIC IRRATIONALS: AN INTRODUCTION TO CLASSICAL NUMBER THEORY
Franz Holter-Koch

REAL ANALYSIS AND FOUNDATIONS, THIRD EDITION
Steven G. Krantz

RISK ANALYSIS IN ENGINEERING AND ECONOMICS, SECOND EDITION
Bilal M. Ayyub

RISK MANAGEMENT AND SIMULATION
Aparna Gupta

TEXTBOOKS in MATHEMATICS

LINEAR ALGEBRA
GEOMETRY AND TRANSFORMATION

BRUCE SOLOMON
INDIANA UNIVERSITY, BLOOMINGTON
USA

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2015 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Version Date: 20141103

International Standard Book Number-13: 978-1-4822-9930-4 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

To my teachers...
...and to my students

Contents

Preface

> "*The eyes of the mind, by which it sees and observes things, are none other than proofs.*"
>
> —Baruch Spinoza

The organizing concept of this book is this: every topic should bring students closer to a solid geometric grasp of *linear transformations.*

Even more specifically, we aim to build a strong foundation for two enormously important results that no undergraduate math student should miss:

- The Spectral Theorem for symmetric transformations, and
- The Inverse/Implicit Function Theorem for differentiable mappings, or even better, the strong form of that result, sometimes called the Rank Theorem.

Every student who continues in math or its applications will encounter both these results in many contexts. The Spectral Theorem belongs to Linear Algebra proper; a course in the subject is simply remiss if it fails to get there. The Rank Theorem actually belongs to multivariable calculus, so we don't state or prove it here. Roughly, it says that a differentiable map of constant rank can be locally approximated by—and indeed, behaves geometrically just like—a *linear* map of the same rank. A student *cannot* understand this without a solid grasp of the *linear* case, which we do formulate and prove here as the *Linear Rank Theorem* in Chapter 7, making it, and the Spectral Theorem, key goals of our text.

The primacy we give those results motivates an unconventional start to our book, one that moves quickly to a first encounter with multivariable mappings and to the basic questions they raise about images, pre-images, injectivity, surjectivity, and distortion. While these are fundamental concerns throughout mathematics, they can be frustratingly difficult to analyze in general. The beauty and power of Linear

Algebra stem in large part from the utter transparency of these problems in the linear setting. A student who follows our discussion will apprehend them with a satisfying depth, and find them easy to apply in other areas of mathematical pursuit.

Of course, we cover all the standard topics of a first course in Linear Algebra—linear systems, vector geometry, matrix algebra, subspaces, independence, dimension, orthogonality, eigenvectors, and diagonalization. In our view, however, these topics mean more when they are directed toward the motivating results listed above.

We therefore introduce linear mappings and the basic questions they raise in our very first chapter, and aim the rest of our book toward answering those questions.

Key secondary themes emerge along the way. One is the centrality of the homogeneous system and the version of Gauss-Jordan we teach for solving it—and for expressing its solution as the span of independent "homogeneous generators." The number of such generators, for instance, gives the nullity of the system's coefficient matrix $\mathbf{A}$, which in turn answers basic questions about the structure of solutions to inhomogeneous systems having $\mathbf{A}$ as coefficient matrix, and about the linear transformation represented by $\mathbf{A}$.

Throughout, we celebrate the beautiful dualities that illuminate the subject:

- An $n \times m$ matrix $\mathbf{A}$ is both a list of rows, acting as linear functions on $\mathbf{R}^m$, and a list of columns, representing vectors in $\mathbf{R}^n$. Accordingly, we can interpret matrix/vector multiplication in dual ways: As a transformation of the input vector, or as a linear combination of the matrix columns. We stress the latter viewpoint more than many other authors, for it often delivers surprisingly clear insights.

- Similarly, an $n \times m$ system $\mathbf{Ax} = \mathbf{b}$ asks for the intersection of certain hyperplanes in $\mathbf{R}^m$, while simultaneously asking for ways to represent $\mathbf{b} \in \mathbf{R}^n$ as a linear combination of the columns of $\mathbf{A}$.

- The solution set of a homogeneous system can be alternatively expressed as the image (column-space) of one linear map, or as the pre-image (kernel) of another.

- The ubiquitous operations of addition and scalar multiplication manifest as pure algebra in the numeric vectorspaces $\mathbf{R}^n$, while simultaneously representing pure geometry in 2- and 3-dimensional Euclidean space.

- Every subspace of $\mathbf{R}^n$ can be described in essentially just two dual ways: as a *span*—the span of a generating set, or as an intersection of hyperplanes—what we call a *perp*.

We emphasize the computational and conceptual skills that let students navigate easily back and forth along any of these dualities, since problems posed from one perspective can often be solved with less effort from the dual viewpoint.

Finally, we strive to make all this material a ramp, lifting students from the computational mathematics that dominates their experience before this course, to the conceptual reasoning that often dominates after it. We move very consciously from simple "identity verification" proofs early on (where students check, using the definitions, for instance, that vector addition commutes, or that it distributes over dot products) to constructive and contrapositive arguments—e.g., the proof that the usual algorithm for inverting a matrix fulfills its mission. One can base many such arguments on reasoning about the outcome of the Gauss-Jordan algorithm—i.e., row-reduction and reduced row-echelon form—which students easily master. Linear algebra thus forms an ideal context for fostering and growing students' mathematical sophistication.

Our treatment omits abstract vector spaces, preferring to spend the limited time available in one academic term focusing on $\mathbf{R}^n$ and its subspaces, orthogonality and diagonalization. We feel that when students develop familiarity and the ability to reason well with $\mathbf{R}^n$ and—especially—its subspaces, the transition to abstract vector spaces, if and when they encounter it, will pose no difficulty.

Most of my students have been sophomores or juniors, typically majoring in math, informatics, one of the sciences, or business. The lack of an engineering school here has given my approach more of a liberal arts flavor, and allowed me to focus on the mathematics and omit applications. I know that for these very reasons, my book will not satisfy everyone. Still, I hope that all who read it will find themselves sharing the pleasure I always feel in learning, teaching, and writing about linear algebra.

Acknowledgments. This book springs from decades of teaching linear algebra, usually using other texts. I learned from each of those books, and from every group of students. About 10 years ago, Gilbert Strang's lively and unique introductory text inspired many ideas and

syntheses of my own, and I began to transition away from his book toward my own notes. These eventually took the course over, evolving into the present text. I thank all the authors, teachers, and students with whom I have learned to think about this beautiful subject, starting with the late Prof. Richard F. Arens, my undergraduate linear algebra teacher at UCLA.

Sincere thanks also go to CRC Press for publishing this work, and especially editor Bob Ross, who believed in the project and advocated for me within CRC.

I could not have reached this point without the unflagging support of my wife, family, and friends. I owe them more than I can express.

Indiana University and its math department have allowed me a life of continuous mathematical exploration and communication. A greater privilege is hard to imagine, and I am deeply grateful.

On a more technical note, I was lucky to have excellent software tools: TeXShop and LaTeX for writing and typesetting, along with Wolfram *Mathematica*®,[1] which I used to create all figures except Figure 28 in Chapter 3. The latter image of M.C. Escher's striking 1938 woodcut *Day and Night* (which also graces the cover) comes from the Official M.C. Escher website (`www.mcescher.com`).

Bruce Solomon
Indiana University
Bloomington, Indiana

[1]Wolfram *Mathematica*® is a registered trademark of Wolfram Research, Inc.

CHAPTER 1

Vectors, Mappings, and Linearity

1. Numeric Vectors

The overarching goal of this book is to impart a sure grasp of the numeric vector functions known as *linear transformations.* Students will have encountered *functions* before. We review and expand that familiarity in Section 2 below, and we define *linearity* in Section 4. Before we can properly discuss these matters though, we must introduce *numeric vectors* and their basic arithmetic.

Definition 1.1 (Vectors and scalars). A **numeric vector** (or just **vector** for short) is an ordered n-tuple of the form $(x_1, x_2, \ldots, x_n)$. Here, each x_i—the i**th entry** (or ith **coordinate**) of the vector—is a real number.

The (x, y) pairs often used to label points in the plane are familiar examples of vectors with $n = 2$, but we allow more than two entries as well. For instance, the triple $(3, -1/2, 2)$, and the 7-tuple $(1, 0, 2, 0, -2, 0, -1)$ are also numeric vectors.

In the linear algebraic setting, we usually call single numbers **scalars**. This helps highlight the difference between numeric vectors and individual numbers. □

Vectors can have many entries, so to clarify and save space, we often label them with single bold letters instead of writing out all their entries. For example, we might define

$$\begin{aligned}
\mathbf{x} &:= (x_1, x_2, \ldots, x_n)\\
\mathbf{a} &:= (a_1, a_2, a_3, a_4)\\
\mathbf{b} &:= (-5, 0, 1)
\end{aligned}$$

and then use $\mathbf{x}$, $\mathbf{a}$, or $\mathbf{b}$ to indicate the associated vector. We use boldface to distinguish vectors from scalars. For instance, the same letters, without boldface, would typically represent scalars, as in $x = 5$, $a = -4.2$, or $b = \pi$.

Often, we write numeric vectors vertically instead of horizontally, in which case $\mathbf{x}$, $\mathbf{a}$, and $\mathbf{b}$ above would look like this:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}, \quad \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} -5 \\ 0 \\ 1 \end{pmatrix}$$

In our approach to the subject (unlike some others) we draw absolutely no distinction between

$$(x_1, x_2, \ldots, x_n) \quad \text{and} \quad \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

These are merely different notations for the same vector—the very same mathematical object.

Definition 1.2. We denote the set of all scalars—also known as the real number line—by $\mathbf{R}^1$ or simply $\mathbf{R}$.

Similarly, $\mathbf{R}^n$ denotes the collection of **all** numeric vectors with n entries; that is, all $(x_1, x_2, \ldots, x_n)$. The "all zero" vector $(0, 0, \ldots, 0) \in \mathbf{R}^n$ is called the **origin**, and denoted by $\mathbf{0}$. □

As examples, the vectors $\mathbf{x}$, $\mathbf{a}$, and $\mathbf{b}$ above belong to $\mathbf{R}^m$, $\mathbf{R}^4$, and $\mathbf{R}^3$, respectively. We express this symbolically with the "element of" symbol "$\in$":

$$\mathbf{x} \in \mathbf{R}^m, \ \mathbf{a} \in \mathbf{R}^4, \ \text{and } \mathbf{b} \in \mathbf{R}^3$$

If $\mathbf{a}$ does **not** lie in $\mathbf{R}^5$, we can write $\mathbf{a} \notin \mathbf{R}^5$.

$\mathbf{R}^m$ is more than just a *set*, though, because it supports two important algebraic operations: *vector addition* and *scalar multiplication.*

1.3. Vector addition. To add (or subtract) vectors in $\mathbf{R}^m$, we simply add (or subtract) coordinates, entry-by-entry. This is best depicted vertically. Here are two examples, one numeric and one symbolic:

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ -5 \\ 6 \end{pmatrix} = \begin{pmatrix} 1+4 \\ 2-5 \\ 3+6 \end{pmatrix} = \begin{pmatrix} 5 \\ -3 \\ 9 \end{pmatrix}$$

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} - \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 - c_1 \\ a_2 + b_2 - c_2 \\ a_3 + b_3 - c_3 \\ a_4 + b_4 - c_4 \end{pmatrix}$$

Adding the origin $\mathbf{0} \in \mathbf{R}^m$ to any vector obviously leaves it unchanged: $\mathbf{0} + \mathbf{x} = \mathbf{x}$ for any $\mathbf{x} \in \mathbf{R}^m$. For this reason, $\mathbf{0}$ is called the *additive identity* in $\mathbf{R}^m$.

Recall that addition of *scalars* is commutative and associative. That is, for any scalars x, y, and z we have

$$\begin{aligned} x + y &= y + x && \text{(Commutativity)} \\ (x + y) + z &= x + (y + z) && \text{(Associativity)} \end{aligned}$$

It follows easily that *vector* addition has these properties too:

PROPOSITION 1.4. *Given any three vectors* $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{R}^m$, *we have*

$$\begin{aligned} \mathbf{x} + \mathbf{y} &= \mathbf{y} + \mathbf{x} && \text{(Commutativity)} \\ (\mathbf{x} + \mathbf{y}) + \mathbf{z} &= \mathbf{x} + (\mathbf{y} + \mathbf{z}) && \text{(Associativity)} \end{aligned}$$

PROOF. We prove associativity, and leave commutativity as an exercise.

The associativity statement is an *identity*: it asserts that two things are equal. Our approach is a basic and useful one for proving such assertions: *Expand both sides of the identity to show individual entries, then simplify using the familiar algebra of scalars. If the simplified expressions can be made equal using legal algebraic moves, we have a proof.*

Here, we start with the left-hand side, labeling the coordinates of $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ using x_i, y_i, and z_i, and then using the definition of vector addition twice:

$$\begin{aligned} (\mathbf{x} + \mathbf{y}) + \mathbf{z} &= \left[\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \right] + \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{pmatrix} \\ &= \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_m + y_m \end{pmatrix} + \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{pmatrix} \\ &= \begin{pmatrix} (x_1 + y_1) + z_1 \\ (x_2 + y_2) + z_2 \\ \vdots \\ (x_m + y_m) + z_m \end{pmatrix} \end{aligned}$$

Similarly, for the right-hand side of the identity, we get

$$
\begin{aligned}
\mathbf{x} + (\mathbf{y} + \mathbf{z}) &= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} + \left[\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} + \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{pmatrix} \right] \\
&= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} + \begin{pmatrix} y_1 + z_1 \\ y_2 + z_2 \\ \vdots \\ y_m + z_m \end{pmatrix} \\
&= \begin{pmatrix} x_1 + (y_1 + z_1) \\ x_2 + (y_2 + z_2) \\ \vdots \\ x_m + (y_m + z_m) \end{pmatrix}
\end{aligned}
$$

The simplified expressions for the two sides are now very similar. The parentheses don't line up the same way on both sides, but we can fix that by using the associative law for *scalars*. The two sides then agree, exactly, and we have a proof.

In short, the associative law for vectors boils down, after simplification, to the associative law for scalars, which we already know. □

1.5. Scalar multiplication. The second fundamental operation in $\mathbf{R}^n$ is even simpler than vector addition. *Scalar multiplication* lets us multiply any vector $\mathbf{x} \in \mathbf{R}^m$ by an arbitrary scalar t to get a new vector $t\,\mathbf{x}$. As with vector addition, we execute it entry-by-entry:

$$
t\,\mathbf{x} = t \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} := \begin{pmatrix} t\,x_1 \\ t\,x_2 \\ \vdots \\ t\,x_m \end{pmatrix}
$$

For instance, $2\,(1,3,5) = (2,6,10)$ and $-3\,(1,1,0,1) = (-3,-3,0,-3)$, while $0\,\mathbf{x} = (0,0,\ldots,0)$ no matter what $\mathbf{x}$ is.

Recall that for *scalars*, multiplication distributes over addition. This means that for any scalars t, x, and y, we have

$$
t(x+y) = tx + ty
$$

Since scalar multiplication and vector addition both operate entry-by-entry, scalar multiplication distributes over *vector* addition too. This simple relationship between the two operations is truly fundamental in linear algebra. Indeed, we shall see in Section 4 below, that it models the concept of *linearity.*

PROPOSITION 1.6. *Scalar multiplication distributes over vector addition. That is, if* t *is any scalar and* $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_k$ *are arbitrary vectors in* $\mathbf{R}^m$, *we have*

$$t\,(\mathbf{x}_1+\mathbf{x}_2+\cdots+\mathbf{x}_k)=t\,\mathbf{x}_1+t\,\mathbf{x}_2+\cdots+t\,\mathbf{x}_k$$

PROOF. To keep things simple, we prove this for just two vectors $\mathbf{x},\mathbf{y}\in\mathbf{R}^m$. The argument for k vectors works exactly the same way.

Using the same approach we used in proving the associativity identity in Proposition 1.4, we expand both sides of the identity in individual entries, simplify, and observe that we get the same result either way.

Let $\mathbf{x}=(x_1,x_2,\ldots,x_m)$ and $\mathbf{y}=(y_1,y_2,\ldots,y_m)$ be any two vectors in $\mathbf{R}^m$. Then for each scalar t, the left-hand side of the identity expands like this:

$$\begin{aligned} & t\,(\mathbf{x}+\mathbf{y}) \\ &= t\begin{pmatrix} x_1+y_1\\ x_2+y_2\\ \vdots\\ x_m+y_m\end{pmatrix}=\begin{pmatrix} t(x_1+y_1)\\ t(x_2+y_2)\\ \vdots\\ t(x_m+y_m)\end{pmatrix}=\begin{pmatrix} tx_1+ty_1\\ tx_2+ty_2\\ \vdots\\ tx_m+ty_m\end{pmatrix}\end{aligned}$$

While the right-hand side expands thus:

$$\begin{aligned} t\mathbf{x}+t\mathbf{y} &= t\begin{pmatrix} x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix}+t\begin{pmatrix} y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix}=\begin{pmatrix} tx_1\\ tx_2\\ \vdots\\ tx_n\end{pmatrix}+\begin{pmatrix} ty_1\\ ty_2\\ \vdots\\ ty_n\end{pmatrix}\\ &= \begin{pmatrix} tx_1+ty_1\\ tx_2+ty_2\\ \vdots\\ tx_m+ty_m\end{pmatrix}\end{aligned}$$

We get the same result either way, so the identity holds. □

1.7. Linear combination. We now define a third operation that combines scalar multiplication and vector addition. Actually, scalar multiplication and vector addition can be seen as mere special cases of this new operation:

DEFINITION 1.8. Given vectors $\mathbf{a_1}, \mathbf{a_2}, \ldots, \mathbf{a}_m \in \mathbf{R}^n$ and equally many scalars $x_1, x_2, \ldots, x_m$, the "weighted sum"

$$x_1 \, \mathbf{a}_1 \, + \, x_2 \, \mathbf{a}_2 \, + \, \cdots \, + \, x_m \, \mathbf{a}_m$$

is again a vector in $\mathbf{R}^n$. We call it a **linear combination** of the $\mathbf{a}_i$'s. We say that x_i is the **coefficient** of $\mathbf{a}_i$ in the linear combination. $\square$

EXAMPLE 1.9. Suppose $\mathbf{a}_1 = (1, -1, 0)$, $\mathbf{a}_2 = (0, 1, -1)$ and $\mathbf{a}_3 = (1, 0, -1)$. If we multiply these by the scalar coefficients $x_1 = 2$, $x_2 = -3$, and $x_3 = 4$, respectively and then add, we get the linear combination

$$\begin{aligned} 2\mathbf{a}_1 - 3\mathbf{a}_2 + 4\mathbf{a}_3 &= 2\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} - 3\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} + 4\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \\ &= \begin{pmatrix} 2-0+4 \\ -2-3+0 \\ 0+3-4 \end{pmatrix} \\ &= \begin{pmatrix} 6 \\ -5 \\ -1 \end{pmatrix} \end{aligned}$$

$\square$

Ultimately, many (perhaps most!) problems in linear algebra reduce to that of finding coefficients that linearly combine several given vectors to make a specified target vector. Here's an example. Because it involves just two vectors in $\mathbf{R}^2$, we can solve it by elementary methods.

EXAMPLE 1.10. Does some linear combination of $(2, 1)$ and $(-1, 2)$ add up to $(8, -1)$?

This is equivalent to asking if we can find coefficients x and y such that

$$x\begin{pmatrix} 2 \\ 1 \end{pmatrix} + y\begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 8 \\ -1 \end{pmatrix}$$

After performing the indicated scalar multiplications and vector addition, this becomes

$$\begin{pmatrix} 2x - y \\ x + 2y \end{pmatrix} = \begin{pmatrix} 8 \\ -1 \end{pmatrix}$$

Solving this for x and y is now clearly the same as simultaneously solving

$$\begin{aligned} 2x - y &= 8 \\ x + 2y &= 1 \end{aligned}$$

To do so, we can multiply the second equation by 2 and subtract it from the first to get

$$-3y = 6 \quad \text{hence} \quad y = -2$$

Setting $y = -2$ now reduces the first equation to $2x+2 = 8$, so $x = 3$. This solves our problem: With $x = 3$ and $y = -2$, we get a linear combination of the given vectors that adds up to $(8, -1)$:

$$3\begin{pmatrix} 2 \\ 1 \end{pmatrix} - 2\begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 8 \\ -1 \end{pmatrix}$$

□

We end our introductory discussion of linear combination by introducing the *standard basis vectors* of $\mathbf{R}^n$. They play key roles later on.

DEFINITION 1.11. The **standard basis vectors** in $\mathbf{R}^n$ are the n numeric vectors

$$\begin{array}{ccl} \mathbf{e}_1 & = & (1,\ 0,\ 0,\ \ldots,\ 0,\ 0) \\ \mathbf{e}_2 & = & (0,\ 1,\ 0,\ \ldots,\ 0,\ 0) \\ \mathbf{e}_3 & = & (0,\ 0,\ 1,\ \ldots,\ 0,\ 0) \\ \vdots & \vdots & \quad\ \vdots \\ \mathbf{e}_n & = & (0,\ 0,\ 0,\ \ldots,\ 0,\ 1) \end{array}$$

□

Simple as they are, these vectors are central to our subject. We introduce them here partly because problems like Example 1.10 and Exercises 6 and 7 become *trivial* when we're combining standard basis vectors, thanks to the following:

OBSERVATION 1.12. *We can express any numeric vector*

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)$$

as a linear combination of standard basis vectors in an obvious way:

$$\mathbf{x} = x_1\,\mathbf{e}_1 + x_2\,\mathbf{e}_2 + x_3\,\mathbf{e}_3 + \cdots + x_n\,\mathbf{e}_n$$

PROOF. Since $x_1\,\mathbf{e}_1 = (x_1, 0, 0, \ldots, 0)$, $x_2\,\mathbf{e}_2 = (0, x_2, 0, \ldots, 0)$ and so forth, the identity is easy to verify. □

1.13. Matrices. One of the most fundamental insights in linear algebra is simply this: *We can view any linear combination as the result of multiplying a vector by a* matrix:

DEFINITION 1.14 (Matrix). An $n \times m$ **matrix** is a rectangular array of scalars, with n horizontal **rows** (each in $\mathbf{R}^m$), and m vertical **columns** (each in $\mathbf{R}^n$). For instance:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & -2.5 \\ \pi & 4 & 1/2 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 0 & 1 \\ 2 & 0 \\ 3 & -3 \end{bmatrix}$$

Here $\mathbf{A}$ has 2 rows and 3 columns, while $\mathbf{B}$ has 3 rows, 2 columns.

We generally label matrices with bold uppercase letters, as with $\mathbf{A}$ and $\mathbf{B}$ above. We *double-subscript* the corresponding *lowercase* letter to address the **entries**—the individual scalars—in the matrix. So if we call a matrix $\mathbf{X}$, then x_{34} names the entry in row 3 and column 4 of $\mathbf{X}$.

With regard to $\mathbf{A}$ and $\mathbf{B}$ above, for example, we have

$$a_{21} = \pi, \quad a_{12} = 0, \quad a_{13} = -2.5, \quad \text{and} \quad b_{11} = b_{22} = 0\,.$$

We sometimes label a matrix $\mathbf{X}$ by $[x_{ij}]$ or write $\mathbf{X} = [x_{ij}]$ to emphasize that the entries of $\mathbf{X}$ will be called x_{ij}.

Finally, if we want to clarify that a matrix $\mathbf{C}$ has, say, 4 rows and 5 columns, we can call it $\mathbf{C}_{4\times 5}$. Just as with entries, the first index refers to rows, while the second refers to columns. □

1.15. Matrix addition and scalar multiplication. Matrices, like numeric vectors, can be scalar multiplied: When k is a scalar and $\mathbf{A}$ is a matrix, we simply multiply each entry in $\mathbf{A}$ by k to get $k\mathbf{A}$.

EXAMPLE 1.16. Suppose

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ -4 & -3 & -2 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix}$$

Then

$$\pi\mathbf{A} = \begin{bmatrix} \pi & 2\pi & 3\pi & 4\pi \\ -4\pi & -3\pi & -2\pi & -\pi \end{bmatrix} \quad \text{while} \quad 5\mathbf{B} = \begin{bmatrix} 5 & -5 & 0 \\ 0 & 5 & -5 \\ -5 & 0 & 5 \end{bmatrix}$$

□

Similarly, matrices *of the same size* can be added together. Again, just as with numeric vectors, we do this entry-by-entry:

EXAMPLE 1.17. If

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

then

$$\mathbf{A} + \mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{bmatrix} \quad \text{while} \quad \mathbf{A} + \mathbf{B} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 4 & 0 \\ 3 & 0 & 3 \end{bmatrix}$$

□

1.18. Matrix/vector products. The matrix/vector product we describe next is an operation much richer than either matrix addition or scalar multiplication. In particular, the matrix/vector product gives us a new and useful way to handle linear combination. The rule is very simple:

We can express any linear combination

$$x_1\,\mathbf{v}_1 + x_2\,\mathbf{v}_2 + \cdots + x_m\mathbf{v}_m$$

as a matrix/vector product, as follows:

Write the vectors $\mathbf{v}_i$ *as the* columns *of a matrix* $\mathbf{A}$*, and stack the coefficients* x_i *up as a vector* $\mathbf{x}$*. The given linear combination then agrees with the product* $\mathbf{Ax}$*.*

EXAMPLE 1.19. To write the linear combination

$$x\begin{pmatrix} 7 \\ -3 \end{pmatrix} + y\begin{pmatrix} -5 \\ 2 \end{pmatrix} + z\begin{pmatrix} 1 \\ -4 \end{pmatrix}$$

as a matrix/vector product, we then take the *vectors* in the linear combination, namely

$$\begin{pmatrix} 7 \\ -3 \end{pmatrix}, \begin{pmatrix} -5 \\ 2 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} 1 \\ -4 \end{pmatrix}$$

and line them up as columns in a matrix

$$\mathbf{A} = \begin{bmatrix} 7 & -5 & 1 \\ -3 & 2 & -4 \end{bmatrix}$$

We then stack the coefficients x, y, and z up as the vector

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

In short, we can now write the original linear combination, which was

$$x \begin{pmatrix} 7 \\ -3 \end{pmatrix} + y \begin{pmatrix} -5 \\ 2 \end{pmatrix} + z \begin{pmatrix} 1 \\ -4 \end{pmatrix}$$

as the matrix/vector product

$$\mathbf{A}\mathbf{x} = \begin{bmatrix} 7 & -5 & 1 \\ -3 & 2 & -4 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

Note that the coefficient vector $\mathbf{x} = (x, y, z)$ here lies in $\mathbf{R}^3$, while $\mathbf{A}\mathbf{x}$ lies in $\mathbf{R}^2$. Indeed, if we actually compute it, we get

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} 7x - 5y + z \\ -3x + 2y - 4z \end{pmatrix} \in \mathbf{R}^2$$

□

With this example in mind, we carefully state the general rule:

Definition 1.20 (Matrix/vector multiplication). If a matrix $\mathbf{A}$ has n rows and m columns, we can multiply it by any vector $\mathbf{x} \in \mathbf{R}^m$ to produce a result $\mathbf{A}\mathbf{x}$ in $\mathbf{R}^n$.

To compute it, we linearly combine the columns of $\mathbf{A}$ (each a vector in $\mathbf{R}^n$), using the entries of $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ as coefficients:

$$\mathbf{A}\mathbf{x} := x_1\,\mathbf{c}_1(\mathbf{A}) + x_2\,\mathbf{c}_2(\mathbf{A}) + \cdots + x_m\,\mathbf{c}_m(\mathbf{A})$$

where $\mathbf{c}_j(\mathbf{A})$ signifies column j of $\mathbf{A}$.

Conversely, any linear combination

$$x_1\,\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_m\mathbf{v}_m$$

can be written as the product $\mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is the matrix with columns $\mathbf{v}_1$, $\mathbf{v}_2$, $\ldots$, $\mathbf{v}_m$ (in that order) and $\mathbf{x} = (x_1, x_2, \ldots, x_m)$. Symbolically,

$$\mathbf{A} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \\ | & | & \cdots & | \end{bmatrix}, \qquad \mathbf{x} = (x_1, x_2, \ldots, x_m)$$

and then

$$\mathbf{A}\mathbf{x} = x_1\,\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_m\mathbf{v}_m$$

□

REMARK 1.21 (Warning!). We can **only** multiply $\mathbf{A}$ by $\mathbf{x}$ when *the number of columns in* $\mathbf{A}$ *equals the number of entries in* $\mathbf{x}$. When the vector $\mathbf{x}$ lies in $\mathbf{R}^m$, the matrix $\mathbf{A}$ must have exactly m columns. On the other hand, $\mathbf{A}$ can have any number n of rows. The product $\mathbf{Ax}$ will then lie in $\mathbf{R}^n$.

REMARK 1.22. It is useful to conceptualize matrix/vector multiplication via the following mnemonic "mantra":

Matrix/vector multiplication = *Linear combination*

Commit this phrase to memory—we will have many opportunities to invoke it. □

EXAMPLE 1.23. If

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ -4 & -3 \\ -1 & -2 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = (-1, 5)$$

then

$$\mathbf{Ax} = -1 \begin{pmatrix} 1 \\ 3 \\ -4 \\ -1 \end{pmatrix} + 5 \begin{pmatrix} 2 \\ 4 \\ -3 \\ -2 \end{pmatrix} = \begin{pmatrix} -1+10 \\ -3+20 \\ 4-15 \\ 1-10 \end{pmatrix} = \begin{pmatrix} 9 \\ 17 \\ -11 \\ -9 \end{pmatrix}$$

More generally, if $\mathbf{x} = (x, y)$, then

$$\mathbf{Ax} = x \begin{pmatrix} 1 \\ 3 \\ -4 \\ -1 \end{pmatrix} + y \begin{pmatrix} 2 \\ 4 \\ -3 \\ -2 \end{pmatrix} = \begin{pmatrix} x+2y \\ 3x+4y \\ -4x-3y \\ -x-2y \end{pmatrix}$$

Note how dramatically we abbreviate the expression on the right above when we write it as simply $\mathbf{Ax}$. □

1.24. Properties of matrix/vector multiplication. To continue our discussion of matrix/vector multiplication we record two crucial properties:

PROPOSITION 1.25. *Matrix/vector multiplication commutes with scalar multiplication, and distributes over vector addition. More precisely, if* $\mathbf{A}$ *is any* $n \times m$ *matrix, the following two facts always hold:*

i) *If* k *is any scalar and* $\mathbf{x} \in \mathbf{R}^m$, *then*

$$\mathbf{A}(k\mathbf{x}) = k(\mathbf{A}\mathbf{x}) = (k\mathbf{A})\mathbf{x}\,.$$

ii) *For any two vectors* $\mathbf{x}, \mathbf{y} \in \mathbf{R}^m$, *we have*

$$\mathbf{A}(\mathbf{x} + \mathbf{y}) = \mathbf{A}\mathbf{x} + \mathbf{A}\mathbf{y}\,.$$

PROOF. For simplicity here, we denote the columns of $\mathbf{A}$, respectively by $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$. We then prove (i) and (ii) in the usual way: we simplify each side of the equation separately and show that they agree.

Start with the first equality in (i). Expanding $\mathbf{x}$ as $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ we know that $k\,\mathbf{x} = k\,(x_1, x_2, \ldots, x_m) = (kx_1, kx_2, \ldots, kx_m)$. The definition of matrix/vector multiplication (Definition 1.20) then gives

$$\mathbf{A}(k\mathbf{x}) = kx_1\mathbf{a}_1 + kx_2\mathbf{a}_2 + \cdots + kx_m\mathbf{a}_m$$

Similarly, we can rewrite the middle expression in (i) as

$$\begin{aligned} k\,(\mathbf{A}\mathbf{x}) &= k\,(x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m) \\ &= kx_1\mathbf{a}_1 + kx_2\mathbf{a}_2 + \cdots + kx_m\mathbf{a}_m \end{aligned}$$

because scalar multiplication distributes over vector addition (Proposition 1.6). This expression matches exactly with what we got before. Since $\mathbf{A}$, k, and $\mathbf{x}$ were completely arbitrary, this proves the first equality in (i). We leave the reader to expand out $(k\mathbf{A})\mathbf{x}$ and show that it takes the same form.

A similar left/right comparison confirms (ii). Given arbitrary vectors $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_m)$ in $\mathbf{R}^m$, we have

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1,\ x_2 + y_2,\ \cdots\ ,\ x_m + y_m)$$

and hence

$$\begin{aligned} \mathbf{A}\,(\mathbf{x} + \mathbf{y}) &= (x_1 + y_1)\mathbf{a}_1 + (x_2 + y_2)\mathbf{a}_2 + \cdots + (x_m + y_m)\mathbf{a}_m \\ &= x_1\mathbf{a}_1 + y_1\mathbf{a}_1 + x_2\mathbf{a}_2 + y_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m + y_m\mathbf{a}_m \end{aligned}$$

by the definition of matrix/vector multiplication, and the distributive property (Proposition 1.6). When we simplify the right side of (ii), namely $\mathbf{A}\mathbf{x} + \mathbf{A}\mathbf{y}$, we get the same thing. (The summands come in a different order, but that's allowed, since vector addition is commutative, by Proposition 1.4). We leave this to the reader. □

1.26. The dot product. As we have noted, the matrix/vector product $\mathbf{Ax}$ makes sense only when the number of columns in $\mathbf{A}$ matches the number of entries in $\mathbf{x}$.

The number of *rows* in $\mathbf{A}$ will then match the number of entries in $\mathbf{Ax}$. So any number of rows is permissible—even just one.

In that case $\mathbf{Ax} \in \mathbf{R}^1 = \mathbf{R}$. So when $\mathbf{A}$ has just one row, $\mathbf{Ax}$ reduces to a single scalar.

EXAMPLE 1.27. Suppose we have

$$\mathbf{A} = \begin{bmatrix} -4 & 1 & 3 & -2 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$$

Then

$$\begin{aligned} \mathbf{Ax} &= \begin{bmatrix} -4 & 1 & 3 & -2 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} \\ &= 1(-4) + 1(1) - 1(3) - 1(-2) \\ &= -3 \end{aligned}$$

□

Note, however, that a $1 \times m$ matrix corresponds in an obvious way to a vector in $\mathbf{R}^m$. Seen in that light, matrix/vector multiplication provides a way to multiply two *vectors* $\mathbf{a}$ and $\mathbf{x}$ in $\mathbf{R}^m$: we just regard the first vector $\mathbf{a}$ as a $1 \times m$ matrix, and multiply it by $\mathbf{x}$ using matrix/vector multiplication. As noted above, this produces a scalar result.

Multiplying two vectors in $\mathbf{R}^m$ this way—by regarding the first vector as a $1 \times m$ matrix—is therefore sometimes called a **scalar product**. We simply call it the **dot product** since we indicate it with a dot.

DEFINITION 1.28 (Dot product). Given any two vectors

$$\mathbf{u} = (u_1, u_2, \ldots, u_m) \quad \text{and} \quad \mathbf{v} = (v_1, v_2, \ldots, v_m)$$

in $\mathbf{R}^m$, we define the **dot product** $\mathbf{u} \cdot \mathbf{v}$ via

$$\mathbf{u} \cdot \mathbf{v} := u_1 v_1 + u_2 v_2 + \cdots u_m v_m \tag{1}$$

bearing in mind that this is exactly what we get if we regard $\mathbf{u}$ as a $1 \times m$ matrix and multiply it by $\mathbf{v}$.

Effectively, however, this simply has us multiply the two vectors entry-by-entry, and then sum up the results. □

EXAMPLE 1.29. In $\mathbf{R}^2$,

$$\begin{pmatrix} 2 \\ -1 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 2 \end{pmatrix} = 2 \cdot 3 \ + \ (-1) \cdot 2 = 6 - 2 = 4$$

while in $\mathbf{R}^4$,

$$\begin{pmatrix} 2 \\ -1 \\ 0 \\ -1 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 2 \\ -1 \\ 1 \end{pmatrix} = 2 \cdot 3 \ + \ (-1) \cdot 2 \ + \ 0 \cdot (-1) \ + \ (-1) \cdot 1 = 3$$

□

PROPOSITION 1.30. *The dot product is commutative. It also commutes with scalar multiplication and distributes over vector addition. Thus, for any vectors* $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w} \in \mathbf{R}^n$, *we have*

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$$

$$\mathbf{u} \cdot (k\mathbf{v}) = k(\mathbf{u} \cdot \mathbf{v}) = (k\mathbf{u}) \cdot \mathbf{v}$$

$$\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$$

PROOF. We leave the proof of the first identity to the reader (Exercise 18). The last two identities follow straight from the matrix identities in Proposition 1.25, since the dot product can be seen as the "$(1 \times n)$ times $(n \times 1)$" case of matrix/vector multiplication. □

1.31. Fast matrix/vector multiplication via dot product. We have seen that the dot product (Definition 1.28) corresponds to matrix/vector multiplication with a one-rowed matrix. We now turn this around to see that the dot product gives an efficient way to compute matrix/vector products—without forming linear combinations. To see how, take any matrix $\mathbf{A}$ and vector $\mathbf{v}$, like these:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1m} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2m} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nm} \end{bmatrix}, \quad \text{and} \quad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix}$$

By the definition of matrix/vector multiplication (as a linear combination) we get

$$\mathbf{A}\mathbf{v} = v_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} + v_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix} + \cdots + v_m \begin{pmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{pmatrix}$$

Now carry out the scalar multiplications and vector additions to rewrite as a single vector:

$$\mathbf{A}\mathbf{v} = \begin{pmatrix} v_1 a_{11} + v_2 a_{12} + \cdots + v_m a_{1m} \\ v_1 a_{21} + v_2 a_{22} + \cdots + v_m a_{2m} \\ \vdots \quad\quad \vdots \quad\quad \vdots \\ v_1 a_{n1} + v_2 a_{n2} + \cdots + v_m a_{nm} \end{pmatrix}$$

Each entry is now a dot product! The first entry dots $\mathbf{v}$ with the first row of $\mathbf{A}$, the second entry dots $\mathbf{v}$ with the second row of $\mathbf{A}$, and so forth. In other words, we have:

Observation 1.32 (Dot-product formula for matrix/vector multiplication). *We can compute the product of any $n \times m$ matrix $\mathbf{A}$ with any vector $\mathbf{v} = (v_1, v_2, \ldots, v_m) \in \mathbf{R}^m$ as a vector of dot products:*

$$\mathbf{A}\mathbf{v} = \begin{pmatrix} \mathbf{r}_1(\mathbf{A}) \cdot \mathbf{v} \\ \mathbf{r}_2(\mathbf{A}) \cdot \mathbf{v} \\ \mathbf{r}_3(\mathbf{A}) \cdot \mathbf{v} \\ \vdots \\ \mathbf{r}_n(\mathbf{A}) \cdot \mathbf{v} \end{pmatrix}$$

where $\mathbf{r}_i(\mathbf{A})$ denotes row i of $\mathbf{A}$.

Example 1.33. Given

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 3 \\ 1 & 4 & -5 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{pmatrix} 3 \\ -2 \\ -7 \end{pmatrix}$$

we compute $\mathbf{A}\mathbf{v}$ using dot products as follows:

$$\mathbf{A}\mathbf{v} = \begin{pmatrix} \mathbf{r}_1(\mathbf{A}) \cdot \mathbf{v} \\ \mathbf{r}_2(\mathbf{A}) \cdot \mathbf{v} \\ \mathbf{r}_3(\mathbf{A}) \cdot \mathbf{v} \\ \vdots \\ \mathbf{r}_n(\mathbf{A}) \cdot \mathbf{v} \end{pmatrix} = \begin{pmatrix} (2, -1, 3) \cdot (3, -2, -7) \\ \\ (1, 4, -5) \cdot (3, -2, -7) \end{pmatrix} = \begin{pmatrix} -13 \\ 30 \end{pmatrix}$$

The reader will easily check that this against our definition of $\mathbf{Av}$, namely

$$3\begin{pmatrix}2\\1\end{pmatrix} - 2\begin{pmatrix}-1\\4\end{pmatrix} - 7\begin{pmatrix}3\\-5\end{pmatrix}$$

□

EXAMPLE 1.34. Similarly, given

$$\mathbf{A} = \begin{bmatrix} 3 & -1 \\ 2 & 2 \\ -1 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{pmatrix} 7 \\ -5 \end{pmatrix}$$

the dot-product approach gives

$$\mathbf{Av} = \begin{pmatrix} (3,-1)\cdot(7,-5) \\ \\ (2,2)\cdot(7,-5) \\ \\ (-1,3)\cdot(7,-5) \end{pmatrix} = \begin{pmatrix} 26 \\ 4 \\ -22 \end{pmatrix}$$

□

1.35. Eigenvectors. Among matrices, *square* matrices—matrices having the same number of rows and columns—are particularly interesting and important. One reason for their importance is this:

When we multiply a vector $\mathbf{x} \in \mathbf{R}^m$ *by a* ***square*** *matrix* $\mathbf{A}_{m\times m}$*, the product* $\mathbf{Ax}$ *lies in the* ***same*** *space as* $\mathbf{x}$ *itself:* $\mathbf{R}^m$*.*

This fact makes possible a phenomenon that unlocks some of the deepest ideas in linear algebra: *The product* $\mathbf{Ax}$ *may actually be a scalar multiple of the original vector* $\mathbf{x}$*.* That is, there may be certain "lucky" vectors $\mathbf{x} \in \mathbf{R}^m$ for which $\mathbf{Ax} = \lambda\mathbf{x}$, where λ (the Greek letter *lambda*) is some scalar.

DEFINITION 1.36 (Eigenvalues and eigenvectors). If $\mathbf{A}$ is an $m \times m$ matrix, and there exists a vector $\mathbf{x} \neq \mathbf{0}$ in $\mathbf{R}^m$ such that $\mathbf{Ax} = \lambda\mathbf{x}$ for some scalar $\lambda \in \mathbf{R}$, we call $\mathbf{x}$ an **eigenvector** of $\mathbf{A}$, and we call λ its **eigenvalue**. □

EXAMPLE 1.37. The vectors $(1,1)$ and $(-3,3)$ in $\mathbf{R}^2$ are eigenvectors of the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

but the vector $(2,1)$ is *not* an eigenvector. To verify these statements, we just multiply each vector by $\mathbf{A}$ and see whether the product is a scalar multiple of $\mathbf{A}$ or not. This is easy to verify using the dot-product method of matrix/vector multiplication:

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1+2 \\ 2+1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Thus, when $\mathbf{x} = (1,1)$, we have $\mathbf{Ax} = 3\mathbf{x}$. This makes $\mathbf{x}$ an eigenvector of $\mathbf{A}$ with eigenvalue $\lambda = 3$.

Similarly, when $\mathbf{x} = (-3,3)$, we have

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{pmatrix} -3 \\ 3 \end{pmatrix} = \begin{pmatrix} -3+6 \\ -6+3 \end{pmatrix} = \begin{pmatrix} 3 \\ -3 \end{pmatrix} = -1 \begin{pmatrix} -3 \\ 3 \end{pmatrix}$$

Thus, when $\mathbf{x} = (-3,3)$, we again have $\mathbf{Ax} = -\mathbf{x}$, which makes $\mathbf{x}$ an eigenvector of $\mathbf{A}$, this time with eigenvalue $\lambda = -1$.

On the other hand, when we multiply $\mathbf{A}$ by $\mathbf{x} = (2,1)$, we get

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2+2 \\ 4+1 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \end{pmatrix}$$

Since $(4,5)$ is *not* a scalar multiple of $(2,1)$, it is *not* an eigenvector of $\mathbf{A}$. $\square$

Example 1.38. The vector $\mathbf{x} = (2,3,0)$ is an eigenvector of the matrix

$$\mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$$

since (again by the dot-product method of matrix/vector multiplication)

$$\mathbf{Bx} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 2+6+0 \\ 0+12+0 \\ 0+0+0 \end{pmatrix} = \begin{pmatrix} 8 \\ 12 \\ 0 \end{pmatrix} = 4 \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}$$

In short, we have $\mathbf{Bx} = 4\mathbf{x}$, and hence $\mathbf{x} = (2,3,0)$ is an eigenvector of $\mathbf{B}$ with eigenvalue $\lambda = 4$.

In an exercise below, we ask the reader to verify that $(1,0,0)$ and $(16,25,10)$ are also eigenvectors of $\mathbf{B}$, and to discover their eigenvalues. Most vectors in $\mathbf{R}^3$, however, are *not* eigenvectors of $\mathbf{B}$. For instance, if we multiply $\mathbf{B}$ by $\mathbf{x} = (1,2,1)$, we get $\mathbf{Bx} = (8,13,6)$ which is clearly not a scalar multiple of $(1,2,1)$. (Scalar multiples of $(1,2,1)$ always have the same first and third coordinates.) $\square$

Eigenvectors and eigenvalues play an truly fundamental role in linear algebra. We won't be prepared to grasp their full importance until Chapter 7, where our explorations all coalesce. We have introduced them here, however, so they can begin to take root in students' minds. We will revisit them off and on throughout the course so that when we reach Chapter 7, they will already be familiar.

– Practice –

1. Find the vector sum and difference $\mathbf{a} \pm \mathbf{b}$, if

   a) $\mathbf{a} = (2, -3, 1)$ and $\mathbf{b} = (0, 0, 0)$

   b) $\mathbf{a} = (1, -2, 0)$ and $\mathbf{b} = (0, 1, -2)$

   c) $\mathbf{a} = (1, 1, 1, 1)$ and $\mathbf{b} = (1, 1, -1, -1)$

2. Guided by the proof of associativity for Proposition 1.4, prove that Proposition's claim that vector addition is also *commutative.*

3. Compute

   a) $5\,(0, 1, 2, 1, 0)$

   b) $-1\,(2, -2)$

   c) $\frac{1}{10}\,(10, 25, 40)$

4. Rework the proof of Proposition 1.6 for the case of three vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ instead of just two vectors $\mathbf{x}$ and $\mathbf{y}$.

5. Compute these additional linear combinations of the vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ in Example 1.9.

   a) $\mathbf{a}_1 + 2\mathbf{a}_2 + \mathbf{a}_3$ b) $-2\mathbf{a}_1 + \mathbf{a}_2 - 2\mathbf{a}_3$ c) $x\,\mathbf{a}_1 + y\,\mathbf{a}_2 + z\,\mathbf{a}_3$

   (In part (c), treat x, y, and z as unevaluated scalars, and leave them that way in your answer.)

6. Find a linear combination of the vectors $\mathbf{v} = (1, 2, 3)$ and $\mathbf{w} = (-2, 3, -1)$ in $\mathbf{R}^3$ that adds up to $(8, -5, 9)$.

7. Find 3 *different* linear combinations of $\mathbf{a} = (1, -2)$, $\mathbf{b} = (2, 3)$, and $\mathbf{c} = (3, -1)$ that add up to $(0, 0)$ in $\mathbf{R}^2$.

8. Without setting the scalars x and y *both* equal to zero, find a linear combination $x(1,1)+y(1,-1)$ that adds up to $(0,0)\in\mathbf{R}^2$, or explain why this cannot be done.

9. Express each vector below as a linear combination of the standard basis vectors:

a) $(1,2,-1)$

b) $(1,-1,-1,1)$

c) $(0,3,0,-4,0)$

10. Write each linear combination below as a matrix/ vector product $\mathbf{Ax}$.

a) $$2\begin{pmatrix}-1\\1\end{pmatrix}-3\begin{pmatrix}1\\-1\end{pmatrix}$$

b) $$\tfrac{1}{2}\begin{pmatrix}1\\1\\0\end{pmatrix}+0.9\begin{pmatrix}0\\1\\1\end{pmatrix}+\pi\begin{pmatrix}1\\0\\1\end{pmatrix}$$

c) $$x_1\begin{pmatrix}1\\0\end{pmatrix}-x_2\begin{pmatrix}0\\1\end{pmatrix}+x_3\begin{pmatrix}1\\3\end{pmatrix}-x_4\begin{pmatrix}2\\4\end{pmatrix}$$

d) $$z\begin{pmatrix}1\\2\\3\\4\end{pmatrix}-w\begin{pmatrix}4\\3\\2\\1\end{pmatrix}$$

11. Expand each matrix/vector product below as a linear combination, then simplify as far as possible, writing each product as a single vector.

a)
$$\begin{bmatrix}1&-2&3\\-4&5&-6\end{bmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix}$$

b)
$$\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{pmatrix}a\\b\\c\end{pmatrix}$$

c)

$$\begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 2 \\ 0 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{pmatrix} x \\ -1 \\ \pi \end{pmatrix}$$

d)

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 1 \\ 3 & 4 & 5 & 1 & 2 \end{bmatrix} \begin{pmatrix} 1 \\ -2 \\ 4 \\ -2 \\ 1 \end{pmatrix}$$

12. Complete the proof of Proposition 1.25 by showing that for any $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_m)$ in $\mathbf{R}^m$, we have

$$\mathbf{Ax} + \mathbf{Ay} = x_1\mathbf{a}_1 + y_1\mathbf{a}_1 + x_2\mathbf{a}_2 + y_2\mathbf{a}_2 + \cdots + x_m\mathbf{a}_m + y_m\mathbf{a}_m \,,$$

the same result we got there for $\mathbf{A}\,(\mathbf{x} + \mathbf{y})$.

13. Compute each matrix/vector product below using dot products, as in Examples 1.33 and 1.34 above.

a)

$$\begin{bmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

b)

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

c)

$$\begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 2 \\ 0 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{pmatrix} x \\ -1 \\ \pi \end{pmatrix}$$

d)
$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 1 \\ 3 & 4 & 5 & 1 & 2 \end{bmatrix} \begin{pmatrix} 1 \\ -2 \\ 4 \\ -2 \\ 1 \end{pmatrix}$$

14. Show that $(1, 0, 0)$ and $(16, 25, 10)$ are both eigenvectors of the matrix $\mathbf{B}$ of Example 1.38. What are the corresponding eigenvalues? Is $(0, 2, 3)$ an eigenvector? How about $(0, -3, 2)$?

15. A 3-by-3 diagonal matrix is a matrix of the form
$$\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$$
where a, b, and c are any (fixed) scalars. Show that the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \in \mathbf{R}^3$ are always eigenvectors of a diagonal matrix. What are the corresponding eigenvalues? Do the analogous statements hold for 2×2 diagonal matrices? How about $n \times n$ diagonal matrices?

16. Consider the matrices
$$\mathbf{Y} = \begin{bmatrix} 2 & -1 & 0 & 3 \\ 0 & -2 & -3 & 1 \end{bmatrix} \qquad \mathbf{Z} = \begin{bmatrix} 1 & z & y \\ 0 & 1 & x \\ 0 & 0 & 1 \end{bmatrix}$$

a) How many rows and columns does each matrix have?

b) What are y_{21}, y_{14}, and y_{23}? Why is there no y_{32}?

c) What are z_{11}, z_{22}, and z_{33}? What is z_{13}? z_{31}?

17. Compute the dot product $\mathbf{x} \cdot \mathbf{y}$ for:

a) $\mathbf{x} = (1, 2, 3)$, $\mathbf{y} = (4, -5, 6)$

b) $\mathbf{x} = (-1, 1, -1, 1)$, $y = (2, 2, 2, 2)$

c) $\mathbf{x} = (\pi, \pi)$, $\mathbf{y} = (\frac{1}{4}, \frac{3}{4})$

18. Prove commutativity of the dot product (i.e., the first identity in Proposition 1.30 of the text).

19. Prove the third identity of Proposition 1.30 (the distributive law) in $\mathbf{R}^2$ and $\mathbf{R}^4$ directly:

a) In $\mathbf{R}^2$, consider arbitrary vectors $\mathbf{u} = (u_1, u_2)$, $\mathbf{v} = (v_1, v_2)$ and $\mathbf{w} = (w_1, w_2)$, and expand out both

$$\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) \quad \text{and} \quad \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$$

to show that they are equal.

b) In $\mathbf{R}^4$, carry out the same argument for vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{R}^4$. Do you see that it would work for any $\mathbf{R}^n$?

20. Suppose $\mathbf{x} \in \mathbf{R}^m$ is an eigenvector of an $m \times m$ matrix $\mathbf{A}$. Show that if $k \in \mathbf{R}$ is any scalar, then $k\mathbf{x}$ is *also* an eigenvector of $\mathbf{A}$, and has the same eigenvalue as $\mathbf{x}$.

Similarly, if both $\mathbf{v}$ and $\mathbf{w}$ are eigenvectors of $\mathbf{A}$, and both have the same eigenvalue λ, show that any linear combination $a\mathbf{v} + b\mathbf{w}$ is *also* an eigenvector of $\mathbf{A}$, again with the same eigenvalue λ.

— ⋆ —

2. Functions

Now that we're familiar with numeric vectors and matrices, we can consider *vector functions*—functions that take numeric vectors as inputs and produce them as outputs. The ultimate goal of this book is to give students a detailed understanding of *linear vector functions*, both algebraically, and geometrically. Here and in Section 3, we lay out the basic vocabulary for the kinds of questions one seeks to answer for any vector function, linear or not. Then, in Section 4, we introduce linearity, and with these building blocks all in place, we can at least *state* the main questions we'll be answering in later chapters.

2.1. Domain, image, and range. Roughly speaking, a **function** is an input-output rule. Here is is a more precise formal definition.

DEFINITION 2.2. A **function** is an input/output relation specified by three data:

i) A **domain** set X containing all allowed inputs,

ii) A **range** set Y containing all allowed outputs, and

iii) A **rule** f that assigns exactly one output $f(x)$ to every input x in the domain.

We typically signal all three of these at once with a simple diagram like this:

$$f\colon X \to Y$$

For instance, if we apply the rule $T(x,y) = x + y$ to any input pair $(x,y) \in \mathbf{R}^2$, we get a scalar output in $\mathbf{R}$, and we can summarize this situation by writing $T\colon \mathbf{R}^2 \to \mathbf{R}$. □

Technically, *function* and *mapping* are synonyms, but we will soon reserve the term *function* for the situation where (as with T above) the range is just $\mathbf{R}$. When the range is $\mathbf{R}^n$ for some $n > 1$, we typically prefer the term *mapping* or *transformation*.

2.3. Image. Suppose S is a subset of the domain X of a function. Notationally, we express this by writing $S \subset X$. This subset S may consist of one point, the entire domain X, or anything in between.

Whatever S is, if we apply f to every $x \in S$, the resulting outputs $f(\mathbf{x})$ form a subset of the range Y called the **image of S under f**, denoted $f(S)$. In particular,

- The image of a domain point $\mathbf{x} \in X$ is the single point $f(\mathbf{x})$ in the range.
- The image of the *entire domain* X, written $f(X)$, is called the **image** of the mapping f.

The image of any subset $S \subset X$ lies *in* the range, of course. But even when $S = X$ (the entire domain), *its image may not fill the entire range.*

Example 2.4. Consider the familiar *squaring* rule $f(x) = x^2$. If we take its domain to be $\mathbf{R}$ (the set of all real numbers), what is its image? What is its range?

Since x^2 cannot be negative, $f(x)$ has no negative outputs. On the other hand, every *non-negative* number $y \geq 0$ *is* an output, since $y = f(\sqrt{y})$. Note that $f(-\sqrt{y}) = y$ too, a fact showing that in general, *different inputs* may produce the *same output.*

In any case, we see that with $\mathbf{R}$ as domain, the squaring function has the half-line $[0,\infty)$ (all $0 \leq y < \infty$) as its image.

We may take the image—*or any larger set*—to serve as the **range** of f. One often takes the range to be all of $\mathbf{R}$, for instance. We would write

$$f\colon \mathbf{R} \to [0,\infty) \qquad \text{or} \qquad f\colon \mathbf{R} \to \mathbf{R}$$

to indicate that we have a rule named f with domain $\mathbf{R}$, and range either $[0, \infty)$ or $\mathbf{R}$, depending on our choice. Technically speaking, each choice yields a different function, since the domain is one of the three data that define the function.

Now consider the subset $S = [-1, 1]$ in the domain $\mathbf{R}$. What is the image of this subset? That is, what is $f(S)$? The answer is $f(S) = [0, 1]$, which the reader may verify as an exercise. □

We thus associate three basic sets with any function:

- **Domain:** The set of all allowed *inputs* to the function f.
- **Range:** The set of all allowed *outputs* to the function.
- **Image:** The collection of all *actual* outputs $f(\mathbf{x})$ as $\mathbf{x}$ runs over the entire domain. It is always *contained* in the range, and *may or may not* fill the entire range.

Remark 2.5. It may seem pointless—perhaps even perverse—to make the range larger than the image. Why should the range include points that never actually arise as outputs?

A simple example illustrates at least part of the reason. Indeed, suppose we have a function given by a somewhat complicated formula like

$$h(t) = 2.7\, t^6 - 1.3\, t^5 + \pi\, t^3 - \sin|t|$$

Determining the exact image of h would be difficult at best. But we can easily see that every output $h(x)$ will be a real number. So we can take $\mathbf{R}$ as the range, and then describe the situation correctly, albeit roughly, by writing

$$h\colon \mathbf{R} \to \mathbf{R}$$

We don't know the image of h, because we can't say exactly which numbers are actual outputs—but we *can* be sure that all outputs are real numbers. So we can't easily specify the image, but we *can* make a valid choice of range. □

2.6. Onto. As emphasized above, the image of a function is always a *subset* of the range, but it may not fill the entire range. When the image *does* equal the entire range, we say the function is *onto*:

Definition 2.7 (Onto). We call a function **onto** if every point in the range also lies in the image—that is, the image fills the entire range. Figures 1 and 2 illustrate the concept. □

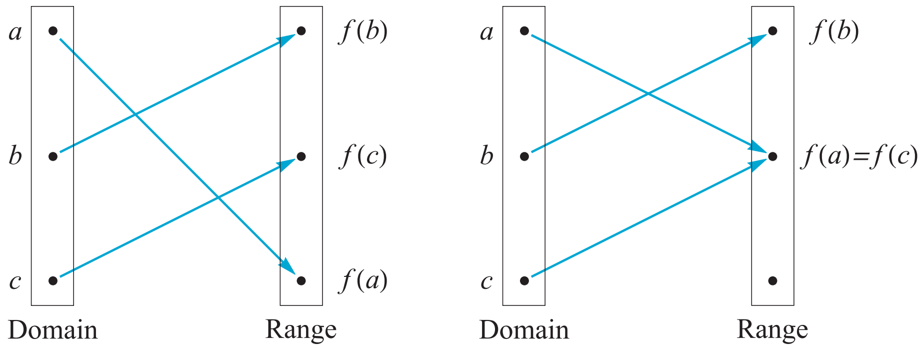


Figure 1. A function sends each point in the domain to one in the range. The function on the left is **onto**—each point in the range also lies in the image. The function on the right is *not* onto: the lowest point in the range does *not* lie in the image. That point has no *pre-image* (Definition 2.9).

Example 2.8. The squaring function with domain $\mathbf{R}$ is *onto* if we take its range to be *just* the interval $[0, \infty)$ of non-negative numbers. If we take its range to be all of $\mathbf{R}$, however, it is *not* onto, because $\mathbf{R}$ contains negative numbers, which do not lie in the squaring function's image. □

There is a useful counterpart to the term *image* which, among other things, makes it easier to discuss the difference between the image and range of a function.

Definition 2.9 (Pre-image). Suppose we have a function $f\colon X \to Y$, and a subset S of the range Y. Notionally, $S \subset Y$. The **pre-image** (or **inverse image**) of S consists of all points x in the domain X that f sends into S—all points whose images lie in S. We denote the pre-image of S by $f^{-1}(S)$ (pronounced "f *inverse of* S"). □

Remark 2.10. We are *not* claiming here that f has an inverse *function* f^{-1}. It may or may not—this is a topic we take up later. In general, "f^{-1}" by itself means *nothing* unless it occurs with a subset of the range, as in $f^{-1}(S)$ or $f^{-1}(x)$, in which case it means the pre-image of that subset, as defined above.

In certain cases, we *can* define an inverse mapping called f^{-1} (see Section 2.17). Then the pre-image $f^{-1}(S)$ equals the *image* of S under the *inverse* mapping f^{-1}, so our notation $f^{-1}(S)$ is consistent.
□

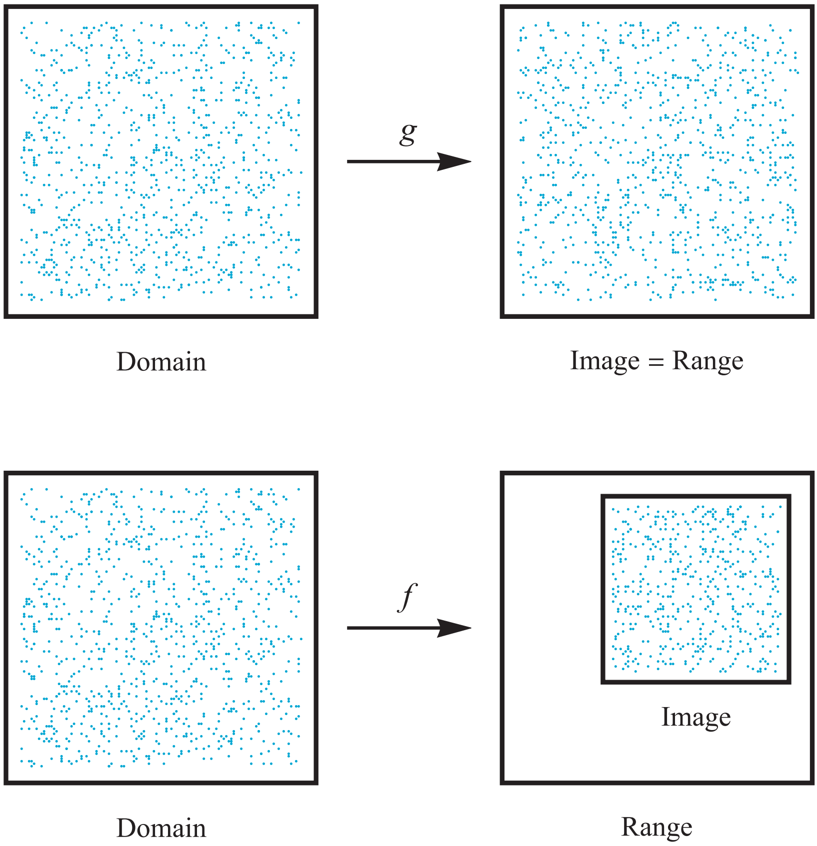


Figure 2. The image of g (above) fills its entire range, so g is *onto*. The image of f (below) does not fill the range, so f is *not* onto.

We can use the concept of *pre-image* to offer an alternate definition of the term *onto*:

OBSERVATION 2.11. *A function is* ***onto*** *exactly when every point in the range has* ***at least*** *one pre-image.*

Indeed, if every point in the range has a pre-image, then every point in the range is the image of some point in the domain. In this case, the image fills the entire range, and our function is indeed onto.

Pre-image is also a useful term because it gives an alternate name for something quite familiar, and very central to mathematics: *the solution of an equation.* The most basic question we ask about any equation $f(x) = y$ is whether we can solve it for x, given y. But this is the same as *finding a pre-image* $f^{-1}(y)$. Solving an equation and finding a pre-image are exactly the same thing.

EXAMPLE 2.12. Again consider the squaring function $f\colon \mathbf{R} \to \mathbf{R}$ given by $f(x) = x^2$. What is the pre-image of 4? Of 0? Of -4?

According to Definition 2.9 above, the pre-image of 4 consists of all x such that $f(x) = 4$. Here f is the squaring function, so that means we seek all x for which $x^2 = 4$. Clearly, this means

$$f^{-1}(4) = \{-2, 2\}$$

Similarly, we get the pre-image of $y = 0$ by solving $x^2 = 0$. Here there is only one solution—only one point in the pre-image:

$$f^{-1}(0) = \{0\}$$

The pre-image of -4, on the other hand, consists of all solutions to $x^2 = -4$. Since this equation has *no* solutions in the domain $\mathbf{R}$ we specified here, -4 has no pre-image; its pre-image is the *empty* set:

$$f^{-1}(-4) = \emptyset$$

Finally, we might ask for the pre-image of a set *larger* than just one point; say the pre-image of the interval $[0, 1]$. Since every number in $[0, 1]$ has a square root in $[0, 1]$, and also a square-root in $[-1, 0]$, it is easy to see that

$$f^{-1}([0, 1]) = [-1, 1]$$

Note that $f^{-1}[-1, 1]$ is *also* $[-1, 1]$. □

2.13. One-to-one. By definition, a *function* $f\colon X \to Y$ sends each point x in the domain X to a point $y = f(x)$ in the range Y. In that case, x belongs to the pre-image of y. But the pre-image of y may contain *other* inputs beside x. This happens, for instance, with the squaring function $f(x) = x^2$. The pre-image of any positive number contains *two* inputs. For example, $f^{-1}(4)$ contains both 2 and -2.

A nicer situation arises with the function $g\colon \mathbf{R} \to \mathbf{R}$ given by $g(x) = 2x + 3$. Here, the pre-image of an output y always contains exactly one point—no more, and no less. We know this because we can easily solve for $2x + 3 = y$ for x, and we always get exactly one solution, namely $x = (y - 3)/2$ (try it).

Since each point in the range of this function *has* a pre-image, it is *onto*. But we also know that pre-images never contain *more* than one point. This makes g *one-to-one*:

Definition 2.14. A function $f\colon X \to Y$ is **one-to-one** if each point in the range has at **most** one point in its pre-image. Examples show (see below) that a function may be one-to-one, onto, both, or neither. □

Example 2.15. The function depicted on the left in Figure 1 (above, not below) is both onto and one-to-one, because each point in the range has one—and only one—point in its pre-image. The function on the *right* in that figure, however, is *neither* one-to-one *nor* onto. It's not one-to-one because the pre-image of the middle point in the range contains *two* points (a and c). Neither is it onto, since the lowest point in the range has no pre-image at all.

The functions in Figure 3 below, on the other hand, each have one of the properties, but not the other. The function on the left is one-to-one but not onto. The one on the right is onto, but not one-to-one. (Make sure you see why.)

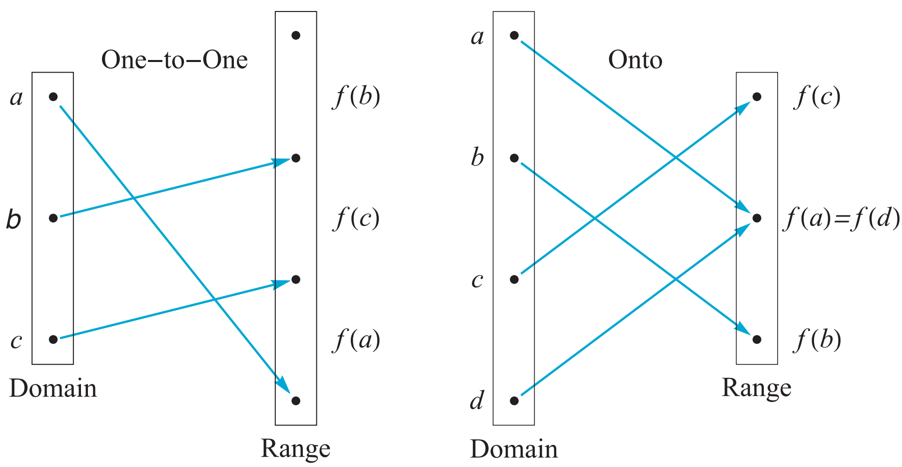


Figure 3. The function on the left is one-to-one, but not onto. The function on the right is onto, but not one-to-one.

□

Remark 2.16 (One-to-one vs. Onto). The definitions of *one-to-one* and *onto* compare and contrast very nicely if we summarize them like this:

- A function is **one-to-one** if every point in the range has **at most** one pre-image.
- A function is **onto** if every point in the range has **at least** one pre-image. □

2.17. Inverse functions. When a function $f\colon X \to Y$ is one-to-one, each y in the image has exactly one pre-image $x = f^{-1}(y)$, hence exactly one solution of $f(x) = y$. If f is also *onto*, then *every* y in the range has a unique pre-image x in this way. In this case, the assignment $y \to x$ defines a new function that "undoes" f:

DEFINITION 2.18 (Inverse mapping). If $f\colon X \to Y$ is *both* one-to-one *and* onto, the mapping that sends each $y \in Y$ to its unique pre-image $x = f^{-1}(y)$, is called the **inverse** of f. We denote it by f^{-1}. □

EXAMPLE 2.19. Consider the mapping $f\colon \mathbf{R} \to \mathbf{R}$ given by $f(x) = x + 1$. The solutions of $f(x) = y$ constitute the pre-image of y, and here that means solving $y = x + 1$. Doing so, we get $x = y - 1$. There's no restriction on y here—this gives a solution for *every* y. In fact, it gives *exactly one* solution for every y, so the function is both one-to-one and onto, and hence has an inverse. The inverse maps each y in the range to its unique pre-image in the domain, and our solution gives a formula for it: $f^{-1}(y) := y - 1$. □

EXAMPLE 2.20. The identity mapping $I(\mathbf{x}) = \mathbf{x}$ on $\mathbf{R}^m$ is its *own* inverse, since each $\mathbf{x}$ is obviously its own pre-image. Thus, $I^{-1} = I$. A slightly less trivial example is given by the *doubling* map on $\mathbf{R}^m$, given by $D(\mathbf{x}) = 2\mathbf{x}$. To solve $D(\mathbf{x}) = \mathbf{y}$, we write $2\mathbf{x} = \mathbf{y}$, which implies $\mathbf{x} = \mathbf{y}/2$. It follows that the inverse of D is the "halving" map: $D^{-1}(\mathbf{y}) = \mathbf{y}/2$. □

– Practice –

21. Define a function $f : X \to Y$ whose domain and range both contain just the first five letters of the alphabet: $X = Y = \{a, b, c, d, e\}$. Define the "rule" f for this function by setting

$$f(a) = b,\ f(b) = c,\ f(c) = a,\ f(d) = b,\ \text{and } f(e) = c$$

a) Find the images of these sets: $\{a, b, c\}$, $\{a, b, d\}$, $\{a, b, e\}$.

b) Find the pre-images of these sets: $\{a\}$, $\{a, b\}$, $\{a, b, c\}$, $\{c\}$, and $\{d, e\}$.

c) What is the image of f?

d) Is this function one-to-one? Is it onto? Explain.

e) Define a function $g : X \to Y$ (same X and Y as above) which is one-to-one, onto, *and* satisfies $g(a) = c$.

22. Suppose X and Y are the sets of young men and young women at a dance where the protocol is that each $x \in X$ chooses a dance partner $y \in Y$. Let $f \colon X \to Y$ be the "choosing" function, so that for each young man x, $y = f(x)$ is the partner he chooses.

a) What does it mean for f to be *onto*?

b) What does it mean for f to be *one-to-one*?

c) What is the image of f?

d) If $S \subset Y$ is a subset of the young woman, what does $f^{-1}(S)$ correspond to?

e) If $S \subset X$ is a subset of the young men, what is $f(S)$?

23. The following questions refer to Figure 4.

a) If we take $\mathbf{R}$ as the domain of the constant function $c(x) \equiv 1$, what is the image of c? (The triple equal sign emphasizes that $c(x) = 1$ for *all* inputs x in the domain.)

b) Assuming 1, 0 and -1 are in the range of c, what are the pre-images $c^{-1}(1)$, $c^{-1}(0)$, and $c^{-1}(-1)$?

c) Could the interval $[0, \infty)$ serve as the range of c? How about the interval $(-\infty, 0]$? How about the entire real line $\mathbf{R}$? Is c *onto* in any of these cases?

24. The following questions refer to Figure 4.

a) Suppose we take $\mathbf{R}$ as both the domain and range of the function $g(x) = 2x + 1$. What is the image of g? Is g onto?

b) What are the pre-images $g^{-1}(1)$, $g^{-1}(0)$, and $g^{-1}(-1)$?

c) If $\mathbf{R}$ is the domain of g, could the interval $[0, \infty)$ serve as its range? How about the interval $(-\infty, 0]$? Why or why not?

d) If we changed the domain of g from $\mathbf{R}$ to just the the interval $[-1, 1]$ (all $-1 \leq x \leq 1$), what would the image be?

25. Devise examples of:

   a) A function $N : \mathbf{Z} \to \mathbf{Z}$ that is one-to-one, but not onto.

   b) A function $F : \mathbf{Z} \to \mathbf{Z}$ that is onto, but not one-to-one.

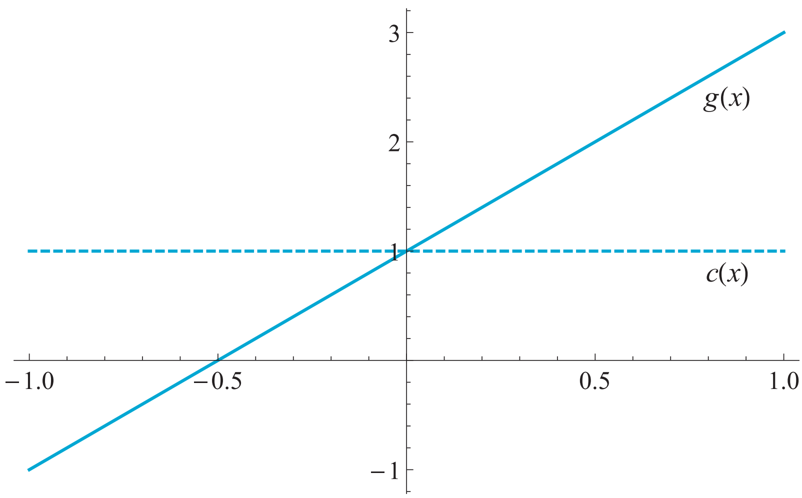


Figure 4. Graphs of the functions in Exercises 23 and 24.

26. Let $\mathbf{Z}$ be the set of all integers (positive and negative), and let $E \subset \mathbf{Z}$ be the subset of *even* integers. Show that the map $H \colon E \to \mathbf{Z}$ given by the rule $H(n) = n/2$ is *onto*.

Thus, even though E is a *proper* subset of $\mathbf{Z}$ (E is not *all* of $\mathbf{Z}$), we can map it *onto* $\mathbf{Z}$. This is only possible because $\mathbf{Z}$ contains infinitely many elements (integers). A proper subset of a *finite* set Y can never map *onto* Y. Can you give a reason for this?

27. Let A, B, C be scalars with $A \neq 0$ and consider the quadratic function

$$Q(x) = Ax^2 + Bx + C$$

   a) Give a precise description of the image of Q (in terms of the coefficients A, B, C). Thinking about the shape of the graph of Q should help you answer.

   b) What is the pre-image $f^{-1}(0)$? How about $f^{-1}(y)$ if $y \neq 0$? (Again, your answers will be functions of the coefficients.)