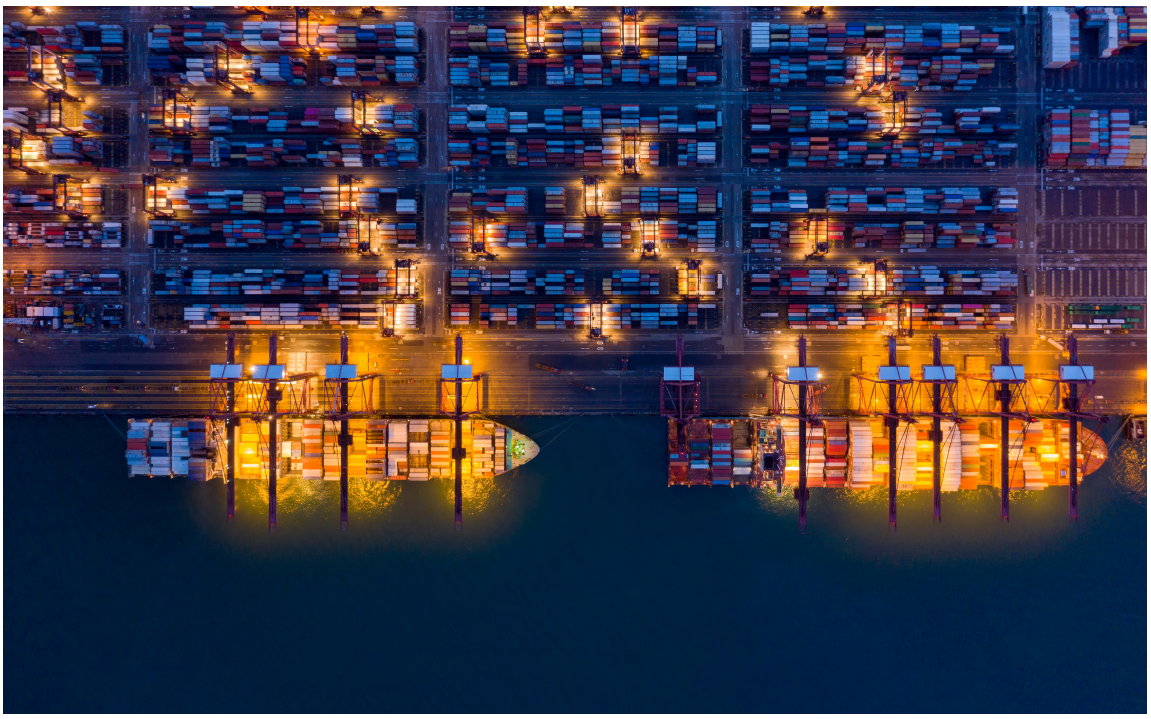
Introduction to Linear Algebra

July 26, 2026



First Edition

To everything I didn't get the chance to express,
no matter how trivial.

Contents

Introduction **5**

1 Foundations **6**
1.1 Introduction to Set Theory . . . 6
1.2 Relations . . . 12
1.3 Vector Algebra . . . 17
1.4 Vector Usage in Linear Algebra . . . 37

2 Linear Equations and Matrix Algebra **59**
2.1 Systems of Linear Equations . . . 59
2.2 Row Reduction . . . 71
2.3 Linear Combinations, Span, and Linear Independence . . . 83
2.4 The Matrix Equation $A\mathbf{x} = \mathbf{b}$ . . . 105
2.5 Classifying Solutions . . . 115
2.6 Applications of Linear Systems . . . 124
2.7 Linear Transformations Through Matrices . . . 140
2.8 Linear Models . . . 164
2.9 Matrix Structure and Operations . . . 177
2.10 Matrix Inverses . . . 192
2.11 Block Matrices . . . 208
2.12 *LU* Factorization . . . 216
2.13 Applications of Matrix Algebra . . . 229

3 Determinants **254**
3.1 Intro to Determinants . . . 254
3.2 Properties of the Determinant . . . 264
3.3 Cramer's Rule and Applications of Determinants . . . 275

4 The Four Fundamental Subspaces **287**
4.1 Vector Spaces and Subspaces . . . 287
4.2 Null Spaces and Column Spaces . . . 298
4.3 Linear Transformations . . . 312
4.4 Linearly Independent Sets and the Basis . . . 316
4.5 Coordinates, Dimension, and Rank . . . 334
4.6 Change of Basis . . . 366

5 Eigenvalues and Eigenvectors **378**
5.1 Foundations of Eigenvalues and Eigenvectors . . . 378
5.2 The Characteristic Equation . . . 394
5.3 Diagonalization . . . 416
5.4 Complex Eigenvalues . . . 441
5.5 Discrete Dynamical Systems . . . 454
5.6 Applications to Differential Equations . . . 472

6 Orthogonality **481**
6.1 Inner Product; Orthogonal Complement, Sets, and Projections 481
6.2 The Gram-Schmidt Process 512
6.3 Least Squares 531
6.4 Applications of Least Squares 555
6.5 Inner Product Spaces and Hilbert Spaces 586

7 Symmetric Matrices and the Singular Value Decomposition **614**
7.1 Diagonalizing Symmetric Matrices 614
7.2 Quadratic and Hermitian Forms 630
7.3 Constrained Optimization 658
7.4 The Singular Value Decomposition 677

8 Chosen Topics **710**
8.1 Markov Chains 710
8.2 Robotics 717
8.3 Chemical Reactors 730
8.4 Quantum Mechanics 736
8.5 PageRank 743

Bibliography **754**

Introduction

The cover features an image of the Port of Hong Kong. Notice how the cargo in ports and on container ships is arranged in rectangular arrays. Linear algebra is used heavily in industrial processes like these, whether it's to control automated cranes or to optimize the times for vessels to arrive and depart. What other things in the real world can be arranged into arrays? Image credit: tampatra, Adobe Stock

Welcome to the first edition of my guide on linear algebra, which is arguably the most powerful branch of modern day mathematics. It is the heart of many applications in engineering, physics, chemistry, biology, computer science, and economics. If you are really good at linear algebra, you will make a lot of money. This is intended for students of the sciences and engineering.

Make sure you fully understand the **Examples**. Then, work through all of the **Problems** afterwards to check your understanding. If you get stuck, glance at the solutions and then reiterate until you can work the problem entirely on your own from the beginning.

As always, I want to reiterate that learning mathematics is a process and not a destination. Use this guide and combine it with other resources over the course of many years. Keep revisiting problems and digest content in a way that makes sense to you. Think about not only what you're learning, but also *how* you are learning. If something seems like a waste of time, then don't spend as much time on it. I recommend storing your notes and practice problems such that you can look back on them in the future. There's basically no way to master linear algebra in a single semester and you don't want to have to relearn everything all over from scratch again! As a matter of fact, basically no one "understands" topics this massive on their first try, so you'll likely go through some of the concepts here in your other courses multiple times until you get a strong grasp of them.

This guide is primarily based on *Linear Algebra and Its Applications* by Lay et al with additional material drawn from the sources in **Bibliography**. Any images not attributed to someone else are my original work, but you are free to use them as long as you give me credit.

This document was created entirely for educational purposes, and I make no money from it. It is not meant to replace the original works, which I encourage you to seek out on your own for more explanations and practice problems.

1 Foundations

Before we jump into linear algebra, there are a few things I want to go over that will help you understand not only linear algebra better but will help you with your other classes too. I am aware that the content here is not explicitly taught in your standard linear algebra course. If you are confident with the material, you can move on to the next chapter.

1.1 Introduction to Set Theory

You may have already seen this notation before and used it intuitively, but it's really important to have the specific language for linear algebra.

Set

A **set** is a collection of different objects known as **elements**.

We have two ways to define a set. The first is called *enumeration* or *exhaustion*. This method involves listing each element in the set.

There are four genera of coronavirus called Alphacoronavirus, Betacoronavirus, Deltacoronavirus, and Gammacoronavirus. Let's take the Greek prefixes and form them into a set. We have $C = \{\alpha, \beta, \delta, \gamma\}$. Notice that the order here does not matter. $C = \{\alpha, \beta, \delta, \gamma\}$ is the same thing as $C = \{\gamma, \delta, \beta, \alpha\}$. If we want to characterize α and μ, we would say that α is an element of C

$$\alpha \in C$$

and that μ is not an element of C

$$\mu \notin C.$$

The second method is called *set-builder notation* and is the standard. All that matters in set-builder notation are the elements. And these elements are subject to propositions. We write it as follows:

$$A = \{x \mid P(x)\}.$$

This would read out as "Set A equals the set of all elements x such that proposition $P(x)$ holds true." You can also replace the vertical bar with a colon if you prefer that. While the colon is easier to distinguish than the vertical bar, it is already reserved for function mapping.

Here is an example. Say we have $S = \{x \in \mathbb{Z} \mid 1 \leq x \leq 100\}$. The first condition is that x must belong to the set of integers. The second condition is that x must be greater than or equal to 1 but less than or equal to 100. All of the following statements are true:

$$\begin{aligned} -1 &\notin S \\ 2 &\in S \\ 100 &\in S \end{aligned}$$

Let's use this to define the number systems. First, we have the set of all natural numbers, or positive whole numbers:

$$\mathbb{N} = \{1, 2, 3, 4, 5, \dots\}$$

We have integers, an extension of the natural numbers that includes negative whole numbers:

$$\mathbb{Z} = \{0, -1, 1, -2, 2, \dots\}$$

We have rational numbers, an extension of the rational numbers to include fractions:

$$\mathbb{Q} = \left\{ \frac{p}{q} \;\middle|\; p \in \mathbb{Z},\ q \in \mathbb{Z},\ q \neq 0 \right\}$$

We have real numbers, an extension of the rational numbers to include irrational numbers:

$$\mathbb{R} = \{x \mid x \text{ is a real number}\}$$

And lastly we have complex numbers, an extension of the real numbers to include imaginary numbers:

$$\mathbb{C} = \{a + bi \mid a \in \mathbb{R}, b \in \mathbb{R}, i^2 = -1\}$$

Let's now move on to **subsets**. If we want to write A is a subset of B, we would write

$$A \subseteq B.$$

This holds true if and only if every element of A is also an element of B. If there is an $x \in A$ with $x \notin B$, we would say

$$A \nsubseteq B.$$

We also have C is a strict subset of B, written as

$$C \subset B.$$

A strict subset must omit at least one element that is in the original set.

For the common number systems, we have

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

Let's come up with some example sets:

$$X = \{1, 2, 3\}$$
$$Y = \{1, 2, 3, 4, 5\}$$
$$Z = \{1, 2, 4, 5\}$$

Here, $X \subseteq Y$ is true because every element of X is in Y. $X \nsubseteq Z$ is true because $3 \in X$, but $3 \notin Z$. And $Z \subset Y$ is true because $3 \in Y$ but $3 \notin Z$.

To prove that $A \subseteq B$, you must show that every element of A is also an element of B. This would be written as

$$A \subseteq B \iff \forall x(x \in A \Rightarrow x \in B).$$

This says that "Set A is a subset of set B if and only if for every element in x, if x is in A, then x is also in B."

And to prove that $A = B$, we need to prove that $A \subseteq B$ and $B \subseteq A$.

Example 1.1.1

Let $A = \left\{\frac{\pi}{4}, \frac{9\pi}{4}\right\}$ and $B = \left\{\theta \in [0, 6\pi] \mid 4\cos(\theta - \frac{\pi}{4}) - 4 = 0\right\}$. Show that $A \subseteq B$.

Solution

Let $\theta \in A$. For $\theta = \frac{\pi}{4}$,

$$4\cos\left(\frac{\pi}{4} - \frac{\pi}{4}\right) - 4 = 4\cos(0) - 4 = 4(1) - 4 = 0.$$

Thus $\frac{\pi}{4} \in B$.

For $\theta = \frac{9\pi}{4}$,

$$4\cos\left(\frac{9\pi}{4} - \frac{\pi}{4}\right) - 4 = 4\cos(2\pi) - 4 = 4(1) - 4 = 0.$$

Thus $\frac{9\pi}{4} \in B$.

We have proved that if $\theta \in A$, then $\theta \in B$. Therefore by definition, $A \subseteq B$.

Example 1.1.2

Let $A = \left\{x \in \mathbb{R} \mid \frac{d}{dx}\left(x^3 - 12x\right) = 0\right\}$ and $B = \{-2, 2\}$. Show that $A = B$.

Solution

Let $x \in A$. First, compute the derivative:

$$\frac{d}{dx}(x^3 - 12x) = 3x^2 - 12$$

Now solve to get

$$3x^2 - 12 = 0$$
$$x^2 = 4 \Rightarrow x = \pm 2.$$

Thus, $x \in B$ and we can conclude that $A \subseteq B$.

Now we need to show that $B \subseteq A$. To prove this, we will let $x \in B$ the following should hold true:

$$\frac{d}{dx}\left[x^3 - 12x\right]\Big|_{x=\pm 2} \stackrel{?}{=} 0$$

Let's see if it does:

$$\frac{d}{dx}\left[x^3 - 12x\right]\Big|_{x=-2} = 3(-2)^2 - 12 = 0$$

$$\frac{d}{dx}\left[x^3 - 12x\right]\Big|_{x=2} = 3(2)^2 - 12 = 0$$

It does, and therefore $B \subseteq A$.

The **empty set**, denoted by $\emptyset$, is the set that contains no elements. It can be written in set-builder notation as

$$\emptyset = \{x : x \neq x\}.$$

We can safely assume that $\emptyset$ exists. Since there is no x that satisfies $x \neq x$, the statement $x \in \emptyset$ is false for any x.

Classically, we have the following:

Let A, B, and C be any sets. Then

1. $\emptyset \subseteq A$ (The empty set is a subset of every set)
2. $A \subseteq A$ (Every set is a subset of itself)
3. If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$ (Subset relation is transitive)

Section 1.1 Problems

1. Let A and B be sets. What does it mean if we say that A is a subset of B?

   (a) A is an element of B.

   (b) Some element x in A is also an element of B.

   (c) Every element y in B is also an element of A.

   (d) Every element x in A is contained in some element y of B.

   (e) Every element x in A is also an element of B.

2. For sets A and B, the statement "if $x \in A$, then $x \in B$" is written using which of the following?

   (a) $A \neq B$

   (b) $A = B$

   (c) $A \leq B$

   (d) $B \subseteq A$

   (e) $A \subseteq B$

3. Let $A = \{n \in \mathbb{N} : n - 4 < 10\}$ and $B = \{m \in \mathbb{N} : m^2 \leq 169\}$. Prove that $A = B$.

Solution

1. (e). This is the definition of a set.

2. (e). This is the definition of a set.

3. We must show that $A \subseteq B$ and $B \subseteq A$.

Let $x \in A$. This means $x \in \mathbb{N}$ and $x - 4 < 10$ which simplifies to $x < 14$. Since natural numbers start at 1, we conclude that $x \in \{1, 2, 3, \ldots, 13\}$. Now consider x^2: squaring any of these values still gives us a natural number, and the largest possible square is $13^2 = 169$. It is true that if $x \in A$, then $x \in B$, and therefore $A \subseteq B$.

Moving on, we let $x \in B$. This means $x \in \mathbb{N}$ and $x^2 \leq 169$. Since x is a natural number and $x^2 \leq 169$, we know $x \leq 13$, because $13^2 = 169$. That tells us $x < 14$, and since $x \in \mathbb{N}$, it also follows that $x - 4 < 10$. It is true that if $x \in B$, then $x \in A$, and therefore $B \subseteq A$.

1.2 Relations

Let's say we have the function $f(x) = A_{m\times n}\mathbf{x}_{n\times 1}$ where m denotes the number of columns and n denotes the number of rows. And suppose the function has the following properties:

$$f : \underbrace{\mathbb{R}^n}_{\text{domain}} \longrightarrow \underbrace{\mathbb{R}^m}_{\text{codomain}}$$

The domain of f is written as $\text{dom}(f) = \mathbb{R}^n$ and the codomain of f is written as $\text{codom}(f) = \mathbb{R}^m$. Generally, $\text{codom}(f)$ is not equal to the range of f which is written as $\text{rng}(f)$; there are some special matrices where the codomain and range are equal, which we will explore later.

We will now define something called the cross product of sets. Let A and B be sets. The cross product of A and B is given by

$$A \times B = \{(a, b) \mid a \in A, b \in B\}.$$

The cross product equals the set of all ordered pairs such that the first coordinate is some $a \in B$ and the second coordinate is some $b \in B$.

Example 1.2.1

Model a basic system that could be used to track how a US senator's personal interests in a company influences their vote regarding a bill that would benefit the company.

Solution

Let $A = \{-1, 0, 1\}$ where -1 means that the senator is against the interests of the company, 0 means that they are neutral, and 1 means they have stakes in the company. Let $B = \{\text{Nay}, \text{Abstain}, \text{Yea}\}$ represent the senator's vote. Then the cross product is given by

$$\begin{aligned} A \times B = \{&(-1, \text{Nay}),\\ &(-1, \text{Abstain}),\\ &(-1, \text{Yea}),\\ &(0, \text{Nay}),\\ &(0, \text{Abstain}),\\ &(0, \text{Yea}),\\ &(1, \text{Nay}),\\ &(1, \text{Abstain}),\\ &(1, \text{Yea})\}. \end{aligned}$$

This represents a list of all possible ways a senator's personal interests in a company correlates with their voting record.

Example 1.2.2

Develop a system for every way we could assign every uppercase English letter a number from 1 to 26 one time per number. How many ordered pairs would this system contain?

Solution

Let $U = \{A, B, C, .., Z\}$ represent the set of uppercase letters and let $N = \{n \in \mathbb{N} \mid 1 \leq n \leq 26\}$ represent the set of natural numbers from 1 to 26. If we were to list the cross product, we would have $U \times N = \{(A, 1), (A, 2), (A, 26), \ldots, (B, 1), (B, 2), (B, 26), \ldots, (Z, 1), (Z, 2), (Z, 26)\}$. This product includes $26^2 = 676$ ordered pairs.

We can also generalize the cross product of two sets to three or more sets. The cross product of three sets A, B, and C is defined as

$$A \times B \times C = \{(a, b, c) \mid a \in A, b \in B, c \in C\}.$$

Each element of the cross product here is an ordered triplet.

And for $n \in N$, we define the cross product of n sets $A_1, A_2, \ldots, A_n$ as

$$A_1 \times A_2 \times \ldots A_n = \{(a_1, a_2, \ldots, a_n) \mid a_i \in A_i \text{ for all } i \in \{1, 2, \ldots, n\}\}.$$

Each element of the cross product here is an n-tuple where two ordered n-tuples $a_1, a_2, \ldots, a_n$ and $x_1, x_2, \ldots, x_n$ are equal if and only if $a_i = x_i$ for every $i = 1, 2, \ldots, n$.

Let's now formally define a relation using set notation:

Relation

Let A and B be sets. The set R is a **relation** from A to B if and only if

$$R \subseteq A \times B.$$

If $(a, b) \in R$, we say $a \; R \; b$.
If $(a, b) \notin R$, we say $a \not R \; b$.

Example 1.2.3

Suppose you are designing a drip irrigation control system to water crops in the daylight. The morning is the optimal time because the crops will be able to fully utilize the water before it gets dark. You have decided that the water will drip slowly from 8:00 AM to 9:00 AM. If the water Model the system using a mathematical relation.

Solution

Let $A = \{8{:}00 \text{ AM}, 9{:}00 \text{ AM}\}$ represent the start and stop times. Let $B = \{\text{Water On}, \text{Water Off}\}$ control whether or not the system is dripping. Then,

$$A \times B = \{(8{:}00 \text{ AM}, \text{Water On}), (8{:}00 \text{ AM}, \text{Water Off}), (9{:}00 \text{ AM}, \text{Water On}), (9{:}00 \text{ AM}, \text{Water Off})\}.$$

We want $R = \{(8{:}00 \text{ AM}, \text{Water On}), (9{:}00 \text{ AM}, \text{Water Off})\}$. Notice how $R \subset A \times B$.

One potentially helpful way to think about this is through the way you can type numbers and letters on your phone's keypad.

Let's now use set notation to formally define domain and range:

Domain and Range

Let A and B be sets and let R be a relation from A to B. The **domain** of the relation R is the set

$$\text{dom}(R) = \{x \in A \mid \text{there is a } y \in B \text{ such that } (x, y) \in \mathbb{R}\}.$$

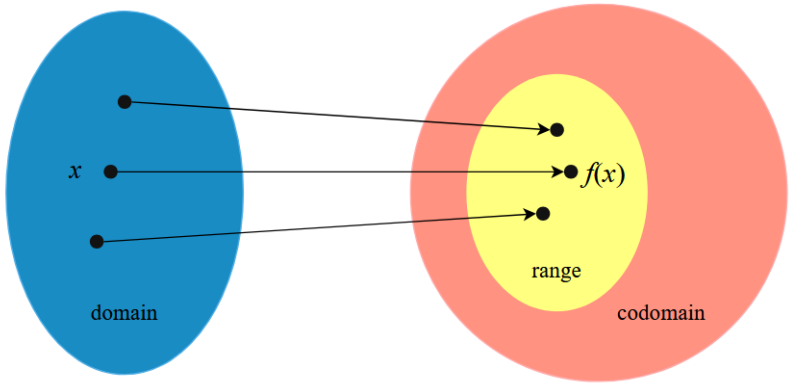
The **codomain** of the relation R is the set

$$\text{codom}(R) = B.$$

The **range** of the relation R is the set

$$\text{rng}(R) = \{y \in B \mid \text{there is an } x \in A \text{ such that } (x, y) \in \mathbb{R}\}.$$

In other words, the domain of R is the set of all first elements from each ordered pair $(a, b) \in R$. The codomain of R is the set from which the second elements in the ordered pairs can come from. And lastly, the range of R is the set of all second elements that we will actually choose for an ordered pair $(a, b) \in R$.



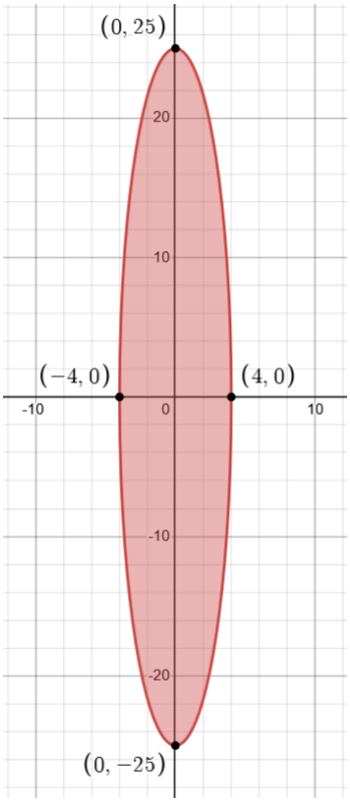
Example 1.2.4

Find the domain, codomain, and range of the ellipse given by the relation

$$E = \left\{ (x, y) \in \mathbb{R} \times \mathbb{R} : \frac{x^2}{16} + \frac{y^2}{625} \leq 1 \right\}.$$

Solution

This is an ellipse centered at the origin:



We have $\text{dom}(D) = [-4, 4] \subset \mathbb{R}$, $\text{codom}(D) = \mathbb{R}$, and $\text{rng}(D) = [-25, 25] \subset \mathbb{R}$.

Let's now formally define what a function is using set notation:

Function

A function from A to B is a relation f from A to B such that both of the following hold:

1. $\text{dom}(f) = A$
2. If $(x, y) \in f$ and $(x, z) \in f$, then $y = z$.

If f is a function from A to B, we write

$$f : A \to B.$$

If $B = A$, we say that f is a function on A.

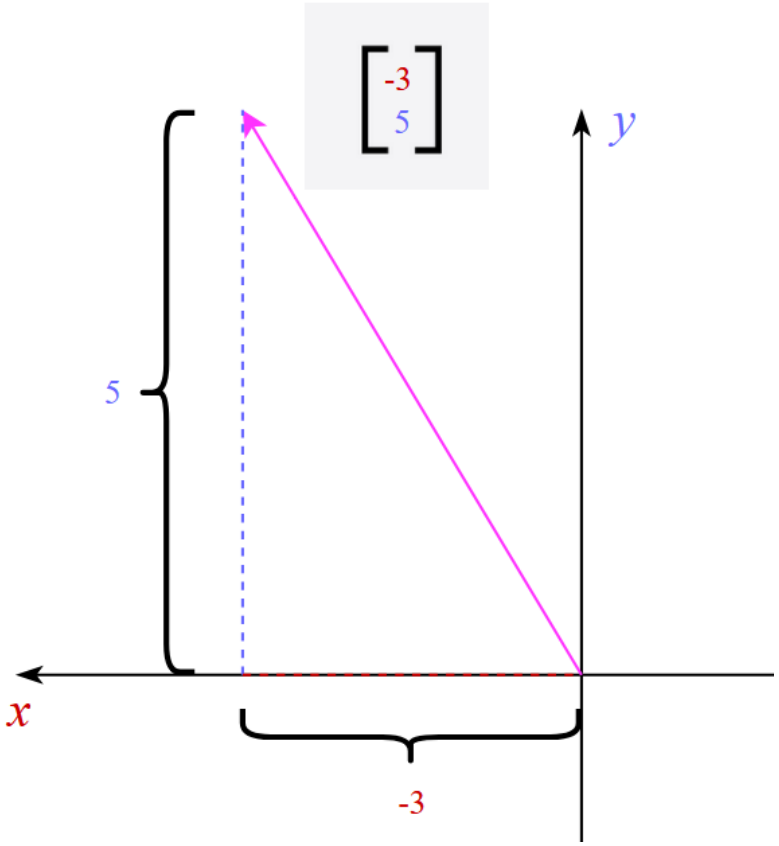
From this, we see that if $f : A \to B$, then $f \subseteq A \times B$. In other words, all functions are relations, but not all relations are functions. By the former, any function is thus a subset of a cross product of sets.

1.3 Vector Algebra

The purpose of this section is to help you review vectors, get you some geometric intuition, and prepare you for some applications that will arise in the rest of this guide. Treat it as a warm-up. I begin with two visualizations of vectors and then get into review problems.

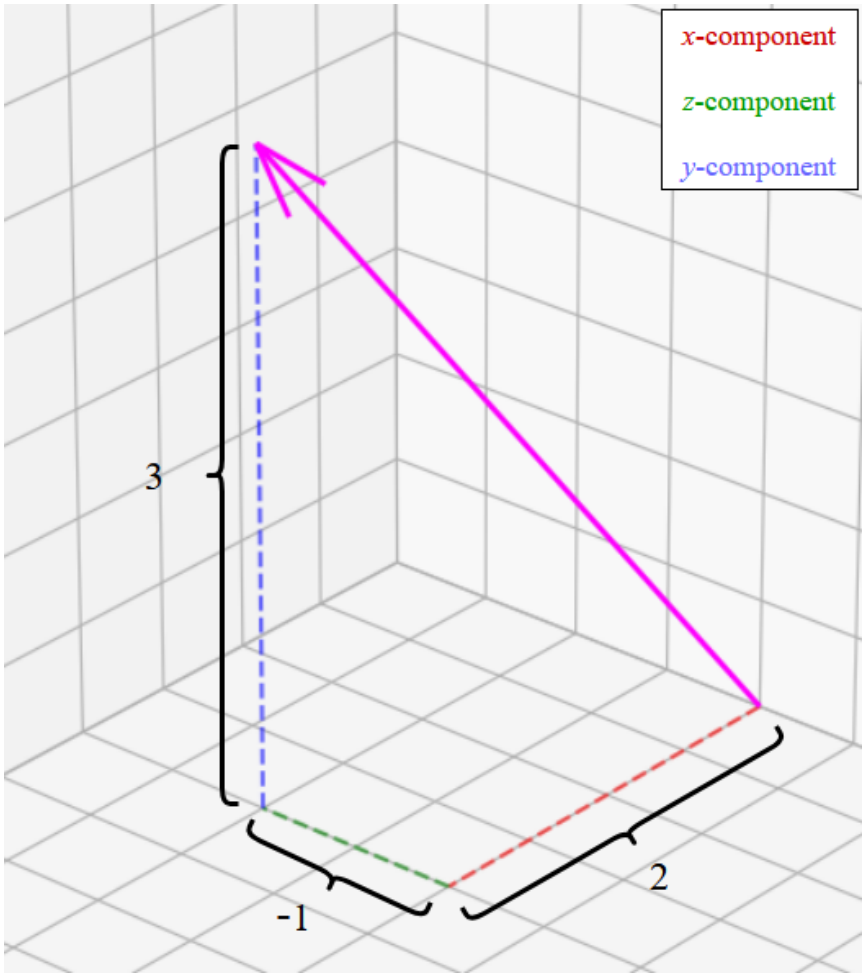
In linear algebra, vectors will generally be rooted at the origin.

Here, we have the vector $\mathbf{v} = \begin{bmatrix} -3 \\ 5 \end{bmatrix} = -3\mathbf{i} + 5\mathbf{j}$ in two-dimensional space.



The starting point, or tail, is $(0, 0)$ and the ending point, or head, is $(-3, 5)$.

Here we have the vector $\mathbf{v} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} = 2\mathbf{i} - \mathbf{j} + 3\mathbf{k}$ in three-dimensional space.



The tail is $(0, 0, 0)$ and the head is $(2, -1, 3)$.

Problem 1.3.1

A fin whale is swimming 35 km/hr at 50° to the east from the southward direction. Given that $\|\mathbf{a}\| = \|\mathbf{b}\| = \|\mathbf{c}\| = \|\mathbf{d}\| = 35$ km/hr, which vectors can represent the whale's velocity?

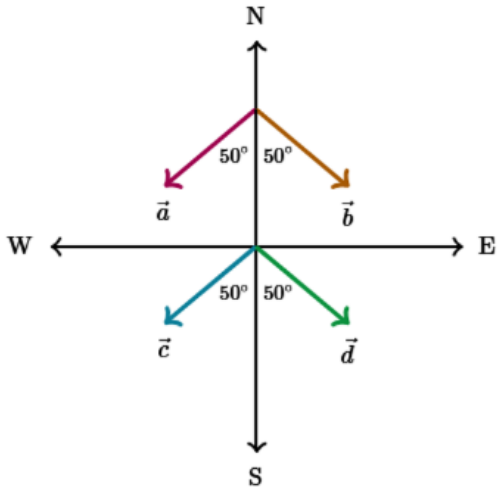


Image credit: Khan Academy

Solution

Vectors $\|\mathbf{b}\|$ and $\|\mathbf{d}\|$ can represent the whale's velocity. Both have negative vertical components, point 50° east, and have a magnitude of 35 km/hr.

Problem 1.3.2

A boat bumps into a rock with the force of 25 N at 35° to the east from the northward direction relative to the rock. Given that $\|\mathbf{a}\| = \|\mathbf{c}\| = 25$ N and $\|\mathbf{b}\| = \|\mathbf{d}\| = 35$ N, which vectors can represent the force the boat exerts on the rock?

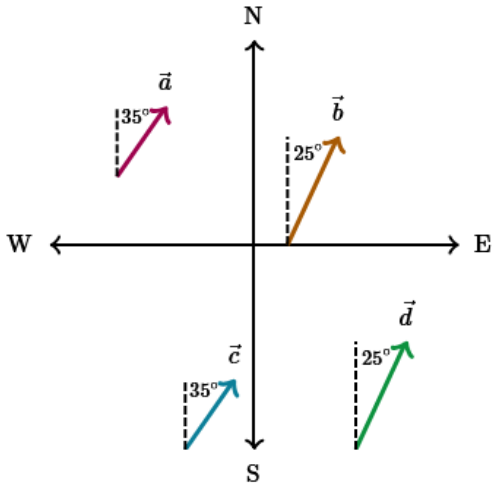


Image credit: Khan Academy

Solution

Vectors $\|\mathbf{a}\|$ and $\|\mathbf{c}\|$ can represent the boat's applied force. Both have positive vertical components, point 35° east, and have a magnitude of 25 N.

Problem 1.3.3

Find the components of vector $\overrightarrow{AB}$.

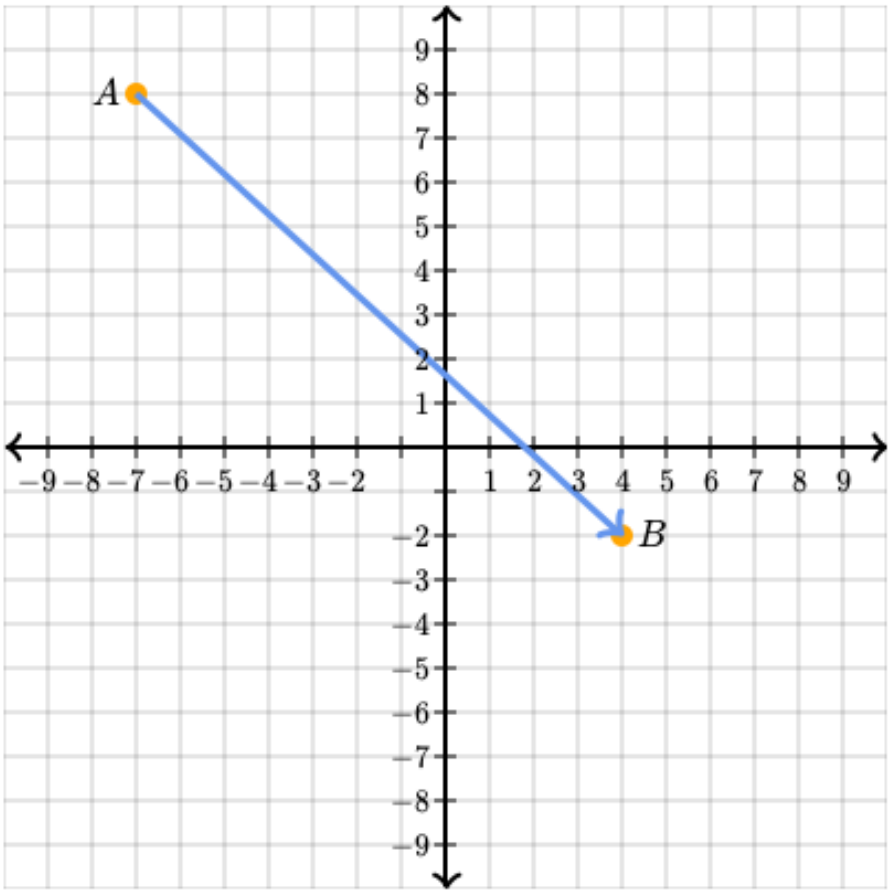


Image credit: Khan Academy

Solution

$$\overrightarrow{AB} = \begin{bmatrix} 11 \\ -10 \end{bmatrix}$$

Problem 1.3.4

(a) Find the magnitude of **u**.

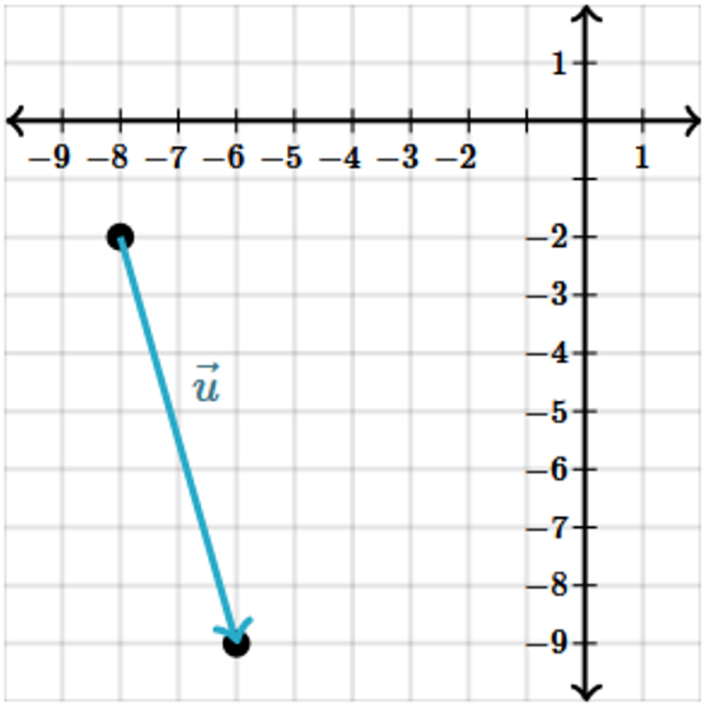


Image credit: Khan Academy

(b) **t** has a tail $(3, 3)$ and a head $(7, 2)$. Find the magnitude of **t**.

Solution

(a) $\|\mathbf{u}\| = \sqrt{(-7)^2 + (2)^2} = \sqrt{53}$

(b) $\|\mathbf{t}\| = \sqrt{(7-3)^2 + (2-3)^2} = \sqrt{17}$

Problem 1.3.5

$\mathbf{v}$ has a tail at the origin and a head at B as shown below. Let B and E be equidistant from the origin.

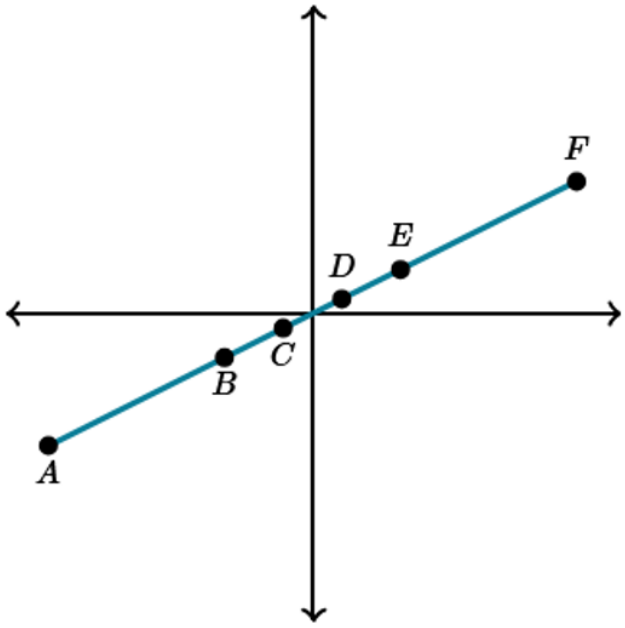


Image credit: Khan Academy

(a) Assuming that $\frac{1}{5}\mathbf{v}$ starts at the origin, which option could be its endpoint?
(b) Assuming that $-\frac{1}{4}\mathbf{v}$ starts at the origin, which option could be its endpoint?
(c) Assuming that $3\mathbf{v}$ starts at the origin, which option could be its endpoint?

Solution

(a) Multiplying $\mathbf{v}$ by positive scalar $\frac{1}{4}$ changes its magnitude but not the direction. The vector becomes shorter and thus its only possible endpoint could be C.

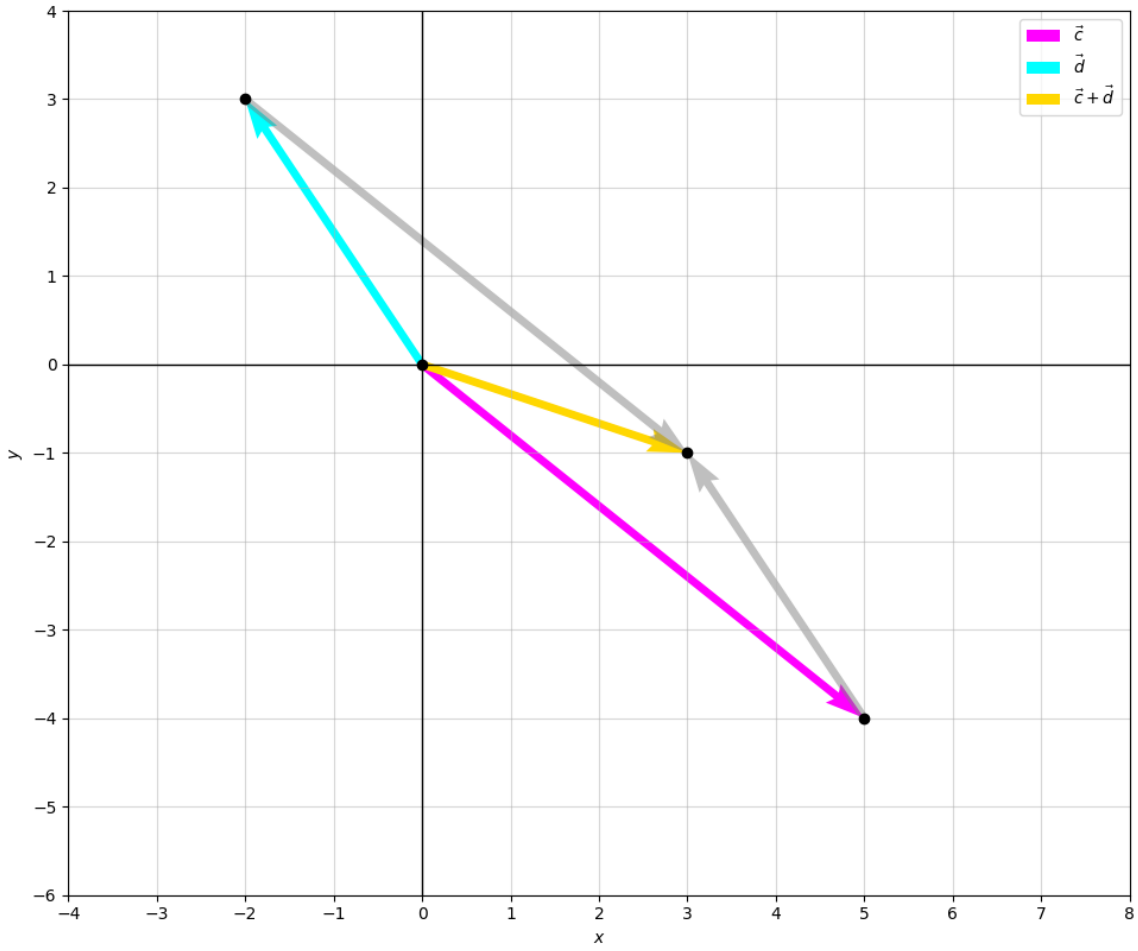
(b) Multiplying $\mathbf{v}$ by negative scalar $-\frac{1}{4}$ changes its magnitude and also flips its direction. The vector becomes shorter and thus the only possible endpoint could be D.

(c) The magnitude increases while maintaining the same direction. Thus the only possible endpoint could be A.

Problem 1.3.6

Let $\mathbf{c} = (5, -4)$ and $\mathbf{d} = (-2, 3)$. Graph $\mathbf{c} + \mathbf{d}$ using the parallelogram rule starting each vector at the origin. Then use the graph to find $\mathbf{c} + \mathbf{d}$.

Solution

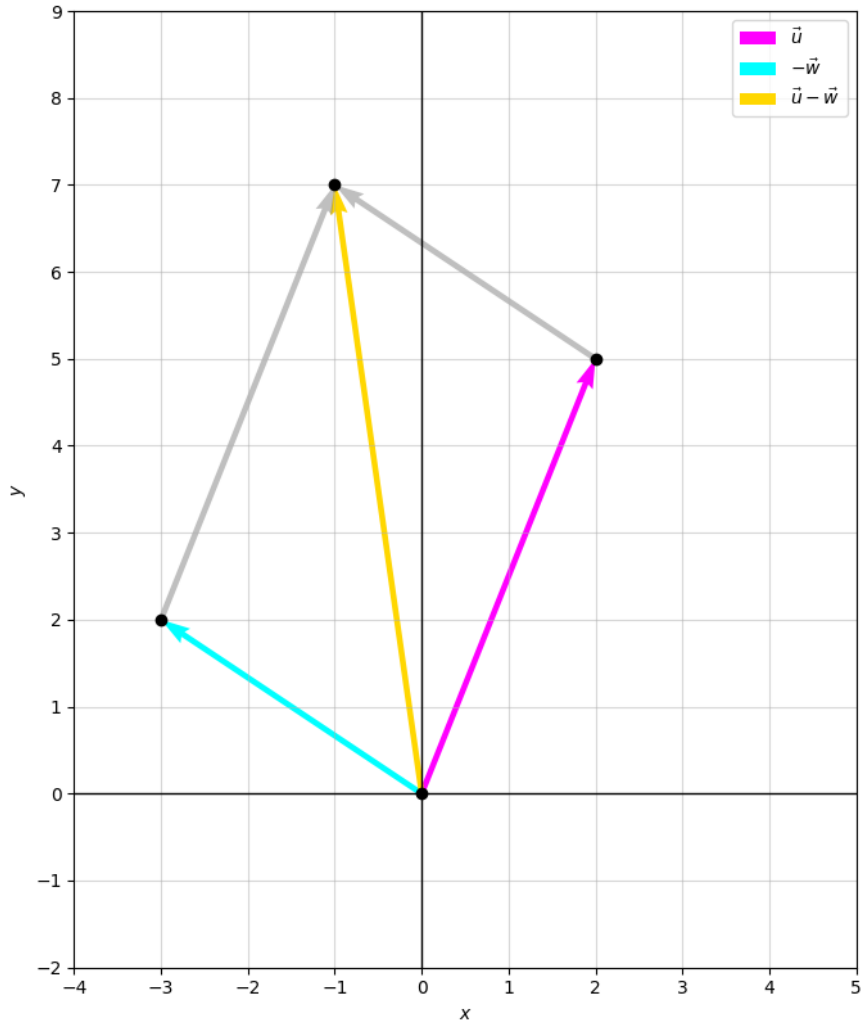


$\mathbf{c} + \mathbf{d} = (3, -1)$

Problem 1.3.7

Let $\mathbf{u} = (2, 5)$ and $\mathbf{w} = (3, -2)$. Graph $\mathbf{u} - \mathbf{w}$ using the parallelogram rule starting each vector at the origin. Then use the graph to find $\mathbf{u} - \mathbf{w}$.

Solution



$\mathbf{u} - \mathbf{w} = (-1, 7)$

Problem 1.3.8

(a) Let $\mathbf{v} = (-9, 3)$. Find the direction angle θ of $\mathbf{v}$.
(b) Let $\mathbf{u} = (-7, -10)$. Find the direction angle θ of $\mathbf{u}$.

Solution

(a) We first compute $\theta = \tan^{-1}\left(\frac{y}{x}\right) = \tan^{-1}\left(\frac{3}{-9}\right) = -18.4°$. Since $\mathbf{v}$ is in the second quadrant and our current θ is in the fourth quadrant, we have to add $180°$. Thus $-18.4° + 180° = 161.6°$.

(b) We first compute $\theta = \tan^{-1}\left(\frac{y}{x}\right) = \tan^{-1}\left(\frac{-10}{-7}\right) = 55.0°$. Since $\mathbf{u}$ is in the third quadrant and our current θ is in the first quadrant, we have to add $180°$. Thus $55.0° + 180° = 235.0°$.

Problem 1.3.9

Find the component form of $\mathbf{u}$ which is shown below:

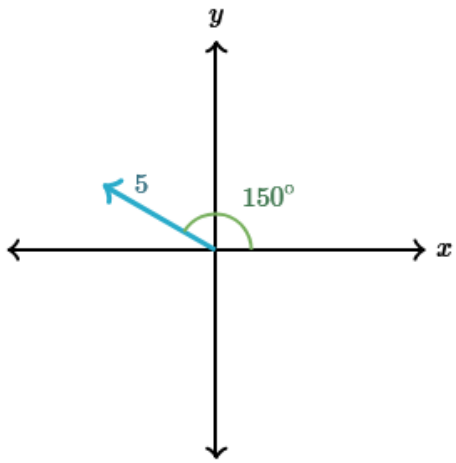


Image credit: Khan Academy

Solution

$\mathbf{u} = (5\cos 150°, 5\sin 150°) = (-4.33, 2.50)$

Problem 1.3.10

Find the magnitude and direction angle θ of $\mathbf{v} + \mathbf{w}$.

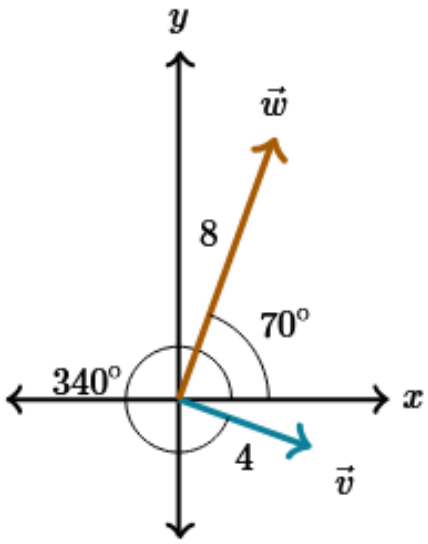


Image credit: Khan Academy

Solution

We have $\mathbf{v} = (4\cos 340°, 4\sin 340°)$ and $\mathbf{w} = (8\cos 70°, 8\sin 70°)$. The magnitude is given by

$$\|\mathbf{v} + \mathbf{w}\| = \sqrt{(4\cos 340° + 8\cos 70°)^2 + (4\sin 340° + 8\sin 70°)^2} = 8.94.$$

The direction angle θ is given by

$$\theta = \tan^{-1}\left(\frac{4\sin 340° + 8\sin 70°}{4\cos 340° + 8\cos 70°}\right) = 43.4°.$$

Problem 1.3.11

A point P divides a line segment AB in the ratio $\lambda : \mu$. Let the position vectors of the points A and B be $\mathbf{a}$ and $\mathbf{b}$, respectively. Find the position vector $\mathbf{p}$ of the point P.

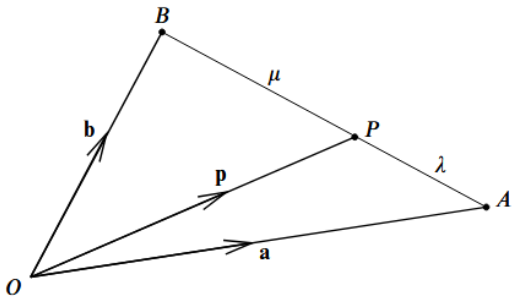


Image credit: Riley et al

Solution

We use a method called the ratio theorem to solve this. Let the vector from point A to point B be $\overrightarrow{AB} = \mathbf{b} - \mathbf{a}$. In order to find $\mathbf{p}$, we need to go from the origin to point A and then to point P. The former is given by $\mathbf{a}$ and the latter is a fraction of $\overrightarrow{AB}$ given by $\frac{\lambda}{\lambda+u}$. Combining those, we have

$$\begin{aligned}\mathbf{p} &= \mathbf{a} + \frac{\lambda}{\lambda+\mu}\overrightarrow{AB} \\ &= \mathbf{a} + \frac{\lambda}{\lambda+\mu}(\mathbf{b}-\mathbf{a}) \\ &= \left(1 - \frac{\lambda}{\lambda+\mu}\right)\mathbf{a} + \frac{\lambda}{\lambda+\mu}\mathbf{b} \\ &= \frac{\mu}{\lambda+\mu}\mathbf{a} + \frac{\lambda}{\lambda+\mu}\mathbf{b}.\end{aligned}$$

If you were to do this for the other path, you would use $\overrightarrow{BA} = (\mathbf{a} - \mathbf{b})$ as the full segment instead of $\overrightarrow{AB}$ and acquire the same result.

Problem 1.3.12

The vertices of triangle ABC have position vectors $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ relative to the origin. Find the position vector of the centroid $\mathbf{G}$ of the triangle.

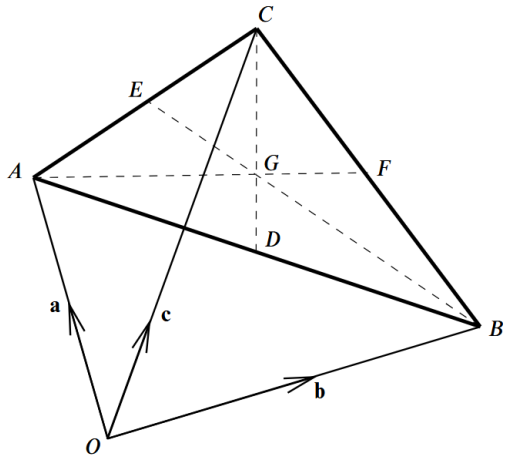


Image credit: Riley et al

Solution

We begin by noting that the centroid is the intersection point of the medians of a triangle.

Let D and E be the midpoints of segments AB and AC, respectively. Then we use our method from **Problem 1.3.11** with $\lambda = \mu = \frac{1}{2}$ to acquire the following:

$$\mathbf{d} = \frac{1}{2}\mathbf{a} + \frac{1}{2}\mathbf{b}$$
$$\mathbf{e} = \frac{1}{2}\mathbf{a} + \frac{1}{2}\mathbf{c}$$

Let point G divide the line CD in the ratio $\lambda : (1 - \lambda)$. Then using the ratio theorem again, we have

$$\begin{aligned}\mathbf{r} &= (1-\lambda)\mathbf{c} + \lambda\mathbf{d} \\ &= (1-\lambda)\mathbf{c} + \lambda\left(\frac{1}{2}\mathbf{a} + \frac{1}{2}\mathbf{b}\right) \\ &= (1-\lambda)\mathbf{c} + \frac{\lambda}{2}(\mathbf{a}+\mathbf{b}).\end{aligned}$$

Let point G also divide the line BE in the ratio $\mu : 1$. Then the position vector of a general point on the line BE is

$$\begin{aligned}\mathbf{r} &= (1-\mu)\mathbf{b} + \mu\mathbf{e} \\ &= (1-\mu)\mathbf{b} + \mu\left(\frac{1}{2}\mathbf{a} + \frac{1}{2}\mathbf{c}\right) \\ &= (1-\mu)\mathbf{b} + \frac{\mu}{2}(\mathbf{a}+\mathbf{c}).\end{aligned}$$

We find the intersection of the lines CD and BE by solving

$$(1-\lambda)\mathbf{c} + \frac{\lambda}{2}(\mathbf{a}+\mathbf{b}) = (1-\mu)\mathbf{b} + \frac{\mu}{2}(\mathbf{a}+\mathbf{c}).$$

We equate the coefficients of $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ to yield

$$\lambda = \mu, \quad \frac{\lambda}{2} = 1 - \mu, \quad 1 - \lambda = \frac{\mu}{2}.$$

Solving these equations, we find $\lambda = \mu = \frac{2}{3}$. Substituting this in to $\mathbf{r}$, we get our final answer:

$$\begin{aligned}\mathbf{G} &= (1-\lambda)\mathbf{c} + \frac{\lambda}{2}(\mathbf{a}+\mathbf{b}) \\ &= \frac{1}{3}\mathbf{c} + \frac{1}{3}(\mathbf{a}+\mathbf{b}) \\ &= \frac{1}{3}(\mathbf{a}+\mathbf{b}+\mathbf{c})\end{aligned}$$

Problem 1.3.13

Particle 1 has velocity given by $\mathbf{v}_1 = \mathbf{i}+3\mathbf{j}+6\mathbf{k}$. Particle 2 has velocity given by $\mathbf{v}_2 = \mathbf{i}-2\mathbf{k}$. Find the velocity of the second particle $\mathbf{u}$ relative to the first.

Solution

$\mathbf{u} = \mathbf{v}_2 - \mathbf{v}_1 = (1-1)\mathbf{i} + (0-3)\mathbf{j} + (-2-6)\mathbf{k} = -3\mathbf{j} - 8\mathbf{k}$

Problem 1.3.14

Find the angle θ between the vectors $\mathbf{a} = \mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$ and $\mathbf{b} = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$.

Solution

The formula is $\cos\theta = \frac{\mathbf{a}\cdot\mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}$.

We now compute:

$$\cos\theta = \left(\frac{1\times 2 + 2\times 3 + 3\times 4}{\sqrt{1^2+2^2+3^2}\sqrt{2^2+3^2+4^2}}\right) = \frac{20}{\sqrt{14}\sqrt{29}} = 0.99 \Rightarrow \theta = 0.12 \text{ rad}$$

Problem 1.3.15

(a) Find the area A of a parallelogram with sides $\mathbf{a} = \mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$ and $\mathbf{b} = 4\mathbf{i} + 5\mathbf{j} + 6\mathbf{k}$.
(b) Suppose we generalize the parallelogram to three dimensions so that it's a parallelepiped. Sides $\mathbf{a}$ and $\mathbf{b}$ remain the same, but we now have side $\mathbf{c} = 7\mathbf{i} + 8\mathbf{j} + 10\mathbf{k}$. Find its volume.

Solution

(a) The cross product is given by $\mathbf{a}\times\mathbf{b} = (2\times6-3\times5)\mathbf{i}+(3\times4-1\times6)\mathbf{j}+(1\times5-2\times4)\mathbf{k} = -3\mathbf{i}+6\mathbf{j}-3\mathbf{k}$. The area of the parallelogram is thus given by

$$A = \|\mathbf{a}\times\mathbf{b}\| = \sqrt{(-3)^2+6^2+(-3)^2} = \sqrt{54}.$$

(b) The volume of the parallelepiped is given by

$$\begin{aligned} V &= \|(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}\| \\ &= \|(-3\mathbf{i}+6\mathbf{j}-3\mathbf{k})\cdot(7\mathbf{i}+8\mathbf{j}+10\mathbf{k})\| \\ &= \|(-3)(7)+(6)(8)+(-3)(10)\| = \|-3\| \\ &= 3. \end{aligned}$$

Problem 1.3.16

A hooked mass lies flat on a lab bench. It is being pulled on by three strings.

- String 1 pulling 1 N directed east and 2 N directed north.
- String 2 is pulling 1.5 N directed west and 1 N directed north.
- String 3 is pulling 2 N directed east and 2.5 N directed south.

Find the direction of the resultant force acting on the hooked mass.

Solution

Let east be positive x, north be positive y, west be negative x, and south be negative y. Then, we break the forces into components:

$$\mathbf{F}_1 = \langle 1, 2 \rangle$$
$$\mathbf{F}_2 = \langle -1.5, 1 \rangle$$
$$\mathbf{F}_3 = \langle 2, -2.5 \rangle$$

We then combine the x- and y-components to get the resultant force vector:

$$\mathbf{F}_{\text{tot}} = \langle (1 + (-1.5) + 2), (2 + 1 + (-2.5)) \rangle = \langle 1.5, 0.5 \rangle$$

The direction angle is given by

$$\theta = \tan^{-1}\left(\frac{0.5}{1.5}\right) = 18.4°.$$

This would be directed east.

1.4 Vector Usage in Linear Algebra

Whether you internalize vectors as physical quantities with magnitude and direction or as lists of numbers largely depends on your background. Because at the end of the day, both are valid. That being said, perhaps the best way to look at it is that vectors encode data. In linear algebra, we use the properties of vectors to explore *vector spaces* and solve linear equations.

Vector addition and scalar-vector multiplication let us combine information to see new insights. The inner product, or more specifically the dot product, measures how aligned two vectors are. Vector norms represent the magnitude of vectors and allow us to determine the distance between points of data.

To begin our study of linear algebra, we will view a vector as a list of data organized either vertically or horizontally. As a matter of fact, the word vector is derived from Latin where it meant "carrier." Vectors are indeed carriers of information. We begin with the vertical way of arranging data:

Column Vector

Definition: A **column vector** $\mathbf{x} \in \mathbb{R}^{n \times 1}$ is a list of n real numbers written vertically with n rows and 1 column:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Each $i \in \{1, 2, \ldots, n\}$ denotes an entry in the ith row of $\mathbf{x}$, written as x_i.

Column vectors are the most common way of storing data through vectors. For instance, the vector

$$\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}$$

represents a point. The first entry $\mathbf{v}_1$ represents the distance along the x-axis and the second entry $\mathbf{v}_2$ represents the distance along the y-axis. The vector

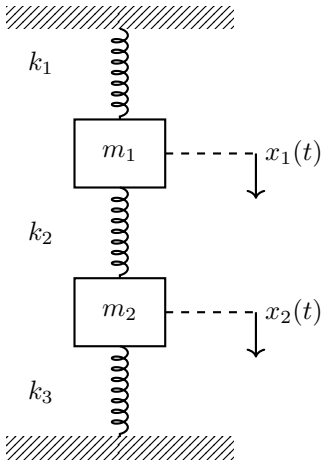
$$\mathbf{x} = \begin{bmatrix} 12 \\ 27 \\ 56 \end{bmatrix}$$

could represent physical quantities like velocity or the amount of people who voted for something.

Let's work through an example.

Example 1.4.1

Suppose we have a coupled mass-spring system. Construct a vector model to represent the positions of the masses at a given point in time.



Solution

All of the given masses m_i measured in kilograms can be represented as

$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}.$$

The spring constants k_j in Newtons per meter (N/m) can be represented as

$$\mathbf{k} = \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} 200 \\ 200 \\ 200 \end{bmatrix}.$$

Let $\mathbf{x}(t) : \mathbb{R} \to \mathbb{R}^2$ encode the vertical positions of the two masses at time t, where

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \quad \text{with} \quad \begin{cases} x_1(t) = \text{height of mass } m_1 \text{ (cm) from the base} \\ x_2(t) = \text{height of mass } m_2 \text{ (cm) from the base} \end{cases}$$

For instance, at time $t = 0$, we have the equilibrium positions:

$$\mathbf{x}(0) = \begin{bmatrix} 21.2 \\ 10.6 \end{bmatrix}$$

We now introduce the horizontal way of representing data through vectors:

Row Vector

A **row vector** $\mathbf{x} \in \mathbb{R}^{1\times n}$ is a list of n real numbers written horizontally with 1 row and n columns:

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

Each $k \in \{1, 2, \ldots, n\}$ denotes an entry in the kth column of $\mathbf{x}$, written as x_k.

It is standard to refer to column vectors as $\mathbb{R}^n$, but row vectors must be written out entirely as $\mathbb{R}^{1\times n}$.

We will now go over vector operations in a way that prepares us for linear algebra. First off, we need to be able to identify when two vectors are equal:

Equal Column Vectors

Let $\mathbf{x} \in \mathbb{R}^{n\times 1}$ and $\mathbf{y} \in \mathbb{R}^{m\times 1}$ be column vectors. These vectors are equal if and only if the vectors have the same number of rows ($n = m$) and $x_i = y_i$ for all $i \in \{1, 2, \ldots, n\}$. In other words, for vectors of identical dimension, $\mathbf{x} = \mathbf{y}$ is true if and only if

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Equal Row Vectors

Let $\mathbf{x} \in \mathbb{R}^{1\times n}$ and $\mathbf{y} \in \mathbb{R}^{1\times m}$ be row vectors. These vectors are equal if and only if the vectors have the same number of columns ($n = m$) and $x_k = y_k$ for all $k \in \{1, 2, \ldots, n\}$. In other words, for vectors of identical dimension, $\mathbf{x} = \mathbf{y}$ is true if and only if

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}.$$

And we will now introduce the most basic vector operations:

Scalar-vector Multiplication

Let $\mathbf{x} \in \mathbb{R}^n$ be a column vector and let $a \in \mathbb{R}$ be a scalar. We define the operation of **scalar-vector multiplication** as

$$a\mathbf{x} = a\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{bmatrix},$$

where the ith coefficient of this product is given by ax_i for all $i \in \{1, 2, \ldots, n\}$.

Column Vector Addition

Let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$ be column vectors. We define the operation of **column vector addition** as

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix},$$

We can only sum vectors that have the same number of rows and the same number of columns ($n = m$).

When we operate on column vectors, here are some of the most important algebraic properties:

Algebraic Properties of Vector Addition and Scalar Multiplication

Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^{n \times 1}$ and $a, b \in \mathbb{R}$. Then, all of the following are properties of column vector addition:

1. Commutativity of vector addition: $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$
2. Associativity of vector addition: $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$
3. Additive identity: $\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$
4. Additive inverses: $\mathbf{x} + (-\mathbf{x}) = -\mathbf{x} + \mathbf{x} = \mathbf{0}$
5. Distributivity over vector addition: $a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}$
6. Distributivity over scalar addition: $(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$
7. Associativity of scalar multiplication: $a(b\mathbf{x}) = (ab)\mathbf{x}$
8. Multiplicative identity: $1\mathbf{x} = \mathbf{x}$

Proof of Commutativity of Vector Addition

We know that $x_i + y_i = y_i + x_i$ since addition of real numbers is commutative. Thus, we have

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} = \begin{bmatrix} y_1 + x_1 \\ y_2 + x_2 \\ \vdots \\ y_n + x_n \end{bmatrix} = \mathbf{y} + \mathbf{x}.$$

□

Proof of Associativity of Vector Addition

Consider that

$$\mathbf{x}+(\mathbf{y}+\mathbf{z})=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}+\left(\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}+\begin{bmatrix}z_1\\z_2\\\vdots\\z_n\end{bmatrix}\right)=\begin{bmatrix}x_1+(y_1+z_1)\\x_2+(y_2+z_2)\\\vdots\\x_n+(y_n+z_n)\end{bmatrix}.$$

Since real number addition is associative, we know that $x_i+(y_i+z_i)=(x_i+y_i)+z_i$. Hence,

$$\mathbf{x}+(\mathbf{y}+\mathbf{z})=\begin{bmatrix}(x_1+y_1)+z_1\\(x_2+y_2)+z_2\\\vdots\\(x_n+y_n)+z_n\end{bmatrix}=\left(\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}+\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}\right)+\begin{bmatrix}z_1\\z_2\\\vdots\\z_n\end{bmatrix}=(\mathbf{x}+\mathbf{y})+\mathbf{z}.$$

□

Proof of Additive Identity

Let $\mathbf{0}\in\mathbb{R}^n$ be the zero vector:

$$\mathbf{x}+\mathbf{0}=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}+\begin{bmatrix}0\\0\\\vdots\\0\end{bmatrix}=\begin{bmatrix}x_1+0\\x_2+0\\\vdots\\x_n+0\end{bmatrix}$$

Since any $x_i + 0 = x_i$, we can say that

$$\begin{bmatrix} x_1 + 0 \\ x_2 + 0 \\ \vdots \\ x_n + 0 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{x}.$$

□

Proof of Distributivity Over Vector Addition

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $a \in \mathbb{R}$. We have

$$a(\mathbf{x} + \mathbf{y}) = a \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} = \begin{bmatrix} a(x_1 + y_1) \\ a(x_2 + y_2) \\ \vdots \\ a(x_n + y_n) \end{bmatrix}.$$

Using the distributive property of real numbers, we rewrite each entry:

$$\begin{bmatrix} ax_1 + ay_1 \\ ax_2 + ay_2 \\ \vdots \\ ax_n + ay_n \end{bmatrix} = \begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{bmatrix} + \begin{bmatrix} ay_1 \\ ay_2 \\ \vdots \\ ay_n \end{bmatrix} = a\mathbf{x} + a\mathbf{y}$$

□

Proof of Associativity of Scalar Multiplication

Let $\mathbf{x} \in \mathbb{R}^1$ and $a, b \in \mathbb{R}$. We have

$$b\mathbf{x} = \begin{bmatrix} bx_1 \\ bx_2 \\ \vdots \\ bx_n \end{bmatrix}.$$

Then,

$$a(b\mathbf{x}) = a \cdot \begin{bmatrix} bx_1 \\ bx_2 \\ \vdots \\ bx_n \end{bmatrix} = \begin{bmatrix} abx_1 \\ abx_2 \\ \vdots \\ abx_n \end{bmatrix}.$$

Since scalar multiplication is associative, we know $abx_i = (ab)x_i$ for all i. Thus

$$\begin{bmatrix} (ab)x_1 \\ (ab)x_2 \\ \vdots \\ (ab)x_n \end{bmatrix} = (ab)\mathbf{x}.$$

□

Let's say we want to change the way we arrange row and column vectors. We have an operation that lets us do this:

Transpose of a Vector

We define the **transpose** of a column vector $\mathbf{x} \in \mathbb{R}^n$ as the $1 \times n$ row vector

$$\mathbf{x}^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix},$$

where the entry in the ith row of $\mathbf{x}$ becomes the entry in the ith column of $\mathbf{x}^T$.

On the other hand, the **transpose** of a row vector $\mathbf{y} \in \mathbb{R}^{1 \times n}$ is the $n \times 1$ column vector

$$\mathbf{y}^T = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

where the coefficient in the kth row of $\mathbf{y}$ becomes the entry in the kth column of $\mathbf{y}^T$.

The transpose also comes with a set of algebraic properties:

Algebraic Properties of Vector Transposes

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n \times 1}$ and $a, b \in \mathbb{R}$. Then, all of the following are properties of vector transposes:

1. $(\mathbf{x}^T)^T = \mathbf{x}$
2. $(\mathbf{x} + \mathbf{y})^T = \mathbf{x}^T + \mathbf{y}^T$
3. $(a\mathbf{x})^T = a\mathbf{x}^T$

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $a \in \mathbb{R}$.

Proof of 1.

$$(\mathbf{x}^T)^T = \left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}^T \right)^T = \left(\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \right)^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{x}$$

□

Proof of 2.

$$(\mathbf{x}+\mathbf{y})^T = \left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \right)^T = \left(\begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} \right)^T = \begin{bmatrix} x_1 + y_1 & x_2 + y_2 & \cdots & x_n + y_n \end{bmatrix}$$

$$= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} + \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \mathbf{x}^T + \mathbf{y}^T$$

□

Proof of 3.

$$(a\mathbf{x})^T = \left(a \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right)^T = \left(\begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{bmatrix} \right)^T = \begin{bmatrix} ax_1 & ax_2 & \cdots & ax_n \end{bmatrix} = a \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} = a\mathbf{x}^T$$

□

Problem 1.4.1

Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^4$ be given by

$$\mathbf{x} = \begin{bmatrix} 1 \\ -2 \\ 3 \\ 0 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 4 \\ -1 \\ 0 \\ 2 \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} -3 \\ 5 \\ 1 \\ -4 \end{bmatrix}$$

Use these vectors to compute the following:

(a) $\mathbf{x} + \mathbf{z}$

(b) $\mathbf{x}^T + \mathbf{y}^T$

(c) $\mathbf{y} + \mathbf{z}^T$

(d) $2\mathbf{x} - \mathbf{1y}$

Solution

(a)

$$\mathbf{x}+\mathbf{z}=\begin{bmatrix}1\\-2\\3\\0\end{bmatrix}+\begin{bmatrix}-3\\5\\1\\-4\end{bmatrix}=\begin{bmatrix}1+(-3)\\-2+5\\3+1\\0+(-4)\end{bmatrix}=\begin{bmatrix}-2\\3\\4\\-4\end{bmatrix}$$

(b)

$$\mathbf{x}^T+\mathbf{y}^T=\begin{bmatrix}1 & -2 & 3 & 0\end{bmatrix}+\begin{bmatrix}4 & -1 & 0 & 2\end{bmatrix}=\begin{bmatrix}1+4 & -2+(-1) & 3+0 & 0+2\end{bmatrix}=\begin{bmatrix}5 & -3 & 3 & 2\end{bmatrix}$$

(c) These dimensions are not identical:

$$\mathbf{y}\in\mathbb{R}^{4\times 1},\quad \mathbf{z}^T\in\mathbb{R}^{1\times 4}$$

Therefore, the operation is not defined.

(d)

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

Then,

$$2\mathbf{x}-\mathbf{y}=\begin{bmatrix}2\\-4\\6\\0\end{bmatrix}-\begin{bmatrix}4\\-1\\0\\2\end{bmatrix}=\begin{bmatrix}2-4\\-4-(-1)\\6-0\\0-2\end{bmatrix}=\begin{bmatrix}-2\\-3\\6\\-2\end{bmatrix}.$$

We now will formalize a few more tools, some of which you used in **Section 1.3**.

Inner Product

Definition: Let $n \in \mathbb{N}$ and suppose that $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then, the inner product between $\mathbf{x}$ and $\mathbf{y}$ is given by

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = \sum_{j=1}^{n} x_j y_j.$$

The dot product is simply the inner product but in Euclidean space. It has domain $\mathbb{R}^n \times \mathbb{R}^n$ and codomain $\mathbb{R}$. It comes with the following properties:

Algebraic Properties of the Dot Product

Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^{n \times 1}$ and $a, b \in \mathbb{R}$. Then, all of the following are algebraic properties of the dot product:

i. **Bilinearity:**
   (a) Linearity in left argument: $(a\mathbf{x} + b\mathbf{y}) \cdot \mathbf{z} = a(\mathbf{x} \cdot \mathbf{z}) + b(\mathbf{y} \cdot \mathbf{z})$
   (b) Linearity in right argument: $\mathbf{x} \cdot (a\mathbf{y} + b\mathbf{z}) = a(\mathbf{x} \cdot \mathbf{y}) + b(\mathbf{x} \cdot \mathbf{z})$

ii. **Symmetry:** $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$

iii. **Positivity:** $\mathbf{x} \cdot \mathbf{x} > 0$ when $\mathbf{x} \neq \mathbf{0}$, while $\mathbf{0} \cdot \mathbf{0} = 0$

We will prove two of these.

Proof of Bilinearity

Let

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}.$$

Then

$$
\begin{aligned}
(a\mathbf{x}+b\mathbf{y})\cdot\mathbf{z} &= \begin{bmatrix} ax_1+by_1 \\ ax_2+by_2 \\ \vdots \\ ax_n+by_n \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \\
&= \sum_{j=1}^{n}(ax_j+by_j)z_j \\
&= a\sum_{j=1}^{n}x_jz_j + b\sum_{j=1}^{n}y_jz_j \\
&= a(\mathbf{x}\cdot\mathbf{z}) + b(\mathbf{y}\cdot\mathbf{z}).
\end{aligned}
$$

□

Proof of Positivity

Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$. Then we can write out the dot product as

$$\mathbf{x}\cdot\mathbf{x} = x_1^2 + x_2^2 + \cdots + x_n^2 = \sum_{j=1}^{n} x_j^2.$$

Since each squared term x_j^2 is always nonnegative, the entire sum is nonnegative:

$$\mathbf{x}\cdot\mathbf{x} \geq 0$$

Moreover, the only way this sum can equal zero is if every component $x_j = 0$. Thus the dot product is always nonnegative, and equals zero only for the zero vector. □

When we computed the magnitude or length of a vector, we used these:

The Magnitude of a Vector

Let $n \in \mathbb{N}$ be a positive integer and let $\mathbf{x} \in \mathbb{R}^n$. Then the 2-norm $\|\mathbf{x}\|_2 \in \mathbb{R}$ or Euclidean norm $\|\mathbf{x}\| \in \mathbb{R}$ represents the magnitude of a vector and is given by

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\sum_{j=1}^{n} x_j^2}.$$

We can also compute this using the inner product:

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}.$$

Algebraic Properties of the Euclidean norm

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n \times 1}$ and $a \in \mathbb{R}$. Then all of the following are algebraic properties of the Euclidean norm:

i. **Positivity:** $\|\mathbf{x}\| \geq 0$ with $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$

ii. **Homogeneity:** $\|a\mathbf{x}\| = |a| \cdot \|\mathbf{x}\|$

iii. **Triangle inequality:** $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

Proof of Homogeneity

Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and let $a \in \mathbb{R}$. Then scalar-vector multiplication gives

$$a\mathbf{x} = a \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{bmatrix}.$$

We now compute the Euclidean norm:

$$\begin{aligned}\|a\mathbf{x}\| &= \sqrt{(ax_1)^2 + (ax_2)^2 + \cdots + (ax_n)^2} \\ &= \sqrt{a^2 \cdot (x_1^2 + x_2^2 + \cdots + x_n^2)} \\ &= \sqrt{a^2} \cdot \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \\ &= |a| \cdot \|\mathbf{x}\|\end{aligned}$$

□

We now formalize some important geometric ideas:

Cosine Formula for Dot Product

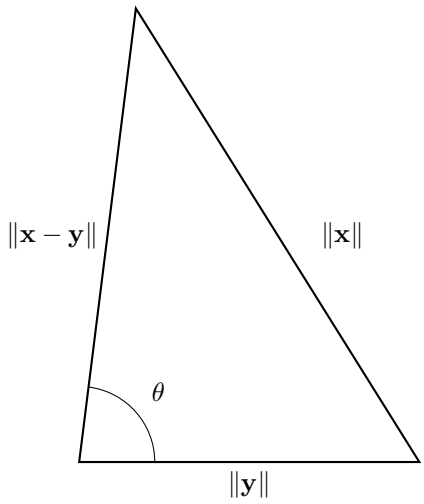
Let $n \in \mathbb{N}$ and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then the dot product satisfies

$$\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\| \cos\theta$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$.

Proof of Cosine Formula for Dot Product

Assume $\mathbf{x}$ and $\mathbf{y}$ are not scalar multiples of each other. We have two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ that define the following triangle:



Using the law of cosines, we get

$$\|\mathbf{x}-\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta.$$

Using the algebraic properties of the dot product, we can simplify:

$$\begin{aligned}\|\mathbf{x}-\mathbf{y}\|^2 &= (\mathbf{x}-\mathbf{y})\cdot(\mathbf{x}-\mathbf{y}) \\ &= \mathbf{x}\cdot\mathbf{x} - 2\mathbf{x}\cdot\mathbf{y} + \mathbf{y}\cdot\mathbf{y} \\ &= \|\mathbf{x}\|^2 - 2\mathbf{x}\cdot\mathbf{y} + \|\mathbf{y}\|^2\end{aligned}$$

We equate both expressions:

$$\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta = \|\mathbf{x}\|^2 - 2\mathbf{x}\cdot\mathbf{y} + \|\mathbf{y}\|^2$$

We subtract $\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ from both sides:

$$-2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta = -2\mathbf{x}\cdot\mathbf{y}$$

Dividing both sides by -2, we obtain

$$\mathbf{x}\cdot\mathbf{y} = \|\mathbf{x}\|\,\|\mathbf{y}\|\cos\theta.$$

And now let's assume that $\mathbf{x}$ and $\mathbf{y}$ are scalar multiples of one another. That is, $\mathbf{y} = a\mathbf{x}$ for some scalar $a \in \mathbb{R}$. This means $\mathbf{y}$ points in the same direction, the opposite direction, or becomes the zero vector depending on the value of a.

We can compute the dot product directly:

$$\mathbf{x}\cdot\mathbf{y} = \mathbf{x}\cdot(a\mathbf{x}) = a(\mathbf{x}\cdot\mathbf{x}) = a\|\mathbf{x}\|^2$$

Now let's relate this to the geometric form of the dot product:

$$\mathbf{x}\cdot\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta$$

We'll verify this by checking different cases for the scalar a.

In the case that $a > 0$, $\mathbf{y}$ points in the same direction as $\mathbf{x}$, so the angle between them is $\theta = 0$, and $\cos\theta = 1$.

Thus we know that

$$\|\mathbf{y}\| = \|a\mathbf{x}\| = |a|\|\mathbf{x}\| = a\|\mathbf{x}\|.$$

Then,

$$\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta = \|\mathbf{x}\|(a\|\mathbf{x}\|)(1) = a\|\mathbf{x}\|^2.$$

In the case that $a < 0$, $\mathbf{y}$ points in the opposite direction as $\mathbf{x}$, so $\theta = \pi$, and $\cos\theta = -1$.

We again have

$$\|\mathbf{y}\| = |a|\|\mathbf{x}\| = -a\|\mathbf{x}\|.$$

Then,

$$\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta = \|\mathbf{x}\|(|a|\|\mathbf{x}\|)(-1) = -|a|\|\mathbf{x}\|^2 = a\|\mathbf{x}\|^2.$$

Lastly, in the case that $a = 0$, we have $\mathbf{y} = \mathbf{0}$, the zero vector. Because $\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot \mathbf{0} = 0$ and $\|\mathbf{y}\| = 0$, we have

$$\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta = 0.$$

□

And now we will use the law of cosines to make a statement about the relationship between the lengths of general column vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{x} - \mathbf{y}$.

Cauchy–Schwarz Inequality

Let $n \in \mathbb{N}$ be a positive integer and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then,

$$|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\| \, \|\mathbf{y}\|.$$

This inequality tells us that the absolute value of the dot product between two vectors is always less than or equal to the product of their magnitudes. Intuitively, the dot product measures how much two vectors point in the same direction. If they point in the exact same direction, the inequality becomes an equality. If they point in orthogonal (perpendicular) directions, the dot product is zero, which still satisfies the inequality. If they point in opposite directions, the dot product becomes negative, and the absolute value still satisfies the bound.

Proof of Cauchy-Schwarz Inequality

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. By the definition of the dot product, we have

$$\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\| \, \|\mathbf{y}\| \cos\theta.$$

The cosine of any real angle θ satisfies

$$-1 \leq \cos\theta \leq 1$$

for all θ. We can multiply the inequality by $\|\mathbf{x}\| \, \|\mathbf{y}\| \geq 0$, giving

$$-\|\mathbf{x}\| \, \|\mathbf{y}\| \leq \mathbf{x} \cdot \mathbf{y} \leq \|\mathbf{x}\| \, \|\mathbf{y}\|.$$

And then we take the absolute value of the expression:

$$|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\| \, \|\mathbf{y}\|.$$

□

And finally, we have an important condition that will reappear and reappear later on:

Orthogonal Vectors

Two vectors are **orthogonal** if and only if their dot product is zero:

$$\mathbf{x} \cdot \mathbf{y} = 0$$

Now that we've established developed some foundational tools, we are ready to begin our journey into linear algebra.

2 Linear Equations and Matrix Algebra

Systems of linear equations lie at the heart of linear algebra. You may wonder: if most of the real world is resembles nonlinear equations, why are linear equations so important? The answer to that is that we often *linearize* nonlinear systems to make them easier to compute and analyze. In this chapter, we will investigate systems of linear equations and learn how representing them through matrices is one of the most powerful ideas in modern day mathematics.

2.1 Systems of Linear Equations

A **linear equation** in the variables $x_1, x_2, \ldots, x_n$ can be written in the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b \tag{1}$$

where $a, b \in \mathbb{C}$ and n is any positive integer.

For instance, $3x_1 - 2x_2 = 5$ is already in Form 1. On the other hand, the equation $x_2 = 2\left(\sqrt{6} - x_1\right) + x + 3$ can be rewritten in Form 1:

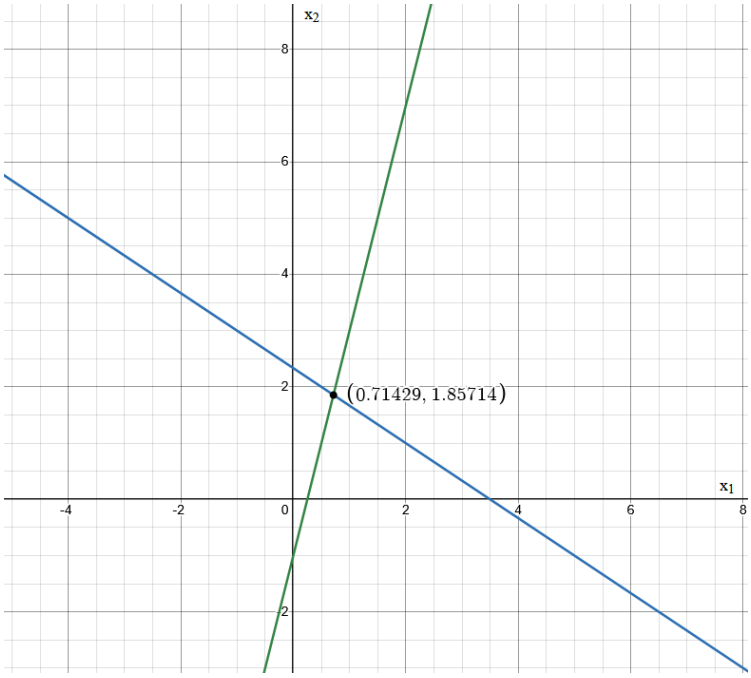
$$2x_1 + x_2 - x_3 = 2\sqrt{6}$$

However, the equation $4x_1 - 3x_2 = x_1x_2$ is not linear because it contains the term x_1x_2. The equation $x_2 = 2\sqrt{x_1} - 6$ is not linear either because it contains a term with $\sqrt{x_1}$.

A **system of linear equations** is simply a collection of one or more linear equations in the variables $x_1, \ldots, x_n$. Here's an example:

$$\begin{cases} 2x_1 + 3x_2 = 7 \\ -4x_1 + x_2 = -1 \end{cases}$$

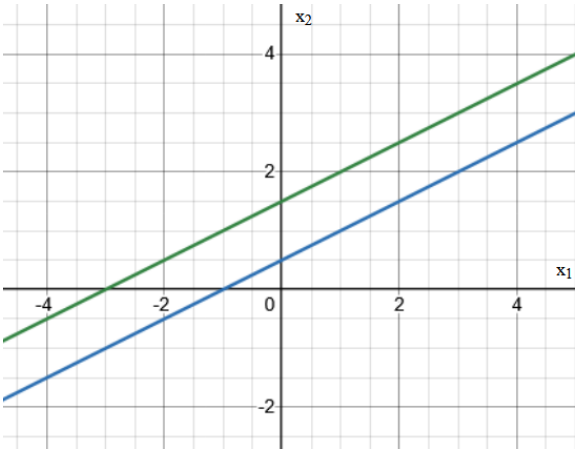
The solution to this would be a list of numbers that makes each equation true. The graph of $2x_1 + 3x_2 = 7$ is the blue line and the graph of $-4x_1 + x_2 = -1$ is the green line. They intersect at one point and thus have one solution:



It is of course possible for systems of linear equations to have no solution. The system

$$\begin{cases} x_1 - 2x_2 = -1 \\ -x_1 + 2x_2 = 3 \end{cases}$$

has no solution as the lines are parallel:



The system

$$\begin{cases} x_1 - 2x_2 = -1 \\ -x_1 + 2x_2 = 1 \end{cases}$$

has two identical linear equations, so the lines overlap at every point and thus the system has infinitely many solutions.

Solutions of Systems of Linear Equations

A system of linear equations has one of the following:

1. no solution
2. exactly one solution
3. infinitely many solutions

If a system has no solution, it is called **inconsistent**. If a system has either one solution or infinitely many solutions, it is called **consistent**.

We can encode systems of linear equations into a rectangular array called a **matrix**. Matrices are special types of vectors that are two-dimensional.

Matrix

Let $m, n \in \mathbb{N}$. A rectangular matrix $\mathbf{A}$ is of dimension $m \times n$ where m is the number of rows and n is the number of columns. We can write the general structure of an $m \times n$ matrix A as:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$$

Each number in a matrix is called an **entry**, **element**, or **coefficient** of the matrix. Each entry has a row index and a column index, which identifies where in the matrix that element is stored.

The system

$$\begin{cases} x_1 - 2x_2 + x_3 = 0 \\ 2x_2 - 8x_3 = 8 \\ 5x_1 - 5x_3 = 10 \end{cases}$$

can be encoded into a matrix where the coefficients are aligned in columns:

$$\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \\ 5 & 0 & -5 \end{bmatrix}$$

This is called the **coefficient matrix**. If we were to add a column containing the constants from the right-hand side of the equation, we would have

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}$$

This is called the **augmented matrix**. In this case, we have a 3×4 matrix. Let's now walk through solving this linear system.

Example 2.1.1

Solve the following system:

$$\begin{cases} x_1 - 2x_2 + x_3 = 0 \\ 2x_2 - 8x_3 = 8 \\ 5x_1 - 5x_3 = 10 \end{cases} \iff \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}$$

Solution

We want to keep x_1 in the first equation and eliminate it from the other two. We begin by multiplying the first equation by -5 and then adding the result to the third equation:

$$\begin{array}{r} -5x_1 + 10x_2 - \ 5x_3 = \ 0 \\ 5x_1 \qquad\quad - \ 5x_3 = 10 \\ \hline 10x_2 - 10x_3 = 10 \end{array}$$

This now becomes our new third equation. Let's update the system:

$$\begin{cases} x_1 - \ 2x_2 + \ x_3 = \ 0 \\ \qquad 2x_2 - \ 8x_3 = \ 8 \\ \qquad 10x_2 - 10x_3 = 10 \end{cases} \iff \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -\ 8 & 8 \\ 0 & 10 & -10 & 10 \end{bmatrix}$$

We continue on by multiplying the second equation by $\frac{1}{2}$ so that the coefficient of x_2 becomes 1:

$$\begin{cases} x_1 - \ 2x_2 + \ x_3 = \ 0 \\ \qquad x_2 - \ 4x_3 = \ 4 \\ \qquad 10x_2 - 10x_3 = 10 \end{cases} \iff \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 10 & -10 & 10 \end{bmatrix}$$

We now multiply the second equation by -10 and then subtract it from the third equation:

$$\begin{array}{r} -10x_2 + 40x_3 \ = -40 \\ 10x_2 - 10x_3 \ = \ 10 \\ \hline 30x_3 \ = -30 \end{array}$$

This eliminates $10x_2$ from the third equation. Let's update the system:

$$\begin{cases} x_1 - 2x_2 + \ x_3 = \ 0 \\ \qquad x_2 - \ 4x_3 = \ 4 \\ \qquad\quad 30x_3 = -30 \end{cases} \iff \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 30 & -30 \end{bmatrix}$$

We can simplify by multiplying the third equation by $\frac{1}{3}$ so that the coefficient of x_3 becomes 1:

$$\begin{cases} x_1 - 2x_2 + \ x_3 = \ 0 \\ \qquad x_2 - 4x_3 = \ 4 \\ \qquad\quad x_3 = -1 \end{cases} \iff \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

Now, we want to eliminate $-4x_3$ and $-x_3$ from the equations above. Multiply the third equation by 4 and add it to the second:

$$\begin{aligned} 4x_3 &= -4 \\ \underline{x_2 - 4x_3} &= 4 \\ x_2 &= 0 \end{aligned}$$

Then, we multiply the third equation by -1 and add it to the first:

$$\begin{aligned} -x_3 &= 1 \\ \underline{x_1 - 2x_2 + x_3} &= 0 \\ x_1 - 2x_2 &= 1 \end{aligned}$$

Let's update the system:

$$\begin{cases} x_1 - 2x_2 = 1 \\ x_2 = 0 \\ x_3 = -1 \end{cases} \iff \begin{bmatrix} 1 & -2 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

We now eliminate the $-2x_2$ term in the first equation by multiplying the second equation by 2 and adding it to the first equation:

$$\begin{aligned} 2x_2 &= 0 \\ \underline{x_1 - 2x_2} &= 1 \\ x_1 &= 1 \end{aligned}$$

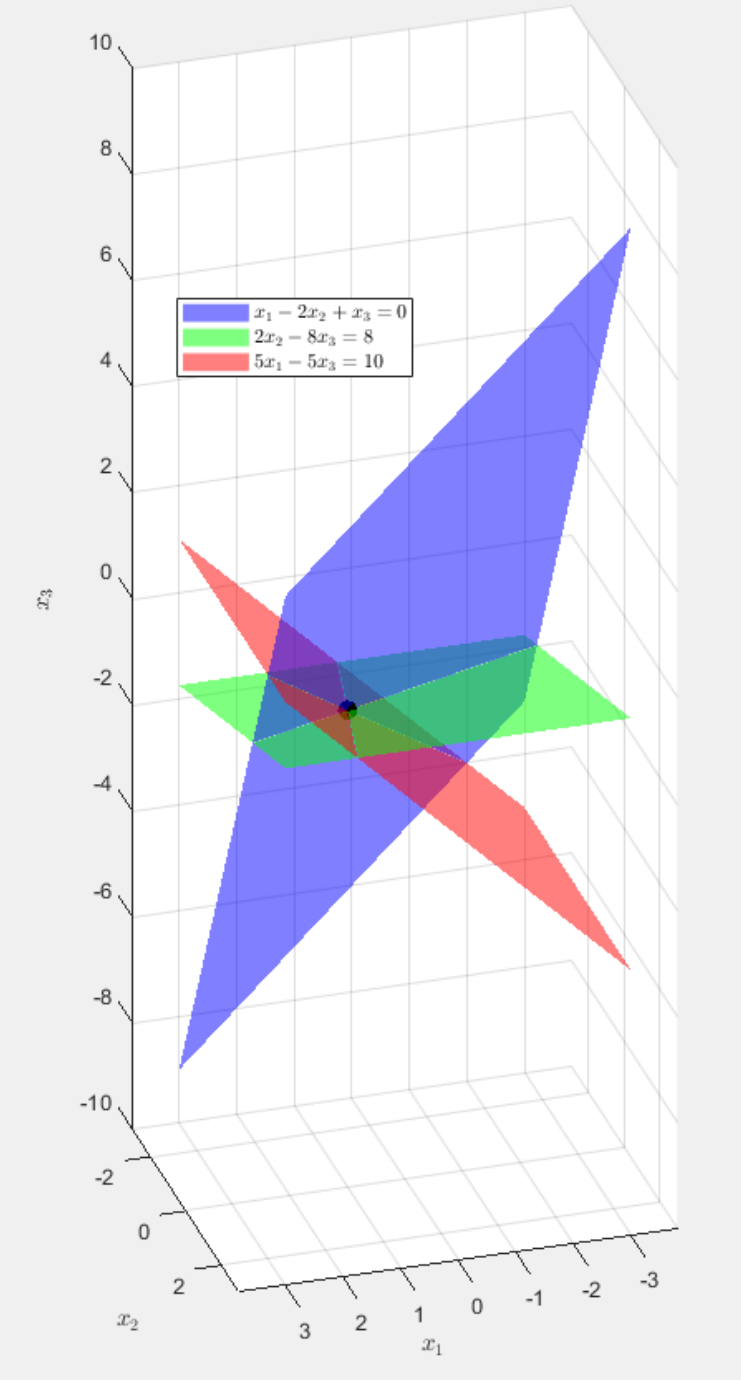
This gives the final system:

$$\begin{cases} x_1 = 1 \\ x_2 = 0 \\ x_3 = -1 \end{cases} \iff \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

The solution is $(x_1, x_2, x_3) = (1, 0, -1)$. We will check our work by plugging the values back in:

$$\begin{cases} 1 - 2(0) + -1 = 1 - 0 - 1 = 0 \\ 2(0) - 8(-1) = 0 + 8 = 8 \\ 5(1) - 5(-1) = 5 + 5 = 10 \end{cases}$$

The results agree. And now let's look at it graphically:



Let's summarize the operations we used to solve that system of linear equations:

Elementary Row Operations

Replacement: Replace one row R_i with the sum of itself and a multiple of another row:

$$R_i + cR_j \to R_i$$

where $i \neq j$ and $c \in \mathbb{R}$.

Interchange: Swap the positions of two rows:

$$R_i \leftrightarrow R_j$$

Scaling: Multiply every entry in a row by a nonzero constant c:

$$cR_i \to R_i$$

These operations can be used on any matrix. Keep in mind when working on problems that these operations are reversible. We say that two matrices are **row equivalent** if there is a series of elementary row operations that transforms one matrix into the other. Elementary row operations *do not* change the solution of a system of linear equations. Thus, we have an important property:

If the augmented matrices of two linear systems are row equivalent, then the two systems have the same solution set.

When you are given a linear system, the first two questions you should ask are

1. Is the system consistent?
2. If a solution exists, is it unique? That is, is it the only one?

You've already seen a consistent linear system. Let's take a look at one that is inconsistent.

Example 2.1.2

Determine if the following system is consistent:

$$\begin{cases} x_2 - 4x_3 = 8 \\ 2x_1 - 3x_2 + 2x_3 = 1 \\ 5x_1 - 8x_2 + 7x_3 = 1 \end{cases}$$

Solution

We first encode it as an augmented matrix:

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

We will interchange rows 1 and 2:

$$\begin{bmatrix} 0 & 1 & -4 & 8 \\ 2 & -3 & 2 & 1 \\ 5 & -8 & 7 & 1 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 5 & -8 & 7 & 1 \end{bmatrix}$$

We did that to bring a nonzero pivot into the first row. We will learn more about this later.

To eliminate the $5x_1$ term in the third equation, we multiply row 1 by $-\frac{5}{2}$ and add it to row 3:

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 5 & -8 & 7 & 1 \end{bmatrix} \xrightarrow{-\frac{5}{2}R_1 + R_3 \to R_3} \begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & -\frac{1}{2} & 2 & -\frac{3}{2} \end{bmatrix}$$

Then we use the x_2 term in the second equation to eliminate the $-\frac{1}{2}x_2$ term from the third equation. We multiply row 2 by $\frac{1}{2}$ and add it to row 3:

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & -\frac{1}{2} & 2 & -\frac{3}{2} \end{bmatrix} \xrightarrow{\frac{1}{2}R_2+R_3\to R_3} \begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & \frac{5}{2} \end{bmatrix}$$

Notice how the matrix is in a triangular form. Keep this in mind. Let's turn our linear system back into equation form:

$$\begin{cases} 2x_1 - 3x_2 + 2x_3 = 1 \\ x_2 - 4x_3 = 8 \\ 0 = \frac{5}{2} \end{cases}$$

As you can see, this system has no solution because $0 \neq \frac{5}{2}$.

Problem 2.1.1

Solve the following system:

$$\begin{cases} x_1 - 3x_2 = 5 \\ -x_1 + x_2 + 5x_3 = 2 \\ x_2 + x_3 = 0 \end{cases}$$

Solution

After writing out the augmented matrix, we see that we want to eliminate the -1 in row 2. To cancel out -1, we can use replacement. We can get this by adding row 1 to row 2:

$$\begin{bmatrix} 1 & -3 & 0 & 5 \\ -1 & 1 & 5 & 2 \\ 0 & 1 & 1 & 0 \end{bmatrix} \xrightarrow{R_1+R_2\to R_2} \begin{bmatrix} 1 & -3 & 0 & 5 \\ 0 & -2 & 5 & 7 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

Now we want to eliminate the 1 in $a_{3,2}$. Let's multiply row 3 by 2 and then add row 2 to it:

$$\begin{bmatrix} 1 & -3 & 0 & 5 \\ 0 & -2 & 5 & 7 \\ 0 & 1 & 1 & 0 \end{bmatrix} \xrightarrow{2R_3+R_2\to R_3} \begin{bmatrix} 1 & -3 & 0 & 5 \\ 0 & -2 & 5 & 7 \\ 0 & 0 & 7 & 7 \end{bmatrix}$$

Now that our matrix is in triangular form, let's write it back out as a system of linear equations and solve:

$$\begin{cases} x_1 - 3x_2 = 5 \\ -2x_2 + 5x_3 = 7 \\ 7x_3 = 7 \end{cases}$$

We start at the bottom. From $7x_3 = 7$, we can see that $x_3 = 1$. If we plug this in to the second equation, we get

$$-2x_2 + 5(1) = 7 \Rightarrow x_2 = -1.$$

Lastly, we substitute this into the first equation to get

$$x_1 - 3(-1) = 5 \Rightarrow x_1 = 2.$$

Thus, our solution is $(x_1, x_2, x_3) = (2, -1, 1)$.

Problem 2.1.2

Determine the value(s) of h that would make the following matrix the augmented matrix of a consistent linear system:

$$\begin{bmatrix} 1 & h & 4 \\ 3 & 6 & 8 \end{bmatrix}$$

Solution

We will eliminate the 3 in row 2 by multiplying row 1 by 3 and then subtracting it from row 2:

$$\begin{bmatrix} 1 & h & 4 \\ 0 & 6-3h & -4 \end{bmatrix}$$

By just looking at this system, we can see that there is a solution. We however, want it to be consistent, so we must avoid contradictions. When you look at row 2, think about what could create a contradiction. That's right, if there were some value that would make $6-3h$ equal to 0, we would have a contradiction:

$$6-3h=0\neq 4$$

Thus, we can solve $6-3h=0$ to get $h=2$. There are no values that could make $0x_2$ equal to -4. Thus this matrix is the augmented matrix of a consistent linear system if and only if $h\neq 2$.

2.2 Row Reduction

In the previous section, we solved systems of linear equations by turning them into augmented matrices and operating until they were in some sort of triangular form. Let's now refine our definition of that so we can develop algorithms for what is known as **row reduction**.

We begin by introducing the idea of a *nonzero* row or column. This means that the row or column contains at least one nonzero element. The **leading entry** of a row refers to its leftmost nonzero element.

Row Echelon Form

A rectangular matrix is in **row echelon form (REF)** if it has the following three properties:

1. All nonzero rows are above any rows of all zeros.
2. Each leading entry of a row is in a column to the right of the leading entry of the row above it.
3. All entries in a column below a leading entry are zeros.

If a matrix in echelon form also satisfies these conditions, then it is in **reduced row echelon form (RREF)**:

4. The leading entry in each nonzero row is 1.
5. Each leading 1 is the only nonzero entry in its column.

This is what we mean when we try to get matrices into that "triangular" form. Let's take a look at two matrices.

The first, which we worked through earlier, is in REF:

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & \frac{5}{2} \end{bmatrix}$$

The second goes a step further and is in RREF:

$$\begin{bmatrix} 1 & 0 & 0 & 29 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

Let the leading entries (■) be any nonzero value, also known as pivot positions. And let the starred entries ($\star$) have any value. The following matrices are in REF:

$$\mathbf{A} = \begin{bmatrix} \blacksquare & \star & \star & \star \\ 0 & \blacksquare & \star & \star \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0 & \blacksquare & \star & \star & \star & \star & \star & \star & \star & \star \\ 0 & 0 & 0 & \blacksquare & \star & \star & \star & \star & \star & \star \\ 0 & 0 & 0 & 0 & \blacksquare & \star & \star & \star & \star & \star \\ 0 & 0 & 0 & 0 & 0 & \blacksquare & \star & \star & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & \star \end{bmatrix}$$

Now, let's change these so that the leading entries are 1s. Then, let's make sure there are 0s below and above each leading 1. These are now in RREF:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \star & \star \\ 0 & 1 & \star & \star \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0 & 1 & \star & \star & \star & \star & \star & \star & \star & \star \\ 0 & 0 & 0 & 1 & \star & \star & \star & \star & \star & \star \\ 0 & 0 & 0 & 0 & 1 & \star & \star & \star & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 1 & \star & \star & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \star \end{bmatrix}$$

Through elementary row operations, our goal is to **row reduce** a matrix into REF or RREF. Any nonzero matrix can be row reduced into more than one matrix in REF, but a matrix reduced to RREF is unique.

Uniqueness of RREF

Each matrix is row equivalent to one and only one RREF matrix.

After you perform elementary row operations to produce an REF matrix, further operations to get an RREF matrix do not change the positions of the leading entries.

With this, we introduce another aspect of our algorithm.

Pivot Position

A **pivot position** in a matrix **A** is a location in **A** that corresponds to a leading 1 in the REF of A. A column of **A** that contains a pivot position is called a **pivot column**.

With this new information, let's try out an example.

Example 2.2.1

Row reduce the following matrix **A** to REF and locate its pivot columns:

$$A = \begin{bmatrix} 0 & -3 & -6 & 4 & 9 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 1 & 4 & 5 & -9 & -7 \end{bmatrix}$$

Solution

We begin with $R_1 \leftrightarrow R_4$ to get

$$= \begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix}.$$

The pivot position is boxed in red (note that a pivot can still be a pivot even if it isn't boxed. I'm just boxing the pivot in that specific step and anything else we've already completed); its column is also now the pivot column. It is properly placed at the top of the leftmost nonzero column. Let's now try to create 0s below the pivot. We can do this by adding multiples of row 1 to the rows below. That is, $R_2 + R_1 \to R_2$ and $R_3 + 2R_1 \to R_3$:

$$\begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 5 & 10 & -15 & -15 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix}$$

By $-\frac{5}{2}R_2 + R_3 \to R_3$ and $\frac{3}{2}R_2 + R_4 \to R_4$, we now have the 2 as the pivot position:

$$\begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & \boxed{2} & 4 & -6 & -6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -5 & 0 \end{bmatrix}$$

There is no way to create a leading entry in column 3. We can't use columns 1 or 2 because we already have those in echelon arrangement. We will instead interchange rows 3 and 4 $R_3 \leftrightarrow R_4$ to bring the -5 into pivot position:

$$\begin{bmatrix} \boxed{1} & 4 & 5 & -9 & -7 \\ 0 & \boxed{2} & 4 & -6 & -6 \\ 0 & 0 & 0 & \boxed{-5} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Let's go over our criteria for an REF matrix:

1. All nonzero rows are above the row of all zeros.
2. The pivot in each nonzero row is to the right of the leading entry in the row above it.
3. All entries below a pivot are zeros.

Thus, the matrix is now in REF. Columns 1, 2, and 4 are pivot columns. It is however, not in RREF because not every pivot is a 1 and not every pivot is the only nonzero entry in its column. Let's now try to explain the algorithm for row reduction, also known as **Gaussian elimination**.

We begin with the following matrix:

$$\begin{bmatrix} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{bmatrix}$$

Begin with the leftmost nonzero column. This is the first pivot column (column 1). **We have to pick which nonzero entry in the pivot column will be the pivot**. The topmost nonzero entry in this column will become the pivot position. We interchange rows 1 and 3 to move the pivot into position:

$$\begin{bmatrix} \boxed{3} & -9 & 12 & -9 & 6 & 15 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}$$

Now we want to use row replacement operations to create zeros in all positions below the pivot. We eliminate the entry below the pivot in column 1 with the row operation $R_2 + (-1)R_1 \to R_2$:

$$\begin{bmatrix} \boxed{3} & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}$$

We now repeat the process until there are no more nonzero rows to modify. Next, we eliminate the entry below the pivot in column 2 with $R_3 + (-\frac{3}{2})R_2 \to R_3$:

$$\begin{bmatrix} \boxed{3} & -9 & 12 & -9 & 6 & 15 \\ 0 & \boxed{2} & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

At this point, the matrix is in REF. To continue to RREF, we **start with the rightmost pivot and work upward to the left until there are zeros above each pivot**. If a pivot is not 1, we should scale it to 1.

We eliminate the entry above the pivot in column 5 with $R_2 + (-2)R_3 \to R_2$ and $R_1 + (-6)R_3 \to R_1$:

$$\begin{bmatrix} \boxed{3} & -9 & 12 & -9 & 0 & -9 \\ 0 & \boxed{2} & -4 & 4 & 0 & -14 \\ 0 & 0 & 0 & 0 & \boxed{1} & 4 \end{bmatrix}$$

Next, we scale row 2 by $\frac{1}{2}$ to make the pivot a 1:

$$\begin{bmatrix} \boxed{3} & -9 & 12 & -9 & 0 & -9 \\ 0 & \boxed{1} & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & \boxed{1} & 4 \end{bmatrix}$$

We eliminate the entry in $a_{1,2}$ by $9R_2 + R_1 \to R_1$:

$$\begin{bmatrix} \boxed{3} & 0 & -6 & 9 & 0 & -72 \\ 0 & \boxed{1} & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & \boxed{1} & 4 \end{bmatrix}$$

Finally, we scale row 1 by $\frac{1}{3}$ to make the pivot a 1:

$$\begin{bmatrix} \boxed{1} & 0 & -2 & 3 & 0 & -24 \\ 0 & \boxed{1} & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & \boxed{1} & 4 \end{bmatrix}$$

This is the RREF of the matrix.

Let's now write out the strategy:

Gaussian Elimination

Step 1: Identify the Pivot Column
Begin with the leftmost nonzero column. This is the first pivot column. Select a nonzero entry in this column to be the pivot position. If necessary, interchange rows to bring this entry into the topmost position of the pivot column.

Step 2: Create Zeros Below the Pivot Position
Use row replacement operations to create zeros in all positions below the pivot. For each row below, subtract an appropriate multiple of the pivot row to eliminate the entry in the pivot column.

Step 3: Move to the Next Pivot Column
Ignore the row containing the pivot position and all rows above it. Move to the rest of the matrix (submatrix) that remains and repeat Steps 1 and 2. Select the leftmost nonzero column of the submatrix as the next pivot column and continue the process. Repeat until there are no more nonzero rows to modify.

Step 4: Achieve REF
Once all pivot positions have been established and zeros are below each pivot, the matrix is in REF. At this point, all non-pivot columns correspond to free variables in the solution of the system.

Step 5: Reduce to RREF
Start with the rightmost pivot and work upward and to the left until you've created zeros above each pivot. If the pivot is not one, make it one by scaling.

Steps 1 through 4 are called the **forward phase** of Gaussian elimination. Step 5, which is optional, is called the **backward phase**.

As you saw previously, when applied to an augmanted matrix, this algorithm leads directly to the solution set of a linear system.

For example, suppose we have arrived at the following augmented matrix in REF:

$$\begin{bmatrix} 1 & 0 & -5 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The associated system of equations is

$$\begin{cases} x_1 - 5x_3 = 1 \\ x_2 + x_3 = 4 \\ 0 = 0 \end{cases}.$$

The variables x_1 and x_2, which correspond to the pivot columns in the matrix, are called **basic variables**. The remaining variable, x_3, corresponds to a non-pivot column and is called a **free variable**.

Whenever a system is consistent, the solution set can be described explicitly by solving for the basic variables in terms of the free variables. This is possible because RREF guarantees that each basic variable appears in exactly one equation.

From the system above, we can solve explicitly for x_1 and x_2 in terms of x_3:

$$\begin{cases} x_1 = 1 + 5x_3 \\ x_2 = 4 - x_3 \\ x_3 \text{ is free} \end{cases} \tag{2}$$

Saying that x_3 is "free" means that we can assign it any value. Once we pick a value for x_3, the corresponding values of x_1 and x_2 are determined by the equations in system (2).

For example, if $x_3 = 0$, then the solution is $(x_1, x_2, x_3) = (1, 4, 0)$. If $x_3 = 1$, then the solution is $(x_1, x_2, x_3) = (6, 3, 1)$. Each different choice of x_3 produces a different solution of the system. Every solution to this system is determined by a choice of x_3.

Example 2.2.2

Find the general solution of the linear system whose augmented matrix has been reduced to

$$\begin{bmatrix} 1 & 6 & 2 & -5 & -2 & -4 \\ 0 & 0 & 2 & -8 & -1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

Solution

The matrix is in REF, but we want it in RREF before solving for the basic variables. The row reduction continues as follows. The symbol $\sim$ signifies that row operations have been performed on a matrix and that the following matrix is row equivalent.

$$\begin{bmatrix} 1 & 6 & 2 & -5 & -2 & -4 \\ 0 & 0 & 2 & -8 & -1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 2 & -5 & 0 & 10 \\ 0 & 0 & 2 & -8 & 0 & 10 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 6 & 2 & -5 & 0 & 10 \\ 0 & 0 & 1 & -4 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 0 & -3 & 0 & 0 \\ 0 & 0 & 1 & -4 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

At this stage, the system has five variables because the augmented matrix has six columns. The corresponding system of equations is:

$$\begin{cases} x_1 + 6x_2 + 3x_4 = 0 \\ x_3 - 4x_4 = 5 \\ x_5 = 7 \end{cases}$$

The pivot columns are 1, 3, and 5. Thus, x_1, x_3, and x_5 are basic variables. The variables x_2 and x_4 correspond to non-pivot columns and are free variables.

We express the basic variables in terms of the free variables to obtain the general solution:

$$\begin{cases} x_1 = -6x_2 - 3x_4 \\ x_2 \text{ is free} \\ x_3 = 5 + 4x_4 \\ x_4 \text{ is free} \\ x_5 = 7 \end{cases}$$

When we solve a system of equations, our goal is to describe the entire set of solutions. If the system has free variables, these free variables act as *parameters* that describe infinitely many solutions. This is called a *parametric description*.

For example, suppose we simplify a system and get the following:

$$\begin{cases} x_1 + 5x_2 = 21 \\ x_2 + x_3 = 4 \end{cases}$$

We can treat x_2 as a free variable (a parameter) and express x_1 and x_3 in terms of x_2. This would give us a complete description of the solution set. Technically, we could pick any variable to parameterize, but to stay consistent, we should always select the free variables.

On the other hand, if a system is inconsistent, the solution set is always empty even if the system has free variables.

You may be tempted to go from matrices back to traditional linear equations when solving a linear system. In other words, you would be doing back-substitution after REF instead of Gaussian elimination all the way to RREF. The reality is that both of these take about the same number of steps, but working with matrices is preferable because of their compactness. It is thus recommended to use RREF all the way to solve a system.

So while a system in REF is not the best for going straight to solving through substitution, it can still be used to answer questions regarding existence and uniqueness. When a system of equations is in REF with no rows that could produce an equation in the form of $0 = b$ where b is a nonzero integer, every nonzero equation involves a basic variable with a nonzero coefficient. This guarantees one and only one of the following:

1. Every leading variable is completely determined. If there are no free variables, every variable has exactly one solution.
2. Every leading variable can be expressed in terms of free variables, thus there are infinitely many solutions.

This leads us to an answer to existence and uniqueness:

Existence and Uniqueness Theorem

A linear system is consistent if and only if the rightmost column of its augmented matrix is *not* a pivot column. That is, a system is consistent if no row in the REF of its augmented matrix is of the form

$$\begin{bmatrix} 0 & \cdots & 0 & b \end{bmatrix}$$

where $b \neq 0$.

If a system is consistent, then one of the following holds:

- It has a **unique solution** if there are no free variables.
- It has **infinitely many solutions** if there is at least one free variable.

With that, we can now outline how to solve a linear system:

Using Row Reduction to Solve a Linear System

1. Write the augmented matrix of the system.
2. Perform row operations to bring it to REF. Determine if the system is consistent. If it is inconsistent, stop.
3. If it is consistent, continue row reducing until RREF.
4. Turn the matrix back into a system of equations.
5. Solve for the basic variables in terms of free variables to describe the solution set.

Problem 2.2.1

Find the general solution of the system:

$$\begin{cases} x_1 - 2x_2 - x_3 + 3x_4 = 0 \\ -2x_1 + 4x_2 + 5x_3 - 5x_4 = 3 \\ 3x_1 - 6x_2 - 6x_3 + 8x_4 = 2 \end{cases}$$

Solution

We have our augmented matrix. We need to eliminate the entries below the first pivot which is already 1:

$$\begin{bmatrix} 1 & -2 & -1 & 3 & 0 \\ -2 & 4 & 5 & -5 & 3 \\ 3 & -6 & -6 & 8 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & -1 & 3 & 0 \\ 0 & 0 & 3 & 1 & 3 \\ 3 & -6 & -6 & 8 & 2 \end{bmatrix}$$

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

We are in REF. As you can see, the system is inconsistent because the third row corresponds to the equation $0 = 5$. There is no solution.

2.3 Linear Combinations, Span, and Linear Independence

Linear Combination

Let $m, n \in \mathbb{N}$. Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \in \mathbb{R}^m$. Then, suppose we have scalars $x_1, x_2, \ldots, x_n \in \mathbb{R}$. A **linear combination** of vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ is a vector $\mathbf{b} \in \mathbb{R}^m$ written as

$$\mathbf{b} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \sum_{k=1}^{n} x_k\mathbf{a}_k.$$

Example 2.3.1

An instrument has saved four points of temperature data in Celsius. Model the relationship between the Celsius and Fahrenheit temperature scales.

Solution

Celsius scale measures 0°C for freezing and 100°C for boiling. The Fahrenheit scale measures 32°F for freezing and 212°F for boiling.

The formula to convert from Celsius to Fahrenheit is:

$$f(c) = 32 + 1.8c$$

where c is the temperature in degrees Celsius and $f(c)$ is the corresponding Fahrenheit temperature.

We can convert the Celsius measurements into Fahrenheit at once using linear combinations. Suppose we have Celsius measurements c_1, c_2, c_3, c_4. The Fahrenheit outputs can be written as:

$$\begin{bmatrix} f(c_1) \\ f(c_2) \\ f(c_3) \\ f(c_4) \end{bmatrix} = 32 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + 1.8 \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix}$$

For example, if $c_1 = 0$, $c_2 = 10$, $c_3 = 20$, and $c_4 = 30$, we get

$$\begin{bmatrix} f(0) \\ f(10) \\ f(20) \\ f(30) \end{bmatrix} = 32 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + 1.8 \begin{bmatrix} 0 \\ 10 \\ 20 \\ 30 \end{bmatrix} = \begin{bmatrix} 32 \\ 50 \\ 68 \\ 86 \end{bmatrix}.$$

The Fahrenheit vector is given by a linear combination of two other vectors and two scalars.

Scalars are seen as weights. Let's visualize this given the vectors $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Every point you see plotted here is some combination of $a\mathbf{v}_1 + b\mathbf{v}_2$.

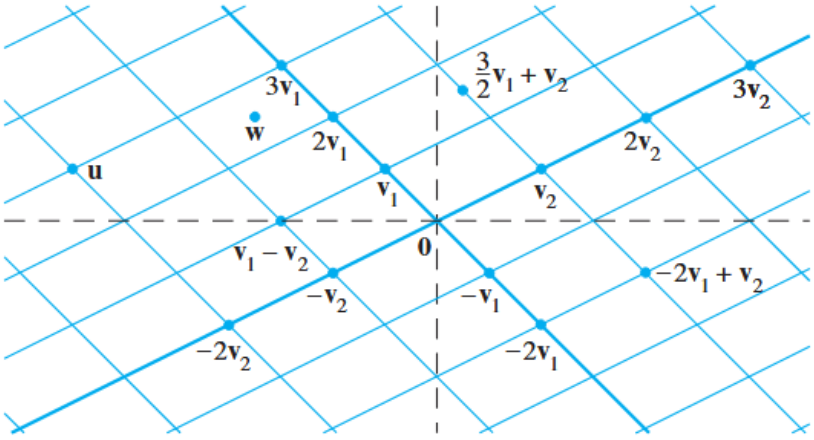


Image credit: Lay et al

Notice how there is a line that goes through $\mathbf{u}$ and $3\mathbf{v}_1$ and another line that goes through $\mathbf{u}$ and $-2\mathbf{v}_2$. Starting at the origin, you move 3 steps in the $\mathbf{v}_1$ direction and then 2 steps down in the $\mathbf{v}_2$ direction to get to $\mathbf{u}$. Indeed, this is the parallelogram rule! Thus, we can say that

$$\mathbf{u} = 3\mathbf{v}_1 - 2\mathbf{v}_2.$$

Let's make an estimate for $\mathbf{w}$. Notice how it is halfway between $2\mathbf{v}_1$ and $2\mathbf{v}_2$. Thus, we would say that it is at $\frac{5}{2}\mathbf{v}_1$.

Along $\mathbf{v}_2$, notice that it is halfway between $-\mathbf{v}_2$ and the origin. Thus, we will say it is at $-\frac{1}{2}\mathbf{v}_2$. Together, this means we have

$$\mathbf{w} = \frac{5}{2}\mathbf{v}_1 - \frac{1}{2}\mathbf{v}_2.$$

In other words, starting at the origin, we would move 2.5 steps in the $\mathbf{v}_1$ direction and half a step back in the $\mathbf{v}_2$ direction.

Example 2.3.2

Let

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}.$$

Determine whether $\mathbf{b}$ can be written as a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$. That is, determine whether scalars x_1 and x_2 exist such that

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b}.$$

If a solution exists, find x_1 and x_2.

Solution

The vector equation

$$x_1 \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} x_1 \\ -2x_1 \\ -5x_1 \end{bmatrix} + \begin{bmatrix} 2x_2 \\ 5x_2 \\ 6x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$$

can be expanded into a system of equations by matching corresponding entries:

$$\begin{cases} x_1 + 2x_2 = 7 \\ -2x_1 + 5x_2 = 4 \\ -5x_1 + 6x_2 = -3 \end{cases}$$

We now solve this system by row reducing the augmented matrix:

$$\begin{bmatrix} 1 & 2 & 7 \\ -2 & 5 & 4 \\ -5 & 6 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 7 \\ 0 & 9 & 18 \\ -5 & 6 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 7 \\ 0 & 9 & 18 \\ 0 & 16 & 32 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 7 \\ 0 & 1 & 2 \\ 0 & 16 & 32 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 7 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

From the second row, we have $x_2 = 2$. Let's substitute this into the first row to get $x_1 + 2(2) = 7 \Rightarrow x_1 = 3$.

Thus, $\mathbf{b}$ is a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$ with $x_1 = 3$ and $x_2 = 2$. This gives

$$3\begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix} + 2\begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}.$$

From this, we can conclude the following:

A vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$$

has the same solution set as the linear system represented by the augmented matrix:

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n & \mathbf{b} \end{bmatrix}$$

In particular, $\mathbf{b}$ can be written as a linear combination of $\mathbf{a}_1, \ldots, \mathbf{a}_n$ if and only if the corresponding linear system has a solution.

One prominent application of linear combinations is in manufacturing and production. Let's say we wanted to model the cost of producing a certain number of units of an item:

$$\left\{\begin{array}{c}\text{number}\\ \text{of units}\end{array}\right\} \cdot \left\{\begin{array}{c}\text{cost}\\ \text{per unit}\end{array}\right\} = \left\{\begin{array}{c}\text{total}\\ \text{cost}\end{array}\right\}$$

Example 2.3.3

A company manufactures two products. For \$1.00 worth of product B, the company spends \$0.45 on materials, \$0.25 on labor, and \$0.15 on overhead. For \$1.00 worth of product C, the company spends \$0.40 on materials, \$0.30 on labor, and \$0.15 on overhead.

Let

$$\mathbf{b} = \begin{bmatrix}0.45\\0.25\\0.15\end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix}0.40\\0.30\\0.15\end{bmatrix}.$$

Then $\mathbf{b}$ and $\mathbf{c}$ represent the "costs per dollar of income" for the two products.

(a) What economic interpretation can be given to the vector $100\mathbf{b}$?

(b) Suppose the company wishes to manufacture x_1 dollars worth of product B and x_2 dollars worth of product C. Give a vector that describes the various costs the company will have for materials, labor, and overhead.

Solution

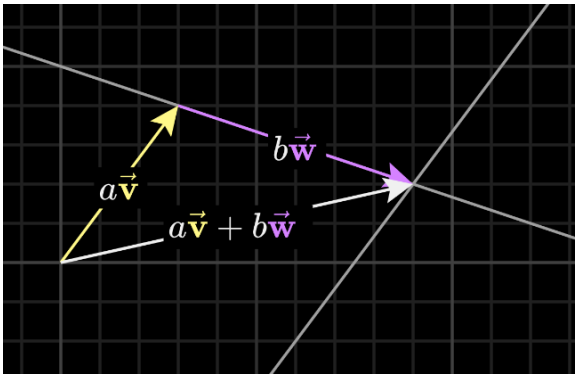
(a) First, let's compute:

$$100\mathbf{b} = 100\begin{bmatrix}0.45\\0.25\\0.15\end{bmatrix} = \begin{bmatrix}45\\25\\15\end{bmatrix}$$

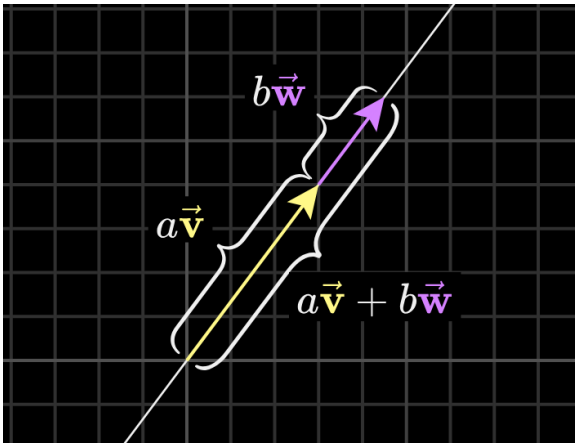
The vector $100\mathbf{b}$ lists the various costs for producing \$100 worth of product B. We have \$45 for materials, \$25 for labor, and \$15 for overhead.

(b) The costs of manufacturing x_1 dollars worth of product B are given by the vector $x_1\mathbf{b}$, and the costs of manufacturing x_2 dollars worth of product C are given by the vector $x_2\mathbf{c}$. Thus, we add them up to get the total costs for both products:

$$x_1\mathbf{b} + x_2\mathbf{c}$$



The linear combination of $\mathbf{v}$ and $\mathbf{w}$ where a and b are scalars. Image credit: 3Blue1Brown



The span of the vector. Image credit: 3Blue1Brown

Suppose we wanted to study the set of all vectors that could be written as a linear combination of a fixed set of vectors.

Span of a Set of Vectors

Let $m, n \in \mathbb{N}$ and let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \in \mathbb{R}^m$. The **span** of the set $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ is the set of all possible linear combinations of these vectors. That is,

$$\text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\} = \left\{ \sum_{k=1}^{n} x_k \mathbf{a}_k \ : \ x_k \in \mathbb{R} \text{ for each } k = 1, 2, \ldots, n \right\}$$

In other words, $\text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is the collection of all vectors that can be written as

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n$$

where $c_1, c_2, \ldots, c_n$ are scalars.

Sometimes, we use angle brackets to denote the span:

$$\langle \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \rangle = \text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \text{Span}\left\{\mathbf{a}_k\right\}_{k=1}^{n}$$

Suppose we have the span of a single vector written as $\text{Span}\{\mathbf{v}\}$ where $\mathbf{v} \in \mathbb{R}^3$. This represents the set of all scalar multiples of $\mathbf{v}$. Geometrically, this is the set of all points in $\mathbb{R}^3$ that pass through $\mathbf{v}$ and the origin, forming a line.

Example 2.3.4

We are given the following vector:

$$\mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

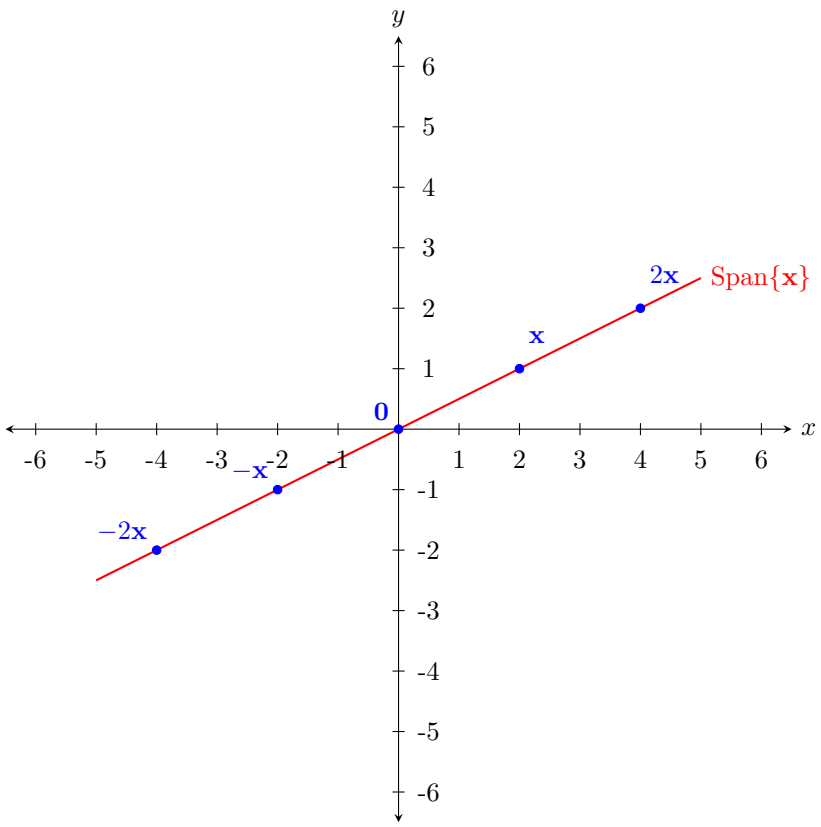
Find the span of the vector and graph it.

Solution

The span of this vector is this set

$$\langle \mathbf{x} \rangle = \{\alpha \mathbf{x} : \alpha \in \mathbb{R}\} = \left\{ \begin{bmatrix} 2\alpha \\ \alpha \end{bmatrix} : \alpha \in \mathbb{R} \right\}$$

Geometrically, this is the line through the origin with slope 0.5, given by the equation $y = 0.5x$:



Example 2.3.5

Consider three vectors in $\mathbb{R}^3$:

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 0 \\ 0 \\ -1 \\ -1 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Generate the span.

Solution

We are looking for $\operatorname{Span}\{\mathbf{a}_k\}_{k=1}^3$. In other words,

$$\left\{\mathbf{b} \in \mathbb{R}^{4\times 1} : \mathbf{b} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 \text{ where } x_k \in \mathbb{R} \text{ for all } k\right\}$$

If $\mathbf{b} \in \operatorname{Span}\{\mathbf{a}_k\}$, then

$$\begin{aligned} \mathbf{b} = \sum_{k=1}^{3} x_k\mathbf{a}_k &= x_1 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 0 \\ -1 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} x_1 \\ x_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ -x_2 \\ -x_2 \end{bmatrix} + \begin{bmatrix} x_3 \\ x_3 \\ x_3 \\ x_3 \end{bmatrix} \\ &= \begin{bmatrix} x_1 + x_3 \\ x_1 + x_3 \\ x_3 - x_2 \\ x_3 - x_2 \end{bmatrix}. \end{aligned}$$

Therefore, we can conclude that all vectors in $\text{Span}\{\mathbf{a}_k\}_{k=1}^3$ must be of the form

$$\mathbf{b} = \begin{bmatrix} \alpha \\ \alpha \\ \beta \\ \beta \end{bmatrix}$$

for some $\alpha, \beta \in \mathbb{R}$. Our solution, the set of all vectors in the span, is

$$\left\{ \begin{bmatrix} \alpha \\ \alpha \\ \beta \\ \beta \end{bmatrix} : \alpha, \beta \in \mathbb{R} \right\}$$

Here are some vectors that could be in the span:

$$\begin{bmatrix} 2 \\ 2 \\ -4 \\ -4 \end{bmatrix}, \begin{bmatrix} 5 \\ 5 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Thus, we conclude that $\text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\} = \text{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$. $\mathbf{a}_3$ was redundant with respect to the linear combination; it adds no new information to the span. Geometrically, this means it doesn't add another dimension. $\mathbf{a}_3$ is a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$.

Example 2.3.6

Suppose we are given the following vectors:

$$\mathbf{a}_1 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

Find the span and graph it.

Solution

The span is

$$\langle \mathbf{a}_1, \mathbf{a}_2 \rangle = \left\{ x_1 \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} : x_1, x_2 \in \mathbb{R} \right\}.$$

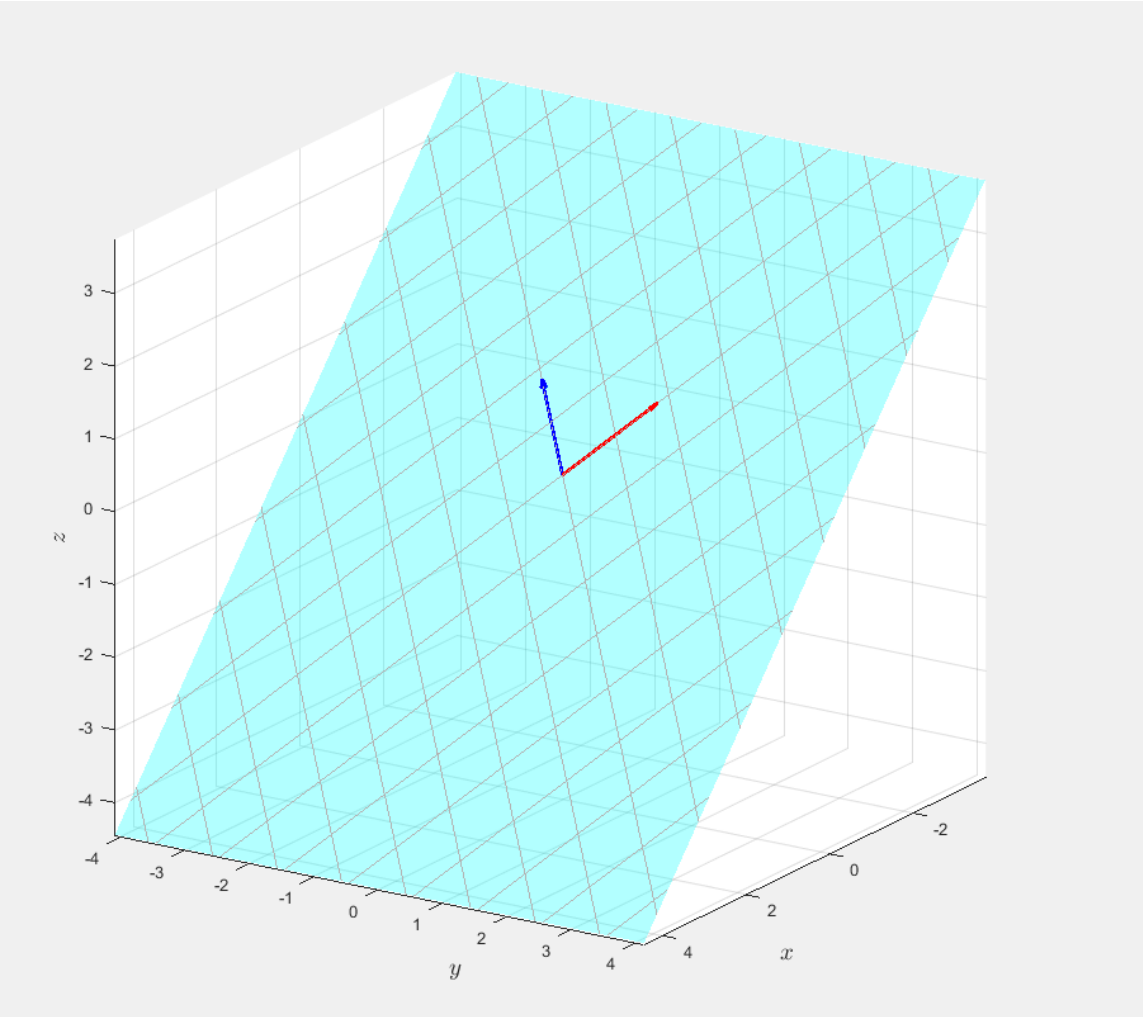
This set of all linear combinations of $\mathbf{a}_1$ and $\mathbf{a}_2$ forms a plane in $\mathbb{R}^3$ passing through the origin. Every point on this plane can be expressed as a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$. For example, the vector:

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

lies in the span, since

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} + \left(-\frac{1}{2} \right) \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}.$$

The graph below illustrates the span of $\mathbf{a}_1$ and $\mathbf{a}_2$ as a light blue plane in $\mathbb{R}^3$, with $\mathbf{a}_1$ shown in red and $\mathbf{a}_2$ in blue.



Before we move on, let's summarize. A linear combination is what you get when you multiply vectors by some scalar and then add them together. The span of a set of vectors is all of the points you can reach by making linear combinations of those vectors.

Problem 2.3.1

Consider the following vectors:

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 5 \\ -13 \\ -3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -3 \\ 8 \\ 1 \end{bmatrix}$$

Given that $\text{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$ is a plane through the origin in $\mathbb{R}^3$, is $\mathbf{b} \in \text{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$?

Solution

To answer this, we are asking "Can we find a solution to $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b}$?" In other words, we are trying to solve

$$x_1 \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 5 \\ -13 \\ -3 \end{bmatrix} = \begin{bmatrix} -3 \\ 8 \\ 1 \end{bmatrix}.$$

This can be written as a linear system:

$$\begin{cases} x_1 + 5x_2 = -3 \\ -2x_1 - 13x_2 = 8 \\ 3x_1 - 3x_2 = 1 \end{cases} \iff \begin{bmatrix} 1 & 5 & -3 \\ -2 & -13 & 8 \\ 3 & -3 & 1 \end{bmatrix}$$

Let's row reduce:

$$\begin{bmatrix} 1 & 5 & -3 \\ -2 & -13 & 8 \\ 3 & -3 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & -3 \\ 0 & -3 & 2 \\ 0 & -18 & 10 \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & -3 \\ 0 & -3 & 2 \\ 0 & 0 & -2 \end{bmatrix}$$

The last row says $0 = -2$, which is a contradiction. Thus, the system has no solution and $\mathbf{b} \notin \text{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$.

Linearly Dependent Set of Vectors

Let $m, n \in \mathbb{N}$. Suppose $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \in \mathbb{R}^m$. We say that the set of vectors $\{\mathbf{a}_k\}_{k=1}^n$ is **linearly dependent** if and only if one of the vectors can be written as a linear combination of the other vectors.

In other words, the set $\{\mathbf{a}_k\}_{k=1}^n$ is linearly dependent if there exists a $j \in \mathbb{N}$ with $1 \leq j \leq n$ and

$$\mathbf{a}_j = \sum_{\substack{k=1 \\ k \neq j}}^{n} x_k \mathbf{a}_k$$

for some choice of scalars $x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n \in \mathbb{R}$.

Example 2.3.7

Consider the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Determine whether this set of vectors is linearly independent or not.

Solution

This set of vectors is linearly dependent because

$$\mathbf{v}_1 = 0 \cdot \mathbf{v}_2 + 0 \cdot \mathbf{v}_3.$$

In other words, we can express $\mathbf{v}_1$ as a linear combination of $\mathbf{v}_2$ and $\mathbf{v}_3$. In fact, any set of vectors that includes the zero vector will always be linearly dependent. For a linearly dependent set of vectors, you can create the zero vector by taking nonzero linear combinations.

Test for Linear Dependence

Let $m, n \in \mathbb{N}$. Suppose that $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \in \mathbb{R}^m$. These vectors are **linearly dependent** if and only if there exist scalars $x_1, x_2, \ldots, x_n$ such that

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{0},$$

where one of the scalars must *not* be zero.

Example 2.3.8

Consider the vectors

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} -3 \\ 9 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 4 \\ 9 \end{bmatrix}.$$

Determine whether this set of vectors is linearly independent or not.

Solution

This set of vectors is linearly dependent. We see this immediately because

$$\mathbf{a}_1 = -\frac{1}{3} \cdot \mathbf{a}_2.$$

Whenever you try to cram more vectors into a space than the space's dimension allows, dependence must occur. This leads to a theorem:

Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a set of vectors in $\mathbb{R}^m$. If $n > m$, then the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly dependent.

That is, given a set of vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \in \mathbb{R}^m$, we can define the matrix

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}.$$

The span of these vectors is the column space of A, which is the set of all linear combinations of $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$.

A set of vectors is linearly dependent if at least one vector can be expressed as a linear combination of the others. This means that one of its columns can be removed without changing the column space. The span remains the same because the removed vector was redundant.

Linearly Independent Set of Vectors

Let $m, n \in \mathbb{N}$. Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ and $x_1, x_2, \ldots, x_n \in \mathbb{R}$. The set $\{\mathbf{a}_k\}_{k=1}^n$ of vectors is **linearly independent** if it is not linearly dependent.

A linearly independent set of vectors has the property that no vector in the set can be written as a linear combination of the other vectors in the set. Each vector in the set brings unique information to the set.

Test for Linear Independence

Let $m, n \in \mathbb{N}$. Suppose that $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \in \mathbb{R}^m$. The set $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ is **linearly independent** if and only if the vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{0}$$

is true when $x_k = 0$ for all $k = 1, 2, \ldots, n$.

Problem 2.3.2

Determine if the following set of vectors is linearly independent.

(a) $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$

Determine if the following set of vectors is linearly independent.

(b) $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$

Then, graph both.

Solution

(a) Notice that $\mathbf{v}_2$ is a multiple of $\mathbf{v}_1$, specifically $\mathbf{v}_2 = 2\mathbf{v}_1$. Hence, $-2\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{0}$. This shows that the set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly dependent.

(b) The vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are not scalar multiples of each other. To determine whether they are linearly dependent, suppose there exist scalars c and d such that

$$c\mathbf{v}_1 + d\mathbf{v}_2 = \mathbf{0}.$$

Then, we solve

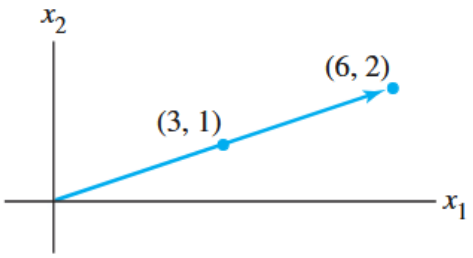
$$c\begin{bmatrix}3\\2\end{bmatrix} + d\begin{bmatrix}6\\2\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}.$$

This yields the system

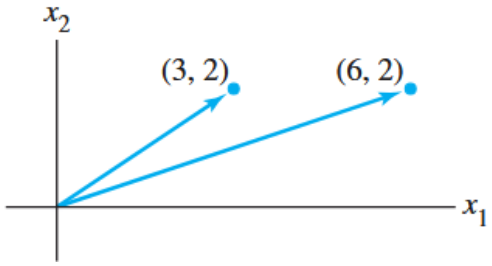
$$3c + 6d = 02c + 2d = 0$$

Solve the second equation yields $c = -d$. Substituting this into the first equation yields $3(-d)+6d = -3d + 6d = 3d = 0 \Rightarrow d = 0$. Thus, $c = 0$ as well. Thus, $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a linearly independent set.

Let's graph:



(a). Image credit: Lay et al



(b). Image credit: Lay et al

Problem 2.3.3

Which of the following sets of vectors is linearly dependent?

(a) $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$

(b) $\left\{ \begin{bmatrix} 1 \\ 4 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}$

(c) $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}$

(d) $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ -4 \\ -4 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}$

(e) $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \right\}$

Solution

We need to find scalars x_1, x_2, x_3 where at least one is nonzero such that a linear combination of the set results in the zero vector.

(a)

$$x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The only solution is $x_1 = x_2 = x_3 = 0$. Thus this set is linearly independent.

(b)

$$x_1 \begin{bmatrix} 1 \\ 4 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ -2 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The only solution is $x_1 = x_2 = x_3 = 0$. Thus this set is linearly independent.

(c)

$$x_1 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The only solution is $x_1 = x_2 = x_3 = 0$. Thus this set is linearly independent.

(d)

$$x_1 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 3 \\ -4 \\ -4 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Notice how

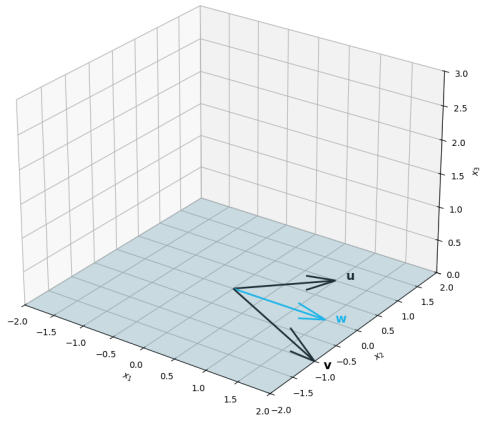
$$3 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} - 4 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ -4 \\ -4 \end{bmatrix}.$$

We have $\mathbf{v}_2 = 3\mathbf{v}_1 - 4\mathbf{v}_3 \Rightarrow \mathbf{v}_2 - 3\mathbf{v}_1 + 4\mathbf{v}_3 = \mathbf{0}$. This means $x_1 = -3$, $x_2 = 1$, and $x_3 = 4$. Thus this set is linearly dependent and is the answer.

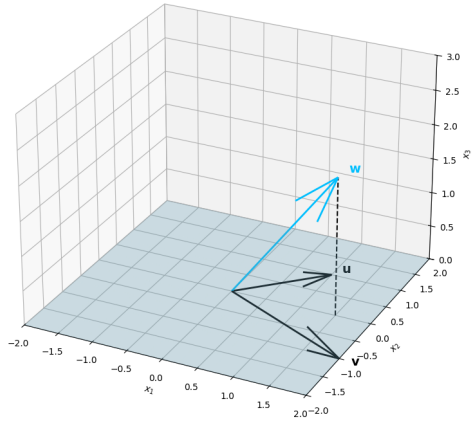
(e)

$$x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The only solution is $x_1 = x_2 = x_3 = 0$. Thus this set is linearly independent.



Linearly dependent, $\mathbf{w} \in \text{Span}\{\mathbf{u}, \mathbf{v}\}$



Linearly independent, $\mathbf{w} \notin \text{Span}\{\mathbf{u}, \mathbf{v}\}$

2.4 The Matrix Equation $A\mathbf{x} = \mathbf{b}$

There are two ways to do **matrix-vector multiplication**. The first is matrix-column-vector-multiplication which means multiplying a matrix $A \in \mathbb{R}^{m \times n}$ by a column vector $\mathbf{x} \in \mathbb{R}^{n \times 1}$ to get a column vector:

$$\mathbf{A}\mathbf{x} = \mathbf{b} \in \mathbb{R}^{m \times 1}.$$

The second is called matrix-row-vector-multiplication which means multiplying a matrix $A \in \mathbb{R}^{m \times n}$ by a row vector $\mathbf{x}^T$ where $\mathbf{x} \in \mathbb{R}^{m \times 1}$ to get a row vector:

$$\mathbf{x}^T\mathbf{A} = \mathbf{b} \in \mathbb{R}^{1 \times n}.$$

We can view a linear combination of vectors as the product of a matrix and a vector:

Matrix-Vector Multiplication

Let $A \in \mathbb{R}^{m \times n}$ be a matrix with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and let $\mathbf{x} \in \mathbb{R}^n$. Then the product of A and $\mathbf{x}$, denoted by $A\mathbf{x}$, is the **linear combination** of the columns of A using the corresponding entries in $\mathbf{x}$ as weights. That is,

$$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n.$$

Example 2.4.1

For $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in \mathbb{R}^m$, write the linear combination $3\mathbf{v}_1 - 5\mathbf{v}_2 + 7\mathbf{v}_3$ using matrix-vector multiplication.

Solution

Place $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ into the columns of a matrix A, and place the weights 3, -5, and 7 into a vector $\mathbf{x}$. That is,

$$3\mathbf{v}_1 - 5\mathbf{v}_2 + 7\mathbf{v}_3 = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} \begin{bmatrix} 3 \\ -5 \\ 7 \end{bmatrix} = A\mathbf{x}.$$

You already know how to write a system of linear equations as a vector equation involving a linear combination of vectors. For example, the system

$$\begin{cases} x_1 + 2x_2 - x_3 = 4 \\ -5x_2 + 3x_3 = 1 \end{cases}$$

is equivalent to

$$x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ -5 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}.$$

As in the earlier example, the linear combination on the left side is a matrix times a vector, so the equation becomes

$$\begin{bmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}.$$

This is called a **matrix equation**.

Let $\mathbf{A}$ be an $m \times n$ matrix with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$. If $\mathbf{b} \in \mathbb{R}^m$, the matrix equation

$$A\mathbf{x} = \mathbf{b}$$

has the same solution set as the vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}.$$

This, in turn, has the same solution set as the system of linear equations whose augmented matrix is

$$[\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n \quad \mathbf{b}]$$

We can write a system of linear equations in three different ways:

1. As a system of linear equations
2. As a vector equation (a linear combination)
3. As a matrix equation

These three forms are equivalent and all solved using row reduction.

The equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is a linear combination of the columns of A.

Example 2.4.2

Let

$$A = \begin{bmatrix} 1 & 3 & 4 \\ -4 & -2 & -6 \\ -3 & -2 & -7 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

Is the equation $A\mathbf{x} = \mathbf{b}$ consistent for all possible b_1, b_2, b_3?

Solution

Row reduce the augmented matrix for $A\mathbf{x} = \mathbf{b}$:

$$\begin{bmatrix} 1 & 3 & 4 & b_1 \\ -4 & -2 & -6 & b_2 \\ -3 & -2 & -7 & b_3 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 4 & b_1 \\ 0 & 10 & 10 & b_2 + 4b_1 \\ 0 & 7 & 5 & b_3 + 3b_1 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 4 & b_1 \\ 0 & 14 & 10 & b_2 + 4b_1 \\ 0 & 0 & 0 & b_3 + 3b_1 - \frac{1}{2}(b_2 + 4b_1) \end{bmatrix}$$

The third row becomes

$$b_1 - \frac{1}{2}b_2 + b_3 = 0.$$

Thus, $A\mathbf{x} = \mathbf{b}$ is not consistent for every $\mathbf{b}$, because some choices of $\mathbf{b}$ can make this expression nonzero. This is the equation of a plane through the origin in $\mathbb{R}^3$ that is the set of all linear combinations of the three columns of A.

When we say that "The columns of A span $\mathbb{R}^m$", we are saying that every vector $\mathbf{b} \in \mathbb{R}^m$ can be built as a linear combination of the columns of A. Thus, generally, if you have a set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\} \in \mathbb{R}^m$ where every vector in $\mathbb{R}^m$ can be expressed as a combination of these vectors, we say they span $\mathbb{R}^m$. That is, $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\} = \mathbb{R}^m$.

Let A be an $m \times n$ coefficient matrix. Then the following statements are logically equivalent. That is, for a particular A, either they are all true statements or they are all false.

(a) For each $\mathbf{b} \in \mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution.

(b) Each $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of A.

(c) The columns of A span $\mathbb{R}^m$.

(d) A has a pivot position in every row.

In other words, if the columns of A span $\mathbb{R}^m$, then every $\mathbf{b}$ can be reached as a linear combination, so $A\mathbf{x} = \mathbf{b}$ has a solution. Then, since A has a pivot in every row, you can pick any $\mathbf{b}$ because there's no row of zeros on the left that could create an inconsistency.

And keep in mind that this only talks about the coefficient matrix of A. An augmented matrix $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$ having a pivot in every row doesn't guarantee that $A\mathbf{x} = \mathbf{b}$ may or may not be consistent.

Example 2.4.3

Compute $A\mathbf{x}$, where

$$A = \begin{bmatrix} 2 & 3 & 4 \\ -1 & 5 & -3 \\ 6 & -2 & 8 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Solution

By definition, multiplying a matrix by a vector is a linear combination of the matrix's columns, using entries of $\mathbf{x}$ as weights:

$$A\mathbf{x} = x_1 \begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 5 \\ -2 \end{bmatrix} + x_3 \begin{bmatrix} 4 \\ -3 \\ 8 \end{bmatrix}$$

Compute the sum:

$$= \begin{bmatrix} 2x_1 + 3x_2 + 4x_3 \\ -x_1 + 5x_2 - 3x_3 \\ 6x_1 - 2x_2 + 8x_3 \end{bmatrix}$$

Alternatively, rather than computing the combination column-by-column, we can think of $A\mathbf{x}$ row-wise:

The first entry of $A\mathbf{x}$ comes from the dot product of the first row of A with $\mathbf{x}$:

$$\begin{bmatrix} 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 2x_1 + 3x_2 + 4x_3$$

Similarly, the second entry is

$$\begin{bmatrix} -1 & 5 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = -x_1 + 5x_2 - 3x_3.$$

And the third entry is

$$\begin{bmatrix} 6 & -2 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 6x_1 - 2x_2 + 8x_3.$$

If the product $A\mathbf{x}$ is defined, then the ith entry in $A\mathbf{x}$ is the sum of the products of corresponding entries from row i of A and from the vector $\mathbf{x}$.

Example 2.4.4

Compute the product:

$$A\mathbf{x} = \begin{bmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \\ 7 \end{bmatrix}$$

Solution

Compute entry-wise products (dot products of rows with $\mathbf{x}$):

$$= \begin{bmatrix} 1 \cdot 4 + 2 \cdot 3 + (-1) \cdot 7 \\ 0 \cdot 4 + (-5) \cdot 3 + 3 \cdot 7 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

Properties of Matrix-Vector Multiplication

If A is an $m \times n$ matrix, $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^n$, and c is a scalar, then

(a) $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$

(b) $A(c\mathbf{u}) = c(A\mathbf{u})$

Let's explain why this is true using a simple example where A has 3 columns. Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix}$ and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$.

Proof of Property (a) of Matrix-Vector Multiplication

We will compute $A(\mathbf{u} + \mathbf{v})$ as a linear combination of the columns of A using the entries in $\mathbf{u} + \mathbf{v}$ as weights:

$$A(\mathbf{u} + \mathbf{v}) = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ u_3 + v_3 \end{bmatrix}$$

$$= (u_1 + v_1)\mathbf{a}_1 + (u_2 + v_2)\mathbf{a}_2 + (u_3 + v_3)\mathbf{a}_3$$

$$= (u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + u_3\mathbf{a}_3) + (v_1\mathbf{a}_1 + v_2\mathbf{a}_2 + v_3\mathbf{a}_3)$$
$$= A\mathbf{u} + A\mathbf{v}$$

□

Proof of Property (b) of Matrix-Vector Multiplication

We will compute $A(c\mathbf{u})$ as a linear combination of the columns of A using the entries in $c\mathbf{u}$ as weights:

$$A(c\mathbf{u}) = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3\end{bmatrix}\begin{bmatrix}cu_1\\ cu_2\\ cu_3\end{bmatrix}$$
$$= cu_1\mathbf{a}_1 + cu_2\mathbf{a}_2 + cu_3\mathbf{a}_3$$
$$= c(u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + u_3\mathbf{a}_3)$$
$$= c(A\mathbf{u}).$$

□

Problem 2.4.1

Consider the system of equations:

$$\begin{cases} 5x_1 - x_2 = 10 \\ 2x_1 + 6x_2 = 3 \\ 10x_1 - x_2 = 1 \end{cases}$$

Write them out as (a) a linear combination of column vectors and (b) a matrix equation.

Solution

(a)

$$x_1 \begin{bmatrix} 5 \\ 2 \\ 10 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ 6 \\ -1 \end{bmatrix} = \begin{bmatrix} 10 \\ 3 \\ 1 \end{bmatrix}$$

(b)

$$\begin{bmatrix} 5 & -1 \\ 2 & 6 \\ 10 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 10 \\ 3 \\ 1 \end{bmatrix}$$

Problem 2.4.2

Suppose A is a 3×3 matrix and $\mathbf{b}$ is a vector in $\mathbb{R}^3$ with the property that $A\mathbf{x} = \mathbf{b}$ has a unique solution. Explain why the columns of A must span $\mathbb{R}^3$.

Solution

If $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b} \in \mathbb{R}^3$, it means that the columns of A are linearly independent and span the entire space $\mathbb{R}^3$.

Since A is a 3×3 matrix, it has 3 columns in $\mathbb{R}^3$. For the equation $A\mathbf{x} = \mathbf{b}$ to have a unique solution, there must be a pivot in every column to avoid free variables and a pivot in every row to guarantee consistency. Having a pivot in every row ensures that every $\mathbf{b}$ is a linear combination of A's columns, meaning the columns of A span $\mathbb{R}^3$.

Thus, the fact that $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$ implies that the columns of A must span $\mathbb{R}^3$.

Problem 2.4.3

Determine whether the columns of the matrix

$$A = \begin{bmatrix} 7 & 2 & -5 & 8 \\ -5 & -3 & 4 & -9 \\ 6 & 10 & -2 & 7 \\ -7 & 9 & 2 & 15 \end{bmatrix}$$

span $\mathbb{R}^4$. In other words, determine if $A\mathbf{x} = \mathbf{b}$ has a solution for every vector $\mathbf{b} \in \mathbb{R}^4$.

Solution

We will row reduce A to see if there is a pivot in every row:

$$\begin{bmatrix} 7 & 2 & -5 & 8 \\ -5 & -3 & 4 & -9 \\ 6 & 10 & -2 & 7 \\ -7 & 9 & 2 & 15 \end{bmatrix} \sim \begin{bmatrix} 1 & \frac{2}{7} & -\frac{5}{7} & \frac{8}{7} \\ -5 & -3 & 4 & -9 \\ 6 & 10 & -2 & 7 \\ -7 & 9 & 2 & 15 \end{bmatrix} \sim \begin{bmatrix} 1 & \frac{2}{7} & -\frac{5}{7} & \frac{8}{7} \\ 0 & -\frac{11}{7} & \frac{5}{7} & -\frac{23}{7} \\ 0 & \frac{62}{7} & \frac{8}{7} & \frac{1}{7} \\ 0 & \frac{83}{7} & -\frac{33}{7} & \frac{71}{7} \end{bmatrix} \sim \begin{bmatrix} 1 & \frac{2}{7} & -\frac{5}{7} & \frac{8}{7} \\ 0 & 1 & -\frac{5}{11} & \frac{23}{11} \\ 0 & \frac{62}{7} & \frac{8}{7} & \frac{1}{7} \\ 0 & \frac{83}{7} & -\frac{33}{7} & \frac{71}{7} \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & \frac{2}{7} & -\frac{5}{7} & \frac{8}{7} \\ 0 & 1 & -\frac{5}{11} & \frac{23}{11} \\ 0 & 0 & 1 & -\frac{189}{50} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The last row is a row of zeros, which means there is no pivot in the fourth row. Thus, the columns of A do not span $\mathbb{R}^4$.

2.5 Classifying Solutions

Let's learn how to classify solutions of linear systems.

A system of linear equations is said to be **homogeneous** if it can be written as

$$A\mathbf{x} = \mathbf{0}$$

where A is an $m \times n$ matrix. Such a system always has at least one solution, which is $\mathbf{0} \in \mathbb{R}^m$. This is called the *trivial* solution. For a given equation $A\mathbf{x} = \mathbf{0}$, we are tasked with determining whether or not there is a *nontrivial* solution.

> The homogeneous equation $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution if and only if the equation has at least one free variable.

Example 2.5.1

Determine if the following homogeneous system has a nontrivial solution. Then describe the solution set.

$$\begin{cases} 3x_1 + 5x_2 - 4x_3 = 0 \\ -3x_1 - 2x_2 + 4x_3 = 0 \\ 6x_1 + x_2 - 8x_3 = 0 \end{cases}$$

Solution

Let A be the coefficient matrix. We now row reduce the augmented matrix $[A\ \mathbf{0}]$:

$$\begin{bmatrix} 3 & 5 & -4 & 0 \\ -3 & -2 & 4 & 0 \\ 6 & 1 & -8 & 0 \end{bmatrix} \sim \begin{bmatrix} 3 & 5 & -4 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & -9 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 3 & 5 & -4 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We are now in REF. We see that x_3 is a free variable because column 3 lacks a pivot.

We continue to RREF:

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

Turning this into a system of linear equations and then solving yields $x_1 = \frac{4}{3}x_3$ and $x_2 = 0$. x_3 is free.

We now express the general solution of the matrix equation as vector equation:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{4}{3}x_3 \\ 0 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix} = x_3\mathbf{v}$$

Every solution is a scalar multiple of $\mathbf{v} = \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix}$. Geometrically, the solution set is a line through the origin in $\mathbb{R}^3$. A nontrivial solution vector $\mathbf{x}$, found by setting $\mathbf{x}_3 = 0$, can have some zero entries, as long as not all entries are zero.

Example 2.5.2

Describe all solutions of the following homogeneous equation:

$$10x_1 - 3x_2 - 2x_3 = 0$$

Solution

This is a single linear equation, but we can treat it like a very simple system of linear equations. This equation has two free variables, x_2 and x_3. Solving yields $x_1 = 0.3x_2 + 0.2x_3$. Thus, the general solution is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.3x_2 + 0.2x_3 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.3x_2 \\ x_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.2x_3 \\ 0 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} 0.3 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0.2 \\ 0 \\ 1 \end{bmatrix}.$$

The solution set is the Span$\{\mathbf{u}, \mathbf{v}\}$, where

$$\mathbf{u} = \begin{bmatrix} 0.3 \\ 1 \\ 0 \end{bmatrix}, \ \mathbf{v} = \begin{bmatrix} 0.2 \\ 0 \\ 1 \end{bmatrix}$$

Every solution is a linear combination of $\mathbf{u}$ and $\mathbf{v}$. Neither $\mathbf{u}$ nor $\mathbf{v}$ is a scalar multiple of the other, confirming the solution set is a plane through the origin in $\mathbb{R}^3$.

Every time you solve a homogeneous equation $\mathbf{Ax} = \mathbf{0}$, the solution set can always be written as a span of vectors Span$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$. If the only solution is the trivial solution, then the solution set is a *single point* at the origin Span$\{\mathbf{0}\}$. If there's only one free variable, the solution set is a *line*

through the origin. If there's two or more free variables, the solution set becomes a *plane* through the origin.

When you have an equation like $10x_1 - 3x_2 - 2x_3 = 0$ from **Example 2.5.2**, it's called an *implicit* equation. If you solve it and express the solution as combinations of vectors, you get an *explicit* description of the plane as the set spanned by $\mathbf{u}$ and $\mathbf{v}$:

$$\mathbf{x} = s\mathbf{u} + t\mathbf{v}$$

This is called a *parametric* vector equation.

Example 2.5.3

Describe all solutions of $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 3 & 5 & -4 \\ -3 & -2 & 4 \\ 6 & 1 & -8 \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} 7 \\ -1 \\ -4 \end{bmatrix}.$$

Solution

Row reduce the augmented matrix $[A \ \mathbf{b}]$:

$$\begin{bmatrix} 3 & 5 & -4 & 7 \\ -3 & -2 & 4 & -1 \\ 6 & 1 & -8 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -\frac{4}{3} & -1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We have

$$\begin{cases} x_1 - \frac{4}{3}x_3 = -1 \\ x_2 = 2 \\ 0 = 0 \end{cases}$$

Thus $x_1 = -1 + \frac{4}{3}x_3$, $x_2 = 2$, and x_3 is free. The general solution vector is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 + \frac{4}{3}x_3 \\ 2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{4}{3}x_3 \\ 0 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix},$$

where

$$\mathbf{p} = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \ \mathbf{v} = \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix}.$$

Then the solution set is $\mathbf{x} = \mathbf{p} + x_3\mathbf{v}$, or generally,

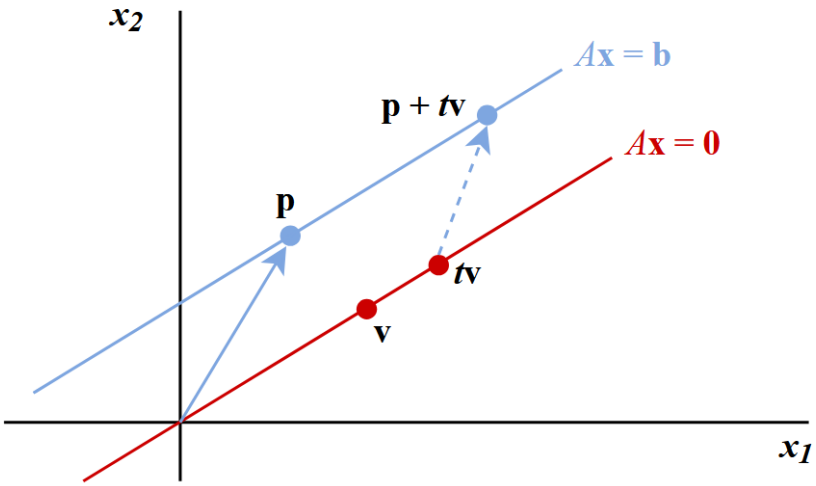
$$\mathbf{x} = \mathbf{p} + t\mathbf{v},$$

where $t \in \mathbb{R}$.

The vector $\mathbf{p}$ is one particular solution to $A\mathbf{x} = \mathbf{b}$. The vector $\mathbf{v}$ is a solution to the corresponding homogeneous equation $A\mathbf{x} = \mathbf{0}$. The general solution set is a line through $\mathbf{p}$, parallel to the line defined by $\mathbf{v}$.

All solutions of $A\mathbf{x} = \mathbf{b}$ are obtained by adding $\mathbf{p}$ to the solution set of $A\mathbf{x} = \mathbf{0}$. The equation $A\mathbf{x} = \mathbf{0}$ gives you a line through the origin, which is the set of all vectors that satisfy $\mathbf{b} = \mathbf{0}$. On the other hand, $A\mathbf{x} = \mathbf{b}$ gives you a solution set that is parallel to the solution set of $A\mathbf{x} = \mathbf{0}$ but shifted away from the origin. This is because you take the line through the origin and translate it by a solution vector $\mathbf{p}$.

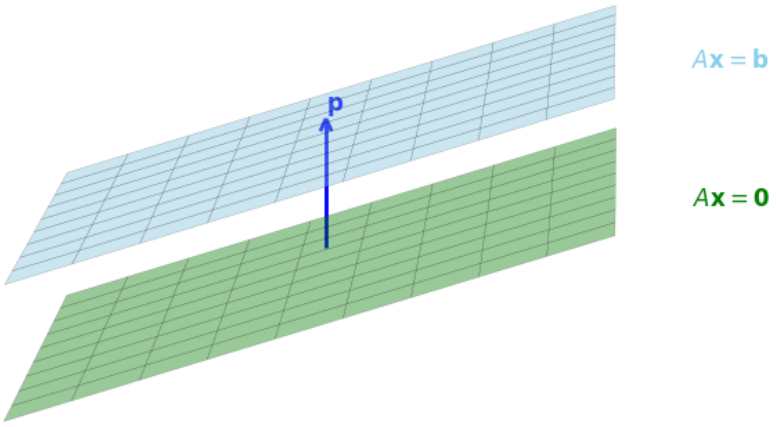
Take a look:



We can generalize this to any consistent matrix equation:

> Suppose the equation $A\mathbf{x} = \mathbf{b}$ is consistent for some given $\mathbf{b}$, and let $\mathbf{p}$ be a solution. Then the solution set of $A\mathbf{x} = \mathbf{b}$ is the set of all vectors of the form $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$ where $\mathbf{v}_h$ is any solution of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

Here is a visualization in three dimensions:



Parallel solution sets of $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{0}$.

Remember that this only applies to consistent matrix equations with at least one nonzero solution $\mathbf{b}$. If there's no solution, then the solution set is empty.

Lastly, let's outline the algorithm:

Writing a Solution Set in Parametric Vector Form

1. Row reduce the augmented matrix to RREF.
2. Express each basic variable in terms of the free variables appearing in an equation.
3. Write a general solution vector $\mathbf{x}$ where entries depend on the free variables, if any.
4. Decompose $\mathbf{x}$ as a linear combination of vectors by using the free variables as parameters. That is, factor them out.

Problem 2.5.1

Solve the system and express the solution set in parametric vector form:

$$\begin{cases} x_1 + 3x_2 + x_3 = 1 \\ -4x_1 - 9x_2 + 2x_3 = -1 \\ -3x_2 - 6x_3 = -3 \end{cases}$$

Solution

Let's row reduce to RREF:

$$\begin{bmatrix} 1 & 3 & 1 & 1 \\ -4 & -9 & 2 & -1 \\ 0 & -3 & -6 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 1 & 1 \\ 0 & 3 & 6 & 3 \\ 0 & -3 & -6 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 1 & 1 \\ 0 & 3 & 6 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 1 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 0 & -5 & -2 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We now have

$$\begin{cases} x_1 - 5x_3 = -2 \\ x_2 + 2x_3 = 1 \\ 0 = 0 \end{cases}.$$

Solving this yields $x_1 = -2 + 5x_3$, $x_2 = 1 - 2x_3$, and x_3 is free.

The solution vector is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2 + 5x_3 \\ 1 - 2x_3 \\ x_3 \end{bmatrix}.$$

Factor out x_3 to get it it in parametric vector form:

$$\mathbf{x} = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}$$

2.6 Applications of Linear Systems

Let's now apply what we've learned so far to some applications. Wassily Leontief was a Nobel prize-winning economist known for his work in modeling how different sectors of an economy depend on each other's outputs as their inputs. We will study this more as we go on. For now, let's come up with a simple economic model.

Example 2.6.1

A country's economy is divided into multiple sectors. Let's focus on the coal, electric, and steel sectors, represented through the following table:

Purchased by	Coal	Electric	Steel
Coal	0.0	0.4	0.6
Electric	0.6	0.1	0.2
Steel	0.4	0.5	0.2

The entries in each column represent the fractional parts of a sector's total output. For instance, column 2 says that the total output of the electric sector is broken up as 40% to coal, 50% to steel, and 10% to electric.

Let p_C = price of coal output, p_E = price of electric output, and p_S = price of steel output. Then, find the equilibrium prices that make it so that each sector earns exactly as much as it spends.

Solution

We begin by setting up a system of equations.

The coal sector pays for 40% of electric and 60% of steel:

$$p_C = 0.4p_E + 0.6p_S$$

The electric sector pays for 60% of coal, 10% of itself, and 20% of steel:

$$p_E = 0.6p_C + 0.1p_E + 0.2p_S$$

The steel sector pays for 40% of coal, 50% of electric, and 10% of itself:

$$p_S = 0.4p_C + 0.5p_E + 0.1p_S$$

We move all variables to the left-hand side to get the following:

$$\begin{cases} p_C - 0.4p_E - 0.6p_S = 0 \\ -0.6p_C + 0.9p_E - 0.2p_S = 0 \\ -0.4p_C - 0.5p_E + 0.9p_S = 0 \end{cases}$$

Now we encode the system as a matrix and row reduce:

$$\begin{bmatrix} 1 & -0.4 & -0.6 & 0 \\ -0.6 & 0.9 & -0.2 & 0 \\ -0.4 & -0.5 & 0.8 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -0.40 & -0.60 & 0 \\ 0 & 0.66 & -0.56 & 0 \\ 0 & -0.66 & 0.56 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -0.40 & -0.60 & 0 \\ 0 & 0.66 & -0.56 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -0.40 & -0.60 & 0 \\ 0 & 1 & -0.85 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 0 & -0.94 & 0 \\ 0 & 1 & -0.85 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We obtain $p_C - 0.94p_S = 0 \Rightarrow p_C = 0.94p_S$, $p_E - 0.85p_S = 0 \Rightarrow p_E = 0.85p_S$, and $0 = 0$ which means that p_S is free. Thus, the general solution vector is

$$\mathbf{p} = \begin{bmatrix} p_C \\ p_E \\ p_S \end{bmatrix} = \begin{bmatrix} 0.94p_S \\ 0.85p_S \\ p_S \end{bmatrix} = p_S \begin{bmatrix} 0.94 \\ 0.85 \\ 1 \end{bmatrix}.$$

Any nonnegative choice of p_S gives an equilibrium price vector. Suppose $p_S = 100$. Then $p_C = 94$, $p_E = 85$, and $p_S = 100$. This means that the coal sector's income equals its expenses at \$94 million and the electric sector's income equals its expenses at \$85 million. The steel sector is set to \$100 million.

We now move on to an application in chemistry. We can represent chemical equations through linear systems.

Example 2.6.2

For instance, let's model the combustion reaction of propane gas burning to heat a barbecue grill. Propane (C_3H_8) combines with molecular oxygen (O_2) to form carbon dioxide (CO_2) and water (H_2O):

$$x_1(C_3H_8) + x_2(O_2) \longrightarrow x_3(CO_2) + x_4(H_2O)$$

To balance this equation, we must find whole numbers $x_1, \ldots, x_4$ such that the total numbers of carbon, hydrogen, and oxygen atoms on the left match the numbers on the right. This is because of the law of conservation of mass.

Solution

We construct vectors listing the numbers of atoms per molecule. C_3H_8 has 3 atoms of carbon, 8 atoms of hydrogen, and 0 atoms of oxygen. Thus, its vector is

$$\begin{bmatrix} 3 \\ 8 \\ 0 \end{bmatrix}.$$

O_2 has 0 atoms of carbon, 0 atoms of hydrogen, and 2 atoms of hydrogen. Thus, its vector is

$$\begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}.$$

CO_2 has 1 atom of carbon, 0 atoms of hydrogen, and 2 atoms of oxygen. Thus its vector is

$$\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}.$$

H_2O has 0 atoms of carbon, 2 atoms of hydrogen, and 1 atom of oxygen. Thus its vector is

$$\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}.$$

We now build the reaction:

$$x_1 \begin{bmatrix} 3 \\ 8 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} = x_3 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$$

Move all terms to the left-hand side:

$$x_1 \begin{bmatrix} 3 \\ 8 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ -2 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ -2 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Row reduce the augmented matrix:

$$\begin{bmatrix} 3 & 0 & -1 & 0 & 0 \\ 8 & 0 & 0 & -2 & 0 \\ 0 & 2 & -2 & -1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & -\frac{1}{4} & 0 \\ 0 & 1 & 0 & -\frac{5}{4} & 0 \\ 0 & 0 & 1 & -\frac{3}{4} & 0 \end{bmatrix}$$

We have $x_1 = \frac{1}{4}x_4$, $x_2 = \frac{5}{4}x_4$, $x_3 = \frac{3}{4}x_4$, and x_4 is free.

To obtain integer coefficients, we will let $x_4 = 4$. This gives us $x_1 = 1$, $x_2 = 5$, $x_3 = 3$, and $x_4 = 4$.

Thus, the balanced chemical equation is

$$C_3H_8 + 5\,O_2 \longrightarrow 3\,CO_2 + 4\,H_2O.$$

Problem 2.6.1

Balance the chemical reaction for the combustion of ethanol in the engine of a flex-fuel vehicle.

$$x_1(C_2H_5OH) + x_2(O_2) \rightarrow x_3(CO_2) + x_4(H_2O)$$

Solution

We now build the reaction:

$$x_1 \begin{bmatrix} 2 \\ 6 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} = x_3 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$$

Move all terms to the left-hand side:

$$x_1 \begin{bmatrix} 2 \\ 6 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ -2 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ -2 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Row reduce the augmented matrix:

$$\begin{bmatrix} 2 & 0 & -1 & 0 & 0 \\ 6 & 0 & 0 & -2 & 0 \\ 1 & 2 & -2 & -1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & -\frac{1}{3} & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -\frac{2}{3} & 0 \end{bmatrix}$$

We have $x_1 = \frac{1}{3}x_4$, $x_2 = x_4$, $x_3 = \frac{2}{3}x_4$, and x_4 is free.

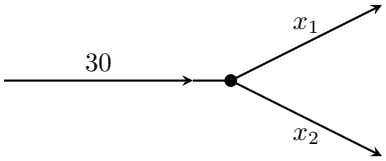
To obtain integer coefficients, we will let $x_4 = 3$. This gives us $x_1 = 1$, $x_2 = 3$, $x_3 = 2$, and $x_4 = 3$.

Thus, the balanced chemical equation is

$$\mathrm{C_2H_5OH} + 3\,\mathrm{O_2} \rightarrow 2\,\mathrm{CO_2} + 3\,\mathrm{H_2O}.$$

Our final application for this section is network flow, which is a part of graph theory. When things move through an interconnected system, they can be modeled using systems of linear equations. Whether it's traffic in a grid of city streets, data packets in a computer network, fluid flow in pipes, or a conveyor belt system in a factory, we can assume that whatever goes into a node comes out.

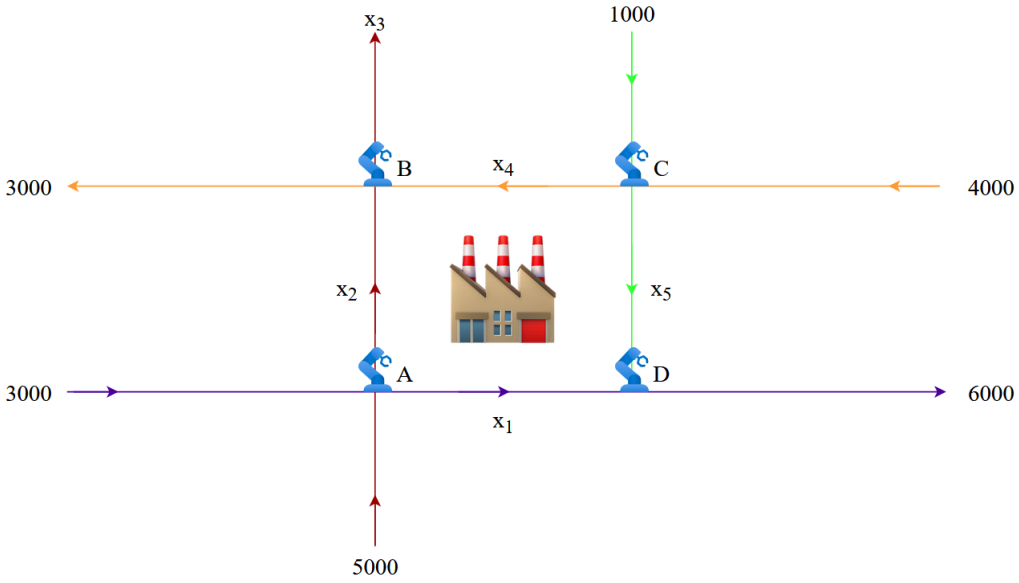
Here, you can see that we have a branch carrying 30 units going into a node (dot) with x_1 and x_2 as branches flowing out of the node:



This would be written as $x_1 + x_2 = 30$.

Example 2.6.3

The conveyor belt system in a battery factory is shown below. The robotic arms represent nodes. x_1, x_2, x_3, x_4, x_5 represent unknown flows. Model the system using a system of linear equations.



Solution

At each robotic arm (node), we will set the units of batteries that come in equal to the units of batteries that come out. Let's begin by breaking down each node:

Node	Batteries in	Batteries out
A	$3000 + 5000$	$x_1 + x_2$
B	$x_2 + x_4$	$3000 + x_3$
C	$1000 + 4000$	$x_4 + x_5$
D	$x_1 + x_5$	6000

The number of batteries into the network $3000 + 5000 + 1000 + 4000 = 13000$ must equal the number of batteries out of the network $3000 + x_3 + 6000$. Setting them equal and simplifying yields $x_3 = 4000 =$. We will combine everything together into a system of equations:

$$\begin{cases} x_1 + x_2 = 8000 \\ x_2 - x_3 + x_4 = 3000 \\ x_4 + x_5 = 5000 \\ x_1 + x_5 = 6000 \\ x_3 = 4000 \end{cases}$$

Let's encode this as an augmented matrix and row reduce:

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 8000 \\ 0 & 1 & -1 & 1 & 0 & 3000 \\ 0 & 0 & 0 & 1 & 1 & 5000 \\ 1 & 0 & 0 & 0 & 1 & 6000 \\ 0 & 0 & 1 & 0 & 0 & 4000 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 6000 \\ 0 & 1 & 0 & 0 & -1 & 2000 \\ 0 & 0 & 1 & 0 & 0 & 4000 \\ 0 & 0 & 0 & 1 & 1 & 5000 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

If you're working by hand, it's probably faster to use back-substitution. Either way, we get the following system:

$$\begin{cases} x_1 & & & & + x_5 = 6000 \\ & x_2 & & & - x_5 = 2000 \\ & & x_3 & & = 4000 \\ & & & x_4 + x_5 = 5000 \end{cases}$$

Solving yields

$$\begin{cases} x_1 = 6000 - x_5 \\ x_2 = 2000 + x_5 \\ x_3 = 4000 \\ x_4 = 5000 - x_5 \\ x_5 \text{ is free} \end{cases}$$

where $x_5 \leq 5000$.

Problem 2.6.2

A large city's economy is reliant on their drugs for autoimmune diseases and cancer treatment. Their three largest sectors are pharmaceuticals, chemicals, and biotech.

- Pharmaceuticals sells 10% of its output to chemicals, 30% to biotech, and retains the rest.
- Chemicals sells 50% of its output to pharmaceuticals, 40% to biotech, and retains the rest.
- Biotech sells 20% of its output to pharmaceuticals, 20 % to chemistry, and retains the rest.

Use a Leontief input-output table to develop a system of equations that leads to the prices at which each sector's income equals its expenses. Then find a set of equilibrium prices when the price for the biotech output is 40 units.

Solution

Purchased by	Pharmaceuticals	Chemicals	Biotech
Pharmaceuticals	0.6	0.5	0.2
Chemicals	0.1	0.1	0.2
Biotech	0.3	0.4	0.6

We now set up a system of equations.

The pharmaceuticals sector pays for 60% of itself, 50% of chemicals, and 20% of biotech:

$$p_P = 0.6p_P + 0.5p_C + 0.2p_B$$

The chemicals sector pays for 10% of pharmaceuticals, 10% of itself, and 20% of biotech:

$$p_C = 0.1p_P + 0.1p_C + 0.2p_B$$

The biotech sector pays for 30% of pharmaceuticals, 40% of chemicals, and 60% of itself:

$$p_B = 0.3p_P + 0.4p_C + 0.6p_B$$

We move all variables to the left-hand side to get the following:

$$\begin{cases} 0.4p_P - 0.5p_C - 0.2p_B = 0 \\ -0.1p_P + 0.9p_C - 0.2p_B = 0 \\ -0.3p_P - 0.4p_C + 0.4p_B = 0 \end{cases}$$

Now we encode the system as a matrix and row reduce:

$$\begin{bmatrix} 0.4 & -0.5 & -0.2 & 0 \\ -0.1 & 0.9 & -0.2 & 0 \\ -0.3 & -0.4 & 0.4 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -0.9032 & 0 \\ 0 & 1 & -0.3226 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We obtain $p_P = 0.9032t$, $p_C = 0.3226t$, and p_B is free. Thus, the general solution vector is

$$\mathbf{p} = \begin{bmatrix} p_P \\ p_C \\ p_B \end{bmatrix} = \begin{bmatrix} 0.9032p_B \\ 0.3226p_B \\ p_S \end{bmatrix} = p_B \begin{bmatrix} 0.9032 \\ 0.3226 \\ 1 \end{bmatrix}.$$

If $p_B = 40$, then $p_P = 36.13$, $p_C = 12.90$, and $p_S = 40$. This means that the biotech sector's income equals its expenses at \$36.13 billion and the chemicals sector's income equals its expenses at \$12.90 billion. The biotech sector is set to \$40 billion.

Problem 2.6.3

France leads the world in roundabouts per capita. Find the general solution of the traffic flow in the network below and find the smallest possible value for x_6. The nodes represent intersections.

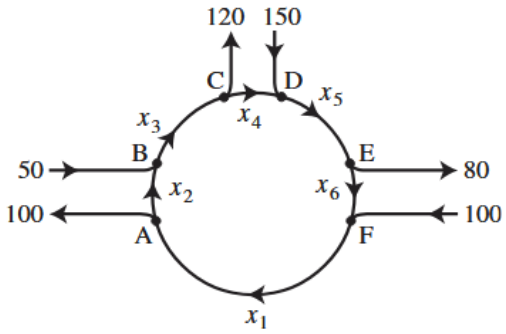


Image credit: Lay et al

Solution

Intersection	Flow in	Flow out
A	x_1	$x_2 + 100$
B	$x_2 + 50$	x_3
C	x_3	$x_4 + 120$
D	$x_4 + 150$	x_5
E	x_5	$x_6 + 80$
F	$x_6 + 100$	x_1

Rearrange the equations to get

$$\begin{cases} x_1 - x_2 = 100 \\ x_2 - x_3 = 50 \\ x_3 - x_4 = 120 \\ x_4 - x_5 = -150 \\ x_5 - x_6 = 80 \\ -x_1 + x_6 = -100 \end{cases}.$$

We form and row reduce the augmented matrix:

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 100 \\ 0 & 1 & -1 & 0 & 0 & 0 & -50 \\ 0 & 0 & 1 & -1 & 0 & 0 & 120 \\ 0 & 0 & 0 & 1 & -1 & 0 & -150 \\ 0 & 0 & 0 & 0 & 1 & -1 & 80 \\ -1 & 0 & 0 & 0 & 0 & 1 & -100 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 & 100 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 50 \\ 0 & 0 & 0 & 1 & 0 & -1 & -70 \\ 0 & 0 & 0 & 0 & 1 & -1 & 80 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The general solution is

$$\begin{cases} x_1 = 100 + x_6 \\ x_2 = x_6 \\ x_3 = 50 + x_6 \\ x_4 = -70 + x_6 \\ x_5 = 80 + x_6 \\ x_6 \text{ is free} \end{cases}$$

where $x_6 \geq 70$.

Problem 2.6.4

Gasoline is delivered to terminals T_1 and T_2. They distribute the fuel to three storage depots S_1, S_2, S_3. The following network diagram shows what fraction of the fuel goes from each terminal to the three storage depots. Then, the three depots supply fuel to four Phillips 66 gas stations P_1, P_2, P_3, P_4.

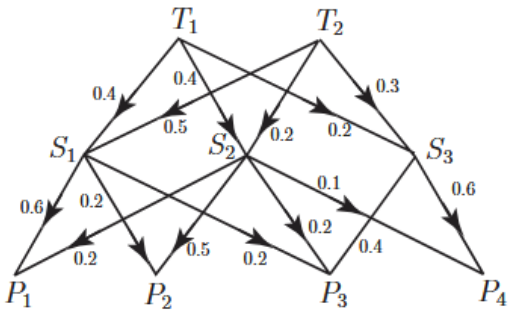


Image credit: Loughborough University

Let the amount of fuel in liters flowing from T_1 be t_1 and let the amount of fuel flowing from T_2 be t_2. Let the quantities received by storage depots S_1, S_2, S_3 are denoted by s_1, s_2, s_3. Let p_1, p_2, p_3, p_4 denote the amounts received by gas stations P_1, P_2, P_3, P_4.

(a) Express the relationship between terminal flows t_1, t_2 and storage depot quantities s_1, s_2, s_3 using a matrix equation.

(b) Express the relationship between storage depot quantities s_1, s_2, s_3 and gas station quantities p_1, p_2, p_3, p_4 using a matrix equation.

(c) Combine these two systems into a single matrix equation that expresses gas station quantities p_1, p_2, p_3, p_4 directly in terms of terminal flows t_1, t_2.

(d) Terminal T_1 delivers 1200 liters of fuel and terminal T_2 delivers 800 liters. Determine the amount of fuel delivered to each gas station P_1, P_2, P_3, P_4.

Solution

(a) From the diagram we have the following:

$$\begin{aligned} s_1 &= 0.4t_1 + 0.5t_2 \\ s_2 &= 0.4t_1 + 0.2t_2 \\ s_3 &= 0.2t_1 + 0.3t_2 \end{aligned}$$

Let's encode this as a matrix:

$$\begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} 0.4 & 0.5 \\ 0.4 & 0.2 \\ 0.2 & 0.3 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$$

(b) From the diagram we have the following:

$$\begin{aligned} p_1 &= 0.6s_1 + 0.2s_2 \\ p_2 &= 0.2s_1 + 0.5s_2 \\ p_3 &= 0.2s_1 + 0.2s_2 + 0.4s_3 \\ p_4 &= 0.1s_2 + 0.6s_3 \end{aligned}$$

Let's encode this as a matrix:

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix} = \begin{bmatrix} 0.6 & 0.2 & 0.0 \\ 0.2 & 0.5 & 0.0 \\ 0.2 & 0.2 & 0.4 \\ 0.0 & 0.1 & 0.6 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix}$$

(c) We will substitute the expressions for s_1, s_2, s_3 in terms of t_1, t_2 into the gas station equations:

$$\begin{aligned} p_1 &= 0.6s_1 + 0.2s_2 = 0.6(0.4t_1 + 0.5t_2) + 0.2(0.4t_1 + 0.2t_2) = 0.32t_1 + 0.34t_2 \\ p_2 &= 0.2s_1 + 0.5s_2 = 0.2(0.4t_1 + 0.5t_2) + 0.5(0.4t_1 + 0.2t_2) = 0.28t_1 + 0.20t_2 \\ p_3 &= 0.2s_1 + 0.2s_2 + 0.4s_3 = 0.2(0.4t_1 + 0.5t_2) + 0.2(0.4t_1 + 0.2t_2) + 0.4(0.2t_1 + 0.3t_2) = 0.24t_1 + 0.26t_2 \\ p_4 &= 0.1s_2 + 0.6s_3 = 0.1(0.4t_1 + 0.2t_2) + 0.6(0.2t_1 + 0.3t_2) = 0.16t_1 + 0.20t_2 \end{aligned}$$

Thus, the combined system in matrix form is

$$\begin{aligned}
\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix} &= \begin{bmatrix} 0.6 & 0.2 & 0 \\ 0.2 & 0.5 & 0 \\ 0.2 & 0.2 & 0.4 \\ 0 & 0.1 & 0.6 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} \\
&= \begin{bmatrix} 0.6 & 0.2 & 0 \\ 0.2 & 0.5 & 0 \\ 0.2 & 0.2 & 0.4 \\ 0 & 0.1 & 0.6 \end{bmatrix} \begin{bmatrix} 0.4 & 0.5 \\ 0.4 & 0.2 \\ 0.2 & 0.3 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \\
&= \begin{bmatrix} 0.32 & 0.34 \\ 0.28 & 0.20 \\ 0.24 & 0.26 \\ 0.16 & 0.20 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \\
&= \begin{bmatrix} 0.32t_1 + 0.34t_2 \\ 0.28t_1 + 0.20t_2 \\ 0.24t_1 + 0.26t_2 \\ 0.16t_1 + 0.20t_2 \end{bmatrix}.
\end{aligned}$$

(d) We have $t_1 = 1200$ and $t_2 = 800$. Let's substitute these values into the system:

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix} = \begin{bmatrix} 0.32 & 0.34 \\ 0.28 & 0.20 \\ 0.24 & 0.26 \\ 0.16 & 0.20 \end{bmatrix} \begin{bmatrix} 1200 \\ 800 \end{bmatrix}$$

And now we compute:

$$\begin{aligned}
p_1 &= 0.32(1200) + 0.34(800) = 656 \text{ liters} \\
p_2 &= 0.28(1200) + 0.20(800) = 496 \text{ liters}
\end{aligned}$$

$$p_3 = 0.24(1200) + 0.26(800) = 496 \text{ liters}$$
$$p_4 = 0.16(1200) + 0.20(800) = 352 \text{ liters}$$

2.7 Linear Transformations Through Matrices

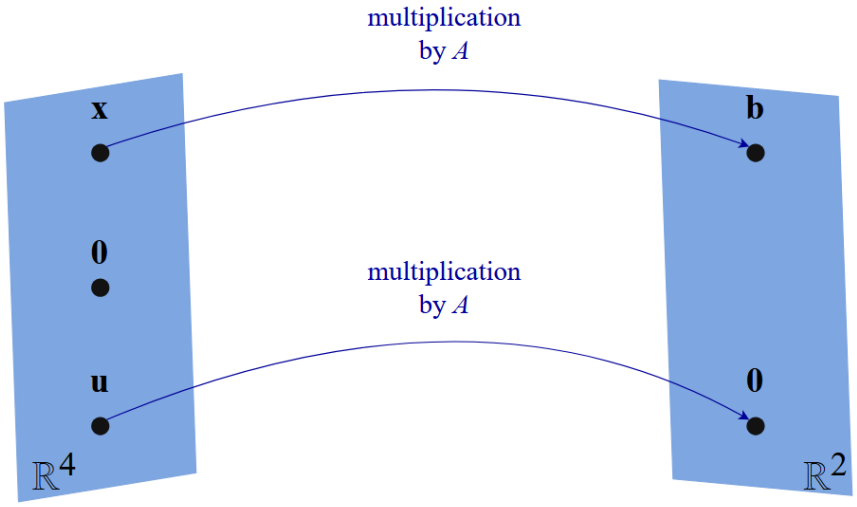
A matrix equation $A\mathbf{x} = \mathbf{b}$ and its associated vector equation $x_1\mathbf{x}_1 + \cdots + x_n\mathbf{a}_n$ are different ways of encoding the same data. However, let's look at different case. Say we have the following matrix equation:

$$\underbrace{\begin{bmatrix} 4 & -3 & 1 & 3 \\ 2 & 0 & 5 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 5 \\ 8 \end{bmatrix}}_{\mathbf{b}}$$

And then we have a new equation:

$$\underbrace{\begin{bmatrix} 4 & -3 & 1 & 3 \\ 2 & 0 & 5 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 1 \\ 4 \\ -1 \\ 3 \end{bmatrix}}_{\mathbf{u}} = \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{\mathbf{0}}$$

As you can see, $\mathbf{x}$ transformed into $\mathbf{b}$ and $\mathbf{u}$ transformed into $\mathbf{0}$. The vectors transformed through matrix multiplication.



Thus, solving $A\mathbf{x} = \mathbf{b}$ means finding all $\mathbf{x} \in \mathbb{R}^4$ that are transformed into the vector $\mathbf{b} \in \mathbb{R}^2$. You can think of A almost like a function that takes in vector inputs and transforms them. This is a mapping we can write as

$$T : \mathbb{R}^4 \to \mathbb{R}^2.$$

We now introduce some new language. Suppose we have a general mapping $T : \mathbb{R}^n \to \mathbb{R}^m$. Recall that $\mathbb{R}^n$ is the domain of T and $\mathbb{R}^m$ is the codomain. For $\mathbf{x} \in \mathbb{R}^n$, the vector $T(\mathbf{x}) \in \mathbb{R}^m$ is called the **image** of $\mathbf{x}$ under the operation of T. The set of all images $T(\mathbf{x})$ is the range of T. This is called a **linear transformation**. We will explore linear transformations in more depth as we go on, but for now we will learn about how they are represented through matrices.

Matrix Transformations

Matrix transformations are mappings associated with matrix multiplication. We have $T : \mathbb{R}^n \to \mathbb{R}^m$ which can also be written as $\mathbf{x} \mapsto A\mathbf{x}$. For each $\mathbf{x} \in \mathbb{R}^n$, $T(\mathbf{x})$ is computed as $A\mathbf{x}$ where A is an $m \times n$ matrix. The domain of T is $\mathbb{R}^n$ and the codomain of T is $\mathbb{R}^m$. The range of T is the set of all linear combinations of the columns of A because each image $T(\mathbf{x})$ is of the form $A\mathbf{x}$.

Example 2.7.1

Let $A = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}$, and $\mathbf{c} = \begin{bmatrix} 2 \\ 2 \\ 5 \end{bmatrix}$.

We will define a transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ by $T(\mathbf{x}) = A\mathbf{x}$ so that

$$T(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - 3x_2 \\ 3x_1 + 5x_2 \\ -x_1 + 7x_2 \end{bmatrix}.$$

(a) Compute the image of $\mathbf{u}$ under T.

(b) Find a vector $\mathbf{x} \in \mathbb{R}^2$ whose image under T is $\mathbf{b}$.

(c) Is there more than one $\mathbf{x}$ whose image under T is $\mathbf{b}$?

(d) Determine whether $\mathbf{c}$ is in the range of the transformation T.

Solution

(a) We find the image of $\mathbf{u}$ under T by computing $T(\mathbf{u})$:

$$T(\mathbf{u}) = A\mathbf{u} = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1(2) + (-3)(-1) \\ 3(2) + 5\ (-1) \\ -1(2) + 7\ (-1) \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ -9 \end{bmatrix}$$

(b) We solve $T(\mathbf{x}) = \mathbf{b}$ for $\mathbf{x}$ which means solving $A\mathbf{x} = \mathbf{b}$:

$$\begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}$$

Row reduce the augmented matrix:

$$\begin{bmatrix} 1 & -3 & 3 \\ 3 & 5 & 2 \\ -1 & 7 & -5 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 14 & -7 \\ 0 & 4 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}$$

From this, we get $x_2 = -\frac{1}{2} \Rightarrow x_1 = \frac{3}{2}$ and thus $\mathbf{x} = \begin{bmatrix} 1.5 \\ -0.5 \end{bmatrix}$.

(c) The system has a unique solution. Hence, there is exactly one vector $\mathbf{x}$ whose image under T is $\mathbf{b}$.

(d) $\mathbf{c}$ is in the range of T if $\mathbf{c}$ is in the range of some $\mathbf{x} \in \mathbb{R}^2$. In other words, is there some $\mathbf{x}$ that makes $\mathbf{c} = T(\mathbf{x})$ true? This is just another way of asking if $A\mathbf{x} = \mathbf{c}$ is consistent, so we will try to find a solution:

$$\begin{bmatrix} 1 & -3 & 2 \\ 3 & 5 & 2 \\ -1 & 7 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 2 \\ 0 & 14 & -4 \\ 0 & 4 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 2 \\ 0 & 1 & -\frac{2}{7} \\ 0 & 0 & -5 \end{bmatrix}$$

As you can see here, we have $0 = -5$ which is a contradiction. Thus, the system is inconsistent and $\mathbf{c}$ is not in the range of T.

The previous question is all about addressing existence and uniqueness but encoded as a matrix transformation rather than a system of linear equations. You are asking if there exists an $\mathbf{x}$ whose image is $\mathbf{c}$ and if $\mathbf{b}$ is the image of a unique $\mathbf{x} \in \mathbb{R}^n$.

Example 2.7.2

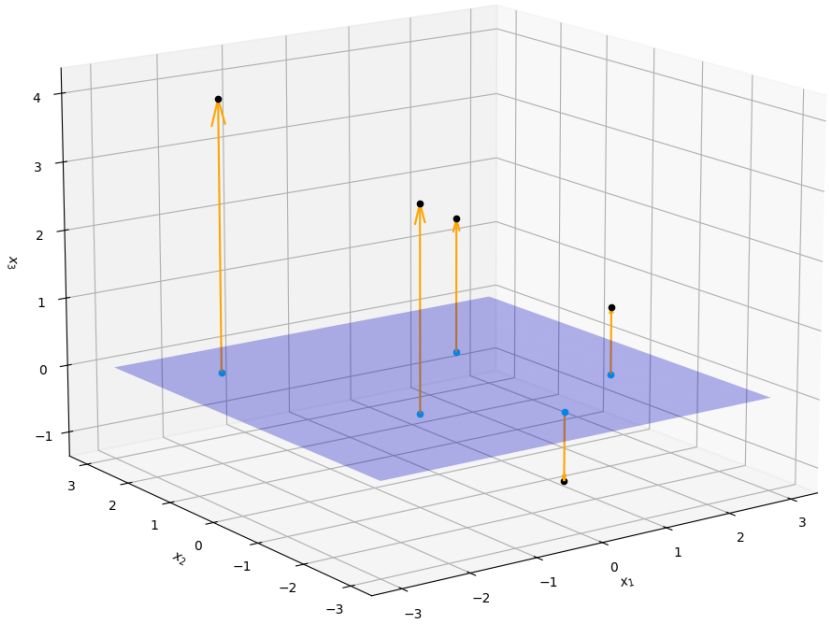
Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. We define a transformation $T(\mathbf{x}) = A\mathbf{x}$ where $\mathbf{x} \in \mathbb{R}^3$. Describe what this transformation does to a general point $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3$.

Solution

We have

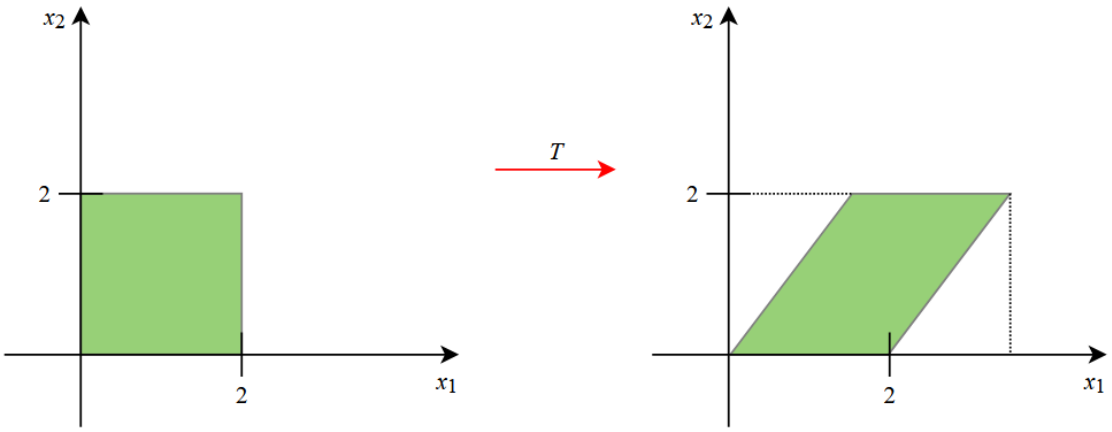
$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix}.$$

The transformation $\mathbf{x} \mapsto A\mathbf{x}$ projects points $\mathbf{x} \in \mathbb{R}^3$ onto the x_1x_2-plane. It "flattens" any 3D points down to 2D and as a result, the x_3-component becomes 0.



A projection onto the x_1x_2-plane for various points

Let's discuss a geometric transformation. The transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(\mathbf{x}) = A\mathbf{x}$ is called a **shear transformation**. A shear transformation distorts shapes by pushing the points sideways or vertically:



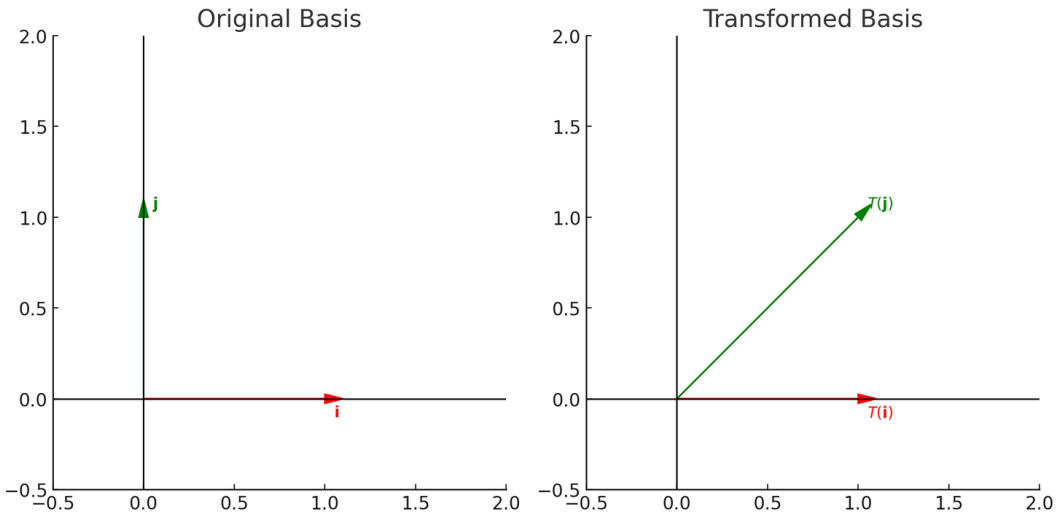
Notice how the base of the parallelogram remains the same.

The shear matrix here would be $A = \begin{bmatrix} 1 & \frac{3}{4} \\ 0 & 1 \end{bmatrix}$. The image of $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$ is given by

$$\begin{bmatrix} 1 & \frac{3}{4} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 3.5 \\ 2 \end{bmatrix}.$$

And as you can see, the tip of the parallelogram after the transformation is at 3.5.

Problem 2.7.1



Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation defined by the matrix A. Compute the image of the vector $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ under T, and explain the transformation in terms of basis vectors.

Solution

We begin with $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Then we have a horizontal shear that fixes $\mathbf{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and maps $\mathbf{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ to $\mathbf{i} + \mathbf{j} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Our new matrix is $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

We begin by observing that any vector in $\mathbb{R}^2$ can be written as a linear combination of $\mathbf{i}$ and $\mathbf{j}$:

$$\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 2\mathbf{i} + 3\mathbf{j} = 2\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Under the shear transformation T, we apply the matrix A by linearity:

$$T(\mathbf{v}) = A\mathbf{v} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2(1) + 3(3) \\ 0(1) + 3(3) \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

Properties of Linear Transformations

A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be linear if it satisfies the following two properties for all vectors $\mathbf{u}, \mathbf{v}$ in the domain of T and all scalars $c \in \mathbb{R}$:

(i) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ (Additivity)

(ii) $T(c\mathbf{u}) = cT(\mathbf{u})$ (Homogeneity)

These two properties ensure that a linear transformation preserves the operations of vector addition and scalar multiplication.

While every linear transformation here is a matrix transformation, there are many important linear transformations that are not matrix transformations we will discuss later. From the previous two properties, we have some more useful facts:

If T is a linear transformation, then

$$T(\mathbf{0}) = \mathbf{0} \tag{3}$$

and

$$T(c\mathbf{u} + d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v}) \tag{4}$$

for all vectors $\mathbf{u}, \mathbf{v}$ in the domain of T and scalars $c, d \in \mathbb{R}$.

This shows that linear transformations preserve linear combinations. Observe how the property of additivity results in property (3) because

$$T(\mathbf{0}) = T(0\mathbf{u}) = 0T(\mathbf{u}) = \mathbf{0}.$$

Property (4) requires both the properties of additivity and homogeneity:

$$T(c\mathbf{u} + d\mathbf{v}) = T(c\mathbf{u}) + T(d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v})$$

Any transformation that satisfies property (4) must be linear. Applying the property (4) repeatedly yields

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_pT(\mathbf{v}_p)$$

This result is called the *superposition principle* in physics and engineering. It states that if an input is expressed as a linear combination of signals, the system's response is the same linear combination of the individual responses to those signals.

Example 2.7.3

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $T(\mathbf{x}) = r\mathbf{x}$ for a scalar $r \in \mathbb{R}$. This transformation is called

- a **contraction** if $0 \leq r \leq 1$.
- a **dilation** if $r > 1$,

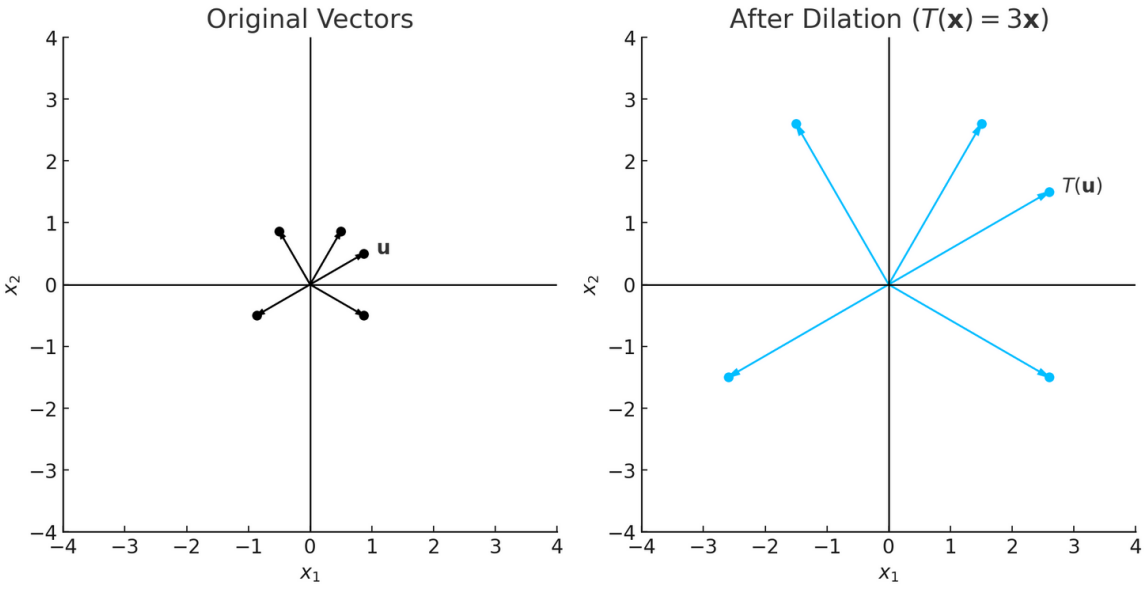
Suppose $r = 3$. Show that T is a linear transformation.

Solution

We need to make sure it satisfies property (4) which is the linearity condition:

$$T(c\mathbf{u}+d\mathbf{v}) = 3(c\mathbf{u}+d\mathbf{v})$$
$$= 3c\mathbf{u}+3d\mathbf{v} = c(3\mathbf{u})+d(3\mathbf{v}) = cT(\mathbf{u})+dT(\mathbf{v})$$

T satisfies (4) and thus is a linear transformation.



Example 2.7.4

Define a transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$T(\mathbf{x}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}.$$

For $\mathbf{u} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, and $\mathbf{u}+\mathbf{v} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$, find the images under T:

Solution

$$T(\mathbf{u}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$$

$$T(\mathbf{v}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}$$

$$T(\mathbf{u}+\mathbf{v}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 6 \\ 4 \end{bmatrix} = \begin{bmatrix} -4 \\ 6 \end{bmatrix}$$

Since $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$, this confirms that T is a linear transformation.

Geometrically, this transformation rotates the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u}+\mathbf{v}$ by $90°$ counterclockwise. The parallelogram formed by $\mathbf{u}$ and $\mathbf{v}$ is mapped to a new parallelogram formed by $T(\mathbf{u})$ and $T(\mathbf{v})$.

We now shift away from geometric interpretations and into one that's all about how linear mappings can be used to transform the type of data we are encoding.

Suppose a company manufactures two products B and C. The cost of producing one dollar's worth of each product is described by the following unit cost matrix $U = \begin{bmatrix} \mathbf{b} & \mathbf{c} \end{bmatrix}$:

$$U = \begin{bmatrix} 0.45 & 0.40 \\ 0.25 & 0.30 \\ 0.15 & 0.15 \end{bmatrix} \begin{matrix} \text{Materials} \\ \text{Labor} \\ \text{Overhead} \end{matrix}$$

Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$ be a production vector, where x_1 is the number of dollars allocated to product B and x_2 is the number of dollars allocated to product C. We define a transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ that outputs the total cost vector:

$$T(\mathbf{x}) = U\mathbf{x} = x_1 \begin{bmatrix} 0.45 \\ 0.25 \\ 0.15 \end{bmatrix} + x_2 \begin{bmatrix} 0.40 \\ 0.30 \\ 0.15 \end{bmatrix} = \begin{bmatrix} \text{Total cost of materials} \\ \text{Total cost of labor} \\ \text{Total cost of overhead} \end{bmatrix}$$

The matrix U transforms the production allocation vector $\mathbf{x}$ into a vector of associated costs. Since this transformation is matrix multiplication, it is linear. There are two interpretations of this mapping:

1. If the production vector $\mathbf{x}$ is scaled by a constant c, then the total costs also scale:

$$T(c\mathbf{x}) = cT(\mathbf{x})$$

   For example, if production is quadrupled from $\mathbf{x}$ to $4\mathbf{x}$, then the cost vector becomes $4T(\mathbf{x})$.

2. If $\mathbf{x}$ and $\mathbf{y}$ are separate production vectors, then the total cost of the combined production is the sum of individual costs:

$$T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$$

   This reflects the additivity of matrix transformations and makes sense intuitively.

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is defined by what it does to the columns of the $n \times n$ identity matrix I_n. These columns represent the standard basis vectors.

Example 2.7.5

Let the columns of $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ be $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Then, suppose $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear transformation such that

$$T(\mathbf{e}_1) = \begin{bmatrix} 5 \\ -7 \\ 2 \end{bmatrix}, \; T(\mathbf{e}_2) = \begin{bmatrix} -3 \\ 8 \\ 0 \end{bmatrix}$$

Find a formula for the image of any $\mathbf{x} \in \mathbb{R}^2$.

Solution

We begin by expressing $\mathbf{x} \in \mathbb{R}^2$ as a linear combination of the standard basis vectors:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$$

We know that T is a linear transformation, so we can apply linearity. For any $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$, we have

$$T(\mathbf{x}) = x_1 T(\mathbf{e}_1) + x_2 T(\mathbf{e}_2) = x_1 \begin{bmatrix} 5 \\ -7 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} -3 \\ 8 \\ 0 \end{bmatrix} = \begin{bmatrix} 5x_1 - 3x_2 \\ -7x_1 + 8x_2 \\ 2x_1 \end{bmatrix}.$$

We can also put these vectors into the columns of a matrix A to rewrite $T(\mathbf{x})$ as $\begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A\mathbf{x}$.

Let's generalize this:

Standard Matrix of a Linear Transformation

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there exists a unique matrix A such that

$$T(\mathbf{x}) = A\mathbf{x}$$

for all $\mathbf{x} \in \mathbb{R}^n$.

A is the $m \times n$ matrix whose jth column is the vector $T(\mathbf{e}_j)$ where $\mathbf{e}_j$ is the jth column of the identity matrix $I_n \in \mathbb{R}^n$:

$$A = \begin{bmatrix} T(\mathbf{e}_1) & \dots & T(\mathbf{e}_2) \end{bmatrix}$$

This is called the **standard matrix for the linear transformation** T.

The formula is justified by writing $\mathbf{x}$ as a linear combination of basis vectors:

$$\mathbf{x} = I_n \mathbf{x} = \begin{bmatrix} \mathbf{e}_1 & \dots & \mathbf{e}_2 \end{bmatrix} \mathbf{x} = x_1 \mathbf{e}_1 + \cdots + x_n \mathbf{e}_n$$

And then by linearity,

$$T(\mathbf{x}) = T(x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n) = x_1T(\mathbf{e}_1) + \cdots + x_nT(\mathbf{e}_n) = \begin{bmatrix} T(\mathbf{e}_1) & \dots & T(\mathbf{e}_2) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_2 \end{bmatrix} = A\mathbf{x}.$$

Example 2.7.6

Find the standard matrix A for the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T(\mathbf{x}) = 3\mathbf{x}$$

for $\mathbf{x} \in \mathbf{R}^2$.

Solution

A dilation multiplies every vector by a constant factor. In this case, each input vector is scaled by 3. Since the transformation is linear, we construct the matrix A by applying T to the standard basis vectors.

$$T(\mathbf{e}_1) = 3\mathbf{e}_1 = 3\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

$$T(\mathbf{e}_2) = 3\mathbf{e}_2 = 3\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

We now form the matrix A by writing these outputs as columns:

$$A = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

Example 2.7.7

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that rotates every vector about the origin by an angle ϕ counterclockwise. Find the standard matrix A for this transformation.

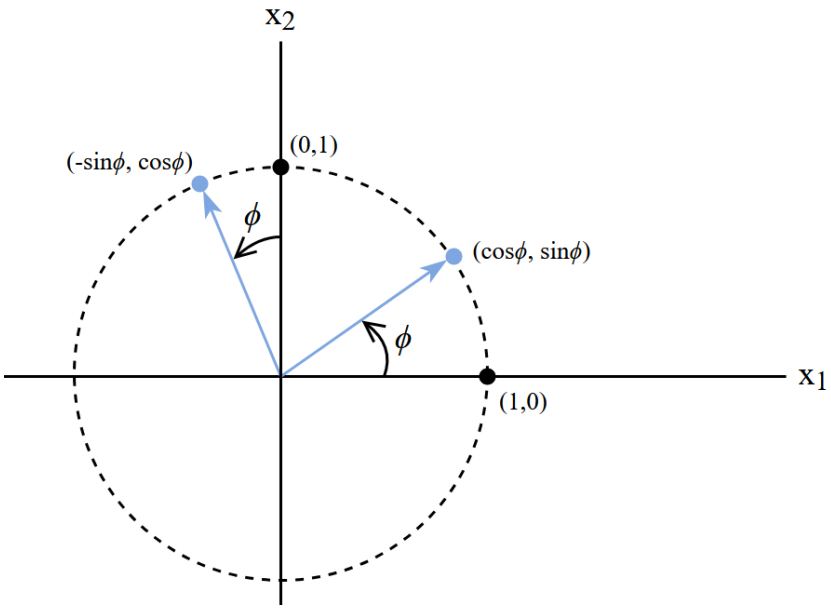
Solution

We apply T to the standard basis vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Geometrically, rotating by an angle ϕ transforms $\mathbf{e}_1$ into

$$T(\mathbf{e}_1) = \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix}$$

and $\mathbf{e}_2$ into

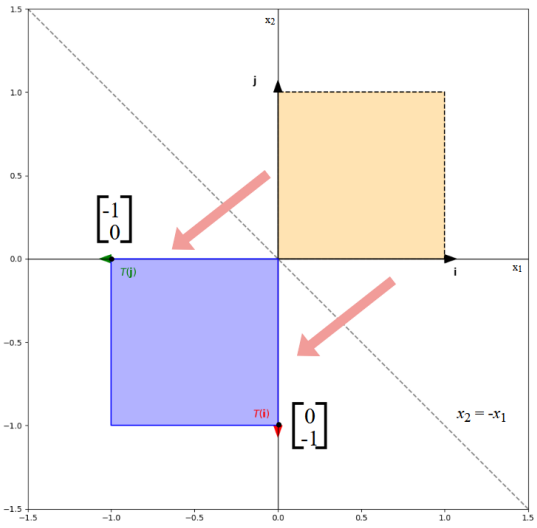
$$T(\mathbf{e}_2) = \begin{bmatrix} -\sin\phi \\ \cos\phi \end{bmatrix}.$$



We now have four tables of common geometric linear transformations.

Transformation	**Standard Matrix**
Reflection through the x_1-axis	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
Reflection through the x_2-axis	$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$
Reflection through the line $x_2 = x_1$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
Reflection through the line $x_2 = -x_1$	$\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$
Reflection through the origin	$\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$

Table 1: Reflections



Reflection through the line $x_2 = -x_1$

Transformation	Instances	Standard Matrix
Horizontal contraction or expansion	$0 < k < 1$ or $k > 1$	$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$
Vertical contraction or expansion	$0 < k < 1$ or $k > 1$	$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$

Table 2: Contractions and Expansions

Transformation	Instances	Standard Matrix
Horizontal shear	$k < 0$ and $k > 0$	$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$
Vertical shear	$k < 0$ and $k > 0$	$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$

Table 3: Shears

Transformation	Standard Matrix
Projection onto x_1-axis	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
Projection onto x_2-axis	$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$

Table 4: Projections

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be onto $\mathbb{R}^m$ if each $\mathbf{b} \in \mathbb{R}^m$ is the image of at least one $\mathbf{x} \in \mathbb{R}^n$.

In other words, the range of T must be all of the codomain of $\mathbb{R}^m$. That is, for every $\mathbf{b}$ in the codomain of $\mathbb{R}^m$, the equation $T(\mathbf{x}) = \mathbf{b}$ has at least one solution. From every input $\mathbf{x} \in \mathbb{R}^m$, you should be able to reach every $\mathbf{b} \in \mathbb{R}^m$ in your target space. If there is any value that makes this false, then T is *not* onto. This is how we address *existence*.

Recall that we call a mapping *one-to-one* if each $\mathbf{b} \in \mathbb{R}^m$ is the image of at most one $\mathbf{x} \in \mathbb{R}^n$. That is, for each $\mathbf{b} \in \mathbb{R}^m$, the equation $T(\mathbf{x}) = \mathbf{b}$ has at most one solution. This is how we address *uniqueness*.

The transformations in Table 4 are not one-to-one and do not map $\mathbb{R}^2$ onto $\mathbb{R}^2$. On the other hand, the transformations in Tables 1, 2, and 3 are one-to-one and do map $\mathbb{R}^2$ onto $\mathbb{R}^2$.

Example 2.7.8

We have a transformation $T : \mathbb{R}^4 \to \mathbb{R}^3$ with the standard matrix $A = \begin{bmatrix} 1 & -4 & 8 & 1 \\ 0 & 2 & -1 & 3 \\ 0 & 0 & 0 & 5 \end{bmatrix}$.

Determine (a) if T maps $\mathbb{R}^4$ onto $\mathbb{R}^3$ and (b) whether or not T is a one-to-one mapping.

Solution

(a) This matrix is already in REF, and we can see that there's a pivot in every row. That means the equation $A\mathbf{x} = \mathbf{b}$ is solvable for every $\mathbf{b} \in \mathbb{R}^3$. Thus, T is onto $\mathbb{R}^3$.

(b) The matrix has 4 columns but only 3 pivots, so there's a free variable. Thus, the system has infinitely many solutions. Each $\mathbf{b}$ is the image of more than one $\mathbf{x}$. Therefore, T is not one-to-one.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T is one-to-one if and only if $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.

This makes sense because if you want every output to come from only one input, then the only way to map to $\mathbf{0}$ should be from $\mathbf{0}$. This gives us an easy way to test.

We now have a new theorem that connects linear transformations to earlier concepts.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation where the standard matrix of T is A.

(a) T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if and only if the columns of A span $\mathbb{R}^m$

(b) T is one-to-one if and only if the columns of A are linearly independent

Example 2.7.9

Let $T(\mathbf{x}) = T(x_1, x_2) = (3x_1 + x_2,\ 5x_1 + 7x_2,\ x_1 + 3x_2)$. Show that T is a one-to-one linear transformation. Then, determine if T maps $\mathbb{R}^2$ onto $\mathbb{R}^3$.

Solution

$T(\mathbf{x})$ can be rewritten as

$$T(\mathbf{x}) \begin{bmatrix} 3x_1 + x_2 \\ 5x_1 + 7x_2 \\ x_1 + 3x_2 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 5 & 7 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A\mathbf{x}.$$

So T is a linear transformation with standard matrix

$$A = \begin{bmatrix} 3 & 1 \\ 5 & 7 \\ 1 & 3 \end{bmatrix}.$$

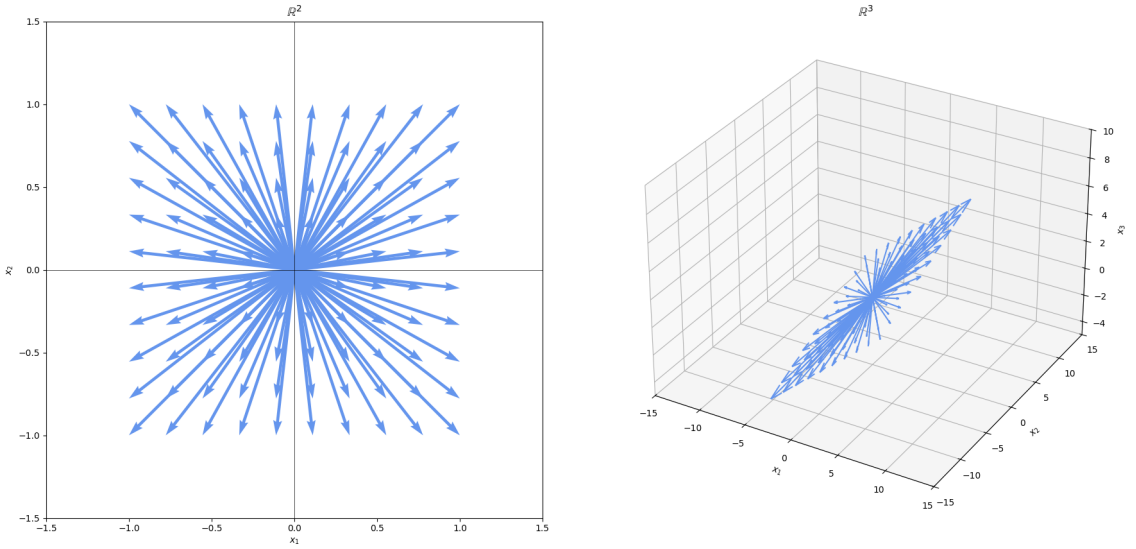
To determine whether T is one-to-one or onto, we must analyze A.

Recall that a linear transformation is one-to-one if and only if the columns of the matrix A are linearly independent. Looking at the columns of A,

$$\mathbf{a}_1 = \begin{bmatrix} 3 \\ 5 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{a}_2 = \begin{bmatrix} 1 \\ 7 \\ 3 \end{bmatrix},$$

we can see that $\mathbf{a}_1$ and $\mathbf{a}_2$ are not scalar multiples of one another, so they are linearly independent. Therefore, T is one-to-one.

To be onto $\mathbb{R}^3$, the columns of A must span $\mathbb{R}^3$. But A is a 3×2 matrix; it only has 2 columns and thus can span a plane in $\mathbb{R}^3$, but not the entire space. T is not onto $\mathbb{R}^3$.



The left graph shows vectors in $\mathbb{R}^2$, and the right graph shows their images in $\mathbb{R}^3$. The image of T lies within a two-dimensional subspace of $\mathbb{R}^3$, so T is one-to-one but not onto $\mathbb{R}^3$.

Problem 2.7.2

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that first performs a horizontal shear that maps $\mathbf{e}_2$ into $\mathbf{e}_2 - 0.5\,\mathbf{e}_1$ and then reflects the result through the x_2-axis. Assuming that T is linear, find its standard matrix. In other words, determine the final location of the images of $\mathbf{e}_1$ and $\mathbf{e}_2$.

Solution

Let's track what happens to each of the standard basis vectors.

(1) $\mathbf{e}_1$ is unaffected by the shear, and then it gets reflected across the x_2-axis. That flips its x_1-component. So

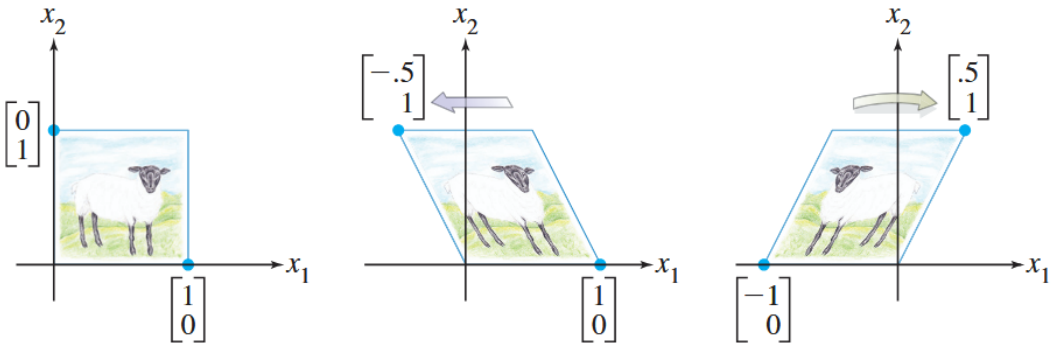
$$T(\mathbf{e}_1) = -\mathbf{e}_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

(2) $\mathbf{e}_2$ gets sheared to $\mathbf{e}_2 - 0.5\,\mathbf{e}_1$ which is

$$\begin{bmatrix} -0.5 \\ 1 \end{bmatrix}.$$

(3) This result is reflected across the x_2-axis, flipping the x_1-component again. That gives

$$T(\mathbf{e}_2) = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}.$$



(1) Unaffected, (2) shear transformation, and (3) reflection through the x_2-axis. Image credit: Lay et al

So the images of the standard basis vectors are

$$T(\mathbf{e}_1) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \quad T(\mathbf{e}_2) = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}.$$

Thus, the standard matrix of T is

$$A = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} -\mathbf{e}_1 & \mathbf{e}_2 + 0.5\mathbf{e}_1 \end{bmatrix} = \begin{bmatrix} -1 & 0.5 \\ 0 & 1 \end{bmatrix}.$$

Problem 2.7.3

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that first reflects points through the vertical x_2-axis and then rotates points counterclockwise by $\frac{\pi}{2}$ radians. Find the standard matrix of T.

Solution

Let R be the reflection through the x_2-axis and let Q be the rotation by $\frac{\pi}{2}$ radians.

Reflecting across the vertical axis negates the x_1-coordinate and leaves x_2 unchanged:

$$R(\mathbf{x}) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x}$$

The standard rotation matrix is

$$Q(\mathbf{x}) = \begin{bmatrix} \cos\left(\frac{\pi}{2}\right) & -\sin\left(\frac{\pi}{2}\right) \\ \sin\left(\frac{\pi}{2}\right) & \cos\left(\frac{\pi}{2}\right) \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

We first apply the reflection and then the rotation:

$$T(\mathbf{x}) = Q(R(\mathbf{x})) = Q \cdot R \cdot \mathbf{x}$$

$$A = Q \cdot R = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$$

Alternatively, we could have done that in vector notation. We begin by reflecting across the x_2-axis:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \to \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \to \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

And then we rotate by $\frac{\pi}{2}$ radians counterclockwise.

Rotation by $\frac{\pi}{2}$ maps each vector $\begin{bmatrix} a \\ b \end{bmatrix}$ to $\begin{bmatrix} -b \\ a \end{bmatrix}$. Let's apply that to our vectors:

$$\mathbf{e}_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \to \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \to \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

Using these as columns in A, we get

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

Problem 2.7.4

Let $T : \mathbb{R}^4 \to \mathbb{R}^4$ be the linear transformation with standard matrix

$$A = \begin{bmatrix} -5 & 10 & -5 & 4 \\ 8 & 3 & -4 & 7 \\ 4 & -9 & 5 & -3 \\ -3 & -2 & 5 & 4 \end{bmatrix}.$$

Determine whether or not T is one-to-one.

Solution

We row reduce to

$$\sim \begin{bmatrix} 1 & 0 & 0 & 1.2571 \\ 0 & 1 & 0 & 2.2571 \\ 0 & 0 & 1 & 2.4571 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

There is no pivot in the 4th column, so the columns are linearly dependent. There are free variables and thus nontrivial solutions. T is not one-to-one.

In this section, we united much of what we have learned into the a cohesive framework. Going forward, we will continue to develop our geometric intuition. This begins with interpreting matrices as tools that transform space.

2.8 Linear Models

We now have some more real-world applications. Linear models let us represent phenomena where the end result can be expressed as a linear combination of individual components. That is, each input variable contributes proportionally to the outcome. Remember that while most of the world isn't linear, much of it can be approximated as linear within reasonable bounds. What makes this idea so powerful is the fact that linear systems are simple to compute and analyze.

The Mediterranean diet is one of the most famous dietary patterns in the world. It was introduced in 1975 by American biologist Ancel Keys and chemist Margaret Keys. It is based off the traditional foods eaten in Greece, Italy, Eastern Spain, and Southern France. It is based around fruits, vegetables, whole grains, legumes, seafood, and healthy fats like olive oil. There are studies linking it to reduced risk of heart disease, stroke, and diabetes. Let's model this through a breakdown of nutrients supplied by 100 grams of three staples of the Mediterranean diet:

Nutrient (per 100g)	**Farro**	**Olive Oil**	**Sardines**	**Daily Target (g)**
Protein	15	0	20	60
Carbohydrate	70	0	0	130
Fat	2	100	10	50

Table 5: Nutritional breakdown for farro, olive oil, and sardines

Example 2.8.1

Using Table 5, find some combination of farro, olive oil, and sardines that match the daily targets for nutrients.

Solution

Let x_1, x_2, and x_3 be the number of 100 g servings of farro, olive oil, and sardines, respectively.

For instance,

$$\left\{\begin{array}{c} x_1 \text{ units of} \\ \text{farro} \end{array}\right\} \cdot \left\{\begin{array}{c} \text{protein per unit} \\ \text{of farro} \end{array}\right\}$$

gives the amount of protein supplied by x_1 units of farro. To this amount, we would then add similar products for olive oil and sardines and set the resulting sum equal to the amount of protein we need. We would then repeat for each nutrient.

A more efficient method is to consider a "nutrient vector" for each food item and build a single vector equation. The amount of nutrients supplied by x_1 units of farrow is the scalar multiple

$$\underbrace{\left\{\begin{array}{c} x_1 \text{ units of} \\ \text{farro} \end{array}\right\}}_{\text{Scalar}} \cdot \underbrace{\left\{\begin{array}{c} \text{nutrients per unit} \\ \text{of farro} \end{array}\right\}}_{\text{Vector}} = x_1\mathbf{a}_1$$

where $\mathbf{a}_1$ is the first column in Table 5. Let $\mathbf{a}_2$ and $\mathbf{a}_3$ be the corresponding vectors for olive oil and sardines, respectively, and let $\mathbf{b}$ be the vector that lists the daily targets for total nutrients. Then $x_2\mathbf{a}_2$ and $x_3\mathbf{a}_3$ give the nutrients supplied by x_2 units of olive oil and x_3 units of sardines, respectively. We then combine everything to get

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = \mathbf{b}.$$

We encode our data as an augmented matrix and row reduce to RREF:

$$\begin{bmatrix} 15 & 0 & 20 & 60 \\ 70 & 0 & 0 & 130 \\ 2 & 100 & 10 & 50 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & \frac{13}{7} \\ 0 & 1 & 0 & \frac{155}{513} \\ 0 & 0 & 1 & \frac{45}{28} \end{bmatrix}$$

Thus, your daily nutrient targets could be met by $x_1 = 1.8571$ units of farro, $x_2 = 0.3021$ units of olive oil, and $x_3 = 1.6071$ units of sardines. Multiplying each of these by 100 yields 185.71 g of farro, 30.21 g of olive oil, and 160.71 g of sardines. While these numbers are reasonable, a realistic diet has more than three items and the model would thus have to factor in more food.

Each food's nutrient content can be expressed as a scalar multiple of a vector. Thus, the nutrients contributed by each food are proportional to the amount of each food item included in the total, which is acquired by adding together the contributions from each food item. This is exactly why it works as a linear system.

Current flow in a circuit is one of the most prominent applications of linear algebra in engineering. In a circuit, a battery or some other source of voltage V enables current to flow through components like resistors. By Ohm's law, the voltage drop across a resistor is $V = IR$ where R is resistance in ohms and I is current in amperes.

We can break a circuit up into loops that current flow around. The directions in which current flows is arbitrary. If current is negative, then the actual direction of current flow is opposite to what is depicted in the figure. In Figure 1, the current flowing in loops 1, 2, and 3 are denoted by I_1, I_2, and I_3, respectively.

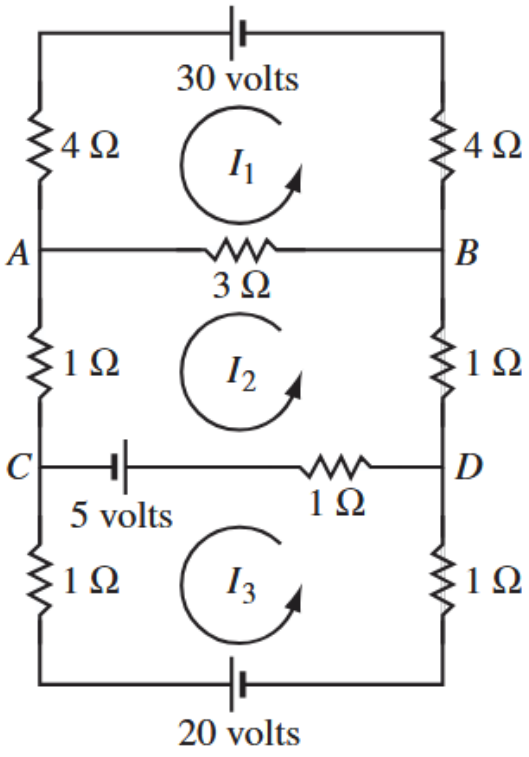


Figure 1: Image credit: Lay et al

Kirchhoff's law says that the sum of all voltage drop going around a closed loop is equal to the total voltage from all sources around that loop in the same direction. Using this, let's solve a circuit problem.

Example 2.8.2

Calculate the values of each loop current in Figure 1.

Solution

By Ohm's law, each resistor contributes a voltage drop IR.

For loop 1, we have

$$4I_1 + 4I_1 + 3I_1 = 11I_1.$$

Current from loop 2 also flows in part of loop 1. The associated voltage drop is $3I_2$ V. However, the direction for branch AB is opposite to loop 2's direction, so we have

$$11I_1 - 3I_2 = 30 \text{ V}$$

because the total votage in loop 1 is 30 volts.

For loop 2, we have

$$-3I_1 + 6I_2 - I_3 = 5.$$

$-3I_1$ comes from the flow of loop 1's current through AB, $6I_2$ is the sum of all resistance in loop 2 multiplied by the corresponding current, and $-I_3$ comes from loop 3's current flowing through the resistor in CD.

For loop 3, we have

$$-I_2 + 3I_3 = -25 \text{ V}.$$

We now write these as a linear system:

$$\begin{cases} 11I_1 - 3I_2 = 30 \\ -3I_1 + 6I_2 - I_3 = 5 \\ -I_2 + 3I_3 = -25 \end{cases}$$

We row reduce the augmented matrix:

$$\begin{bmatrix} 11 & -3 & 0 & 30 \\ -3 & 6 & -1 & 5 \\ 0 & -1 & 3 & -25 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -8 \end{bmatrix}$$

Our solution is $I_1 = 3$ A, $I_2 = 1$ A, and $I_3 = -8$ A.

We could have also written this system out in vector form:

$$I_1 \underbrace{\begin{bmatrix} 11 \\ -3 \\ 0 \end{bmatrix}}_{\mathbf{r}_1} + I_2 \underbrace{\begin{bmatrix} -3 \\ 6 \\ -1 \end{bmatrix}}_{\mathbf{r}_2} + I_3 \underbrace{\begin{bmatrix} 0 \\ -1 \\ 3 \end{bmatrix}}_{\mathbf{r}_3} = \underbrace{\begin{bmatrix} 30 \\ 5 \\ -25 \end{bmatrix}}_{\mathbf{v}} \tag{5}$$

Then we have Ohm's law:

$$R\mathbf{i} = \mathbf{v}$$

where $R = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 \end{bmatrix}$ and $\mathbf{i} = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix}$.

This is where the linearity of the system is most apparent. If the voltage vector is doubled, then the current vector will also be doubled. We also have a *superposition principle* here where the solution to (5) is the sum of the solutions to the following equations:

$$R\mathbf{i} = \begin{bmatrix} 30 \\ 0 \\ 0 \end{bmatrix}, \quad R\mathbf{i} = \begin{bmatrix} 0 \\ 5 \\ 0 \end{bmatrix}, \quad R\mathbf{i} = \begin{bmatrix} 0 \\ 0 \\ -25 \end{bmatrix}$$

Ohm's law and Kirchhoff's laws are linear. Voltage drop across a resistor is proportional to the current flowing through it and the sum of the voltage drops in a closed loop is equal to the sum of the voltage sources in that loop. This is why circuits work so well with linear algebra. In your physics and electrical engineering problems, you may find it easier at first to just not use linear algebra. But when it comes to more complex circuits, linear algebra will become necessary.

When systems change dynamically, we need to be able to take measurements at regular time intervals. This creates a sequence of vectors:

$$\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$$

Each vector records the *state* of the system at that moment. If there is a matrix A such that

$$\mathbf{x}_{k+1} = A\mathbf{x}_k$$

for $k = 0, 1, 2, \ldots$, then we have a **linear difference equation**. If you know A and $\mathbf{x}_0$, you can compute the state of the system at any given time. For instance, $\mathbf{x}_1 = A\mathbf{x}_0$.

Let's model the population of a city and its suburbs over time. Let

$$\mathbf{x}_0 = \begin{bmatrix} r_0 \\ s_0 \end{bmatrix}$$

be the population vector where r_0 is the city population and s_0 is the suburban population in one year. Then, the population in the following years is given by the following:

$$\mathbf{x}_1 = \begin{bmatrix} r_1 \\ s_1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} r_2 \\ s_2 \end{bmatrix}, \ldots$$

Let's try this model out. Suppose that 5% of the city's population moves to the suburbs every year (95% stays) while 3% of the suburban population moves to the city (97% stays). After the first year, we have the vector representing the city:

$$\begin{bmatrix} 0.95r_0 \\ 0.05r_0 \end{bmatrix} = r_0 \begin{bmatrix} 0.95 \\ 0.05 \end{bmatrix}.$$

Then, we have the vector for the suburbs:

$$s_0 \begin{bmatrix} 0.03 \\ 0.97 \end{bmatrix}$$

We can combine these to get the total population in the following year:

$$\begin{bmatrix} r_1 \\ s_1 \end{bmatrix} = r_0 \begin{bmatrix} 0.95 \\ 0.05 \end{bmatrix} + s_0 \begin{bmatrix} 0.03 \\ 0.97 \end{bmatrix} = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} r_0 \\ s_0 \end{bmatrix}$$

This can be written as

$$\mathbf{x}_1 = M\mathbf{x}_0.$$

where $M = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix}$. The next year would be given by

$$\mathbf{x}_2 = M\mathbf{x}_1.$$

In general,

$$\mathbf{x}_{k+1} = M\mathbf{x}_k$$

for $k = 0, 1, 2, \ldots, n$.

Problem 2.8.1

The population of the city in 2025 was 600000 and 400000 in the suburbs. Assuming the immigration numbers remain nearly constant, compute the population of the region in 2026 and 2027.

Solution

The initial population in 2025 is $\mathbf{x}_0 = \begin{bmatrix} 600000 \\ 400000 \end{bmatrix}$. For 2026 we have

$$\mathbf{x}_1 = M\mathbf{x}_0 = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} 600000 \\ 400000 \end{bmatrix} = \begin{bmatrix} 582000 \\ 418000 \end{bmatrix}.$$

For 2027 we have

$$\mathbf{x}_2 = M\mathbf{x}_1 = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} 582000 \\ 418000 \end{bmatrix} = \begin{bmatrix} 565440 \\ 434560 \end{bmatrix}.$$

Problem 2.8.2

Figure 2: Cheerios and Cinnamon Toast Crunch. Image credit: General Mills

The nutrition facts for two breakfast cereals are shown below. Suppose we want a mixture of the two that contains exactly 310 calories, 7 g protein, 62 g of carbohydrate, and 7 g fat.

(a) Set up a vector equation for this problem and include a statement of what the variables in your equation represent.

(b) Write an equivalent matrix equation and determine whether or not the mixture can be made.

(c) Suppose we we want a mixture of the two that has 790 calories, 16 g protein, 157 g of carbohydrate, and 18 g fat. Determine how many servings of both cereals we need.

Nutrient (per serving)	**Cheerios**	**Cinnamon Toast Crunch**
Calories	140	170
Protein (g)	5	2
Carbohydrate (g)	29	33
Fat (g)	3	4

Solution

(a) Let x_1 be servings of Cheerios and x_2 be servings of Cinnamon Toast Crunch.

The vector equation is

$$x_1 \begin{bmatrix} 140 \\ 5 \\ 29 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 170 \\ 2 \\ 33 \\ 4 \end{bmatrix} = \begin{bmatrix} 320 \\ 9 \\ 62 \\ 7 \end{bmatrix}.$$

(b) The matrix equation is

$$\begin{bmatrix} 140 & 170 \\ 5 & 2 \\ 29 & 33 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 310 \\ 7 \\ 62 \\ 7 \end{bmatrix}.$$

We row reduce the augmented matrix to RREF:

$$\begin{bmatrix} 140 & 170 & 310 \\ 5 & 2 & 7 \\ 29 & 33 & 62 \\ 3 & 4 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

This yields $x_1 = 1$ and $x_2 = 2$. Thus, it is possible to create the mixture with exactly 1 serving of both cereals.

(c)

$$\begin{bmatrix} 140 & 170 & 790 \\ 5 & 2 & 16 \\ 29 & 33 & 157 \\ 3 & 4 & 18 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

This yields $x_1 = 2$ and $x_2 = 3$. Thus, it is possible to create the mixture with 2 servings of Cheerios and 3 servings of Cinnamon Toast Crunch.

Problem 2.8.3

Given the following circuit, determine each of its loop currents:

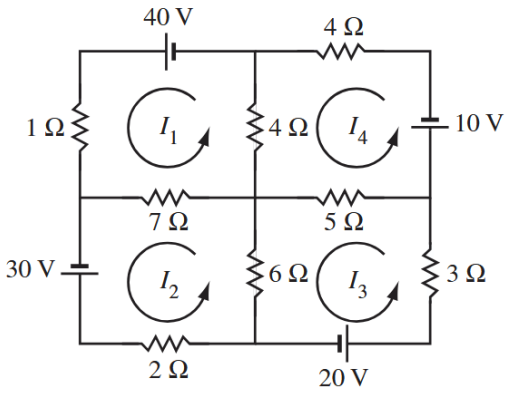


Image credit: Lay et al

Solution

We begin with loop 1. The total voltage drop is $(1 + 4 + 7)\Omega = 12\ \Omega$. It shares current with loop 2 that flows in the opposite direction thus we have -7 V. Current I_3 does not flow into loop 1. Current I_4 flows into loop 1 in the opposite direction thus we have -4 V. Putting all of these together, we have the resistance vector:

$$\mathbf{r}_1 = \begin{bmatrix} 12 \\ -7 \\ 0 \\ -4 \end{bmatrix}$$

We repeat the same process for the other loops to get

$$\mathbf{r}_2 = \begin{bmatrix} -7 \\ 15 \\ -6 \\ 0 \end{bmatrix}, \quad \mathbf{r}_3 = \begin{bmatrix} 0 \\ -6 \\ 14 \\ -5 \end{bmatrix}, \quad \mathbf{r}_4 = \begin{bmatrix} -4 \\ 0 \\ -5 \\ 13 \end{bmatrix}.$$

Then, we assemble the matrix:

$$R = \mathbf{r}_1 = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{r}_4 \end{bmatrix} = \begin{bmatrix} 12 & -7 & 0 & -4 \\ -7 & 15 & -6 & 0 \\ 0 & -6 & 14 & -5 \\ -4 & 0 & -5 & 13 \end{bmatrix}$$

Now we build the system fully by setting $R\mathbf{i} = \mathbf{v}$:

$$\begin{bmatrix} 12 & -7 & 0 & -4 \\ -7 & 15 & -6 & 0 \\ 0 & -6 & 14 & -5 \\ -4 & 0 & -5 & 13 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{bmatrix} = \begin{bmatrix} 40 \\ 30 \\ 20 \\ -10 \end{bmatrix}$$

And now we row reduce the augmented matrix:

$$\begin{bmatrix} 12 & -7 & 0 & -4 & 40 \\ -7 & 15 & -6 & 0 & 30 \\ 0 & -6 & 14 & -5 & 20 \\ -4 & 0 & -5 & 13 & -10 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 & 11.4342 \\ 0 & 1 & 0 & 0 & 10.5502 \\ 0 & 0 & 1 & 0 & 8.0357 \\ 0 & 0 & 0 & 1 & 5.8396 \end{bmatrix}$$

Thus $I_1 = 11.4342$ A, $I_2 = 10.5502$ A, $I_3 = 8.0357$ A, and $I_4 = 5.8396$ A.

2.9 Matrix Structure and Operations

When we can analyze the structure of matrices and perform operations on them, we improve our ability to solve systems of linear equations.

For $m, n \in \mathbb{N}$, a matrix $A \in \mathbb{R}^{m \times n}$ can be put into three categories:

A matrix is *tall and narrow* if $m > n$.

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ a_{2,1} & \cdots & a_{2,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix}$$

A matrix is *square* if $m = n$.

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix}$$

A matrix is *short and wide* if $m < n$.

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,m} & \cdots & a_{1,n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,m} & \cdots & a_{m,n} \end{bmatrix}$$

The **diagonal entries** of A are elements where we have $i = j$ for $a_{i,j}$.

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} & a_{1,5} & a_{1,6} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} & a_{2,5} & a_{2,6} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} & a_{3,5} & a_{3,6} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} & a_{4,5} & a_{4,6} \\ a_{5,1} & a_{5,2} & a_{5,3} & a_{5,4} & a_{5,5} & a_{5,6} \\ a_{6,1} & a_{6,2} & a_{6,3} & a_{6,4} & a_{6,5} & a_{6,6} \end{bmatrix}$$

All of the other entries are **nondiagonal entries** where $i \neq j$.

Let's introduce a specific case:

Diagonal Matrix

A matrix $A \in \mathbb{R}^{n \times n}$ is a **diagonal matrix** if $a_{i,j} = 0$ for all $i \neq j$. In other words, it is a square matrix whose nondiagonal entries are zero:

$$\begin{bmatrix} a_{1,1} & 0 & \cdots & 0 \\ 0 & a_{2,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_{n,n} \end{bmatrix}$$

One example is the identity matrix $I_n \in \mathbb{R}^{n \times n}$.

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

The lower triangular entries (pink) of a matrix are all entries $a_{i,j}$ with $i \geq j$.

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} & a_{1,5} & a_{1,6} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} & a_{2,5} & a_{2,6} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} & a_{3,5} & a_{3,6} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} & a_{4,5} & a_{4,6} \\ a_{5,1} & a_{5,2} & a_{5,3} & a_{5,4} & a_{5,5} & a_{5,6} \\ a_{6,1} & a_{6,2} & a_{6,3} & a_{6,4} & a_{6,5} & a_{6,6} \end{bmatrix}$$

Lower Triangular Matrix

A matrix $A \in \mathbb{R}^{n \times n}$ is a **lower triangular matrix** if $a_{i,j} = 0$ for all $i < j$. In other words, it is a square matrix where all entries above the main diagonal are zero:

$$A = \begin{bmatrix} a_{1,1} & 0 & \cdots & 0 \\ a_{2,1} & a_{2,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix}$$

The upper triangular entries (blue) of a matrix are all entries $a_{i,j}$ with $i \leq j$.

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} & a_{1,5} & a_{1,6} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} & a_{2,5} & a_{2,6} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} & a_{3,5} & a_{3,6} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} & a_{4,5} & a_{4,6} \\ a_{5,1} & a_{5,2} & a_{5,3} & a_{5,4} & a_{5,5} & a_{5,6} \\ a_{6,1} & a_{6,2} & a_{6,3} & a_{6,4} & a_{6,5} & a_{6,6} \end{bmatrix}$$

Upper Triangular Matrix

A matrix $A \in \mathbb{R}^{n \times n}$ is an **upper triangular matrix** if $a_{i,j} = 0$ for all $i > j$. In other words, it is a square matrix where all entries below the main diagonal are zero:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ 0 & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{n,n} \end{bmatrix}$$

Let's now move on to operations. We begin by formalizing some matrix arithmetic. We say that two matrices are equal if their dimensions and corresponding entries are equal. For matrix addition, you can only add matrices of the same size and you have to add them entry-by-entry.

Example 2.9.1

Let

$$A = \begin{bmatrix} 3 & 2 & 4 \\ 0 & 5 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 6 & -2 & 1 \\ 4 & 0 & 3 \end{bmatrix}, \quad C = \begin{bmatrix} 7 & 8 \\ 1 & -5 \end{bmatrix}.$$

(a) If possible, compute $A + B$.

(b) If possible, compute $A + C$.

Solution

(a) $A + B = \begin{bmatrix} 3+ & 6 & 2+(-2) & 4+ & 1 \\ 0+ & 4 & 5+ & 0 & -1+ & 3 \end{bmatrix} = \begin{bmatrix} 9 & 0 & 5 \\ 4 & 5 & 2 \end{bmatrix}$

(b) $A + C$ is not possible because A and C have different dimensions.

Scalar-matrix multiplication is pretty straightforward.

Properties of Scalar-Matrix Multiplication

Let A, B, and C be matrices of the same size, and let r and s be scalars.

(i) $A + B = B + A$

(ii) $(A + B) + C = A + (B + C)$

(iii) $A + 0 = A$

(iv) $r(A + B) = rA + rB$

(v) $(r + s)A = rA + sA$

(vi) $r(sA) = (rs)A$

We now move on to **matrix multiplication (matrix-matrix multiplication)**, which is one of the most important operations in modern day science and engineering.

If a matrix B is multiplied by a vector $\mathbf{x}$, it produces the transformed vector $B\mathbf{x}$. Then, multiplying this by another matrix A yields $A(B\mathbf{x})$. This is a *composition of mappings*. We can represent this as multiplication by a single matrix AB so that

$$A(B\mathbf{x}) = (AB)\mathbf{x}.$$

For $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$, and $\mathbf{x} \in \mathbb{R}^p$, we have the columns of B:

$$\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_p \end{bmatrix}$$

Then

$$B\mathbf{x} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots + x_p\mathbf{b}_p.$$

We now multiply by A. Because of the property of linearity,

$$\begin{aligned} A(B\mathbf{x}) &= A(x_1\mathbf{b}_1) + A(x_2\mathbf{b}_2) + \cdots + A(x_p\mathbf{b}_p) \\ &= x_1A\mathbf{b}_1 + x_2A\mathbf{b}_2 + \cdots + x_pA\mathbf{b}_p. \end{aligned}$$

The result is a linear combination of the vectors $A\mathbf{b}_1, \dots, A\mathbf{b}_p$ where the entries in $\mathbf{x}$ are weights. We could rewrite these vectors as the columns of a matrix whose columns are

$$A(B\mathbf{x}) = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \dots & A\mathbf{b}_p \end{bmatrix} \mathbf{x}.$$

Notice how multiplication by AB transformed $\mathbf{x}$ into $A(B\mathbf{x})$.

If A is an $m \times n$ matrix and B is an $n \times p$ matrix with columns $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_p$, then the product AB is an $m \times p$ matrix whose columns are $A\mathbf{b}_1, A\mathbf{b}_2, \dots, A\mathbf{b}_p$:

$$AB = A\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_p \end{bmatrix} = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_p \end{bmatrix}$$

Example 2.9.2

Compute AB where

$$A = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{bmatrix}.$$

Solution

We begin by writing B in terms of its column vectors. That is, $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 \end{bmatrix}$ where

$$\mathbf{b}_1 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}, \quad \mathbf{b}_3 = \begin{bmatrix} 6 \\ 3 \end{bmatrix}.$$

We compute each $A\mathbf{b}_j$ separately:

$$A\mathbf{b}_1 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 8+3 \\ 4-5 \end{bmatrix} = \begin{bmatrix} 11 \\ -1 \end{bmatrix}$$

$$A\mathbf{b}_2 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 6-6 \\ 3+10 \end{bmatrix} = \begin{bmatrix} 0 \\ 13 \end{bmatrix}$$

$$A\mathbf{b}_3 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 6 \\ 3 \end{bmatrix} = \begin{bmatrix} 12+9 \\ 6-15 \end{bmatrix} = \begin{bmatrix} 21 \\ -9 \end{bmatrix}$$

Finally, we form AB by writing these as the columns of AB:

$$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & A\mathbf{b}_3 \end{bmatrix} = \begin{bmatrix} 11 & 0 & 21 \\ -1 & 13 & -9 \end{bmatrix}.$$

Notice that each column $A\mathbf{b}_j$ of AB is a linear combination of the columns of A that uses the corresponding entries in b_j as weights.

For two matrices A and B, the product AB is defined only if the number of columns of A equal the number of rows of B.

$$A_{m\times n}B_{p\times q}$$

is only defined when $n = p$.

In other words, the inner dimensions must agree!

Problem 2.9.1

We have $A \in \mathbb{R}^{3\times 5}$ and $B \in \mathbb{R}^{5\times 2}$.

(a) Is AB defined? If so, what size is it?

(b) Is BA defined? If so, what size is it?

Solution

(a) AB is defined because A has 5 columns and B has 5 rows. The product is a 3×2 matrix.

(b) BA is not defined because A has 3 rows and B has 2 columns.

We now have more practical ways to calculate entries or rows in AB:

Row-Column Method for Computing AB

Let $A \in \mathbb{R}^{m \times n}$. If the product AB is defined, then the entry in row i and column j of AB is the sum of the products of corresponding entries from row i of A and column j of B:

$$(AB)_{i,j} = a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \cdots + a_{i,n}b_{n,j}$$

Matrix Multiplication via Linear Combination of Rows

Let $A \in \mathbb{R}^{m \times p}$ and $B \in \mathbb{R}^{p \times n}$. If the product AB is defined such that $AB = C$, then the ith row of C is given by

$$C_{i,j} = \sum_{k=1}^{p} A_{i,k}B_{k,j}.$$

In other words, the ith row of C is equal to the ith row of A multiplied by B. Equivalently, each entry of $C = AB$ is the dot product of the ith row of A and the jth column of B.

Example 2.9.3

Let $A = \begin{bmatrix} 2 & -5 & 0 \\ -1 & 3 & -4 \\ 6 & -8 & -7 \\ -3 & 0 & -9 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & -6 \\ 7 & 1 \\ 3 & 2 \end{bmatrix}$. Find the second row of AB.

Solution

The second row of AB comes from row 2 of A and the columns of B. We multiply row 2 of A which is $\begin{bmatrix}-1 & 3 & -4\end{bmatrix}$ by the first column of B:

$$\begin{bmatrix}-1 & 3 & -4\end{bmatrix}\begin{bmatrix}4\\7\\3\end{bmatrix} = -4 + 21 - 12 = 5$$

And then we multiply row 2 of A by the second column of B:

$$\begin{bmatrix}-1 & 3 & -4\end{bmatrix}\begin{bmatrix}-6\\1\\2\end{bmatrix} = 6 + 3 - 8 = 1$$

And then together we have row 2 of AB:

$$\begin{bmatrix} 5 & 1 \end{bmatrix}$$

We now will summarize the properties of matrix multiplication. Recall that I_m is the identity matrix where $I_m\mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^m$.

Algebraic Properties of Matrix Multiplication

Let A, B, and C be matrices for which the following operations are defined. Let $\alpha \in \mathbb{R}$ be any scalar. Then all of the following hold:

1. **Associativity:** $(AB)C = A(BC)$
2. **Left distributivity:** $A(B \pm C) = AB \pm AC$
3. **Right distributivity:** $(A \pm B)C = AC \pm BC$
4. **Identity matrix:** $I_m A = A = AI_n$
5. **Zero matrix:** $A0 = 0 = 0A$
6. **Scalar multiplication:** $(\alpha A)B = A(\alpha B) = \alpha(AB)$

A matrix product can be grouped in any way. For instance, $ABCD$ can be computed as $A(BCD)$, $(ABC)D$, $A(BC)D$ and so on. However, the left-to-right order of the matrices does matter because AB is usually not equivalent to BA. That is, matrix multiplication is not commutative. This makes sense because the columns of AB are linear combinations of A whereas the columns of BA are constructed from the columns of B. We can denote this by saying that A is *right-multiplied* by B or that B is *left-multiplied* by A.

In the case that $AB = BA$, we say that A and B *commute* with one another.

Problem 2.9.2

Let $A = \begin{bmatrix} 5 & 1 \\ 3 & -2 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix}$. Determine whether or not these two matrices commute.

Solution

$$AB = \begin{bmatrix} 5 & 1 \\ 3 & -2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 14 & 3 \\ -2 & -6 \end{bmatrix}$$

$$BA = \begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 5 & 1 \\ 3 & -2 \end{bmatrix} = \begin{bmatrix} 10 & 2 \\ 29 & -2 \end{bmatrix}$$

Since $AB \neq BA$, these matrices do not commute.

One case where matrices *do* commute are when two diagonal matrices are multiplied together. Suppose $A = \begin{bmatrix} 5 & 0 \\ 0 & 7 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$. Then,

$$AB = \begin{bmatrix} 20 & 0 \\ 0 & 14 \end{bmatrix} = BA.$$

That being said, remember that order generally matters.

Another thing to keep in mind is that the cancellation laws generally do not hold for matrix multiplication. If $AB = AC$, you cannot always conclude that $B = C$.

Lastly, a zero product doesn't always mean a zero factor. If $AB = 0$, you cannot always conclude that $A = 0$ or $B = 0$.

Let's now discuss the powers of a matrix. Suppose $A \in \mathbb{R}^{n \times n}$ and let k be a positive integer. Then, A^k means multiplying A by itself k times:

$$A^k = A \cdot A \cdot \ldots \cdot A$$

If A is nonero and $\mathbf{x} \in \mathbb{R}^n$, then $A^k\mathbf{x}$ means you are left-multiplying $\mathbf{x}$ by A for a total of k times. If $k = 0$, then $A^0\mathbf{x} = \mathbf{x}$. Indeed, A^0 is the identity matrix I_n.

We can also take the transpose of a matrix by flipping it over its diagonal. This is done by swapping the rows and the columns. If $A \in \mathbb{R}^{m \times n}$, then $A^T \in \mathbb{R}^{n \times m}$.

Example 2.9.4

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Find A^T.

Solution

$$A^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

Problem 2.9.3

Let $A = \begin{bmatrix} -5 & 2 \\ 1 & -3 \\ 0 & 4 \end{bmatrix}$. Find A^T.

Solution

$$A^T = \begin{bmatrix} -5 & 1 & 0 \\ 2 & -3 & 4 \end{bmatrix}$$

Properties of the Transpose of a Matrix

Let A and B be matrices for which the following operations are defined. Let $r \in \mathbb{R}$ be any scalar. Then all of the following hold:

1. $(A^T)^T = A$
2. $(A + B)^T = A^T + B^T$
3. $(rA)^T = rA^T$
4. $(AB)^T = B^T A^T$

Problem 2.9.4

(a) Show that $I_m A = A$ where $A \in \mathbb{R}^{m \times n}$. Assume that $I_m \mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^m$.

(b) Show that $AI_n = A$ where $A \in \mathbb{R}^{m \times n}$.

Solution

(a) We apply I_m to A column by column:

$$I_m A = I_m \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} I_m\mathbf{a}_1 & I_m\mathbf{a}_2 & \cdots & I_m\mathbf{a}_n \end{bmatrix}$$

Using $I_m\mathbf{a}_j = \mathbf{a}_j$ for each j, we obtain

$$I_m A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} = A.$$

(b)

$$AI_n = A \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{e}_1 & A\mathbf{e}_2 & \cdots & A\mathbf{e}_n \end{bmatrix}.$$

Since $A\mathbf{e}_j$ is exactly the jth column of A, it follows that

$$AI_n = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} = A.$$

Problem 2.9.5

Give a formula for $(AB\mathbf{x})^T$, where $\mathbf{x}$ is a vector and A and B are matrices of appropriate sizes.

Solution

First, group $AB\mathbf{x}$ as

$$(AB\mathbf{x})^T = (A(B\mathbf{x}))^T .$$

Apply the transpose rule:

$$(A(B\mathbf{x}))^T = (B\mathbf{x})^T A^T$$

We then apply the transpose rule again to $(B\mathbf{x})^T$ and it becomes $\mathbf{x}^T B^T$.

Thus we have

$$(A(B\mathbf{x}))^T = (AB\mathbf{x})^T = \mathbf{x}^T B^T A^T.$$

2.10 Matrix Inverses

Before we define the inverse of a matrix, recall what the multiplicative inverse of a number is. For a nonzero scalar α, its inverse α^{-1} or $\frac{1}{\alpha}$ is defined by

$$\alpha^{-1}\alpha = 1 \quad \text{and} \quad \alpha\alpha^{-1} = 1.$$

For example,

$$5^{-1} \cdot 5 = 1 \quad \text{and} \quad 5 \cdot 5^{-1} = 1.$$

We now extend this to matrices:

Inverse of a Matrix

A matrix $A \in \mathbb{R}^{n\times n}$ is *invertible* if and only if there exists a matrix $C \in \mathbb{R}^{n\times n}$ such that

$$AC = CA = I_n$$

where $I_n \in \mathbb{R}^{n\times n}$ is the identity matrix. We say that C is the inverse of A.

C is uniquely determined by A. Suppose there was $B \in \mathbb{R}^{n\times n}$, another inverse of A. Then,

$$B = BI_n = B(AC) = (BA)C = I_nC = C.$$

The *unique* inverse is denoted by A^{-1} such that

$$A^{-1}A = I \quad \text{and} \quad AA^{-1} = I.$$

A matrix that *is* invertible is called a *nonsingular* matrix. A matrix that is *not* invertible is called a *singular* matrix.

Example 2.10.1

We have the matrices $A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$. Is C the inverse of A?

Solution

$$AC = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_n$$

Therefore C is the inverse of A.

We now have a simple formula for determining if the inverse of a square matrix exists:

Rule for the Inverse Matrix

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. If $ad - bc \neq 0$, then A is invertible and

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

If $ad - bc = 0$, then A is not invertible.

Proof of the Rule for the Inverse Matrix

We must transform A into I_n using elementary matrices. For this purpose, we will denote A as

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix}.$$

Then, we will let $ad - bc = \det(A)$ and assume that $a_{1,1} \neq 0$. The **determinant** is a function we will discuss later. A square matrix is invertible if and only if $\det(A) \neq 0$.

The **elementary matrix** will be denoted by S. An elementary matrix is obtained by performing a single row operation on the identity matrix. When we use an elementary matrix E to perform a row operation on A, multiplying by E^{-1} undoes that row operation.

Then the product is given by

$$S_{2,1}\left(-\frac{a_{2,1}}{a_{1,1}}\right) A = \begin{bmatrix} 1 & 0 \\ -\frac{a_{2,1}}{a_{1,1}} & 1 \end{bmatrix} \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ 0 & \frac{\det(A)}{a_{1,1}} \end{bmatrix}$$

We will denote diagonal matrices by D. Then the diagonal matrix is given by

$$D_2\left(\frac{a_{1,1}}{\det(A)}\right) \cdot D_1\left(\frac{1}{a_{1,1}}\right) = \begin{bmatrix} \frac{1}{a_{1,1}} & 0 \\ 0 & \frac{a_{1,1}}{\det(A)} \end{bmatrix}$$

where D_1 scales row 1 so that its corresponding pivot becomes 1 and D_2 scales row 2 so that its corresponding pivot becomes 1.

We will multiply our product by this diagonal matrix:

$$\begin{bmatrix} \frac{1}{a_{1,1}} & 0 \\ 0 & \frac{a_{1,1}}{\det(A)} \end{bmatrix} \begin{bmatrix} a_{1,1} & a_{1,2} \\ 0 & \frac{\det(A)}{a_{1,1}} \end{bmatrix} = \begin{bmatrix} 1 & -\frac{a_{2,1}}{a_{1,1}} \\ 0 & 1 \end{bmatrix}$$

We now compute $A^{-1}A$

$$\begin{bmatrix} 1 & -\frac{a_{2,1}}{a_{1,1}} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{a_{2,1}}{a_{1,1}} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

and see that $A^{-1}A = I_n$ with

$$A^{-1} = S_{1,2}\left(-\frac{a_{2,1}}{a_{1,1}}\right) \cdot D_2\left(\frac{a_{1,1}}{\det(A)}\right) \cdot D_1\left(\frac{1}{a_{1,1}}\right) \cdot S_{2,1}\left(-\frac{a_{2,1}}{a_{1,1}}\right)$$

We can explicitly calculate A^{-1} by reducing A into I_n. We begin by finding the product

$$\left(D_2\left(\frac{a_{1,1}}{\det(A)}\right) \cdot D_1\left(\frac{1}{a_{1,1}}\right)\right) \cdot S_{2,1}\left(-\frac{a_{2,1}}{a_{1,1}}\right).$$

This is given as

$$\begin{bmatrix} \frac{1}{a_{1,1}} & 0 \\ 0 & \frac{a_{1,1}}{\det(A)} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\frac{a_{2,1}}{a_{1,1}} & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{a_{1,1}} & 0 \\ -\frac{a_{2,1}}{\det(A)} & \frac{a_{1,1}}{\det(A)} \end{bmatrix}.$$

We can calculate A^{-1} using the product:

$$\begin{bmatrix} 1 & -\frac{a_{1,2}}{a_{1,1}} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{a_{1,1}} & 0 \\ -\frac{a_{2,1}}{\det(A)} & \frac{a_{1,1}}{\det(A)} \end{bmatrix} = \begin{bmatrix} \frac{1}{a_{1,1}} + \frac{a_{2,1}a_{1,2}}{a_{1,1}\det(A)} & -\frac{a_{1,2}}{a_{1,1}} + \frac{a_{1,1}}{\det(A)} \\ -\frac{a_{2,1}}{\det(A)} & \frac{a_{1,1}}{\det(A)} \end{bmatrix}$$

Simplifying, we have

$$\begin{bmatrix} \frac{a_{2,2}}{\det(A)} & -\frac{a_{1,2}}{\det(A)} \\ -\frac{a_{2,1}}{\det(A)} & \frac{a_{1,1}}{\det(A)} \end{bmatrix} = \frac{1}{\det(A)} \begin{bmatrix} a_{2,2} & -a_{1,2} \\ -a_{2,1} & a_{1,1} \end{bmatrix}.$$

□

There is a connection between invertible matrices and row operations that leads to methods for more efficient computation. An invertible matrix A is row equivalent to an identity matrix, and we can find A^{-1} through the row reduction of A to I

Example 2.10.2

Find the inverse of $A = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}$.

Solution

Since $\det(A) = 3(6) - 4(5) = 18 - 20 = -2 \neq 0$, A is invertible. We now compute the inverse:

$$A^{-1} = \frac{1}{-2}\begin{bmatrix} 6 & -4 \\ -5 & 3 \end{bmatrix} = \begin{bmatrix} \frac{6}{-2} & \frac{-4}{-2} \\ \frac{-5}{-2} & \frac{3}{-2} \end{bmatrix} = \begin{bmatrix} -3 & 2 \\ \frac{5}{2} & -\frac{3}{2} \end{bmatrix}.$$

Inverse matrices can be quite useful when it comes to computation in linear algebra.

If A is an invertible $n \times n$ matrix, then for each $\mathbf{b} \in \mathbb{R}^n$, the equation

$$A\mathbf{x} = \mathbf{b}$$

has the unique solution

$$\mathbf{x} = A^{-1}\mathbf{b}.$$

Proof

Take any $\mathbf{b} \in \mathbb{R}^n$. A solution exists because if $A^{-1}\mathbf{b}$ is substituted for $\mathbf{x}$, then

$$A\mathbf{x} = A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I_n\mathbf{b} = \mathbf{b}.$$

So $\mathbf{x} = A^{-1}\mathbf{b}$ is a solution.

To prove uniqueness, suppose $\mathbf{u}$ is any solution of $A\mathbf{u} = \mathbf{b}$. Multiplying both sides by A^{-1} gives

$$A^{-1}A\mathbf{u} = A^{-1}\mathbf{b}, \quad I_n\mathbf{u} = A^{-1}\mathbf{b}, \quad \mathbf{u} = A^{-1}\mathbf{b}.$$

Thus the solution is unique. □

Example 2.10.3

A horizontal elastic beam is supported at each end and subjected to forces at points $1, 2$, and 3. Let $\mathbf{f} \in \mathbb{R}^3$ be the vector of forces at these points and let $\mathbf{y} \in \mathbb{R}^3$ be the vector of deflections (displacements) of the beam at the three points. Using Hooke's law, we can model the situation as

$$\mathbf{y} = D\mathbf{f},$$

where D is called the **flexibility matrix** of the beam. The inverse D^{-1} is called the **stiffness matrix**. Describe what the columns of D and D^{-1} represent physically.

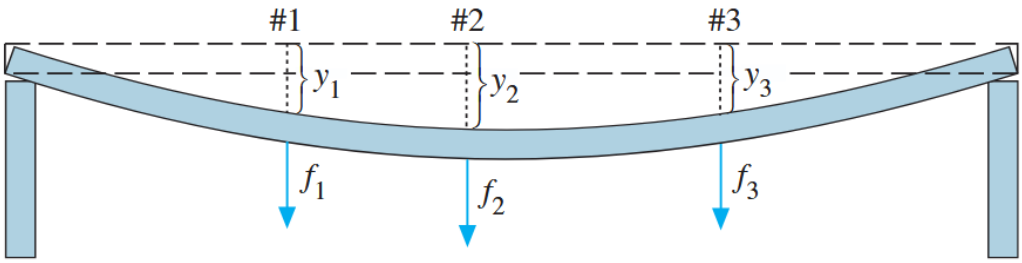


Image credit: Lay et al

Solution

The matrix D tells us how much the beam will deflect at each point for given forces applied at the points. Write the 3×3 identity matrix as

$$I_3 = [\mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{e}_3],$$

where each $\mathbf{e}_j \in \mathbb{R}^3$ is a unit force applied at the corresponding point on the beam with zero force at the other two points. Then

$$D = DI_3 = [D\mathbf{e}_1 \quad D\mathbf{e}_2 \quad D\mathbf{e}_3].$$

Here, $D\mathbf{e}_1$ is the first column of D, showing the beam deflections when we apply a unit force at point 1 and no force at points 2 or 3. $D\mathbf{e}_2$ is the second column, showing deflections from a unit force at point 2. $D\mathbf{e}_3$ is the third column, showing deflections from a unit force at point 3.

For the stiffness matrix D^{-1}, notice that $\mathbf{f} = D^{-1}\mathbf{y}$ computes a force vector $\mathbf{f}$. We compute

$$D^{-1} = D^{-1}I_3 = [D^{-1}\mathbf{e}_1 \quad D^{-1}\mathbf{e}_2 \quad D^{-1}\mathbf{e}_3].$$

Here, $D^{-1}\mathbf{e}_1$ is the force vector needed to produce a unit deflection at point 1 and zero deflections at the other points. $D^{-1}\mathbf{e}_2$ and$D^{-1}\mathbf{e}_3$ list the same at their corresponding points.

Because the beam is supported at both ends, some forces in $D^{-1}\mathbf{e}_j$ will be negative, meaning they push upward, to produce a unit deflection at the desired point and zero deflections at the other two points. Thus, the columns of D describe how the beam responds (deflects) to forces and the columns of D^{-1} describe the forces required to create specific deflections.

If flexibility is measured in inches of deflection per pound of load, then stiffness entries are in pounds of load per inch of deflection.

Example 2.10.4

Use the inverse of the matrix A from Example 2.10.2 to solve the following system:

$$\begin{cases} 3x_1 + 4x_2 = 3 \\ 5x_1 + 6x_2 = 7 \end{cases}$$

Solution

We can write the system as $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}.$$

We already know that

$$A^{-1} = \begin{bmatrix} -3 & 2 \\ \frac{5}{2} & -\frac{3}{2} \end{bmatrix}.$$

Thus

$$\mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} -3 & 2 \\ \frac{5}{2} & -\frac{3}{2} \end{bmatrix} \begin{bmatrix} 3 \\ 7 \end{bmatrix} = \begin{bmatrix} -3(3) + 2(7) \\ \frac{5}{2}(3) - \frac{3}{2}(7) \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}.$$

Therefore, the solution is $x_1 = 5$ and $x_2 = -3$.

We now introduce some properties of invertible matrices:

Properties of Inverse Matrices

Let A, $N \in \mathbb{R}^{n \times n}$ invertible matrices. Then the following properties hold:

(a) A^{-1} is itself invertible, and $(A^{-1})^{-1} = A$.

(b) The product AB is invertible, and $(AB)^{-1} = B^{-1}A^{-1}$.

(c) The transpose A^T is invertible, and $(A^T)^{-1} = (A^{-1})^T$.

To generalize (b), we say that the product of square invertible matrices is invertible, and the inverse is the product of their inverses in the reverse order.

We now will formally go over **elementary matrices**. Recall that an elementary matrix is obtained by performing a single elementary row operation on an identity matrix.

Example 2.10.5

Let

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix},$$

and

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}.$$

Compute E_1A, E_2A, and E_3A, and describe how these products can be obtained via elementary row operations on A.

Solution

$$E_1A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g-4a & h-4b & i-4c \end{bmatrix}$$

$$E_2A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} d & e & f \\ a & b & c \\ g & h & i \end{bmatrix}$$

$$E_3A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 5g & 5h & 5i \end{bmatrix}$$

E_1A corresponds to adding -4 multiplied by row 1 of A to row 3, which is row replacement. E_2A corresponds to interchanging rows 1 and 2 of A, which is a row swap operation. E_3A corresponds to multiplying row 3 of A by 5, which is a row scaling operation.

This leads to some important statements.

When you perform an elementary row operation on an $m \times n$ matrix A, the result can be expressed as EA where E is an $m \times m$ *elementary matrix* obtained by performing the same row operation on the identity matrix I_m. In other words, after performing the elementary row operation, multiplying A on the left by E applies that row operation to A.

Every elementary matrix E is invertible. Its inverse E^{-1} is the same type of elementary matrix that undoes the original row operation and transforms E back into I_m.

Let's take a look at how we can undo elementary matrices:

Inverses of Elementary Matrices

Let $i, k, n \in \mathbb{N}$ such that $1 \leq i$ and $k \leq n$. There are three types of elementary matrices and the inverses of each are given below:

1. **Shear matrices** (row replacement): $S_{i,k}(c)$ means "add c times row k to row i." To undo it, just add $-c$ multiplied by row k to row i:

$$(S_{i,k}(c))^{-1} = S_{i,k}(-c)$$

2. **Transposition matrices** (row swap): $P_{i,k}$ means "swap row i with row k." Doing it again swaps them back:

$$(P_{i,k})^{-1} = P_{i,k}^{T}$$

3. **Dilation matrices** (row scaling): $D_i(c)$ means "multiply row i by c." To undo it, multiply that row by $\frac{1}{c}$:

$$(D_i(c))^{-1} = D_i\left(\frac{1}{c}\right)$$

In short, to reverse a row operation, do the same type of operation but with the opposite effect: negate c for row replacements, swap back for transpositions, and take the reciprocal for row scalings.

Example 2.10.6

Find the inverse of $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}$.

Solution

To transform E_1 into I_3, add 4 multiplied by row 1 to row 3:

$$E_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix}.$$

This leads to another statement about finding the inverse of a matrix:

An $n \times n$ matrix A is invertible if and only if A is row equivalent to I_n. In this case, any sequence of elementary row operations that reduces A to I_n will also transform I_n into A^{-1}.

If we form an augmented matrix, then row operations on it result in the same operations on A and I. If there aren't row operations that transform A to I_n and I_n to A^{-1}, then A is not invertible.

Algorithm for Finding A^{-1}

To find A^{-1}, row reduce the augmented matrix $[A\ I_n]$. If A is row equivalent to I_n, then

$$\begin{bmatrix} A & I_n \end{bmatrix} \sim \begin{bmatrix} I_n & A^{-1} \end{bmatrix}.$$

If A is not row equivalent to I_n, then A is not invertible.

Example 2.10.7

Find the inverse of $A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix}$.

Solution

$$\begin{bmatrix} A & I_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & -\frac{9}{2} & 7 & -\frac{3}{2} \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & \frac{3}{2} & -2 & \frac{1}{2} \end{bmatrix}.$$

Since $A \sim I_3$ and

$$A^{-1} = \begin{bmatrix} -\frac{9}{2} & 7 & -\frac{3}{2} \\ -2 & 4 & -1 \\ \frac{3}{2} & -2 & \frac{1}{2} \end{bmatrix},$$

we can conclude that A is invertible. We can check by computing

$$AA^{-1} = I_3 = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix} \begin{bmatrix} -\frac{9}{2} & 7 & -\frac{3}{2} \\ -2 & 4 & -1 \\ \frac{3}{2} & -2 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Let's quickly go over an alternative perspective. When you find A^{-1} by row reducing $\begin{bmatrix} A & I_n \end{bmatrix}$ to $\begin{bmatrix} A & I_n A^{-1} \end{bmatrix}$, we are really just solving n separate systems of equations all at once:

$$A\mathbf{x} = \mathbf{e}_1, \quad A\mathbf{x} = \mathbf{e}_2, \quad \ldots, \quad A\mathbf{x} = \mathbf{e}_n.$$

In general, the jth system is

$$A\mathbf{x} = \mathbf{e}_j$$

where $\mathbf{e}_j$ is the jth column of the identity matrix. These "augmented" columns can all be placed next to A to form

$$\begin{bmatrix} A & \mathbf{e}_1 & \mathbf{e}_2 & \ldots \mathbf{e}_n \end{bmatrix} = \begin{bmatrix} A & I_n \end{bmatrix}.$$

Each column of A^{-1} is the solution to one of these corresponding systems.

In real-world computation, A^{-1} is not computed unless you need the actual entries of A^{-1}. Solving $A\mathbf{x} = \mathbf{b}$ by computing both A^{-1} and $A^{-1}\mathbf{b}$ typically takes about three times more operations than solving directly by row reduction.

We can now unite everything we have learned. We begin with the following theorem:

The Invertible Matrix Theorem

Let $A \in \mathbb{R}^{n\times n}$. Then following statements are equivalent. That is, for a given A, either all are true or all are false:

(a) A is an invertible matrix.

(b) A is row equivalent to I_n.

(c) A has n pivot positions.

(d) The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(e) The columns of A form a linearly independent set.

(f) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.

(g) The equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b} \in \mathbb{R}^n$.

(h) The columns of A span $\mathbb{R}^n$.

(i) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.

(j) There exists an $n \times n$ matrix C such that $CA = I$.

(k) There exists an $n \times n$ matrix D such that $AD = I$.

(l) A^T is an invertible matrix.

Example 2.10.8

Use the invertible matrix theorem to decide if

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 3 & 1 & -2 \\ -5 & -1 & 9 \end{bmatrix}$$

is invertible.

Solution

Row reduce A:

$$\begin{bmatrix} 1 & 0 & -2 \\ 3 & 1 & -2 \\ -5 & -1 & 9 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 4 \\ 0 & -1 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 4 \\ 0 & 0 & 3 \end{bmatrix}.$$

A has 3 pivot positions, so by the invertible matrix theorem, A is invertible.

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is said to be invertible if there exists a function $S : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$S(T(\mathbf{x})) = \mathbf{x} \tag{6}$$

and

$$T(S(\mathbf{x})) = \mathbf{x} \tag{7}$$

for all $\mathbf{x} \in \mathbb{R}^n$.

The function S is called the inverse of T and is denoted T^{-1}. If such an S exists, it is unique and must itself be a linear transformation.

Relation Between T^{-1} and A^{-1}

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation and let A be its standard matrix. Then T is invertible if and only if A is an invertible matrix. In this case, the inverse transformation S is given by

$$S(\mathbf{x}) = A^{-1}\mathbf{x},$$

which is the unique function satisfying equations (6) and (7).

Problem 2.10.1

Use the invertible matrix theorem to answer the following.

(a) What can you say about a one-to-one linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$?

(b) Suppose that for a certain $n \times n$ matrix A, $A\mathbf{x} = \mathbf{b}$ does not have at least one solution for each $\mathbf{b} \in \mathbb{R}^n$. What can you say about $A\mathbf{x} = \mathbf{b}$?

(c) Suppose that A and B are $n \times n$ matrices and the equation $AB\mathbf{x} = \mathbf{0}$ has a nontrivial solution. What can you say about the matrix AB?

Solution

(a) The columns of the standard matrix A of T are linearly independent. A is invertible, so T maps $\mathbb{R}^n$ onto $\mathbb{R}^n$ and is invertible.

(b) Then $A\mathbf{x} = \mathbf{b}$ does not have a solution for at least one $\mathbf{b} \in \mathbb{R}^n$. And thus A is not invertible.

(c) If $AB\mathbf{x} = \mathbf{0}$ has a nontrivial solution, there exists some $\mathbf{x} \neq \mathbf{0}$ such that $AB\mathbf{x} = \mathbf{0}$. This means the equation $AB\mathbf{x} = \mathbf{0}$ does not have only the trivial solution. Therefore, AB is not invertible.

2.11 Block Matrices

As you know, matrices are powerful tools for encoding data in a compact way. To further build, we will now introduce notation that can highlight structure in linear systems. If we have a matrix and divide it up into sections, we call these new rectangular subregions **blocks**. And they function as **submatrices**.

Suppose we have the following matrix:

$$A = \begin{bmatrix} 3 & 0 & -1 & 5 & 9 & -2 \\ -5 & 2 & 4 & 0 & -3 & 1 \\ -8 & -6 & 3 & 1 & 7 & -4 \end{bmatrix}$$

We can use lines to turn it into a block matrix:

$$A = \left[\begin{array}{ccc|c|cc} 3 & 0 & -1 & 5 & 9 & -2 \\ -5 & 2 & 4 & 0 & -3 & 1 \\ \hline -8 & -6 & 3 & 1 & 7 & -4 \end{array}\right]$$

Thus we have

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} \\ A_{2,1} & A_{2,2} & A_{2,3} \end{bmatrix}$$

where each $A_{i,j}$ is a block or submatrix of A:

$$A_{1,1} = \begin{bmatrix} 3 & 0 & -1 \\ -5 & 2 & 4 \end{bmatrix}, \qquad A_{1,2} = \begin{bmatrix} 5 & 9 \\ 0 & -3 \end{bmatrix}, \qquad A_{1,3} = \begin{bmatrix} -2 \\ 1 \end{bmatrix},$$

$$A_{2,1} = \begin{bmatrix} -8 & -6 & 3 \end{bmatrix}, \qquad A_{2,2} = \begin{bmatrix} 1 & 7 \end{bmatrix}, \qquad A_{2,3} = \begin{bmatrix} -4 \end{bmatrix}.$$

Block matrices are especially useful in power systems, signal processing, and machine learning. You could, for instance, divide a microprocessor up into the number of microchips it has.

When it comes to operations with block matrices, please note that they are largely for humans to understand. The actual matrix still remains the same and calculations on it as a whole do not change. However, you can perform operations *block by block*.

If two matrices A and B are the same size and are partitioned in exactly the same way, then the sum $A + B$ is obtained by adding corresponding blocks. Similarly, multiplying a matrix by a scalar k is done by multiplying each block by k.

Provided that the block sizes are compatible, partitioned matrices can be multiplied using the usual row–column rule where you treat each block as if it were a single matrix entry. For AB to be defined in block form, the *column partition* of A must match the *row partition* of B.

If A is partitioned as $\begin{bmatrix} A_{i,j} \end{bmatrix}$ and B as $\begin{bmatrix} B_{i,j} \end{bmatrix}$, the block product is

$$C_{i,j} = \sum_{k=1}^{q} A_{i,k} B_{k,j}$$

where q is the number of block columns in A and the number of block rows in B.

Example 2.11.1

Let

$$A = \left[\begin{array}{ccc|cc} 2 & -3 & 1 & 0 & -4 \\ 1 & 5 & -2 & 3 & -1 \\ \hline 0 & -4 & -2 & 7 & -1 \end{array}\right] = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} \quad \text{and} \quad B = \left[\begin{array}{cc} 6 & 4 \\ -2 & 1 \\ -3 & 7 \\ \hline -1 & 3 \\ 5 & 2 \end{array}\right] = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$$

Compute the top and bottom blocks of AB.

Solution

Here, the 5 columns of A are partitioned into a group of 3 columns and then a set of 2 columns. The 5 rows of B are split into 3 rows and 2 rows. Since they are partitioned the same way, we say that the partitions are *comfortable* for block multiplication.

The block multiplication rule gives:

$$AB = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} A_{1,1}B_1 + A_{1,2}B_2 \\ A_{2,1}B_1 + A_{2,2}B_2 \end{bmatrix}.$$

In each term, the block from A must be on the left because matrix multiplication is not commutative.

We first compute the top block:

$$A_{1,1}B_1 = \begin{bmatrix} 2 & -3 & 1 \\ 1 & 5 & -2 \end{bmatrix} \begin{bmatrix} 6 & 4 \\ -2 & 1 \\ -3 & 7 \end{bmatrix} = \begin{bmatrix} 15 & 12 \\ 2 & -5 \end{bmatrix} \quad \text{and} \quad A_{1,2}B_2 = \begin{bmatrix} 0 & -4 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} -1 & 3 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} -20 & -8 \\ -8 & 7 \end{bmatrix}$$

Then

$$A_{1,1}B_1 + A_{1,2}B_2 = \begin{bmatrix} 15 & 12 \\ 2 & -5 \end{bmatrix} + \begin{bmatrix} -20 & -8 \\ -8 & 7 \end{bmatrix} = \begin{bmatrix} -5 & 4 \\ -6 & 2 \end{bmatrix}.$$

$$A_{2,1}B_1 = \begin{bmatrix} 0 & -4 & -2 \end{bmatrix} \begin{bmatrix} 6 & 4 \\ -2 & 1 \\ -3 & 7 \end{bmatrix} = \begin{bmatrix} 14 & -18 \end{bmatrix} \quad \text{and} \quad A_{2,2}B_2 = \begin{bmatrix} 7 & -1 \end{bmatrix} \begin{bmatrix} -1 & 3 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} -12 & 19 \end{bmatrix}.$$

Then

$$A_{2,1}B_1 + A_{2,2}B_2 = \begin{bmatrix} 14 & -18 \end{bmatrix} + \begin{bmatrix} -12 & 19 \end{bmatrix} = \begin{bmatrix} 2 & 1 \end{bmatrix}.$$

If you combine the top and bottom blocks into one matrix, you get the result as if you simply computed AB:

$$AB = \left[\begin{array}{cc} -5 & 4 \\ -6 & 2 \\ \hline 2 & 1 \end{array}\right]$$

Note that all of our previous rules for matrix multiplication can be described using partitions. Indeed, the definition of $A\mathbf{x}$ using columns of A, the column definition of AB, the row-column rule for computing AB, the rows of AB as products of the rows of A and the matrix B all technically use partitions.

Example 2.11.2

Let

$$A = \begin{bmatrix} -3 & 1 & 2 \\ 1 & -4 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}.$$

Compute AB. In particular, verify that

$$AB = \sum_{k=1}^{3} \begin{bmatrix} A_{1,k} \\ A_{2,k} \end{bmatrix} \begin{bmatrix} B_{k,1} & B_{k,2} \end{bmatrix}.$$

In other words, multiply column 1 of A by row 1 of B to form a matrix. Then multiply column 2 of A by row 2 of B to form another matrix. Then multiply column 3 of A by row 3 of B to form another matrix. Then, add all of these together to get the product AB.

Solution

$$AB = \underbrace{\begin{bmatrix} -3 \\ 1 \end{bmatrix} \begin{bmatrix} a & b \end{bmatrix}}_{k=1} + \underbrace{\begin{bmatrix} 1 \\ -4 \end{bmatrix} \begin{bmatrix} c & d \end{bmatrix}}_{k=2} + \underbrace{\begin{bmatrix} 2 \\ 5 \end{bmatrix} \begin{bmatrix} e & f \end{bmatrix}}_{k=3}$$

$$= \begin{bmatrix} -3a & -3b \\ a & b \end{bmatrix} + \begin{bmatrix} c & d \\ -4c & -4d \end{bmatrix} + \begin{bmatrix} 2e & 2f \\ 5e & 5f \end{bmatrix}$$

$$= \begin{bmatrix} -3a + c + 2e & -3b + d + 2f \\ a - 4c + 5e & b - 4d + 5f \end{bmatrix}$$

Column-Row Expansion of AB

If $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$, then

$$AB = \begin{bmatrix} \text{col}_1(A) & \text{col}_2(A) & \cdots & \text{col}_n(A) \end{bmatrix} \begin{bmatrix} \text{row}_1(B) \\ \text{row}_2(B) \\ \vdots \\ \text{row}_n(B) \end{bmatrix}$$

$$= \text{col}_1(A)\,\text{row}_1(B) + \cdots + \text{col}_n(A)\,\text{row}_n(B).$$

A matrix of the form

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ 0 & A_{2,2} \end{bmatrix}$$

is called *block upper triangular*. Let $A_{1,1} \in \mathbb{R}^{p \times q}$ and $A_{2,2} \in \mathbb{R}^{q \times q}$. Then, assuming that A is invertible, we will denote A^{-1} by B such that

$$\underbrace{\begin{bmatrix} A_{1,1} & A_{1,2} \\ 0 & A_{2,2} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{bmatrix}}_{B} = \underbrace{\begin{bmatrix} I_p & 0 \\ 0 & I_q \end{bmatrix}}_{I}. \tag{8}$$

In matrix Equation 8, matching the blocks on the left-hand side to the identity matrix yields four equations:

$$A_{1,1}B_{1,1} + A_{1,2}B_{2,1} = I_p \tag{9}$$

$$A_{1,1}B_{1,2} + A_{1,2}B_{2,2} = 0 \tag{10}$$

$$A_{2,2}B_{2,1} = 0 \tag{11}$$

$$A_{2,2}B_{2,2} = I_q \tag{12}$$

Since $A_{2,2}$ is square, Equation 12 in tandem with the invertible matrix theorem shows that $A_{2,2}$ is invertible. It also shows that $B_{2,2} = A_{2,2}^{-1}$.

We can multiply both sides of (11) by $A_{2,2}^{-1}$ to get

$$B_{2,1} = A_{2,2}^{-1}0 = 0.$$

Then, we can simplify (9):

$$B_{2,1} = 0 \Rightarrow A_{1,1}B_{1,1} + 0 = I_p$$

Since $A_{1,1}$ is square, it is also invertible and $B_{1,1} = A_{1,1}^{-1}$. Then, we combine everything with (10) to get

$$A_{1,1}B_{1,2} = -A_{1,2}B_{2,2} = -A_{1,2}A_{2,2}^{-1}$$

and

$$B_{1,2} = -A_{1,1}^{-1}A_{1,2}A_{2,2}^{-1}.$$

Thus,

$$A^{-1} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ 0 & A_{2,2} \end{bmatrix}^{-1} = \begin{bmatrix} A_{1,1}^{-1} & -A_{1,1}^{-1}A_{1,2}A_{2,2}^{-1} \\ 0 & A_{2,2}^{-1} \end{bmatrix}.$$

A **block diagonal matrix** is a matrix where all the nonzero blocks are along the main diagonals of their blocks and every block not on the main diagonal is zero. Such a matrix is invertible if and only if each diagonal block is invertible.

$$A = \begin{bmatrix} A_{1,1} & 0 & \cdots & 0 \\ 0 & A_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{n,n} \end{bmatrix}, \quad A^{-1} = \begin{bmatrix} A_{1,1}^{-1} & 0 & \cdots & 0 \\ 0 & A_{2,2}^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{n,n}^{-1} \end{bmatrix}.$$

Problem 2.11.1

Compute $X^T X$ where X is partitioned as $\begin{bmatrix} X_1 & X_2 \end{bmatrix}$.

Solution

$$X^T = \begin{bmatrix} X_1^T \\ X_2^T \end{bmatrix}$$

$$X^T X = \begin{bmatrix} X_1^T \\ X_2^T \end{bmatrix} \begin{bmatrix} X_1 & X_2 \end{bmatrix} = \begin{bmatrix} X_1^T X_1 & X_1^T X_2 \\ X_2^T X_1 & X_2^T X_2 \end{bmatrix}$$

Problem 2.11.2

Let

$$A = \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix},$$

where B and C are square. Show that A is invertible if and only if both B and C are invertible.

Solution

Suppose A is invertible and let

$$A^{-1} = \begin{bmatrix} X_{1,1} & X_{1,2} \\ X_{2,1} & X_{2,2} \end{bmatrix}.$$

Then

$$AA^{-1} = \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix} \begin{bmatrix} X_{1,1} & X_{1,2} \\ X_{2,1} & X_{2,2} \end{bmatrix} = \begin{bmatrix} BX_{1,1} & BX_{1,2} \\ CX_{2,1} & CX_{2,2} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}.$$

Hence $BX_{1,1} = I$ and $CX_{2,2} = I$, so B and C are invertible.

If B and C are invertible, then

$$A^{-1} = \begin{bmatrix} B^{-1} & 0 \\ 0 & C^{-1} \end{bmatrix}$$

satisfies

$$\begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix} \begin{bmatrix} B^{-1} & 0 \\ 0 & C^{-1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} B^{-1} & 0 \\ 0 & C^{-1} \end{bmatrix} \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}.$$

Because of this and the invertible matrix theorem, A is invertible.

2.12 *LU* Factorization

The *factorization* of a matrix refers to when we break down a matrix into a product of two or more matrices.

Suppose we need to solve several systems of linear equations that share the same coefficient matrix:

$$A\mathbf{x} = \mathbf{b}_1, \quad A\mathbf{x} = \mathbf{b}_2, \quad \ldots, \quad A\mathbf{x} = \mathbf{b}_p.$$

Based on the previous section, your instinct may be to compute A^{-1} and then compute

$$\mathbf{x}_k = A^{-1}\mathbf{b}_k$$

for $k = 1, 2, \ldots, p$. However, this is quite an inefficient method. A more practical method is to factor A once into the product of two simpler matrices and reuse that factorization to solve all systems. This would be done by row reduction and obtaining an *LU factorization*.

LU Factorization

Let $A \in \mathbb{R}^{n\times n}$ be an invertible matrix with all nonzero diagonal elements. An LU factorization of A is given by

$$A = LU$$

where L is a unit lower triangular matrix with $l_{1,1} = 1$ and $U \in \mathbb{R}^{n\times n}$ is an upper triangular matrix where $u_{i,i} \neq 0$.

We first assume that we have $A\mathbb{R}^{m\times n}$ that can be row reduced to REF without row interchanges. Then we have an example of A written in the form LU:

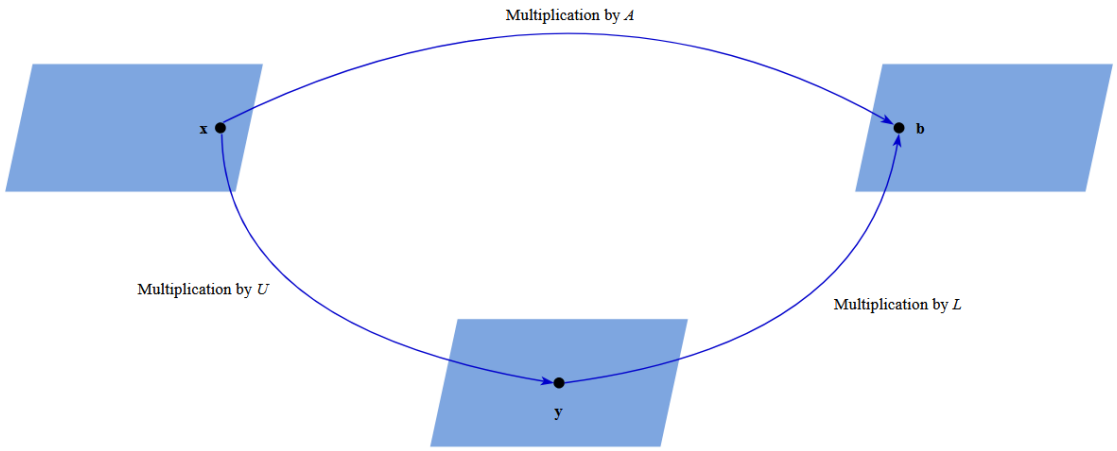
$$A = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ \star & 1 & 0 & 0 \\ \star & \star & 1 & 0 \\ \star & \star & \star & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} \blacksquare & \star & \star & \star & \star \\ 0 & \blacksquare & \star & \star & \star \\ 0 & 0 & 0 & \blacksquare & \star \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{U}$$

Once A is factored, the matrix equation$A\mathbf{x} = \mathbf{b}$ becomes $L(U\mathbf{x}) = \mathbf{b}$. We can then solve the following pair of equations:

$$L\mathbf{y} = \mathbf{b} \tag{13}$$

$$U\mathbf{x} = \mathbf{y} \tag{14}$$

You first solve (13) for $\mathbf{y}$ and then solve (14) for $\mathbf{x}$. This a relatively efficient method because L and U are triangular.



Factorization of the mapping $\mathbf{x} \mapsto A\mathbf{x}$

Example 2.12.1

Consider the following:

$$A = \begin{bmatrix} 3 & -7 & -2 & 2 \\ -3 & 5 & 1 & 0 \\ 6 & -4 & 0 & -5 \\ -9 & 5 & -5 & 12 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 2 & -5 & 1 & 0 \\ -3 & 8 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & -7 & -2 & 2 \\ 0 & -2 & 1 & 2 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix} = LU$$

Use this LU factorization of A to solve $A\mathbf{x} = \mathbf{b}$, where

$$\mathbf{b} = \begin{bmatrix} -9 \\ 5 \\ 7 \\ 11 \end{bmatrix}.$$

Solution

First solve $L\mathbf{y} = \mathbf{b}$ using forward substitution:

$$\begin{bmatrix} L & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & -9 \\ -1 & 1 & 0 & 0 & 5 \\ 2 & -5 & 1 & 0 & 7 \\ -3 & 8 & 3 & 1 & 11 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 & -9 \\ 0 & 1 & 0 & 0 & -4 \\ 0 & 0 & 1 & 0 & 5 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} I & \mathbf{y} \end{bmatrix}$$

Thus

$$\mathbf{y} = \begin{bmatrix} -9 \\ -4 \\ 5 \\ 1 \end{bmatrix}.$$

Next solve $U\mathbf{x} = \mathbf{y}$ using back substitution:

$$\begin{bmatrix} U & \mathbf{y} \end{bmatrix} = \begin{bmatrix} 3 & -7 & -2 & 2 & -9 \\ 0 & -2 & -1 & 2 & -4 \\ 0 & 0 & -1 & 1 & 5 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 & -6 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} = \begin{bmatrix} I & \mathbf{x} \end{bmatrix}$$

Thus

$$\mathbf{x} = \begin{bmatrix} 3 \\ 4 \\ -6 \\ -1 \end{bmatrix}.$$

In Gaussian elimination, each step that eliminates entries below a pivot can be represented as

multiplication by a special unit lower triangular matrix called an elimination matrix. This method is sometimes referred to as Gauss transform.

We begin by writing the kth elimination matrix

$$L_k = I_n - \tau_k \mathbf{e}_k^T$$

where

$$\tau_k = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \tau_{k+1} \\ \vdots \\ \tau_n \end{bmatrix}$$

is a column vector of the multipliers used in elimination for the kth step.

Consider the following elimination matrix with the multipliers inserted in column 1 below the pivot:

$$L_1 = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} = I_3 - \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

We can factor the second term:

$$\begin{aligned} L_1 &= I_3 - \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \\ &= I_3 - \boldsymbol{\tau}_1 \mathbf{e}_1^T \end{aligned}$$

with

$$\boldsymbol{\tau}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{a_{2,1}}{a_{1,1}} \\ \frac{a_{3,1}}{a_{1,1}} \end{bmatrix}.$$

If we carry on with elimination, we can write the process out as

$$L_2 L_1 A = U$$

where L_1 eliminates entries in column 1 and L_2 eliminates entries in column 2. And then we multiply on the left by $L_1^{-1}L_2^{-1}$ to get A:

$$\begin{aligned} L_1^{-1}L_2^{-1}L_2L_1A &= L_1^{-1}L_2^{-1}U \\ L_1^{-1}I_3L_1A &= L_1^{-1}L_2^{-1}U \\ L_1^{-1}L_1A &= L_1^{-1}L_2^{-1}U \\ I_3A &= L_1^{-1}L_2^{-1}U \end{aligned}$$

Recognizing that $L = L_1^{-1}L_2^{-1}$ is also unit lower triangular, we arrive at

$$A = LU.$$

Recall that each elimination matrix can be expressed as

$$L_k = I_n - \boldsymbol{\tau}_k \mathbf{e}_k^T$$

where $\boldsymbol{\tau}_k$ contains the multipliers in column k below the pivot.

Since L_k is unit lower triangular, its inverse is also unit lower triangular and is particularly simple to compute. Noting that $\boldsymbol{\tau}_k \mathbf{e}_k^T$ has nonzero entries only in column k below the diagonal, we have

$$L_k^{-1} = \left(I_n - \boldsymbol{\tau}_k \mathbf{e}_k^T\right)^{-1} = I_n + \boldsymbol{\tau}_k \mathbf{e}_k^T.$$

In other words, inverting L_k simply changes the signs of the multipliers in column k.

For example, in the 3×3 case with $k = 1$:

$$L_1^{-1} = \left(I_3 - \boldsymbol{\tau}_1 \mathbf{e}_1^T\right)^{-1} = I_3 + \boldsymbol{\tau}_1 \mathbf{e}_1^T.$$

Using this fact, the L matrix in the LU factorization is obtained by multiplying together the inverses of the elimination matrices in reverse order:

$$L = L_1^{-1} L_2^{-1} \cdots L_{n-1}^{-1}.$$

For instance, with $n = 3$ and two elimination steps:

$$\begin{aligned} L &= L_1^{-1} L_2^{-1} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix}. \end{aligned}$$

We will now generalize the LU factorization algorithm. Suppose A can be reduced to an REF matrix U using only row replacements that add a multiple of one row to another row below it (no row swaps). Then there exists a product of multiple lower triangular elementary matrices $L_p \cdots L_1$ such that

$$L_p \cdots L_2 L_1 A = U$$

We now place entries in L so that the same sequence of row operations reduces L to I. Taking inverses and using the fact that $(L_k)^{-1}$ is also unit lower triangular, we obtain

$$A = L_1^{-1} L_2^{-1} \cdots L_p^{-1} U$$

where

$$L = L_1^{-1} L_2^{-1} \cdots L_p^{-1}$$

is unit lower triangular. This gives the LU factorization $A = LU$.

Example 2.12.2

Find an LU factorization of

$$A = \begin{bmatrix} \boxed{2} & 4 & -1 & 5 & -2 \\ \boxed{-4} & -5 & 3 & -8 & 1 \\ \boxed{2} & -5 & -4 & 1 & 8 \\ \boxed{-6} & 0 & 7 & -3 & 1 \end{bmatrix}.$$

Solution

We perform Gaussian elimination without row swaps, storing the multipliers (boxed in blue) in each L. Then, the entries that determine the sequence of row operations are boxed in red. We have multipliers $m_{2,1} = -\frac{4}{2} = \boxed{-2}$, $m_{3,1} = \frac{2}{2} = \boxed{1}$, and $m_{4,1} = -\frac{6}{2} = \boxed{-3}$ stored in column 1 of L below the diagonal:

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \boxed{-2} & 1 & 0 & 0 \\ \boxed{1} & 0 & 1 & 0 \\ \boxed{-3} & 0 & 0 & 1 \end{bmatrix}$$

We then eliminate column 1:

$$A \sim A_1 = \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & \boxed{3} & 1 & 2 & -3 \\ 0 & \boxed{-9} & -3 & -4 & 10 \\ 0 & \boxed{12} & 4 & 12 & -5 \end{bmatrix}$$

We now have $m_{3,2} = \frac{-9}{3} = \boxed{-3}$ and $m_{4,2} = \frac{12}{3} = \boxed{4}$. Let's update L:

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \boxed{-2} & 1 & 0 & 0 \\ \boxed{1} & \boxed{-3} & 1 & 0 \\ \boxed{-3} & \boxed{4} & 0 & 1 \end{bmatrix}$$

We then eliminate column 2:

$$A_1 \sim A_2 = \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & 0 & 0 & \boxed{2} & 1 \\ 0 & 0 & 0 & \boxed{4} & 7 \end{bmatrix}$$

Lastly we have $m_{4,3} = \frac{4}{2} = \boxed{2}$. We place each of the multipliers divided by their corresponding pivots into L. The final L is

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \boxed{-2} & 1 & 0 & 0 \\ \boxed{1} & \boxed{-3} & 1 & 0 \\ \boxed{-3} & \boxed{4} & \boxed{2} & 1 \end{bmatrix}.$$

We eliminate column 3 in A_2 to get U:

$$A_2 \sim U = \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & \boxed{5} \end{bmatrix}$$

The forward elimination process transforms a matrix A into an upper triangular matrix U. At each elimination step $E_{i,j}$, we subtract multiplier $m_{i,j}$ times row j from row i, creating a zero in the (i, j) position.

To reverse this process, we replace subtraction with addition. This gives us the lower triangular factors L_{ij} that build A back from U.

For instance,

$$E_{3,1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -m_{3,1} & 0 & 1 \end{bmatrix} \Rightarrow L_{3,1} = E_{3,1}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ m_{3,1} & 0 & 1 \end{bmatrix}.$$

In general, inverses in reverse order are

$$E^{-1} = E_{2,1}^{-1} E_{3,1}^{-1} E_{3,2}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ m_{2,1} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ m_{3,1} & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & m_{3,2} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ m_{2,1} & 1 & 0 \\ m_{3,1} & m_{3,2} & 1 \end{bmatrix} = L.$$

When we form U from A, the pivot rows, from which we subtract multiples of, never change. They are exactly the same in U as they were in A after previous steps. However, the rows we eliminate are changed versions of A's original rows.

Consider computing the third row of U:

$$\text{Row 3 of } U = (\text{Row 3 of } A) - m_{3,1} \cdot (\text{Row 1 of } U) - m_{3,2} \cdot (\text{Row 2 of } U)$$

Rewriting this to solve for the original row of A:

$$\text{Row 3 of } A = m_{3,1} \cdot (\text{Row 1 of } U) + m_{3,2} \cdot (\text{Row 2 of } U) + 1 \cdot (\text{Row 3 of } U)$$

This shows that Row 3 of A is a linear combination of the rows of U, with coefficients $m_{3,1}, m_{3,2}, 1$. These coefficients are exactly the entries in row 3 of L.

Every row of A can be written this way. The pivot row has a 1 in L on the diagonal. The multipliers $m_{i,j}$ from elimination fill the entries below the diagonal.

This structure ensures $A = LU$.

Example 2.12.3

We are given the LU factorization of the matrix A and vector $\mathbf{b}$:

$$A = \begin{bmatrix} 2 & -1 & 1 \\ 6 & -4 & 2 \\ 4 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & -4 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 1 \\ 0 & -1 & -1 \\ 0 & 0 & -5 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix}.$$

Use this LU factorization to solve $A\mathbf{x} = \mathbf{b}$ by solving the two systems $L\mathbf{y} = \mathbf{b}$ and $U\mathbf{x} = \mathbf{y}$. Then compute $\mathbf{y}^T\mathbf{x}$.

Solution

First solve $L\mathbf{y} = \mathbf{b}$ by forward substitution:

$$\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & -4 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix}$$

From the first equation, we have $y_1 = 1$. From the second, we have $3y_1 + y_2 = 6 \Rightarrow y_2 = 3$. From the third, we have $2y_1 - 4y_2 + y_3 = 0 \Rightarrow y_3 = 10$. Thus

$$\mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 10 \end{bmatrix}.$$

Next solve $U\mathbf{x} = \mathbf{y}$ by back substitution:

$$\begin{bmatrix} 2 & -1 & 1 \\ 0 & -1 & -1 \\ 0 & 0 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 10 \end{bmatrix}.$$

From the third equation, we have $-5x_3 = 10 \Rightarrow x_3 = -2$. From the second, we have $-x_2 - x_3 = 3 \Rightarrow x_2 = -1$. From the first, we have $2x_1 - (-1) + (-2) = 1 \Rightarrow x_1 = 1$. Thus

$$\mathbf{x} = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}.$$

Finally,

$$\mathbf{y}^T\mathbf{x} = \begin{bmatrix} 1 & 3 & 10 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix} = 1 - 3 - 20 = -22.$$

Problem 2.12.1

Find the LU factorization of $A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 5 & -2 \\ 2 & -2 & 0 \end{bmatrix}$.

Solution

We are trying to eliminate below the pivot in column 1. We have $m_{2,1} = \frac{4}{2} = 2$ and $m_{3,1} = \frac{2}{2} = 1$. Thus, we perform row operations to get

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

We are trying to eliminate below the pivot in column 2. We have $m_{3,2} = \frac{-3}{3} = -1$. Thus, we perform row operations to get

$$U = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & -4 \\ 0 & 0 & -5 \end{bmatrix}.$$

L has 1s on the diagonal with the multipliers below it:

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

Problem 2.12.2

Let

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}.$$

(a) Solve $L\mathbf{y} = \mathbf{b}$ to find $\mathbf{y}$ and (b) solve $U\mathbf{x} = \mathbf{y}$ to find $\mathbf{y}$. (c) Recover the matrix A.

Solution

(a)

$$\begin{cases} y_1 & = 4 \\ y_1 + y_2 & = 5 \\ y_1 + y_2 + y_3 & = 6 \end{cases}$$

Then $y_2 = 1$ and $y_3 = 1$. Thus $\mathbf{y} = \begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix}$.

(b)

$$\begin{cases} x_3 = 1 \\ x_2 + x_3 = 1 \\ x_1 + x_2 + x_3 = 6 \end{cases}$$

Then $x_2 = 0$ and $x_1 = 3$. Thus $\mathbf{x} = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$.

(c)

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

2.13 Applications of Matrix Algebra

Our first application follows from our previous section on matrix factorizations. Let's analyze an electrical network to learn about how matrix factorization is connected to circuit analysis.

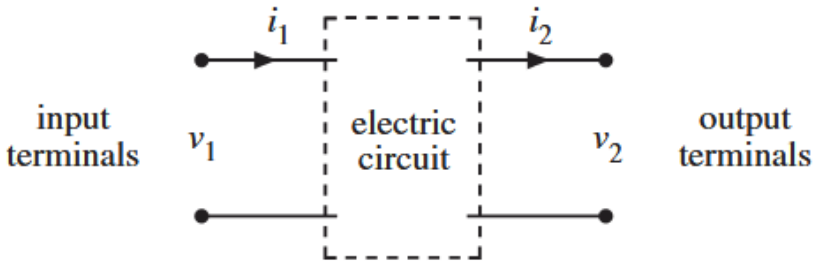


Figure 3: Image credit: Lay et al

In Figure 2, we have an electrical circuit with an input voltage vector

$$\begin{bmatrix} v_1 \\ i_1 \end{bmatrix}$$

and an output voltage vector

$$\begin{bmatrix} v_2 \\ i_2 \end{bmatrix}.$$

Often, the mapping $\begin{bmatrix} v_1 \\ i_1 \end{bmatrix} \mapsto \begin{bmatrix} v_2 \\ i_2 \end{bmatrix}$ between these quantities is linear, thus exists a *transfer matrix* A such that

$$\begin{bmatrix} v_2 \\ i_2 \end{bmatrix} = A \begin{bmatrix} v_1 \\ i_1 \end{bmatrix}.$$

In Figure 3, we have a ladder network which consists of two circuits connected in series so that the output of one circuit becomes the input of the next circuit.

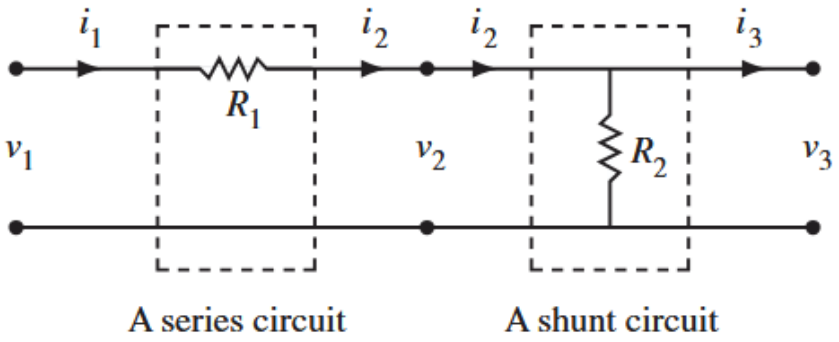


Figure 4: Image credit: Lay et al

The series circuit's transfer matrix is

$$A_1 = \begin{bmatrix} 1 & -R_1 \\ 0 & 1 \end{bmatrix}.$$

The shunt circuit's transfer matrix is

$$A_2 = \begin{bmatrix} 1 & 0 \\ -\frac{1}{R_2} & 1 \end{bmatrix}.$$

An input vector $\mathbf{x}$ is transformed into $A_1\mathbf{x}$ and then into $A_2(A_1\mathbf{x})$. The series connection of the circuits corresponds to the composition of the corresponding linear transformations. Thus the overall transfer matrix is

$$A_2A_1 = \begin{bmatrix} 1 & 0 \\ -\frac{1}{R_2} & 1 \end{bmatrix} \begin{bmatrix} 1 & -R_1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -R_1 \\ -\frac{1}{R_2} & 1 + \frac{R_1}{R_2} \end{bmatrix}.$$

Example 2.13.1

Design a ladder network with transfer matrix $A = \begin{bmatrix} 1 & -8 \\ -0.5 & 5 \end{bmatrix}$.

Solution

We need to factor A into the product of transfer matrices. To do this, we need to find values of R_1 and R_2 such that

$$\begin{bmatrix} 1 & -R_1 \\ -\frac{1}{R_2} & 1+\frac{R_1}{R_2} \end{bmatrix} = \begin{bmatrix} 1 & -8 \\ -0.5 & 5 \end{bmatrix}.$$

We match the entries as follows:

$$-R_1 = -8 \Rightarrow R_1 = 8\ \Omega$$
$$-\frac{1}{R_2} = -0.5 \Rightarrow R_2 = 2\ \Omega$$

By encoding a two-port circuit as a transfer matrix, we can see that connecting blocks in series means we are composing their transformations, and thus multiplying their corresponding transfer matrices in signal order. We do this by designing a matrix that specifies the specific physical behavior we want. If we can factor this matrix into smaller circuits, the network can then be manufactured. Then, through optimization, engineers often try to find a *minimal realization* that uses the least electrical components.

We can also use matrix factorizations to solve heat flow problems.

Example 2.13.2

The following band matrix (that is, a matrix where all its nonzero elements are stored within a "band" along its main diagonal) is used to estimate the unsteady conduction of heat in a rod when the temperatures at equally spaced points $p_1, \ldots, p_5$ on the rod change with time.

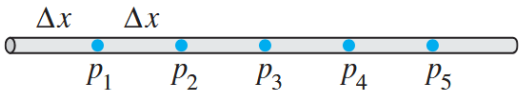


Image credit: Lay et al

$$A = \begin{bmatrix} 1+2C & -C & 0 & 0 & 0 \\ -C & 1+2C & -C & 0 & 0 \\ 0 & -C & 1+2C & -C & 0 \\ 0 & 0 & -C & 1+2C & -C \\ 0 & 0 & 0 & -C & 1+2C \end{bmatrix}$$

The constant C in the matrix depends on the material of the rod, the distance Δx between the points on the rod, and the length of time Δt between successive temperature measurements.

Suppose that for $k = 0, 1, 2, \ldots$, a vector $\mathbf{t}_k \in \mathbb{R}^5$ lists the temperatures at time $k\Delta t$. If two ends of the rod are a constant $0°$, then the temperature vectors are modeled by the equation

$$A\mathbf{t}_{k+1} = \mathbf{t}_k$$

(a) Find the LU factorization of A when $C = 1$.

(b) Suppose $C = 1$ and $\mathbf{t}_0 = (10, 12, 12, 12, 10)$. Use your LU factorization from (a) to find the temperature distributions $\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3$, and $\mathbf{t}_4$.

Solution

(a) Because of its three nonzero diagonals, A is what we call a *tridiagonal* matrix with L unit lower bidiagonal and U upper bidiagonal. At $C = 1$, we have

$$A = \begin{bmatrix} 3 & -1 & 0 & 0 & 0 \\ -1 & 3 & -1 & 0 & 0 \\ 0 & -1 & 3 & -1 & 0 \\ 0 & 0 & -1 & 3 & -1 \\ 0 & 0 & 0 & -1 & 3 \end{bmatrix}.$$

We now seek $A = LU$ with L unit lower bidiagonal and U upper bidiagonal. It's obviously better to do this using MATLAB, but you can do it by hand if you want.

We initialize with $u_{1,1} = 3$. For the L and U entries, we have

$$m_i = \frac{b}{u_{i-1,i-1}} \quad \text{and} \quad u_{i,i} = a - m_i c$$

Now we compute the entries:

$$\begin{aligned} m_2 &= \frac{-1}{3} = -\frac{1}{3} \Rightarrow m_{2,2} = 3 - \left(-\frac{1}{3}\right)(-1) = \frac{8}{3}, \\ m_3 &= \frac{-1}{8/3} = -\frac{3}{8} \Rightarrow m_{3,3} = 3 - \left(-\frac{3}{8}\right)(-1) = \frac{21}{8}, \\ m_4 &= \frac{-1}{21/8} = -\frac{8}{21} \Rightarrow m_{4,4} = 3 - \left(-\frac{8}{21}\right)(-1) = \frac{55}{21}, \\ m_5 &= \frac{-1}{55/21} = -\frac{21}{55} \Rightarrow m_{5,5} = 3 - \left(-\frac{21}{55}\right)(-1) = \frac{144}{55}. \end{aligned}$$

We can now construct L and U:

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -\frac{1}{3} & 1 & 0 & 0 & 0 \\ 0 & -\frac{3}{8} & 1 & 0 & 0 \\ 0 & 0 & -\frac{8}{21} & 1 & 0 \\ 0 & 0 & 0 & -\frac{21}{55} & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 3 & -1 & 0 & 0 & 0 \\ 0 & \frac{8}{3} & -1 & 0 & 0 \\ 0 & 0 & \frac{21}{8} & -1 & 0 \\ 0 & 0 & 0 & \frac{55}{21} & -1 \\ 0 & 0 & 0 & 0 & \frac{144}{55} \end{bmatrix}.$$

(b) With the factorization $A = LU$ from part (a), advance one time step by

$$L\mathbf{s}_{k+1} = \mathbf{t}_k$$

and

$$U\mathbf{t}_{k+1} = \mathbf{s}_{k+1}.$$

Because L and U are bidiagonal, these reduce to simple forward and backward substitutions.

For forward, we solve $L\,\mathbf{s}_{k+1} = \mathbf{t}_k$:

$$\begin{aligned}
s_1^{(k+1)} &= t_1^{(k)},\\
s_2^{(k+1)} &= t_2^{(k)} - m_2\, s_1^{(k+1)},\\
s_3^{(k+1)} &= t_3^{(k)} - m_3\, s_2^{(k+1)},\\
s_4^{(k+1)} &= t_4^{(k)} - m_4\, s_3^{(k+1)},\\
s_5^{(k+1)} &= t_5^{(k)} - m_5\, s_4^{(k+1)},
\end{aligned}$$

where

$$m_2 = -\frac{1}{3}, \quad m_3 = -\frac{3}{8}, \quad m_4 = -\frac{8}{21}, \quad m_5 = -\frac{21}{55}.$$

For backward, we solve $U\,\mathbf{t}_{k+1} = \mathbf{s}_{k+1}$:

$$\begin{aligned}
t_5^{(k+1)} &= \frac{s_5^{(k+1)}}{u_{5,5}},\\
t_4^{(k+1)} &= \frac{s_4^{(k+1)} - (-1)\, t_5^{(k+1)}}{u_{4,4}},\\
t_3^{(k+1)} &= \frac{s_3^{(k+1)} - (-1)\, t_4^{(k+1)}}{u_{3,3}},\\
t_2^{(k+1)} &= \frac{s_2^{(k+1)} - (-1)\, t_3^{(k+1)}}{u_{2,2}},\\
t_1^{(k+1)} &= \frac{s_1^{(k+1)} - (-1)\, t_2^{(k+1)}}{u_{1,1}},
\end{aligned}$$

where

$$u_{1,1} = 3, \quad u_{2,2} = \frac{8}{3}, \quad u_{3,3} = \frac{21}{8}, \quad u_{4,4} = \frac{55}{21}, \quad u_{5,5} = \frac{144}{55}.$$

Carrying out the two substitutions for each step

We begin with step $k = 0$, from $\mathbf{t}_0$ to $\mathbf{s}_1$, $\mathbf{t}_1$. Here is the forward:

$$
\begin{aligned}
s_1^{(1)} &= 10, \\
s_2^{(1)} &= 12 - m_2\, s_1^{(1)} = 12 - \left(-\frac{1}{3}\right)(10) = 15.3333 \\
s_3^{(1)} &= 12 - m_3\, s_2^{(1)} = 12 - \left(-\frac{3}{8}\right)(15.3333) = 17.7500 \\
s_4^{(1)} &= 12 - m_4\, s_3^{(1)} = 12 - \left(-\frac{8}{21}\right)(17.7500) = 18.7619 \\
s_5^{(1)} &= 10 - m_5\, s_4^{(1)} = 10 - \left(-\frac{21}{55}\right)(18.7619) = 17.1636
\end{aligned}
$$

And here is the backward:

$$
\begin{aligned}
t_5^{(1)} &= \frac{s_5^{(1)}}{u_{5,5}} = \frac{17.1636}{144/55} = 6.5556 \\
t_4^{(1)} &= \frac{s_4^{(1)} + t_5^{(1)}}{u_{4,4}} = \frac{18.7619 + 6.5556}{55/21} = 9.6667 \\
t_3^{(1)} &= \frac{s_3^{(1)} + t_4^{(1)}}{u_{3,3}} = \frac{17.7500 + 9.6667}{21/8} = 10.4444 \\
t_2^{(1)} &= \frac{s_2^{(1)} + t_3^{(1)}}{u_{2,2}} = \frac{15.3333 + 10.4444}{8/3} = 9.6667 \\
t_1^{(1)} &= \frac{s_1^{(1)} + t_2^{(1)}}{u_{1,1}} = \frac{10 + 9.6667}{3} = 6.5556
\end{aligned}
$$

We have step $k = 1$, from $\mathbf{t}_1$ to $\mathbf{s}_2$, $\mathbf{t}_2$. Here is the forward:

$$
\begin{aligned}
s_1^{(2)} &= 6.5556, \\
s_2^{(2)} &= 9.6667 - m_2\, s_1^{(2)} = 9.6667 - \left(-\frac{1}{3}\right)(6.5556) = 11.8519 \\
s_3^{(2)} &= 8.5926 - m_3\, s_2^{(2)} = 10.4444 - \left(-\frac{3}{8}\right)(11.8519) = 14.8889 \\
s_4^{(2)} &= 9.6667 - m_4\, s_3^{(2)} = 9.6667 - \left(-\frac{8}{21}\right)(14.8889) = 15.3386 \\
s_5^{(2)} &= 6.5556 - m_5\, s_4^{(2)} = 6.5556 - \left(-\frac{21}{55}\right)(15.3386) = 12.4121
\end{aligned}
$$

And here is the backward:

$$
\begin{aligned}
t_5^{(2)} &= \frac{12.4121}{144/55} = 4.7407 \\
t_4^{(2)} &= \frac{15.3386 + 4.7407}{55/21} = 7.6667 \\
t_3^{(2)} &= \frac{14.8889 + 7.6667}{21/8} = 8.5926 \\
t_2^{(2)} &= \frac{11.8519 + 8.5926}{8/3} = 7.6667 \\
t_1^{(2)} &= \frac{6.5556 + 7.6667}{3} = 4.7407
\end{aligned}
$$

We have step $k = 2$, from $\mathbf{t}_2$ to $\mathbf{s}_3$, $\mathbf{t}_3$. Here is the forward:

$$
\begin{aligned}
s_1^{(3)} &= 4.7407, \\
s_2^{(3)} &= 7.6667 - \left(-\frac{1}{3}\right)(4.7407) = 9.2469 \\
s_3^{(3)} &= 8.5926 - \left(-\frac{3}{8}\right)(9.2469) = 12.0602 \\
s_4^{(3)} &= 7.6667 - \left(-\frac{8}{21}\right)(12.0602) = 12.2610 \\
s_5^{(3)} &= 4.7407 - \left(-\frac{21}{55}\right)(12.2610) = 9.4222
\end{aligned}
$$

And here is the backward:

$$
\begin{aligned}
t_5^{(3)} &= \frac{9.4222}{144/55} = 3.5988 \\
t_4^{(3)} &= \frac{12.2610 + 3.5988}{55/21} = 6.0556 \\
t_3^{(3)} &= \frac{12.0602 + 6.0556}{21/8} = 6.9012 \\
t_2^{(3)} &= \frac{9.2469 + 6.9012}{8/3} = 6.0556 \\
t_1^{(3)} &= \frac{4.7407 + 6.0556}{3} = 3.5988
\end{aligned}
$$

We have step $k = 3$, from $\mathbf{t}_3$ to $\mathbf{s}_4$, $\mathbf{t}_4$. Here is the forward:

$$
\begin{aligned}
s_1^{(4)} &= 3.5988 \\
s_2^{(4)} &= 6.0556 - \left(-\frac{1}{3}\right)(3.5988) = 7.2551 \\
s_3^{(4)} &= 6.9012 - \left(-\frac{3}{8}\right)(7.2551) = 9.6219 \\
s_4^{(4)} &= 6.0556 - \left(-\frac{8}{21}\right)(9.6219) = 9.7210 \\
s_5^{(4)} &= 3.5988 - \left(-\frac{21}{55}\right)(9.7210) = 7.3104
\end{aligned}
$$

And here is the backward:

$$
\begin{aligned}
t_5^{(4)} &= \frac{7.3104}{144/55} = 2.7922 \\
t_4^{(4)} &= \frac{9.7210 + 2.7922}{55/21} = 4.7778 \\
t_3^{(4)} &= \frac{9.6219 + 4.7778}{21/8} = 5.4856 \\
t_2^{(4)} &= \frac{7.2551 + 5.4856}{8/3} = 4.7778 \\
t_1^{(4)} &= \frac{3.5988 + 4.7778}{3} = 2.7922
\end{aligned}
$$

And finally we put all the results together:

$$
\mathbf{s}_1 = \begin{bmatrix} 10.0000 \\ 15.3333 \\ 17.7500 \\ 18.7619 \\ 17.1636 \end{bmatrix}, \quad \mathbf{t}_1 = \begin{bmatrix} 6.5556 \\ 9.6667 \\ 10.4444 \\ 9.6667 \\ 6.5556 \end{bmatrix}
$$

$$
\mathbf{s}_2 = \begin{bmatrix} 6.5556 \\ 11.8519 \\ 14.8889 \\ 15.3386 \\ 12.4121 \end{bmatrix}, \quad \mathbf{t}_2 = \begin{bmatrix} 4.7407 \\ 7.6667 \\ 8.5926 \\ 7.6667 \\ 4.7407 \end{bmatrix}
$$

$$\mathbf{s}_3 = \begin{bmatrix} 4.7407 \\ 9.2469 \\ 12.0602 \\ 12.2610 \\ 9.4222 \end{bmatrix}, \quad \mathbf{t}_3 = \begin{bmatrix} 3.5988 \\ 6.0556 \\ 6.9012 \\ 6.0556 \\ 3.5988 \end{bmatrix}$$

$$\mathbf{s}_4 = \begin{bmatrix} 3.5988 \\ 7.2551 \\ 9.6219 \\ 9.7210 \\ 7.3104 \end{bmatrix}, \quad \mathbf{t}_4 = \begin{bmatrix} 2.7922 \\ 4.7778 \\ 5.4856 \\ 4.7778 \\ 2.7922 \end{bmatrix}$$

We can now create a more elaborate Leontief input-output model.

Suppose the economy is divided into n sectors, each producing goods or services. Let $\mathbf{x} \in \mathbb{R}^n$ be the *production vector*, where x_i represents the total output of sector i per year. Let $\mathbf{d} \in \mathbb{R}^n$ be the *final demand vector*, which lists the amounts of goods demanded from each sector by the *open* (nonproductive) part of the economy. This includes consumers and surplus production.

As sectors produce goods to meet this final demand, they also require goods from other sectors to use as inputs in their own production. These internal requirements are called *intermediate demand*.

Leontief's model finds a production level $\mathbf{x}$ such that the amounts produced will balance the total demand for that production such that

$$\left\{ \begin{array}{c} \text{amount} \\ \text{produced} \\ \mathbf{x} \end{array} \right\} = \left\{ \begin{array}{c} \text{intermediate} \\ \text{demand} \end{array} \right\} + \left\{ \begin{array}{c} \text{final} \\ \text{demand} \\ \mathbf{d} \end{array} \right\}.$$

For each sector j there is a *unit consumption vector* $\mathbf{c}_j \in \mathbb{R}^n$ that lists the inputs needed *per unit output* of that sector. All quantities are measured in monetary units.

For example, suppose the economy is divided into manufacturing, agriculture, and services sectors with the following input requirements:

Purchased from	Manufacturing	Agriculture	Services
Manufacturing	0.50	0.40	0.20
Agriculture	0.20	0.30	0.10
Services	0.10	0.10	0.30

Here, column j is the unit consumption vector $\mathbf{c}_j$ for sector j. We will let manufacturing be $\mathbf{c}_1$, agriculture be $\mathbf{c}_2$, and services be $\mathbf{c}_3$.

Example 2.13.3

What amounts will be consumed by the manufacturing sector if it produces 100 units?

Solution

$$100\mathbf{c}_1 = 100\begin{bmatrix}0.50\\0.20\\0.10\end{bmatrix} = \begin{bmatrix}50\\20\\10\end{bmatrix}$$

Producing 100 units of manufacturing output requires 50 units from manufacturing itself, 20 from agriculture, and 10 from services.

If $\mathbf{x} = \begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$ is the production vector, then intermediate demand is given by

$$x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + x_3\mathbf{c}_3 = C\mathbf{x}.$$

If we combine it all, we get

$$\mathbf{x} = C\mathbf{x} + \mathbf{d}.$$

This can be rewritten as $(I - C)\mathbf{x} = \mathbf{d}$ or $(I - C)\mathbf{x} = \mathbf{d}$.

Given the final demand vector $\mathbf{d}$ and the consumption matrix C, solving this system yields the production vector $\mathbf{x}$ that exactly meets both intermediate and final demands.

Example 2.13.4

Suppose the final demand is 50 units for manufacturing, 30 units for agriculture, and 20 units for services. Find the production level $\mathbf{x}$ that will satisfy this demand.

Solution

$$I - C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.50 & 0.40 & 0.20 \\ 0.20 & 0.30 & 0.10 \\ 0.10 & 0.10 & 0.30 \end{bmatrix} = \begin{bmatrix} 0.50 & -0.40 & -0.20 \\ -0.20 & 0.70 & -0.10 \\ -0.10 & -0.10 & 0.70 \end{bmatrix}$$

We now form the augmented matrix and row reduce:

$$\begin{bmatrix} 0.50 & -0.40 & -0.20 & 50 \\ -0.20 & 0.70 & -0.10 & 30 \\ -0.10 & -0.10 & 0.70 & 20 \end{bmatrix} \sim \begin{bmatrix} 5 & -4 & -2 & 500 \\ -2 & 7 & -1 & 300 \\ -1 & -1 & 7 & 200 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 225.9 \\ 0 & 1 & 0 & 118.5 \\ 0 & 0 & 1 & 77.8 \end{bmatrix}$$

If we round, manufacturing will produce 226 units, agriculture will produce 119 units, and services will produce 78 units.

If the entries of C are nonnegative and each column sum of C is less than 1, then $(I-C)$ is invertible and the production vector

$$\mathbf{x} = (I - C)^{-1}\mathbf{d}$$

has nonnegative entries and is the solution of $\mathbf{x} = C\mathbf{x} = \mathbf{d}$.

When the final demand vector $\mathbf{d}$ is given, industries first produce exactly $\mathbf{x} = \mathbf{d}$ to meet it. However, producing $\mathbf{d}$ requires additional inputs from other sectors. The 1st round intermediate demand is given by $C\mathbf{d}$. Producing $C\mathbf{d}$ requires more inputs. The 2nd round of intermediate demand is given by

$$C(C\mathbf{d}) = C^2\mathbf{d},$$

and so on:

Round	Demand That Must Be Met	Inputs Needed to Meet This Demand
Final demand	$\mathbf{d}$	$C\mathbf{d}$
1st round	$C\mathbf{d}$	$C(C\mathbf{d}) = C^2\mathbf{d}$
2nd round	$C^2\mathbf{d}$	$C(C^2\mathbf{d}) = C^3\mathbf{d}$
3rd round	$C^3\mathbf{d}$	$C(C^3\mathbf{d}) = C^4\mathbf{d}$
$\vdots$	$\vdots$	$\vdots$

The total production needed is therefore

$$\mathbf{x} = \mathbf{d} + C\mathbf{d} + C^2\mathbf{d} + C^3\mathbf{d} + \cdots = (I + C + C^2 + C^3 + \cdots)\mathbf{d}. \tag{15}$$

This can be simplified algebraically as

$$(I - C)(I + C + C^2 + \cdots + C^m) = I - C^{m+1}.$$

If each column sum of C is strictly less than 1, then $C^m \to 0$ as $m \to \infty$ and the infinite sum converges to

$$I + C + C^2 + \cdots = (I - C)^{-1}.$$

For practical computation, we can approximate:

$$(I - C)^{-1} \approx I + C + C^2 + \cdots + C^m \tag{16}$$

for a sufficiently large m, which converges quickly when C has small column sums.

Equation 15 is used to compute $(I - C)\mathbf{x} = \mathbf{d}$ practically and Equation 16 is used to compute $(I - C)^{-1}$ practically.

Each entry in $(I - C)^{-1}$ tells you how much extra output each sector must produce if the final demand for a particular sector changes. For instance, the entries in column j of $(I - C)^{-1}$ are the

increased amounts the various sectors must produce to compensate for an increase of 1 unit in the final demand for output from sector j.

Note that numerically, any system $A\mathbf{x} = \mathbf{b}$ can be written as $(I - C)\mathbf{x} = \mathbf{b}$ with $C = I - A$. If the system is large and sparse (mostly zeros), and the column sums of the absolute values in C are less than 1, then $C^m \to 0$ as $m \to \infty$. If C^m approaches zero quickly, Equations 15 and 16 provide practical ways to solve $A\mathbf{x} = \mathbf{b}$ and compute A^{-1}.

Problem 2.13.1

Suppose an economy has two sectors: goods and services. One unit of output from goods requires inputs of 0.2 unit from goods and 0.5 unit from services. One unit of output from services requires input of 0.4 unit from goods and 0.3 unit from services. The final demand is 20 units of goods and 30 units of services. Set up the Leontief model for this situation.

Solution

Purchased from	Goods	Services	Final Demand
Goods	0.20	0.40	20
Services	0.50	0.30	30

We have $\mathbf{x} = C\mathbf{x} + \mathbf{d}$ where

$$C = \begin{bmatrix} 0.20 & 0.40 \\ 0.50 & 0.30 \end{bmatrix}$$

and

$$\mathbf{d} = \begin{bmatrix} 20 \\ 30 \end{bmatrix}.$$

Then, we get it in the form of $(I - C)\mathbf{x} = \mathbf{d}$:

$$I - C = \begin{bmatrix} 1 - 0.20 & -0.40 \\ -0.50 & 1 - 0.30 \end{bmatrix} = \begin{bmatrix} 0.80 & 0.40 \\ -0.50 & 0.70 \end{bmatrix}$$

Let $\mathbf{x} = \begin{bmatrix} x_g \\ x_s \end{bmatrix}$ be the production vector. Then we have the system:

$$\begin{bmatrix} 0.80 & 0.40 \\ -0.50 & 0.70 \end{bmatrix} \begin{bmatrix} x_g \\ x_s \end{bmatrix} = \begin{bmatrix} 30 \\ 20 \end{bmatrix}$$

For our final application, we have computer graphics.

Consider the following image, which represents the letter N in a program.

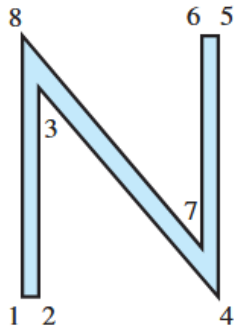


Figure 5: Image credit: Lay et al

It is determined by eight labeled vertices which we can encode in a matrix D:

$$D = \begin{bmatrix} 0 & 0.50 & 0.50 & 6 & 6 & 5.50 & 5.50 & 0 \\ 0 & 0 & 6.42 & 0 & 8 & 8 & 1.58 & 8 \end{bmatrix}$$

Computer programs map line segments onto line segments, so describing them as collections of straight line segments is convenient.

Example 2.13.5

Given $A = \begin{bmatrix} 1 & 0.25 \\ 0 & 1 \end{bmatrix}$, apply the transformation $\mathbf{x} \mapsto A\mathbf{x}$ to the letter N in Figure 4.

Solution

The matrix product AD has columns equal to the transformed vertices:

$$AD = \begin{bmatrix} 0 & 0.5 & 2.105 & 6 & 8 & 7.5 & 5.895 & 2 \\ 0 & 0 & 6.420 & 0 & 8 & 8 & 1.580 & 8 \end{bmatrix}.$$

N now is slanted:

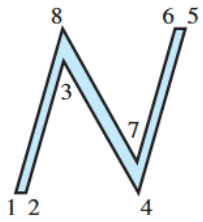


Image credit: Lay et al

While the mathematics behind computer graphics run on matrix multiplication, the translation of an object is not directly matrix multiplication because a translation is not a linear transformation. To avoid this, there is something called *homogeneous coordinates*. For 2D homogeneous coordinates, we can take a point $(x, y) \in \mathbb{R}^2$ and write it as $(x, y, 1) \in \mathbb{R}^3$. This extra dimension lets us perform transformations using a 3×3 matrix. In 3D homogeneous coordinates, a point $(x, y, z) \in \mathbb{R}^3$ becomes $(x, y, x, 1) \in \mathbb{R}^4$. In general, (X, Y, Z, H) are homogeneous coordinates for (x, y, z) if $h \neq 0$ and

$$x = \frac{X}{H}, \quad y = \frac{Y}{H}, \quad \text{and} \quad z = \frac{Z}{H}.$$

If we scale each value by a nonzero scalar, we can produce homogeneous coordinates where the true point stays the same.

While you may immediately think of CAD, FEM, or video games, there are also applications of this in biology. The following example is all about molecular engineering, where we model drug and protein design.

Example 2.13.6

Give 4×4 matrices for the following transformations:

(a) Rotation about the y-axis by an angle of $30°$. By convention, a positive angle is the counterclockwise direction when looking toward the origin from the positive half of the axis of rotation.

(b) Translation by the vector $\mathbf{p} = (-6, 4, 5)$.

Solution

(a) The rotation about the y-axis leaves $\mathbf{e}_2$ fixed, rotates $\mathbf{e}_1$ toward the negative z-axis, and rotates $\mathbf{e}_3$ toward the positive x-axis.

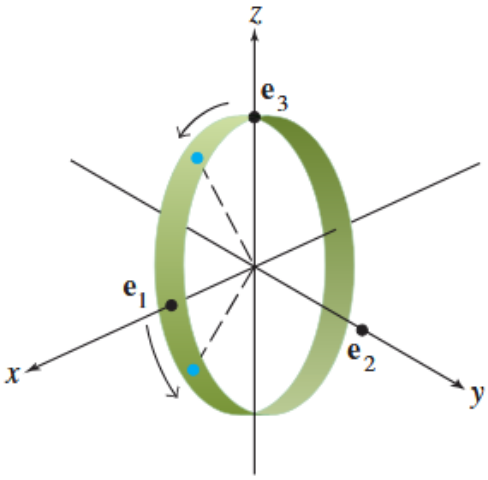


Image credit: Lay et al

We plug 30° into

$$R_y = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$$

to get

$$R_y = \begin{bmatrix} \frac{\sqrt{3}}{2} & 0 & 0.5 \\ 0 & 1 & 0 \\ -0.5 & 0 & \frac{\sqrt{3}}{2} \end{bmatrix}.$$

We now convert this to homogenous coordinates by adding a 4th row and column:

$$A = \begin{bmatrix} \frac{\sqrt{3}}{2} & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 0 \\ -0.5 & 0 & \frac{\sqrt{3}}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(b) To map $(x, y, z, 1)$ to $(x-6,\, y+4,\, z+5,\, 1)$, we add this vector to the standard translation matrix:

$$T = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow T \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x + t_x \\ y + t_y \\ z + t_z \\ 1 \end{bmatrix}$$

We have $t_x = -6$, $t_y = 4$, and $t_z = 5$:

$$T = \begin{bmatrix} 1 & 0 & 0 & -6 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The rotation can reorient a molecule to align a functional group with a target and the translation positions the molecule in the binding site.

We see three-dimensional objects on our two-dimensional screens by mapping them onto a *viewing plane*. For simplicity, let the xy-plane represent the screen, and suppose your is at $(0, 0, d)$ along the z-axis, looking toward the origin. Each point (x, y, z) is mapped to $(x^*, y^*, 0)$ on the viewing plane so that the eye position, the point, and its projection are collinear. The point where everything converges is called the *center of projection*.

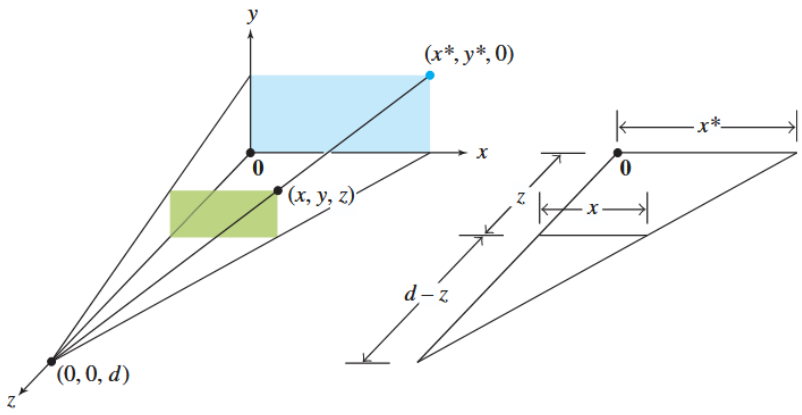


Image credit: Lay et al

From the xz-plane diagram, similar triangles give:

$$\frac{x^*}{d} = \frac{x}{d-z} \quad \text{and} \quad x^* = \frac{dx}{d-z} = \frac{x}{1-z/d}.$$

Similarly, in the yz-plane,

$$y^* = \frac{y}{1-z/d}.$$

Thus, the perspective projection is

$$(x, y, z) \mapsto \left(\frac{x}{1-z/d},\ \frac{y}{1-z/d},\ 0 \right),$$

where the factor $\frac{1}{1-z/d}$ accounts for the depth-based scaling. Points with z close to d appear very large (near the eye) and points with z far behind the viewing plane appear smaller.

We can represent this mapping with a 4×4 matrix in homogeneous coordinates.

We want

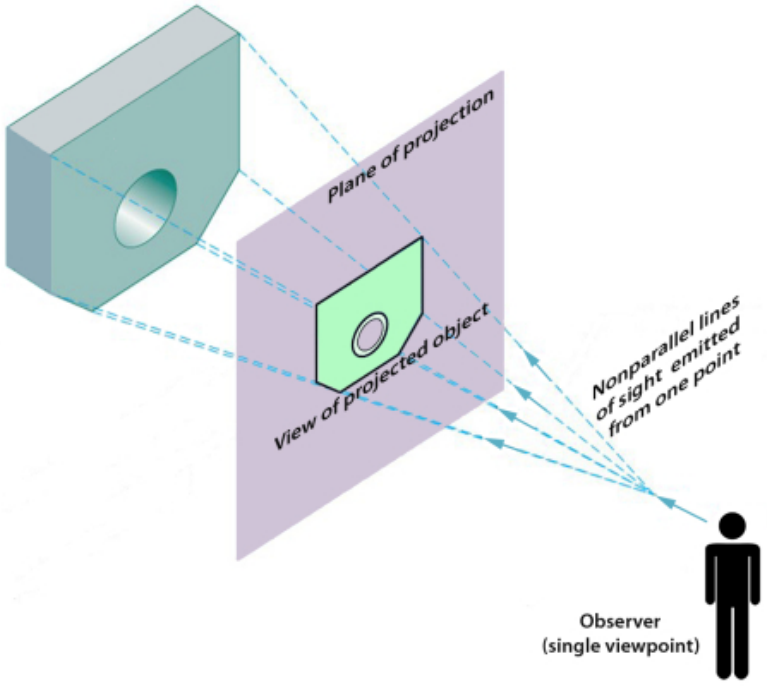
$$(x, y, z, 1) \mapsto \left(\frac{x}{1 - z/d},\ \frac{y}{1 - z/d},\ 0,\ 1 \right).$$

We can now display P:

$$P = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{d} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ 0 \\ 1 - \frac{z}{d} \end{bmatrix}.$$

In homogeneous coordinates, we divide by the fourth entry:

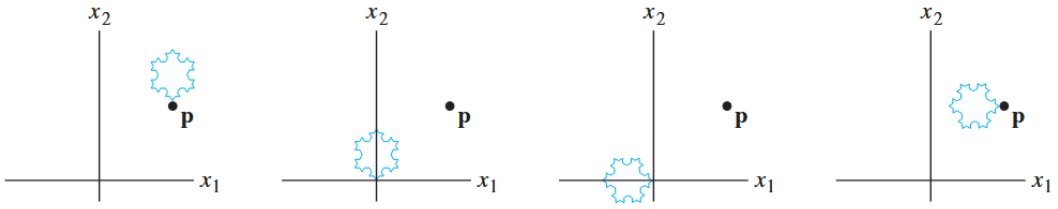
$$\left[x,\ y,\ 0,\ 1 - \frac{z}{d} \right]^T \mapsto \left[\frac{x}{1 - z/d},\ \frac{y}{1 - z/d},\ 0,\ 1 \right]^T.$$



A projection in CAD. Image credit: McGill University

Problem 2.13.2

We want to rotate a figure about a point $\mathbf{p} \in \mathbb{R}^2$. The process is to translate the figure by $-\mathbf{p}$, rotate it about the origin, and then translate it back by $\mathbf{p}$:



(a) Original figure. (b) Translated to origin by $-\mathbf{p}$. (c) Rotated about the origin. (d) Translated back by $\mathbf{p}$.

Image credit: Lay et al

Construct the 3×3 matrix that rotates points $-30°$ about the point $(-2, 6)$ using homogeneous coordinates.

Solution

First, we have translation by $-\mathbf{p}$:

$$T_{-\mathbf{p}} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -6 \\ 0 & 0 & 1 \end{bmatrix}$$

Second, we have rotation by $-30°$. Given $\cos(-30°) = \frac{\sqrt{3}}{2}$ and $\sin(-30°) = -\frac{1}{2}$,

$$R_{-30°} = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Third, we have translation by $\mathbf{p}$:

$$T_{\mathbf{p}} = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 6 \\ 0 & 0 & 1 \end{bmatrix}$$

We combine the transformations:

$$
\begin{aligned}
M &= T_{\mathbf{p}} \cdot R_{-30^\circ} \cdot T_{-\mathbf{p}} \\
&= \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -6 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} & \sqrt{3}-3 \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} & -1-3\sqrt{3} \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} & \sqrt{3}-5 \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} & -3\sqrt{3}+5 \\ 0 & 0 & 1 \end{bmatrix}
\end{aligned}
$$

3 Determinants

In the early 19th century, it was Cauchy who formalized the idea of the determinant, building on earlier work* by Leibniz and Cramer. Back then, matrices were small and determinants were an essential tool in the likes of analytic geometry. However, determinants are not very useful in modern day computing because they are simply too inefficient to compute. So while most of the time we resort to matrix factorizations, determinants still reveal important information about data encoded in matrices and help us develop more intuition about linear algebra.

*A Japanese mathematician named Seki Takakazu also independently discovered the determinant.

3.1 Intro to Determinants

A **permutation** of a nonempty set A is a bijective function

$$\pi : A \to A$$

In other words, a permutation is a map. If π is one-to-one and $\pi(a) = \pi(b)$, then $a = b$. Furthermore, if π is onto, then the codomain of π is equal to the range of π.

Let

$$[n] = \{j \in \mathbb{N} : 1 \leq j \leq n\}$$

The set of all permutations of $[n]$ is denoted as S_n, and contains all one-to-one maps from $[n]$ to itself. For determinants, we focus on permutations of the set

$$[n] = \{1, 2, \ldots, n\}.$$

The number of elements in S_n is $n!$ since there are n choices for $\pi(1)$, $n-1$ choices for $\pi(2)$, and so on. These choices of course get multiplied together:

If S_n is the set of all bijections from $[n]$ to $[n]$, then

$$n! = n \cdot (n-1) \cdots 2 \cdot 1.$$

Cauchy's two-line notation for a permutation of the integers

$$\pi : \{1, 2, \cdots, n\} \to \{1, 2, \cdots, n\}$$

is given as

$$\begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ \pi(1) & \pi(2) & \pi(3) & \cdots & \pi(n) \end{pmatrix}.$$

As you can see, the image of i is $\pi(i)$ under the map π.

For example, we can list all of the permutations of S_2, the permutation group on a set with two elements. We want to find all bijective maps given by

$$\pi : \{1, 2\} \to \{1, 2\}.$$

There are $2! = 2$ such permutations. They are

$$\pi_1 = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$$

and

$$\pi_2 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Let $\pi \in S_n$. From $[n]$, we can select an i, j with $1 \leq i < j \leq n$.The pair (i, j) is called an **inversion** of π if and only if

$$\pi(i) > \pi(j).$$

One way to check if (i, j) is an inversion with respect to π is to calculate the ratio

$$\frac{\pi(i) - \pi(j)}{i - j}.$$

If the ratio is less than 0, it means that (i, j) is an inversion with respect to π. If the ratio is greater than 0, then the pair is not an inversion with respect to π.

The set of all inversions of π is denoted

$$\text{Inv}(\pi) = \{(i,j) : i,j \in [n], i < j,\ \pi(i) > \pi(j)\}.$$

The number of inversions is $n(\pi) = |\text{Inv}(\pi)|$.

The sign of π is defined through a function:

$$\text{sgn}(\pi) = (-1)^{n(\pi)} = \begin{cases} +1, & \text{if } n(\pi) \text{ is even,} \\ -1, & \text{if } n(\pi) \text{ is odd.} \end{cases}$$

Now that we have set the foundation for the *permutation* definition of the determinant, we will actually take a break and explain the *recursive definition* as it is more intuitive.

Previously, we learned about how a 2×2 matrix is invertible if and only if its determinant is nonzero. We can extend this to a general $n \times n$ matrix by first looking at a 3×3 matrix. Consider $A = [a_{i,j}]$ with $a_{1,1} \neq 0$. If we multiply the second and third rows of A by $a_{1,1}$ and then subtract the proper multiples of the first row from the other two rows, we get

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} \sim \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ 0 & a_{1,1}a_{2,2} - a_{1,2}a_{2,1} & a_{1,1}a_{2,3} - a_{1,3}a_{2,1} \\ 0 & a_{1,1}a_{3,2} - a_{1,2}a_{3,1} & a_{1,1}a_{3,3} - a_{1,3}a_{3,1} \end{bmatrix}. \tag{17}$$

Since A is invertible, either the $(2,2)$-entry or the $(3,2)$-entry on the right in (17) is nonzero. If we assume that the $(2,2)$-entry is nonzero, we can multiply row 3 by $a_{1,1}a_{2,2} - a_{1,2}a_{2,1}$. Then, we add add $-a_{1,1}a_{3,2} - a_{1,2}a_{3,1}$ multiplied by row 2 to the new row 3. We do this to eliminate the entry $(3,2)$-entry:

$$A \sim \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ 0 & a_{1,1}a_{2,2} - a_{1,2}a_{2,1} & a_{1,1}a_{2,3} - a_{1,3}a_{2,1} \\ 0 & 0 & a_{1,1}\Delta \end{bmatrix}$$

where $\Delta = a_{1,1}a_{2,2}a_{3,3} + a_{1,2}a_{2,3}a_{3,1} + a_{1,3}a_{2,1}a_{3,2} - a_{1,1}a_{2,3}a_{3,2} - a_{1,2}a_{2,1}a_{3,3} - a_{1,3}a_{2,2}a_{3,1}$.

Since A is invertible, the determinant of this 3×3 matrix Δ must be nonzero. The converse of this is also true.

Recall the case for the 2×2 matrix. If

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

then

$$\det(A) = a_{1,1}a_{2,2} - a_{1,2}a_{2,1}.$$

We can group the terms in Δ and thus write

$$\Delta = a_{1,1} \cdot \det \begin{bmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{bmatrix} - a_{1,2} \cdot \det \begin{bmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{bmatrix} + a_{1,3} \cdot \det \begin{bmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{bmatrix}.$$

This leads to the cofactor expansion:

$$\det(A) = a_{1,1} \det(A_{1,1}) - a_{1,2} \det(A_{1,2}) + a_{1,3} \det(A_{1,3}), \tag{18}$$

where $A_{1,1}$, $A_{1,2}$, and $A_{1,3}$ are obtained from A by deleting the first row and one of the three columns. In general, for any square matrix A, $A_{i,j}$ is the submatrix obtained by deleting the ith row and jth column of A.

For instance, consider the following matrix:

$$A = \begin{bmatrix} 1 & -2 & 5 & 0 \\ 2 & 0 & 4 & -1 \\ 3 & 1 & 0 & 7 \\ 0 & 4 & -2 & 0 \end{bmatrix}$$

By deleting row 3 and column 2, we get the submatrix

$$A_{3,2} = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}.$$

When $n = 3$, the determinant is defined using determinants of the 2×2 submatrices. If $n = 4$, the determinant would be defined using determinants of the 3×3 submatrices. In general, an $n \times n$

determinant is defined using determinants of the $(n-1)\times(n-1)$ submatrices. We now arrive at our *recursive* definition of the determinant.

Recursive Definition of the Determinant

For $n \geq 2$, the **determinant** of an $n \times n$ matrix $A = [a_{i,j}]$ is the sum of n terms of the form $\pm a_{1,j}\ \det(A_{1,j})$, with plus and minus signs alternating, where the entries $a_{1,1}, a_{1,2}, \ldots, a_{1,n}$ are from the first row of A. This is written as

$$\begin{aligned}\det(A) &= a_{1,1}\det(A_{1,1}) - a_{1,2}\det(A_{1,2}) + \cdots + (-1)^{1+n}a_{1,n}\det(A_{1,n}) \\ &= \sum_{j=1}^{n}(-1)^{1+j}\, a_{1,j}\ \det(A_{1,j}).\end{aligned}$$

Example 3.1.1

Compute the determinant of $A = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}$.

Solution

We expand along the first row:

$$\det(A) = a_{1,1}\det(A_{1,1}) - a_{1,2}\det(A_{1,2}) + a_{1,3}\det(A_{1,3})$$

Here,

$$\det(A) = 1 \cdot \det\begin{bmatrix} 4 & -1 \\ -2 & 0 \end{bmatrix} - 5 \cdot \det\begin{bmatrix} 2 & -1 \\ 0 & 0 \end{bmatrix} + 0 \cdot \det\begin{bmatrix} 2 & 4 \\ 0 & -2 \end{bmatrix}.$$

Let's compute each 2×2 determinant:

$$\det\begin{bmatrix} 4 & -1 \\ -2 & 0 \end{bmatrix} = (4)(0) - (-1)(-2) = 0 - 2 = -2$$

$$\det\begin{bmatrix} 2 & -1 \\ 0 & 0 \end{bmatrix} = (2)(0) - (-1)(0) = 0 - 0 = 0$$

$$\det\begin{bmatrix} 2 & 4 \\ 0 & -2 \end{bmatrix} = (2)(-2) - (4)(0) = -4 - 0 = -4$$

And thus we have

$$\det(A) = 1(-2) - 5(0) + 0(-4) = -2 - 0 + 0 = -2.$$

The most common notation for the determinant uses vertical bars instead of brackets:

$$\det(A) = 1\begin{vmatrix} 4 & -1 \\ -2 & 0 \end{vmatrix} - 5\begin{vmatrix} 2 & -1 \\ 0 & 0 \end{vmatrix} + 0\begin{vmatrix} 2 & 4 \\ 0 & -2 \end{vmatrix} = -2$$

Let's discuss the permutation definition of the determinant. Let $A \in \mathbb{R}^{n\times n}$. Then there is a unique determinant function

$$\det : \mathbb{R}^{n\times n} \to \mathbb{R}$$

given by

$$\det(A) = \sum_{\pi\in S_n} \operatorname{sgn}(\pi) \cdot a_{1,\pi(1)} \cdot a_{2,\pi(2)} \cdots a_{n,\pi(n)},$$

where S_n is the set of all permutations of $\{1, 2, \ldots, n\}$ and $\operatorname{sgn}(\pi)$ is the sign of the permutation π.

Then, for any 2×2 matrix, we see that

$$\begin{aligned}\det(A) &= \sum_{\pi \in S_n} \operatorname{sgn}(\pi) \cdot a_{1,\pi(1)} \cdot a_{2,\pi(2)} \\ &= \sum_{i=1}^{2} \operatorname{sgn}(\pi_i) \cdot a_{1,\pi_i(1)} \cdot a_{2,\pi_i(2)} \\ &= \operatorname{sgn}(\pi_1) \cdot a_{1,\pi_1(1)} \cdot a_{2,\pi_1(2)} + \operatorname{sgn}(\pi_2) \cdot a_{1,\pi_2(1)} \cdot a_{2,\pi_2(2)} \\ &= a_{1,1} \cdot a_{2,2} - a_{1,2} \cdot a_{2,1}\end{aligned}$$

The permutation definition is too tedious for practical computation, but it still is the most general way to represent the determinant. A permutation provide a systemic way to encode all possible ways of picking one entry from each row and each column in a square matrix, which is what the determinant sums over. The sign of each permutation accounts for orientation too, which is important because the determinant needs to change sign when rows are swapped.

We can rewrite the recursive determinant formula in terms of **cofactors**. For a matrix $A = [a_{i,j}]$, the (i, j)-cofactor of A is

$$C_{i,j} = (-1)^{i+j} \det(A_{i,j}),$$

where $A_{i,j}$ is the submatrix obtained by deleting the ith row and jth column from A.

Using cofactors, the determinant can be computed by cofactor expansion along any row or column.

Expansion along the ith row is given by

$$\det(A) = a_{i,1}C_{i,1} + a_{i,2}C_{i,2} + \cdots + a_{i,n}C_{i,n}.$$

Expansion along the jth column is given by

$$\det(A) = a_{1,j}C_{1,j} + a_{2,j}C_{2,j} + \cdots + a_{n,j}C_{n,j}.$$

The sign factor $(-1)^{i+j}$ produces a checkerboard pattern of signs:

$$\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \cdots \\ + & - & + & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Example 3.1.2

Let

$$A = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}.$$

Compute $\det(A)$ by expanding along row 3.

Solution

$$\begin{aligned}
\det(A) &= a_{3,1}C_{3,1} + a_{3,2}C_{3,2} + a_{3,3}C_{3,3} \\
&= a_{3,1}(-1)^{3+1}\det(A_{3,1}) + a_{3,2}(-1)^{3+2}\det(A_{3,2}) + a_{3,3}(-1)^{3+3}\det(A_{3,3}) \\
&= 0(1)\begin{vmatrix} 5 & 0 \\ 4 & -1 \end{vmatrix} + (-2)(-1)\begin{vmatrix} 1 & 0 \\ 2 & -1 \end{vmatrix} + 0(1)\begin{vmatrix} 1 & 5 \\ 2 & 4 \end{vmatrix} \\
&= 0 + 2\big((1)(-1) - (0)(2)\big) + 0 \\
&= 2(-1) \\
&= -2.
\end{aligned}$$

Example 3.1.3

Consider the following matrix:

$$A = \begin{bmatrix} 3 & -7 & 8 & 9 & -6 \\ 0 & 2 & -5 & 7 & 3 \\ 0 & 0 & 1 & 5 & 0 \\ 0 & 0 & 2 & 4 & -1 \\ 0 & 0 & 0 & -2 & 0 \end{bmatrix}$$

Compute $\det(A)$ using cofactor expansion.

Solution

We notice that the first column of A has four zeros, so a cofactor expansion down this column leaves only the first term:

$$\det(A) = 3\begin{vmatrix} 2 & -5 & 7 & 3 \\ 0 & 1 & 5 & 0 \\ 0 & 2 & 4 & -1 \\ 0 & 0 & -2 & 0 \end{vmatrix}.$$

The first column of this 4×4 determinant has three zeros, so we expand down this column

$$\det(A) = 3(2)\begin{vmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{vmatrix}.$$

The 3×3 determinant as computed in Example 3.1.2 is -2. Thus,

$$\det(A) = 3(2)(-2) = -12.$$

If A is a triangular matrix, than $\det(A)$ is the product of the entries on the main diagonal of A. If an entire row or column consists of zeros, the cofactor expansion along it is simply the sum of zeros and thus the determinant is zero. Unfortunately, most cofactor expansions are not calculated this quickly. For a 25×25 matrix, it would require $n! = 26! = 1.5 \times 10^{25}$ multiplication operations to compute the determinant by cofactor expansion. If we assume a computer performs one trillion operations per second, it would still have to run for more than $500,000$ years to compute this matrix. And this is a small matrix in comparison to what you will likely see in the real world.

Problem 3.1.1

Compute $\begin{vmatrix} 5 & -7 & 2 & 2 \\ 0 & 3 & 0 & -4 \\ -5 & -8 & 0 & 3 \\ 0 & 5 & 0 & -6 \end{vmatrix}$.

Solution

Since there are several zeros in the third column, we should expand along it first. Begin with a cofactor expansion down the third column to obtain a 3×3 matrix, which may be evaluated by an expansion down its first column:

$$\begin{vmatrix} 5 & -7 & 2 & 2 \\ 0 & 3 & 0 & -4 \\ -5 & -8 & 0 & 3 \\ 0 & 5 & 0 & -6 \end{vmatrix} = (-1)^{1+3} \cdot 2 \begin{vmatrix} 0 & 3 & -4 \\ -5 & -8 & 3 \\ 0 & 5 & -6 \end{vmatrix}$$

$$= 2 \cdot (-1)^{2+1}(-5) \begin{vmatrix} 3 & -4 \\ 5 & -6 \end{vmatrix}$$

$$= 2(-1)(-5) \cdot (3(-6) - (-4)(5))$$

$$= 10 \cdot 2 = 20.$$

3.2 Properties of the Determinant

We can use the following properties to analyze matrices and compute determinants more efficiently.

How Row Operations Affect the Determinant

Let A be a square matrix.

(a) If a multiple of one row of A is added to another row to produce a matrix B, then

$$\det(B) = \det(A).$$

(b) If two rows of A are interchanged to produce B, then

$$\det(B) = -\det(A).$$

(c) If one row of A is multiplied by k to produce B, then

$$\det(B) = k \cdot \det(A).$$

Example 3.2.1

Compute $\det(A)$ where $A = \begin{bmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{bmatrix}$

Solution

We need to reduce A to REF and then use the fact that the determinant of a triangular matrix is just the product of its diagonal entries. We begin with row operations that do not change the determinant:

$$\det(A) = \begin{vmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{vmatrix} = \begin{vmatrix} 1 & -4 & 2 \\ 0 & 0 & -5 \\ -1 & 7 & 0 \end{vmatrix} = \begin{vmatrix} 1 & -4 & 2 \\ 0 & 0 & -5 \\ 0 & 3 & 2 \end{vmatrix}$$

Now we have to swap rows 2 and 3 so that we get zeros below the main diagonal in column 2. This flips the sign of the determinant.

$$\det(A) = -\begin{vmatrix} 1 & -4 & 2 \\ 0 & 3 & 2 \\ 0 & 0 & -5 \end{vmatrix}$$

The matrix is now upper triangular, so we compute the product of its diagonal entries:

$$\det(A) = -(1)(3)(-5) = 15$$

Example 3.2.2

Compute $\det(A)$ where $A = \begin{bmatrix} 2 & -8 & 6 & 8 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{bmatrix}$.

Solution

We want a 1 in the upper-left corner to make row operations easier. Instead of swapping rows, we will factor out 2 from the top row and then use row replacement to zero out the entries below it in column 1.

$$\det(A) = 2\begin{vmatrix} 1 & -4 & 3 & 4 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{vmatrix} = 2\begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & -12 & 10 & 10 \\ 0 & 0 & -3 & 2 \end{vmatrix}$$

We choose 3 as the pivot and then multiply row 2 by 4 and add it to row 3 to zero out the entry below the pivot.

$$\det(A) = 2\begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & 0 & -6 & 2 \\ 0 & 0 & -3 & 2 \end{vmatrix}$$

Then we multiply row 3 by $-\frac{1}{2}$ and add it to row 4 to zero out the -3 in row 4. This results in an upper triangular matrix:

$$\det(A) = 2\begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & 0 & -6 & 2 \\ 0 & 0 & 0 & 1 \end{vmatrix}$$

Finally, we can compute the determinant by multiplying the diagonal entries:

$$2 \cdot (1)(3)(-6)(1) = -36$$

If you row reduce a matrix A to an REF matrix U using a total of r row swaps, then

$$\det(A) = (-1)^r \det(U).$$

Since U is in REF, it is triangular, so

$$\det(U) = \prod_{i-1}^{n} u_{i,i} = u_{1,1}u_{2,2}\cdots u_{n,n}. \tag{19}$$

In other words, $\det(U)$ is the product of the diagonal entries. And if A is invertible, all of the entries $u_{i,i}$ are pivots. Thus, we can refer to Equation 19 as the product of pivots. Remember that if any pivot is 0, the product is 0. To summarize, we have the following:

$$\det(A) = \begin{cases} (-1)^r \prod_{i=1}^{n} u_{i,i}, & \text{when } A \text{ is invertible,} \\ 0, & \text{when } A \text{ is not invertible.} \end{cases}$$

U is not unique because it's only in REF, and neither are its pivots. That being said, the product of pivots does have a unique value. It can however have its sign changed. Here we have the typical forms of REF square matrices:

$$U = \begin{bmatrix} \blacksquare & \star & \star & \star \\ 0 & \blacksquare & \star & \star \\ 0 & 0 & \blacksquare & \star \\ 0 & 0 & 0 & \blacksquare \end{bmatrix} \tag{20}$$

$$U = \begin{bmatrix} \blacksquare & \star & \star & \star \\ 0 & \blacksquare & \star & \star \\ 0 & 0 & 0 & \blacksquare \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{21}$$

In Equation 20, all pivots are nonzero and thus $\det(U) \neq 0$. In Equation 21, one pivot is zero and thus $\det(U) = 0$.

A square matrix A is invertible if and only if $\det(A) \neq 0$. If $\det(A) = 0$, it means the columns of A are linearly dependent. As a reminder, this means that one of them can be written as a linear combination of the others. Since the rows of A are columns of A^T, A^T has linearly dependent columns which make A^T singular. By the invertible matrix theorem, when A^T is singular, so is A. Linear dependence is only clear when two columns or two rows are identical or a column or row is entirely zeros.

Example 3.2.1

Compute $\det(A)$ for $A = \begin{bmatrix} 3 & -1 & 2 & -5 \\ 0 & 5 & -3 & -6 \\ -6 & 7 & -7 & 4 \\ -5 & -8 & 0 & 9 \end{bmatrix}$.

Solution

We add multiply row 1 by 2 and add it to row 3 to obtain

$$\det(A) = \det \begin{bmatrix} 3 & -1 & 2 & -5 \\ 0 & 5 & -3 & -6 \\ 0 & 5 & -3 & -6 \\ -5 & -8 & 0 & 9 \end{bmatrix}.$$

Since rows 2 and 3 are identical, the determinant is zero:

$$\det(A) = 0$$

So far, we have discussed how row operations affect the determinant of a matrix. Because the determinant of a matrix is equal to the determinant of its transpose, the similar properties also apply to column operations.

Transpose of a Determinant

If A is an $n \times n$ matrix, then $\det(A^T) = \det(A)$.

The determinant function treats rows and columns in the same way, so the number you get is the same before and after transposing. When $n = 1$, the transpose is of course the same number.

Let's assume our transpose of a determinant theorem is true for all $k \times k$ matrices, and consider a $(k+1) \times (k+1)$ matrix A. Suppose we expand $\det(A)$ along the first row and expand $\det(A^T)$ along the first column. The cofactor of $a_{1,j}$ in A equals the cofactor of a_{j1} in A^T because these cofactors involve $k \times k$ determinants, which are equal by the induction hypothesis. Therefore, the cofactor expansion of $\det(A)$ along the first row equals the cofactor expansion of $\det(A^T)$ down the first column. This shows that A and A^T have equal determinants for any $n \geq 1$.

Multiplicative Property of Determinants

If A and B are $n \times n$ matrices, then

$$\det(AB) = \det(A)\ \det(B).$$

Proof of the Multiplicative Property of Determinants

If A is not invertible, then neither is AB; if you compose a linear map with a linear map that is not invertible, the composition cannot be invertible. Thus

$$\det(AB) = \det(A)\det(B) = 0.$$

Now, let's assume that A *is* invertible. Then, A and the identity matrix I_n are row equivalent by the invertible matrix theorem. We need elementary matrices because for any square matrix A,

$$|EX| = |E||X|.$$

Therefore there exists elementary matrices $E_1, \ldots, E_p$ such that

$$A = E_p E_{p-1} \cdots E_1 I_n = E_p E_{p-1} \cdots E_1.$$

This means that multiplying B on the left side by A is the same as multiplying B successively by $E_1, E_2, \ldots, E_p$.

We apply the determinant rules one matrix at a time:

$$\begin{aligned} |AB| &= |E_p E_{p-1} \cdots E_1 B| \\ &= |E_p|\,|E_{p-1} \cdots E_1 B| \\ &= |E_p|\,|E_{p-1}||E_{p-2} \cdots E_1 B| \\ &\vdots \\ &= |E_p| \cdots |E_1||B|. \end{aligned}$$

Since

$$|A| = |E_p E_{p-1} \cdots E_1| = |E_p| \cdots |E_1|,$$

we conclude

$$|AB| = |A||B|.$$

□

Example 3.2.2

Test out the multiplicative property of determinants for $A = \begin{bmatrix} 6 & 1 \\ 3 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}$.

Solution

$$AB = \begin{bmatrix} 6 & 1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} (6)(4)+(1)(1) & (6)(3)+(1)(2) \\ (3)(4)+(2)(1) & (3)(3)+(2)(2) \end{bmatrix} = \begin{bmatrix} 25 & 20 \\ 14 & 13 \end{bmatrix}$$

Now we compute the determinants:

$$\det(AB) = (25)(13) - (20)(14) = 45$$
$$\det(A) = (6)(2) - (1)(3) = 9 \quad \text{and} \quad \det(B) = (4)(2) - (3)(1) = 5$$

Finally, we can verify that

$$(\det(A))(\det(B)) = (9)(5) = 45 = \det(AB).$$

A common misconception is to think a similar rule holds for addition, but generally,

$$\det(A + B) \neq \det(A) + \det(B).$$

For an $n \times n$ matrix A, $\det(A)$ is a function of the n column vectors in A. Furthermore, if all column vectors except one are fixed, the determinant is a linear function of one variable which is a vector.

Suppose we hold all columns fixed except the jth. Then,

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{j-1} & \cdots & \mathbf{x} & \mathbf{a}_{j+1} & \cdots & \mathbf{a}_1 \end{bmatrix}.$$

We can define a transformation $T : \mathbb{R}^n \to \mathbb{R}$ by

$$T(\mathbf{x}) = \det \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{j-1} & \cdots & \mathbf{x} & \mathbf{a}_{j+1} & \cdots & \mathbf{a}_1 \end{bmatrix}.$$

So $\mathbf{x}$ is the column vector you put in the jth spot and T is just a function that takes in $\mathbf{x}$ and outputs the determinant of the whole matrix. For this, linearity rules apply:

For all scalars $c \in \mathbb{R}^n$ and all $\mathbf{x} \in \mathbb{R}^n$,

$$T(c\mathbf{x}) = cT(\mathbf{x}).$$

In other words, scaling a single column vector by c scales the determinant.

And for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}).$$

In other words, splitting one column into two parts splits the determinant into two determinants.

Problem 3.2.1

Use a determinant to decide if

$$\mathbf{v}_1 = \begin{bmatrix} 5 \\ -7 \\ 9 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -3 \\ 3 \\ -5 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2 \\ -7 \\ 5 \end{bmatrix}$$

are linearly independent.

Solution

Form the matrix with these vectors as columns:

$$A = \begin{bmatrix} 5 & -3 & 2 \\ -7 & 3 & -7 \\ 9 & -5 & 5 \end{bmatrix}.$$

The columns are linearly independent therefore $\det(A) \neq 0$.

$$\det(A) = \begin{vmatrix} 5 & -3 & 2 \\ -7 & 3 & -7 \\ 9 & -5 & 5 \end{vmatrix}.$$

We add row 1 to row 2:

$$= \begin{vmatrix} 5 & -3 & 2 \\ -2 & 0 & -5 \\ 9 & -5 & 5 \end{vmatrix}$$

Expand along the second column:

$$\begin{aligned} \det(A) &= -(-3)\begin{vmatrix} -2 & -5 \\ 9 & 5 \end{vmatrix} - (-5)\begin{vmatrix} 5 & 2 \\ -2 & -5 \end{vmatrix} \\ &= 3((-2)(5) - (9)(-5)) - (-5)((5)(-5) - (-2)(2)) \\ &= 3(-10 + 45) - (-5)(-25 + 4) \\ &= 3(35) - (-5)(-21) \\ &= 105 - 105 \\ &= 0. \end{aligned}$$

A is not invertible. Since $\det(A) = 0$, the the vectors are linearly dependent by the invertible matrix theorem.

Problem 3.2.2

Let A and B be 3×3 matrices with $\det(A) = -3$ and $\det(B) = 4$. Use properties of determinants to compute:

(a) $\det(AB)$

(b) $\det(5A)$

(c) $\det(B^T)$

(d) $\det(A^{-1})$

(e) $\det(A^3)$

Solution

(a) $\det(AB) = \det(A)\det(B) = (-3)(4) = -12$

(b) $\det(5A) = 5^3 \det(A)$, since scaling a 3×3 matrix by 5 multiplies the determinant by 5^3. Thus $125(-3) = -375$.

(c) $\det(B^T) = \det(B) = 4$

(d) $\det(A^{-1}) = \frac{1}{\det(A)} = \frac{1}{-3}$

(e) $\det(A^3) = (\det(A))^3 = (-3)^3 = -27$

Problem 3.2.3

Determine if each of the following are true or false and justify your answer briefly.

(a) $\det[(A+B)(A-B)] = \det(A^2 - B^2)$

(b) If $\det(A) \neq 0$, then there will always be a solution to $A\mathbf{x} = \mathbf{b}$.

(c) If $A \in \mathbb{R}^{n \times n}$ is nonsingular, then $\det(I) = \det(A)\det(A^{-1})$.

(d) If $\det(A) = 0$ for $A \in \mathbb{R}^{n \times n}$, then the corresponding system $A\mathbf{x} = \mathbf{b}$ will be inconsistent for all $\mathbf{b} \in \mathbb{R}^n$

(e) For $A \in \mathbb{R}^{n \times n}$, we have $\det(-A) = -\det(A)$.

Solution

(a) False, matrix multiplication is generally not commutative.

(b) True, because $\det(A) \neq 0$ means that A is invertible. We can multiply both sides by A^{-1} to get the solution $\mathbf{x} = A^{-1}\mathbf{b}$.

(c) True, because if A is invertible then $AA^{-1} = I$ and then we can take the determinants to get $\det(AA^{-1}) = \det(I)$. By the multiplicative property, $\det(AA^{-1}) = \det(A)\det(A^{-1})$ which means that $1 = \det(A)\det(A^{-1})$.

(d) False, because $\det(A) = 0$ means that A is singular and thus not invertible. This does not mean that every system is inconsistent. It simply means that A cannot produce every possible $\mathbf{b}$.

(e) False, the correct statement would be $\det(-A) = (-1)^n \det(A)$.

3.3 Cramer's Rule and Applications of Determinants

Cramer's rule is a way for us to solve systems of linear equations using determinants. While this method isn't really common in real-world applications, it helps us develop our theoretical understanding of the relationship between determinants and solutions of $A\mathbf{x} = \mathbf{b}$. Thus, we mostly use it for 2×2 and 3×3 matrices.

For any $n \times n$ matrix A and any $\mathbf{b} \in \mathbb{R}^n$, let $A_i(\mathbf{b})$ denote the matrix obtained from A by replacing its ith column with $\mathbf{b}$:

$$Ai(\mathbf{b}) = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{b} & \cdots & a_n \end{bmatrix}$$

We can build one of these matrices for each variable x_i. Then, we arrive at the following:

Cramer's rule

Let $A \in \mathbb{R}^n$ is an invertible matrix. Then for any $\mathbf{b} \in \mathbb{R}^n$, the unique solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ has entries given by

$$x_i = \frac{\det(Ai(\mathbf{b}))}{\det(A)}$$

for $i = 1, 2, \ldots, n$.

Proof of Cramer's rule

We will denote the columns of A by $\mathbf{a}_1, \dots, \mathbf{a}_n$ and the columns of $I \in \mathbb{R}^{n \times n}$ by $\mathbf{e}_1, \dots, \mathbf{e}_n$. Thus,

$$I_i(\mathbf{x}) = \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{x} & \cdots & \mathbf{e}_n \end{bmatrix}$$

If $A\mathbf{x} = \mathbf{b}$, matrix multiplication shows that

$$\begin{aligned} A \cdot I_i(\mathbf{x}) &= A \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{x} & \cdots & \mathbf{e}_n \end{bmatrix} \\ &= \begin{bmatrix} A\mathbf{e}_1 & \cdots & A\mathbf{x} & \cdots & A\mathbf{e}_n \end{bmatrix} \\ &= \begin{bmatrix} a_1 & \cdots & \mathbf{b} & \cdots & a_n \end{bmatrix} \\ &= A_i(\mathbf{b}). \end{aligned}$$

Taking determinants and using the fact that $\det(AB) = \det(A)\det(B)$ gives

$$\det(A)\, \det\big(I_i(\mathbf{x})\big) = \det\big(A\, I_i(\mathbf{x})\big) = \det\big(A_i(\mathbf{b})\big).$$

Now we compute $\det\big(I_i(\mathbf{x})\big)$ by expanding along the ith column:

$$\det(I_i(\mathbf{x})) = \sum_{k=1}^{n} x_k\, C_{k,i} = x_i,$$

because the cofactors of the identity satisfy $C_{k,i} = 0$ for $k \neq i$ and $C_{i,i} = 1$.

Hence

$$\det(A)\, x_i = \det\big(A_i(\mathbf{b})\big) \quad \text{and} \quad x_i = \frac{\det\big(A_i(\mathbf{b})\big)}{\det(A)} \quad \text{for} \quad i = 1, \dots, n.$$

□

Example 3.3.1

Use Cramer's rule to solve the following system.

$$\begin{cases} 3x_1 - 2x_2 = 6 \\ -5x_1 + 4x_2 = 8. \end{cases}$$

Solution

We write this system as $A\mathbf{x} = \mathbf{b}$ with

$$A = \begin{bmatrix} 3 & -2 \\ -5 & 4 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 6 \\ 8 \end{bmatrix}.$$

First compute $\det(A)$:

$$\det(A) = 3 \cdot 4 - (-2)(-5) = 12 - 10 = 2.$$

Since $\det(A) \neq 0$, the system has a unique solution.

Now form the matrices $A_1(\mathbf{b})$ and $A_2(\mathbf{b})$ by replacing the first and second columns with $\mathbf{b}$:

$$A_1(\mathbf{b}) = \begin{bmatrix} 6 & -2 \\ 8 & 4 \end{bmatrix}, \quad A_2(\mathbf{b}) = \begin{bmatrix} 3 & 6 \\ -5 & 8 \end{bmatrix}$$

Compute their determinants:

$$\det(A_1(\mathbf{b})) = 6 \cdot 4 - (-2)(8) = 24 + 16 = 40, \quad \det(A_2(\mathbf{b})) = 3 \cdot 8 - (6)(-5) = 24 + 30 = 54.$$

By Cramer's rule,

$$x_1 = \frac{\det(A_1(\mathbf{b}))}{\det(A)} = \frac{40}{2} = 20, \quad x_2 = \frac{\det(A_2(\mathbf{b}))}{\det(A)} = \frac{54}{2} = 27$$

In engineering, systems of linear differential equations are often transformed into systems of linear equations using Laplace transforms. The parameter s naturally arises in this process, and the

condition $\det(A(s)) = 0$ corresponds to values of s (called *poles*) where the system fails to have a solution. In the future, I will cover Laplace transforms and the rest of differential equations, but for now, here's an example that illustrates this.

Example 3.3.2

Consider the system of which s is a parameter. Determine the values of s for which the system has a unique solution and then use Cramer's rule to describe the solution.

$$\begin{cases} 3s\,x_1 - 2x_2 = 4 \\ -6x_1 + s\,x_2 = 1 \end{cases}$$

Solution

We write this system as a matrix equation:

$$\begin{bmatrix} 3s & -2 \\ -6 & s \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

Then,

$$A = \begin{bmatrix} 3s & -2 \\ -6 & s \end{bmatrix}, \quad A_1(\mathbf{b}) = \begin{bmatrix} 4 & -2 \\ 1 & s \end{bmatrix}, \quad A_2(\mathbf{b}) = \begin{bmatrix} 3s & 4 \\ -6 & 1 \end{bmatrix}$$

We compute $\det(A) = 3s^2 - 12 = 3(s+2)(s-2)$ and thus the system has a unique solution when $s \neq \pm 2$. For such an s, the solution is (x_1, x_2) where

$$x_1 = \frac{\det A_1(\mathbf{b})}{\det A} = \frac{4s+2}{3(s+2)(s-2)}, \quad x_2 = \frac{\det A_2(\mathbf{b})}{\det A} = \frac{3s+24}{3(s+2)(s-2)} = \frac{s+8}{(s+2)(s-2)}.$$

Cramer's rule can also be used to the general formula for the inverse of a matrix $A \in \mathbb{R}^{n\times n}$. The jth column of A^{-1} is a solution to

$$A\mathbf{x} = \mathbf{e}_j$$

where $\mathbf{e}_j$ is the jth column of the identity matrix and the ith entry of $\mathbf{x}$ is the (i,j)-entry of A^{-1}. Then, by Cramer's rule, we will write the (i,j)-entry as

$$x_i = \frac{\det(A_i(\mathbf{e}_j))}{\det(A)}.$$

A cofactor expansion down the ith column of $A_i(\mathbf{e}_j)$ shows that

$$\det(A_i(\mathbf{e}_j)) = (-1)^{i+j} \det(A_{j,i}) = C_{j,i}$$

where $C_{j,i}$ is a cofactor of A. Notice how the (i,j)-th entry of A^{-1} is $\frac{C_{j,i}}{\det(A)}$. Thus our formula is

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} C_{1,1} & C_{2,1} & \cdots & C_{n,1} \\ C_{1,2} & C_{2,2} & \cdots & C_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1,n} & C_{2,n} & \cdots & C_{n,n} \end{bmatrix}. \tag{22}$$

In other words, each entry of A^{-1} can be written in terms of the determinants of submatrices (cofactors) of A. This matrix of cofactors in (22) is called the **adjugate** of A, written as $\text{adj}(A)$.

Adjugate Formula for the Inverse of a Matrix

$$A^{-1} = \frac{1}{\det(A)} \text{adj}(A)$$

Example 3.3.3

Find the inverse of $A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & -1 & 1 \\ 1 & 4 & -2 \end{bmatrix}$.

Solution

We begin by computing the nine cofactors. The (i,j)-cofactor $C_{i,j}$ is $(-1)^{i+j}$ times the determinant of the submatrix obtained by deleting row i and column j:

$$C_{1,1} = \det \begin{bmatrix} -1 & 1 \\ 4 & -2 \end{bmatrix} = (-1)(-2) - (1)(4) = -2$$

$$C_{1,2} = -\det \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} = -((1)(-2) - (1)(1)) = 3$$

$$C_{1,3} = \det \begin{bmatrix} 1 & -1 \\ 1 & 4 \end{bmatrix} = (1)(4) - (-1)(1) = 5$$

$$C_{2,1} = -\det \begin{bmatrix} 1 & 3 \\ 4 & -2 \end{bmatrix} = 14, \qquad C_{2,2} = +\det \begin{bmatrix} 2 & 3 \\ 1 & -2 \end{bmatrix} = -7, \qquad C_{2,3} = -\det \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} = -7,$$

$$C_{3,1} = +\det \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix} = 4, \qquad C_{3,2} = -\det \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix} = 1, \qquad C_{3,3} = +\det \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} = -3$$

Carrying out all nine, we get

$$\begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} \\ C_{2,1} & C_{2,2} & C_{2,3} \\ C_{3,1} & C_{3,2} & C_{3,3} \end{bmatrix} = \begin{bmatrix} -2 & 3 & 5 \\ 14 & -7 & -7 \\ 4 & 1 & -3 \end{bmatrix}.$$

The adjugate is the transpose of the cofactor matrix:

$$\operatorname{adj}(A) = \begin{bmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{bmatrix}$$

Instead of computing $\det(A)$, we can check that $(\operatorname{adj}(A))A = (\det A)I$:

$$(\operatorname{adj}(A))A = \begin{bmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{bmatrix} \begin{bmatrix} 2 & 1 & 3 \\ 1 & -1 & 1 \\ 1 & 4 & -2 \end{bmatrix} = \begin{bmatrix} 14 & 0 & 0 \\ 0 & 14 & 0 \\ 0 & 0 & 14 \end{bmatrix} = 14I$$

so $\det(A) = 14$. Finally,

$$A^{-1} = \frac{1}{\det(A)}\text{adj}(A) = \frac{1}{14}\begin{bmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{7} & 1 & \frac{2}{7} \\ \frac{3}{14} & -\frac{1}{2} & \frac{1}{14} \\ \frac{5}{14} & -\frac{1}{2} & -\frac{3}{14} \end{bmatrix}.$$

This formula for A^{-1} lets us determine properties of the inverse without calculating it, but keep in mind that this is mostly used in theoretical work because computing determinants is very inefficient. Cramer's rule is also generally theoretical and can be used to analyze how sensitive the solution of $A\mathbf{x} = \mathbf{b}$ is to changes in an entry in $\mathbf{b}$ or in A. But once again, Cramer's rule is awful for computing large matrices. Computing one determinant takes nearly as much time as just solving the matrix equation by row reduction.

Determinants also have a geometric representation.

Determinants as Area or Volume

If $A \in \mathbb{R}^{2\times 2}$, the area of the parallelogram spanned by its columns is given by $|\det(A)|$.

If $A \in \mathbb{R}^{3\times 3}$, the area of the parallelepiped spanned by its columns is given by $|\det(A)|$.

For

$$\left|\det\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}\right| = |ad| = \left\{\begin{array}{c} \text{area of} \\ \text{rectangle} \end{array}\right\}$$

This works because if you take two nonzero vectors $\mathbf{a}_1$ and $\mathbf{a}_2$, they form a parallelogram. Furthermore, the area of this parallelogram doesn't change if you replace $\mathbf{a}_2$ with $\mathbf{a}_2 + c\mathbf{a}_1$ for any scalar c. In other words, translating a parallelogram does not change its area.

A parallelepiped is really just a 3D box spanned by three vectors. Its volume is given by the area of the base, formed by two of the three vectors, multiplied by the height of the third vector. Volume here is also preserved under column replacement operations since swapping columns only changes orientation.

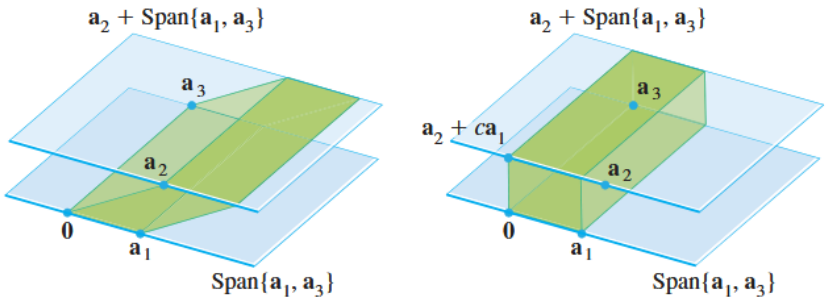


Image credit: Lay et al

Example 3.3.4

Calculate the area of the parallelogram determined by the points $(-2,-2)$, $(0,3)$, $(4,-1)$, and $(6,4)$.

Solution

We first translate the parallelogram so that one vertex is at the origin. Subtracting $(-2,-2)$ from each of the given vertices gives the new vertices:

$$(0,0),\ (2,5),\ (6,1),\ (8,6)$$

This parallelogram is determined by the two vectors

$$\mathbf{a}_1 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 6 \\ 1 \end{bmatrix}.$$

So the corresponding matrix is

$$A = \begin{bmatrix} 2 & 6 \\ 5 & 1 \end{bmatrix}.$$

The area is the absolute value of the determinant:

$$|\det(A)| = |(2)(1) - (6)(5)| = |2 - 30| = 28$$

Perhaps the most important aspect of determinants is how they connect to linear transformations.

Linear Transformations

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation determined by a matrix $A \in \mathbb{R}^{2\times 2}$. If $S \in \mathbb{R}^2$ is a parallelogram, then the area A is given by

$$A_{T(S)} = |\det(A)| \cdot A_S.$$

If T is determined by a matrix $A \in \mathbb{R}^{3\times 3}$ and $S \in \mathbb{R}^2$ is a parallelepiped, then the volume V is given by

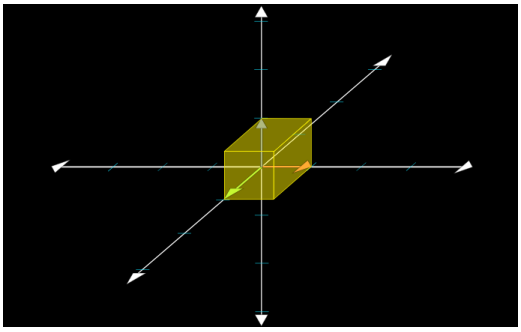
$$V_{T(S)} = |\det(A)| \cdot V_S.$$

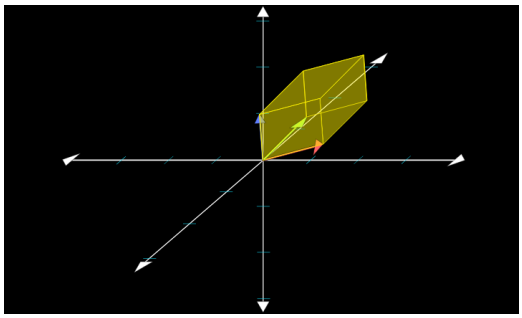
The determinant is actually a measurement of how much a linear transformation scales space.

When we extend this to more general regions in $\mathbb{R}^2$ or $\mathbb{R}^3$, we cannot rely on parallelograms or parallelepipeds exactly. Instead, we can approximate a region by covering it up with squares in 2D or cubes in 3D. This is similar to the way you approximated integrals in calculus.

Now suppose T is a linear transformation with matrix A. The image of each small square under T is a parallelogram where its area gets multiplied by $|\det(A)|$. If we extend this to three dimensions, volume is the same thing but multiplied by volume instead of area. Thus, generally, for a region S with finite area or volume, we can say that its area or volume is given by

$$|\det(A)| \cdot A \quad \text{or} \quad |\det(A)| \cdot V.$$





A cube is transformed into a parallelepiped. The magnitude of the determinant is the factor by which the volume of the cube is scaled and thus the area of the parallelepiped. Image credit: 3Blue1Brown

Problem 3.3.1

Let S be the parallelogram determined by $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$, and let $A = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}$.

Compute the area of $T(S)$ under the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

Solution

The area of S is

$$A_S = \left|\det\begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2\end{bmatrix}\right| = \left|\det\begin{bmatrix}1 & 5\\ 3 & 1\end{bmatrix}\right| = |1\cdot 1 - 5\cdot 3| = |-14| = 14.$$

The scaling factor of T is

$$|\det(A)| = \left|\det\begin{bmatrix}1 & -1\\ 0 & 2\end{bmatrix}\right| = |1\cdot 2 - (-1)\cdot 0| = 2.$$

Therefore, the area of the image of S under the mapping is

$$A_{T(S)} = |\det(A)| \cdot A_S = 2 \cdot 14 = 28.$$

Problem 3.3.2

Now think about why when multiplying two square matrices A, B together, the following holds:

$$\det(AB) = \det(A)\det(B)$$

Solution

Multiplying two matrices together leads to a new linear transformation. If we apply B and then A, B scales space by $\det(B)$ and then A scales space as $\det(A)\det(B)$. If we apply both, we simply get $\det(AB)$. These both describe the same overall transformation.

4 The Four Fundamental Subspaces

In this chapter, we will expand everything we have learned up to this point into a larger and more powerful idea called the **vector space**. The heart of this chapter is the study of **subspaces**, which are smaller vector spaces that live inside larger ones. Every matrix A has four fundamental subspaces: the column space, the row space, the null space, and the left null space.

4.1 Vector Spaces and Subspaces

Vector Space

A **vector space** is a nonempty set V of objects, called vectors, equipped with two operations:

1. **Vector addition**: Adding any pair of vectors $\mathbf{u}, \mathbf{v} \in V$ gives another vector in the set $\mathbf{u} + \mathbf{v} \in V$. This is referred to as being *closed under addition*.
2. **Scalar multiplication**: The product of any scalar $c \in \mathbb{R}$ and any $\mathbf{u} \in V$ gives another vector in the set $c\mathbf{u} \in V$. This is referred to as being *closed under scalar multiplication*.

These operations must satisfy the following properties for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and scalars $c, d \in \mathbb{R}$:

1. **Commutativity of addition**: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
2. **Associativity of addition**: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
3. **Additive identity**: There exists a zero vector $\mathbf{0} \in V$ such that $\mathbf{u} + \mathbf{0} = \mathbf{u} = \mathbf{0} + \mathbf{u}$.
4. **Additive inverse**: For each $\mathbf{u} \in V$, there exists a $-\mathbf{u} \in V$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.
5. **Distributivity over vector addition**: $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
6. **Distributivity over scalar addition**: $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$
7. **Associativity of scalar multiplication**: $c(d\mathbf{u}) = (cd)\mathbf{u}$
8. **Multiplicative identity of scalars**: $1\mathbf{u} = \mathbf{u}$

The zero vector and additive inverses are unique. From the properties it also follows that for all $\mathbf{u} \in V$ and scalar $c \in \mathbb{R}$:

$$0\mathbf{u} = \mathbf{0}$$

$$c\mathbf{0} = \mathbf{0}$$

$$-\mathbf{u} = (-1)\mathbf{u}$$

We begin with the most familiar example of a real vector space. This is $\mathbb{R}^n$, the set of all n-dimensional column vectors with real entries. For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and scalar $c \in \mathbb{R}$, we have

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}, \quad c\mathbf{x} = \begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{bmatrix}.$$

The zero vector $\mathbf{0} \in \mathbb{R}^n$ is the vector with all entries equal to zero. This space satisfies all the algebraic properties of a vector space. Notice that we do not consider extra operations (like dot products or cross products) when defining the vector space itself. The only required operations are vector addition and scalar-vector multiplication.

For another example, consider the following. The set of all $m \times n$ real matrices $\mathbb{R}^{m \times n}$ also forms a vector space under matrix addition and scalar multiplication. The zero vector in this space is the $m \times n$ zero matrix:

$$0 = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$

Recall that $\mathbb{R}^{m \times 1}$ is just a special case of a vector space of matrices with one column. In this case, we once again only needed matrix addition and scalar-matrix multiplication.

So far, most of our problems have been dependent on straightforward results of arithmetic. Now, we can use the algebra of vector spaces $\mathbb{R}^n$ and $\mathbb{R}^{m \times n}$ to solve, compute, and understand these problems. Outside of linear algebra, there are many problems that depend on vector spaces of functions. The difference between $\mathbb{R}^{m \times n}$ and a vector space of functions V is analogous to the difference between arithmetic and calculus.

Vector spaces exist in many different forms. Let's consider a few:

- All infinite sequences $\{y_k\} = (\dots, y_{-2}, y_{-1}, y_0, y_1, y_2, \dots)$ of real numbers, with addition and scalar multiplication defined term by term. This is important in modeling discrete-time signals.
- The set P_n of all polynomials of degree $\leq n$, closed under addition and scalar multiplication. This is useful in approximation and numerical methods.
- The set of all real-valued functions on a domain D, with operations defined pointwise. That is, $(\mathbf{f}+\mathbf{g})(t)\mathbf{f}(t)+\mathbf{g}(t)$ and $(c\mathbf{f})(t) = c\mathbf{f}(t)$. This appears all the time in physics and engineering.

The objects are different, but the properties they satisfy remain the same. Let's look at the polynomial example. We define

$$P_n(I) = \{f : I \to \mathbb{R} \mid n \in \mathbb{N}\},$$

where f is a polynomial of degree less than or equal to n, denoted as $\deg(f) \leq n$, and $I \subseteq \mathbb{R}$. This space forms a vector space under polynomial addition and scalar multiplication. Let $f, g \in P_n(I)$. Any $f \in P_n(I)$ can be written as

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

with coefficients $a_i \in \mathbb{R}$ for $i = 0, 1, \ldots, n$.

Similarly, any $g \in P_n(I)$ can be written as

$$g(x) = b_n x^n + b_{n-1} x^{n-1} + \cdots + b_1 x + b_0.$$

We begin by defining polynomial addition as $(f + g)(x) = f(x) + g(x)$. Then

$$(f + g)(x) = (a_n + b_n)x^n + (a_{n-1} + b_{n-1})x^{n-1} + \cdots + (a_1 + b_1)x + (a_0 + b_0).$$

Since the degree of $f + g$ is still less than or equal to n, $P_n(I)$ is closed under addition.

Let's now look at scalar multiplication. For $c \in \mathbb{R}$,

$$(cf)(x) = c \cdot f(x) = (ca_n)x^n + (ca_{n-1})x^{n-1} + \cdots + (ca_1)x + ca_0.$$

This is also a polynomial of degree less than or equal to n, so $P_n(I)$ is closed under scalar multiplication.

Therefore, $P_n(I)$ satisfies all vector space axioms and is a vector space.

However, notice that the set of polynomials of exactly degree n is *not* a vector space. For example, consider

$$f(x) = x^2 + 3x + 4, \quad g(x) = 2 - x^2.$$

Both f and g are degree 2, but

$$f(x) + g(x) = (x^2 + 3x + 4) + (2 - x^2) = 3x + 6,$$

which has degree $1 \neq 2$. Two polynomials of degree 2 combined to form a polynomial of degree 1. That means the sum of two elements in the set does not necessarily stay in the set. Hence, the set is not closed under addition.

Be careful with notation. The zero polynomial $f(x) = 0$ is a function whose value is 0 for all $x \in I$. The scalar 0 is just the real number zero. These are different objects, even though both are written as 0.

In this setting, we can treat each polynomial in $P_n(I)$ as a vector. Going forward, vectors are simply elements of vector spaces. Thus, the polynomials in the vector space $P_n(I)$ can be considered vectors in this more abstract context.

A **subspace** is a smaller vector space living inside a larger one. Formally, we have the following:

Subspace of a Vector Space

A nonempty subset $W \subseteq V$ of a vector space V is called a **subspace** if the following three properties hold:

1. **Zero vector**: $\mathbf{0} \in W$.
2. **Closed under addition**: For all $\mathbf{u}, \mathbf{v} \in W$, the sum $\mathbf{u} + \mathbf{v} \in W$.
3. **Closed under scalar multiplication**: For all $\mathbf{u} \in W$ and scalars $c \in \mathbb{R}$, the vector $c\mathbf{u} \in W$.

Then, W automatically inherits the other vector space axioms from V, and is a vector space itself.

Let's take a look at some examples.

Example 4.1.1

Is the set consisting only of the zero vector $\{\mathbf{0}\}$ a subspace of a vector space V?

Solution

$\{\mathbf{0}\}$ is nonempty and contains $\mathbf{0}$. It is closed under addition:

$$\mathbf{0} + \mathbf{0} = \mathbf{0} \in \{\mathbf{0}\}$$

And it is closed under scalar multiplication. For any $c \in \mathbb{R}$,

$$c\mathbf{0} = \mathbf{0} \in \{\mathbf{0}\}.$$

Hence $\{\mathbf{0}\}$ is a subspace of V.

Problem 4.1.1

Let P be the set of all real polynomials as functions on $\mathbb{R}$.

(a) Prove P is a subspace of the vector space of all real-valued functions on $\mathbb{R}$.

(b) For $n \geq 0$, let $P_n = \{\, f : \deg(f) \leq n \,\}$. Prove P_n is a subspace of P.

(c) Show that the set of polynomials of exactly degree n is not a subspace of P.

Solution

(a) The zero function $f(x) \equiv 0$ is a polynomial, so $0 \in P$. If $f, g \in P$, then $(f+g)(x) = f(x)+g(x)$ is a polynomial and thus $f+g \in P$. If $c \in \mathbb{R}$ and $f \in P$, then $(cf)(x) = c\,f(x)$ is a polynomial and thus $cf \in P$. Hence P is a subspace of the function space.

(b) If $f, g \in P_n$, then $\deg(f+g) \leq n$, so $f+g \in P_n$. For $c \in \mathbb{R}$, $\deg(cf) \leq \deg(f) \leq n$, so $cf \in P_n$. Also $0 \in P_n$. Hence P_n is a subspace of P.

(c) Let $S = \{\, f : \deg(f) = n \,\}$. Take, for $n = 2$,

$$f(x) = x^2 + 3x + 4, \quad g(x) = 2 - x^2.$$

Both have degree 2, but

$$f(x) + g(x) = (x^2 + 3x + 4) + (2 - x^2) = 3x + 6,$$

which has degree $1 \neq 2$. Thus $f, g \in S$ but $f+g \notin S$, so S is not closed under addition and hence not a subspace.

It is not closed under scalar multiplication either. For any $n \geq 0$, if $f \in S$ and $c = 0$, then $cf = 0$ has no degree n.

Example 4.1.2

(a) Is $\mathbb{R}^2$ a subspace of $\mathbb{R}^3$?

(b) Let $H = \left\{ \begin{bmatrix} s \\ t \\ 0 \end{bmatrix} : s, t \in \mathbb{R} \right\} \subset \mathbb{R}^3$. Show that H is a subspace of $\mathbb{R}^3$.

Solution

(a) No. A subspace must be a subset of the ambient space. Elements of $\mathbb{R}^2$ are vectors with two components, not three components, so $\mathbb{R}^2 \not\subset \mathbb{R}^3$. Hence it cannot be a subspace of $\mathbb{R}^3$.

(b) Let $s = t = 0$. H is nonempty and contains $\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. If $\mathbf{u} = \begin{bmatrix} s_1 \\ t_1 \\ 0 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} s_2 \\ t_2 \\ 0 \end{bmatrix}$ are in H,

then

$$\mathbf{u}+\mathbf{v}=\begin{bmatrix} s_1+s_2 \\ t_1+t_2 \\ 0 \end{bmatrix} \in H.$$

For any $c \in \mathbb{R}$,

$$c\,\mathbf{u}=\begin{bmatrix} cs_1 \\ ct_1 \\ 0 \end{bmatrix} \in H.$$

Thus H is closed under addition and scalar multiplication. The third entries in the vector are also always 0. Hence H is a subspace of $\mathbb{R}^3$.

Example 4.1.3

(a) Is a plane in $\mathbb{R}^3$ that does not pass through the origin a subspace of $\mathbb{R}^3$?

(b) Similarly, is a line in $\mathbb{R}^2$ that does not go through the origin a subspace of $\mathbb{R}^2$?

Solution

(a) No, a plane in $\mathbb{R}^3$ not through the origin is not a subspace of $\mathbb{R}^3$ because a subspace must contain the zero vector.

(b) Similarly, if a plane in $\mathbb{R}^3$ or a line in $\mathbb{R}^2$ does not pass through the origin, then $\mathbf{0}$ is not in the set either. Therefore, these sets cannot be subspaces.

Any subspace *must* pass through the origin. This is because subspaces are closed under scalar multiplication, so they must stretch all the way back to the origin.

One of the most common ways to describe a subspace is through the idea of a linear combination. Let's do a bit of review. Recall that a linear combination means adding together scalar multiples of given vectors. If we are given vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$, then the set of all possible linear combinations of them is called the span:

$$\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\} = \{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p : c_i \in \mathbb{R} \text{ for all } i \in \{1, 2, \ldots, p\}\}$$

This span represents all vectors that can be built from $\mathbf{v}_1, \ldots, \mathbf{v}_p$ by scaling and adding them together. Every span is automatically a subspace and many important subspaces are described this way.

If $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are vectors in a vector space V, then

$$\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\} = \{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p : c_i \in \mathbb{R} \text{ for all } i \in \{1, 2, \ldots, p\}\}.$$

is a subspace of V.

We refer to $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ as the subspace spanned by $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$.

Example 4.1.4

Given $\mathbf{v}_1$ and $\mathbf{v}_2$ in a vector space V, let $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$. Show that H is a subspace of V.

Solution

In set-builder notation, we have

$$H = \{\, s\mathbf{v}_1 + t\mathbf{v}_2 : s, t \in \mathbb{R} \,\}.$$

The zero vector is in H since

$$\mathbf{0} = 0\mathbf{v}_1 + 0\mathbf{v}_2.$$

If $\mathbf{u} = s_1\mathbf{v}_1 + t_1\mathbf{v}_2$ and $\mathbf{w} = s_2\mathbf{v}_1 + t_2\mathbf{v}_2$, then

$$\mathbf{u} + \mathbf{w} = (s_1\mathbf{v}_1 + s_2\mathbf{v}_2) + (t_1\mathbf{v}_1 + t_2\mathbf{v}_2) = (s_1 + t_1)\mathbf{v}_1 + (s_2 + t_2)\mathbf{v}_2.$$

Thus it is closed under addition.

If scalar $c \in \mathbb{R}$, then

$$c\mathbf{u} = c(s_1\mathbf{v}_1 + s_2\mathbf{v}_2) = (cs_1)\mathbf{v}_1 + (cs_2)\mathbf{v}_2.$$

So H is closed under addition and scalar multiplication, and therefore H is a subspace of V.

Example 4.1.5

Let $H = \{(a - 3b,\ b - a,\ a,\ b) : a, b \in \mathbb{R}\} \subseteq \mathbb{R}^4$. Show that H is a subspace of $\mathbb{R}^4$.

Solution

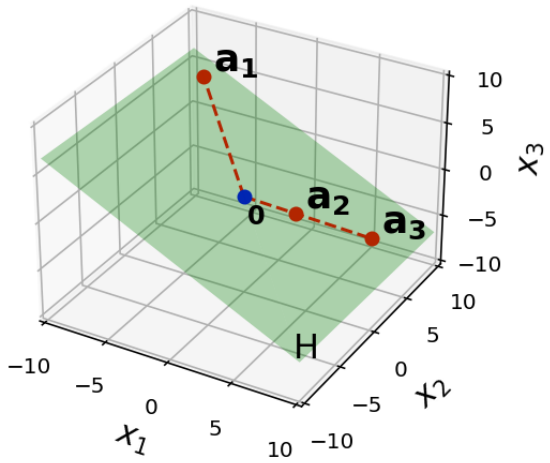
Write the general elements (vectors) of H in column form and then factor:

$$\begin{bmatrix} a - 3b \\ b - a \\ a \\ b \end{bmatrix} = a \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} -3 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

Thus

$$H = \text{Span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Sine the set can be expressed as a span, H is a subspace of $\mathbb{R}^4$.



$H = \text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$. Image credit: Crovella, Boston University

In $\mathbb{R}^3$, there are four possible subspaces:

1. Any line through $(0, 0, 0)$
2. Any plane through $(0, 0, 0)$
3. The whole space
4. The single vector $(0, 0, 0)$

Example 4.1.6

Determine the value(s) of h where $\mathbf{y}$ in the subspace of $\mathbb{R}^3$ spanned by $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, if

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 5 \\ -4 \\ -7 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} -4 \\ 3 \\ h \end{bmatrix}$$

Solution

$\mathbf{y} \in \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ if and only if the system

$$\begin{bmatrix} 1 & 5 & -3 \\ -1 & -4 & 1 \\ -2 & -7 & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} -4 \\ 3 \\ h \end{bmatrix}$$

is consistent. We will row reduce the augmented matrix to get a value for h.

$$\begin{bmatrix} 1 & 5 & -3 & -4 \\ -1 & -4 & 1 & 3 \\ -2 & -7 & 0 & h \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & -3 & -4 \\ 0 & 1 & -2 & -1 \\ 0 & 3 & -6 & h-8 \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & -3 & -4 \\ 0 & 1 & -2 & -1 \\ 0 & 0 & 0 & h-5 \end{bmatrix}$$

Consistency requires the last row to be $[0\ 0\ 0\ 0]$, so $h - 5 = 0 \Rightarrow h = 5$.

Problem 4.1.2

An $n \times n$ matrix A is symmetric if $A^T = A$. Let S be the set of all 3×3 symmetric matrices. Show that S is a subspace of $M_{3\times 3}$, the vector space of all 3×3 real matrices.

Solution

First, the zero matrix 0 satisfies $0^T = 0$, so $0 \in S$.

Second, if $A, B \in S$, then $A^T = A$ and $B^T = B$. Hence

$$(A + B)^T = A^T + B^T = A + B,$$

so $A + B \in S$. It is closed under addition.

If $A \in S$ and $c \in \mathbb{R}$, then

$$(cA)^T = cA^T = cA,$$

so $cA \in S$. It is closed under scalar multiplication.

Since the three axioms are met, S is a subspace of $M_{3\times 3}$.

4.2 Null Spaces and Column Spaces

To begin, let's go over when the subspaces of $\mathbb{R}^n$ arise. The first is in the set of all solutions to a system of homogeneous linear equations and the second is as the set of all linear combinations of certain specified vectors.

Null Space

The **null space** of A, written $\text{Nul}(A)$, is the set of all solutions to the homogeneous equation

$$A\mathbf{x} = \mathbf{0}.$$

That is,

$$\text{Nul}(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}.$$

$\text{Nul}(A)$ is a subspace of $\mathbb{R}^n$ that describes all the vectors that map A to $\mathbf{0}$. It is the solution set to the homogeneous linear systems problem

$$A\mathbf{x} = \mathbf{0}.$$

In other words, $\text{Nul}(A)$ is the set of all $\mathbf{x} \in \mathbb{R}^n$ that are mapped into the zero vector of $\mathbb{R}^m$ through a linear transformation $\mathbf{x} \mapsto A\mathbf{x}$. See Figure 5.

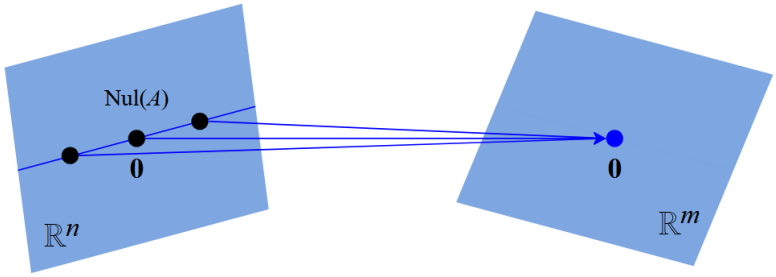


Figure 6

Example 4.2.1

Find the null space of $A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$.

Solution

We begin by getting A in RREF:

$$A \sim U = \begin{bmatrix} 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We have five columns and two pivot columns. The non-pivot columns are columns 2, 4, and 5. Now we use linear combinations to produce zero output vectors.

$$\text{Nul}(A) = \{\mathbf{x} \in \mathbb{R}^5 : A\mathbf{x} = \mathbf{0}\}$$

We see that we have

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

Notice how we have an equation relating column 1 of A to column 2 of A. We can use this to create the zero vector.

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

This shows that columns 1 and 2 of A are linearly dependent. We can write the zero vector as a linear combination of all five columns of A:

$$2 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} -2 \\ 0 \\ 0 \end{bmatrix} + 0 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 0 \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} + 0 \begin{bmatrix} 3 \\ -2 \\ 0 \end{bmatrix} = \mathbf{0}$$

In matrix-vector form, this is expressed as

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

Thus the first vector for the null space, which corresponds with the first non-pivot column, is

$$\mathbf{u}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The process continues. First, we take the non-pivot column, express it in terms of the pivot columns, and flip the coefficients into a vector with a 1 in the slot of the non-pivot column.

Let's now consider column 4 of A. We have

$$1\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + (-2)\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 1\begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} = \mathbf{0}.$$

In matrix-vector form, this is expressed as

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

Thus the second vector for the null space is

$$\mathbf{u}_2 = \begin{bmatrix} 0 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}.$$

Let's now consider column 5 of A. We can express it as a linear combination of the pivot columns:

$$3\begin{bmatrix}1\\0\\0\end{bmatrix}+(-2)\begin{bmatrix}0\\1\\0\end{bmatrix}=\begin{bmatrix}3\\-2\\0\end{bmatrix}\Rightarrow\begin{bmatrix}3\\-2\\0\end{bmatrix}-3\begin{bmatrix}1\\0\\0\end{bmatrix}+2\begin{bmatrix}0\\1\\0\end{bmatrix}=\mathbf{0}$$

This is the same as saying

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

These coefficients form the vector $\mathbf{u}_3$ that satisfies this relationship:

$$U\begin{bmatrix}-3\\0\\2\\0\\1\end{bmatrix}=\mathbf{0}$$

This can be written as a linear combination:

$$\mathbf{x}=x_2\begin{bmatrix}2\\1\\0\\0\\0\end{bmatrix}+x_4\begin{bmatrix}1\\0\\-2\\1\\0\end{bmatrix}+x_5\begin{bmatrix}-3\\0\\2\\0\\1\end{bmatrix}.$$

Therefore, a basis for the null space is

$$\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} \right\},$$

and the null space itself is

$$\text{Nul}(A) = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \text{Span}\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

We can use the same steps to determine traits regarding the linear independence of a matrix.

Example 4.2.2

The system of homogeneous equations

$$\begin{cases} x_1 - 3x_2 - 2x_3 = 0 \\ -5x_1 + 9x_2 + x_3 = 0 \end{cases}$$

can be written in matrix form as $A\mathbf{x} = \mathbf{0}$, where

$$A = \begin{bmatrix} 1 & -3 & -2 \\ -5 & 9 & 1 \end{bmatrix}.$$

Let $\mathbf{u} = \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix}$. Determine if $\mathbf{u}$ belongs to the null space of A.

Solution

We are trying to test if $\mathbf{u}$ satisfies $A\mathbf{u} = \mathbf{0}$:

$$A\mathbf{u} = \begin{bmatrix} 1 & -3 & -2 \\ -5 & 9 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix} = \begin{bmatrix} (1)(5) + (-3)(3) + (-2)(-2) \\ (-5)(5) + (9)(3) + (1)(-2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Thus $\mathbf{u} \in \text{Nul}(A)$.

Example 4.2.3

Find the linear dependence relations between the columns of the matrix

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$$

.

Solution

The process is similar to before. We need A in RREF:

$$A \sim U = \begin{bmatrix} 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We can see that columns 2, 4, and 5 are non-pivot columns. Thus columns 1 and 3 are pivot columns. The pivot columns of U are used to determine where the linearly independent columns of A are:

$$\left\{ \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix} \right\}$$

Conversely, the non-pivot columns of U identify the linearly dependent columns of A. In U, column 2 is equal to column 1 of multiplied by -2. Then, in A, we know that column 2 of is equal to column 1 of multiplied by -2. This is the first linear dependence relation.

Now we move on to the second linear dependence equation. In U, column 4 is equal to column 1 multiplied by -1 added to column 3 multiplied by 2. That is, column 4 is a linear combination of them. In A, column 4 is equal to column 1 multiplied by -1 added to column 3 multiplied by 2.

Notice how the same linear dependence relationships that exist in RREF also exist in the original matrix.

Problem 4.2.1

Prove that if $n, m \in \mathbb{N}$ and $A \in \mathbb{R}^{m \times n}$, then $\text{Nul}(A)$ is a subspace of $\mathbb{R}^n$.

Solution

To do this, we need to prove the three axioms.

(i) First, we need to prove that $\mathbf{0} \in \text{Nul}(A)$. If we consider the fact that $A_{m\times n} \cdot \mathbf{0}_{n\times 1} = \mathbf{0}_{m\times 1}$, we can say by definition that $\mathbf{0}_{m\times 1} \in \text{Nul}(A)$.

(ii) Second, we need to prove that the null space is closed under addition. That is, if $\mathbf{x}, \mathbf{y} \in \text{Nul}(A)$, then $\mathbf{x} + \mathbf{y} \in \text{Nul}(A)$. If we let $\mathbf{x}, \mathbf{y} \in \text{Nul}(A)$, then $A\mathbf{x} = \mathbf{0}$ because vectors in the null space of A multiplied by A produce $\mathbf{0}$. Additionally, $A\mathbf{y} = \mathbf{0}$. Then consider $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}$. Since $\mathbf{x}$ and $\mathbf{y}$ are in the null space, we know that $A\mathbf{x} = \mathbf{0}$ and $A\mathbf{y} = \mathbf{0}$. Thus $A(\mathbf{x} + \mathbf{y}) = \mathbf{0}$.

(iii) Third, we need to prove that the null space is closed under scalar multiplication. That is, if any scalar $c \in \mathbb{R}$ and any vector $\mathbf{x} \in \text{Nul}(A)$, then $c\mathbf{x} \in \text{Nul}(A)$. If we let $\mathbf{x} \in \text{Nul}(A)$, then $A\mathbf{x} = \mathbf{0}$. Furthermore, if we let $c \in \mathbb{R}$ and consider $c\mathbf{x}$, then $A(c\mathbf{x}) = cA\mathbf{x} = c\mathbf{0} = \mathbf{0}$. This implies that $c\mathbf{x} \in \text{Nul}(A)$.

Thus, $\text{Nul}(A) \subseteq \mathbb{R}^n$ is a subspace of $\mathbb{R}^n$.

Keep in mind that the linear equations defining a set are homogeneous. If not, the set of solutions will definitely not be a subspace because the zero vector is not a solution of a homogeneous solution. Furthermore, in some cases, the set of solutions could be empty.

$\text{Nul}(A)$ is defined implicitly because it is defined by a condition that must be checked; no explicit list of the entries in the null space are given. However, solving the linear systems problem means we are producing an explicit description of $\text{Nul}(A)$.

Problem 4.2.2

Find a spanning set for the null space of the matrix

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & 7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$$

Solution

We need to find the general solution of the linear systems problem in terms of free variables. Thus, we row reduce A to REF so we can write the basic variables in terms of the free variables:

$$\sim \begin{bmatrix} 1 & -2 & 0 & -1 & 3 & 0 \\ 0 & 0 & 1 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \iff \begin{cases} x_1 - 2x_2 - x_4 + 3x_5 = 0 \\ x_3 + 2x_4 - 2x_5 = 0 \\ 0 = 0 \end{cases}$$

The general solution is $x_1 = 2x_2 + x_4 - 3x_5$, $x_3 = -2x_4 + 2x_5$, and x_2, x_4, x_5 are free. Now, we decompose the vector to express the general solution as a linear combination of vectors where the weights are the free variables:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}$$
$$= x_2\mathbf{u} + x_4\mathbf{v} + x_5\mathbf{w}$$

Since every linear combination of $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ is an element of $\text{Nul}(A)$, we can conclude that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a spanning set for $\text{Nul}(A)$.

The spanning set from Problem 4.2.2 is actually automatically linearly independent because of the fact that we made the free variables the weights on the spanning vectors. Given the second, fourth, and fifth entries in the solution vector, $x_2\mathbf{u} + x_4\mathbf{v} + x_5\mathbf{w}$ can be $\mathbf{0}$ if and only if the weights are all zero.

Additionally, when $\text{Nul}(A)$ contains nonzero vectors, the number of vectors in the spanning set of $\text{Nul}(A)$ is equal to the number of free variables in $A\mathbf{x} = \mathbf{b}$.

We now move on to the second of the four fundamental subspaces.

Let A be an $m \times n$ matrix with columns $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$. The **column space** of A, written $\text{Col}(A)$, is the set of all linear combinations of the columns of A:

$$\text{Col}(A) = \text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\} = \text{Span}\left\{\begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{bmatrix}, \begin{bmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{m,2} \end{bmatrix}, \ldots, \begin{bmatrix} a_{1,n} \\ a_{2,n} \\ \vdots \\ a_{m,n} \end{bmatrix}\right\}$$

Equivalently,

$$\text{Col}(A) = \{\mathbf{b} \in \mathbb{R}^m : A\mathbf{x} = \mathbf{b} \text{ for some } \mathbf{x} \in \mathbb{R}^n\}.$$

The column space, that is, the set of all vectors that can be created taking linear combinations of the columns of A must be in $\mathbb{R}^m$.

Thus, $\text{Col}(A)$ is the subspace of $\mathbb{R}^m$ consisting of all vectors $\mathbf{b}$ that can be expressed as $A\mathbf{x}$.

At this point, let's step back and look at a paradigm. Let's begin by defining a function $f(\mathbf{x}) = A\mathbf{x}$. We know that

$$f : \mathbb{R}^n \to \mathbb{R}^m$$

We know that $\text{Col}(A) \subseteq \mathbb{R}^m$ where $\mathbb{R}^m$ is the space in which the column space lives.

We also know that the range of f is

$$\{\mathbf{b} \in \mathbb{R}^m : A\mathbf{x} = \mathbf{b} \text{ for some } \mathbf{x} \in \mathbb{R}^n\},$$

which equivalently is the span of the columns of A because each input vector gives you a linear combination.

Thus, the range of the linear transformation $f(\mathbf{x}) = A\mathbf{x}$ is exactly the column space of A:

$$\text{Rng}(f) = \text{Col}(A)$$

This is a powerful idea that will help us determine whether or not linear systems problems have solutions, and later on, if we should use least-squares to solve $A\mathbf{x} = \mathbf{b}$.

Example 4.2.4

Find the column space of the matrix

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}.$$

Solution

From before, we have

$$A \sim U = \begin{bmatrix} 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Since the pivot columns indicate the position of the linearly independent columns of A, we know that columns 1 and 3 are linearly independent. This implies that the column space can be written as the span of these two linearly independent columns:

$$\text{Col}(A) = \text{Span}\left\{ \begin{bmatrix} -3 \\ 1 \\ -2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ -5 \end{bmatrix} \right\}$$

Example 4.2.5

Let

$$W = \left\{ \begin{bmatrix} 6a - b \\ a + b \\ -7a \end{bmatrix} : a, b \in \mathbb{R} \right\}.$$

Find a matrix A such that $W = \text{Col}(A)$

Solution

We write W as a set of linear combinations:

$$W = \left\{ a \begin{bmatrix} 6 \\ 1 \\ -7 \end{bmatrix} + b \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} : a, b \in \mathbb{R} \right\} = \text{Span} \left\{ \begin{bmatrix} 6 \\ 1 \\ -7 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

The vectors of the spanning set then become the columns of A:

$$A = \begin{bmatrix} 6 & -1 \\ 1 & 1 \\ -7 & 0 \end{bmatrix}$$

Now, $W = \text{Col}(A)$.

The Column Space is a Subspace

Let $A \in \mathbb{R}^{m \times n}$ where $m, n \in \mathbb{N}$. Then, $\text{Col}(A)$ is a subspace of $\mathbb{R}^m$.

Proof

If $A \in \mathbb{R}^{m \times n}$, we know that the column space of A is just the span of the columns of A. To show that $\text{Col}(A)$ is a subspace, we need the three axioms to hold true.

(i) We first need to determine that $\mathbf{0} \in \text{Col}(A)$. Since each element of the column space is a linear combination of the columns of A which are $m \times 1$ therefore every element of the column space must be of size m. We can conclude that $\mathbf{0}$ is $m \times 1$.

Let's start with the fact that $\mathbf{0} \in \text{Col}(A)$ if and only if $\mathbf{0} = A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$. This is simply the definition of the column space, the set of all outputs such that there exists at least one input where we can take A multiplied by the input and produce the output. We know that at least one solution to the homogeneous linear system is the zero vector. So for $\mathbf{x} = \mathbf{0} \in \mathbb{R}^n$, it is true that $A\mathbf{x} = A\mathbf{0} = \mathbf{0}_{m \times 1}$. This means that $\mathbf{0} \in \text{Col}(A)$.

(ii) We now need to prove that the column space is closed under vector addition. That is, if we let $b_1, b_2 \in \text{Col}(A)$, we need to show that $\mathbf{b}_1 + \mathbf{b}_2 \in \text{Col}(A)$. By the assumption that $x, y \in \text{Col}(A)$, it follows that that $\mathbf{b}_1$ and $\mathbf{b}_2$ are in the span of the columns of A by definition of the span. This implies that there exists some $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ such that $A\mathbf{x}_1 = \mathbf{b}_1$ and $A\mathbf{x}_2 = \mathbf{b}_2$. For $\mathbf{b}_1$ to be in the span, we must be able to write it as a linear combination of the columns of A. This implies that we should be able to produce scalar weights on each of the columns of A to get $\mathbf{b}_1$. This is what the $\mathbf{x}_1$ term does for us in the matrix-vector multiplication. The same thing applies for $\mathbf{b}_2$ and $\mathbf{x}_2$.

If we consider $\mathbf{b}_1 + \mathbf{b}_2$, we know that for $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$, we can write that $A\mathbf{x} = A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2 = \mathbf{b}_1 + \mathbf{b}_2$. This implies that $\mathbf{b}_1 + \mathbf{b}_2 \in \text{Col}(A)$.

(iii) We now need to prove that the column space is closed under scalar multiplication. That is, if we let $c \in \mathbb{R}$ and $\mathbf{b} \in \text{Col}(A)$, we need to show that $c\mathbf{b} \in \text{Col}(A)$. If we consider the fact that $\mathbf{b} \in \text{Col}(A)$ means that $\mathbf{b} = A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$, we can write that $c\mathbf{b} = cA\mathbf{x} \Rightarrow c\mathbf{b} = A(c\mathbf{x})$. This means that $c\mathbf{x} \in \mathbb{R}^n$ and thus the output $c\mathbf{b}$ has a pre-image. This implies that $c\mathbf{b}$ is in the span of the columns of A which means that $c\mathbf{b} \in \text{Col}(A)$.

Thus, $\text{Col}(A) \subseteq \mathbb{R}^m$ is a subspace of $\mathbb{R}^m$. We've been thinking about $\mathbb{R}^m$ as the codomain of the map defined by A. Hence, $\text{Col}(A)$ is a subspace of its codomain. □

Example 4.2.6

Let $A = \begin{bmatrix} 2 & 4 & -2 & 1 \\ -2 & -5 & 7 & 3 \\ 3 & 7 & -8 & 6 \end{bmatrix}$.

(a) If the column space of A is a subspace of $\mathbb{R}^k$, what is k?

(b) If the null space of A is a subspace of $\mathbb{R}^k$, what is k?

Solution

(a) $\text{Col}(A) \subseteq \mathbb{R}^m$. Since A has $m = 3$ rows, each column of A is a vector in $\mathbb{R}^3$, and every linear combination of the columns is still in $\mathbb{R}^3$. Therefore, $k = 3$.

(b) $\text{Nul}(A) \subseteq \mathbb{R}^n$. Since A has $n = 4$ columns, any vector $\mathbf{x}$ in $A\mathbf{x}$ must lie in $\mathbb{R}^4$. The null space is the set of all such $\mathbf{x}$ with $A\mathbf{x} = \mathbf{0}$. Therefore, $k = 4$.

Next, we have the **row space**. The row space of $A_{m\times n}$ is simply the span of the rows of A. It is given by $\text{Col}(A^T)$ and lives in $\mathbb{R}^n$.

Then, we have the **left null space**, which lives in $\mathbb{R}^m$. It is simply the null space of its transpose. That is, $\text{Nul}(A^T)$ for all vectors $\mathbf{y} \in \mathbb{R}^m$ that satisfy $A^T\mathbf{y} = \mathbf{0}$. Think about why the left null space is named as such. It is because you are multiplying vectors on the left rather than the right.

To summarize, the dimension of the column space is given by the rank r. The dimension of the row space is also r. Then, the dimension of the null space is given by $n - r$ and the dimension of the left null space is given by $m - r$.

4.3 Linear Transformations

We will now step back and look at transformations. This is one of the most important parts of linear algebra, so much so that Gil Strang's linear algebra textbook, the most famous linear algebra textbook of all time, uses the four fundamental subspaces as the cover.

A transformation can be solved using linear algebra if we can prove it is a matrix transformation. Let's begin with a quick review from before.

Linear Transformation

Let V and W be vector spaces. A function

$$T : V \to W$$

is called a **linear transformation** if for all vectors $\mathbf{u}, \mathbf{v} \in V$ and all scalars $c \in \mathbb{R}$, the following properties hold:

(i) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$

(ii) $T(c\mathbf{u}) = cT(\mathbf{u})$

Thus, every matrix transformation is a linear transformation. We now introduce two important properties that hold true for any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$:

(i) $T(\mathbf{0}) = \mathbf{0}$.

(ii) For any vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ and any scalars $c_1, c_2, \ldots, c_k \in \mathbb{R}$,

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_kT(\mathbf{v}_k).$$

In physics and engineering, property (ii) is known as the superposition principle. It states that the output of a linear system in response to a linear combination of inputs is the same linear combination of the corresponding outputs.

Example 4.3.1

Let $T : \mathbb{R} \to \mathbb{R}$ be defined by $T(x) = x + 1$. Is T a linear transformation?

Solution

Any linear transformation must map the zero vector to the zero vector. Thus, since $T(\mathbf{0}) \neq \mathbf{0}$, T cannot be a linear transformation.

Example 4.3.2

Let $T : \mathbb{R} \to \mathbb{R}$ be defined by $T(x) = 1.5x$. Is T a linear transformation?

Solution

We must verify the defining properties of a linear transformation for all vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}$ and scalar $c \in \mathbb{R}$:

$$T(\mathbf{u}+\mathbf{v}) = 1.5(\mathbf{u}+\mathbf{v}) = 1.5\mathbf{u} + 1.5\mathbf{v} = T(\mathbf{u}) + T(\mathbf{v})$$

$$T(c\mathbf{u}) = 1.5(c\mathbf{u}) = c(1.5\mathbf{u}) = cT(\mathbf{u})$$

Since both properties are satisfied, T is a linear transformation.

Now, we step back and work with the fact that subspaces of vector spaces other than $\mathbb{R}^n$ are often described as linear transformations rather than matrices. The kernel, or null space, of a linear transformation $T : V \to W$ is the set of all vectors $\mathbf{u} \in V$ such that $T(\mathbf{u}) = \mathbf{0}$. That is, the kernel consists of all inputs such that T transforms into $\mathbf{0}$. Similarly, the range of T is the set of all vectors in W of the form $T(x)$ for some $\mathbf{x} \in V$. If T is a matrix transformation $T(\mathbf{x}) = A\mathbf{x}$ for some matrix A, then the kernel of T is the null space and the range of T is the column space of A. In applications, subspaces often arise in the form of the kernel or the range of a linear transformation. For example, the set of all solutions to a homogeneous differential equation forms a subspace, which happens to be the kernel of a linear transformation.

Let's now summarize. The column space represents the range of a transformation, or the set of all vectors you can get out of a system. The null space is the kernel of the transformation, or the set of all inputs that collapse into the zero vector. The row space is the span of the rows of the matrix. And lastly, the left null space is the set of all vectors that give linear combinations of the rows of the matrix that yield the zero vector.

Problem 4.3.1

Let A be an $m \times n$ matrix. Recall that an element of $\text{Col}(A)$ has the form $A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$. Let $A\mathbf{x}$ and $A\mathbf{w}$ represent any two vectors in $\text{Col}(A)$.

(a) Explain why the zero vector is in $\text{Col}(A)$.

(b) Show that $A\mathbf{x} + A\mathbf{w}$ is in $\text{Col}(A)$.

(c) Given a scalar $c \in \mathbb{R}$, show that $c(A\mathbf{x})$ is in $\text{Col}(A)$.

Solution

(a) The zero vector in $\mathbb{R}^m$ can be written as

$$\mathbf{0} = A\mathbf{0},$$

where $\mathbf{0} \in \mathbb{R}^n$. Since $\mathbf{0}$ is of the form $A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$, it follows that $\mathbf{0} \in \text{Col}(A)$.

(b) Since $A\mathbf{x}, A\mathbf{w} \in \text{Col}(A)$, there exist vectors $\mathbf{x}, \mathbf{w} \in \mathbb{R}^n$ such that these vectors are their images under A. Using the property of linearity of matrix multiplication,

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

Because $\mathbf{x} + \mathbf{w} \in \mathbb{R}^n$, the vector $A(\mathbf{x} + \mathbf{w})$ lies in $\text{Col}(A)$.

(c) Let $c \in \mathbb{R}$. Then

$$c(A\mathbf{x}) = A(c\mathbf{x}).$$

Since $c\mathbf{x} \in \mathbb{R}^n$, it follows that $A(c\mathbf{x}) \in \text{Col}(A)$, and therefore $c(A\mathbf{x}) \in \text{Col}(A)$.

Problem 4.3.2

Let A be an $m \times n$ matrix. Mark each of the following statements as true or false and justify each answer.

(a) The null space of A is the solution set of the equation $A\mathbf{x} = \mathbf{0}$.

(b) The null space of an A matrix is in $\mathbb{R}^m$.

(c) The column space of A is the range of the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

(d) If the equation $A\mathbf{x} = \mathbf{b}$ is consistent, then $\text{Col}(A) = \mathbb{R}^m$.

(e) The kernel of a linear transformation is a vector space.

(f) $\text{Col}(A)$ is the set of all vectors that can be written as $A\mathbf{x}$ for some $\mathbf{x}$.

Solution

(a) True. By definition, the null space of A is

$$\text{Nul}(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\},$$

which is exactly the solution set of the homogeneous equation $A\mathbf{x} = 0$.

(b) False. The null space consists of input vectors $\mathbf{x}$, so

$$\text{Nul}(A) \subseteq \mathbb{R}^n,$$

not $\mathbb{R}^m$.

(c) True. The column space of A is the set

$$\text{Col}(A) = \{\mathbf{b} \in \mathbb{R}^m : A\mathbf{x} = \mathbf{b} \text{ for some } \mathbf{x} \in \mathbb{R}^n\},$$

which is precisely the range of the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

(d) False. If $A\mathbf{x} = \mathbf{b}$ is consistent for a single vector $\mathbf{b}$, then $\mathbf{b} \in \text{Col}(A)$. However, $\text{Col}(A) = \mathbb{R}^m$ requires that $A\mathbf{x} = \mathbf{b}$ be consistent for *every* $\mathbf{b} \in \mathbb{R}^m$.

(e) True. The kernel of a linear transformation T is defined as

$$\{\mathbf{u} : T(\mathbf{u}) = \mathbf{0}\}.$$

Since T is linear, the kernel contains $\mathbf{0}$ and is closed under addition and scalar multiplication, so it is a vector space.

(f) True. By definition,

$$\text{Col}(A) = \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\},$$

which represents the set of all vectors that can be written as $A\mathbf{x}$ for some $\mathbf{x}$.

4.4 Linearly Independent Sets and the Basis

Up to this point, we have described subspaces by listing the vectors that span them. However, spanning sets are not unique, and the vectors they contain often provide no new information. That is, they can be written as a linear combination of the other vectors in the set and they can be removed without changing the overall span. How can we determine if a set of vectors contains redundancies?

In Figure 6, we can see that it is a set containing one vector $\mathbf{v}$. It is linearly independent because it cannot be expressed as a linear combination of other vectors. The only case in which it would be linearly dependent is if it were the zero vector. In Figure 7, we add on the vector $\mathbf{w}$ to the set. Since it is not in Span$\{\mathbf{v}\}$, it is linearly dependent. Notice how the span gets bigger.

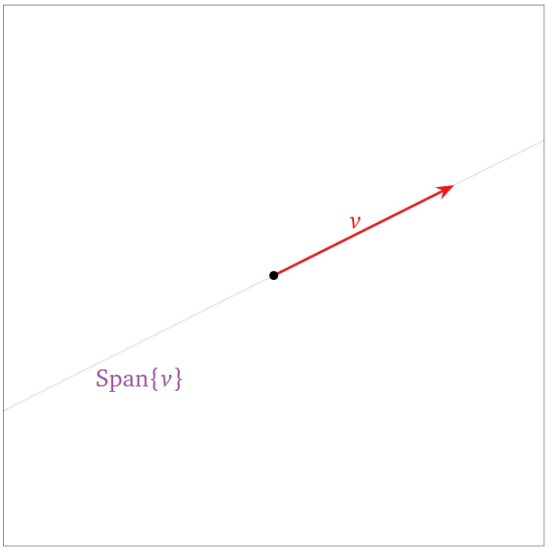


Figure 7: A set containing one vector. Image credit: Margalit and Rabinoff

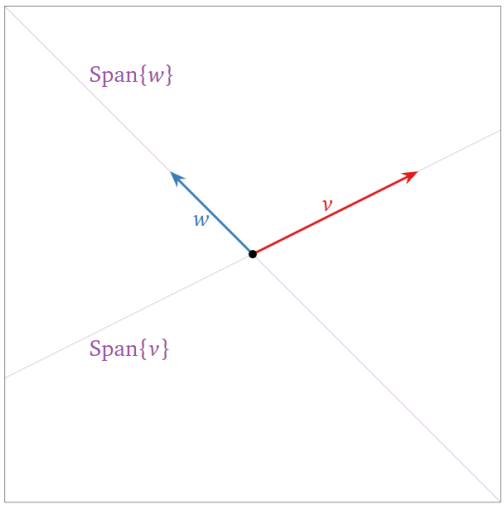


Figure 8: Linearly independent set of vectors. Image credit: Margalit and Rabinoff

Consider Figure 8. The right parallelogram shows that $\mathbf{w} = \mathbf{u} + \mathbf{v}$. Thus, $\mathbf{w} \in \text{Span}\{\mathbf{u}, \mathbf{v}\}$, meaning that $\mathbf{w}$ can already be constructed as a linear combination of $\mathbf{u}$ and $\mathbf{v}$. Adding $\mathbf{w}$ therefore introduces no new direction and does not change the span:

$$\text{Span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\} = \text{Span}\{\mathbf{u}, \mathbf{v}\}.$$

Rearranging the relation also gives $\mathbf{v} = \mathbf{w} - \mathbf{u}$ and $\mathbf{u} = \mathbf{w} - \mathbf{v}$. Therefore, each vector can be written as a linear combination of the other two, so any one of the three vectors may be removed without changing the span.

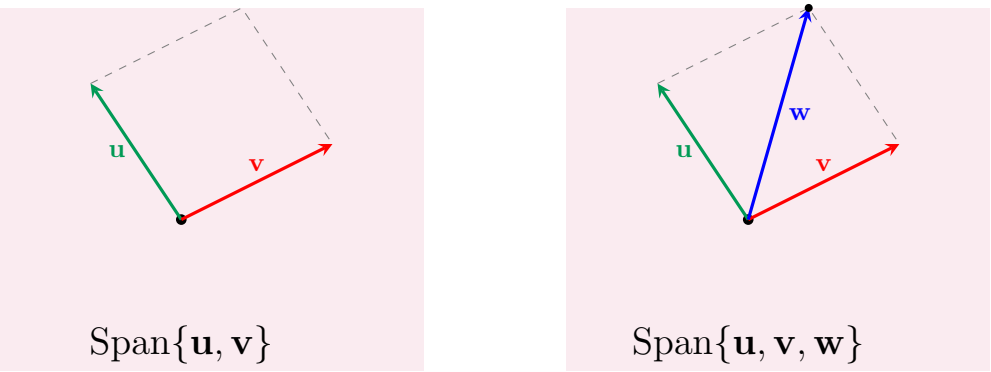


Figure 9: Adding a vector that is already a linear combination of the others does not enlarge the span.

Example 4.4.1

Which of the following vectors makes the set in Figure 9 linearly dependent?

(a)

$$\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}$$

(b)

$$\mathbf{x} = \begin{bmatrix} -2 \\ 4 \\ 0 \end{bmatrix}$$

(c)

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

(d)

$$\mathbf{x} = \begin{bmatrix} 0 \\ -2 \\ 4 \end{bmatrix}$$

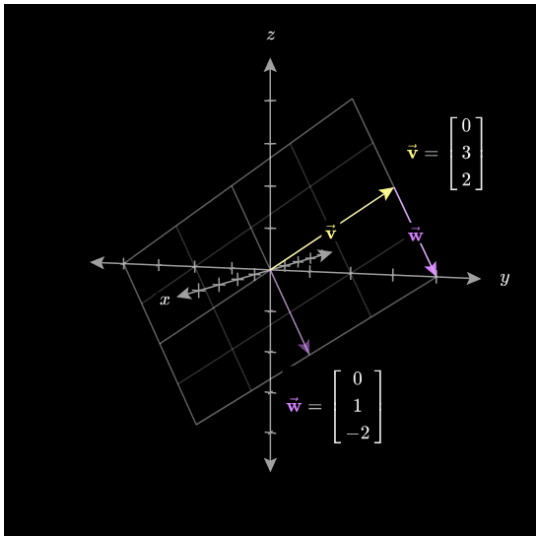


Figure 10: The span of two vectors. Image credit: 3Blue1Brown

Solution

Both of the vectors lie in the yz-plane. The vector must be expressed only using $\mathbf{v}$ and $\mathbf{w}$. The answer is (d).

Now that we've illustrated the central problem, hopefully you have gained some more geometric intuition. A spanning set must contain enough vectors to generate the entire subspace, but it may also contain vectors that are redundant. We therefore want to describe a subspace using enough vectors to span it, but no more than are needed. Let's take a look at a formal definition:

Linear Independence and Linear Dependence

Let V be a vector space, and let

$$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$$

be a set of vectors in V. The set S is **linearly independent** if the vector equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$$

has only the trivial solution $c_1 = c_2 = \cdots = c_p = 0$.

The set S is **linearly dependent** if the equation has a nontrivial solution. That is, S is linearly dependent if there exist some nonzero scalars $c_1, c_2, \ldots, c_p$ such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0}.$$

This equation is called a **linear dependence relation** among the vectors.

Linear independence means that the zero vector can be formed from the vectors only by assigning a coefficient of zero to every vector. Linear dependence means that some nonzero choice of coefficients also produces the zero vector. In that case, at least one vector can be constructed from the others and therefore contributes no new information to the span.

Let's take a look at the null space interpretation of this. Suppose that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p \in \mathbb{R}^n$, and place the vectors into the columns of a matrix:

$$A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_p \end{bmatrix}$$

Then

$$A \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p.$$

Therefore, the equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0},$$

is equivalent to the homogeneous matrix equation

$$A\mathbf{c} = \mathbf{0},$$

where

$$\mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}.$$

Thus, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ being linearly independent implies $\text{Nul}(A) = \{\mathbf{0}\}$ and vice versa.

Equivalently, the columns of A are linearly independent if and only if $A\mathbf{c} = \mathbf{0}$ has no free variables and therefore only the trivial solution.

Let's take a look at another equivalent interpretation:

Linear Dependence and Redundancy

A set of two or more vectors

$$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$$

is linearly dependent if and only if at least one vector in the set can be written as a linear combination of the others.

To see why, suppose that the vectors satisfy a linear dependence relation:

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$$

Because the relation is nontrivial, at least one coefficient is nonzero. If $c_j \neq 0$, then we can solve for $\mathbf{v}_j$:

$$\mathbf{v}_j = -\frac{c_1}{c_j}\mathbf{v}_1 - \cdots - \frac{c_{j-1}}{c_j}\mathbf{v}_{j-1} - \frac{c_{j+1}}{c_j}\mathbf{v}_{j+1} - \cdots - \frac{c_p}{c_j}\mathbf{v}_p$$

Thus, removing $\mathbf{v}_j$ does not change the span and thus is a linear combination of the remaining vectors.

Conversely, if one vector is a linear combination of the others, then moving all terms to one side produces a nontrivial linear dependence relation.

With the definition of linear independence and its two equivalent interpretations, let's try a problem:

Problem 4.4.1

Determine whether each of the following sets of vectors is linearly independent or linearly dependent. Briefly justify your answer.

(a)
$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

(b)
$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0.00001 \end{bmatrix} \right\}$$

(c)
$$\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix} \right\}$$

(d)
$$\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}$$

(e) Any set of three vectors in $\mathbb{R}^2$.

Solution

(a) The set is linearly independent. Neither vector is a scalar multiple of the other, so neither can be written as a linear combination of the other.

(b) The set is linearly independent. Neither vector is a scalar multiple of the other, so neither can be written as a linear combination of the other.

(c) The set is linearly dependent because

$$\begin{bmatrix} -1 \\ -1 \end{bmatrix} = -1 \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Thus, one vector is a scalar multiple of the other.

(d) The set is linearly dependent. Any set containing the zero vector is linearly dependent since

$$1 \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \mathbf{0},$$

which is a nontrivial linear dependence relation.

(e) Any set of three vectors in $\mathbb{R}^2$ is linearly dependent. Since $\mathbb{R}^2$ has dimension 2, at most two vectors can be linearly independent. Therefore, every set containing three or more vectors in $\mathbb{R}^2$ must be linearly dependent.

A spanning set contains enough vectors to generate a space, while a linearly independent set contains no redundant vectors. A **basis** combines these two ideas. It is a set of vectors that spans a vector space (or subspace) while containing no unnecessary vectors. Consequently, every vector in the space can be written uniquely as a linear combination of the basis vectors.

Basis

Let H be a subspace of a vector space V. A set of vectors

$$\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\}$$

is called a **basis** for H if

(i) $\mathcal{B}$ is linearly independent, and

(ii) $H = \text{Span}\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\}$.

Example 4.4.2

Let A be an invertible $n \times n$ matrix with columns

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}.$$

Show that the columns of A form a basis for $\mathbb{R}^n$.

Solution

To show that the columns of A form a basis, we must verify the two conditions in the definition of a basis.

First, the columns must be linearly independent. Since A is invertible, the invertible matrix theorem tells us that the homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Therefore, the columns of A are linearly independent.

Next, the columns must span $\mathbb{R}^n$. Again by the invertible matrix theorem, the equation $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b} \in \mathbb{R}^n$. Thus every vector in $\mathbb{R}^n$ can be expressed as a linear combination of the columns of A. Equivalently,

$$\text{Col}(A) = \mathbb{R}^n.$$

Since the columns of A are linearly independent and span $\mathbb{R}^n$, they form a basis for $\mathbb{R}^n$.

Let's show that the columns of the identity matrix I_n form a basis for $\mathbb{R}^n$.

The columns of the identity matrix are

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

Since I_n is invertible, the invertible matrix theorem tells us that its columns are linearly independent and span $\mathbb{R}^n$. Therefore,

$$\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$$

is a basis for $\mathbb{R}^n$. This set is called the **standard basis** for $\mathbb{R}^n$.

Furthermore, every vector in $\mathbb{R}^n$ can be written uniquely as a linear combination of the standard basis vectors. For example,

$$\begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix} = 3\mathbf{e}_1 - \mathbf{e}_2 + 5\mathbf{e}_3.$$

Example 4.4.5

Determine whether the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 0 \\ -6 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -4 \\ 1 \\ 7 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -2 \\ 1 \\ 5 \end{bmatrix}$$

form a basis for $\mathbb{R}^3$.

Solution

We place the vectors into the columns of a matrix:

$$A = \begin{bmatrix} 3 & -4 & -2 \\ 0 & 1 & 1 \\ -6 & 7 & 5 \end{bmatrix}.$$

Row reducing gives

$$\sim \begin{bmatrix} 3 & -4 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

Since there is a pivot in every column, the columns of A are linearly independent. Further, because A is a 3×3 matrix with a pivot in every column, it is invertible. By the invertible matrix theorem, the columns also span $\mathbb{R}^3$.

Therefore, $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis for $\mathbb{R}^3$.

In practice, we often begin with a spanning set that contains more vectors than are actually needed and simplify it into a basis. The spanning set theorem shows that any redundant vectors can be removed without changing the span, allowing us to systematically construct a basis.

Spanning Set Theorem

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ be a set in V. And let H be a vector space spanned by a set of vectors

$$H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}.$$

Then, the following is true:

(i) If a vector $\mathbf{v}_k \in S$ can be written as a linear combination of the others in S, the set created without $\mathbf{v}_k$ would still spans H.

(ii) If $H \neq \{\mathbf{0}\}$, then removing redundant vectors from a spanning set eventually produces a subset of the original spanning set S that is a basis for H.

Proof

Suppose one of the vectors, say $\mathbf{v}_p$, can be written as a linear combination of the others:

$$\mathbf{v}_p = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_{p-1}\mathbf{v}_{p-1}$$

Take any vector $\mathbf{x} \in H$. Since the vectors span H,

$$\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_{p-1}\mathbf{v}_{p-1} + c_p\mathbf{v}_p.$$

Substituting the expression for $\mathbf{v}_p$ shows that $\mathbf{x}$ can be written entirely as a linear combination of

$$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{p-1}.$$

Therefore removing $\mathbf{v}_p$ does not change the span.

If the remaining set is still linearly dependent, we simply repeat the process. Eventually no redundant vectors remain. The resulting set is linearly independent and still spans H, so it is a basis. □

We will now work on identifying **bases** for each subspace.

For the null space, nothing new is required. Solving the homogeneous system $A\mathbf{x} = \mathbf{0}$ produces the general solution as a linear combination of vectors associated with the free variables. These vectors are automatically linearly independent, so they not only span the null space, but also form a basis for $\text{Nul}(A)$.

Finding a basis for the column space is slightly different. The columns of A already span $\text{Col}(A)$, but some of these columns may be redundant. By the spanning set theorem, every non-pivot column can be removed without changing the span. The remaining pivot columns are linearly independent and still span the column space, so they form a basis.

There is something subtle to remember here. Although row operations preserve the linear dependence relationships among the columns of a matrix, they do *not* preserve the column space itself. Consequently, a basis for $\text{Col}(A)$ must always be selected from the *original* matrix, not from its REF.

Suppose

$$A \sim U,$$

where U is the REF (or RREF) of A. The pivot columns of U identify which columns of the original matrix are linearly independent. Thus, the procedure for finding a basis for the column space is:

1. Row reduce A to REF (or RREF).
2. Identify the pivot columns.
3. Return to the original matrix A.
4. The corresponding pivot columns of A form a basis for $\text{Col}(A)$.

Here is one thing to keep in mind. A vector space generally has infinitely many possible bases. For example, if a vector belongs to a basis, then any nonzero scalar multiple of that vector may be used instead without changing the span. The pivot columns simply provide one convenient and systematic choice of basis.

When finding a basis for the column space of a matrix, always choose the pivot columns from the *original* matrix, not from its REF. Although row operations preserve the pivot locations and the linear dependence relationships among the columns, please remember once again they do *not* preserve the column space itself.

To conclude, here are two equivalent ways to think about a basis.

A basis is indeed a spanning set that is *as small as possible*. Once a spanning set has been reduced to a basis, removing any additional vector causes the remaining vectors to no longer span the space.

Equivalently, a basis is a linearly independent set that is *as large as possible*. Once a basis has been obtained, adding any additional vector from the vector space necessarily introduces linear dependence, since every vector in the space can already be written as a linear combination of the basis vectors.

A basis is simultaneously a minimal spanning set and a maximal linearly independent set.

Problem 4.4.2

Let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -3 \\ 4 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 6 \\ 2 \\ -1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} -4 \\ -8 \\ 9 \end{bmatrix}.$$

Find a basis for the subspace $W = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$.

Solution

Place the spanning vectors into the columns of a matrix:

$$A = \begin{bmatrix} 1 & 6 & 2 & -4 \\ -3 & 2 & -2 & -8 \\ 4 & -1 & 3 & 9 \end{bmatrix}.$$

Row reducing to REF gives

$$\sim \begin{bmatrix} 1 & 6 & 2 & -4 \\ 0 & 20 & 4 & -20 \\ 0 & -25 & -5 & 25 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 2 & -4 \\ 0 & 5 & 1 & -5 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The pivot columns are columns one and two. Therefore, the corresponding columns of the *original* matrix form a basis for the column space of A, which is exactly the subspace W.

Hence, a basis for W is

$$\left\{ \begin{bmatrix} 1 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} 6 \\ 2 \\ -1 \end{bmatrix} \right\}.$$

Problem 4.4.3

Consider the subspace

$$V = \text{Span} \left\{ \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 5 \\ -2 \end{bmatrix} \right\}.$$

Find a basis for V that is different from the one obtained using the pivot columns.

Solution

The pivot-column method produces one basis for V, but a vector space generally has infinitely many different bases. To construct another basis, we simply need to find two linearly independent vectors that lie in V.

Since every linear combination of the spanning vectors is still in V, consider

$$\mathbf{w}_1 = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ -5 \\ 6 \end{bmatrix},$$

and

$$\mathbf{w}_2 = -\begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ 5 \\ -4 \end{bmatrix}.$$

Since both vectors are linear combinations of the original spanning vectors, $\mathbf{w}_1, \mathbf{w}_2 \in V$.

Furthermore, $\mathbf{w}_2 \neq c\mathbf{w}_1$ for any scalar c, so the vectors are linearly independent.

Therefore,

$$\left\{ \begin{bmatrix} 3 \\ -5 \\ 6 \end{bmatrix}, \begin{bmatrix} -2 \\ 5 \\ -4 \end{bmatrix} \right\}$$

is another basis for V.

This example illustrates an important fact: a vector space generally has infinitely many different bases. The pivot-column method produces one convenient basis, but it is by no means the only one.

Problem 4.4.4

Find a basis for the subspace of $\mathbb{R}^3$ defined by the plane $x + 2y + z = 0$.

Solution

The equation can be viewed as the homogeneous system

$$\begin{bmatrix} 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{0}.$$

Thus, we seek a basis for the null space of the matrix

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

Writing the system,

$$x + 2y + z = 0,$$

we solve for the pivot variable:

$$x = -2y - z.$$

Since y and z are free variables, let $y = s$ and $z = t$. Then

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -2s - t \\ s \\ t \end{bmatrix} = s \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

Therefore,

$$\text{Nul}(A) = \text{Span} \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\},$$

so a basis for the plane is

$$\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

4.5 Coordinates, Dimension, and Rank

After introducing the concept of a basis, we now turn our attention to how a basis can be used to describe vectors. Just as the standard basis provides the familiar coordinate system for $\mathbb{R}^n$, any basis defines its own coordinate system for a vector space. Relative to a chosen basis, every vector can be represented by a unique collection of scalars called its coordinates.

This idea rests on one of the most important ideas in linear algebra: every vector has a unique representation as a linear combination of the vectors in a basis.

We will focus on vector spaces that possess a *finite* basis. Such spaces are called **finite-dimensional vector spaces**. We will learn to assign coordinates to vectors, define the dimension of a space, and measure the size of important subspaces such as the column space and null space through the concepts of rank and nullity.

Finite-Dimensional Vector Space

A vector space is called **finite-dimensional** if it can be spanned by a finite number of vectors. Equivalently, a vector space is finite-dimensional if it has a basis consisting of finitely many vectors. Otherwise, the vector space is called **infinite-dimensional**.

After choosing a basis for a vector space, every vector can be written as a unique linear combination of the basis vectors. The coefficients in this linear combination tell us exactly how much of each basis vector is needed to construct the vector. These coefficients are called the **coordinates** of the vector relative to the chosen basis.

Just as the standard basis assigns the familiar coordinates to vectors in $\mathbb{R}^n$, any basis defines its own coordinate system. Although the numerical values of the coordinates depend on the chosen basis, the vector itself will remain unchanged.

Coordinates Relative to a Basis

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ be a basis for a vector space V and let $\mathbf{x} \in V$. If

$$\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n,$$

then the scalars $c_1, c_2, \ldots, c_n$ are called the **coordinates of x relative to the basis** $\mathcal{B}$.

The vector

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

is called the **coordinate vector** of $\mathbf{x}$ relative to $\mathcal{B}$, or the $\mathcal{B}$-coordinate vector.

The function

$$\mathbf{x} \mapsto [\mathbf{x}]_{\mathcal{B}}$$

is called the **coordinate mapping**. It assigns to each vector in V its coordinate vector relative to $\mathcal{B}$.

Keep in mind that $\mathbf{x}$ and $[\mathbf{x}]_{\mathcal{B}}$ are different objects. The vector $\mathbf{x}$ belongs to the vector space V, whereas $[\mathbf{x}]_{\mathcal{B}}$ is the coordinate vector that represents $\mathbf{x}$ relative to the basis $\mathcal{B}$.

Example 4.5.1

Let

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$$

be a basis for $\mathbb{R}^3$. Find the coordinate vector of

$$\mathbf{x} = \begin{bmatrix} 0 \\ 2 \\ 6 \end{bmatrix}$$

relative to the basis $\mathcal{B}$.

Solution

We want scalars c_1, c_2, and c_3 such that

$$c_1 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 6 \end{bmatrix}.$$

These scalars will be the $\mathcal{B}$-coordinates.

Writing this as a matrix equation,

$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 6 \end{bmatrix},$$

and row reducing the augmented matrix gives

$$\left[\begin{array}{ccc|c} 1 & 2 & 0 & 0 \\ 0 & 1 & 1 & 2 \\ 2 & 1 & 1 & 6 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 3 \end{array}\right].$$

We can get

$$c_1 = 2, \quad c_2 = -1, \quad c_3 = 3,$$

Thus the coordinate vector of $\mathbf{x}$ relative to $\mathcal{B}$ is

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}.$$

And indeed,

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

Coordinates depend on the choice of basis. With the standard basis, the entries of a vector are its coordinates. If we choose a different basis, the vector itself does not change, but its coordinate vector will change to. In this sense, a basis defines a new coordinate system for the vector space.

The standard basis provides the simplest coordinate system for $\mathbb{R}^n$. Since every vector is already written as a linear combination of the standard basis vectors, its coordinates are simply its entries.

Example 4.5.2

Express the vector

$$\mathbf{x} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$$

as a linear combination of the standard basis vectors for $\mathbb{R}^2$.

Solution

The standard basis for $\mathbb{R}^2$ is

$$\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2\} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}.$$

Thus,

$$\mathbf{x} = \begin{bmatrix} 1 \\ 6 \end{bmatrix} = 1\mathbf{e}_1 + 6\mathbf{e}_2.$$

Therefore, the coordinate vector of $\mathbf{x}$ relative to the standard basis is

$$[\mathbf{x}]_{\mathcal{E}} = \begin{bmatrix} 1 \\ 6 \end{bmatrix} = \mathbf{x}.$$

Relative to the standard basis $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2\}$,

$$[\mathbf{x}]_{\mathcal{E}} = \mathbf{x}.$$

If $\mathbf{x} \in \mathbb{R}^n$, then the coordinate vector of $\mathbf{x}$ relative to the standard basis is simply $\mathbf{x}$ itself.

Figure 10 illustrates that every basis defines a coordinate system for a vector space. The blue vector $\mathbf{x}$ represents the same physical vector in each panel, while the green and red vectors define different

bases. Although the vector itself does not change, its coordinate vector depends on the basis that is chosen. The relationship between the vector and its coordinate vectors is

$$B \begin{bmatrix} (\mathbf{x}_{\mathcal{B}})_1 \\ (\mathbf{x}_{\mathcal{B}})_2 \end{bmatrix} = \mathbf{x}_{\mathcal{E}} = A \begin{bmatrix} (\mathbf{x}_{\mathcal{A}})_1 \\ (\mathbf{x}_{\mathcal{A}})_2 \end{bmatrix}.$$

$$\mathbf{x} = B[\mathbf{x}]_{\mathscr{B}} = [\mathbf{x}]_{\mathscr{E}} = A[\mathbf{x}]_{\mathscr{A}}$$

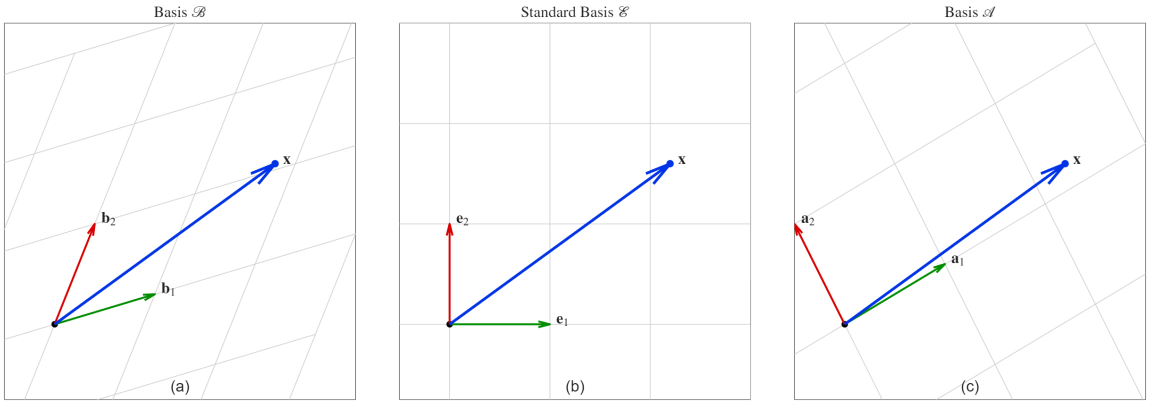


Figure 11: The same vector expressed in different bases

Figure 11 shows the same idea but in 3D.

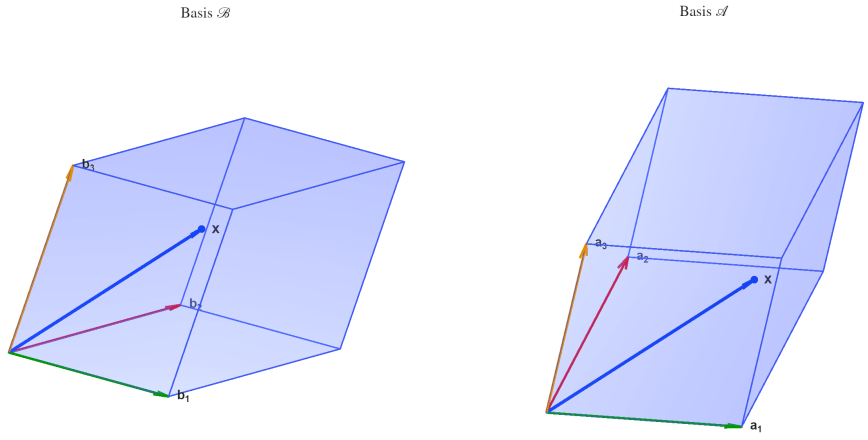


Figure 12: The same vector expressed in different bases

A basis not only provides a coordinate system for a vector space, but it also allows us to convert between a coordinate vector and the vector itself using matrix multiplication.

Suppose

$$\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$$

is a basis for $\mathbb{R}^n$. Write the matrix whose columns are the basis vectors:

$$P_{\mathcal{B}} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix}$$

If

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix},$$

then

$$\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n,$$

which can be written compactly as

$$\mathbf{x} = P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}.$$

We call $P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$ the change-of-coordinates-matrix from $\mathcal{B}$ to $\mathcal{E} \in \mathbb{R}^n$. As you can see, multiplying the coordinate vector by $P_{\mathcal{B}}$ reconstructs the original vector in the standard coordinate system.

Since the columns of $P_{\mathcal{B}}$ form a basis, they are linearly independent, and thus $P_{\mathcal{B}}$ is invertible. Multiplying by its inverse recovers the coordinate vector:

$$[\mathbf{x}]_{\mathcal{B}} = P_{\mathcal{B}}^{-1}\mathbf{x} \tag{23}$$

Equation 23 defines a function that assigns every vector in $\mathbb{R}^n$ its coordinate vector relative to the basis $\mathcal{B}$:

$$\mathbf{x} \longmapsto [\mathbf{x}]_{\mathcal{B}}$$

Recall that this is the coordinate mapping. It converts vectors written in the standard coordinate system into their coordinates relative to the basis $\mathcal{B}$.

Since $P_{\mathcal{B}}$ is invertible, its inverse $P_{\mathcal{B}}^{-1}$ is also invertible. Therefore, the coordinate mapping is a **one-to-one** mapping from $\mathbb{R}^n$ onto $\mathbb{R}^n$; every vector has exactly one coordinate vector, and every coordinate vector corresponds to exactly one vector. Furthermore, because the coordinate mapping is given by multiplication by the matrix $P_{\mathcal{B}}^{-1}$, it is a linear transformation.

The Coordinate Mapping

Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis for a vector space V. Then the coordinate mapping $\mathbf{x} \longmapsto [\mathbf{x}]_{\mathcal{B}}$ is a one-to-one linear transformation from V onto $\mathbb{R}^n$.

This theorem has several important consequences:

- Every vector in V has a unique coordinate vector in $\mathbb{R}^n$.
- Every coordinate vector in $\mathbb{R}^n$ corresponds to exactly one vector in V.
- The coordinate mapping preserves vector addition and scalar multiplication, making it a linear transformation.
- Computations in a finite-dimensional abstract vector space can be carried out using coordinate vectors and familiar matrix operations in $\mathbb{R}^n$.

Because the coordinate mapping is a one-to-one linear transformation, the vector space V and $\mathbb{R}^n$ have exactly the same linear structure. Such vector spaces are called **isomorphisms**. Generally, a one-to-one linear transformation from a vector space V onto a vector space W is called an isomorphism from V onto W. Although the vectors may look different, every linear algebra calculation performed in V can be carried out using coordinate vectors in W and vice versa.

Example 4.5.3

Express the polynomial $p(t) = 3 - 2t + 5t^2 + t^3$ as a coordinate vector relative to the standard basis $\mathcal{B} = \{1, t, t^2, t^3\}$ for the polynomial vector space $\mathbb{P}_3$. A typical element $\mathbf{p} \in \mathbb{R}^3$ has the form

$$\mathbf{p}(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3.$$

Solution

Since $\mathbf{p}(t)$ is already written as a linear combination of the basis vectors,

$$\mathbf{p}(t) = 3(1) - 2t + 5t^2 + 1t^3,$$

its coordinate vector is simply the list of coefficients:

$$[\mathbf{p}]_{\mathcal{B}} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \\ 5 \\ 1 \end{bmatrix}$$

Thus, the coordinates of a polynomial relative to the standard basis are simply it's coefficients. We can also conclude that the coordinate mapping $\mathbf{p} \mapsto [\mathbf{p}]_{\mathcal{B}}$ is an isomorphism from $\mathbb{P}_3$ onto $\mathbb{R}^4$, or that every polynomial in $\mathbb{P}_3$ can be represented by a unique vector in $\mathbb{R}^4$. Furthermore, since all vector space operations in $\mathbb{P}_3$ correspond to operations in $\mathbb{R}^4$, operations on polynomials can be done by performing the corresponding operations on their coordinate vectors.

Problem 4.5.1

Let

$$\mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} -3 \\ 4 \\ 0 \end{bmatrix}, \quad \mathbf{b}_3 = \begin{bmatrix} 3 \\ -6 \\ 3 \end{bmatrix},$$

and let

$$\mathbf{x} = \begin{bmatrix} -8 \\ 2 \\ 3 \end{bmatrix}.$$

Use the basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ to answer the following questions.

(a) Show that $\mathcal{B}$ is a basis for $\mathbb{R}^3$.

(b) Find the change-of-coordinates matrix $P_{\mathcal{B}}$ from $\mathcal{B}$ to the standard basis.

(c) Write the equation relating $\mathbf{x} \in \mathbb{R}^3$ to $[\mathbf{x}]_{\mathcal{B}}$.

(d) Find the coordinate vector $[\mathbf{x}]_{\mathcal{B}}$.

Solution

(a) Place the vectors of $\mathcal{B}$ into the columns of a matrix:

$$P_{\mathcal{B}} = \begin{bmatrix} 1 & -3 & 3 \\ 0 & 4 & -6 \\ 0 & 0 & 3 \end{bmatrix}$$

Since $P_{\mathcal{B}}$ is upper triangular with nonzero diagonal entries, it is invertible. Therefore, its columns are linearly independent and span $\mathbb{R}^3$, so $\mathcal{B}$ is a basis for $\mathbb{R}^3$.

(b) The change-of-coordinates matrix from $\mathcal{B}$ to the standard basis is

$$P_{\mathcal{B}} = \begin{bmatrix} 1 & -3 & 3 \\ 0 & 4 & -6 \\ 0 & 0 & 3 \end{bmatrix}.$$

(c) The relationship between $\mathbf{x}$ and its coordinate vector is $\mathbf{x} = P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$.

We have

$$\begin{bmatrix} -8 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & -3 & 3 \\ 0 & 4 & -6 \\ 0 & 0 & 3 \end{bmatrix} [\mathbf{x}]_{\mathcal{B}}.$$

(d) Solve the matrix equation by row reducing the augmented matrix:

$$\begin{bmatrix} 1 & -3 & 3 & -8 \\ 0 & 4 & -6 & 2 \\ 0 & 0 & 3 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & -5 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Therefore,

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} -5 \\ 2 \\ 1 \end{bmatrix}.$$

We can check our work to see that

$$-5\mathbf{b}_1 + 2\mathbf{b}_2 + \mathbf{b}_3 = \begin{bmatrix} -8 \\ 2 \\ 3 \end{bmatrix} = \mathbf{x}.$$

Problem 4.5.2

Let $\mathcal{B} = \{1 + t,\ 1 + t^2,\ t + t^2\}$ be a basis for $\mathbb{P}_2$.

Find the coordinate vector of $\mathbf{p}(t) = 6 + 3t - t^2$ relative to $\mathcal{B}$.

Solution

Express $\mathbf{p}(t)$ as a linear combination of the basis vectors:

$$c_1(1+t)+c_2(1+t^2)+c_3(t+t^2)=6+3t-t^2.$$

Expanding the left-hand side gives

$$(c_1+c_2)+(c_1+c_3)t+(c_2+c_3)t^2=6+3t-t^2.$$

Equating the coefficients of like powers of t,

$$\begin{cases} c_1+c_2 = 6 \\ c_1+c_3 = 3 \\ c_2+c_3 = -1 \end{cases}$$

Solving the system yields

$$c_1=5, \quad c_2=1, \quad c_3=-2$$

Therefore,

$$[\mathbf{p}]_{\mathcal{B}}=\begin{bmatrix} 5 \\ 1 \\ -2 \end{bmatrix}.$$

Problem 4.5.3

Let $\mathcal{B}=\{\mathbf{b}_1,\mathbf{b}_2,\ldots,\mathbf{b}_n\}$ be a basis for a vector space V.

Explain why the $\mathcal{B}$-coordinate vectors of $\mathbf{b}_1,\mathbf{b}_2,\ldots,\mathbf{b}_n$ are the columns $\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_n$ of the $n\times n$ identity matrix.

Solution

For each $k = 1, 2, \ldots, n$, the basis vector $\mathbf{b}_k$ can be written as a linear combination of the vectors in $\mathcal{B}$ by assigning the coefficient 1 to $\mathbf{b}_k$ and the coefficient 0 to every other basis vector:

$$\mathbf{b}_k = 0\mathbf{b}_1 + \cdots + 0\mathbf{b}_{k-1} + 1\mathbf{b}_k + 0\mathbf{b}_{k+1} + \cdots + 0\mathbf{b}_n.$$

The coordinate vector $[\mathbf{b}_k]_{\mathcal{B}}$ is formed from these coefficients, listed in the same order as the basis vectors in $\mathcal{B}$. Therefore,

$$[\mathbf{b}_k]_{\mathcal{B}} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{e}_k,$$

where the entry 1 appears in the kth position and all other entries are zero:

$$[\mathbf{b}_1]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{e}_1, \quad [\mathbf{b}_2]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{e}_2, \quad \ldots, \quad [\mathbf{b}_n]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} = \mathbf{e}_n.$$

Hence,

$$[\mathbf{b}_1]_{\mathcal{B}} = \mathbf{e}_1, \quad [\mathbf{b}_2]_{\mathcal{B}} = \mathbf{e}_2, \quad \ldots, \quad [\mathbf{b}_n]_{\mathcal{B}} = \mathbf{e}_n.$$

The vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ are the columns of the $n \times n$ identity matrix:

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix}.$$

Therefore, the $\mathcal{B}$-coordinate vectors of $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n$ are the columns of I_n.

The coordinate mapping established an important connection between an abstract vector space and the familiar space $\mathbb{R}^n$. We can now use this relationship to prove general results about vector spaces by applying properties that are already known to hold in $\mathbb{R}^n$.

One such property is that no more than n vectors in $\mathbb{R}^n$ can be linearly independent. This fact extends to every vector space with a basis containing n vectors.

Maximum Size of a Linearly Independent Set

Let V be a vector space with basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$. Then any set in V containing more than n vectors must be linearly dependent.

Proof

Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ be a set of vectors in V with $p > n$.

Since $\mathcal{B}$ is a basis, each vector has a unique coordinate vector relative to $\mathcal{B}$. Thus,

$$[\mathbf{u}_1]_{\mathcal{B}},\ [\mathbf{u}_2]_{\mathcal{B}},\ \ldots,\ [\mathbf{u}_p]_{\mathcal{B}}$$

are p vectors in $\mathbb{R}^n$.

Since $p > n$, these coordinate vectors must be linearly dependent. Therefore, there exist scalars $c_1, c_2, \ldots, c_p$, not all zero, such that

$$c_1[\mathbf{u}_1]_{\mathcal{B}} + c_2[\mathbf{u}_2]_{\mathcal{B}} + \cdots + c_p[\mathbf{u}_p]_{\mathcal{B}} = \mathbf{0}.$$

Because the coordinate mapping is linear,

$$[c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p]_{\mathcal{B}} = \mathbf{0}.$$

The only vector whose coordinate vector is $\mathbf{0}$ is the zero vector. Hence,

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p = \mathbf{0}.$$

Since the scalars are not all zero, the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ is linearly dependent. □

We now use this result to establish one of the most important facts in linear algebra: although a vector space may have many different bases, they all contain the same number of vectors.

Basis Theorem

If a vector space V has a basis consisting of n vectors, then every basis of V must consist of exactly n vectors.

Proof of Basis Theorem

Let $\mathcal{B}_1$ be a basis for V consisting of n vectors, and let $\mathcal{B}_2$ be any other basis for V.

Since $\mathcal{B}_2$ is linearly independent, we know that $\mathcal{B}_2$ can contain no more than n vectors.

Likewise, because $\mathcal{B}_2$ is a basis and $\mathcal{B}_1$ is linearly independent, we know that $\mathcal{B}_1$ can contain no more vectors than $\mathcal{B}_2$. Therefore, $\mathcal{B}_2$ must contain at least n vectors.

Hence, $\mathcal{B}_2$ contains both at most and at least n vectors, so it must contain exactly n vectors. □

The basis theorem guarantees that every basis of a finite-dimensional vector space contains the same number of vectors. Consequently, the number of vectors in a basis is an intrinsic property of the vector space rather than a property of any particular basis. This allows us to define the dimension of a vector space.

Dimension of a Vector Space

A vector space V is said to be **finite-dimensional** if it is spanned by a finite set. The **dimension** of V, denoted by $\dim(V)$ is the number of vectors in any basis for V.

The zero vector space $\{\mathbf{0}\}$ has dimension zero. A vector space that is not spanned by a finite set is called **infinite-dimensional**.

Some vector spaces have standard bases, making their dimensions easy to determine. Since the standard basis of $\mathbb{R}^n$ consists of n vectors,

$$\dim(\mathbb{R}^n) = n.$$

Likewise, the standard basis for the polynomial space $\mathbb{P}_n$ is

$$\{1, t, t^2, \ldots, t^n\},$$

which contains $n+1$ vectors. Therefore,

$$\dim(\mathbb{P}_n) = n + 1.$$

Similarly, the standard basis for the space of all $m \times n$ matrices consists of the $m \times n$ matrices having a single entry equal to one and all remaining entries equal to zero. Thus,

$$\dim(M_{m,n}) = mn.$$

Finally, the vector space $\mathbb{P}$ of all polynomials is infinite-dimensional since no finite set of polynomials can span polynomials of arbitrarily high degree.

Example 4.5.4

Find the dimension of the subspace

$$H = \left\{ \begin{bmatrix} a - 3b + 6c \\ 5a + 4d \\ b - 2c - d \\ 5d \end{bmatrix} : a, b, c, d \in \mathbb{R} \right\}.$$

Solution

Rewrite the vector in terms of the parameters:

$$\begin{bmatrix} a-3b+6c \\ 5a+4d \\ b-2c-d \\ 5d \end{bmatrix} = a\begin{bmatrix} 1 \\ 5 \\ 0 \\ 0 \end{bmatrix} + b\begin{bmatrix} -3 \\ 0 \\ 1 \\ 0 \end{bmatrix} + c\begin{bmatrix} 6 \\ 0 \\ -2 \\ 0 \end{bmatrix} + d\begin{bmatrix} 0 \\ 4 \\ -1 \\ 5 \end{bmatrix}.$$

Thus, $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ where

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 5 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -3 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 6 \\ 0 \\ -2 \\ 0 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 0 \\ 4 \\ -1 \\ 5 \end{bmatrix}.$$

Notice that $\mathbf{v}_3 = -2\mathbf{v}_2$, so $\mathbf{v}_3$ is redundant and may be removed without changing the span. The remaining vectors $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}$ are linearly independent, so they form a basis for H.

Since this basis contains three vectors, $\dim(H) = 3$.

In $\mathbb{R}^3$, the dimension of a subspace has a simple geometric interpretation:

- A zero-dimensional subspace is the zero subspace, $\{\mathbf{0}\}$.
- A one-dimensional subspace is any line through the origin, spanned by a single nonzero vector. See Figure 12a.
- A two-dimensional subspace is any plane through the origin, spanned by two linearly independent vectors. See Figure 12b.
- A three-dimensional subspace is $\mathbb{R}^3$ itself. See Figure 12c.

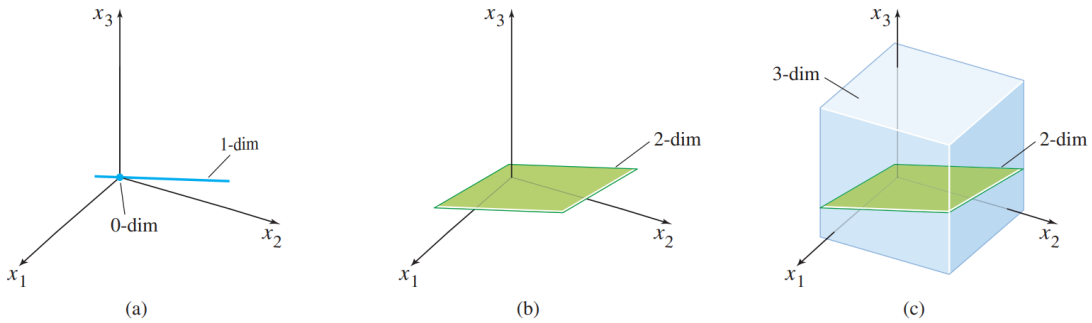


Figure 13: Some subspaces in $\mathbb{R}^3$. Image credit: Lay et al

The previous theorems established that the dimension of a finite-dimensional vector space is well-defined. We now turn our attention to its subspaces. It makes sense that a subspace cannot be "larger" than the space that contains it. The next theorem makes this idea precise by showing that every subspace of a finite-dimensional vector space is finite-dimensional, and that its dimension can never exceed that of the ambient vector space.

Subspaces of a Finite-Dimensional Vector Space

Let H be a subspace of a finite-dimensional vector space V. Then every linearly independent set in H can be extended, if necessary, to form a basis for H. Consequently, H is finite-dimensional and

$$\dim(H) \leq \dim(V).$$

Proof

In the case that $H = \{\mathbf{0}\}$, it is clear to see that $\dim(H) = 0 \leq \dim(V)$.

Otherwise, let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ be any linearly independent set in H.

If S already spans H, then S is a basis for H.

If not, there exists a vector in H that is not contained in $\text{Span}(S)$. Adding this vector to S preserves linear independence, since it cannot be written as a linear combination of the existing vectors.

We may continue this process, adding vectors from H whenever the current set does not yet span H. However, we know that no linearly independent set in V can contain more than $\dim(V)$ vectors. Therefore, this process must eventually terminate.

When it terminates, the resulting set is both linearly independent and spans H, so it is a basis

for H. Thus, H is finite-dimensional. Since this basis is also a linearly independent set in V, it contains at most $\dim(V)$ vectors. Hence,

$$\dim(H) \leq \dim(V).$$

□

We now know that every subspace of a finite-dimensional vector space has a well-defined dimension. Knowing this dimension provides a powerful way to identify bases. Rather than verifying both linear independence and spanning, the next theorem shows that it is enough to check just one of these properties, provided the set contains the correct number of vectors.

Basis Criterion

Let V be a p-dimensional vector space, where $p \geq 1$.

1. Any linearly independent set containing exactly p vectors is a basis for V.
2. Any spanning set containing exactly p vectors is a basis for V.

Proof

Suppose S is a linearly independent set containing exactly p vectors.

Since $\dim(V) = p$, every basis of V contains exactly p vectors. We know that every linearly independent set can be extended to a basis. However, S already contains the maximum possible number of linearly independent vectors, so no additional vectors can be added. Therefore, S is already a basis for V.

Now suppose S spans V and contains exactly p vectors.

If S were linearly dependent, then one of its vectors could be removed while the remaining vectors would still span V. This would produce a spanning set containing fewer than p vectors, contradicting the fact that every basis of V has exactly p vectors.

Hence, S must be linearly independent. Since it both spans V and is linearly independent, S is a basis for V.

The concept of dimension becomes especially useful when applied to the fundamental subspaces of a matrix. Recall that the pivot columns of a matrix form a basis for its column space, while the special solutions corresponding to the free variables form a basis for its null space. As a result, the dimensions of these subspaces can be determined directly from the REF of the matrix, without explicitly constructing each basis.

Dimensions of the Fundamental Subspaces

Let A be an $m \times n$ matrix.

- The dimension of $\text{Col}(A)$ equals the number of pivot columns of A.
- The dimension of $\text{Nul}(A)$ equals the number of free variables in the homogeneous system $A\mathbf{x} = \mathbf{0}$.

Example 4.5.5

Find the dimensions of the null space and the column space of

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}.$$

Solution

Row reduce the augmented matrix to REF:

$$\left[\begin{array}{ccccc|c} -3 & 6 & -1 & 1 & -7 & 0 \\ 1 & -2 & 2 & 3 & -1 & 0 \\ 2 & -4 & 5 & 8 & -4 & 0 \end{array}\right] \sim \left[\begin{array}{ccccc|c} 1 & -2 & 2 & 3 & -1 & 0 \\ 0 & 0 & 1 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

The pivot columns are columns 1 and 3, so the column space has a basis consisting of two vectors. Therefore, $\dim(\text{Col}(A)) = 2$.

The remaining variables, x_2, x_4, and x_5, are free variables. Since each free variable contributes one basis vector to the null space, $\dim(\text{Nul}(A)) = 3$.

Problem 4.5.4

For a vector space V, determine whether each statement is true or false. If a statement is false, briefly explain why.

(a) The number of pivot columns in a matrix equals the dimension of its column space.

(b) A plane in $\mathbb{R}^3$ is a two-dimensional subspace of $\mathbb{R}^3$.

(c) The dimension of the vector space $\mathbb{P}_4$ is 4.

(d) If $\dim(V) = n$ and S is a linearly independent set in V, then S is a basis for V.

(e) If a set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ spans a finite-dimensional vector space V, and if $T \subseteq V$ contains more than p vectors, then T is linearly dependent.

Solution

(a) True. The pivot columns of a matrix form a basis for its column space. Therefore, the number of pivot columns equals the dimension of $\text{Col}(A)$.

(b) False. A plane is a subspace of $\mathbb{R}^3$ only if it passes through the origin. The zero vector cannot satisfy $x + y + z = 1$.

(c) False. The standard basis for $\mathbb{P}_4$ is

$$\{1, x, x^2, x^3, x^4\},$$

which contains five vectors. Therefore,

$$\dim(\mathbb{P}_4) = 5.$$

(d) False. A linearly independent set is a basis only if it contains exactly n vectors. A linearly independent set with fewer than n vectors does not span V.

(e) True. Since $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ spans V, every basis of V contains at most p vectors. Any subset $T \subseteq V$ containing more than p vectors must therefore be linearly dependent.

The column space is formed by taking all linear combinations of the columns of a matrix. By symmetry, we may also consider all linear combinations of its rows. This leads to the row space, which we discussed earlier. Although it is defined using the rows of a matrix rather than its columns, the row space shares many of the same fundamental properties. We can apply it now because it provides a natural way to measure the amount of independent information contained in a matrix, leading to the important concept of **rank**. We continue our study of the four fundamental subspaces by defining it in greater detail.

Row Space

Let A be an $m \times n$ matrix with row vectors

$$\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m.$$

The **row space** of A, denoted $\text{Row}(A)$, is the set of all linear combinations of the row vectors of A:

$$\text{Row}(A) = \text{Span}\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m\}.$$

Since each row of A has n entries, every row vector belongs to $\mathbb{R}^n$. Therefore,

$$\text{Row}(A) \subseteq \mathbb{R}^n.$$

Example 4.5.6

Find a basis for the row space of the matrix

$$A = \begin{bmatrix} -2 & -5 & 8 & 0 & -17 \\ 1 & 3 & -5 & 1 & 5 \\ 3 & 11 & -19 & 7 & 1 \\ 1 & 7 & -13 & 5 & -3 \end{bmatrix}.$$

Solution

Row reducing A to REF gives

$$\sim \begin{bmatrix} 1 & 3 & -5 & 1 & 5 \\ 0 & 1 & -2 & 2 & -7 \\ 0 & 0 & 0 & -4 & 20 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The nonzero rows of an echelon form of a matrix form a basis for its row space. Therefore, a basis for $\text{Row}(A)$ is

$$\left\{\begin{bmatrix} 1 & 3 & -5 & 1 & 5 \end{bmatrix}, \begin{bmatrix} 0 & 1 & -2 & 2 & -7 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & -4 & 20 \end{bmatrix}\right\}.$$

The previous example illustrates how we can efficiently find a basis for the row space. Although row operations change the individual rows of a matrix, they do not change the subspace spanned by those rows. It then follows that row reducing a matrix provides a convenient way to determine a basis for its row space.

Basis for the Row Space

Let matrices A and B be row equivalent. Then both of the following statements are true.

1. $\text{Row}(A) = \text{Row}(B)$.
2. If B is in REF, then the nonzero rows of B form a basis for both $\text{Row}(A)$ and $\text{Row}(B)$.

The nonzero rows of a matrix in REF generally are *not* rows of the original matrix. Although they form a basis for $\text{Row}(A)$, it is incorrect to conclude that the corresponding rows of A are linearly independent. Row operations preserve the row space, but they may change the linear dependence relations among the rows.

If a basis consisting entirely of rows from the original matrix is desired, row reduce A^T to identify its pivot columns. These pivot columns of A^T correspond to rows of A, and those rows form a basis for $\text{Row}(A)$.

Although the column and row spaces are defined in different ways, it turns out they have exactly the same dimension. This common dimension is the rank of the matrix. Together with the dimension of the null space (nullity), rank provides one of the most fundamental relationships in linear algebra.

Rank

Let A be an $m \times n$ matrix. The **rank** of A, denoted $\text{rank}(A)$, is the dimension of its column space:

$$\text{rank}(A) = \dim(\text{Col}(A)).$$

The Rank Theorem

Let A be an $m \times n$ matrix.

1. $\dim(\text{Row}(A)) = \dim(\text{Col}(A)) = \text{rank}(A)$.
2. The rank of A equals the number of pivot positions in any REF of A.
3. $\text{rank}(A) + \dim(\text{Nul}(A)) = n$, where n is the number of columns of A.

Proof of the Rank Theorem

Since the pivot columns of a matrix form a basis for its column space, the rank of A is equal to the number of pivot columns in A. Equivalently, the rank is the number of pivot positions in any REF of A.

Furthermore, every pivot position corresponds to a nonzero row in the REF, and these nonzero rows form a basis for $\text{Row}(A)$. Therefore,

$$\dim(\text{Row}(A)) = \dim(\text{Col}(A)) = \text{rank}(A).$$

Next, recall that the dimension of the null space is equal to the number of free variables in the homogeneous system $A\mathbf{x} = \mathbf{0}$.

Since every nonpivot column corresponds to one free variable, $\dim(\text{Nul}(A))$ is equal to the number of nonpivot columns.

Finally, every column of A is either a pivot column or a nonpivot column. Hence,

$$\underbrace{\text{number of pivot columns}}_{\text{rank}(A)} + \underbrace{\text{number of nonpivot columns}}_{\dim(\text{Nul}(A))} = \underbrace{\text{number of columns}}_{n}$$

$$\text{rank}(A) + \dim(\text{Nul}(A)) = n.$$

□

Example 4.5.7

Answer the following questions using the rank theorem.

(a) If A is a 7×9 matrix with a two-dimensional null space, what is the rank of A?

(b) Can a 6×9 matrix B have a two-dimensional null space? Explain your reasoning.

Solution

(a) Since A has 9 columns,

$$\text{rank}(A) + \dim(\text{Nul}(A)) = 9.$$

Substituting $\dim(\text{Nul}(A)) = 2$ gives

$$\text{rank}(A) + 2 = 9,$$

so

$$\text{rank}(A) = 7.$$

(b) No. Since B has 9 columns,

$$\text{rank}(A) + 2 = 9 \Rightarrow \text{rank}(A) = 7.$$

However, the rank of a matrix cannot exceed the smaller of its number of rows and columns. Since A has only 6 rows,

$$\text{rank}(A) \leq 6,$$

which is a contradiction. Such a matrix cannot exist.

Example 4.5.8

A climate scientist is calibrating an atmospheric model that relates 42 unknown parameters to 40 independent measurements. After analyzing the associated homogeneous system, the scientist finds two linearly independent solutions, and every solution can be written as a linear combination of these two solutions. Can the scientist conclude that, for any collection of 40 measurements, the associated nonhomogeneous system $A\mathbf{x} = \mathbf{b}$ has a solution for any collection of measurements $\mathbf{b} \in \mathbb{R}^{40}$?

Solution

Since the two solutions are linearly independent and span the solution space of the homogeneous system,

$$\dim(\text{Nul}(A)) = 2.$$

Because the coefficient matrix has 42 columns, the rank theorem gives

$$\text{rank}(A) + 2 = 42 \Rightarrow \text{rank}(A) = 40.$$

Since $\text{rank}(A) = \dim(\text{Col}(A))$, it follows that

$$\dim(\text{Col}(A)) = 40.$$

The column space is therefore a 40-dimensional subspace of $\mathbb{R}^{40}$. The only 40-dimensional subspace of $\mathbb{R}^{40}$ is $\mathbb{R}^{40}$ itself, so

$$\text{Col}(A) = \mathbb{R}^{40}.$$

Consequently, every vector $\mathbf{b} \in \mathbb{R}^{40}$ belongs to the column space of A. Therefore, the system $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b} \in \mathbb{R}^{40}$.

The concepts of dimension and rank provide several new characterizations of invertible matrices. As a result, we will have to expand on the invertible matrix theorem.

The Invertible Matrix Theorem (continued)

Let A be an $n \times n$ matrix. The following statements are equivalent to the statement that A is invertible.

(m) The columns of A form a basis for $\mathbb{R}^n$.

(n) $\text{Col}(A) = \mathbb{R}^n$.

(o) $\dim(\text{Col}(A)) = n$.

(p) $\text{rank}(A) = n$.

(q) $\text{Nul}(A) = \{\mathbf{0}\}$.

(r) $\dim(\text{Nul}(A)) = 0$.

For brevity, you may see some abbreviations. Let A be an $m \times n$ matrix with rank r, where

- m is the number of rows of A,
- n is the number of columns of A, and
- $r = \text{rank}(A)$ is the number of pivot positions.

For example,

$$\dim(\text{Col}(A)) = r, \quad \dim(\text{Nul}(A)) = n - r, \quad \dim(\text{Nul}(A^T)) = m - r.$$

Thus, the dimensions of all four fundamental subspaces are completely determined by the three quantities m, n, and r. See Figure 13 and Table 6.

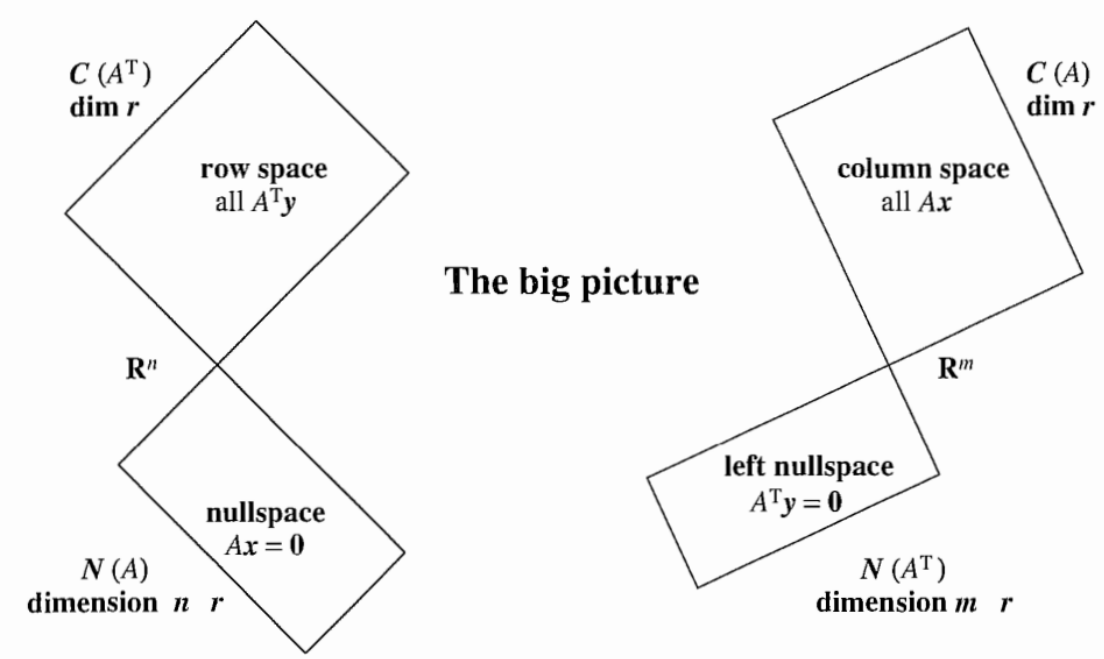


Figure 14: The Four Fundamental Subspaces. Image credit: Strang

Table 6: Summary of The Four Fundamental Subspaces

Subspace	**Lives In**	**Dimension**	**Basis**
$\text{Row}(A)$	$\mathbb{R}^n$	r	Nonzero rows of A in REF
$\text{Nul}(A)$	$\mathbb{R}^n$	$n-r$	Special solution vectors
$\text{Col}(A)$	$\mathbb{R}^m$	r	Pivot columns of A
$\text{Nul}(A^T)$	$\mathbb{R}^m$	$m-r$	Special solution vectors of $A^T\mathbf{y}=\mathbf{0}$

Problem 4.5.5

Let

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

(a) Determine the rank of $\text{rank}(A)$.

(b) Find a basis for $\text{Col}(A)$.

(c) Find a basis for $\text{Row}(A)$.

(d) Find a basis for $\text{Nul}(A)$.

(e) Find a basis for $\text{Nul}(A^T)$.

(f) State the dimensions of all four fundamental subspaces.

(g) Verify that the rank theorem is satisfied.

Solution

(a) Since the matrix has two pivot positions, $\operatorname{rank}(A) = 2$.

(b) The pivot columns are columns 1 and 3 of the original matrix, so

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \right\}$$

forms a basis for $\operatorname{Col}(A)$.

(c) The nonzero rows of the REF form a basis for the row space:

$$\left\{ \begin{bmatrix} 1 & 2 & 0 & -1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 & 1 & 2 \end{bmatrix} \right\}$$

(d) Since columns 2, 4, and 5 are free, let

$$x_2 = s, \quad x_4 = t, \quad x_5 = u.$$

Then $x_3 = -t - 2u$ and $x_1 = -2s + t - u$.

Hence,

$$\mathbf{x} = s \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + u \begin{bmatrix} -1 \\ 0 \\ -2 \\ 0 \\ 1 \end{bmatrix}.$$

Therefore,

$$\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a basis for $\text{Nul}(A)$.

(e) To find the left null space, solve

$$A^T\mathbf{y} = \mathbf{0},$$

where

$$A^T = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 3 & 2 \\ 0 & 1 & 1 \\ 3 & 8 & 5 \end{bmatrix}.$$

The system reduces to

$$y_1 = -y_2, \quad y_3 = -y_2.$$

Letting $y_2 = t$ gives

$$\mathbf{y} = t \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}.$$

Therefore,

$$\left\{ \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} \right\}$$

is a basis for $\text{Nul}(A^T)$.

(f) Since A is a 3×5 matrix with rank $r = 2$,

$$\begin{aligned} \dim(\text{Row}(A)) &= 2, \\ \dim(\text{Col}(A)) &= 2, \\ \dim(\text{Nul}(A)) &= 5 - 2 = 3, \\ \dim(\text{Nul}(A^T)) &= 3 - 2 = 1. \end{aligned}$$

(g) Since

$$\text{rank}(A) + \dim(\text{Nul}(A)) = 2 + 3 = 5,$$

the rank theorem is satisfied.

4.6 Change of Basis

When we introduced coordinate vectors, we saw that every vector can be described relative to a chosen basis. Once a basis is fixed, every vector has a unique list of coordinates. However, a vector space does not have only one possible basis. In fact, infinitely many different bases can describe the same vector space. Changing the basis does not change the vector itself. Instead, it changes the numbers used to describe that vector.

Imagine measuring the same distance using inches instead of centimeters. The physical length remains identical, but its numerical value changes because the measuring system has changed.

The same idea applies to vectors. A vector can have one coordinate vector relative to a basis $\mathcal{B}$ and a completely different coordinate vector relative to a basis $\mathcal{C}$, even though both represent exactly the same vector.

For example, suppose

$$\mathbf{x} = 3\mathbf{b}_1 + \mathbf{b}_2.$$

Then the coordinates of $\mathbf{x}$ relative to $\mathcal{B}$ are

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

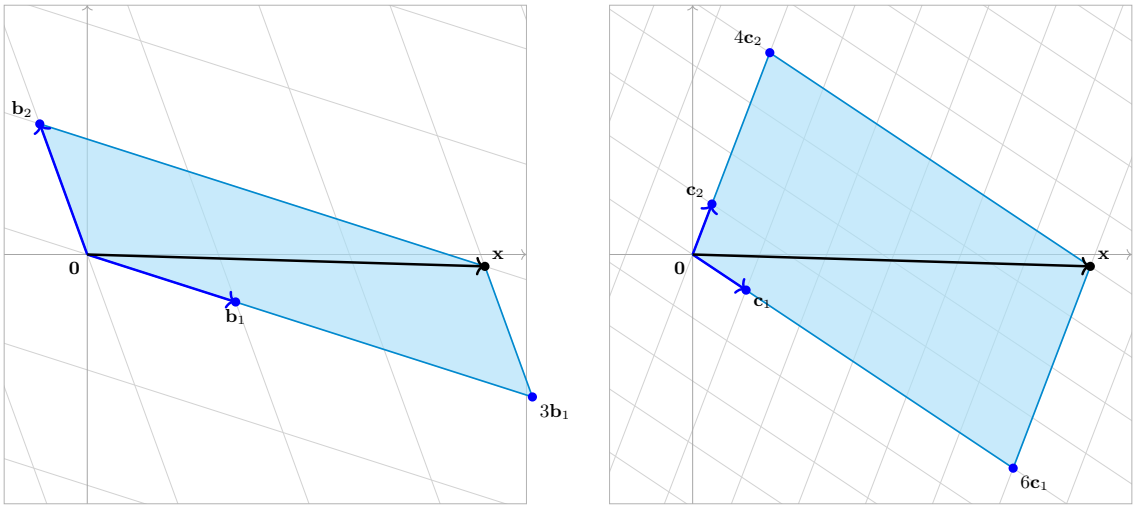
Now suppose we choose a different basis $\mathcal{C}$. The same vector might instead be written as

$$\mathbf{x} = 6\mathbf{c}_1 + 4\mathbf{c}_2,$$

so that

$$[\mathbf{x}]_{\mathcal{C}} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}.$$

Although these coordinate vectors are different, they both describe the same geometric vector. See Figure 14a for $[\mathbf{x}]_{\mathcal{B}}$ and Figure 14b for $[\mathbf{x}]_{\mathcal{C}}$.



(a) (b)

Figure 15: The same vector space described using two different coordinate systems.

We will now develop a systematic method for converting coordinate vectors from one basis to another.

Change-of-Coordinates Matrix Theorem

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$, and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ be bases for an n-dimensional vector space V. Then there exists a unique $n \times n$ matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$ such that, for every $\mathbf{x} \in V$,

$$[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}.$$

The matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$ converts the $\mathcal{B}$-coordinates of a vector into its $\mathcal{C}$-coordinates. Its columns are the $\mathcal{C}$-coordinate vectors of the basis vectors in $\mathcal{B}$:

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{bmatrix}[\mathbf{b}_1]_{\mathcal{C}} & [\mathbf{b}_2]_{\mathcal{C}} & \cdots & [\mathbf{b}_n]_{\mathcal{C}}\end{bmatrix}$$

In Figure 15, the maps $[\,\cdot\,]_{\mathcal{B}}$ and $[\,\cdot\,]_{\mathcal{C}}$ send a vector $\mathbf{x} \in V$ to its coordinate vectors $[\mathbf{x}]_{\mathcal{B}}$ and $[\mathbf{x}]_{\mathcal{C}}$ in $\mathbb{R}^n$, respectively. The change-of-coordinates matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$ converts $\mathcal{B}$-coordinates into $\mathcal{C}$-coordinates according to

$$[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}.$$

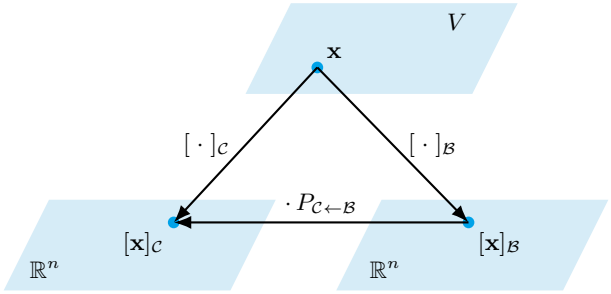


Figure 16: Two coordinate systems for the vector space V.

Example 4.6.1

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$ where

$$\mathbf{b}_1 = 4\mathbf{c}_1 + \mathbf{c}_2, \quad \mathbf{b}_2 = -6\mathbf{c}_1 + \mathbf{c}_2.$$

If

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix},$$

find $[\mathbf{x}]_{\mathcal{C}}$.

Solution

The coordinates of the basis vectors in $\mathcal{C}$ are

$$[\mathbf{b}_1]_{\mathcal{C}} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \quad [\mathbf{b}_2]_{\mathcal{C}} = \begin{bmatrix} -6 \\ 1 \end{bmatrix}.$$

Place these coordinate vectors into the columns of the change-of-coordinates matrix:

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{bmatrix} [\mathbf{b}_1]_{\mathcal{C}} & [\mathbf{b}_2]_{\mathcal{C}} \end{bmatrix} = \begin{bmatrix} 4 & -6 \\ 1 & 1 \end{bmatrix}.$$

Therefore,

$$[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 4 & -6 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}.$$

Since $P_{\mathcal{C}\leftarrow\mathcal{B}}$ is a change-of-coordinates matrix, it is invertible. Multiplying both sides of

$$[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$$

by $P^{-1}_{\mathcal{C}\leftarrow\mathcal{B}}$ gives

$$P^{-1}_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{x}]_{\mathcal{C}} = [\mathbf{x}]_{\mathcal{B}}.$$

Therefore, $P^{-1}_{\mathcal{C}\leftarrow\mathcal{B}}$ converts $\mathcal{C}$-coordinates into $\mathcal{B}$-coordinates. We define this inverse matrix as

$$P_{\mathcal{B}\leftarrow\mathcal{C}} = P^{-1}_{\mathcal{C}\leftarrow\mathcal{B}}.$$

When the target basis is the standard basis $\mathcal{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, the change-of-coordinates matrix simplifies to

$$P_{\mathcal{E}\leftarrow\mathcal{B}} = P_{\mathcal{B}} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix},$$

since the standard-coordinate vector of each basis vector is the vector itself.

Let's think about the change-of-coordinates matrix in $\mathbb{R}^n$.

Suppose $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ are bases for $\mathbb{R}^n$. Their basis matrices are

$$P_{\mathcal{B}} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix}, \quad P_{\mathcal{C}} = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{bmatrix}.$$

The basis matrix $P_{\mathcal{B}}$ converts $\mathcal{B}$-coordinates into standard coordinates:

$$\mathbf{x} = P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}} \tag{24}$$

Similarly, $P_{\mathcal{C}}^{-1}$ converts standard coordinates into $\mathcal{C}$-coordinates:

$$[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C}}^{-1}\mathbf{x} \tag{25}$$

Substituting Equation 24 into Equation 25 gives

$$[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C}}^{-1}P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}. \tag{26}$$

We have already established this from the change-of-coordinates matrix theorem:

$$[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$$

Notice how both this matrix and Equation 26 produce the same output from the same input for every coordinate vector $[\mathbf{x}]_{\mathcal{B}}$. Therefore, the matrices themselves must be equal:

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = P_{\mathcal{C}}^{-1}P_{\mathcal{B}}$$

The order of the matrices reflects the path of the conversion. We go from $\mathcal{B}$-coordinates to standard coordinates to $\mathcal{C}$-coordinates. Thus, $P_{\mathcal{B}}$ is applied first, followed by $P_{\mathcal{C}}^{-1}$.

The matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$ can also be found without computing $P_{\mathcal{C}}^{-1}$ separately. Begin with the augmented matrix

$$\left[\begin{array}{c|c} P_{\mathcal{C}} & P_{\mathcal{B}} \end{array}\right] = \left[\begin{array}{cccc|cccc} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n & \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{array}\right].$$

Row reducing $P_{\mathcal{C}}$ to I applies the same row operations to the entire augmented matrix. The product of the elementary matrices representing these row operations is $P_{\mathcal{C}}^{-1}$, since

$$P_{\mathcal{C}}^{-1}P_{\mathcal{C}} = I.$$

Therefore,

$$P_{\mathcal{C}}^{-1}\left[\begin{array}{c|c} P_{\mathcal{C}} & P_{\mathcal{B}} \end{array}\right] = \left[\begin{array}{c|c} P_{\mathcal{C}}^{-1}P_{\mathcal{C}} & P_{\mathcal{C}}^{-1}P_{\mathcal{B}} \end{array}\right].$$

Writing this transformation column by column gives

$$\begin{aligned} &P_{\mathcal{C}}^{-1}\left[\begin{array}{cccc|cccc} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n & \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{array}\right] \\ =&\left[\begin{array}{cccc|cccc} P_{\mathcal{C}}^{-1}\mathbf{c}_1 & P_{\mathcal{C}}^{-1}\mathbf{c}_2 & \cdots & P_{\mathcal{C}}^{-1}\mathbf{c}_n & P_{\mathcal{C}}^{-1}\mathbf{b}_1 & P_{\mathcal{C}}^{-1}\mathbf{b}_2 & \cdots & P_{\mathcal{C}}^{-1}\mathbf{b}_n \end{array}\right]. \end{aligned}$$

Since $P_{\mathcal{C}}^{-1}$ converts standard coordinates into $\mathcal{C}$-coordinates,

$$P_{\mathcal{C}}^{-1}\mathbf{c}_1 = \mathbf{e}_1, \quad P_{\mathcal{C}}^{-1}\mathbf{c}_2 = \mathbf{e}_2, \quad \cdots, \quad P_{\mathcal{C}}^{-1}\mathbf{c}_n = \mathbf{e}_n,$$

and

$$P_{\mathcal{C}}^{-1}\mathbf{b}_1 = [\mathbf{b}_1]_{\mathcal{C}}, \quad P_{\mathcal{C}}^{-1}\mathbf{b}_2 = [\mathbf{b}_2]_{\mathcal{C}}, \quad \cdots, \quad P_{\mathcal{C}}^{-1}\mathbf{b}_n = [\mathbf{b}_n]_{\mathcal{C}}.$$

Thus, the resulting augmented matrix is

$$\left[\begin{array}{cccc|cccc} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n & [\mathbf{b}_1]_{\mathcal{C}} & [\mathbf{b}_2]_{\mathcal{C}} & \cdots & [\mathbf{b}_n]_{\mathcal{C}} \end{array}\right] = \left[\begin{array}{c|c} I & P_{\mathcal{C}\leftarrow\mathcal{B}} \end{array}\right].$$

Therefore,

$$\left[\begin{array}{c|c} P_{\mathcal{C}} & P_{\mathcal{B}} \end{array}\right] \sim \left[\begin{array}{c|c} I & P_{\mathcal{C}\leftarrow\mathcal{B}} \end{array}\right].$$

To find $P_{\mathcal{C}\leftarrow\mathcal{B}}$, place the target basis $\mathcal{C}$ on the left and the source basis $\mathcal{B}$ on the right. After row reduction, the right-hand block is the desired change-of-coordinates matrix:

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = \Big[[\mathbf{b}_1]_{\mathcal{C}} \quad [\mathbf{b}_2]_{\mathcal{C}} \quad \cdots \quad [\mathbf{b}_n]_{\mathcal{C}}\Big]$$

A useful way to remember the construction is

$$\left[\begin{array}{c|c} \text{target basis} & \text{source basis} \end{array}\right] \sim \left[\begin{array}{c|c} I & \text{change-of-coordinates matrix} \end{array}\right].$$

Example 4.6.2

Let

$$\mathbf{b}_1 = \begin{bmatrix} -9 \\ 1 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} -5 \\ -1 \end{bmatrix}, \quad \mathbf{c}_1 = \begin{bmatrix} 1 \\ -4 \end{bmatrix}, \quad \mathbf{c}_2 = \begin{bmatrix} 3 \\ -5 \end{bmatrix}.$$

The sets $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$ are bases for $\mathbb{R}^2$. Find the change-of-coordinates matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$.

Solution

Because the conversion is from $\mathcal{B}$-coordinates to $\mathcal{C}$-coordinates, place the target basis $\mathcal{C}$ on the left and the source basis $\mathcal{B}$ on the right:

$$\left[\begin{array}{c|c} P_{\mathcal{C}} & P_{\mathcal{B}} \end{array}\right] = \left[\begin{array}{cc|cc} 1 & 3 & -9 & -5 \\ -4 & -5 & 1 & -1 \end{array}\right]$$

Row reduce the left block to I:

$$\left[\begin{array}{cc|cc} 1 & 3 & -9 & -5 \\ -4 & -5 & 1 & -1 \end{array}\right] \sim \left[\begin{array}{cc|cc} 1 & 0 & 6 & 4 \\ 0 & 1 & -5 & -3 \end{array}\right]$$

The right-hand block is therefore the change-of-coordinates matrix:

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{bmatrix} 6 & 4 \\ -5 & -3 \end{bmatrix}$$

where its columns are the $\mathcal{C}$-coordinate vectors of $\mathbf{b}_1$ and $\mathbf{b}_2$:

$$[\mathbf{b}_1]_{\mathcal{C}} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}, \quad [\mathbf{b}_2]_{\mathcal{C}} = \begin{bmatrix} 4 \\ -3 \end{bmatrix}$$

Example 4.6.3

Let

$$\mathbf{b}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} -2 \\ 4 \end{bmatrix}, \quad \mathbf{c}_1 = \begin{bmatrix} -7 \\ 9 \end{bmatrix}, \quad \mathbf{c}_2 = \begin{bmatrix} -5 \\ 7 \end{bmatrix}.$$

The sets $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$ are bases for $\mathbb{R}^2$.

(a) Find $P_{\mathcal{B}\leftarrow\mathcal{C}}$.

(b) Use part (a) to find $P_{\mathcal{C}\leftarrow\mathcal{B}}$.

Solution

(a) Because the conversion is from $\mathcal{C}$-coordinates to $\mathcal{B}$-coordinates, place the target basis $\mathcal{B}$ on the left and the source basis $\mathcal{C}$ on the right:

$$\left[\begin{array}{c|c} P_{\mathcal{B}} & P_{\mathcal{C}} \end{array}\right] = \left[\begin{array}{cc|cc} 1 & -2 & -7 & -5 \\ -3 & 4 & 9 & 7 \end{array}\right]$$

Row reducing the left block to I gives

$$\left[\begin{array}{cc|cc} 1 & -2 & -7 & -5 \\ -3 & 4 & 9 & 7 \end{array}\right] \sim \left[\begin{array}{cc|cc} 1 & 0 & 5 & 3 \\ 0 & 1 & 6 & 4 \end{array}\right].$$

Therefore,

$$P_{\mathcal{B}\leftarrow\mathcal{C}} = \begin{bmatrix} 5 & 3 \\ 6 & 4 \end{bmatrix}.$$

(b) The reverse change-of-coordinates matrix is the inverse:

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = P_{\mathcal{B}\leftarrow\mathcal{C}}^{-1}$$

Thus,

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{bmatrix} 5 & 3 \\ 6 & 4 \end{bmatrix}^{-1} = \frac{1}{2}\begin{bmatrix} 4 & -3 \\ -6 & 5 \end{bmatrix} = \begin{bmatrix} 2 & -\frac{3}{2} \\ -3 & \frac{5}{2} \end{bmatrix}.$$

See Table 7 for a summary of the coordinate conversion matrices.

Conversion	Matrix	Formula	Valid In
$\mathcal{B}$-coordinates to standard coordinates	$P_{\mathcal{B}}$	$\mathbf{x} = P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$	$\mathbb{R}^n$
Standard coordinates to $\mathcal{B}$-coordinates	$P_{\mathcal{B}}^{-1}$	$[\mathbf{x}]_{\mathcal{B}} = P_{\mathcal{B}}^{-1}\mathbf{x}$	$\mathbb{R}^n$
$\mathcal{B}$-coordinates to $\mathcal{C}$-coordinates	$P_{\mathcal{C}\leftarrow\mathcal{B}} = P_{\mathcal{B}\leftarrow\mathcal{C}}^{-1}$	$[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$	V
$\mathcal{C}$-coordinates to $\mathcal{B}$-coordinates	$P_{\mathcal{B}\leftarrow\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}^{-1}$	$[\mathbf{x}]_{\mathcal{B}} = P_{\mathcal{B}\leftarrow\mathcal{C}}[\mathbf{x}]_{\mathcal{C}}$	V
$\mathcal{B}$-coordinates to $\mathcal{C}$-coordinates	$P_{\mathcal{C}}^{-1}P_{\mathcal{B}}$	$[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C}}^{-1}P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$	$\mathbb{R}^n$
$\mathcal{C}$-coordinates to $\mathcal{B}$-coordinates	$P_{\mathcal{B}}^{-1}P_{\mathcal{C}}$	$[\mathbf{x}]_{\mathcal{B}} = P_{\mathcal{B}}^{-1}P_{\mathcal{C}}[\mathbf{x}]_{\mathcal{C}}$	$\mathbb{R}^n$

Table 7: Assume V is a finite-dimensional vector space.

Problem 4.6.1

Let $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2\}$ and $\mathcal{G} = \{\mathbf{g}_1, \mathbf{g}_2\}$ be bases for a vector space V, and let P be the matrix whose columns are $[\mathbf{f}_1]_{\mathcal{G}}$ and $[\mathbf{f}_2]_{\mathcal{G}}$.

Which of the following equations is satisfied by P for every $\mathbf{v} \in V$?

(a) $[\mathbf{v}]_{\mathcal{F}} = P[\mathbf{v}]_{\mathcal{G}}$

(b) $[\mathbf{v}]_{\mathcal{G}} = P[\mathbf{v}]_{\mathcal{F}}$

Solution

Since the columns of P are $[\mathbf{f}_1]_{\mathcal{G}}$ and $[\mathbf{f}_2]_{\mathcal{G}}$, the matrix P is the change-of-coordinates matrix from the basis $\mathcal{F}$ to the basis $\mathcal{G}$. Therefore, multiplying a coordinate vector expressed relative to $\mathcal{F}$ by P produces the corresponding coordinate vector relative to $\mathcal{G}$. The correct answer is (b).

Problem 4.6.2

Determine whether each statement is true or false. Justify your answer.

(a) The columns of the change-of-coordinates matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$ are linearly independent.

(b) Let $V = \mathbb{R}^2$, $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$, and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$. Suppose row reduction transforms

$$\left[\begin{array}{cc|cc} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{b}_1 & \mathbf{b}_2 \end{array}\right]$$

into

$$\left[\begin{array}{c|c} I & P \end{array}\right].$$

Then $[\mathbf{x}]_{\mathcal{B}} = P[\mathbf{x}]_{\mathcal{C}}$ for every $\mathbf{x} \in V$.

Solution

(a) True.

The change-of-coordinates matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible because every basis vector in $\mathcal{B}$ has a unique coordinate vector relative to $\mathcal{C}$. An invertible matrix has linearly independent columns, so the columns of $P_{\mathcal{C}\leftarrow\mathcal{B}}$ are linearly independent.

(b) False.

Row reduction of

$$\left[\begin{array}{c|c} P_{\mathcal{C}} & P_{\mathcal{B}} \end{array}\right]$$

produces

$$\left[\begin{array}{c|c} I & P_{\mathcal{C}}^{-1}P_{\mathcal{B}} \end{array}\right] = \left[\begin{array}{c|c} I & P_{\mathcal{C}\leftarrow\mathcal{B}} \end{array}\right].$$

Thus, the resulting matrix P converts $\mathcal{B}$-coordinates into $\mathcal{C}$-coordinates, so

$$[\mathbf{x}]_{\mathcal{C}} = P[\mathbf{x}]_{\mathcal{B}}.$$

5 Eigenvalues and Eigenvectors

5.1 Foundations of Eigenvalues and Eigenvectors

Up to this point, we have studied the structure of vector spaces and the matrices that act on them. We have learned how to describe vectors using different bases, determine the dimension of a space, and understand how linear transformations act throughout that space.

Some transformations possess special directions that are preserved under their action. A nonzero vector lying along one of these directions is mapped to a scalar multiple of itself. Such vectors are called **eigenvectors**, and their corresponding scale factors are called **eigenvalues**.

While their definition might seem simple, eigenvalues and eigenvectors reveal the underlying structure of a linear transformation. Rather than describing how a matrix acts on arbitrary vectors, they identify its invariant directions and describe its behavior along those directions. This perspective forms the foundation of **spectral theory**, one of the central themes of modern linear algebra.

Spectral theory has profound applications throughout mathematics, science, and engineering. Eigenvalues and eigenvectors appear in differential equations, dynamical systems, mechanical vibrations, spectroscopy, quantum mechanics, data science, LLM training, and many other fields. They are among the most important and widely used ideas in linear algebra.

In this chapter, we will develop the theory of eigenvalues and eigenvectors, learn how to compute them, and uncover how they are used.

Consider a multistory building disturbed by strong winds or an earthquake. Each floor may move horizontally by a different amount, so the displacement of an n-story building at any instant can be represented by the vector

$$\mathbf{r}(t) = \begin{bmatrix} r_1(t) \\ r_2(t) \\ \vdots \\ r_n(t) \end{bmatrix},$$

where $r_i(t)$ denotes the horizontal displacement of the i-th floor.

When the structure vibrates freely, certain deformation patterns arise naturally. In one pattern, the floors may sway together in the same direction. In another, the upper floors may move in the opposite direction from the lower floors. These characteristic patterns are called **mode shapes**.

For an undamped (continuous motion without energy loss) structure with no external force, the motion is governed by

$$M\ddot{\mathbf{r}}(t) + K\mathbf{r}(t) = \mathbf{0},$$

where M is the mass matrix of each floor and K is the stiffness matrix determined by the stiffness of the columns between adjacent floors. Suppose that the building vibrates harmonically with angular frequency ω while maintaining a fixed deformation pattern $\boldsymbol{\phi}$:

$$\begin{aligned} \mathbf{r}(t) &= \boldsymbol{\phi}\sin(\omega t) \\ \ddot{\mathbf{r}}(t) &= -\omega^2\boldsymbol{\phi}\sin(\omega t) \end{aligned}$$

Substituting these expressions into the governing equation gives

$$-\omega^2 M\boldsymbol{\phi}\sin(\omega t) + K\boldsymbol{\phi}\sin(\omega t) = \mathbf{0},$$

and therefore

$$K\boldsymbol{\phi} = \omega^2 M\boldsymbol{\phi} \Rightarrow \left(K - \omega^2 M\right)\boldsymbol{\phi} = \mathbf{0}.$$

When M is invertible, this may be rewritten as

$$M^{-1}K\boldsymbol{\phi} = \omega^2\boldsymbol{\phi}.$$

This is an eigenvalue equation. The mode shape $\boldsymbol{\phi}$ is an **eigenvector** of $M^{-1}K$, while the corresponding eigenvalue

$$\lambda = \omega^2$$

is the square of the natural angular frequency of the structure.

The entries of $\boldsymbol{\phi}$ describe the relative motion of the floors. If one floor moves twice as far as another, that ratio remains fixed while the building oscillates in that mode. Multiplying $\boldsymbol{\phi}$ by any nonzero scalar changes only the overall amplitude, not the mode shape itself.

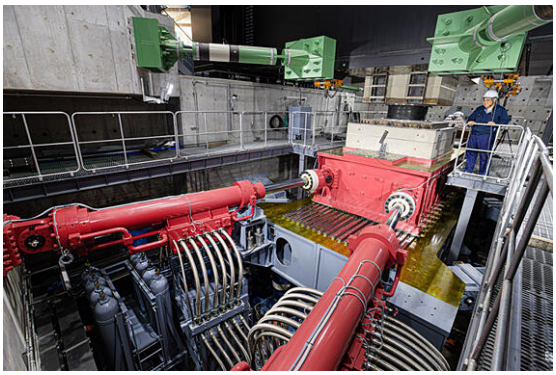
A general vibration of the building can be expressed as a combination of its natural modes:

$$\mathbf{r}(t) = \boldsymbol{\phi}_1 q_1(t) + \boldsymbol{\phi}_2 q_2(t) + \cdots + \boldsymbol{\phi}_n q_n(t)$$

Thus, the eigenvectors provide a coordinate system naturally adapted to the structure. Rather than describing the motion using the displacement of each individual floor, we may describe it by how much of each natural mode is present. In these modal coordinates, a complicated coupled system can be separated into simpler independent motions.

In the real world, engineers use eigenvalues and eigenvectors to study how structures respond to earthquakes, wind, and other dynamic loads. At two of the world's largest seismic testing facilities, UC San Diego's outdoor shake table and the National Research Institute for Earth Science and Disaster Resilience's E-Defense facility in Japan, full-scale structures are subjected to controlled ground motions. This allows civil engineers to identify their natural frequencies and characteristic vibration patterns.

(a) A full-scale 10-story building constructed for seismic testing at the UC San Diego outdoor shake table. Image credit: David Baillot, UC San Diego Jacobs School of Engineering



(b) The seismic isolation components beneath a large-scale indoor shake table at NIED's E-Defense facility in Japan. Image credit: Japan Science and Technology Agency

Eigenvectors and Eigenvalues

Let A be an $n \times n$ matrix. A nonzero vector $\mathbf{x}$ is an **eigenvector** of A if there exists a scalar λ such that

$$A\mathbf{x} = \lambda\mathbf{x}.$$

The scalar λ is called the **eigenvalue** corresponding to $\mathbf{x}$. Equivalently, λ is an eigenvalue of A if $A\mathbf{x} = \lambda\mathbf{x}$has a nonzero solution $\mathbf{x}$.

Consider

$$A = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Multiplying $\mathbf{v}$ by A gives

$$A\mathbf{v} = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 2\mathbf{v}.$$

Thus, $\mathbf{v}$ is an eigenvector of A corresponding to the eigenvalue $\lambda = 2$.

In contrast,

$$A\mathbf{u} = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -5 \\ -1 \end{bmatrix},$$

which is not a scalar multiple of $\mathbf{u}$. Therefore, $\mathbf{u}$ is not an eigenvector of A.

Because $A\mathbf{v} = 2\mathbf{v}$, multiplication by A preserves the direction of $\mathbf{v}$ and doubles its length. Geometrically, the transformation stretches $\mathbf{v}$ by a factor of 2 along the line that it spans. This preserved direction is what makes $\mathbf{v}$ an eigenvector, while the stretching factor 2 is its corresponding eigenvalue.

Example 5.1.1

Let

$$A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}.$$

Determine whether $\mathbf{u}$ and $\mathbf{v}$ are eigenvectors of A.

Solution

For $\mathbf{u}$,

$$A\mathbf{u} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ -5 \end{bmatrix} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = -4 \begin{bmatrix} 6 \\ -5 \end{bmatrix} = -4\mathbf{u}.$$

Thus, $\mathbf{u}$ is an eigenvector of A corresponding to the eigenvalue $\lambda = -4$.

For $\mathbf{v}$,

$$A\mathbf{v} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 11 \end{bmatrix}.$$

Since

$$\begin{bmatrix} -9 \\ 11 \end{bmatrix} \neq \lambda \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

for any scalar λ, $\mathbf{v}$ is not an eigenvector of A.

Example 5.1.2

Let

$$A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}.$$

Show that $\lambda = 7$ is an eigenvalue of A, and find all corresponding eigenvectors.

Solution

The scalar $\lambda = 7$ is an eigenvalue of A if the equation

$$A\mathbf{x} = 7\mathbf{x}$$

has a nonzero solution. Rearranging gives

$$A\mathbf{x} - 7\mathbf{x} = \mathbf{0} \Rightarrow (A - 7I)\mathbf{x} = \mathbf{0}.$$

Now,

$$A - 7I = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} - \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} -6 & 6 \\ 5 & -5 \end{bmatrix}.$$

Solving the homogeneous system,

$$\begin{bmatrix} -6 & 6 & 0 \\ 5 & -5 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus $x_1 - x_2 = 0$, so $x_1 = x_2$. Letting $x_2 = t$, the solution is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = t \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Because this equation has nonzero solutions, $\lambda = 7$ is an eigenvalue of A. The corresponding eigenvectors are all nonzero vectors of the form

$$\mathbf{x} = t \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

where $t \neq 0$.

Example 5.1.2 illustrates an important principle.

Eigenvalue Criterion

A scalar λ is an eigenvalue of an $n \times n$ matrix A if and only if

$$(A - \lambda I)\mathbf{x} = \mathbf{0}$$

has a nonzero solution. The set of all solutions is the null space of $A - \lambda I$.

Eigenspace

Let λ be an eigenvalue of an $n \times n$ matrix A. The **eigenspace** corresponding to λ is

$$E_\lambda = \text{Nul}(A - \lambda I).$$

The eigenspace consists of the zero vector together with all eigenvectors corresponding to λ.

To find a basis for E_λ, solve

$$(A - \lambda I)\mathbf{x} = \mathbf{0}$$

and write the solution in parametric vector form. The vectors multiplying the free variables form a basis for E_λ, and the number of free variables equals $\dim(E_\lambda)$.

Keep in mind that row reduction may be used to solve

$$(A - \lambda I)\mathbf{x} = \mathbf{0},$$

and thereby find the eigenvectors corresponding to a known eigenvalue λ. However, row reducing A itself generally changes its eigenvalues. Therefore, the eigenvalues of A usually **cannot** be determined from an REF or RREF of A.

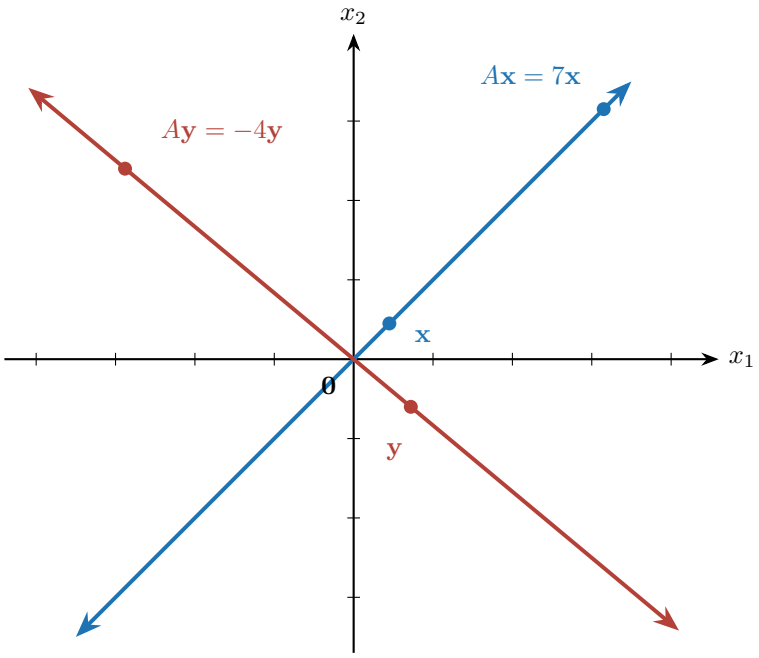


Figure 18: The eigenspaces corresponding to $\lambda = 7$ and $\lambda = -4$. Along E_7 (pictured in blue), vectors are stretched without changing direction. Along E_{-4} (pictured in red), vectors are stretched and reversed.

Example 5.1.3

Let

$$A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}.$$

Given that $\lambda = 2$ is an eigenvalue of A, find a basis for the corresponding eigenspace. Then, interpret the result geometrically.

Solution

The eigenspace corresponding to $\lambda = 2$ is $E_2 = \text{Nul}(A - 2I)$.

First,

$$A - 2I = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix} - \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 6 \end{bmatrix}.$$

Thus, we solve

$$(A - 2I)\mathbf{x} = \mathbf{0}.$$

Row reduce:

$$\begin{bmatrix} 2 & -1 & 6 & 0 \\ 2 & -1 & 6 & 0 \\ 2 & -1 & 6 & 0 \end{bmatrix} \sim \begin{bmatrix} 2 & -1 & 6 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The resulting equation is

$$2x_1 - x_2 + 6x_3 = 0 \Rightarrow x_1 = \frac{1}{2}x_2 - 3x_3.$$

Since x_2 and x_3 are free variables,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} \frac{1}{2} \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}.$$

Therefore,

$$E_2 = \text{span}\left\{\begin{bmatrix}1\\2\\0\end{bmatrix}, \begin{bmatrix}-3\\0\\1\end{bmatrix}\right\}.$$

A basis for the eigenspace corresponding to $\lambda = 2$ is

$$\left\{\begin{bmatrix}1\\2\\0\end{bmatrix}, \begin{bmatrix}-3\\0\\1\end{bmatrix}\right\}.$$

Every vector $\mathbf{x} \in E_2$ satisfies

$$A\mathbf{x} = 2\mathbf{x}.$$

Thus, every vector in the eigenspace is stretched by a factor of 2 without changing direction. In other words, when restricted to the eigenspace E_2, the linear transformation A acts as a dilation by a factor of 2.

Example 5.1.4

Consider the projection matrix

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

Determine the eigenvalues and eigenspaces of A without performing any row reduction.

Solution

The matrix projects every vector onto the x_1-axis.

Suppose first that

$$\mathbf{x} = \begin{bmatrix}a\\0\end{bmatrix},$$

so that $\mathbf{x}$ already lies on the x_1-axis. Then

$$A\mathbf{x} = \begin{bmatrix} a \\ 0 \end{bmatrix} = \mathbf{x}.$$

Thus every nonzero vector on the x_1-axis is an eigenvector with eigenvalue $\lambda = 1$.

Next suppose

$$\mathbf{y} = \begin{bmatrix} 0 \\ b \end{bmatrix},$$

which lies on the x_2-axis. Applying the projection gives

$$A\mathbf{y} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0\mathbf{y}.$$

Therefore every nonzero vector on the x_2-axis is an eigenvector with eigenvalue $\lambda = 0$.

Hence,

$$E_1 = \text{span}\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\}, \quad E_0 = \text{span}\left\{\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}.$$

Geometrically, vectors on the x_1-axis remain fixed, while vectors on the x_2-axis collapse to the origin. Every other vector changes direction and is therefore not an eigenvector.

Example 5.1.5

Consider the shear matrix

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

Determine the eigenvalues and eigenspaces of A without performing any row reduction.

Solution

The matrix shears the plane horizontally. Points are shifted parallel to the x_1-axis while their vertical coordinates remain unchanged.

First, consider a vector on the x_1-axis

$$\mathbf{x} = \begin{bmatrix} a \\ 0 \end{bmatrix}$$

Then

$$A\mathbf{x} = \begin{bmatrix} a \\ 0 \end{bmatrix} = \mathbf{x}.$$

Thus every nonzero vector on the x_1-axis is an eigenvector with eigenvalue $\lambda = 1$.

Now consider a vector that is not on the x_1-axis,

$$\mathbf{y} = \begin{bmatrix} a \\ b \end{bmatrix},$$

where $b \neq 0$.

Then

$$A\mathbf{y} = \begin{bmatrix} a+b \\ b \end{bmatrix}.$$

The transformed vector no longer lies on the same line through the origin as $\mathbf{y}$, so it is not a scalar multiple of $\mathbf{y}$. Therefore $\mathbf{y}$ is not an eigenvector.

Hence,

$$E_1 = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}.$$

Geometrically, the shear preserves only one direction: the x_1-axis. Every other vector changes direction, so the shear has only one eigenspace.

Example 5.1.6

Consider the rotation matrix

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

which rotates every vector by 90° counterclockwise. Determine the real eigenvalues and eigenspaces of A without performing any row reduction.

Solution

Suppose $\mathbf{x}$ is any nonzero vector in $\mathbb{R}^2$. Multiplication by A rotates $\mathbf{x}$ by 90°, so $A\mathbf{x}$ is always perpendicular to $\mathbf{x}$.

Since an eigenvector must satisfy $A\mathbf{x} = \lambda\mathbf{x}$, the vectors $A\mathbf{x}$ and $\mathbf{x}$ must lie on the same line through the origin. However, a 90° rotation changes the direction of every nonzero vector, so no nonzero vector is mapped to a scalar multiple of itself.

As a result, A has no *real* eigenvectors.

Since every real eigenvalue must have a corresponding real eigenvector, it follows that A has no *real* eigenvalues either.

Although eigenvectors are defined individually, they become most powerful when grouped together. In particular, eigenvectors corresponding to distinct eigenvalues possess a remarkable property.

Linear Independence of Eigenvectors

If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ are eigenvectors corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_r$ of an $n \times n$ matrix A, then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a linearly independent set.

Problem 5.1.1

Determine whether

$$\mathbf{v} = \begin{bmatrix} -1+\sqrt{2} \\ 1 \end{bmatrix}$$

is an eigenvector of

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}.$$

If it is, determine the corresponding eigenvalue.

Solution

To determine whether $\mathbf{v}$ is an eigenvector of A, we compute

$$A\mathbf{v} = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} -1+\sqrt{2} \\ 1 \end{bmatrix} = \begin{bmatrix} 2(-1+\sqrt{2})+1 \\ (-1+\sqrt{2})+4 \end{bmatrix} = \begin{bmatrix} -1+2\sqrt{2} \\ 3+\sqrt{2} \end{bmatrix}.$$

To check whether this is a scalar multiple of $\mathbf{v}$, observe that

$$(3+\sqrt{2})\mathbf{v} = (3+\sqrt{2}) \begin{bmatrix} -1+\sqrt{2} \\ 1 \end{bmatrix} = \begin{bmatrix} (3+\sqrt{2})(-1+\sqrt{2}) \\ 3+\sqrt{2} \end{bmatrix}.$$

Since $(3+\sqrt{2})(-1+\sqrt{2}) = -3+3\sqrt{2}-\sqrt{2}+2 = -1+2\sqrt{2}$, we have

$$A\mathbf{v} = (3+\sqrt{2})\mathbf{v}.$$

Thus, $\mathbf{v}$ is an eigenvector of A corresponding to the eigenvalue

$$\lambda = 3+\sqrt{2}.$$

Problem 5.1.2

Determine whether each statement is true or false for an $n \times n$ matrix A. Justify each answer.

(a) If $A\mathbf{x} = \lambda\mathbf{x}$ for some vector $\mathbf{x}$, then λ is an eigenvalue of A.

(b) A is not invertible if and only if 0 is an eigenvalue of A.

(c) A number c is an eigenvalue of A if and only if $(A-cI)\mathbf{x} = \mathbf{0}$ has a nontrivial solution.

(d) Finding an eigenvector of A is much more difficult than checking whether a given vector is an eigenvector.

(e) To find the eigenvalues of A, reduce A to REF.

Solution

(a) False. $\mathbf{x}$ must be nonzero. For any scalar λ,

$$A\mathbf{0} = \mathbf{0} = \lambda\mathbf{0},$$

but the zero vector is not an eigenvector.

(b) True. The number 0 is an eigenvalue of A if and only if there exists a nonzero vector $\mathbf{x}$ such that

$$A\mathbf{x} = 0\mathbf{x} = \mathbf{0}.$$

This means that $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution, which occurs if and only if A is not invertible.

(c) True. The equation $A\mathbf{x} = c\mathbf{x}$ can be rewritten as

$$A\mathbf{x} - cI\mathbf{x} = \mathbf{0} \Rightarrow (A - cI)\mathbf{x} = \mathbf{0}.$$

Thus, c is an eigenvalue of A when this equation has a nontrivial solution.

(d) True. To check whether a given nonzero vector $\mathbf{x}$ is an eigenvector, compute $A\mathbf{x}$ and determine whether the result is a scalar multiple of $\mathbf{x}$.

(e) False. Row operations generally change the eigenvalues of a matrix. Therefore, reducing A to REF does not preserve the eigenvalues of A.

5.2 The Characteristic Equation

So far, we have found the eigenvectors corresponding to a known eigenvalue by solving

$$(A - \lambda I)\mathbf{x} = \mathbf{0}.$$

We will now develop a general method for determining the eigenvalues themselves. The key is to identify the values of λ for which this homogeneous system has a nonzero solution.

The Characteristic Equation

Let A be an $n \times n$ matrix. A scalar λ is an eigenvalue of A if and only if

$$\det(A - \lambda I) = 0.$$

This equation is called the **characteristic equation** of A. In other words, the solutions of the characteristic equation are the eigenvalues of A.

Recall the following result from Section 3.2.

Let A be an $n \times n$ matrix, and let U be an REF of A obtained using only row operations. Then,

$$\det(A) = \begin{cases} (-1)^r \prod_{i=1}^{n} u_{i,i}, & \text{when } A \text{ is invertible,} \\ 0, & \text{when } A \text{ is not invertible,} \end{cases}$$

where $\prod_{i=1}^{n} u_{i,i}$ is simply the product of the pivots of U.

We have yet another addition to the invertible matrix theorem.

The Invertible Matrix Theorem (continued)

Let A be an $n \times n$ matrix. Then A is invertible if and only if

(s) 0 is not an eigenvalue of A.

(t) $\det(A) \neq 0$.

Example 5.2.1

Find the eigenvalues of

$$A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}.$$

Solution

By the characteristic equation, $\det(A - \lambda I) = 0$.

First,

$$A - \lambda I = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix}.$$

A square matrix is not invertible if and only if its determinant is zero. Therefore, the eigenvalues of A satisfy $\det(A - \lambda I) = 0$.

We have

$$\det \begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix} = 0.$$

Expanding the determinant gives

$$\begin{aligned} (2-\lambda)(-6-\lambda) - (3)(3) &= 0, \\ \lambda^2 + 4\lambda - 21 &= 0, \\ (\lambda - 3)(\lambda + 7) &= 0. \end{aligned}$$

Thus, $\lambda = 3$ and $\lambda = -7$.

The Characteristic Polynomial

Let A be an $n \times n$ matrix. The n-th order polynomial

$$p_A(\lambda) = \det(A - \lambda I)$$

is called the **characteristic polynomial** of A. Its roots are the eigenvalues of A.

Before developing a general method for computing eigenvalues, we present an important special case. If a matrix is triangular, its eigenvalues can be read directly from its main diagonal.

Eigenvalues of a Triangular Matrix

The eigenvalues of a triangular matrix are precisely the entries on its main diagonal.

Proof

Suppose A is an upper triangular matrix. Then

$$A - \lambda I = \begin{bmatrix} a_{1,1} - \lambda & * & \cdots & * \\ 0 & a_{2,2} - \lambda & \cdots & * \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{n,n} - \lambda \end{bmatrix},$$

which is also upper triangular. Since the determinant of a triangular matrix is the product of its diagonal entries,

$$\det(A - \lambda I) = (a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda).$$

Therefore, the characteristic equation holds if and only if

$$\lambda = a_{1,1},\ a_{2,2},\ \ldots,\ a_{n,n}.$$

Hence, the eigenvalues of A are exactly the entries on its main diagonal. □

Example 5.2.2

Find the characteristic equation and eigenvalues of

$$A = \begin{bmatrix} 5 & -2 & 6 & -1 \\ 0 & 3 & -8 & 0 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Solution

Since A is upper triangular, $A - \lambda I$ is also upper triangular:

$$A - \lambda I = \begin{bmatrix} 5-\lambda & -2 & 6 & -1 \\ 0 & 3-\lambda & -8 & 0 \\ 0 & 0 & 5-\lambda & 4 \\ 0 & 0 & 0 & 1-\lambda \end{bmatrix}$$

The determinant of a triangular matrix is the product of its diagonal entries. Therefore,

$$\begin{aligned} p_A(\lambda) &= \det(A - \lambda I) \\ &= (5-\lambda)(3-\lambda)(5-\lambda)(1-\lambda) \\ &= (5-\lambda)^2(3-\lambda)(1-\lambda). \end{aligned}$$

Thus, the characteristic equation is

$$(5-\lambda)^2(3-\lambda)(1-\lambda) = 0 \Rightarrow (\lambda-5)^2(\lambda-3)(\lambda-1) = 0.$$

Therefore, the eigenvalues of A are

$$\lambda = 5, \quad \lambda = 3, \quad \lambda = 1.$$

Example 5.2.2 brings up an important idea.

Algebraic Multiplicity

Let A be an $n \times n$ matrix. If the characteristic polynomial can be written as

$$p_A(\lambda) = (\lambda - \lambda_0)^m q(\lambda),$$

where $q(\lambda_0) \neq 0$, then λ_0 is an eigenvalue of A with **algebraic multiplicity** m. In other words, the algebraic multiplicity of an eigenvalue is the number of times it occurs as a root of the characteristic polynomial.

Trace

Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ listed according to their algebraic multiplicities. The **trace** of A, denoted $\operatorname{tr}(A)$, is the sum of its diagonal entries.

Since the eigenvalues satisfy

$$\operatorname{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n,$$

the trace is the sum of the eigenvalues. Thus,

$$\operatorname{tr}(A) = \sum_{i=1}^{n} \lambda_i = a_{1,1} + a_{2,2} + \cdots + a_{n,n}.$$

Example 5.2.3

Determine the algebraic multiplicity of the eigenvalues in Example 5.2.2.

Solution

Since the factor $(\lambda - 5)$ appears twice, the eigenvalue $\lambda = 5$ has algebraic multiplicity 2. The remaining eigenvalues each appear only once, so $\lambda = 3$ and $\lambda = 1$ both have algebraic multiplicity one.

Example 5.2.4

Find the eigenvalues of

$$A = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix}.$$

and find a basis for each corresponding eigenspace.

Solution

$$A - \lambda I = \begin{bmatrix} 5 - \lambda & 2 \\ 2 & 1 - \lambda \end{bmatrix}.$$

Therefore,

$$\begin{aligned}\det(A-\lambda I) &= (5-\lambda)(1-\lambda)-(2)(2)\\ &= \lambda^2-6\lambda+1.\end{aligned}$$

The eigenvalues must satisfy $\lambda^2-6\lambda+1=0$. Using the quadratic formula,

$$\begin{aligned}\lambda &= \frac{6\pm\sqrt{(-6)^2-4(1)(1)}}{2}\\ &= \frac{6\pm\sqrt{32}}{2}\\ &= 3\pm 2\sqrt{2}.\end{aligned}$$

Thus, the eigenvalues are

$$\lambda_1 = 3+2\sqrt{2}, \quad \lambda_2 = 3-2\sqrt{2}.$$

To find the eigenspace corresponding to $\lambda_1 = 3+2\sqrt{2}$, solve

$$(A-\lambda_1 I)\mathbf{x} = \mathbf{0}.$$

We have

$$A-\lambda_1 I = \begin{bmatrix} 2-2\sqrt{2} & 2 \\ 2 & -2-2\sqrt{2}\end{bmatrix}.$$

The first equation of the homogeneous system is

$$\begin{gathered}(2-2\sqrt{2})x_1+2x_2=0\\ (1-\sqrt{2})x_1+x_2=0 \Rightarrow x_1=(1+\sqrt{2})x_2.\end{gathered}$$

Letting $x_2=t$, we obtain

$$\mathbf{x} = t\begin{bmatrix}1+\sqrt{2}\\ 1\end{bmatrix}.$$

Therefore,

$$E_{\lambda_1} = \text{Span}\left\{\begin{bmatrix} 1+\sqrt{2} \\ 1 \end{bmatrix}\right\}.$$

For $\lambda_2 = 3 - 2\sqrt{2}$,

$$A - \lambda_2 I = \begin{bmatrix} 2+2\sqrt{2} & 2 \\ 2 & -2+2\sqrt{2} \end{bmatrix}.$$

The first equation is

$$(2+2\sqrt{2})x_1 + 2x_2 = 0.$$
$$(1+\sqrt{2})x_1 + x_2 = 0 \Rightarrow x_1 = (1-\sqrt{2})x_2$$

Letting $x_2 = t$, we obtain

$$\mathbf{x} = t\begin{bmatrix} 1-\sqrt{2} \\ 1 \end{bmatrix}.$$

Therefore,

$$E_{\lambda_2} = \text{Span}\left\{\begin{bmatrix} 1-\sqrt{2} \\ 1 \end{bmatrix}\right\}.$$

Hence, bases for the two eigenspaces are

$$\left\{\begin{bmatrix} 1+\sqrt{2} \\ 1 \end{bmatrix}\right\}, \quad \left\{\begin{bmatrix} 1-\sqrt{2} \\ 1 \end{bmatrix}\right\}.$$

Eigenvalues are often easier to determine after replacing a matrix with a simpler one. However, we cannot use arbitrary row operations, since row operations generally change the eigenvalues of a

matrix. Instead, we need a transformation that simplifies the matrix while preserving its essential structure.

Similar Matrices

Let A and B be $n \times n$ matrices. A is **similar** to B if there exists an invertible matrix P such that

$$B = P^{-1}AP.$$

Replacing A with $P^{-1}AP$ is called a **similarity transformation**.

Similar matrices represent the same linear transformation with respect to different bases. The matrix P converts coordinates from one basis to another, while P^{-1} converts them back. Thus, A and B may look different, but they describe the same underlying transformation from two different coordinate systems.

Similarity is also symmetric. If

$$B = P^{-1}AP,$$

then

$$A = PBP^{-1}.$$

Therefore, if A is similar to B, then B is also similar to A, and we simply say that A and B are similar.

Similarity and the Characteristic Polynomial

If A and B are similar $n \times n$ matrices, then they have the same characteristic polynomial. Consequently, they have the same eigenvalues with the same algebraic multiplicities.

Proof

Suppose A and B are similar. Then there exists an invertible matrix P such that

$$B = P^{-1}AP.$$

Subtracting λI gives

$$\begin{aligned} B - \lambda I &= P^{-1}AP - \lambda I \\ &= P^{-1}AP - \lambda P^{-1}P \\ &= P^{-1}(A - \lambda I)P. \end{aligned}$$

Taking determinants,

$$\begin{aligned} \det(B - \lambda I) &= \det\left(P^{-1}(A - \lambda I)P\right) \\ &= \det(P^{-1})\det(A - \lambda I)\det(P). \end{aligned}$$

Since

$$\det(P^{-1})\det(P) = \det(P^{-1}P) = \det(I) = 1,$$

we obtain

$$\det(B - \lambda I) = \det(A - \lambda I).$$

Therefore,

$$p_B(\lambda) = p_A(\lambda),$$

so A and B have the same characteristic polynomial. Hence, they have the same eigenvalues with the same algebraic multiplicities. □

However, having the same eigenvalues does *not* necessarily mean that two matrices are similar. For example,

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

both have the single eigenvalue $\lambda = 2$ with algebraic multiplicity two, but they are not similar. Indeed, since $B = 2I$,

$$P^{-1}BP = P^{-1}(2I)P = 2P^{-1}P = 2I = B$$

for every invertible matrix P. Thus, B is similar only to itself.

Similarity is also not the same as row equivalence. If A is row equivalent to B, then

$$B = EA$$

for some invertible matrix E. A similarity transformation instead has the form

$$B = P^{-1}AP.$$

Because row operations do not generally preserve eigenvalues, an REF or RREF of A will usually not have the same eigenvalues as A.

Consider a discrete dynamical system of the form

$$\mathbf{x}_{k+1} = A\mathbf{x}_k, \qquad k = 0, 1, 2, \ldots$$

where $\mathbf{x}_0$ is the initial state. Repeated substitution gives

$$\mathbf{x}_k = A^k\mathbf{x}_0.$$

Computing large powers of A directly is usually inefficient. Eigenvectors provide a much simpler description because if $A\mathbf{v} = \lambda\mathbf{v}$, then $A^k\mathbf{v} = \lambda^k\mathbf{v}$.

Thus, after expressing the initial state as a linear combination of eigenvectors, each component evolves independently according to a power of its corresponding eigenvalue.

Example 5.2.5

Let

$$A = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix}, \quad \mathbf{x}_0 = \begin{bmatrix} 0.60 \\ 0.40 \end{bmatrix}.$$

Analyze the long-term behavior of the dynamical system

$$\mathbf{x}_{k+1} = A\mathbf{x}_k.$$

Solution

$$\begin{aligned}
\det \begin{bmatrix} 0.95-\lambda & 0.03 \\ 0.05 & 0.97-\lambda \end{bmatrix} &= 0, \\
(0.95-\lambda)(0.97-\lambda) - (0.03)(0.05) &= 0, \\
\lambda^2 - 1.92\lambda + 0.92 &= 0, \\
(\lambda - 1)(\lambda - 0.92) &= 0.
\end{aligned}$$

Therefore, the eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = 0.92$.

Solving the corresponding homogeneous systems gives

$$E_{\lambda_1} = \text{Span}\left\{ \begin{bmatrix} 3 \\ 5 \end{bmatrix} \right\}, \quad E_{\lambda_2} = \text{Span}\left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}.$$

Let

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Since $\mathbf{v}_1$ and $\mathbf{v}_2$ correspond to distinct eigenvalues, they are linearly independent and therefore form a basis for $\mathbb{R}^2$. Hence, the initial state can be written uniquely as

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2.$$

Substituting the known vectors gives

$$\begin{bmatrix} 0.60 \\ 0.40 \end{bmatrix} = c_1 \begin{bmatrix} 3 \\ 5 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Equivalently,

$$\begin{bmatrix} 3 & 1 \\ 5 & -1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0.60 \\ 0.40 \end{bmatrix}.$$

Solving gives

$$c_1 = 0.125, \quad c_2 = 0.225.$$

Therefore,

$$\mathbf{x}_0 = 0.125\mathbf{v}_1 + 0.225\mathbf{v}_2.$$

We now apply A^k to both sides:

$$\begin{aligned}\mathbf{x}_k &= A^k\mathbf{x}_0 \\ &= 0.125A^k\mathbf{v}_1 + 0.225A^k\mathbf{v}_2\end{aligned}$$

Because $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors,

$$A^k\mathbf{v}_1 = \lambda_1^k\mathbf{v}_1 = 1^k\mathbf{v}_1 = \mathbf{v}_1$$

and

$$A^k\mathbf{v}_2 = \lambda_2^k\mathbf{v}_2 = (0.92)^k\mathbf{v}_2.$$

Thus,

$$\mathbf{x}_k = 0.125\begin{bmatrix}3\\5\end{bmatrix} + 0.225(0.92)^k\begin{bmatrix}1\\-1\end{bmatrix}.$$

This expression separates the state into two independent modes. The first mode remains unchanged because its eigenvalue is 1, while the second mode gradually disappears because $(0.92)^k$ tends towards zero as k approaches infinity.

Therefore,

$$\lim_{k\to\infty}\mathbf{x}_k = 0.125\begin{bmatrix}3\\5\end{bmatrix} = \begin{bmatrix}0.375\\0.625\end{bmatrix}.$$

The system approaches the steady state

$$\begin{bmatrix} 0.375 \\ 0.625 \end{bmatrix}.$$

In general, if the initial state can be written as

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n,$$

where $A\mathbf{v}_i = \lambda_i\mathbf{v}_i$, then

$$\mathbf{x}_k = c_1\lambda_1^k\mathbf{v}_1 + \cdots + c_n\lambda_n^k\mathbf{v}_n.$$

The long-term behavior is therefore determined by the magnitudes of the eigenvalues.

A widely used numerical method for approximating the eigenvalues of a matrix is the **QR algorithm**. The method begins by factoring A as

$$A = Q_1R_1,$$

where Q_1 is an orthogonal matrix and R_1 is upper triangular. The factors are then reversed to form

$$A_1 = R_1Q_1.$$

The process is repeated by factoring

$$A_1 = Q_2R_2$$

and defining

$$A_2 = R_2Q_2.$$

Continuing in this way produces a sequence

$$A,\ A_1,\ A_2,\ \ldots$$

of matrices that are all similar to A and therefore have the same eigenvalues. Under suitable conditions, these matrices approach an upper triangular matrix whose diagonal entries approximate the eigenvalues of A.

Problem 5.2.1

Suppose an $n \times n$ matrix A can be factored as $A = QR$ where Q is invertible. Show that $A_1 = RQ$ is similar to A.

Solution

Since Q is invertible, we may form the similarity transformation $Q^{-1}AQ$.

Substituting $A = QR$,

$$\begin{aligned} Q^{-1}AQ &= Q^{-1}(QR)Q \\ &= (Q^{-1}Q)RQ \\ &= RQ \\ &= A_1. \end{aligned}$$

Thus A_1 is similar to A. A and A_1 have the same characteristic polynomial and the same eigenvalues.

Although eigenvectors corresponding to a known eigenvalue can be found by solving $(A-\lambda I)\mathbf{x} = \mathbf{0}$, modern software (like NumPy, MATLAB, or Julia) almost always computes the eigenvalues and eigenvectors simultaneously using highly optimized numerical algorithms rather than row reduction. The most important, which is the QR algorithm, is remarkably accurate and can efficiently analyze matrices containing thousands or even millions of unknowns.

Problem 5.2.2

The algebraic multiplicity of an eigenvalue is always greater than or equal to the dimension of its corresponding eigenspace. Consider

$$A = \begin{bmatrix} 5 & -2 & 6 & -1 \\ 0 & 3 & h & 0 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Find the value of h for which the eigenspace corresponding to $\lambda = 5$ is two-dimensional.

Solution

We row reduce the augmented matrix from $(A - 5I)\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} 0 & -2 & 6 & -1 & 0 \\ 0 & -2 & h & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & -4 & 0 \end{bmatrix} \sim \begin{bmatrix} 0 & -2 & 6 & 0 & 0 \\ 0 & 0 & h-6 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The variable x_1 is always free. For the eigenspace to be two-dimensional, x_3 must also be free, so the entry $h - 6$ cannot be a pivot. Therefore,

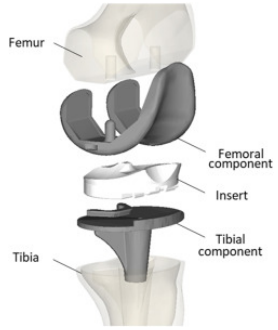
$$h - 6 = 0 \Rightarrow h = 6$$

To verify the result, substitute $h = 6$. The system becomes

$$\begin{bmatrix} 0 & -2 & 6 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

so x_1 and x_3 are free variables. Therefore $\dim(E_5) = 2$ which confirms that $h = 6$.

Problem 5.2.3



Model of an artificial knee joint. Image credit: Putame, Aldieri, et al., Politecnico di Torino

At a point on a polymeric insert in an artificial knee joint, the normal stresses are

$$\sigma_x = -2 \text{ MPa}, \quad \sigma_y = -2 \text{ MPa}, \quad \sigma_z = -3 \text{ MPa},$$

and the shear stresses are

$$\tau_{xy} = 0, \qquad \tau_{yz} = 1 \text{ MPa}, \qquad \tau_{zx} = 1 \text{ MPa}.$$

A stress tensor describes the normal and shear stresses acting at a point in a material. In three dimensions, it can be represented by a 3×3 matrix. The normal stresses appear on the main diagonal of the stress tensor, while the shear stresses appear off the main diagonal. Since complementary shear stresses are equal,

$$\tau_{xy} = \tau_{yx}, \quad \tau_{yz} = \tau_{zy}, \quad \tau_{zx} = \tau_{xz},$$

and the stress tensor is symmetric. Therefore,

$$\boldsymbol{\sigma} = \begin{bmatrix} -2 & 0 & 1 \\ 0 & -2 & 1 \\ 1 & 1 & -3 \end{bmatrix} \text{ MPa}.$$

The eigenvalues of $\boldsymbol{\sigma}$ are the principal normal stresses, and the corresponding unit eigenvectors are the principal directions.

(a) Find the three principal normal stresses.

(b) Find a unit vector in each principal direction.

Solution

(a) The principal normal stresses are the eigenvalues of the stress tensor. Therefore, they satisfy the characteristic equation $\det(\boldsymbol{\sigma} - \lambda I) = 0$. We have

$$\boldsymbol{\sigma} - \lambda I = \begin{bmatrix} -2-\lambda & 0 & 1 \\ 0 & -2-\lambda & 1 \\ 1 & 1 & -3-\lambda \end{bmatrix}.$$

Expanding the determinant along the first row gives

$$\begin{aligned} \det(\boldsymbol{\sigma} - \lambda I) &= (-2-\lambda)\begin{vmatrix} -2-\lambda & 1 \\ 1 & -3-\lambda \end{vmatrix} + \begin{vmatrix} 0 & -2-\lambda \\ 1 & 1 \end{vmatrix} \\ &= (-2-\lambda)\left[(-2-\lambda)(-3-\lambda) - 1\right] + (2+\lambda) \\ &= -(\lambda+2)\left[(\lambda+2)(\lambda+3) - 2\right] \\ &= -(\lambda+2)\left(\lambda^2 + 5\lambda + 4\right) \\ &= -(\lambda+1)(\lambda+2)(\lambda+4). \end{aligned}$$

Thus, the characteristic equation is

$$-(\lambda+1)(\lambda+2)(\lambda+4) = 0,$$

so the principal normal stresses are

$$\sigma_1 = -1 \text{ MPa}, \quad \sigma_2 = -2 \text{ MPa}, \quad \sigma_3 = -4 \text{ MPa}.$$

(b) We now find the principal direction corresponding to each eigenvalue.

For $\sigma_1 = -1$ MPa, solve $(\boldsymbol{\sigma} + I)\mathbf{p}_1 = \mathbf{0}$.

Since

$$\begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & 1 & -2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix},$$

we obtain

$$x_1 = x_3, \quad x_2 = x_3.$$

Therefore,

$$E_{-1} = \text{Span}\left\{\begin{bmatrix}1\\1\\1\end{bmatrix}\right\}.$$

Normalizing this eigenvector gives the first principal direction:

$$\hat{\mathbf{p}}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}$$

For $\sigma_2 = -2$ MPa, solve $(\boldsymbol{\sigma} + 2I)\mathbf{p}_2 = \mathbf{0}$. Since

$$\begin{bmatrix}0 & 0 & 1\\0 & 0 & 1\\1 & 1 & -1\end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix}1 & 1 & 0\\0 & 0 & 1\\0 & 0 & 0\end{bmatrix},$$

we obtain

$$x_1 = -x_2, \quad x_3 = 0.$$

Therefore,

$$E_{-2} = \text{Span}\left\{\begin{bmatrix}1\\-1\\0\end{bmatrix}\right\}.$$

Normalizing gives the second principal direction:

$$\hat{\mathbf{p}}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

For $\sigma_3 = -4$ MPa, solve $(\boldsymbol{\sigma} + 4I)\mathbf{p}_3 = \mathbf{0}$.

Since

$$\begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix},$$

we obtain

$$x_1 = -\frac{1}{2}x_3, \quad x_2 = -\frac{1}{2}x_3.$$

Choosing $x_3 = -2$ gives

$$E_{-4} = \text{Span}\left\{ \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \right\}.$$

Normalizing gives the third principal direction:

$$\hat{\mathbf{p}}_3 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

Thus, the principal stresses and corresponding unit principal directions are

$$\sigma_1 = -1\ \text{MPa}, \quad \hat{\mathbf{p}}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

$$\sigma_2 = -2\ \text{MPa}, \quad \hat{\mathbf{p}}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix},$$

$$\sigma_3 = -4\ \text{MPa}, \quad \hat{\mathbf{p}}_3 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}.$$

5.3 Diagonalization

A square matrix is often easier to analyze when it is similar to a diagonal matrix. More specifically, we can gain insight into the properties of square matrices via the factorization $A = PDP^{-1}$ where D is a diagonal matrix.

Diagonalizable Matrices

An $n \times n$ matrix A is **diagonalizable** if it is similar to a diagonal matrix.

The Diagonalization Theorem

An $n \times n$ matrix A is diagonalizable if and only if it has n linearly independent eigenvectors. If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly independent eigenvectors of A corresponding to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then

$$A = PDP^{-1},$$

where

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$$

and

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Conversely, if $A = PDP^{-1}$, where P is invertible and D is diagonal, then the *columns* of P are linearly independent eigenvectors of A, and the corresponding *diagonal entries* of D are their eigenvalues.

Proof

Suppose A is diagonalizable, so $A = PDP^{-1}$ where P is invertible and D is diagonal. Multiplying on the right by P gives

$$AP = PD.$$

Write

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}, \quad D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

We have

$$AP = PD,$$

$$\begin{bmatrix} A\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n \end{bmatrix}.$$

And thus their corresponding columns are equal. Therefore,

$$A\mathbf{v}_i = \lambda_i\mathbf{v}_i, \qquad i = 1, 2, \ldots, n.$$

Thus, the columns of P are eigenvectors of A. Since P is invertible, its columns are linearly independent.

Conversely, suppose A has n linearly independent eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.

Construct P and D from these eigenvectors and eigenvalues. Since $A\mathbf{v}_i = \lambda_i\mathbf{v}_i$ for each eigenvector, it follows that $AP = PD$. Because the columns of P are linearly independent, P is invertible. Multiplying by P^{-1} gives

$$A = PDP^{-1}.$$

Therefore, A is diagonalizable if and only if it has n linearly independent eigenvectors. □

Eigenvector Basis

An $n \times n$ matrix A is diagonalizable if and only if $\mathbb{R}^n$ has a basis consisting entirely of eigenvectors of A. Such a basis is called an **eigenvector basis** for $\mathbb{R}^n$.

Example 5.3.1

Let

$$D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}.$$

Compute D^2 and D^3. Then, write the general pattern.

Solution

$$D^2 = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix}$$

$$D^3 = DD^2 = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix} = \begin{bmatrix} 5^3 & 0 \\ 0 & 3^3 \end{bmatrix}$$

In general,

$$D^k = \begin{bmatrix} 5^k & 0 \\ 0 & 3^k \end{bmatrix}, \qquad k \geq 1$$

Example 5.3.2

Let

$$A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}.$$

Find a formula for A^k, given that $A = PDP^{-1}$ where

$$P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}, \quad D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}.$$

Solution

First,

$$P^{-1} = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}.$$

Since $A = PDP^{-1}$,

$$\begin{aligned} A^2 &= (PDP^{-1})(PDP^{-1}) \\ &= PD(P^{-1}P)DP^{-1} \\ &= PD^2P^{-1}. \end{aligned}$$

Similarly, $A^3 = (PDP^{-1})(PD^2P^{-1}) = PD^3P^{-1}$.

In general, for $k \geq 1$,

$$\begin{aligned} A^k = PD^kP^{-1} &= \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 5^k & 0 \\ 0 & 3^k \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 2 \cdot 5^k - 3^k & 5^k - 3^k \\ 2 \cdot 3^k - 2 \cdot 5^k & 2 \cdot 3^k - 5^k \end{bmatrix}. \end{aligned}$$

Matrices with Distinct Eigenvalues

An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

Proof

Eigenvectors corresponding to distinct eigenvalues are linearly independent. Therefore, an $n \times n$ matrix with n distinct eigenvalues has n linearly independent eigenvectors. By the diagonalization theorem, the matrix is diagonalize. □

Note that the diagonalization theorem is the most important criterion, as a square matrix does not *necessarily* need to have n distinct eigen values to be diagonalizable.

How to Diagonalize a Matrix

Let A be an $n \times n$ matrix. To diagonalize A,

1. Find the eigenvalues of A from the characteristic equation

$$\det(A - \lambda I) = 0.$$

2. For each distinct eigenvalue λ, find a basis for the eigenspace $E_\lambda = \text{Nul}(A - \lambda I)$.

3. Count the total number of vectors in these eigenspace bases. If there are fewer than n, then A is not diagonalizable. Otherwise, arrange the n eigenvectors as the columns of

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}.$$

4. Construct D by placing the eigenvalue corresponding to each eigenvector along the diagonal in the same order:

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Then P is invertible and $A = PDP^{-1}$. If an eigenvalue corresponds to more than one column of P, it must be repeated along the diagonal of D once for each corresponding eigenvector.

Repeated eigenvalues do not necessarily prevent a matrix from being diagonalizable. The central question is whether A has enough linearly independent eigenvectors to form a basis for $\mathbb{R}^n$. When eigenvalues repeat, this means checking whether each repeated eigenvalue contributes enough independent eigenvectors.

Diagonalizability with Repeated Eigenvalues

Let A be an $n \times n$ matrix with distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$. For each eigenvalue λ_k, let m_k denote its algebraic multiplicity and let

$$E_{\lambda_k} = \text{Nul}(A - \lambda_k I)$$

be its corresponding eigenspace. Then,

(i) The dimension of each eigenspace cannot exceed the algebraic multiplicity of its eigenvalue:

$$\dim(E_{\lambda_k}) \leq m_k$$

(ii) The matrix A is diagonalizable if and only if

$$\dim(E_{\lambda_1}) + \dim(E_{\lambda_2}) + \cdots + \dim(E_{\lambda_p}) = n.$$

Equivalently, A is diagonalizable if and only if its characteristic polynomial factors completely into linear factors and

$$\dim(E_{\lambda_k}) = m_k$$

for every eigenvalue λ_k.

(iii) If A is diagonalizable and $\mathcal{B}_k$ is a basis for E_{λ_k}, then the combined collection of vectors in $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_p$ forms an eigenvector basis for $\mathbb{R}^n$.

Thus, a repeated eigenvalue causes a matrix to fail to be diagonalizable only when its eigenspace is too small. In particular, if $\dim(E_{\lambda_k}) < m_k$ for even one eigenvalue, then A is not diagonalizable.

Example 5.3.3

Diagonalize the following matrix, if possible:

$$A = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}$$

Solution

1. Find the eigenvalues of A.

The characteristic equation is

$$\begin{aligned} 0 &= \det(A - \lambda I) \\ &= -\lambda^3 - 3\lambda^2 + 4 \\ &= -(\lambda - 1)(\lambda + 2)^2. \end{aligned}$$

Thus, the distinct eigenvalues are $\lambda = 1$ and $\lambda = -2$, where $\lambda = -2$ has algebraic multiplicity 2.

2. Find a basis for each eigenspace.

For $\lambda = 1$, solve $(A - I)\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} 0 & 3 & 3 & 0 \\ -3 & -6 & -3 & 0 \\ 3 & 3 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus, $x_1 = x_3$ and $x_2 = -x_3$. Letting $x_3 = t$,

$$\mathbf{x} = t \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}.$$

Therefore, a basis for E_1 is

$$\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}.$$

For $\lambda = -2$, solve $(A + 2I)\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} 3 & 3 & 3 & 0 \\ -3 & -3 & -3 & 0 \\ 3 & 3 & 3 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus, $x_1 = -x_2 - x_3$. Letting $x_2 = s$ and $x_3 = t$,

$$\mathbf{x} = s \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

Therefore, a basis for E_{-2} is

$$\mathcal{B}_{-2} = \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

3. Count the eigenvectors and construct P.

The two eigenspace bases contain $1 + 2 = 3$ eigenvectors in total. Since A is a 3×3 matrix, these vectors form an eigenvector basis for $\mathbb{R}^3$, so A is diagonalizable.

Arrange the eigenvectors as the columns of P:

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

4. Construct D from the corresponding eigenvalues.

The columns $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ correspond to the eigenvalues 1, -2, and -2, respectively. Therefore,

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

The repeated eigenvalue $\lambda = -2$ appears twice along the diagonal of D, once for each corresponding eigenvector placed in P.

Since $\det(P) = 1$, the matrix P is invertible. As a check,

$$AP = \begin{bmatrix} 1 & 2 & 2 \\ -1 & -2 & 0 \\ 1 & 0 & -2 \end{bmatrix} = PD.$$

Therefore, the diagonalization of A is

$$A = PDP^{-1}.$$

Geometric Multiplicity

Let λ be an eigenvalue of A. The **geometric multiplicity** of λ is the dimension of its eigenspace:

$$\mathrm{GM}(\lambda) = \dim(E_\lambda) = \dim(\mathrm{Nul}(A - \lambda I))$$

Geometric multiplicity counts how many linearly independent eigenvectors are available for the corresponding columns of P. If the geometric multiplicity is smaller than the algebraic multiplicity, then there are not enough eigenvectors to diagonalize A.

For every eigenvalue,

$$1 \leq \mathrm{GM}(\lambda) \leq \mathrm{AM}(\lambda).$$

Example 5.3.4

Determine whether the matrix

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

is diagonalizable.

The characteristic equation is

$$\begin{aligned} 0 &= \det(A - \lambda I) \\ &= \det \begin{bmatrix} -\lambda & 1 \\ 0 & -\lambda \end{bmatrix} \\ &= \lambda^2. \end{aligned}$$

Thus, $\lambda = 0$ is the only eigenvalue and has algebraic multiplicity 2.

To find its geometric multiplicity, solve $A\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This gives $x_2 = 0$, while x_1 is free. Therefore,

$$E_0 = \text{Span} \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\},$$

so the geometric multiplicity of $\lambda = 0$ is 1.

Since the geometric multiplicity is less than the algebraic multiplicity, A has only one linearly independent eigenvector, while two are required to form a basis for $\mathbb{R}^2$. Therefore, A is not diagonalizable.

The diagonalizability of a matrix is independent of whether the matrix is invertible. The following matrices represent all four possible cases:

$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$
diagonalizable and invertible	diagonalizable but not invertible	invertible but not diagonalizable	neither invertible nor diagonalizable

Invertibility depends on whether 0 is an eigenvalue, while diagonalizability depends on whether the matrix has enough linearly independent eigenvectors to form a basis.

A diagonal matrix is easy to understand geometrically because it scales each coordinate axis independently. For example, let

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

Then

$$D\mathbf{e}_1 = \mathbf{e}_1, \quad D\mathbf{e}_2 = 2\mathbf{e}_2, \quad D\mathbf{e}_3 = 3\mathbf{e}_3.$$

Thus, D leaves each coordinate axis unchanged in direction while scaling it by the corresponding diagonal entry.

A diagonalizable matrix behaves in the same way, but with respect to an eigenvector basis rather than the standard basis. Suppose A has n linearly independent eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Recall that

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i, \qquad i = 1, 2, \ldots, n.$$

This means that A scales the $\mathbf{v}_i$-direction by the factor λ_i. These eigenvector directions act as a new set of coordinate axes in which the transformation becomes diagonal.

If

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n,$$

then linearity gives

$$A\mathbf{x} = c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \cdots + c_n\lambda_n\mathbf{v}_n.$$

Repeated application of A gives

$$A^k\mathbf{x} = c_1\lambda_1^k\mathbf{v}_1 + c_2\lambda_2^k\mathbf{v}_2 + \cdots + c_n\lambda_n^k\mathbf{v}_n.$$

The long-term behavior of each component is therefore determined by the magnitude and sign of its eigenvalue:

- If $|\lambda_i| > 1$, the component in the $\mathbf{v}_i$-direction grows.
- If $|\lambda_i| < 1$, the component in the $\mathbf{v}_i$-direction decays toward zero.
- If $|\lambda_i| = 1$, the magnitude of the component remains constant.
- If $\lambda_i < 0$, the component reverses direction after each application of A.
- If $\lambda_i = 0$, the component is eliminated after one application of A.

Thus, diagonalization reveals the geometry of a linear transformation by separating its action into independent scaling along eigenvector directions.

Example 5.3.5

Describe how the matrix

$$A = \frac{1}{10}\begin{bmatrix} 11 & 6 \\ 9 & 14 \end{bmatrix}$$

acts on the plane.

Solution

First, diagonalize A. Its characteristic polynomial is

$$\begin{aligned} \det(A - \lambda I) &= \lambda^2 - \frac{5}{2}\lambda + 1 \\ &= (\lambda - 2)\left(\lambda - \frac{1}{2}\right). \end{aligned}$$

Thus, the eigenvalues are $\lambda_1 = 2$ and $\lambda_2 = \frac{1}{2}$.

For $\lambda_1 = 2$,

$$A - 2I = \frac{1}{10}\begin{bmatrix} -9 & 6 \\ 9 & -6 \end{bmatrix} \sim \begin{bmatrix} 3 & -2 \\ 0 & 0 \end{bmatrix}.$$

Therefore, $3x_1 - 2x_2 = 0$, so $x_1 = \frac{2}{3}x_2$. Choosing $x_2 = 3$ gives the eigenvector

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

For $\lambda_2 = \frac{1}{2}$,

$$A - \frac{1}{2}I = \frac{1}{10}\begin{bmatrix} 6 & 6 \\ 9 & 9 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}.$$

Therefore, $x_1 + x_2 = 0$, so $x_1 = -x_2$. Choosing $x_2 = 1$ gives the eigenvector

$$\mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Since the two eigenvectors are linearly independent, they form an eigenvector basis for $\mathbb{R}^2$. Thus,

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix}, \qquad D = \begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{bmatrix},$$

and $A = PDP^{-1}$.

Now write an arbitrary vector as $\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$. Then

$$A\mathbf{x} = 2c_1\mathbf{v}_1 + \frac{1}{2}c_2\mathbf{v}_2.$$

Therefore, A stretches the component in the $\mathbf{v}_1$-direction by a factor of 2 and contracts the component in the $\mathbf{v}_2$-direction by a factor of $\frac{1}{2}$.

Repeated application gives

$$A^k\mathbf{x} = 2^k c_1\mathbf{v}_1 + \left(\frac{1}{2}\right)^k c_2\mathbf{v}_2.$$

Thus, when $c_1 \neq 0$, the direction of $A^k\mathbf{x}$ approaches the eigenspace E_2, while the component in $E_{\frac{1}{2}}$ decays toward zero.

In eigenvector coordinates, the product of the two coordinates is preserved because

$$(2c_1)\left(\frac{1}{2}c_2\right) = c_1c_2.$$

Consequently, A moves vectors along curves of the form $c_1c_2 = C$ when the plane is described using the eigenvector basis $\{\mathbf{v}_1, \mathbf{v}_2\}$.

Example 5.3.6

Consider the dynamical system

$$\mathbf{x}_{k+1} = A\mathbf{x}_k,$$

where

$$A = \frac{1}{70}\begin{bmatrix} 61 & 4 \\ -8 & 79 \end{bmatrix}.$$

Describe how A acts on the plane and explain the long-term behavior of the system.

Solution

The entries of A do not immediately reveal how the system behaves. To expose its geometry, we diagonalize A.

The characteristic polynomial is

$$\begin{aligned}\det(A - \lambda I) &= \lambda^2 - 2\lambda + \frac{99}{100} \\ &= \left(\lambda - \frac{11}{10}\right)\left(\lambda - \frac{9}{10}\right).\end{aligned}$$

Thus, the eigenvalues are $\lambda_1 = \frac{11}{10}$ and $\lambda_2 = \frac{9}{10}$.

For $\lambda_1 = \frac{11}{10}$,

$$A - \frac{11}{10}I = \frac{1}{35}\begin{bmatrix} -8 & 2 \\ -4 & 1 \end{bmatrix} \sim \begin{bmatrix} 4 & -1 \\ 0 & 0 \end{bmatrix}.$$

Therefore, $4x_1 - x_2 = 0$, so $x_2 = 4x_1$. Choosing $x_1 = 1$ gives

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix},$$

and hence

$$E_{\frac{11}{10}} = \text{Span}\left\{ \begin{bmatrix} 1 \\ 4 \end{bmatrix} \right\}.$$

For $\lambda_2 = \frac{9}{10}$,

$$A - \frac{9}{10}I = \frac{1}{35}\begin{bmatrix} -1 & 2 \\ -4 & 8 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 \\ 0 & 0 \end{bmatrix}.$$

Therefore, $x_1 - 2x_2 = 0$, so $x_1 = 2x_2$. Choosing $x_2 = 1$ gives

$$\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

and hence

$$E_{\frac{9}{10}} = \text{Span}\left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\}.$$

The eigenvectors are linearly independent, so they form an eigenvector basis for $\mathbb{R}^2$. Therefore,

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 4 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} \frac{11}{10} & 0 \\ 0 & \frac{9}{10} \end{bmatrix},$$

and $A = PDP^{-1}$.

Now write the initial state in the eigenvector basis as

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2.$$

Repeatedly applying the recurrence relation gives

$$\mathbf{x}_k = A^k\mathbf{x}_0 = c_1\left(\frac{11}{10}\right)^k\mathbf{v}_1 + c_2\left(\frac{9}{10}\right)^k\mathbf{v}_2.$$

Because $\frac{11}{10} > 1$, the component in the $\mathbf{v}_1$-direction grows by a factor of $\frac{11}{10}$ at each step. Because $0 < \frac{9}{10} < 1$, the component in the $\mathbf{v}_2$-direction shrinks by a factor of $\frac{9}{10}$ at each step.

Therefore, unless $c_1 = 0$, the growing component eventually dominates and the trajectory approaches the direction of $E_{\frac{11}{10}}$. If $c_1 = 0$, the trajectory remains in $E_{\frac{9}{10}}$ and converges to the origin.

For example, consider

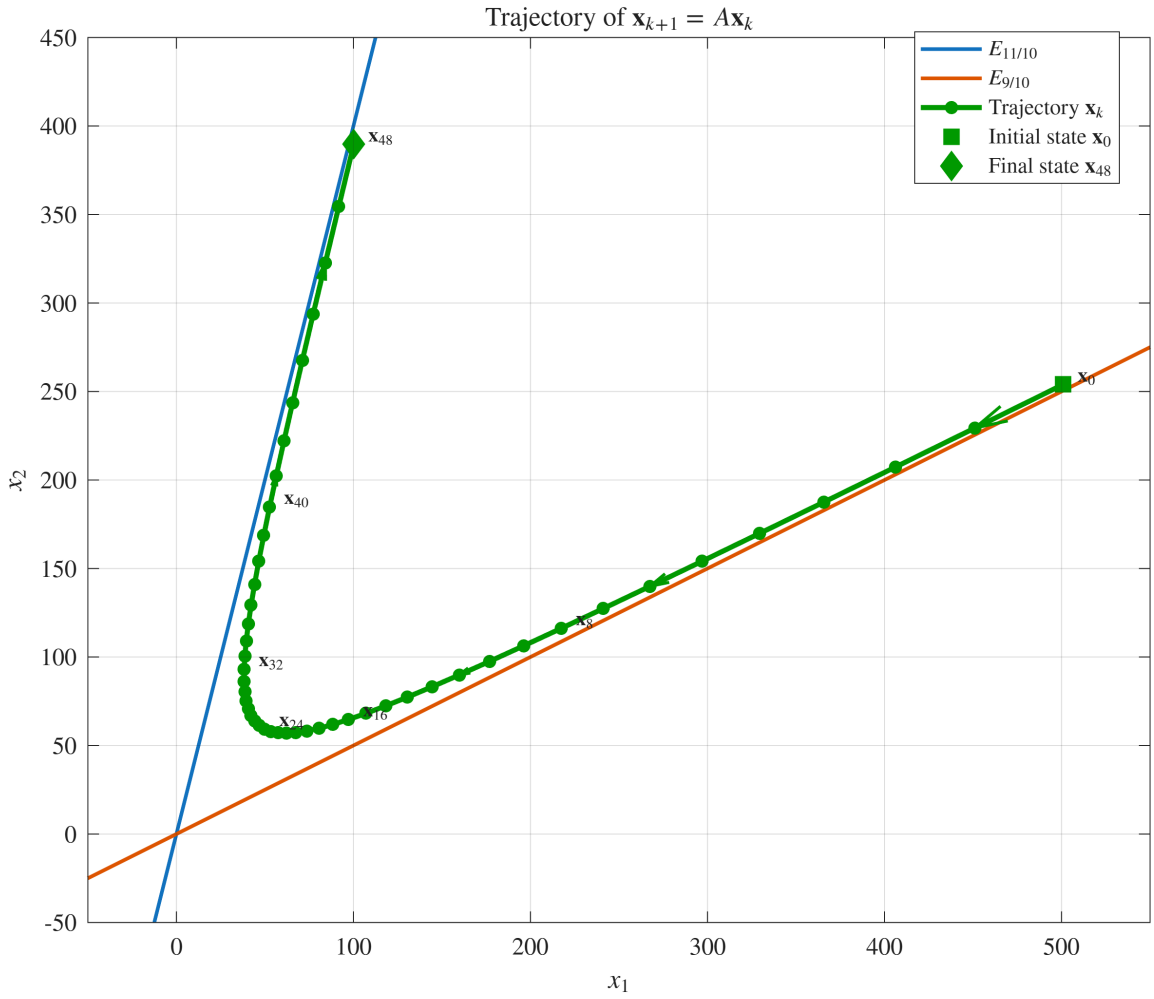
$$\mathbf{x}_0 = \mathbf{v}_1 + 250\mathbf{v}_2 = \begin{bmatrix}501\\254\end{bmatrix}.$$

Then

$$\mathbf{x}_k = \left(\frac{11}{10}\right)^k\mathbf{v}_1 + 250\left(\frac{9}{10}\right)^k\mathbf{v}_2.$$

Initially, the large component in the $\mathbf{v}_2$-direction dominates, so the trajectory moves toward the origin approximately along $E_{\frac{9}{10}}$. As this component decays and the component in the $\mathbf{v}_1$-direction grows, the trajectory turns and moves away from the origin along $E_{\frac{11}{10}}$.

Thus, diagonalization separates the dynamics into one contracting eigendirection (path of the eigenvector) and one expanding eigendirection, making the behavior of the system immediately visible.


Trajectory of $\mathbf{x}_{k+1} = A\mathbf{x}_k$
$E_{11/10}$
$E_{9/10}$
Trajectory $\mathbf{x}_k$
Initial state $\mathbf{x}_0$
Final state $\mathbf{x}_{48}$
$\mathbf{x}_0$
$\mathbf{x}_8$
$\mathbf{x}_{16}$
$\mathbf{x}_{24}$
$\mathbf{x}_{32}$
$\mathbf{x}_{40}$
$\mathbf{x}_{48}$
x_1
x_2
450
400
350
300
250
200
150
100
50
0
-50
0
100
200
300
400
500

Problem 5.3.1

Let

$$A = \begin{bmatrix} 1 & -1 & 2 & 1 \\ 0 & 2 & a & 1 \\ 0 & 0 & 1 & b \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

(a) Determine the values of a and b for which A is diagonalizable. Justify your answer.

(b) For these values of a and b, find a basis for each eigenspace.

(c) Construct matrices P and D such that $A = PDP^{-1}$.

(d) Verify the diagonalization by showing that $AP = PD$.

Solution

(a) Since A is upper triangular, its eigenvalues are the entries along its main diagonal. Thus,

$$\det(A - \lambda I) = (1 - \lambda)^2(2 - \lambda)^2,$$

so $\lambda = 1$ and $\lambda = 2$ each have algebraic multiplicity 2. For A to be diagonalizable, both eigenspaces must have dimension 2.

For $\lambda = 1$,

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

The equation $(A - I)\mathbf{x} = \mathbf{0}$ gives $x_4 = 0$. The first two equations then become $x_2 = 2x_3$ and $(a + 2)x_3 = 0$.

The variable x_1 is always free. For E_1 to have dimension 2, x_3 must also be free. Therefore, $a = -2$.

For $\lambda = 2$,

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

The variable x_2 is always free. For E_2 to have dimension 2, the equations

$$ax_3 + x_4 = 0, \quad -x_3 + bx_4 = 0$$

must have a nonzero solution. Therefore, the coefficient matrix must be singular:

$$\det \begin{bmatrix} a & 1 \\ -1 & b \end{bmatrix} = ab + 1 = 0.$$

Substituting $a = -2$ gives

$$-2b + 1 = 0,$$

so $b = \frac{1}{2}$. Therefore, A is diagonalizable when

$$a = -2, \quad b = \frac{1}{2}.$$

(b) For $a = -2$ and $b = \frac{1}{2}$, the equations for E_1 reduce to $x_2 = 2x_3$ and $x_4 = 0$. Letting $x_1 = s$ and $x_3 = t$,

$$\mathbf{x} = \begin{bmatrix} s \\ 2t \\ t \\ 0 \end{bmatrix} = s \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix}.$$

Thus,

$$\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

For E_2, the equations reduce to $x_4 = 2x_3$ and $x_1 = -x_2 + 4x_3$. Letting $x_2 = s$ and $x_3 = t$,

$$\mathbf{x} = \begin{bmatrix} -s + 4t \\ s \\ t \\ 2t \end{bmatrix} = s \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 4 \\ 0 \\ 1 \\ 2 \end{bmatrix}.$$

Thus,

$$\mathcal{B}_2 = \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 1 \\ 2 \end{bmatrix} \right\}.$$

(c) Arrange the eigenvectors from $\mathcal{B}_1$ first, followed by the eigenvectors from $\mathcal{B}_2$:

$$P = \begin{bmatrix} 1 & 0 & -1 & 4 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

The corresponding eigenvalues must appear along the diagonal of D in the same order:

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

Since $\det(P) = -2 \neq 0$, the matrix P is invertible. Therefore,

$$A = PDP^{-1}.$$

(d) Substituting $a = -2$ and $b = \frac{1}{2}$ gives

$$AP = \begin{bmatrix} 1 & 0 & -2 & 8 \\ 0 & 2 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 4 \end{bmatrix}, \quad PD = \begin{bmatrix} 1 & 0 & -2 & 8 \\ 0 & 2 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 4 \end{bmatrix}.$$

Hence, $AP = PD$.

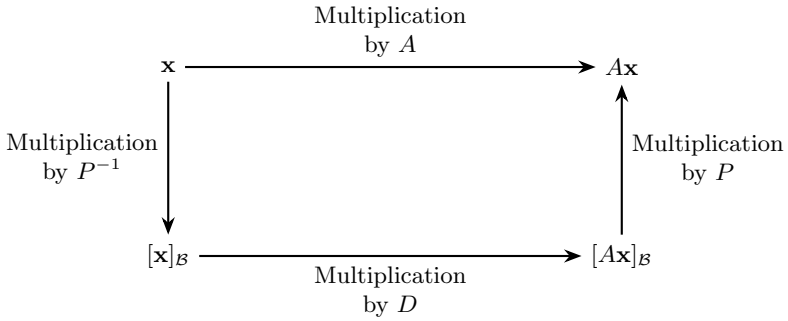
Diagonalization as a Matrix Representation

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be the linear transformation defined by $T(\mathbf{x}) = A\mathbf{x}$. Suppose $A = PDP^{-1}$, where the columns of P form an eigenvector basis $\mathcal{B}$ for $\mathbb{R}^n$. Then

$$[T]_{\mathcal{B}} = P^{-1}AP = D.$$

Thus, diagonalization expresses the same linear transformation in a basis of eigenvectors. The matrix A represents the transformation in the standard basis, while D represents it in the eigenvector basis. More generally, similar matrices are different matrix representations of the same linear transformation.

The relationship between these representations is summarized by the following diagram. Starting with $\mathbf{x}$, we may either apply A directly or convert $\mathbf{x}$ to eigenvector coordinates using P^{-1}, apply D, and then return to standard coordinates using P. Both paths produce $A\mathbf{x}$.



Example 5.3.7

Let $T : \mathbb{P}_2 \to \mathbb{P}_2$ be a linear transformation whose matrix relative to the basis $\mathcal{B} = \{1, t, t^2\}$ is

$$[T]_{\mathcal{B}} = \begin{bmatrix} 3 & 4 & 0 \\ 0 & 5 & -1 \\ 1 & -2 & 7 \end{bmatrix}.$$

Find $T(a_0 + a_1 t + a_2 t^2)$.

Solution

Let $\mathbf{p}(t) = a_0 + a_1 t + a_2 t^2$. Since $\mathcal{B} = \{1, t, t^2\}$, its coordinate vector is

$$[\mathbf{p}]_{\mathcal{B}} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}.$$

The coordinate vector of $T(\mathbf{p})$ is therefore

$$\begin{aligned} [T(\mathbf{p})]_{\mathcal{B}} &= [T]_{\mathcal{B}}[\mathbf{p}]_{\mathcal{B}} \\ &= \begin{bmatrix} 3 & 4 & 0 \\ 0 & 5 & -1 \\ 1 & -2 & 7 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} \\ &= \begin{bmatrix} 3a_0 + 4a_1 \\ 5a_1 - a_2 \\ a_0 - 2a_1 + 7a_2 \end{bmatrix}. \end{aligned}$$

These entries are the coefficients of 1, t, and t^2, respectively. Thus,

$$T(\mathbf{p}) = (3a_0 + 4a_1) + (5a_1 - a_2)t + (a_0 - 2a_1 + 7a_2)t^2.$$

Problem 5.3.2

Let $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ be the standard basis for $\mathbb{R}^3$, let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ be a basis for a vector space V, and let $T : \mathbb{R}^3 \to V$ be the linear transformation defined by

$$T(x_1, x_2, x_3) = (x_3 - x_2)\mathbf{b}_1 - (x_1 + x_3)\mathbf{b}_2 + (x_1 - x_2)\mathbf{b}_3.$$

(a) Compute $T(\mathbf{e}_1)$, $T(\mathbf{e}_2)$, and $T(\mathbf{e}_3)$.

(b) Compute $[T(\mathbf{e}_1)]_{\mathcal{B}}$, $[T(\mathbf{e}_2)]_{\mathcal{B}}$, and $[T(\mathbf{e}_3)]_{\mathcal{B}}$.

(c) Find the matrix for T relative to the bases $\mathcal{E}$ and $\mathcal{B}$.

Solution

(a) Substituting each standard basis vector into the definition of T gives

$$\begin{aligned} T(\mathbf{e}_1) &= T(1,0,0) \\ &= -\mathbf{b}_2 + \mathbf{b}_3, \end{aligned}$$

$$\begin{aligned} T(\mathbf{e}_2) &= T(0,1,0) \\ &= -\mathbf{b}_1 - \mathbf{b}_3, \end{aligned}$$

$$\begin{aligned} T(\mathbf{e}_3) &= T(0,0,1) \\ &= \mathbf{b}_1 - \mathbf{b}_2. \end{aligned}$$

(b) The coefficients of $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$ give the corresponding $\mathcal{B}$-coordinate vectors:

$$[T(\mathbf{e}_1)]_{\mathcal{B}} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \quad [T(\mathbf{e}_2)]_{\mathcal{B}} = \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix}, \quad [T(\mathbf{e}_3)]_{\mathcal{B}} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

(c) The columns of the matrix for T are the $\mathcal{B}$-coordinate vectors of the images of the basis vectors in $\mathcal{E}$. Therefore,

$$\begin{aligned} [T]_{\mathcal{B}\leftarrow\mathcal{E}} &= \begin{bmatrix} [T(\mathbf{e}_1)]_{\mathcal{B}} & [T(\mathbf{e}_2)]_{\mathcal{B}} & [T(\mathbf{e}_3)]_{\mathcal{B}} \end{bmatrix} \\ &= \begin{bmatrix} 0 & -1 & 1 \\ -1 & 0 & -1 \\ 1 & -1 & 0 \end{bmatrix}. \end{aligned}$$

Indeed, for any $\mathbf{x} = (x_1, x_2, x_3)$,

$$[T(\mathbf{x})]_{\mathcal{B}} = \begin{bmatrix} 0 & -1 & 1 \\ -1 & 0 & -1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_3 - x_2 \\ -(x_1 + x_3) \\ x_1 - x_2 \end{bmatrix}.$$

5.4 Complex Eigenvalues

A real matrix can have eigenvalues that are not real. This happens because the characteristic equation of an $n \times n$ matrix is a polynomial of degree n, and every polynomial of degree n has exactly n complex roots when multiplicities are counted. Thus, allowing complex numbers ensures that none of the eigenvalues are missing.

Complex eigenvalues are especially important because they reveal behavior that may be hidden when a matrix is viewed only over $\mathbb{R}^n$. Real matrices associated with rotation, vibration, oscillation, and other periodic motion often have complex eigenvalues, despite the fact that every entry of the matrix is real.

To capture this information, we allow A to act on vectors in $\mathbb{C}^n$, whose entries may be complex numbers. The only difference is that λ and $\mathbf{x}$ may now contain complex numbers. All of the eigenvalue methods developed earlier still apply.

Complex Eigenvalues and Eigenvectors

Let A be an $n \times n$ real matrix. A complex number λ is a complex eigenvalue of A if there exists a nonzero vector $\mathbf{x} \in \mathbb{C}^n$ such that

$$A\mathbf{x} = \lambda\mathbf{x}.$$

The vector $\mathbf{x}$ is a complex eigenvector corresponding to λ. Equivalently, λ is an eigenvalue of A if

$$\det(A - \lambda I) = 0.$$

Although the matrix A may contain only real entries, its complex eigenvalues can describe real geometric behavior. In particular, complex eigenvalues often indicate that the transformation contains a combination of rotation and scaling.

Example 5.4.1

Let

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

Describe the transformation $\mathbf{x} \mapsto A\mathbf{x}$ and find its complex eigenvalues and eigenvectors.

Solution

The matrix A rotates every vector in $\mathbb{R}^2$ counterclockwise through an angle of $90°$. Repeated application gives

$$A^2 = -I, \quad A^4 = I,$$

so every vector returns to its original position after four applications of A.

No nonzero real vector remains on the same line after a $90°$ rotation. Therefore, A has no real eigenvectors and hence no real eigenvalues.

The characteristic equation is

$$\begin{aligned} 0 &= \det(A - \lambda I) \\ &= \det \begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix} \\ &= \lambda^2 + 1. \end{aligned}$$

Thus, $\lambda^2 = -1$. The complex eigenvalues are

$$\lambda_1 = i, \quad \lambda_2 = -i.$$

For $\lambda_1 = i$, solve $(A - iI)\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The first equation gives $x_2 = -ix_1$. Choosing $x_1 = 1$ gives

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -i \end{bmatrix}.$$

Indeed,

$$A\mathbf{v}_1 = \begin{bmatrix} i \\ 1 \end{bmatrix} = i \begin{bmatrix} 1 \\ -i \end{bmatrix} = i\mathbf{v}_1.$$

For $\lambda_2 = -i$, solve $(A + iI)\mathbf{x} = \mathbf{0}$. This gives $x_2 = ix_1$, so choosing $x_1 = 1$ gives

$$\mathbf{v}_2 = \begin{bmatrix} 1 \\ i \end{bmatrix}.$$

Similarly,

$$A\mathbf{v}_2 = \begin{bmatrix} -i \\ 1 \end{bmatrix} = -i \begin{bmatrix} 1 \\ i \end{bmatrix} = -i\mathbf{v}_2.$$

Therefore,

$$\lambda_1 = i, \quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ -i \end{bmatrix},$$

and

$$\lambda_2 = -i, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ i \end{bmatrix}.$$

This rotational behavior could not have been described by real eigendirections.

Example 5.4.2

Let

$$A = \begin{bmatrix} \frac{1}{2} & -\frac{3}{5} \\ \frac{3}{4} & \frac{11}{10} \end{bmatrix}.$$

Find the eigenvalues of A and a basis for each eigenspace.

Solution

The characteristic equation is

$$\begin{aligned} 0 &= \det(A - \lambda I) \\ &= \det \begin{bmatrix} \frac{1}{2} - \lambda & -\frac{3}{5} \\ \frac{3}{4} & \frac{11}{10} - \lambda \end{bmatrix} \\ &= \left(\frac{1}{2} - \lambda\right)\left(\frac{11}{10} - \lambda\right) + \frac{9}{20} \\ &= \lambda^2 - \frac{8}{5}\lambda + 1. \end{aligned}$$

Using the quadratic formula,

$$\begin{aligned} \lambda &= \frac{\frac{8}{5} \pm \sqrt{\left(-\frac{8}{5}\right)^2 - 4}}{2} \\ &= \frac{4}{5} \pm \frac{3}{5}i. \end{aligned}$$

Thus, the eigenvalues are

$$\lambda_1 = \frac{4}{5} - \frac{3}{5}i, \quad \lambda_2 = \frac{4}{5} + \frac{3}{5}i.$$

For $\lambda_1 = \frac{4}{5} - \frac{3}{5}i$, solve $(A - \lambda_1 I)\mathbf{x} = \mathbf{0}$. Multiplying the coefficient matrix by 20 removes the fractions:

$$\begin{bmatrix} -6+12i & -12 \\ 15 & 6+12i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Since λ_1 is an eigenvalue, the two equations describe the same relationship between x_1 and x_2. Using the second equation gives

$$15x_1 + (6+12i)x_2 = 0,$$

so

$$x_1 = -\left(\frac{2}{5} + \frac{4}{5}i\right) x_2.$$

Choosing $x_2 = 5$ gives the eigenvector

$$\mathbf{v}_1 = \begin{bmatrix} -2-4i \\ 5 \end{bmatrix}.$$

Therefore,

$$E_{\frac{4}{5}-\frac{3}{5}i} = \text{Span}\left\{\begin{bmatrix} -2-4i \\ 5 \end{bmatrix}\right\}.$$

Because A has real entries, the complex conjugate of an eigenvector corresponds to the conjugate eigenvalue. Thus,

$$\mathbf{v}_2 = \begin{bmatrix} -2+4i \\ 5 \end{bmatrix}$$

is an eigenvector corresponding to $\lambda_2 = \frac{4}{5} + \frac{3}{5}i$. Therefore,

$$E_{\frac{4}{5}+\frac{3}{5}i} = \text{Span}\left\{\begin{bmatrix} -2+4i \\ 5 \end{bmatrix}\right\}.$$

We can check our work:

$$A\begin{bmatrix}-2+4i\\5\end{bmatrix}=\begin{bmatrix}-4+2i\\4+3i\end{bmatrix}=\left(\frac{4}{5}+\frac{3}{5}i\right)\begin{bmatrix}-2+4i\\5\end{bmatrix}$$

A useful consequence of working with real matrices is that complex eigenvalues and eigenvectors appear in conjugate pairs. This means that once one complex eigenpair is known, the other can be obtained immediately by taking complex conjugates.

Complex Conjugate Eigenpairs

Let A be a real $n\times n$ matrix. If $\lambda\in\mathbb{C}$ is an eigenvalue of A with corresponding eigenvector $\mathbf{v}\in\mathbb{C}^n$, then $\overline{\lambda}$ is also an eigenvalue of A with corresponding eigenvector $\overline{\mathbf{x}}$.

Suppose

$$A\mathbf{x}=\lambda\mathbf{x}.$$

Taking the complex conjugate of both sides gives

$$\overline{A\mathbf{x}}=\overline{\lambda\mathbf{x}}.$$

Because every entry of A is real, taking the complex conjugate does not change A. Therefore,

$$A\overline{\mathbf{x}}=\overline{\lambda}\,\overline{\mathbf{x}},$$

and $\overline{\mathbf{x}}$ is an eigenvector corresponding to $\overline{\lambda}$.

Consequently, if a real matrix has the complex eigenvalue $\lambda=a+bi$, where $b\neq 0$, then it must also have the eigenvalue $\overline{\lambda}=a-bi$. The corresponding eigenvectors are complex conjugates as well.

The eigenvectors found in Example 5.4.2 were complex conjugates and thus $\mathbf{v}_2=\overline{\mathbf{v}}_1$.

A complex conjugate pair $a \pm bi$ has a direct geometric interpretation. The real part a and imaginary part b determine a scaling factor and a rotation angle.

Example 5.4.3

Let

$$C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix},$$

where $a, b \in \mathbb{R}$ and a and b are not both zero. Find the eigenvalues of C and describe how the transformation $\mathbf{x} \mapsto C\mathbf{x}$ acts on the plane.

Solution

The characteristic equation is

$$\begin{aligned} 0 &= \det(C - \lambda I) \\ &= \det \begin{bmatrix} a-\lambda & -b \\ b & a-\lambda \end{bmatrix} \\ &= (a-\lambda)^2 + b^2. \end{aligned}$$

Thus, $(a-\lambda)^2 = -b^2 \Rightarrow a - \lambda = \pm bi$. Therefore the eigenvalues are $\lambda = a \pm bi$.

Now let $r = \sqrt{a^2 + b^2}$. Factoring r from C gives

$$C = r \begin{bmatrix} \frac{a}{r} & -\frac{b}{r} \\ \frac{b}{r} & \frac{a}{r} \end{bmatrix}.$$

Since $\left(\frac{a}{r}\right)^2 + \left(\frac{b}{r}\right)^2 = 1$, there exists an angle φ such that $\cos\varphi = \frac{a}{r}$ and $\sin\varphi = \frac{b}{r}$. Therefore,

$$C = r \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}.$$

The matrix on the right rotates the plane through φ, while multiplication by r scales every vector by the factor r. Therefore, C represents a rotation followed by a scaling. See Figure 19.

The magnitude of each eigenvalue,

$$|\lambda| = \sqrt{a^2 + b^2} = r,$$

gives the scaling factor, while the argument of $a + bi$ gives the rotation angle.

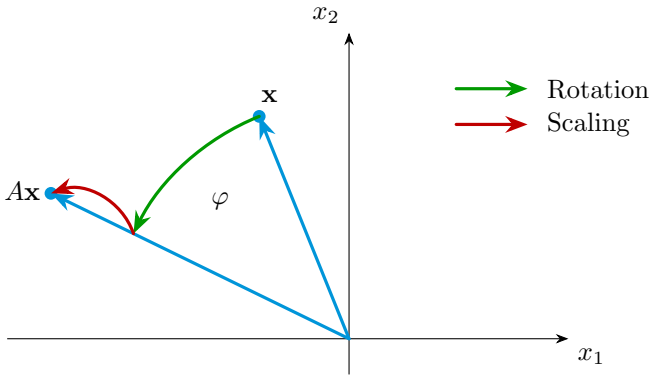


Figure 19: The transformation $A\mathbf{x}$ can be interpreted as first rotating $\mathbf{x}$ through the angle φ, and then scaling its length.

Example 5.4.4

Let

$$A = \begin{bmatrix} \frac{1}{2} & -\frac{3}{5} \\ \frac{3}{4} & \frac{11}{10} \end{bmatrix}.$$

$\lambda = \frac{4}{5} - \frac{3}{5}i$ is an eigenvalue of A with corresponding eigenvector

$$\mathbf{v} = \begin{bmatrix} -2 - 4i \\ 5 \end{bmatrix}.$$

Show that A is similar to a real rotation matrix and interpret the transformation geometrically.

Solution

Separate the eigenvector into its real and imaginary parts:

$$\mathbf{v} = \mathbf{p} + i\mathbf{q},$$

where

$$\mathbf{p} = \mathrm{Re}(\mathbf{v}) = \begin{bmatrix} -2 \\ 5 \end{bmatrix}, \quad \mathbf{q} = \mathrm{Im}(\mathbf{v}) = \begin{bmatrix} -4 \\ 0 \end{bmatrix}.$$

Construct

$$P = \begin{bmatrix} \mathbf{p} & \mathbf{q} \end{bmatrix} = \begin{bmatrix} -2 & -4 \\ 5 & 0 \end{bmatrix}.$$

Since $\det(P) = 20 \neq 0$, the vectors $\mathbf{p}$ and $\mathbf{q}$ are linearly independent, so P is invertible.

Because $A\mathbf{v} = \lambda\mathbf{v}$,

$$A(\mathbf{p} + i\mathbf{q}) = \left(\frac{4}{5} - \frac{3}{5}i\right)(\mathbf{p} + i\mathbf{q}).$$

Expanding the left-hand side gives

$$A\mathbf{p} + iA\mathbf{q} = \left(\frac{4}{5}\mathbf{p} + \frac{3}{5}\mathbf{q}\right) + i\left(-\frac{3}{5}\mathbf{p} + \frac{4}{5}\mathbf{q}\right).$$

Equating the real and imaginary parts,

$$A\mathbf{p} = \frac{4}{5}\mathbf{p} + \frac{3}{5}\mathbf{q}, \quad A\mathbf{q} = -\frac{3}{5}\mathbf{p} + \frac{4}{5}\mathbf{q}.$$

Therefore,

$$\begin{aligned} AP &= \begin{bmatrix} A\mathbf{p} & A\mathbf{q} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{p} & \mathbf{q} \end{bmatrix} \begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix} \\ &= PC, \end{aligned}$$

where

$$C = \begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix}.$$

Since P is invertible, $C = P^{-1}AP$ or equivalently $A = PCP^{-1}$.

The matrix C is of the form

$$C = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix},$$

where

$$\cos\varphi = \frac{4}{5}, \quad \sin\varphi = \frac{3}{5}.$$

Thus, $\varphi = \tan^{-1}\left(\frac{3}{4}\right)$, and C represents a counterclockwise rotation through approximately $36.9°$. Since $\left|\frac{4}{5} - \frac{3}{5}i\right| = 1$, there is no scaling.

To understand the role of P, write $\mathbf{x} = P\mathbf{u}$. Then $A\mathbf{x} = AP\mathbf{u} = PC\mathbf{u}$. The transformation may be viewed as first converting $\mathbf{x}$ into the coordinate system determined by $\mathbf{p}$ and $\mathbf{q}$, applying the rotation C, and then returning to standard coordinates.

The rotation traces circles in the $\mathbf{u}$-coordinate system. However, because P is not orthogonal, these circles generally appear as ellipses in the standard coordinate system. Thus, A contains a rotation that becomes visible only *after* changing coordinates. See Figure 20 for a summary.

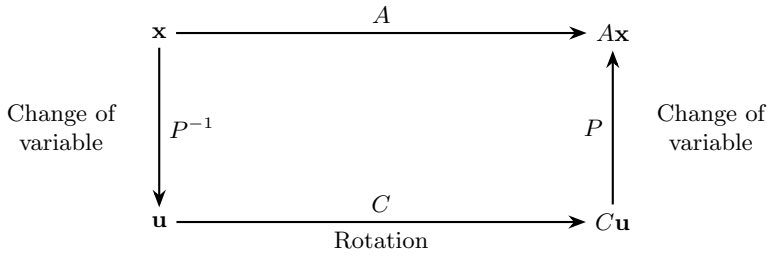


Figure 20: The transformation A is expressed in the new coordinates as the rotation due to complex eigenvalues $C = P^{-1}AP$.

Example 5.4.4 illustrates a result that applies to every real 2×2 matrix with nonreal eigenvalues. Although such a matrix cannot be diagonalized over $\mathbb{R}$, it can still be transformed into a real matrix that clearly displays its rotation and scaling.

Real Form of a Matrix with Complex Eigenvalues

Let A be a real 2×2 matrix with a complex eigenvalue

$$\lambda = a - bi, \qquad b \neq 0,$$

and let $\mathbf{v} \in \mathbb{C}^2 = \mathbf{p} + i\mathbf{q}$ be a corresponding eigenvector, where

$$\mathbf{p} = \mathrm{Re}(\mathbf{v}), \quad \mathbf{q} = \mathrm{Im}(\mathbf{v}).$$

Then $\mathbf{p}$ and $\mathbf{q}$ are linearly independent, so

$$P = \begin{bmatrix} \mathbf{p} & \mathbf{q} \end{bmatrix}$$

is invertible. Moreover,

$$A = PCP^{-1},$$

where

$$C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

The same structure may also appear as part of a larger matrix. For example,

$$A = \begin{bmatrix} \frac{4}{5} & -\frac{3}{5} & 0 \\ \frac{3}{5} & \frac{4}{5} & 0 \\ 0 & 0 & 1.07 \end{bmatrix}$$

rotates vectors in the x_1x_2-plane through the angle $\varphi = \tan^{-1}\left(\frac{3}{4}\right)$, while independently scaling the x_3-coordinate by 1.07. Thus, a real matrix may combine rotational behavior in one invariant

subspace with ordinary real scaling in another.

Problem 5.4.1

Let

$$A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix},$$

where $a, b \in \mathbb{R}$. Show that the eigenvalues of A are $a \pm bi$, with corresponding eigenvectors

$$\begin{bmatrix} 1 \\ -i \end{bmatrix}, \quad \begin{bmatrix} 1 \\ i \end{bmatrix}.$$

Solution

To verify the first eigenpair, compute

$$\begin{aligned} A\begin{bmatrix} 1 \\ -i \end{bmatrix} &= \begin{bmatrix} a & -b \\ b & a \end{bmatrix}\begin{bmatrix} 1 \\ -i \end{bmatrix} \\ &= \begin{bmatrix} a+bi \\ b-ai \end{bmatrix} \\ &= (a+bi)\begin{bmatrix} 1 \\ -i \end{bmatrix}. \end{aligned}$$

Therefore, $\lambda_1 = a + bi$ is an eigenvalue of A with corresponding eigenvector

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -i \end{bmatrix}.$$

Similarly,

$$\begin{aligned} A\begin{bmatrix} 1 \\ i \end{bmatrix} &= \begin{bmatrix} a & -b \\ b & a \end{bmatrix}\begin{bmatrix} 1 \\ i \end{bmatrix} \\ &= \begin{bmatrix} a-bi \\ b+ai \end{bmatrix} \\ &= (a-bi)\begin{bmatrix} 1 \\ i \end{bmatrix}. \end{aligned}$$

Therefore, $\lambda_2 = a - bi$ is an eigenvalue of A with corresponding eigenvector

$$\mathbf{v}_2 = \begin{bmatrix} 1 \\ i \end{bmatrix}.$$

5.5 Discrete Dynamical Systems

In the redwood and Douglas-fir forests of coastal California, dusky-footed woodrats are prey for spotted owls. We will use a very simplified model to study this discrete dynamical system.

(a) Spotted owl. Image credit: Heather Jensen, National Park Service.

(b) Dusky-footed woodrat. Image credit: Peterson B. Moose, U.S. Fish and Wildlife Service.

A **discrete dynamical system** describes how a state changes at equally spaced time steps. If $\mathbf{x}_k$ represents the state of the system at time k, then

$$\mathbf{x}_{k+1} = A\mathbf{x}_k.$$

Repeatedly applying the recurrence relation gives

$$\mathbf{x}_k = A^k\mathbf{x}_0.$$

The entries of $\mathbf{x}_k$ may represent populations, concentrations, economic quantities, or any other variables whose values evolve over time. Although the matrix A determines the system completely, its entries do not always make the long-term behavior immediately clear. Eigenvalues and eigenvectors separate the system into simpler modes.

Suppose A is diagonalizable with eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ and corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Since the eigenvectors form a basis, the initial state can be written uniquely as

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n.$$

Therefore,

$$\begin{aligned}\mathbf{x}_k &= A^k\mathbf{x}_0 \\ &= c_1\lambda_1^k\mathbf{v}_1 + c_2\lambda_2^k\mathbf{v}_2 + \cdots + c_n\lambda_n^k\mathbf{v}_n.\end{aligned}$$

Each eigenvector describes a mode of the system, while the corresponding eigenvalue determines how that mode changes.

- If $|\lambda_i| < 1$, the corresponding mode decays toward zero.
- If $|\lambda_i| = 1$, the magnitude of the corresponding mode remains constant.
- If $|\lambda_i| > 1$, the corresponding mode grows.

If one eigenvalue has strictly greater magnitude than all the others and its coefficient in $\mathbf{x}_0$ is nonzero, then its mode eventually dominates. This allows us to make approximations. For example, if

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|,$$

then, for sufficiently large k,

$$\mathbf{x}_k \approx c_1\lambda_1^k\mathbf{v}_1$$

and

$$\mathbf{x}_{k+1} \approx \lambda_1\mathbf{x}_k.$$

Thus, the dominant eigenvalue determines the eventual growth or decay rate, while the dominant eigenvector determines the eventual proportions among the entries of the state vector.

Complex eigenvalues fit naturally into this framework. A conjugate pair $a \pm bi$ produces oscillatory behavior, while its magnitude $\sqrt{a^2 + b^2}$ determines whether that behavior grows, decays, or remains constant.

In dynamical systems, the origin is called an attractor when every trajectory approaches it, a repeller when every nonzero trajectory moves away from it, and a saddle point when some directions approach the origin while others move away. These behaviors are determined by the magnitudes of the eigenvalues.

Example 5.5.1

Let

$$\mathbf{x}_k = \begin{bmatrix} O_k \\ R_k \end{bmatrix},$$

where O_k is the number of owls in a region after k months and R_k is the number of woodrats in thousands. Suppose the populations satisfy

$$O_{k+1} = 0.5O_k + 0.4R_k, \quad R_{k+1} = -0.104O_k + 1.1R_k.$$

Describe the long-term behavior of the system.

Solution

The system can be written as

$$\mathbf{x}_{k+1} = A\mathbf{x}_k, \quad A = \begin{bmatrix} 0.5 & 0.4 \\ -0.104 & 1.1 \end{bmatrix}.$$

The eigenvalues of A are $\lambda_1 = 1.02$ and $\lambda_2 = 0.58$, with corresponding eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} 10 \\ 13 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

Since the eigenvectors are linearly independent, any initial state can be written as $\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$.

Therefore,

$$\mathbf{x}_k = c_1(1.02)^k \begin{bmatrix} 10 \\ 13 \end{bmatrix} + c_2(0.58)^k \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

As $k \to \infty$, the factor $(0.58)^k$ approaches zero. Provided $c_1 \neq 0$, the first term eventually dominates, so

$$\mathbf{x}_k \approx c_1(1.02)^k \begin{bmatrix} 10 \\ 13 \end{bmatrix}.$$

Consequently,

$$\mathbf{x}_{k+1} \approx 1.02\mathbf{x}_k.$$

The model therefore predicts that both populations eventually grow by approximately 2% per month. The state vector also approaches the direction of $\mathbf{v}_1$, so

$$\frac{R_k}{O_k} \approx \frac{13}{10}.$$

Thus, the long-term population distribution approaches approximately 13 thousand rats for every 10 owls.

Example 5.5.2

Let

$$\mathbf{x}_k = \begin{bmatrix} j_k \\ s_k \\ a_k \end{bmatrix},$$

where j_k, s_k, and a_k are the numbers of female spotted owls in the juvenile, subadult, and adult life stages at time k. Consider the matrix

$$A = \begin{bmatrix} 0 & 0 & 0.33 \\ 0.18 & 0 & 0 \\ 0 & 0.71 & 0.94 \end{bmatrix},$$

which models how a population is distributed among different stages of life and how individuals move between those stages from one time step to the next.

Determine the long-term behavior predicted by the model. Then determine how the prediction changes if the juvenile survival entry 0.18 is increased to 0.30.

Solution

We will use MATLAB.

```
>> A = [0,    0,    0.33;
        0.18, 0,    0;
        0,    0.71, 0.94];

>> eig(A)

ans =

   -0.0218 + 0.2059i
   -0.0218 - 0.2059i
    0.9836 + 0.0000i
```

The magnitude of the complex pair is

$$\begin{aligned} |\lambda_{2,3}| &= \sqrt{(-0.0218)^2 + (0.2059)^2} \\ &\approx 0.2071. \end{aligned}$$

All three eigenvalues have magnitude less than one. Therefore, every mode in the eigenvector decomposition of $\mathbf{x}_0$ decays toward zero, and hence $\mathbf{x}_k \to \mathbf{0}$.

The model therefore predicts that the spotted owl population eventually disappears. The dominant eigenvalue 0.9836 is close to one, so the decline is slow, but it remains a decline.

Now we will replace the juvenile survival entry 0.18 with 0.30. The new matrix is

$$\widetilde{A} = \begin{bmatrix} 0 & 0 & 0.33 \\ 0.30 & 0 & 0 \\ 0 & 0.71 & 0.94 \end{bmatrix}.$$

```
>> A = [0,    0,    0.33;
        0.30, 0,    0;
        0,    0.71, 0.94];

>> eig(A)

ans =
```

```
-0.0345 + 0.2617i
-0.0345 - 0.2617i
 1.0090 + 0.0000i
```

The magnitude of the complex pair is

$$\begin{aligned} |\lambda_{2,3}| &= \sqrt{(-0.0345)^2 + (0.2617)^2} \\ &\approx 0.2639. \end{aligned}$$

Since this magnitude is less than one, the two oscillatory modes decay toward zero. However, the dominant eigenvalue satisfies $\lambda_1 > 1$. A corresponding eigenvector is approximately proportional to

$$\mathbf{v}_1 = \begin{bmatrix} 10 \\ 3 \\ 31 \end{bmatrix}.$$

For an initial population with a nonzero component in this direction,

$$\mathbf{x}_k \approx c_1 (1.0090)^k \begin{bmatrix} 10 \\ 3 \\ 31 \end{bmatrix}$$

for sufficiently large k. Thus,

$$\mathbf{x}_{k+1} \approx 1.0090 \mathbf{x}_k.$$

The modified model predicts slow long-term growth of approximately 0.9% per time step. It also predicts that the population eventually stabilizes at approximately 10 juveniles, 3 subadults, and 31 adults for every 44 owls in the population. In other words, although the total population continues to grow, the relative proportions of each life stage become nearly constant over time.

Thus, a relatively small change in one survival rate moves the dominant eigenvalue from slightly below one to slightly above one, changing the long-term prediction from eventual extinction to gradual population growth.

Problem 5.5.1

A herd of American bison can be modeled using a stage matrix similar to the spotted owl model. The female population is divided into three life stages:

- Calves (up to 1 year old)
- Yearlings (1 to 2 years old)
- Adults

Each year,

- each adult female produces an average of 0.42 female calves,
- 60% of calves survive to become yearlings,
- 75% of yearlings survive to become adults,
- 95% of adults survive to the next year.

Let

$$\mathbf{x}_k = \begin{bmatrix} c_k \\ y_k \\ a_k \end{bmatrix}$$

denote the population after k years.

1. Construct the stage matrix A satisfying $\mathbf{x}_{k+1} = A\mathbf{x}_k$.
2. Use MATLAB to compute the eigenvalues of A. Determine whether the buffalo population grows or declines in the long run, estimate the annual growth rate, and determine the expected numbers of calves and yearlings per 100 adults.

Solution

The number of calves in year $k+1$ depends on the number of adult females in year k. Since each adult female produces an average of 0.42 female calves,

$$c_{k+1} = 0.42a_k.$$

Since 60% of the calves survive to become yearlings,

$$y_{k+1} = 0.60c_k.$$

Finally, 75% of the yearlings survive to become adults, while 95% of the adults survive to the following year. Therefore,

$$a_{k+1} = 0.75y_k + 0.95a_k.$$

Thus,

$$\begin{bmatrix} c_{k+1} \\ y_{k+1} \\ a_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0.42 \\ 0.60 & 0 & 0 \\ 0 & 0.75 & 0.95 \end{bmatrix} \begin{bmatrix} c_k \\ y_k \\ a_k \end{bmatrix},$$

so the matrix is

$$A = \begin{bmatrix} 0 & 0 & 0.42 \\ 0.60 & 0 & 0 \\ 0 & 0.75 & 0.95 \end{bmatrix}.$$

Using MATLAB,

```
>> A = [0,    0,    0.42;
        0.60, 0,    0;
        0,    0.75, 0.95];

>> eig(A)
```

```
ans =

  -0.0774 + 0.4063i
  -0.0774 - 0.4063i
   1.1048 + 0.0000i
```

The complex eigenvalues have magnitude $\sqrt{(0.0774)^2 + (0.4063)^2} = 0.4136$, so the modes associated with the complex eigenvalues decay. The dominant eigenvalue is

$$\lambda_1 \approx 1.1048.$$

Since $\lambda_1 > 1$, the buffalo population grows in the long run. After many years, the population is multiplied by approximately 1.1048 each year, corresponding to an annual growth rate of approximately

$$(1.1048 - 1)(100\%) \approx 10.5\%.$$

A corresponding eigenvector for the dominant eigenvalue is approximately

$$\mathbf{v}_1 = \begin{bmatrix} 0.3801 \\ 0.2064 \\ 1 \end{bmatrix}.$$

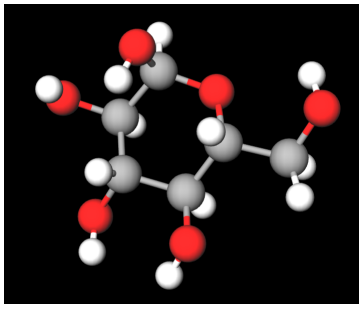
The third entry has been normalized to 1, so the vector describes the long-term numbers of calves and yearlings per adult. Multiplying by 100 gives

$$100\mathbf{v}_1 = \begin{bmatrix} 38.01 \\ 20.64 \\ 100 \end{bmatrix}.$$

Therefore, for every 100 adult females, the model predicts approximately 38 calves and 21 yearlings. Although the total herd continues to grow, these relative numbers will eventually remain nearly constant.

Birds will perch on the backs of bison.
Image credit: Kim Acker, NPS

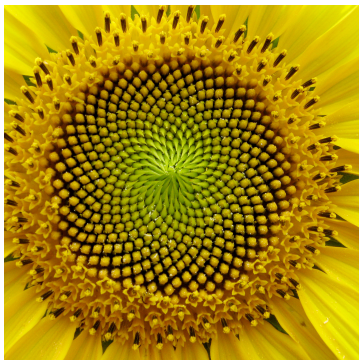
Eight bison walking by the Grand Prismatic Spring.
Image credit: Yueru Hao



A glucose molecule.
Image credit: PubChem

A bison and calf at Rocky Mountain Arsenal National Wildlife Refuge in Colorado.
Image credit: Rich Keen

Bison standing in the snow at the National Elk Refuge in Wyoming.
Image credit: U.S. Fish and Wildlife Service



The spiral arrangement of seeds in a sunflower represents the Fibonacci sequence.
Image credit: Herbert F. Johnson Museum of Art

Figure 22: Our chosen applications of discrete dynamical systems

Problem 5.5.2

Between meals, the glucose concentration and excess hormone concentration in the blood are modeled by the state vector

$$\mathbf{x}_k = \begin{bmatrix} g_k \\ h_k \end{bmatrix},$$

where g_k is the glucose concentration and h_k is the excess hormone concentration after k time intervals.

During each time interval,

- 97.8% of the current glucose concentration remains, while the excess hormone decreases the next glucose concentration by an amount equal to 0.6% of its current value;
- 99.2% of the current excess hormone concentration remains, while the current glucose concentration increases the next hormone concentration by an amount equal to 0.4% of its current value.

1. Write equations for g_{k+1} and h_{k+1} in terms of g_k and h_k.
2. Construct a matrix A such that $\mathbf{x}_{k+1} = A\mathbf{x}_k$.
3. Verify that

$$\begin{bmatrix} -1 \\ 2 \end{bmatrix}, \quad \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

   are eigenvectors of A, and find their corresponding eigenvalues.
4. Use the eigenvalues to describe the long-term behavior of the glucose and excess hormone concentrations.

Solution

The next glucose concentration is determined by the glucose and excess hormone concentrations during the current time interval. Since 97.8% of the glucose remains and the hormone decreases the next glucose concentration by 0.6% of its current value,

$$g_{k+1} = 0.978g_k - 0.006h_k.$$

Similarly, 99.2% of the excess hormone remains, while the current glucose concentration increases the next hormone concentration by 0.4% of its current value. Therefore,

$$h_{k+1} = 0.004g_k + 0.992h_k.$$

We have

$$\begin{bmatrix} g_{k+1} \\ h_{k+1} \end{bmatrix} = \begin{bmatrix} 0.978 & -0.006 \\ 0.004 & 0.992 \end{bmatrix} \begin{bmatrix} g_k \\ h_k \end{bmatrix},$$

so

$$A = \begin{bmatrix} 0.978 & -0.006 \\ 0.004 & 0.992 \end{bmatrix}.$$

To verify the first eigenvector, compute

$$\begin{aligned} A \begin{bmatrix} -1 \\ 2 \end{bmatrix} &= \begin{bmatrix} 0.978 & -0.006 \\ 0.004 & 0.992 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} -0.978 - 0.012 \\ -0.004 + 1.984 \end{bmatrix} \\ &= \begin{bmatrix} -0.990 \\ 1.980 \end{bmatrix} \\ &= 0.99 \begin{bmatrix} -1 \\ 2 \end{bmatrix}. \end{aligned}$$

Therefore,

$$\begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

is an eigenvector corresponding to $\lambda_1 = 0.99$.

For the second vector,

$$\begin{aligned} A \begin{bmatrix} 3 \\ -1 \end{bmatrix} &= \begin{bmatrix} 0.978 & -0.006 \\ 0.004 & 0.992 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix} \\ &= \begin{bmatrix} 2.934 + 0.006 \\ 0.012 - 0.992 \end{bmatrix} \\ &= \begin{bmatrix} 2.940 \\ -0.980 \end{bmatrix} \\ &= 0.98 \begin{bmatrix} 3 \\ -1 \end{bmatrix}. \end{aligned}$$

Therefore,

$$\begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

is an eigenvector corresponding to $\lambda_2 = 0.98$.

Since the two eigenvectors are linearly independent, any initial state can be written as

$$\mathbf{x}_0 = c_1 \begin{bmatrix} -1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 3 \\ -1 \end{bmatrix}.$$

After k time intervals,

$$\mathbf{x}_k = c_1(0.99)^k \begin{bmatrix} -1 \\ 2 \end{bmatrix} + c_2(0.98)^k \begin{bmatrix} 3 \\ -1 \end{bmatrix}.$$

Because both eigenvalues have magnitude less than one, $(0.99)^k \to 0$ and $(0.98)^k \to 0$. Therefore,

$$\mathbf{x}_k \to \mathbf{0}.$$

The model predicts that both the glucose concentration and excess hormone concentration gradually return to their equilibrium levels between meals. Since 0.99 is the dominant eigenvalue, the component in the direction of

$$\begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

decays more slowly and eventually governs the path to equilibrium.

Problem 5.5.3

The Fibonacci sequence is one of the most famous sequences in mathematics. It begins with

$$x_0 = 0, \quad x_1 = 1,$$

and each subsequent term is obtained by adding the two previous terms:

$$x_{n+1} = x_n + x_{n-1}, \qquad n \geq 1$$

The first several terms are

$$0,\ 1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ 34,\ \ldots$$

Although this recurrence is usually studied in number theory, it can also be viewed as a discrete dynamical system by tracking two consecutive terms at a time.

Define the state vector

$$\mathbf{x}_n = \begin{bmatrix} x_n \\ x_{n-1} \end{bmatrix},$$

and construct a matrix A such that $\mathbf{x}_{n+1} = A\mathbf{x}_n$.

Then, find the eigenvalues and eigenvectors of A. Finally, diagonalize A and use your result to derive a closed-form expression for the nth Fibonacci number.

Solution

Since $x_{n+1} = x_n + x_{n-1}$, the state vector satisfies

$$\begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}}_{A} \begin{bmatrix} x_n \\ x_{n-1} \end{bmatrix}.$$

The characteristic equation is

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{vmatrix} \\ &= (1-\lambda)(-\lambda) - 1 \\ &= \lambda^2 - \lambda - 1. \end{aligned}$$

Therefore $\lambda^2 - \lambda - 1 = 0$, so the eigenvalues are

$$\lambda_1 = \frac{1+\sqrt{5}}{2}, \quad \lambda_2 = \frac{1-\sqrt{5}}{2}.$$

Let

$$\phi = \frac{1+\sqrt{5}}{2}, \qquad \psi = \frac{1-\sqrt{5}}{2}.$$

For an eigenvalue λ, we have $(A - \lambda I)\mathbf{v} = \mathbf{0}$ and thus

$$\begin{bmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \mathbf{0}.$$

From the second equation, $v_1 = \lambda v_2$. Choosing $v_2 = 1$ gives the eigenvector

$$\mathbf{v} = \begin{bmatrix} \lambda \\ 1 \end{bmatrix}.$$

Thus, corresponding eigenvectors are

$$\mathbf{v}_1 = \begin{bmatrix} \phi \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} \psi \\ 1 \end{bmatrix}.$$

Therefore, $A = PDP^{-1}$ where

$$P = \begin{bmatrix} \phi & \psi \\ 1 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} \phi & 0 \\ 0 & \psi \end{bmatrix}.$$

Since

$$\mathbf{x}_1 = \begin{bmatrix} x_1 \\ x_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

we will write the initial state as a linear combination of the eigenvectors:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = c_1 \begin{bmatrix} \phi \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} \psi \\ 1 \end{bmatrix}.$$

This gives the system

$$\begin{cases} \phi c_1 + \psi c_2 = 1, \\ c_1 + \quad c_2 = 0. \end{cases}$$

Since $c_2 = -c_1$, we have $c_1(\phi - \psi) = 1$. And because $\phi - \psi = \sqrt{5}$, we obtain

$$c_1 = \frac{1}{\sqrt{5}}, \quad c_2 = -\frac{1}{\sqrt{5}}.$$

After $n - 1$ iterations,

$$\begin{aligned}\mathbf{x}_n &= A^{n-1}\mathbf{x}_1 \\ &= \frac{1}{\sqrt{5}}\phi^{n-1}\begin{bmatrix}\phi\\1\end{bmatrix} - \frac{1}{\sqrt{5}}\psi^{n-1}\begin{bmatrix}\psi\\1\end{bmatrix} \\ &= \frac{1}{\sqrt{5}}\begin{bmatrix}\phi^n - \psi^n\\ \phi^{n-1} - \psi^{n-1}\end{bmatrix}.\end{aligned}$$

The first entry is x_n. Therefore, the nth Fibonacci number is

$$x_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right].$$

This formula is known as Binet's formula.

5.6 Applications to Differential Equations

Linear differential equations arise whenever a quantity changes continuously with time. Examples include vibrating structures, electrical circuits, fluid mechanics, heat transfer, chemical reactions, and control systems.

A system of n first-order linear differential equations can be written compactly as

$$\mathbf{x}'(t) = A\mathbf{x}(t), \tag{27}$$

where A is a constant $n \times n$ matrix and $\mathbf{x}(t)$ is a vector whose components describe the state of the system at time t.

Eigenvalues and eigenvectors provide a direct way to solve these systems. The key observation comes from a familiar equation in single-variable calculus.

Review from Calculus

The differential equation

$$\frac{du}{dt} = \lambda u$$

has the solution

$$u(t) = Ce^{\lambda t},$$

where C is determined by the initial condition.

This suggests that exponential functions may also appear when solving matrix differential equations. Since the unknown is now a vector rather than a scalar, we naturally guess a solution of the form

$$\mathbf{x}(t) = e^{\lambda t}\mathbf{v},$$

where $\mathbf{v}$ is a constant vector and λ is an unknown scalar.

Substituting this expression into Equation 27 gives

$$\lambda e^{\lambda t}\mathbf{v} = Ae^{\lambda t}\mathbf{v}.$$

Since $e^{\lambda t} \neq 0$ for every t, we may divide both sides by the exponential to obtain

$$A\mathbf{v} = \lambda\mathbf{v},$$

which is the eigenvalue equation.

Eigenvalues and Differential Equations

A solution of the differential equation

$$\mathbf{x}'(t) = A\mathbf{x}(t)$$

can be written in the form

$$\mathbf{x}(t) = e^{\lambda t}\mathbf{v}$$

if and only if

$$A\mathbf{v} = \lambda\mathbf{v}.$$

Thus, every eigenvalue-eigenvector pair of A produces a solution of the differential equation.

If A has n linearly independent eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then linearity allows these individual solutions to be combined into the general solution

$$\mathbf{x}(t) = c_1 e^{\lambda_1 t}\mathbf{v}_1 + c_2 e^{\lambda_2 t}\mathbf{v}_2 + \cdots + c_n e^{\lambda_n t}\mathbf{v}_n,$$

where the constants $c_1, \ldots, c_n$ are determined by the initial condition.

This result closely mirrors what we saw for repeated matrix multiplication. In discrete dynamical systems,

$$\mathbf{x}_k = c_1\lambda_1^k\mathbf{v}_1 + \cdots + c_n\lambda_n^k\mathbf{v}_n.$$

For differential equations, the powers λ_i^k are simply replaced by exponential functions $e^{\lambda_i t}$. The eigenvectors determine the independent modes of motion, while the eigenvalues determine how each mode evolves over time.

Example 5.6.1

Solve the initial value problem

$$\frac{d\mathbf{u}}{dt} = A\mathbf{u}, \quad A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{u}(0) = \begin{bmatrix} 4 \\ 2 \end{bmatrix}.$$

Solution

We seek solutions of the form

$$\mathbf{u}(t) = e^{\lambda t}\mathbf{x},$$

where λ is an eigenvalue of A and $\mathbf{x}$ is the corresponding eigenvector.

The characteristic polynomial is

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} -\lambda & 1 \\ 1 & -\lambda \end{vmatrix} \\ &= \lambda^2 - 1. \end{aligned}$$

Therefore,

$$\lambda_1 = 1, \quad \lambda_2 = -1.$$

For $\lambda = 1$, we get

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

For $\lambda = -1$, we get

$$\mathbf{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Thus, the general solution is

$$\mathbf{u}(t) = C_1 e^{t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + C_2 e^{-t} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Apply the initial condition to get

$$\begin{bmatrix} 4 \\ 2 \end{bmatrix} = C_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + C_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

which represents

$$\begin{cases} C_1 + C_2 = 4, \\ C_1 - C_2 = 2. \end{cases}$$

Solving, we get

$$C_1 = 3, \quad C_2 = 1.$$

Therefore,

$$\mathbf{u}(t) = 3e^{t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + e^{-t} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Example 5.6.2

Solve the initial value problem

$$y'' + y = 0, \quad y(0) = 1, \quad y'(0) = 0.$$

Solution

Since this is a second-order differential equation, we first rewrite it as a first-order system by defining the state vector

$$\mathbf{u} = \begin{bmatrix} y \\ y' \end{bmatrix}.$$

Then,

$$\mathbf{u}' = \begin{bmatrix} y' \\ y'' \end{bmatrix}.$$

Using the differential equation,

$$y'' = -y,$$

we get

$$\mathbf{u}' = \begin{bmatrix} y' \\ -y \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}_{A} \mathbf{u}.$$

Thus, we seek solutions of the form

$$\mathbf{u}(t) = e^{\lambda t}\mathbf{x},$$

where λ is an eigenvalue of A and $\mathbf{x}$ is the corresponding eigenvector.

The characteristic polynomial is

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} -\lambda & 1 \\ -1 & -\lambda \end{vmatrix} \\ &= \lambda^2 + 1. \end{aligned}$$

Therefore,

$$\lambda_1 = i, \quad \lambda_2 = -i.$$

For $\lambda = i$, we get

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ i \end{bmatrix}.$$

Similarly, for $\lambda = -i$,

$$\mathbf{x}_2 = \begin{bmatrix} 1 \\ -i \end{bmatrix}.$$

Thus, the general solution is

$$\mathbf{u}(t) = C_1 e^{it} \begin{bmatrix} 1 \\ i \end{bmatrix} + C_2 e^{-it} \begin{bmatrix} 1 \\ -i \end{bmatrix}.$$

Applying the initial condition,

$$\mathbf{u}(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

we obtain

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = C_1 \begin{bmatrix} 1 \\ i \end{bmatrix} + C_2 \begin{bmatrix} 1 \\ -i \end{bmatrix},$$

which represents

$$\begin{cases} C_1 + C_2 = 1, \\ iC_1 - iC_2 = 0. \end{cases}$$

Solving, we get

$$C_1 = C_2 = \frac{1}{2},$$

Substitute into the general solution:

$$\mathbf{u}(t) = \frac{1}{2}e^{it}\begin{bmatrix}1\\ i\end{bmatrix} + \frac{1}{2}e^{-it}\begin{bmatrix}1\\ -i\end{bmatrix}.$$

Use Euler's formulas,

$$e^{it} = \cos t + i\sin t, \quad e^{-it} = \cos t - i\sin t,$$

to obtain

$$\mathbf{u}(t) = \begin{bmatrix}\cos t\\ -\sin t\end{bmatrix}.$$

Since the first component of $\mathbf{u}(t)$ is $y(t)$,

$$y(t) = \cos t.$$

Problem 5.6.1

Solve the coupled differential equations

$$\frac{dx}{dt} = 4x + y, \quad \frac{dy}{dt} = 2x + 3y,$$

subject to the initial conditions

$$x(0) = 2, \quad y(0) = -1.$$

Solution

Writing the system in matrix form gives

$$\frac{d}{dt}\begin{bmatrix}x\\ y\end{bmatrix} = \underbrace{\begin{bmatrix}4 & 1\\ 2 & 3\end{bmatrix}}_{A}\begin{bmatrix}x\\ y\end{bmatrix}.$$

The characteristic polynomial of A is

$$\begin{aligned}\det(A-\lambda I) &= \begin{vmatrix} 4-\lambda & 1 \\ 2 & 3-\lambda \end{vmatrix} \\ &= (4-\lambda)(3-\lambda)-2 \\ &= \lambda^2 - 7\lambda + 10 \\ &= (\lambda-5)(\lambda-2).\end{aligned}$$

Thus, the eigenvalues are

$$\lambda_1 = 5, \quad \lambda_2 = 2.$$

For $\lambda_1 = 5$,

$$(A-5I)\mathbf{v}_1 = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} \mathbf{v}_1 = \mathbf{0},$$

so one corresponding eigenvector is

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

For $\lambda_2 = 2$,

$$(A-2I)\mathbf{v}_2 = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \mathbf{v}_2 = \mathbf{0},$$

so one corresponding eigenvector is

$$\mathbf{v}_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

Therefore, the general solution is

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = C_1 e^{5t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + C_2 e^{2t} \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

Using the initial conditions,

$$\begin{bmatrix} 2 \\ -1 \end{bmatrix} = C_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + C_2 \begin{bmatrix} 1 \\ -2 \end{bmatrix},$$

which represents

$$\begin{cases} C_1 + \ C_2 = 2, \\ C_1 - 2C_2 = -1. \end{cases}$$

Solving the system gives

$$C_1 = 1, \quad C_2 = 1.$$

Thus,

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = e^{5t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + e^{2t} \begin{bmatrix} 1 \\ -2 \end{bmatrix},$$

and therefore,

$$x(t) = e^{5t} + e^{2t}, \quad y(t) = e^{5t} - 2e^{2t}.$$

6 Orthogonality

6.1 Inner Product; Orthogonal Complement, Sets, and Projections

In Sections 1.3 and 1.4, we introduced many of the geometric tools needed for linear algebra. While they have been implicitly used throughout the entire guide, we now return to these ideas because they provide the foundation for this chapter. Recall the inner product:

Inner Product

Let

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

be vectors in $\mathbb{R}^n$. The **inner product** of $\mathbf{u}$ and $\mathbf{v}$ is

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{j=1}^{n} u_j v_j.$$

The inner product is a function

$$\mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}.$$

In Euclidean space, the inner product is the familiar dot product, so the two terms will be used interchangeably throughout this chapter.

Recall the condition for two vectors to be considered orthogonal:

Orthogonal Vectors

Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are **orthogonal** if $\mathbf{u} \cdot \mathbf{v} = 0$.

In $\mathbb{R}^2$ and $\mathbb{R}^3$, orthogonal vectors are perpendicular. The inner product extends this notion of

perpendicularity to $\mathbb{R}^n$, where a direct geometric picture may no longer be available.

The zero vector is orthogonal to every vector because

$$\mathbf{0} \cdot \mathbf{x} = 0$$

for every $\mathbf{x} \in \mathbb{R}^n$.

We may also ask whether a vector is orthogonal not merely to one vector, but to every vector in an entire subspace.

Orthogonal Complement

Let W be a subspace of $\mathbb{R}^n$. A vector $\mathbf{x} \in \mathbb{R}^n$ is **orthogonal to** W if $\mathbf{x} \cdot \mathbf{w} = 0$ for every $\mathbf{w} \in W$.

The set of all vectors orthogonal to W is called the **orthogonal complement** of W and is denoted by $W^\perp$. Thus,

$$W^\perp = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \cdot \mathbf{w} = 0 \text{ for every } \mathbf{w} \in W\}.$$

For example, suppose $W \in \mathbb{R}^3$ is a plane through the origin in and L is a line through the origin that is perpendicular to W. Then $L = W^\perp$. Let $\mathbf{w} \in W$, and $\mathbf{z} \in W^\perp$. Since $W^\perp$ is perpendicular to every vector in W, we have $\mathbf{w} \cdot \mathbf{z} = 0$. See Figure 23.

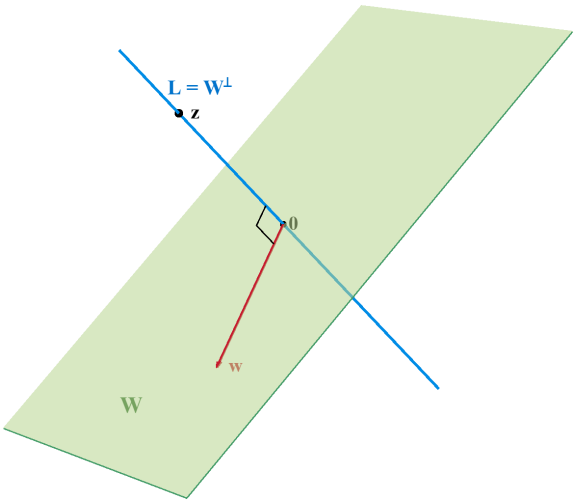


Figure 23: A plane and the line through the origin that forms its orthogonal complement in three-dimensional space.

At first, the definition of $W^\perp$ appears to require checking infinitely many vectors in W. In practice, it is enough to check only a spanning set.

Orthogonality to a Spanning Set

Let $W = \text{Span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_p\}$. Then a vector $\mathbf{x} \in W^\perp$ if and only if

$$\mathbf{x} \cdot \mathbf{w}_j = 0$$

for every $j = 1, 2, \ldots, p$. Moreover, $W^\perp$ is a subspace of $\mathbb{R}^n$.

Proof

Suppose first that $\mathbf{x} \in W^\perp$. Since every vector $\mathbf{w}_j \in W$, it follows directly that

$$\mathbf{x} \cdot \mathbf{w}_j = 0$$

for every $j = 1, 2, \ldots, p$.

Conversely, suppose that

$$\mathbf{x} \cdot \mathbf{w}_j = 0$$

for every j. Any vector $\mathbf{w} \in W$ may be written as $\mathbf{w} = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_p\mathbf{w}_p$.

Using linearity of the inner product,

$$\begin{aligned}\mathbf{x} \cdot \mathbf{w} &= \mathbf{x} \cdot (c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_p\mathbf{w}_p) \\ &= c_1(\mathbf{x} \cdot \mathbf{w}_1) + c_2(\mathbf{x} \cdot \mathbf{w}_2) + \cdots + c_p(\mathbf{x} \cdot \mathbf{w}_p) \\ &= 0.\end{aligned}$$

Therefore, $\mathbf{x}$ is orthogonal to every vector in W, so $\mathbf{x} \in W^\perp$.

Now form the matrix

$$B = \begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_p\end{bmatrix}.$$

The equations $\mathbf{x} \cdot \mathbf{w}_j = 0$ for every j are equivalent to

$$B^T\mathbf{x} = \mathbf{0}.$$

Hence,

$$W^\perp = \text{Nul}(B^T).$$

Since the null space of a matrix is a subspace, $W^\perp$ is a subspace of $\mathbb{R}^n$.

Example 6.1.1

Let

$$W = \text{Span}\left\{\begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}\right\}.$$

Find $W^\perp$.

Solution

Let

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in W^{\perp}.$$

Then $\mathbf{x}$ must be orthogonal to both vectors that span W. Therefore,

$$\mathbf{x} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 0, \quad \mathbf{x} \cdot \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = 0.$$

These equations give

$$x_1 + x_2 = 0, \qquad x_2 + x_3 = 0.$$

Let $x_2 = t$. Then

$$x_1 = -t, \qquad x_3 = -t.$$

Thus,

$$\mathbf{x} = t \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}.$$

Therefore,

$$W^{\perp} = \text{Span} \left\{ \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} \right\}.$$

Properties of Orthogonal Complements

Let A be an $m \times n$ matrix. Then the following statements are true.

(i) $\text{Row}(A)^{\perp} = \text{Nul}(A)$, $\quad \text{Nul}(A)^{\perp} = \text{Row}(A)$

(ii) $\text{Col}(A)^{\perp} = \text{Nul}(A^T)$, $\quad \text{Nul}(A^T)^{\perp} = \text{Col}(A)$

(iii) Moreover, if W is a subspace of $\mathbb{R}^n$, then $\dim(W) + \dim(W^{\perp}) = n$.

Proof

(i) Let the rows of A be $\mathbf{r}_1^T, \mathbf{r}_2^T, \ldots, \mathbf{r}_m^T$.

Then

$$A = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \vdots \\ \mathbf{r}_m^T \end{bmatrix}.$$

A vector $\mathbf{x} \in \mathbb{R}^n$ belongs to $\text{Nul}(A)$ if and only if

$$A\mathbf{x} = \begin{bmatrix} \mathbf{r}_1^T\mathbf{x} \\ \mathbf{r}_2^T\mathbf{x} \\ \vdots \\ \mathbf{r}_m^T\mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1 \cdot \mathbf{x} \\ \mathbf{r}_2 \cdot \mathbf{x} \\ \vdots \\ \mathbf{r}_m \cdot \mathbf{x} \end{bmatrix} = \mathbf{0}.$$

Thus, $\mathbf{x} \in \text{Nul}(A)$ if and only if $\mathbf{r}_j \cdot \mathbf{x} = 0$ for every $j = 1, 2, \ldots, m$.

Since the rows of A span $\text{Row}(A)$, this is equivalent to $\mathbf{x}$ being orthogonal to every vector in $\text{Row}(A)$. Therefore,

$$\text{Row}(A)^{\perp} = \text{Nul}(A).$$

To prove the reverse relationship, observe that every row of A is orthogonal to every vector in $\text{Nul}(A)$. Therefore,

$$\text{Row}(A) \subseteq \text{Nul}(A)^{\perp}.$$

By the rank-nullity theorem,

$$\dim\left(\text{Nul}(A)\right) = n - \text{rank}(A).$$

Hence,

$$\begin{aligned}\dim\left(\text{Nul}(A)^{\perp}\right) &= n - \dim\left(\text{Nul}(A)\right) \\ &= n - [n - \text{rank}(A)] \\ &= \text{rank}(A) \\ &= \dim\left(\text{Row}(A)\right).\end{aligned}$$

Since $\text{Row}(A)$ is contained in $\text{Nul}(A)^{\perp}$ and the two subspaces have the same dimension,

$$\text{Nul}(A)^{\perp} = \text{Row}(A).$$

(ii) A vector $\mathbf{y} \in \mathbb{R}^m$ belongs to $\text{Nul}(A^T)$ if and only if $A^T\mathbf{y} = \mathbf{0}$.

Let the columns of A be $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$. Then

$$A = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n\end{bmatrix},$$

so

$$A^T\mathbf{y} = \begin{bmatrix}\mathbf{a}_1^T\mathbf{y} \\ \mathbf{a}_2^T\mathbf{y} \\ \vdots \\ \mathbf{a}_n^T\mathbf{y}\end{bmatrix} = \begin{bmatrix}\mathbf{a}_1 \cdot \mathbf{y} \\ \mathbf{a}_2 \cdot \mathbf{y} \\ \vdots \\ \mathbf{a}_n \cdot \mathbf{y}\end{bmatrix}.$$

Therefore, $A^T\mathbf{y} = \mathbf{0}$ if and only if $\mathbf{a}_j \cdot \mathbf{y} = 0$ for every $j = 1, 2, \ldots, n$.

Since the columns of A span $\text{Col}(A)$, this means that $\mathbf{y}$ is orthogonal to every vector in $\text{Col}(A)$. Thus,

$$\text{Col}(A)^{\perp} = \text{Nul}(A^T).$$

To prove the reverse relationship, observe that every column of A is orthogonal to every vector in $\text{Nul}(A^T)$. Therefore

$$\text{Col}(A) \subseteq \text{Nul}(A^T)^\perp.$$

Since A^T has m columns, the rank-nullity theorem gives

$$\dim\left(\text{Nul}(A^T)\right) = m - \text{rank}(A^T).$$

Because $\text{rank}(A^T) = \text{rank}(A)$, we have

$$\begin{aligned}\dim\left(\text{Nul}(A^T)^\perp\right) &= m - \dim\left(\text{Nul}(A^T)\right) \\ &= m - [m - \text{rank}(A)] \\ &= \text{rank}(A) \\ &= \dim\left(\text{Col}(A)\right).\end{aligned}$$

Since $\text{Col}(A)$ is contained in $\text{Nul}(A^T)^\perp$ and the two subspaces have the same dimension,

$$\text{Nul}(A^T)^\perp = \text{Col}(A).$$

(iii) Finally, let W be a subspace of $\mathbb{R}^n$ with basis $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_p\}$.

Form the matrix

$$B = \begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_p\end{bmatrix}.$$

Because the columns of B form a basis for W,

$$\text{Col}(B) = W$$

and

$$\text{rank}(B) = p = \dim(W).$$

From part (ii),

$$W^\perp = \text{Col}(B)^\perp = \text{Nul}(B^T).$$

The matrix B^T has n columns, so the rank-nullity theorem gives

$$\begin{aligned}\dim(W^\perp) &= \dim\left(\mathrm{Nul}(B^T)\right)\\ &= n - \mathrm{rank}(B^T)\\ &= n - \mathrm{rank}(B)\\ &= n - \dim(W).\end{aligned}$$

Therefore,

$$\dim(W) + \dim(W^\perp) = n.$$

□

Example 6.1.2

Find the orthogonal complement of the eigenspace corresponding to the eigenvalue $\lambda = 5$ for the matrix

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

Solution

The eigenspace of A is $W = \mathrm{Nul}(A - 5I)$.

Since

$$A - 5I = \begin{bmatrix} -3 & 4 & -1 \\ 3 & -3 & 0 \\ -2 & 4 & -2 \end{bmatrix},$$

we have

$$W = \text{Nul}\left(\begin{bmatrix} -3 & 4 & -1 \\ 3 & -3 & 0 \\ -2 & 4 & -2 \end{bmatrix}\right).$$

Because the orthogonal complement of a null space is the corresponding row space,

$$W^\perp = \text{Nul}(A - 5I)^\perp = \text{Row}(A - 5I).$$

Thus, the rows of $A - 5I$ span $W^\perp$:

$$W^\perp = \text{Span}\left\{\begin{bmatrix} -3 \\ 4 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ -3 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 4 \\ -2 \end{bmatrix}\right\}$$

However, these three vectors are linearly dependent. To obtain a basis, row reduce $A - 5I$:

$$\begin{bmatrix} -3 & 4 & -1 \\ 3 & -3 & 0 \\ -2 & 4 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Elementary row operations preserve the row space, so the nonzero rows of the RREF form a basis for $W^\perp$. Therefore,

$$W^\perp = \text{Span}\left\{\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}\right\}.$$

To understand the geometry, observe that

$$W = \text{Nul}(A - 5I) = \text{Span}\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right\}.$$

Thus, W is a line in $\mathbb{R}^3$, while $W^\perp$ is the plane perpendicular to that line. Indeed,

$$\dim(W^\perp) = 3 - \dim(W) = 3 - 1 = 2.$$

An orthogonal set extends the idea of two orthogonal vectors to an entire collection. Every pair of distinct vectors in the set must be orthogonal.

Orthogonal Set

Let

$$S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$$

be a set of vectors in $\mathbb{R}^n$. The set S is an **orthogonal set** if

$$\mathbf{u}_i \cdot \mathbf{u}_j = 0$$

whenever $i \neq j$.

Example 6.1.3

Show that $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set, where

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -\frac{1}{2} \\ -2 \\ \frac{7}{2} \end{bmatrix}.$$

Solution

There are three possible pairs of distinct vectors. Computing their inner products gives

$$\begin{aligned}
\mathbf{u}_1 \cdot \mathbf{u}_2 &= 3(-1) + 1(2) + 1(1) = 0, \\
\mathbf{u}_1 \cdot \mathbf{u}_3 &= 3\left(-\frac{1}{2}\right) + 1(-2) + 1\left(\frac{7}{2}\right) = 0, \\
\mathbf{u}_2 \cdot \mathbf{u}_3 &= (-1)\left(-\frac{1}{2}\right) + 2(-2) + 1\left(\frac{7}{2}\right) = 0.
\end{aligned}$$

Each pair of distinct vectors is orthogonal. Therefore, $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set.

Orthogonality imposes a strong relationship among the vectors in a set. In particular, nonzero orthogonal vectors cannot be linearly dependent.

Linear Independence of Orthogonal Sets

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ be an orthogonal set of nonzero vectors in $\mathbb{R}^n$. Then S is linearly independent and therefore forms a basis for $\text{Span}(S)$.

Proof

Suppose

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p = \mathbf{0}.$$

Take the inner product of both sides with $\mathbf{u}_j$. Then

$$(c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p) \cdot \mathbf{u}_j = \mathbf{0} \cdot \mathbf{u}_j.$$

Using linearity,

$$c_1(\mathbf{u}_1 \cdot \mathbf{u}_j) + c_2(\mathbf{u}_2 \cdot \mathbf{u}_j) + \cdots + c_p(\mathbf{u}_p \cdot \mathbf{u}_j) = 0.$$

Because the set is orthogonal, every term vanishes except the term containing $\mathbf{u}_j$. Therefore,

$$c_j(\mathbf{u}_j \cdot \mathbf{u}_j) = 0.$$

Since $\mathbf{u}_j \neq \mathbf{0}$,

$$\mathbf{u}_j \cdot \mathbf{u}_j = \|\mathbf{u}_j\|^2 > 0,$$

so $c_j = 0$. This argument applies to every $j = 1, 2, \ldots, p$. Hence,

$$c_1 = c_2 = \cdots = c_p = 0,$$

and S is linearly independent. □

Orthogonal Basis

Let W be a subspace of $\mathbb{R}^n$. An **orthogonal basis** for W is a basis for W that is also an orthogonal set.

An orthogonal basis makes coordinate calculations a lot easier. Ordinarily, finding the coefficients of a vector relative to a basis requires solving a system of equations. For an orthogonal basis, each coefficient may instead be computed independently using an inner product.

Coordinates in an Orthogonal Basis

Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. If $\mathbf{y} \in W$, then

$$\mathbf{y} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p,$$

where

$$c_j = \frac{\mathbf{y} \cdot \mathbf{u}_j}{\mathbf{u}_j \cdot \mathbf{u}_j}, \qquad j = 1, 2, \ldots, p.$$

Proof

Since the vectors form a basis for W, any vector $\mathbf{y} \in W$ may be written uniquely as

$$\mathbf{y} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p.$$

Take the inner product of both sides with $\mathbf{u}_j$:

$$\begin{aligned}\mathbf{y} \cdot \mathbf{u}_j &= (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p) \cdot \mathbf{u}_j \\ &= c_1(\mathbf{u}_1 \cdot \mathbf{u}_j) + c_2(\mathbf{u}_2 \cdot \mathbf{u}_j) + \cdots + c_p(\mathbf{u}_p \cdot \mathbf{u}_j).\end{aligned}$$

Because the basis is orthogonal, every term vanishes except the term containing $\mathbf{u}_j$. Thus,

$$\mathbf{y} \cdot \mathbf{u}_j = c_j(\mathbf{u}_j \cdot \mathbf{u}_j).$$

Since $\mathbf{u}_j \neq \mathbf{0}$, we may divide by $\mathbf{u}_j \cdot \mathbf{u}_j$ to obtain

$$c_j = \frac{\mathbf{y} \cdot \mathbf{u}_j}{\mathbf{u}_j \cdot \mathbf{u}_j}.$$

□

An orthogonal basis allows us to extend this idea from a line to an entire subspace.

Orthogonal Projection onto a Subspace

Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For any $\mathbf{y} \in \mathbb{R}^n$, the **orthogonal projection of $\mathbf{y}$** onto W is

$$\text{proj}_W \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\mathbf{u}_p.$$

Moreover, $\mathbf{y} = \text{proj}_W \mathbf{y} + \mathbf{z}$ where $\mathbf{z} = \mathbf{y} - \text{proj}_W \mathbf{y} \in W^\perp$.

The projection $\text{proj}_W \mathbf{y}$ is the component of $\mathbf{y}$ that lies in W, while $\mathbf{y} - \text{proj}_W \mathbf{y}$ is the component perpendicular to W.

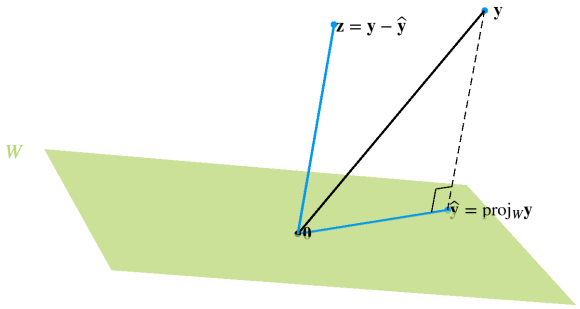


Figure 24: Orthogonal decomposition of $\mathbf{y}$ into $\hat{\mathbf{y}} = \text{proj}_W \mathbf{y}$ and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} \in W^\perp$

Proof

Let

$$\hat{\mathbf{y}} = \text{proj}_W \mathbf{y} = \sum_{i=1}^{p} \frac{\mathbf{y} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i}\mathbf{u}_i.$$

Since $\hat{\mathbf{y}}$ is a linear combination of the basis vectors of W,

$$\hat{\mathbf{y}} \in W.$$

We must show that $\mathbf{y} - \hat{\mathbf{y}} \in W^{\perp}$.

For each $j = 1, 2, \ldots, p$,

$$\begin{aligned}(\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{u}_j &= \mathbf{y} \cdot \mathbf{u}_j - \left(\sum_{i=1}^{p} \frac{\mathbf{y} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i} \mathbf{u}_i \right) \cdot \mathbf{u}_j \\ &= \mathbf{y} \cdot \mathbf{u}_j - \sum_{i=1}^{p} \frac{\mathbf{y} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i} (\mathbf{u}_i \cdot \mathbf{u}_j).\end{aligned}$$

Because the basis is orthogonal, every term in the sum vanishes except the term for which $i = j$. Therefore,

$$\begin{aligned}(\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{u}_j &= \mathbf{y} \cdot \mathbf{u}_j - \frac{\mathbf{y} \cdot \mathbf{u}_j}{\mathbf{u}_j \cdot \mathbf{u}_j} (\mathbf{u}_j \cdot \mathbf{u}_j) \\ &= 0.\end{aligned}$$

Thus, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to every vector in a basis for W. It is therefore orthogonal to every vector in W, so

$$\mathbf{y} - \text{proj}_W \mathbf{y} \in W^{\perp}.$$

Hence,

$$\mathbf{y} = \text{proj}_W \mathbf{y} + (\mathbf{y} - \text{proj}_W \mathbf{y})$$

decomposes $\mathbf{y}$ into one component in W and another in $W^{\perp}$. □

Example 6.1.4

Let

$$W = \text{Span}\left\{\begin{bmatrix}3\\1\\1\end{bmatrix}, \begin{bmatrix}-1\\2\\1\end{bmatrix}\right\}, \quad \mathbf{y} = \begin{bmatrix}1\\1\\1\end{bmatrix}.$$

Find $\text{proj}_W \mathbf{y}$ and the component of $\mathbf{y}$ in $W^\perp$.

Solution

Let

$$\mathbf{u}_1 = \begin{bmatrix}3\\1\\1\end{bmatrix}, \qquad \mathbf{u}_2 = \begin{bmatrix}-1\\2\\1\end{bmatrix}.$$

Since

$$\mathbf{u}_1 \cdot \mathbf{u}_2 = 3(-1) + 1(2) + 1(1) = 0,$$

the vectors form an orthogonal basis for W. Their inner products with $\mathbf{y}$ are

$$\mathbf{y} \cdot \mathbf{u}_1 = 5, \quad \mathbf{y} \cdot \mathbf{u}_2 = 2,$$

while

$$\mathbf{u}_1 \cdot \mathbf{u}_1 = 11, \quad \mathbf{u}_2 \cdot \mathbf{u}_2 = 6.$$

Therefore,

$$\begin{aligned}\operatorname{proj}_W \mathbf{y} &= \frac{5}{11}\mathbf{u}_1 + \frac{2}{6}\mathbf{u}_2 \\ &= \frac{5}{11}\begin{bmatrix}3\\1\\1\end{bmatrix} + \frac{1}{3}\begin{bmatrix}-1\\2\\1\end{bmatrix} \\ &= \begin{bmatrix}\frac{34}{33}\\\frac{37}{33}\\\frac{26}{33}\end{bmatrix}.\end{aligned}$$

The component of $\mathbf{y}$ in $W^\perp$ is

$$\begin{aligned}\mathbf{z} &= \mathbf{y} - \operatorname{proj}_W \mathbf{y} \\ &= \begin{bmatrix}1\\1\\1\end{bmatrix} - \begin{bmatrix}\frac{34}{33}\\\frac{37}{33}\\\frac{26}{33}\end{bmatrix} \\ &= \frac{1}{33}\begin{bmatrix}-1\\-4\\7\end{bmatrix}.\end{aligned}$$

Indeed,

$$\mathbf{z}\cdot\mathbf{u}_1 = 0, \quad \mathbf{z}\cdot\mathbf{u}_2 = 0,$$

so $\mathbf{z} \in W^\perp$. Thus,

$$\mathbf{y} = \underbrace{\begin{bmatrix}\frac{34}{33}\\\frac{37}{33}\\\frac{26}{33}\end{bmatrix}}_{\operatorname{proj}_W \mathbf{y}} + \underbrace{\frac{1}{33}\begin{bmatrix}-1\\-4\\7\end{bmatrix}}_{\mathbf{z}\in W^\perp}.$$

The orthogonal projection is not merely one vector in W that approximates $\mathbf{y}$. It is the unique vector in W closest to $\mathbf{y}$.

Example 6.1.5

Find the distance from $\mathbf{y}$ to the line L shown in Figure 25.

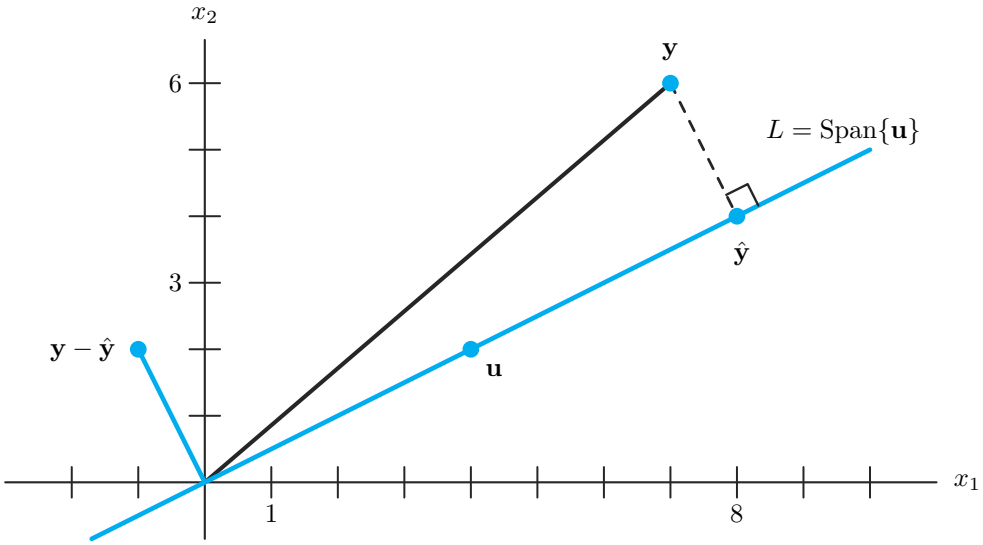


Figure 25: The orthogonal projection of $\mathbf{y}$ onto the line $L = \text{Span}\{\mathbf{u}\}$, which passes through the origin in the direction of $\mathbf{u}$.

Solution

The distance from $\mathbf{y}$ to L is the length of the perpendicular segment from $\mathbf{y}$ to its orthogonal projection $\hat{\mathbf{y}}$ which is $\|\mathbf{y} - \hat{\mathbf{y}}\|$

From Figure 25,

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

Thus,

$$\|\mathbf{y} - \hat{\mathbf{y}}\| = \sqrt{(-1)^2 + 2^2} = \sqrt{5}.$$

Best Approximation Theorem

Let W be a subspace of $\mathbb{R}^n$, let $\mathbf{y} \in \mathbb{R}^n$, and let $\hat{\mathbf{y}} = \text{proj}_W \mathbf{y}$. Then

$$\|\mathbf{y} - \hat{\mathbf{y}}\| \leq \|\mathbf{y} - \mathbf{w}\|$$

for every $\mathbf{w} \in W$. The expression is equal only when $\mathbf{w} = \hat{\mathbf{y}}$.

Thus, $\text{proj}_W \mathbf{y}$ is the unique vector in W closest to $\mathbf{y}$.

Proof

For any $\mathbf{w} \in W$,

$$\mathbf{y} - \mathbf{w} = (\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \mathbf{w}).$$

The first vector lies in $W^\perp$, while the second lies in W. Therefore, the two vectors are orthogonal. By the Pythagorean theorem,

$$\|\mathbf{y} - \mathbf{w}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \mathbf{w}\|^2.$$

Since

$$\|\hat{\mathbf{y}} - \mathbf{w}\|^2 \geq 0,$$

it follows that

$$\|\mathbf{y} - \mathbf{w}\|^2 \geq \|\mathbf{y} - \hat{\mathbf{y}}\|^2.$$

Taking square roots gives

$$\|\mathbf{y} - \mathbf{w}\| \geq \|\mathbf{y} - \hat{\mathbf{y}}\|.$$

Equality holds only when

$$\|\hat{\mathbf{y}} - \mathbf{w}\| = 0,$$

which occurs only when $\mathbf{w} = \hat{\mathbf{y}}$. □

An orthogonal set becomes especially useful when each vector has length 1.

Orthonormal Sets and Bases

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ be a set of vectors in $\mathbb{R}^n$. The set S is an **orthonormal set** if it is orthogonal and every vector has unit length. Thus,

$$\mathbf{u}_i \cdot \mathbf{u}_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

If S spans a subspace W, then S is called an **orthonormal basis** for W.

Since every vector in an orthonormal set is nonzero, every orthonormal set is automatically linearly independent.

The simplest example is the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ for $\mathbb{R}^n$. Each standard basis vector has length one, and distinct standard basis vectors are orthogonal. Therefore, the standard basis is an orthonormal basis for $\mathbb{R}^n$. Any nonempty subset of the standard basis is also an orthonormal set.

Example 6.1.5

Show that $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis for $\mathbb{R}^3$, where

$$\mathbf{v}_1 = \frac{1}{\sqrt{11}} \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \frac{1}{\sqrt{66}} \begin{bmatrix} -1 \\ -4 \\ 7 \end{bmatrix}.$$

Solution

We first verify that the vectors are orthogonal to each other:

$$\begin{aligned}
\mathbf{v}_1 \cdot \mathbf{v}_2 &= \frac{1}{\sqrt{66}}\left[3(-1) + 1(2) + 1(1)\right] \\
&= \frac{-3 + 2 + 1}{\sqrt{66}} \\
&= 0,
\end{aligned}$$

$$\begin{aligned}
\mathbf{v}_1 \cdot \mathbf{v}_3 &= \frac{1}{\sqrt{726}}\left[3(-1) + 1(-4) + 1(7)\right] \\
&= \frac{-3 - 4 + 7}{\sqrt{726}} \\
&= 0,
\end{aligned}$$

$$\begin{aligned}
\mathbf{v}_2 \cdot \mathbf{v}_3 &= \frac{1}{\sqrt{396}}\left[(-1)(-1) + 2(-4) + 1(7)\right] \\
&= \frac{1 - 8 + 7}{\sqrt{396}} \\
&= 0.
\end{aligned}$$

Therefore, S is an orthogonal set.

We next verify that each vector has unit length:

$$\begin{aligned}
\mathbf{v}_1 \cdot \mathbf{v}_1 &= \frac{3^2 + 1^2 + 1^2}{11} = \frac{11}{11} = 1, \\
\mathbf{v}_2 \cdot \mathbf{v}_2 &= \frac{(-1)^2 + 2^2 + 1^2}{6} = \frac{6}{6} = 1, \\
\mathbf{v}_3 \cdot \mathbf{v}_3 &= \frac{(-1)^2 + (-4)^2 + 7^2}{66} = \frac{66}{66} = 1.
\end{aligned}$$

Thus, each vector has length one, so S is an orthonormal set. Since every orthonormal set is linearly independent and S contains three vectors in $\mathbb{R}^3$, the vectors form a basis for $\mathbb{R}^3$. Therefore, S is an orthonormal basis for $\mathbb{R}^3$.

Norm of a Linear Combination in an Orthonormal Set

Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\} \in \mathbb{R}^n$ be an orthonormal set. Then, for any scalars $a_1, a_2, \ldots, a_p \in \mathbb{R}$,

$$\|a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_p\mathbf{u}_p\|^2 = a_1^2 + a_2^2 + \cdots + a_p^2.$$

Proof

Let $\mathbf{x} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_p\mathbf{u}_p$. Then

$$\begin{aligned}\|\mathbf{x}\|^2 &= \mathbf{x} \cdot \mathbf{x} \\ &= \left(\sum_{i=1}^{p} a_i\mathbf{u}_i\right) \cdot \left(\sum_{j=1}^{p} a_j\mathbf{u}_j\right) \\ &= \sum_{i=1}^{p}\sum_{j=1}^{p} a_i a_j \left(\mathbf{u}_i \cdot \mathbf{u}_j\right).\end{aligned}$$

Because the set is orthonormal,

$$\mathbf{u}_i \cdot \mathbf{u}_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Therefore, every cross term vanishes, leaving

$$\|\mathbf{x}\|^2 = a_1^2 + a_2^2 + \cdots + a_p^2.$$

As an immediate consequence, an orthonormal set containing n vectors in $\mathbb{R}^n$ is an orthonormal basis for $\mathbb{R}^n$. Indeed, the vectors are automatically linearly independent, and any linearly independent set of n vectors in $\mathbb{R}^n$ forms a basis. □

Any orthogonal set of nonzero vectors can be converted into an orthonormal set by normalizing each vector. If $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ is orthogonal, then the vectors

$$\mathbf{v}_j = \frac{\mathbf{u}_j}{\|\mathbf{u}_j\|}, \qquad j = 1, 2, \ldots, p,$$

still point in the same directions and remain orthogonal to each other. Since each $\mathbf{v}_j$ has length one, the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is orthonormal.

Matrices whose columns form an orthonormal set have a simple algebraic structure.

Matrices with Orthonormal Columns

Let

$$U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$$

be an $m \times n$ matrix. The columns of U form an orthonormal set if and only if $U^T U = I_n$.

Proof

Since

$$U^T = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix},$$

we have

$$\begin{aligned} U^T U &= \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{u}_1 \cdot \mathbf{u}_1 & \mathbf{u}_1 \cdot \mathbf{u}_2 & \cdots & \mathbf{u}_1 \cdot \mathbf{u}_n \\ \mathbf{u}_2 \cdot \mathbf{u}_1 & \mathbf{u}_2 \cdot \mathbf{u}_2 & \cdots & \mathbf{u}_2 \cdot \mathbf{u}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_n \cdot \mathbf{u}_1 & \mathbf{u}_n \cdot \mathbf{u}_2 & \cdots & \mathbf{u}_n \cdot \mathbf{u}_n \end{bmatrix}. \end{aligned}$$

If the columns of U are orthonormal, then

$$\mathbf{u}_i \cdot \mathbf{u}_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Therefore $U^T U = I_n$.

Conversely, suppose $U^T U = I_n$. The diagonal entries of $U^T U$ give $\mathbf{u}_j \cdot \mathbf{u}_j = 1$, so every column has length one. The off-diagonal entries give $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$. Thus, the columns of U form an orthonormal set. □

The projection formula becomes particularly simple when W has an orthonormal basis.

Projection onto a Subspace with an Orthonormal Basis

Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ be an orthonormal basis for a subspace W of $\mathbb{R}^n$, and let

$$U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p \end{bmatrix}.$$

Then, for every $\mathbf{y} \in \mathbb{R}^n$,

$$\text{proj}_W \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{y} \cdot \mathbf{u}_p)\mathbf{u}_p.$$

Equivalently, $\text{proj}_W \mathbf{y} = UU^T \mathbf{y}$.

Proof

Since the basis is orthonormal,

$$\mathbf{u}_j \cdot \mathbf{u}_j = 1$$

for every $j = 1, 2, \ldots, p$. Therefore, the general projection formula simplifies to

$$\text{proj}_W \mathbf{y} = \sum_{j=1}^{p} (\mathbf{y} \cdot \mathbf{u}_j)\mathbf{u}_j.$$

Moreover,

$$U^T\mathbf{y} = \begin{bmatrix} \mathbf{u}_1^T\mathbf{y} \\ \mathbf{u}_2^T\mathbf{y} \\ \vdots \\ \mathbf{u}_p^T\mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{y}\cdot\mathbf{u}_1 \\ \mathbf{y}\cdot\mathbf{u}_2 \\ \vdots \\ \mathbf{y}\cdot\mathbf{u}_p \end{bmatrix}.$$

Thus,

$$\begin{aligned} UU^T\mathbf{y} &= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p \end{bmatrix} \begin{bmatrix} \mathbf{y}\cdot\mathbf{u}_1 \\ \mathbf{y}\cdot\mathbf{u}_2 \\ \vdots \\ \mathbf{y}\cdot\mathbf{u}_p \end{bmatrix} \\ &= (\mathbf{y}\cdot\mathbf{u}_1)\mathbf{u}_1 + (\mathbf{y}\cdot\mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{y}\cdot\mathbf{u}_p)\mathbf{u}_p \\ &= \text{proj}_W\,\mathbf{y}. \end{aligned}$$

□

Properties of Matrices with Orthonormal Columns

Let U be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then the following statements are true.

(i) $\|U\mathbf{x}\| = \|\mathbf{x}\|$.

(ii) $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$.

(iii) $U\mathbf{x} \perp U\mathbf{y} \iff \mathbf{x} \perp \mathbf{y}$.

Proof

Since the columns of U are orthonormal,

$$U^TU = I_n.$$

(i)

$$\begin{aligned}\|U\mathbf{x}\|^2 &= (U\mathbf{x})^T(U\mathbf{x}) \\ &= \mathbf{x}^T U^T U\mathbf{x} \\ &= \mathbf{x}^T I_n \mathbf{x} \\ &= \mathbf{x}^T\mathbf{x} \\ &= \|\mathbf{x}\|^2.\end{aligned}$$

Since norms are nonnegative,

$$\|U\mathbf{x}\| = \|\mathbf{x}\|.$$

(ii)

$$\begin{aligned}(U\mathbf{x})\cdot(U\mathbf{y}) &= (U\mathbf{x})^T(U\mathbf{y}) \\ &= \mathbf{x}^T U^T U\mathbf{y} \\ &= \mathbf{x}^T I_n \mathbf{y} \\ &= \mathbf{x}^T\mathbf{y} \\ &= \mathbf{x}\cdot\mathbf{y}.\end{aligned}$$

(iii) Finally,

$$U\mathbf{x} \perp U\mathbf{y}$$

if and only if

$$(U\mathbf{x})\cdot(U\mathbf{y}) = 0.$$

By part (ii), this is equivalent to $\mathbf{x}\cdot\mathbf{y} = 0$ which holds if and only if

$$\mathbf{x} \perp \mathbf{y}.$$

Thus, multiplication by a matrix with orthonormal columns preserves lengths, inner products, and orthogonality. □

Example 6.1.6

Let

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & -\frac{2}{3} \\ 0 & \frac{1}{3} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix}.$$

Verify that the columns of U are orthonormal and that

$$\|U\mathbf{x}\| = \|\mathbf{x}\|.$$

Solution

We first compute

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

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

Thus,

$$U^T U = I_2,$$

so the columns of U are orthonormal.

Now,

$$U\mathbf{x} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & -\frac{2}{3} \\ 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix}$$
$$= \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}.$$

Therefore,

$$\|U\mathbf{x}\| = \sqrt{3^2 + (-1)^2 + 1^2} = \sqrt{11},$$

while

$$\|\mathbf{x}\| = \sqrt{(\sqrt{2})^2 + 3^2} = \sqrt{11}.$$

Hence,

$$\|U\mathbf{x}\| = \|\mathbf{x}\|.$$

When dealing with square matrices, we have a special case. A square matrix with orthonormal columns is called an **orthogonal matrix**. For such a matrix,

$$U^T U = I.$$

Since U is square, this implies that U is invertible and

$$U^{-1} = U^T.$$

Consequently,

$$UU^T = I,$$

so the rows of an orthogonal matrix are also orthonormal.

Problem 6.1.1

Let

$$\mathbf{u}_1 = \begin{bmatrix} -7 \\ 1 \\ 4 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \\ -2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} -9 \\ 1 \\ 6 \end{bmatrix},$$

and let $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Use the fact that $\mathbf{u}_1 \perp \mathbf{u}_2$ to compute $\text{proj}_W \mathbf{y}$.

Solution

First,

$$\mathbf{u}_1 \cdot \mathbf{u}_2 = (-7)(-1) + 1(1) + 4(-2) = 0,$$

so $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for W. Therefore,

$$\text{proj}_W \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2.$$

The required inner products are

$$\mathbf{y} \cdot \mathbf{u}_1 = 88, \quad \mathbf{u}_1 \cdot \mathbf{u}_1 = 66,$$
$$\mathbf{y} \cdot \mathbf{u}_2 = -2, \quad \mathbf{u}_2 \cdot \mathbf{u}_2 = 6.$$

Thus,

$$\begin{aligned}\operatorname{proj}_W \mathbf{y} &= \frac{88}{66}\mathbf{u}_1 - \frac{2}{6}\mathbf{u}_2 \\ &= \frac{4}{3}\begin{bmatrix}-7\\1\\4\end{bmatrix} - \frac{1}{3}\begin{bmatrix}-1\\1\\-2\end{bmatrix} \\ &= \begin{bmatrix}-9\\1\\6\end{bmatrix} \\ &= \mathbf{y}.\end{aligned}$$

Therefore, $\operatorname{proj}_W \mathbf{y} = \mathbf{y}$

In this case, $\mathbf{y}$ already belongs to W, so its projection onto W is itself.

Problem 6.1.2

Let W be a subspace of $\mathbb{R}^n$, and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Define $\mathbf{z} = \mathbf{x} + \mathbf{y}$.

Suppose $\mathbf{u} = \operatorname{proj}_W \mathbf{x}$ and $\mathbf{v} = \operatorname{proj}_W \mathbf{y}$. Show that $\operatorname{proj}_W \mathbf{z} = \mathbf{u} + \mathbf{v}$.

Solution

Let U be a matrix whose columns form an orthonormal basis for W. Then

$$\operatorname{proj}_W \mathbf{x} = UU^T\mathbf{x}, \quad \operatorname{proj}_W \mathbf{y} = UU^T\mathbf{y}.$$

Since $\mathbf{z} = \mathbf{x} + \mathbf{y}$,

$$\begin{aligned}\operatorname{proj}_W \mathbf{z} &= UU^T\mathbf{z} \\ &= UU^T(\mathbf{x} + \mathbf{y}) \\ &= UU^T\mathbf{x} + UU^T\mathbf{y} \\ &= \operatorname{proj}_W \mathbf{x} + \operatorname{proj}_W \mathbf{y} \\ &= \mathbf{u} + \mathbf{v}.\end{aligned}$$

Therefore, $\operatorname{proj}_W(\mathbf{x} + \mathbf{y}) = \operatorname{proj}_W \mathbf{x} + \operatorname{proj}_W \mathbf{y}$.

Problem 6.1.3

Let

$$\mathbf{u}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Show that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthonormal basis for $\mathbb{R}^2$.

Solution

First, the vectors are orthogonal because

$$\begin{aligned} \mathbf{u}_1 \cdot \mathbf{u}_2 &= \frac{1}{\sqrt{5}} \begin{bmatrix} -1 \\ 2 \end{bmatrix} \cdot \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix} \\ &= \frac{1}{5} \left[(-1)(2) + 2(1) \right] \\ &= 0. \end{aligned}$$

Next,

$$\begin{aligned} \|\mathbf{u}_1\|^2 &= \mathbf{u}_1 \cdot \mathbf{u}_1 \\ &= \frac{(-1)^2 + 2^2}{5} \\ &= 1, \end{aligned}$$

and

$$\begin{aligned} \|\mathbf{u}_2\|^2 &= \mathbf{u}_2 \cdot \mathbf{u}_2 \\ &= \frac{2^2 + 1^2}{5} \\ &= 1. \end{aligned}$$

Thus, the vectors are orthogonal and each has unit length, so $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthonormal set.

6.2 The Gram-Schmidt Process

Orthogonal and orthonormal bases make computation much easier, but subspaces usually give us arbitrary ones. The Gram–Schmidt process provides a systematic way to construct such a basis from any existing basis of a nonzero subspace of $\mathbb{R}^n$. It begins with any basis and systematically removes from each vector the components that lie in the directions already constructed. The resulting vectors span the same subspace as the original basis but are mutually orthogonal.

Example 6.2.1

Let $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$ where

$$\mathbf{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

Construct an orthogonal basis for W.

Solution

We keep the first vector unchanged:

$$\mathbf{v}_1 = \mathbf{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$$

To construct the second vector, subtract from $\mathbf{x}_2$ its projection onto $\mathbf{v}_1$:

$$\mathbf{v}_2 = \mathbf{x}_2 - \text{proj}_{\mathbf{v}_1} \mathbf{x}_2$$

Since

$$\mathbf{x}_2 \cdot \mathbf{v}_1 = 1(3) + 2(6) + 2(0) = 15$$

and

$$\mathbf{v}_1 \cdot \mathbf{v}_1 = 3^2 + 6^2 = 45,$$

we obtain

$$\begin{aligned}
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \\
&= \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \frac{15}{45}\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}.
\end{aligned}$$

Indeed,

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = 3(0) + 6(0) + 0(2) = 0.$$

Thus, $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal. Moreover,

$$\mathbf{v}_1 = \mathbf{x}_1, \quad \mathbf{v}_2 = \mathbf{x}_2 - \frac{1}{3}\mathbf{x}_1,$$

so the new vectors span the same subspace as the original vectors. Therefore,

$$\left\{ \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} \right\}$$

is an orthogonal basis for W.

The Gram–Schmidt Process

Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ be a basis for a nonzero subspace $W \in \mathbb{R}^n$. Define

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1, \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1, \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2, \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\mathbf{v}_{p-1}.
\end{aligned}$$

Or equivalently, for $k = 2, 3, \ldots, p$,

$$\mathbf{v}_k = \mathbf{x}_k - \sum_{j=1}^{k-1} \frac{\mathbf{x}_k \cdot \mathbf{v}_j}{\mathbf{v}_j \cdot \mathbf{v}_j}\mathbf{v}_j.$$

Then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is an orthogonal basis for W. Moreover,

$$\operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} = \operatorname{Span}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}, \qquad 1 \leq k \leq p.$$

Normalizing each vector,

$$\mathbf{q}_j = \frac{\mathbf{v}_j}{\|\mathbf{v}_j\|}, \qquad j = 1, 2, \ldots, p,$$

produces the orthonormal basis $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_p\}$ for W.

Proof

For each $k = 1, 2, \ldots, p$, define $W_k = \operatorname{Span}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$. We will prove recursively that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an orthogonal basis for W_k.

For $k = 1$,

$$\mathbf{v}_1 = \mathbf{x}_1,$$

so $\text{Span}\{\mathbf{v}_1\} = \text{Span}\{\mathbf{x}_1\} = W_1$.

Now suppose that, for some $k < p$, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an orthogonal basis for W_k. And define

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - \text{proj}_{W_k} \mathbf{x}_{k+1}.$$

Because the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ form an orthogonal basis for W_k,

$$\text{proj}_{W_k} \mathbf{x}_{k+1} = \sum_{j=1}^{k} \frac{\mathbf{x}_{k+1} \cdot \mathbf{v}_j}{\mathbf{v}_j \cdot \mathbf{v}_j} \mathbf{v}_j.$$

Thus,

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - \sum_{j=1}^{k} \frac{\mathbf{x}_{k+1} \cdot \mathbf{v}_j}{\mathbf{v}_j \cdot \mathbf{v}_j} \mathbf{v}_j.$$

By the orthogonal decomposition theorem,

$$\mathbf{v}_{k+1} \in W_k^{\perp}.$$

Therefore,

$$\mathbf{v}_{k+1} \cdot \mathbf{v}_j = 0$$

for every $j = 1, 2, \ldots, k$. Hence, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k+1}\}$ is an orthogonal set.

We have

$$\text{proj}_{W_k} \mathbf{x}_{k+1} \in W_k \subseteq W_{k+1},$$

where $\mathbf{x}_{k+1} \in W_{k+1}$. Since W_{k+1} is closed under subtraction,

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - \text{proj}_{W_k} \mathbf{x}_{k+1} \in W_{k+1}.$$

Also, $\mathbf{v}_{k+1} \neq \mathbf{0}$. Otherwise,

$$\mathbf{x}_{k+1} = \text{proj}_{W_k} \mathbf{x}_{k+1} \in W_k,$$

which would contradict the linear independence of $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$.

Thus, the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k+1}$ are nonzero and orthogonal, so they are linearly independent.

Finally, the definition of $\mathbf{v}_{k+1}$ gives

$$\mathbf{x}_{k+1} = \mathbf{v}_{k+1} + \operatorname{proj}_{W_k} \mathbf{x}_{k+1}.$$

Since the projection lies in $\operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$, we have

$$\mathbf{x}_{k+1} \in \operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k+1}\}.$$

Therefore,

$$\operatorname{Span}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k+1}\} = \operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k+1}\}.$$

Hence, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k+1}\}$ is an orthogonal basis for W_{k+1}. Continuing until $k = p$ gives an orthogonal basis for $W_p = W$.

Normalizing each nonzero vector preserves orthogonality and gives every vector unit length. Therefore,

$$\left\{ \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}, \ldots, \frac{\mathbf{v}_p}{\|\mathbf{v}_p\|} \right\}$$

is an orthonormal basis for W. □

Example 6.2.2

Let

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

The set $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is a basis for a subspace $W \in \mathbb{R}^4$. Use the Gram–Schmidt process to construct an orthogonal basis for W.

Solution

1. Keep the first vector unchanged:

$$\mathbf{v}_1 = \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Thus $W_1 = \operatorname{Span}\{\mathbf{v}_1\}$.

2. Subtract from $\mathbf{x}_2$ its projection onto W_1:

$$\widetilde{\mathbf{v}}_2 = \mathbf{x}_2 - \operatorname{proj}_{W_1} \mathbf{x}_2$$

Since

$$\mathbf{x}_2 \cdot \mathbf{v}_1 = 3, \quad \mathbf{v}_1 \cdot \mathbf{v}_1 = 4,$$

we obtain

$$\begin{aligned} \widetilde{\mathbf{v}}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 \\ &= \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -\frac{3}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{bmatrix}. \end{aligned}$$

Multiplying a vector by a nonzero scalar does not change its direction or its orthogonality to $\mathbf{v}_1$. To eliminate the fractions, choose

$$\mathbf{v}_2 = 4\widetilde{\mathbf{v}}_2 = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Then $W_2 = \operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$.

3. Subtract from $\mathbf{x}_3$ its projection onto W_2. Since $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for W_2,

$$\operatorname{proj}_{W_2} \mathbf{x}_3 = \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2.$$

The required inner products are

$$\mathbf{x}_3 \cdot \mathbf{v}_1 = 2, \quad \mathbf{v}_1 \cdot \mathbf{v}_1 = 4,$$
$$\mathbf{x}_3 \cdot \mathbf{v}_2 = 2, \quad \mathbf{v}_2 \cdot \mathbf{v}_2 = 12.$$

Therefore,

$$\operatorname{proj}_{W_2} \mathbf{x}_3 = \frac{2}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \frac{2}{12}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$
$$= \begin{bmatrix} 0 \\ \frac{2}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}.$$

Hence,

$$\begin{aligned}\widetilde{\mathbf{v}}_3 &= \mathbf{x}_3 - \operatorname{proj}_{W_2} \mathbf{x}_3 \\ &= \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ \frac{2}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ -\frac{2}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}.\end{aligned}$$

Again, scale the vector to eliminate fractions:

$$\mathbf{v}_3 = 3\widetilde{\mathbf{v}}_3 = \begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}$$

The vectors are mutually orthogonal because

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = 0, \quad \mathbf{v}_1 \cdot \mathbf{v}_3 = 0, \quad \mathbf{v}_2 \cdot \mathbf{v}_3 = 0.$$

Each new vector was obtained from an original basis vector by subtracting linear combinations of the preceding vectors, so the new vectors span the same subspace W. Therefore,

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix} \right\}$$

is an orthogonal basis for W.

Once the Gram–Schmidt process has produced an orthogonal basis, an orthonormal basis is obtained by normalizing each vector. For hand calculations, it is usually easier to complete the entire orthogonalization process before normalizing, since this avoids introducing unnecessary square roots into the intermediate steps.

Example 6.2.3

In Example 6.2.1, we constructed the orthogonal basis

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

for the subspace

$$W = \text{Span}\left\{ \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \right\}.$$

Construct an orthonormal basis for W.

Solution

Normalize each vector in the orthogonal basis. Since

$$\|\mathbf{v}_1\| = \sqrt{3^2 + 6^2} = \sqrt{45} = 3\sqrt{5},$$

we obtain

$$\begin{aligned}\mathbf{q}_1 &= \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} \\ &= \frac{1}{3\sqrt{5}} \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \\ 0 \end{bmatrix}.\end{aligned}$$

Similarly, $\|\mathbf{v}_2\| = 2$, so

$$\begin{aligned}\mathbf{q}_2 &= \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} \\ &= \frac{1}{2} \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.\end{aligned}$$

Therefore,

$$\left\{ \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is an orthonormal basis for W.

The Gram–Schmidt process also leads naturally to one of the most important matrix factorizations in numerical linear algebra. It separates a matrix into a matrix with orthonormal columns and an upper triangular matrix.

The QR Factorization

Let A be an $m \times n$ matrix with linearly independent columns. Then A can be factored as

$$A = QR,$$

where Q is an $m \times n$ matrix whose columns form an orthonormal basis for $\mathrm{Col}(A)$, and R is an invertible $n \times n$ upper triangular matrix with positive diagonal entries.
Moreover, $Q^T Q = I_n$ and $R = Q^T A$.

Proof

Write the columns of A as

$$A = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}.$$

Because the columns are linearly independent, they form a basis for $\mathrm{Col}(A)$. Apply the Gram–Schmidt process and normalize the resulting vectors to obtain an orthonormal basis $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n\}$ for $\mathrm{Col}(A)$. Define

$$Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix}.$$

Since the columns of Q are orthonormal,

$$Q^T Q = I_n.$$

The Gram–Schmidt process preserves the successive spans, so

$$\mathrm{Span}\,\{\mathbf{x}_1, \ldots, \mathbf{x}_k\} = \mathrm{Span}\,\{\mathbf{q}_1, \ldots, \mathbf{q}_k\}$$

for every $k = 1, 2, \ldots, n$. Therefore, each column $\mathbf{x}_k$ can be written as

$$\mathbf{x}_k = r_{1,k}\mathbf{q}_1 + r_{2,k}\mathbf{q}_2 + \cdots + r_{k,k}\mathbf{q}_k.$$

Equivalently,

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} r_{1,k} \\ r_{2,k} \\ \vdots \\ r_{k,k} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Let

$$R = \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,n} \\ 0 & r_{2,2} & \cdots & r_{2,n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & r_{n,n} \end{bmatrix}.$$

The kth column of QR is then $\mathbf{x}_k$. Hence,

$$\begin{aligned} QR &= \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} \\ &= A. \end{aligned}$$

Thus,

$$A = QR.$$

In the Gram–Schmidt process, the diagonal entry $r_{k,k}$ is the length of the nonzero orthogonal vector produced at the kth stage. Therefore,

$$r_{k,k} > 0$$

for every $k = 1, 2, \ldots, n$. Since R is upper triangular with nonzero diagonal entries, it is invertible.

Finally, multiply $A = QR$ on the left by Q^T:

$$\begin{aligned} Q^T A &= Q^T QR \\ &= I_n R \\ &= R. \end{aligned}$$

Therefore,

$$R = Q^T A.$$

□

Example 6.2.4

Find a QR factorization of

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

Solution

The columns of A are the vectors from Example 6.2.2:

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

Applying the Gram–Schmidt process produced the orthogonal basis

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}.$$

Their lengths are

$$\|\mathbf{v}_1\| = 2, \quad \|\mathbf{v}_2\| = \sqrt{12}, \quad \|\mathbf{v}_3\| = \sqrt{6}.$$

Therefore, the corresponding orthonormal vectors are

$$\mathbf{q}_1 = \frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \quad \mathbf{q}_2 = \frac{1}{\sqrt{12}}\begin{bmatrix}-3\\1\\1\\1\end{bmatrix}, \quad \mathbf{q}_3 = \frac{1}{\sqrt{6}}\begin{bmatrix}0\\-2\\1\\1\end{bmatrix}.$$

Using these vectors as the columns of Q gives

$$Q = \begin{bmatrix}\frac{1}{2} & -\frac{3}{\sqrt{12}} & 0\\ \frac{1}{2} & \frac{1}{\sqrt{12}} & -\frac{2}{\sqrt{6}}\\ \frac{1}{2} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{6}}\\ \frac{1}{2} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{6}}\end{bmatrix}.$$

Since the columns of Q are orthonormal, $Q^TQ = I_3$. Thus, $R = Q^TA$. Substituting, we get

$$\begin{aligned} R &= \begin{bmatrix}\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2}\\ -\frac{3}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}}\\ 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}}\end{bmatrix}\begin{bmatrix}1&0&0\\1&1&0\\1&1&1\\1&1&1\end{bmatrix}\\ &= \begin{bmatrix}2 & \frac{3}{2} & 1\\ 0 & \frac{3}{\sqrt{12}} & \frac{2}{\sqrt{12}}\\ 0 & 0 & \frac{2}{\sqrt{6}}\end{bmatrix}. \end{aligned}$$

Therefore, a QR factorization of A is $A = QR$ where

$$Q = \begin{bmatrix} \frac{1}{2} & -\frac{3}{\sqrt{12}} & 0 \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{6}} \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{6}} \end{bmatrix}, \quad R = \begin{bmatrix} 2 & \frac{3}{2} & 1 \\ 0 & \frac{3}{\sqrt{12}} & \frac{2}{\sqrt{12}} \\ 0 & 0 & \frac{2}{\sqrt{6}} \end{bmatrix}.$$

In exact arithmetic, the Gram–Schmidt process produces perfectly orthogonal vectors. In floating-point arithmetic, however, roundoff error may accumulate as each new vector is constructed, causing the computed vectors to lose orthogonality; it is numerically unstable. The modified Gram–Schmidt process rearranges the same calculations and is usually more stable than the classical version, although both remain valuable for understanding how orthogonalization works. That being said, neither are used very much nowadays for computation.

For practical numerical computation, QR factorizations are more commonly computed using Householder reflections. A sequence of orthogonal Householder matrices is used to eliminate the entries below the diagonal of A, producing an upper triangular matrix R while preserving numerical stability. Givens rotations are also useful when the matrix is sparse or when individual entries must be eliminated selectively, but Householder QR is generally the standard method for dense matrices. NumPy, MATLAB, Julia, and PyTorch all use Householder reflections.

Problem 6.2.1

Let

$$A = \begin{bmatrix} -10 & 13 & 7 & -11 \\ 2 & 1 & -5 & 3 \\ -6 & 3 & 13 & -3 \\ 16 & -16 & -2 & 5 \\ 2 & 1 & -5 & -7 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 \end{bmatrix}.$$

Suppose that, after k steps, the columns of

$$Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_k \end{bmatrix}$$

form an orthonormal basis for $W_k = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$.

Since $\text{proj}_{W_k} \mathbf{x} = QQ^T\mathbf{x}$, the next orthogonal vector may be computed using

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - Q\left(Q^T\mathbf{x}_{k+1}\right).$$

Normalize this vector by defining

$$\mathbf{q}_{k+1} = \frac{\mathbf{v}_{k+1}}{\|\mathbf{v}_{k+1}\|},$$

and then update

$$Q = \begin{bmatrix} Q & \mathbf{q}_{k+1} \end{bmatrix}.$$

Use this procedure in MATLAB to compute a QR factorization $A = QR$. Then, verify numerically that $Q^TQ = I$ and $QR = A$.

Solution

We construct the columns of Q one at a time. At the kth step, subtract from $\mathbf{x}_k$ its orthogonal projection onto the span of the previously computed columns of Q:

$$\mathbf{v}_k = \mathbf{x}_k - Q_{k-1}\left(Q_{k-1}^T \mathbf{x}_k\right)$$

Then normalize $\mathbf{v}_k$ to obtain the next column of Q. Once all four orthonormal columns have been computed, use $R = Q^T A$.

In MATLAB,

```
>> A = [-10,  13,   7, -11;
          2,   1,  -5,   3;
         -6,   3,  13,  -3;
         16, -16,  -2,   5;
          2,   1,  -5,  -7];

>> [m,n] = size(A);
>> Q = zeros(m,n);

>> for k = 1:n
       v = A(:,k);

       if k > 1
           Qk = Q(:,1:k-1);
           v = v - Qk*(Qk.'*v);
       end

       Q(:,k) = v/norm(v);
   end

>> R = Q.'*A;

>> Q(abs(Q) < 1e-12) = 0;
>> R(abs(R) < 1e-12) = 0;

>> Q

Q =

   -0.5000    0.5000    0.5774         0
    0.1000    0.5000         0    0.7071
   -0.3000   -0.5000    0.5774         0
```

```
    0.8000         0    0.5774         0
    0.1000    0.5000         0   -0.7071

>> R

R =

   20.0000  -20.0000  -10.0000   10.0000
         0    6.0000   -8.0000   -6.0000
         0         0   10.3923   -5.1962
         0         0         0    7.0711
```

Thus, the QR factorization is $A = QR$ where

$$Q = \begin{bmatrix} -0.5000 & 0.5000 & 0.5774 & 0 \\ 0.1000 & 0.5000 & 0 & 0.7071 \\ -0.3000 & -0.5000 & 0.5774 & 0 \\ 0.8000 & 0 & 0.5774 & 0 \\ 0.1000 & 0.5000 & 0 & -0.7071 \end{bmatrix}$$

and

$$R = \begin{bmatrix} 20.0000 & -20.0000 & -10.0000 & 10.0000 \\ 0 & 6.0000 & -8.0000 & -6.0000 \\ 0 & 0 & 10.3923 & -5.1962 \\ 0 & 0 & 0 & 7.0711 \end{bmatrix}.$$

We verify that the columns of Q are orthonormal:

```
>> round(Q.'*Q,12)

ans =

     1     0     0     0
     0     1     0     0
     0     0     1     0
     0     0     0     1
```

Hence,

$$Q^T Q = I_4.$$

Finally, compute the residual of the factorization:

```
>> norm(Q*R - A)

ans =

   4.7260e-15
```

This value is essentially zero to machine precision, confirming that $QR = A$.

6.3 Least Squares

Least squares is one of the most important tools for turning imperfect measurements into useful mathematical models. In science and engineering, data basically never lies exactly on a line, curve, or surface because of noise, measurement error, and natural variation. Instead, we collect data points and seek a function that captures the underlying relationship as accurately as possible. This occurs in applications such as calibrating sensors, estimating material properties, fitting geometric features from coordinate-measuring-machine data, and comparing observed galaxy-clustering patterns with theoretical models to study the accelerating expansion of the universe. Because an exact fit may not exist, we look for the model whose predicted values are closest to the observed data overall. Least squares provides a systematic way to find that best-fitting approximation. Nearly every programming language, scientific package, and spreadsheet software uses least squares.

An inconsistent system $A\mathbf{x} = \mathbf{b}$ has no exact solution, but in many applications we still need a vector $\mathbf{x}$ for which $A\mathbf{x}$ is as close as possible to $\mathbf{b}$. Because every vector $A\mathbf{x}$ lies in $\mathrm{Col}(A)$, the problem is equivalent to finding the point in $\mathrm{Col}(A)$ closest to $\mathbf{b}$. The best approximation theorem tells us that this point is the orthogonal projection of $\mathbf{b}$ onto $\mathrm{Col}(A)$.

Least-Squares Solution

Let A be an $m \times n$ matrix and let $\mathbf{b} \in \mathbb{R}^m$. A vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ is called a **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ if

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

for every $\mathbf{x} \in \mathbb{R}^n$.

The vector $\mathbf{r} = \mathbf{b} - A\hat{\mathbf{x}}$ is called the **residual vector**. The term *least squares* comes from minimizing

$$\|\mathbf{r}\|^2 = r_1^2 + r_2^2 + \cdots + r_m^2,$$

the sum of the squares of the residual components.

Since $A\hat{\mathbf{x}} \in \mathrm{Col}(A)$, the least-squares problem seeks $\hat{\mathbf{x}}$ such that

$$A\hat{\mathbf{x}} = \mathrm{proj}_{\mathrm{Col}(A)}\, \mathbf{b}.$$

If $\mathbf{b} \in \mathrm{Col}(A)$, then the original system is consistent and the minimum residual is $\mathbf{0}$. Otherwise, $A\hat{\mathbf{x}}$ is the closest point in $\mathrm{Col}(A)$ to $\mathbf{b}$.

The projection property gives

$$\mathbf{b} - A\hat{\mathbf{x}} \in \mathrm{Col}(A)^{\perp}.$$

Since

$$\mathrm{Col}(A)^{\perp} = \mathrm{Nul}(A^T),$$

we must have

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}.$$

Expanding,

$$A^T\mathbf{b} - A^TA\hat{\mathbf{x}} = \mathbf{0},$$

so every least-squares solution satisfies

$$A^TA\hat{\mathbf{x}} = A^T\mathbf{b}.$$

These equations are called the **normal equations** for $A\mathbf{x} = \mathbf{b}$.

Least-Squares Solutions and the Normal Equations

Let A be an $m \times n$ matrix and let $\mathbf{b} \in \mathbb{R}^m$. A vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if it satisfies the normal equations

$$A^TA\hat{\mathbf{x}} = A^T\mathbf{b}.$$

The normal equations are always consistent, so every system $A\mathbf{x} = \mathbf{b}$ has at least one least-squares solution.

Proof

Suppose first that $\hat{\mathbf{x}}$ is a least-squares solution. Then $A\hat{\mathbf{x}}$ is the orthogonal projection of $\mathbf{b}$ onto $\mathrm{Col}(A)$. Therefore,

$$\mathbf{b} - A\hat{\mathbf{x}} \in \mathrm{Col}(A)^{\perp} = \mathrm{Nul}(A^T).$$

Hence,

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0},$$

which gives

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}.$$

Conversely, suppose that $\hat{\mathbf{x}}$ satisfies

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}.$$

Then

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0},$$

so

$$\mathbf{b} - A\hat{\mathbf{x}} \in \text{Nul}(A^T) = \text{Col}(A)^\perp.$$

Because $A\hat{\mathbf{x}} \in \text{Col}(A)$, the vector $A\hat{\mathbf{x}}$ is the orthogonal projection of $\mathbf{b}$ onto $\text{Col}(A)$. By the Best Approximation Theorem, $A\hat{\mathbf{x}}$ is the unique vector in $\text{Col}(A)$ closest to $\mathbf{b}$. Therefore, $\hat{\mathbf{x}}$ is a least-squares solution.

Finally, the projection $\hat{\mathbf{b}} = \text{proj}_{\text{Col}(A)}\,\mathbf{b}$ always exists. Since $\hat{\mathbf{b}} \in \text{Col}(A)$, there is at least one vector $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. Thus, the normal equations always have at least one solution. □

Uniqueness of a Least-Squares Solution

Let A be an $m \times n$ matrix. The following statements are equivalent.

(i) For every $\mathbf{b} \in \mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution.

(ii) The columns of A are linearly independent.

(iii) The matrix $A^T A$ is invertible.

When these statements are true, the unique least-squares solution is

$$\hat{\mathbf{x}} = \left(A^T A\right)^{-1} A^T\mathbf{b}.$$

Proof

The matrix A^TA is invertible if and only if

$$\text{Nul}(A^TA) = \{\mathbf{0}\}.$$

For any $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x}^TA^TA\mathbf{x} = (A\mathbf{x})^T(A\mathbf{x}) = \|A\mathbf{x}\|^2.$$

Therefore,

$$A^TA\mathbf{x} = \mathbf{0}$$

if and only if

$$A\mathbf{x} = \mathbf{0}.$$

Thus,

$$\text{Nul}(A^TA) = \text{Nul}(A).$$

Hence, A^TA is invertible if and only if $\text{Nul}(A) = \{\mathbf{0}\}$, which holds if and only if the columns of A are linearly independent.

When A^TA is invertible, the normal equations have the unique solution

$$\hat{\mathbf{x}} = \left(A^TA\right)^{-1}A^T\mathbf{b}.$$

If the columns of A are linearly dependent, then $\text{Nul}(A)$ contains a nonzero vector $\mathbf{z}$. If $\hat{\mathbf{x}}$ is one least-squares solution, then

$$A\left(\hat{\mathbf{x}} + \mathbf{z}\right) = A\hat{\mathbf{x}} + A\mathbf{z} = A\hat{\mathbf{x}},$$

so $\hat{\mathbf{x}} + \mathbf{z}$ is another least-squares solution. Therefore, the least-squares solution is unique precisely when the columns of A are linearly independent. □

Example 6.3.1

Find a least-squares solution of the inconsistent system $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}.$$

Solution

A least-squares solution satisfies the normal equations $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$. First, compute

$$A^TA = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix},$$

and

$$A^T\mathbf{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

$$= \begin{bmatrix} 19 \\ 11 \end{bmatrix}.$$

Thus, the normal equations are

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}.$$

Since A^TA is invertible,

$$\left(A^TA\right)^{-1} = \frac{1}{84}\begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}.$$

Therefore,

$$\begin{aligned} \hat{\mathbf{x}} &= \left(A^TA\right)^{-1}A^T\mathbf{b} \\ &= \frac{1}{84}\begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}\begin{bmatrix} 19 \\ 11 \end{bmatrix} \\ &= \frac{1}{84}\begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}. \end{aligned}$$

Example 6.3.2

Find the least-squares solutions of $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}.$$

Solution

The least-squares solutions satisfy the normal equations $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$. First, compute

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix},$$

and

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}.$$

Thus, the augmented matrix for the normal equations is

$$\begin{bmatrix} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and solving yields

$$\hat{x}_1 = 3 - \hat{x}_4, \quad \hat{x}_2 = -5 + \hat{x}_4, \quad \hat{x}_3 = -2 + \hat{x}_4,$$

where $\hat{x}_4$ is free. Hence, the system has infinitely many least-squares solutions of the form

$$\hat{\mathbf{x}} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + \hat{x}_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

The least-squares solution is not unique because the columns of A are linearly dependent.

When a least-squares solution $\hat{\mathbf{x}}$ is used to approximate $\mathbf{b}$ by $A\hat{\mathbf{x}}$, the distance $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ is called the **least-squares error**. Geometrically, it is the perpendicular distance from $\mathbf{b}$ to $\mathrm{Col}(A)$, so it measures how closely the best possible vector $A\hat{\mathbf{x}}$ approximates $\mathbf{b}$.

Example 6.3.3

For

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix},$$

the least-squares solution is

$$\hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Determine the least-squares error.

Solution

The least-squares solution produces the approximation

$$A\hat{\mathbf{x}} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}.$$

The residual vector is therefore

$$\begin{aligned} \mathbf{r} &= \mathbf{b} - A\hat{\mathbf{x}} \\ &= \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} -2 \\ -4 \\ 8 \end{bmatrix}. \end{aligned}$$

Thus, the least-squares error is

$$\begin{aligned}\|\mathbf{b} - A\hat{\mathbf{x}}\| &= \left\| \begin{bmatrix} -2 \\ -4 \\ 8 \end{bmatrix} \right\| \\ &= \sqrt{(-2)^2 + (-4)^2 + 8^2} \\ &= \sqrt{84} = 2\sqrt{21}.\end{aligned}$$

Therefore, $A\hat{\mathbf{x}}$ lies a distance of $2\sqrt{21}$ from $\mathbf{b}$. No other vector in $\text{Col}(A)$ is closer to $\mathbf{b}$. See Figure 26.

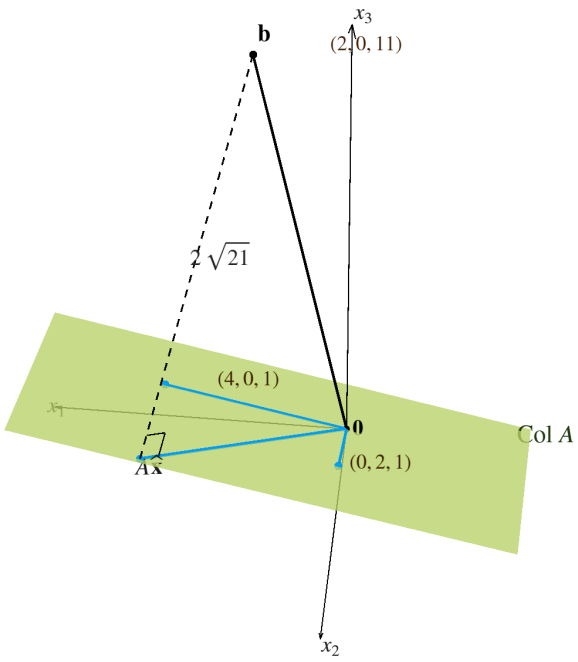


Figure 26: $A\hat{\mathbf{x}}$ is the orthogonal projection of $\mathbf{b}$ onto $\text{Col}\, A$.

Example 6.3.4

Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}.$$

Solution

Let $\mathbf{a}_1$ and $\mathbf{a}_2$ denote the columns of A. Since

$$\mathbf{a}_1 \cdot \mathbf{a}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -6 \\ -2 \\ 1 \\ 7 \end{bmatrix} = -6 - 2 + 1 + 7 = 0,$$

the columns of A form an orthogonal basis for $\text{Col}(A)$. Therefore, the closest vector in $\text{Col}(A)$ to $\mathbf{b}$ is

$$\begin{aligned}\hat{\mathbf{b}} = \operatorname{proj}_{\mathrm{Col}(A)} \mathbf{b} &= \frac{\mathbf{b}\cdot\mathbf{a}_1}{\mathbf{a}_1\cdot\mathbf{a}_1}\mathbf{a}_1 + \frac{\mathbf{b}\cdot\mathbf{a}_2}{\mathbf{a}_2\cdot\mathbf{a}_2}\mathbf{a}_2 \\ &= \frac{8}{4}\mathbf{a}_1 + \frac{45}{90}\mathbf{a}_2 \\ &= 2\begin{bmatrix}1\\1\\1\\1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}-6\\-2\\1\\7\end{bmatrix} \\ &= \begin{bmatrix}-1\\1\\\frac{5}{2}\\\frac{11}{2}\end{bmatrix}.\end{aligned}$$

Because $\hat{\mathbf{b}} = A\hat{\mathbf{x}}$, the entries of $\hat{\mathbf{x}}$ are precisely the coefficients of $\mathbf{a}_1$ and $\mathbf{a}_2$ in this orthogonal expansion; we do not have to solve it. Thus,

$$\hat{\mathbf{x}} = \begin{bmatrix}\frac{8}{4}\\\frac{45}{90}\end{bmatrix} = \begin{bmatrix}2\\\frac{1}{2}\end{bmatrix}.$$

This method avoids forming the normal equations because the columns of A are already orthogonal.

Numerically, the normal equations replaces A with A^TA, which squares the condition number and can significantly amplify rounding error. A QR factorization avoids this loss of accuracy. If $A = QR$ and $Q^TQ = I$, then the least-squares problem reduces to solving the upper-triangular system $R\hat{\mathbf{x}} = Q^T\mathbf{b}$.

Least-Squares Solutions Using a QR Factorization

Let A be an $m \times n$ matrix with linearly independent columns, and suppose $A = QR$ is a QR factorization of A, where the columns of Q are orthonormal and R is invertible and upper triangular. Then, for every $\mathbf{b} \in \mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution given by

$$\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}.$$

Equivalently, $\hat{\mathbf{x}}$ may be found by solving the upper-triangular system

$$R\hat{\mathbf{x}} = Q^T\mathbf{b}.$$

Example 6.3.5

Find the least-squares solution of $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix}.$$

Solution

A QR factorization of A is $A = QR$, where

$$Q = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}, \quad R = \begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix}.$$

The least-squares solution satisfies $R\hat{\mathbf{x}} = Q^T\mathbf{b}$. First, compute

$$Q^T\mathbf{b} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix}$$

$$= \begin{bmatrix} 6 \\ -6 \\ 4 \end{bmatrix}.$$

Therefore, $\hat{\mathbf{x}}$ satisfies

$$\begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \end{bmatrix} = \begin{bmatrix} 6 \\ -6 \\ 4 \end{bmatrix}.$$

Because R is upper triangular, the system can be solved by back substitution. The third equation gives

$$2\hat{x}_3 = 4,$$

so $\hat{x}_3 = 2$. The second equation then gives

$$2\hat{x}_2 + 3(2) = -6,$$

so $\hat{x}_2 = -6$. Finally, the first equation gives

$$2\hat{x}_1 + 4(-6) + 5(2) = 6,$$

so $\hat{x}_1 = 10$. Hence, the least-squares solution is

$$\hat{\mathbf{x}} = \begin{bmatrix} 10 \\ -6 \\ 2 \end{bmatrix}$$

You can also use MATLAB to solve the system.

```
>> R = [2, 4, 5;
        0, 2, 3;
        0, 0, 2];

>> Qtb = [ 6;
          -6;
           4];

>> R \ Qtb

ans =

    10
    -6
     2
```

Problem 6.3.1

Let

$$A = \begin{bmatrix} 1 & -3 & -3 \\ 1 & 5 & 1 \\ 1 & 7 & 2 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 5 \\ -3 \\ -5 \end{bmatrix}.$$

Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ and compute the associated least-squares error.

Solution

Before forming the normal equations, observe that the original system is consistent:

$$\begin{bmatrix} 1 & -3 & -3 & 5 \\ 1 & 5 & 1 & -3 \\ 1 & 7 & 2 & -5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -\frac{3}{2} & 2 \\ 0 & 1 & \frac{1}{2} & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus,

$$x_1 = 2 + \frac{3}{2}x_3, \quad x_2 = -1 - \frac{1}{2}x_3,$$

where x_3 is free. Choosing $x_3 = 0$ gives the least-squares solution

$$\hat{\mathbf{x}} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}.$$

Indeed,

$$A\hat{\mathbf{x}} = \begin{bmatrix} 1 & -3 & -3 \\ 1 & 5 & 1 \\ 1 & 7 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \\ -5 \end{bmatrix} = \mathbf{b}.$$

Therefore, the residual vector is $\mathbf{b} - A\hat{\mathbf{x}} = \mathbf{0}$, so the least-squares error is

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| = 0.$$

Because the original system is consistent, every exact solution is also a least-squares solution.

The following example shows how the classical Gram–Schmidt process can be used to solve a least-squares problem from beginning to end. Instead of forming the normal equations, we first apply Gram–Schmidt to the columns of A to construct an orthogonal basis for $\mathrm{Col}(A)$. This makes it possible to compute the closest vector to $\mathbf{b}$ in the column space by adding its projections onto the orthogonal basis vectors:

$$\hat{\mathbf{b}} = \mathrm{proj}_{\mathrm{Col}(A)}\, \mathbf{b}.$$

Once the closest vector $\hat{\mathbf{b}}$ is known, the least-squares solution is obtained by solving $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. Thus, the calculation separates the least-squares problem into two geometric steps: first find the best approximation to $\mathbf{b}$ in $\mathrm{Col}(A)$, then determine the coefficients that produce that approximation. This is the same underlying idea used by the QR method, but written out directly through Gram–Schmidt and orthogonal projection.

Example 6.3.6

Use the Gram–Schmidt process to find a least-squares solution of $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 3 \\ 1 \\ 2 \end{bmatrix}.$$

Then, compute the corresponding least-squares error.

Solution

Let $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ denote the columns of A:

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

We will first apply Gram–Schmidt to these columns to construct an orthogonal basis for $\text{Col}(A)$.

We need to first construct $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$

The first orthogonal vector is simply

$$\mathbf{v}_1 = \mathbf{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Thus, $\text{Span}\{\mathbf{v}_1\} = \text{Span}\{\mathbf{a}_1\}$.

To remove the component of $\mathbf{a}_2$ in the direction of $\mathbf{v}_1$, compute

$$\begin{aligned}\operatorname{proj}_{\operatorname{Span}\{\mathbf{v}_1\}} \mathbf{a}_2 &= \frac{\mathbf{a}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \\ &= \frac{3}{4}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbf{v}_2 &= \mathbf{a}_2 - \operatorname{proj}_{\operatorname{Span}\{\mathbf{v}_1\}} \mathbf{a}_2 \\ &= \begin{bmatrix}0\\1\\1\\1\end{bmatrix} - \frac{3}{4}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} \\ &= \frac{1}{4}\begin{bmatrix}-3\\1\\1\\1\end{bmatrix}.\end{aligned}$$

As expected,

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = \frac{1}{4}(-3+1+1+1) = 0.$$

Hence, $\mathbf{v}_2$ is orthogonal to $\mathbf{v}_1$.

The third vector must be orthogonal to both $\mathbf{v}_1$ and $\mathbf{v}_2$. First, project $\mathbf{a}_3$ onto $\operatorname{Span}\{\mathbf{v}_1\}$:

$$\begin{aligned}\operatorname{proj}_{\operatorname{Span}\{\mathbf{v}_1\}} \mathbf{a}_3 &= \frac{\mathbf{a}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \\ &= \frac{2}{4}\mathbf{v}_1 \\ &= \frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}.\end{aligned}$$

Next,

$$\mathbf{a}_3 \cdot \mathbf{v}_2 = \frac{1}{4}(1+1) = \frac{1}{2},$$

while

$$\mathbf{v}_2 \cdot \mathbf{v}_2 = \frac{1}{16}(9+1+1+1) = \frac{3}{4}.$$

Therefore,

$$\begin{aligned}\operatorname{proj}_{\operatorname{Span}\{\mathbf{v}_2\}} \mathbf{a}_3 &= \frac{\mathbf{a}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 \\ &= \frac{\frac{1}{2}}{\frac{3}{4}}\mathbf{v}_2 \\ &= \frac{2}{3}\mathbf{v}_2 \\ &= \frac{1}{6}\begin{bmatrix}-3\\1\\1\\1\end{bmatrix}.\end{aligned}$$

Subtracting both projections from $\mathbf{a}_3$ gives

$$\begin{aligned}\mathbf{v}_3 &= \mathbf{a}_3 - \operatorname{proj}_{\operatorname{Span}\{\mathbf{v}_1\}} \mathbf{a}_3 - \operatorname{proj}_{\operatorname{Span}\{\mathbf{v}_2\}} \mathbf{a}_3 \\ &= \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{6} \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}.\end{aligned}$$

The three vectors are mutually orthogonal because

$$\begin{aligned}\mathbf{v}_1 \cdot \mathbf{v}_2 &= 0, \\ \mathbf{v}_1 \cdot \mathbf{v}_3 &= \frac{1}{3}(-2+1+1) = 0, \\ \mathbf{v}_2 \cdot \mathbf{v}_3 &= \frac{1}{12}(-2+1+1) = 0.\end{aligned}$$

Moreover,

$$\operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \operatorname{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\} = \operatorname{Col}(A).$$

Thus, $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal basis for $\operatorname{Col}(A)$.

We now need to project $\mathbf{b}$ onto $\operatorname{Col}(A)$.

Because the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ form an orthogonal basis for $\operatorname{Col}(A)$,

$$\operatorname{proj}_{\operatorname{Col}(A)} \mathbf{b} = \operatorname{proj}_{\operatorname{Span}\{\mathbf{v}_1\}} \mathbf{b} + \operatorname{proj}_{\operatorname{Span}\{\mathbf{v}_2\}} \mathbf{b} + \operatorname{proj}_{\operatorname{Span}\{\mathbf{v}_3\}} \mathbf{b}.$$

For $\mathbf{v}_1$,

$$\mathbf{b} \cdot \mathbf{v}_1 = 1+3+1+2 = 7, \quad \mathbf{v}_1 \cdot \mathbf{v}_1 = 4,$$

so

$$\operatorname{proj}_{\operatorname{Span}\{\mathbf{v}_1\}} \mathbf{b} = \frac{7}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

For $\mathbf{v}_2$,

$$\mathbf{b} \cdot \mathbf{v}_2 = \frac{1}{4}(-3 + 3 + 1 + 2) = \frac{3}{4},$$

and $\mathbf{v}_2 \cdot \mathbf{v}_2 = \frac{3}{4}$. Hence,

$$\begin{aligned} \operatorname{proj}_{\operatorname{Span}\{\mathbf{v}_2\}} \mathbf{b} &= \frac{\frac{3}{4}}{\frac{3}{4}} \mathbf{v}_2 \\ &= \mathbf{v}_2 \\ &= \frac{1}{4} \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}. \end{aligned}$$

For $\mathbf{v}_3$,

$$\mathbf{b} \cdot \mathbf{v}_3 = \frac{1}{3}(-6 + 1 + 2) = -1,$$

while

$$\mathbf{v}_3 \cdot \mathbf{v}_3 = \frac{1}{9}(4 + 1 + 1) = \frac{2}{3}.$$

Therefore,

$$\begin{aligned}\operatorname{proj}_{\operatorname{Span}\{\mathbf{v}_3\}}\mathbf{b} &= \frac{-1}{\frac{2}{3}}\mathbf{v}_3 = -\frac{3}{2}\mathbf{v}_3 \\ &= -\frac{1}{2}\begin{bmatrix}0\\-2\\1\\1\end{bmatrix}.\end{aligned}$$

Combining the three projections, we get

$$\begin{aligned}\hat{\mathbf{b}} = \operatorname{proj}_{\operatorname{Col}(A)}\mathbf{b} &= \frac{7}{4}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} + \frac{1}{4}\begin{bmatrix}-3\\1\\1\\1\end{bmatrix} - \frac{1}{2}\begin{bmatrix}0\\-2\\1\\1\end{bmatrix} \\ &= \begin{bmatrix}1\\3\\\frac{3}{2}\\\frac{3}{2}\end{bmatrix}.\end{aligned}$$

The residual vector is

$$\begin{aligned}\mathbf{r} &= \mathbf{b} - \hat{\mathbf{b}} \\ &= \begin{bmatrix}1\\3\\1\\2\end{bmatrix} - \begin{bmatrix}1\\3\\\frac{3}{2}\\\frac{3}{2}\end{bmatrix} \\ &= \begin{bmatrix}0\\0\\-\frac{1}{2}\\\frac{1}{2}\end{bmatrix}.\end{aligned}$$

Indeed,

$$A^T\mathbf{r} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

which confirms that $\mathbf{r} \in \text{Col}(A)^\perp$.

We now need to solve $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$.

The vector $\hat{\mathbf{b}}$ is the closest vector in $\text{Col}(A)$ to $\mathbf{b}$. We now determine the coefficients $\hat{\mathbf{x}}$ that produce this vector:

$$A\hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 3 \\ \frac{3}{2} \\ \frac{3}{2} \end{bmatrix}.$$

Row reduce the corresponding augmented matrix:

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

Therefore, the least-squares solution is

$$\hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 2 \\ -\frac{3}{2} \end{bmatrix}.$$

The associated least-squares error is

$$\begin{aligned}\|\mathbf{b} - A\hat{\mathbf{x}}\| &= \|\mathbf{r}\| \\ &= \sqrt{\left(-\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2} \\ &= \frac{1}{\sqrt{2}}.\end{aligned}$$

Thus, Gram–Schmidt produces the same least-squares solution as the normal equations by first constructing an orthogonal basis for $\mathrm{Col}(A)$, projecting $\mathbf{b}$ onto that space, then solving for the coefficients of the projected vector.

6.4 Applications of Least Squares

The method of least squares was first published by Adrien-Marie Legendre in 1805 in *Nouvelles méthodes pour la détermination des orbites des comètes* to determine the orbits of comets. Carl Friedrich Gauss later claimed to have used it as early as 1795 and developed much of its statistical foundation, but there is no hard evidence that he used it to find Ceres.

Making a scatter plot is often the first step in analyzing a collection of data because it reveals possible relationships between variables. However, plotting the points alone does not determine a mathematical model. Every measurement contains noise, which can obscure real trends or make false trends look real.

Least squares provides a way for us to determine a model of best fit. After choosing a suitable model family, we determine its coefficients by minimizing the combined discrepancy between the observed values and the values predicted by the model. The resulting model balances the residuals across the entire data set to produce the best overall fit within the chosen family.

After you examine the shape of the data, the next step is to choose a family of functions capable of representing the observed trend. Least squares does not decide whether the relationship should be linear, sinusoidal, quadratic, or exponential; the choice is up to you. See Figure 27.

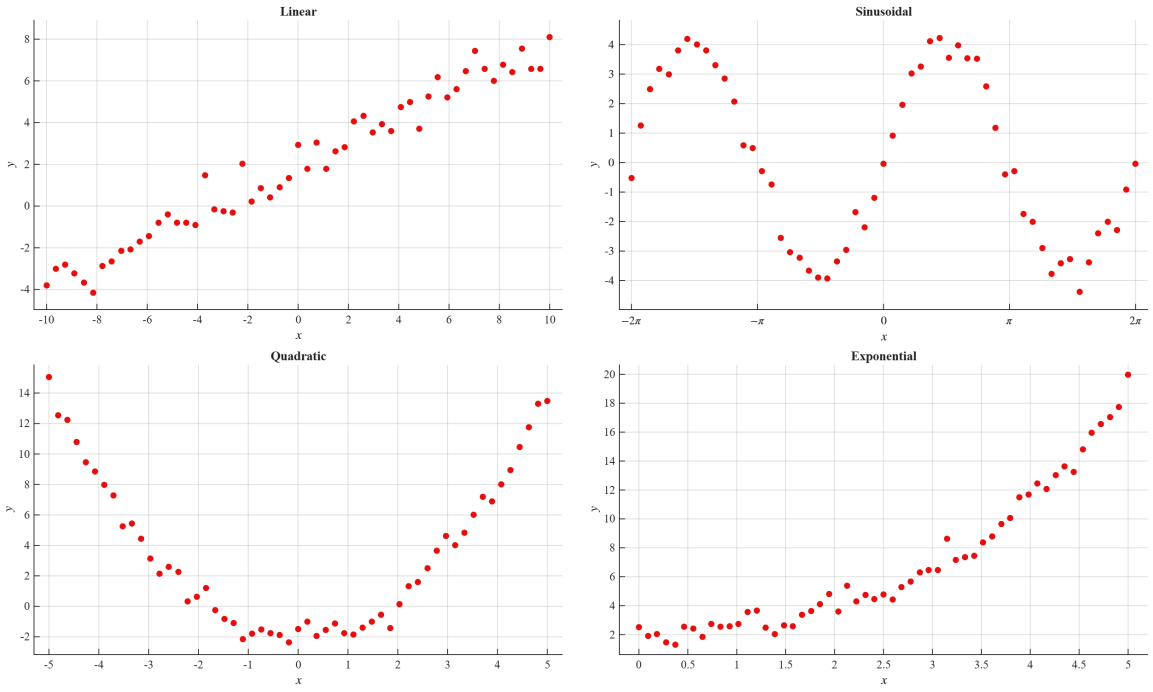


Figure 27: Possible models for data fitting

Many useful models are linear in their unknown coefficients, even when they are nonlinear in x. For example, $a_0 + a_1 \cos x + a_2 \sin x$ contains nonlinear functions of x, but the unknown coefficients a_0, a_1, a_2 still appear linearly. More generally, any model written as a linear combination of known functions can be fit using least squares.

Model Type	**General Equation**
Linear polynomial	$y(x) = a_0 + a_1 x$
Quadratic polynomial	$y(x) = a_0 + a_1 x + a_2 x^2$
nth-degree polynomial	$y(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$
Exponential function	$y(x) = a_0 e^{a_1 x}$
Sinusoidal function	$y(x) = a_0 + a_1 \cos(x) + a_2 \sin(x)$
Linear combination of functions	$y(x) = a_0 f_0(x) + a_1 f_1(x) + \cdots + a_n f_n(x)$

Table 8: Common model families used for least-squares fitting

The polynomial and sinusoidal models are special cases of the linear combination model in Table 8. The functions $f_0, f_1, \ldots, f_n$ are chosen in advance, while least squares determines the coefficients $a_0, a_1, \ldots, a_n$.

Suppose a collection of data points $(x_1, y_1), \ldots, (x_n, y_n)$ appears to follow a linear trend. We model the relationship between the variables using

$$y = \beta_0 + \beta_1 x,$$

where β_0 is the y-intercept and β_1 is the slope.

For each input x_i, the model predicts the value

$$\hat{y}_i = \beta_0 + \beta_1 x_i.$$

The difference between the observed value y_i and the predicted value $\hat{y}_i$ is the residual

$$r_i = y_i - \hat{y}_i = y_i - (\beta_0 + \beta_1 x_i)\,.$$

Because the x_i-values are treated as fixed, these residuals measure the vertical distances between the data points and the line. The least-squares line is the line whose coefficients β_0 and β_1 produce the smallest overall residual error.

If every data point lay exactly on one line, the coefficients would satisfy

$$\begin{aligned}\beta_0 + \beta_1 x_1 &= y_1,\\ \beta_0 + \beta_1 x_2 &= y_2,\\ &\vdots\\ \beta_0 + \beta_1 x_n &= y_n.\end{aligned}$$

In data fitting, this system is commonly written as

$$X\boldsymbol{\beta} = \mathbf{y},$$

where

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

The matrix X is called the design matrix, $\boldsymbol{\beta}$ is the parameter vector, and $\mathbf{y}$ is the observation vector. The first column of X multiplies the intercept β_0, while the second column multiplies the slope β_1.

Real data rarely lie exactly on a line, so the system $X\boldsymbol{\beta} = \mathbf{y}$ is generally inconsistent. For any proposed parameter vector $\boldsymbol{\beta}$, the predicted observations are collected in $X\boldsymbol{\beta}$, and the residual vector is

$$\mathbf{r} = \mathbf{y} - X\boldsymbol{\beta}.$$

The squared length of the residual vector is

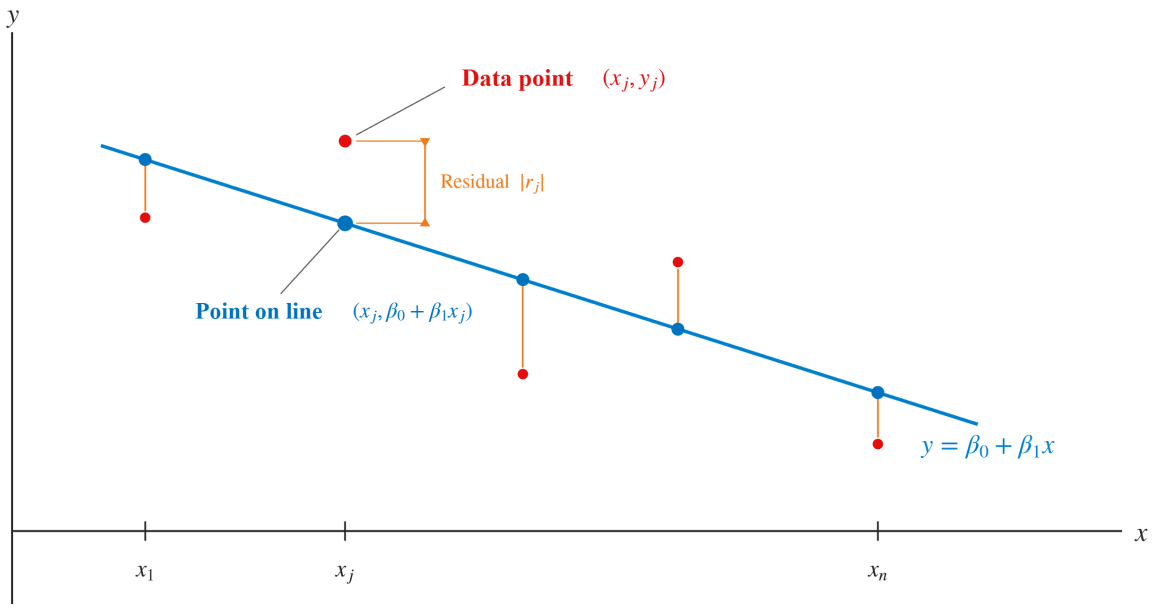
$$\begin{aligned}\|\mathbf{r}\|^2 &= \|\mathbf{y} - X\boldsymbol{\beta}\|^2\\ &= \sum_{i=1}^{n} \left[y_i - (\beta_0 + \beta_1 x_i)\right]^2\\ &= r_1^2 + r_2^2 + \cdots + r_n^2.\end{aligned}$$

Thus, fitting the best line is equivalent to finding the least-squares solution of the system $X\boldsymbol{\beta} = \mathbf{y}$. The optimal parameter vector $\hat{\boldsymbol{\beta}}$ minimizes the sum of the squared residuals and therefore satisfies the normal equations

$$X^T X\hat{\boldsymbol{\beta}} = X^T \mathbf{y}.$$

Once $\hat{\beta}_0$ and $\hat{\beta}_1$ have been determined, the least-squares line is

$$y = \hat{\beta}_0 + \hat{\beta}_1 x.$$



A least-squares line and its residuals

Example 6.4.1

Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points (2,1), (5,2), (7,3), and (8,3). Then, compute $\mathbf{r}$ and explain what it means.

Solution

The x-coordinates form the second column of the design matrix, while the first column accounts for the intercept β_0. The observed y-coordinates form the observation vector:

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}.$$

We must obtain the normal equations $X^T X\hat{\boldsymbol{\beta}} = X^T \mathbf{y}$.

First, compute

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix},$$

and

$$X^T \mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} 9 \\ 57 \end{bmatrix}.$$

Thus, the normal equations are

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}.$$

Since

$$\left(X^T X\right)^{-1} = \frac{1}{84}\begin{bmatrix} 142 & -22 \\ -22 & 4 \end{bmatrix},$$

we obtain

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(X^T X\right)^{-1} X^T \mathbf{y} \\ &= \frac{1}{84}\begin{bmatrix} 142 & -22 \\ -22 & 4 \end{bmatrix}\begin{bmatrix} 9 \\ 57 \end{bmatrix} \\ &= \frac{1}{84}\begin{bmatrix} 24 \\ 30 \end{bmatrix} = \begin{bmatrix} \frac{2}{7} \\ \frac{5}{14} \end{bmatrix}. \end{aligned}$$

Therefore, $\hat{\beta}_0 = \frac{2}{7}$ and $\hat{\beta}_1 = \frac{5}{14}$, so the least-squares line is

$$y = \frac{2}{7} + \frac{5}{14}x.$$

See Figure 28.

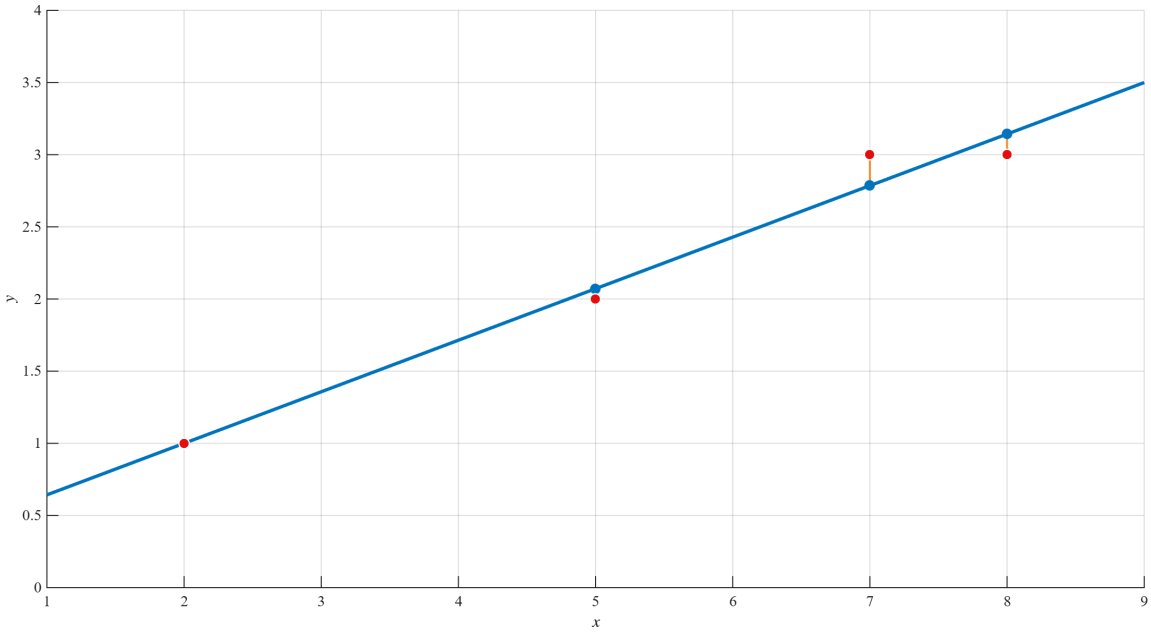


Figure 28: The least-squares line $y = \frac{2}{7} + \frac{5}{14}x$

The least-squares line predicts the y-coordinates

$$X\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} \frac{2}{7} \\ \frac{5}{14} \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{29}{14} \\ \frac{39}{14} \\ \frac{22}{7} \end{bmatrix}.$$

The residual vector is the difference between the observed values and the predicted values:

$$\begin{aligned}\mathbf{r} &= \mathbf{y} - X\hat{\boldsymbol{\beta}} \\ &= \begin{bmatrix}1\\2\\3\\3\end{bmatrix} - \begin{bmatrix}1\\\frac{29}{14}\\\frac{39}{14}\\\frac{22}{7}\end{bmatrix} \\ &= \begin{bmatrix}0\\-\frac{1}{14}\\\frac{3}{14}\\-\frac{1}{7}\end{bmatrix}.\end{aligned}$$

Each entry of $\mathbf{r}$ represents the vertical distance between a data point and the least-squares line. A positive residual indicates that the observed point lies above the line, while a negative residual indicates that it lies below the line. The least-squares method chooses the coefficients of the line so that the sum of the squares of these residuals,

$$\|\mathbf{r}\|^2 = 0^2 + \left(-\frac{1}{14}\right)^2 + \left(\frac{3}{14}\right)^2 + \left(-\frac{1}{7}\right)^2 = \frac{1}{14},$$

is as small as possible.

The least-squares method is not limited to fitting straight lines. Once a suitable model family has been chosen, the same procedure applies regardless of the shape of the function. The only difference is the structure of the design matrix.

Suppose we wish to fit the model

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + \cdots + \beta_k f_k(x),$$

where $f_0, f_1, \ldots, f_k$ are known functions. Although the model may be nonlinear in x, it remains a linear model because the unknown coefficients $\beta_0, \beta_1, \ldots, \beta_k$ appear only as linear combinations.

For the data points $(x_1, y_1), \ldots, (x_n, y_n)$, the design matrix becomes

$$X = \begin{bmatrix} f_0(x_1) & f_1(x_1) & \cdots & f_k(x_1) \\ f_0(x_2) & f_1(x_2) & \cdots & f_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_0(x_n) & f_1(x_n) & \cdots & f_k(x_n) \end{bmatrix},$$

with parameter vector

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}.$$

The least-squares solution is then obtained exactly as before by solving the normal equations

$$X^T X \hat{\boldsymbol{\beta}} = X^T \mathbf{y}.$$

As a simple example, fitting a quadratic polynomial,

$$y = \beta_0 + \beta_1 x + \beta_2 x^2,$$

produces the design matrix

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}.$$

The procedure is identical to fitting a line. Only the known functions, and therefore the columns of the design matrix, have changed.

Example 6.4.2

Suppose a collection of data points $(x_1, y_1), \ldots, (x_n, y_n)$ appears to follow a quadratic trend rather than a linear trend. Construct the corresponding least-squares model for the quadratic polynomial $y = \beta_0 + \beta_1 x + \beta_2 x^2$.

Solution

For each data point (x_i, y_i), the model predicts

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + r_i,$$

where r_i is the residual associated with the ith observation. Thus, the data must satisfy

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + r_1, \\ y_2 &= \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 + r_2, \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_n + \beta_2 x_n^2 + r_n. \end{aligned}$$

Writing these equations in matrix form gives

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}}_{\mathbf{r}}.$$

You would then solve $X^T X \hat{\boldsymbol{\beta}} = X^T \mathbf{y}$.

In many applications, a quantity depends on more than one independent variable. For example, suppose the response variable y depends on two predictors, u and v. A simple linear model is

$$y = \beta_0 + \beta_1 u + \beta_2 v,$$

while a more general model might be

$$y = \beta_0 + \beta_1 u + \beta_2 v + \beta_3 u^2 + \beta_4 uv + \beta_5 v^2.$$

Although these models contain products and powers of the variables, they remain linear in the unknown coefficients $\beta_0, \beta_1, \ldots$. Therefore, they can still be solved using the least-squares method.

For a data set consisting of measurements (u_i, v_i, y_i), the design matrix for the first model is

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1 & u_1 & v_1 \\ 1 & u_2 & v_2 \\ \vdots & \vdots & \vdots \\ 1 & u_n & v_n \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}}_{\mathbf{r}}.$$

As before, the least-squares solution is obtained from the normal equations.

The only difference from ordinary least-squares fitting is that each observation now contributes multiple predictor variables rather than a single x-value.

Example 6.4.3

A drone records the latitude, longitude, and elevation of several points in a small survey region. The first measurement, located at latitude 37.872997° N, longitude 122.255959° W, and elevation 2.1 m, is chosen as the origin of a local Cartesian coordinate system. The variable u denotes the northward distance from this reference point, while v denotes the eastward distance.

Find the least-squares plane of the form

$$y = \beta_0 + \beta_1 u + \beta_2 v$$

that best fits the converted data shown below using MATLAB.

u (m)	0.00	8.91	−7.79	17.81	34.51	24.49
v (m)	0.00	13.18	24.60	35.15	45.69	59.75
y (m)	2.1	4.0	3.8	6.2	8.1	7.9

Solution

The design matrix and observation vector are

$$X = \begin{bmatrix} 1 & 0.00 & 0.00 \\ 1 & 8.91 & 13.18 \\ 1 & -7.79 & 24.60 \\ 1 & 17.81 & 35.15 \\ 1 & 34.51 & 45.69 \\ 1 & 24.49 & 59.75 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 2.1 \\ 4.0 \\ 3.8 \\ 6.2 \\ 8.1 \\ 7.9 \end{bmatrix}.$$

Using MATLAB,

```
>> u = [ 0.00;
         8.91;
        -7.79;
        17.81;
        34.51;
        24.49];

>> v = [ 0.00;
        13.18;
        24.60;
        35.15;
        45.69;
        59.75];

>> y = [2.1;
        4.0;
        3.8;
        6.2;
        8.1;
        7.9];

>> X = [ones(size(u)), u, v];

>> betaHat = X \ y

betaHat =

    2.3910
```

```
    0.0672
    0.0702

>> XTX = X'*X

XTX =

   1.0e+03 *

    0.0060    0.0779    0.1784
    0.0779    2.2480    3.5919
    0.1784    3.5919    7.6720

>> XTy = X'*y

XTy =

   1.0e+03 *

    0.0321
    0.5895
    1.2062

>> yHat = X*betaHat

yHat =

    2.3910
    3.9147
    3.5936
    6.0545
    7.9166
    8.2297

>> r = y - yHat

r =

   -0.2910
    0.0853
    0.2064
    0.1455
    0.1834
   -0.3297

>> norm_r_squared = norm(r)^2
```

```
norm_r_squared =

    0.2981
```

Therefore, the least-squares plane is

$$y = 2.3910 + 0.0672u + 0.0702v,$$

where u is the northward distance, v is the eastward distance, and y is the predicted elevation, all measured in meters.

Figure 29 shows the fitted plane together with the observed data.

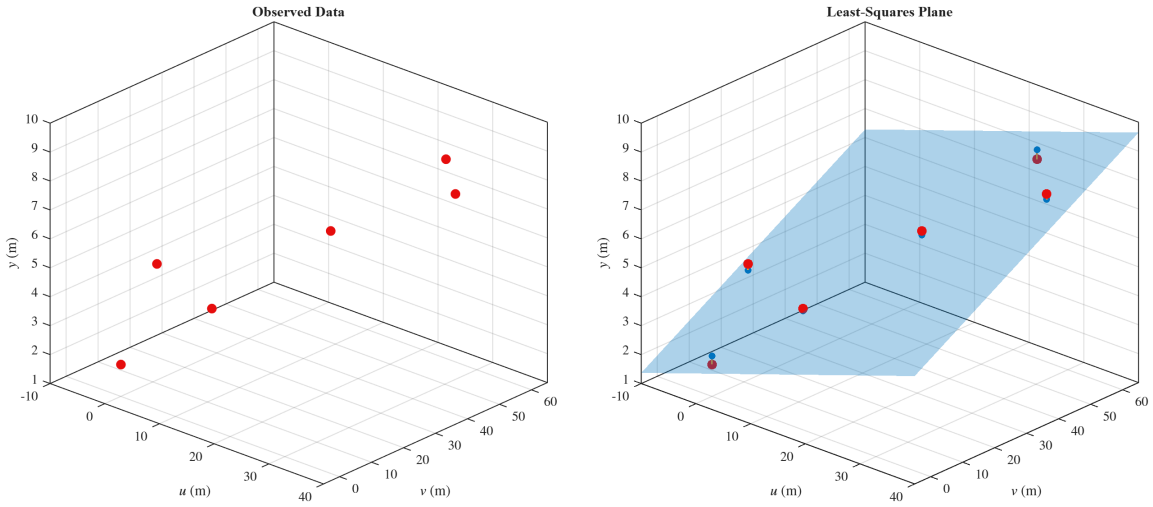


Figure 29

Problem 4.6.1

The monthly sales of a product exhibit both a long-term trend and seasonal fluctuations. Suppose the sales are modeled by

$$y = \beta_0 + \beta_1 x + \beta_2 \sin\left(\frac{2\pi x}{12}\right),$$

where x is measured in months. The term $\beta_0 + \beta_1 x$ represents the underlying sales trend, while the sine term models a seasonal cycle with period 12 months.

Given the data points $(x_1, y_1), \ldots, (x_n, y_n)$, construct the design matrix X and parameter vector $\boldsymbol{\beta}$ for the least-squares model $X\boldsymbol{\beta} \approx \mathbf{y}$.

Solution

For each observation (x_i, y_i), the predicted sales value is

$$\hat{y}_i = \beta_0 + \beta_1 x_i + \beta_2 \sin\left(\frac{2\pi x_i}{12}\right).$$

Thus, each row of the design matrix contains the coefficients multiplying β_0, β_1, and β_2. Therefore,

$$\underbrace{\begin{bmatrix} 1 & x_1 & \sin\left(\frac{2\pi x_1}{12}\right) \\ 1 & x_2 & \sin\left(\frac{2\pi x_2}{12}\right) \\ \vdots & \vdots & \vdots \\ 1 & x_n & \sin\left(\frac{2\pi x_n}{12}\right) \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}_{\boldsymbol{\beta}} \approx \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\mathbf{y}}.$$

See Figure 30 for an example of what this fitting could look like.

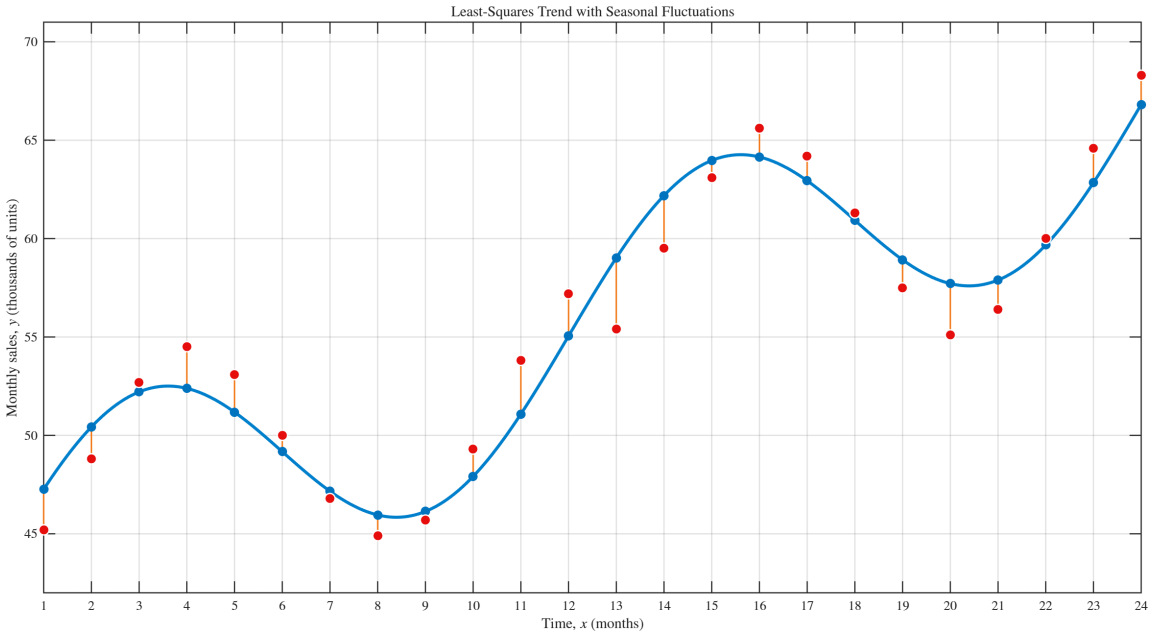


Figure 30: Long-term sales trend with seasonal fluctuations

Problem 4.6.2

A vibration sensor records the following displacement measurements over one cycle:

x	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	π	$\frac{5\pi}{4}$	$\frac{3\pi}{2}$	$\frac{7\pi}{4}$	2π
y (mm)	1.20	0.86	0.10	-0.82	-1.12	-0.76	-0.05	0.91	1.18

Determine the least-squares approximation of the form

$$y = B + C\cos(x) + D\sin(x) + E\cos(2x) + F\sin(2x).$$

Construct the design matrix and solve the problem using MATLAB.

Solution

Let

$$\boldsymbol{\beta} = \begin{bmatrix} B \\ C \\ D \\ E \\ F \end{bmatrix}.$$

The least-squares model can be written as

$$X\boldsymbol{\beta} \approx \mathbf{y},$$

where

$$X = \begin{bmatrix} 1 & \cos(0) & \sin(0) & \cos(0) & \sin(0) \\ 1 & \cos\left(\frac{\pi}{4}\right) & \sin\left(\frac{\pi}{4}\right) & \cos\left(\frac{\pi}{2}\right) & \sin\left(\frac{\pi}{2}\right) \\ 1 & \cos\left(\frac{\pi}{2}\right) & \sin\left(\frac{\pi}{2}\right) & \cos(\pi) & \sin(\pi) \\ 1 & \cos\left(\frac{3\pi}{4}\right) & \sin\left(\frac{3\pi}{4}\right) & \cos\left(\frac{3\pi}{2}\right) & \sin\left(\frac{3\pi}{2}\right) \\ 1 & \cos(\pi) & \sin(\pi) & \cos(2\pi) & \sin(2\pi) \\ 1 & \cos\left(\frac{5\pi}{4}\right) & \sin\left(\frac{5\pi}{4}\right) & \cos\left(\frac{5\pi}{2}\right) & \sin\left(\frac{5\pi}{2}\right) \\ 1 & \cos\left(\frac{3\pi}{2}\right) & \sin\left(\frac{3\pi}{2}\right) & \cos(3\pi) & \sin(3\pi) \\ 1 & \cos\left(\frac{7\pi}{4}\right) & \sin\left(\frac{7\pi}{4}\right) & \cos\left(\frac{7\pi}{2}\right) & \sin\left(\frac{7\pi}{2}\right) \\ 1 & \cos(2\pi) & \sin(2\pi) & \cos(4\pi) & \sin(4\pi) \end{bmatrix},$$

and

$$\mathbf{y} = \begin{bmatrix} 1.20 \\ 0.86 \\ 0.10 \\ -0.82 \\ -1.12 \\ -0.76 \\ -0.05 \\ 0.91 \\ 1.18 \end{bmatrix}.$$

Using MATLAB,

```
>> x = [0; pi/4; pi/2; 3*pi/4; pi; 5*pi/4; 3*pi/2; 7*pi/4; 2*pi];

>> y = [1.20; 0.86; 0.10; -0.82; -1.12; -0.76; -0.05; 0.91; 1.18];

>> X = [ones(size(x)), cos(x), sin(x), cos(2*x), sin(2*x)];

>> betaHat = X \ y

betaHat =

    0.0369
    1.1661
    0.0181
    0.0014
    0.0025
```

Because $\mathbf{x}$ is stored as a column vector, MATLAB evaluates each trigonometric function at every measurement simultaneously. Thus, the command defining `X` constructs the entire design matrix column by column, with one column for each basis function 1, $\cos(x)$, $\sin(x)$, $\cos(2x)$, and $\sin(2x)$. The command `X \ y` then computes the coefficient vector that minimizes the squared residual error. We don't have to enter everything in because MATLAB performs the repeated function evaluations, matrix construction, and least-squares computation automatically.

Therefore,

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 0.0369 \\ 1.1661 \\ 0.0181 \\ 0.0014 \\ 0.0025 \end{bmatrix},$$

and

$$\hat{y}(x) = 0.0369 + 1.1661\cos(x) + 0.0181\sin(x) + 0.0014\cos(2x) + 0.0025\sin(2x) \quad \text{mm}.$$

See Figure 31.

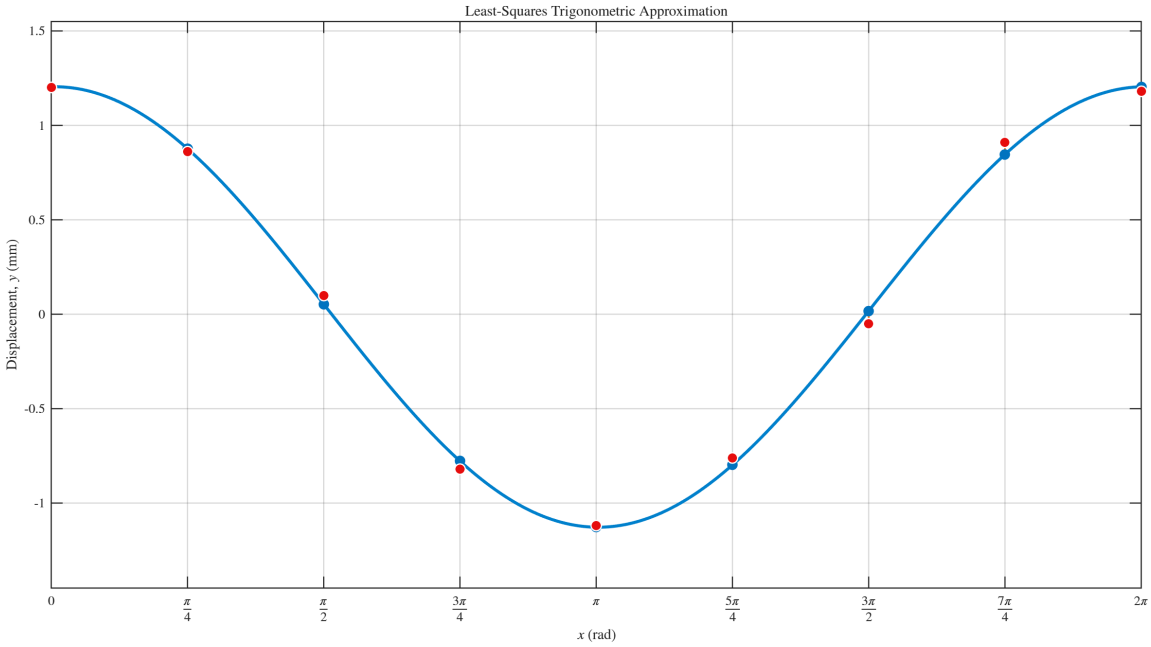
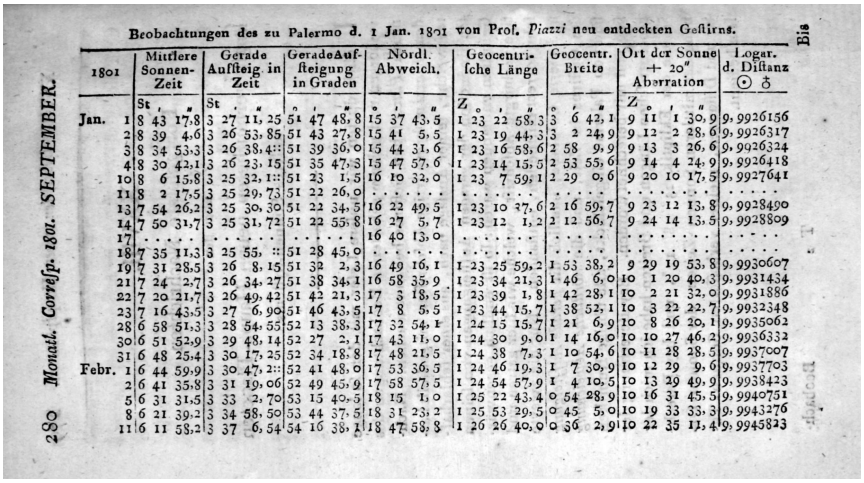


Figure 31

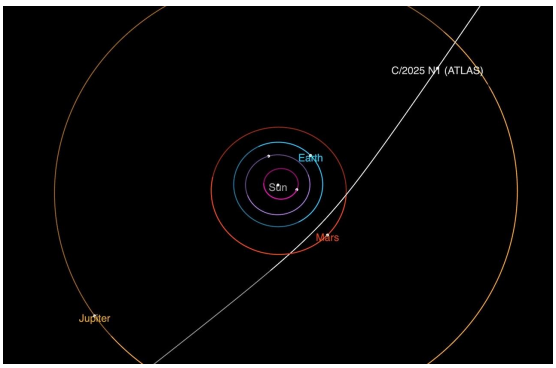


Beobachtungen des zu Palermo d. 1 Jan. 1801 von Prof. *Piazzi* neu entdeckten Gestirns.

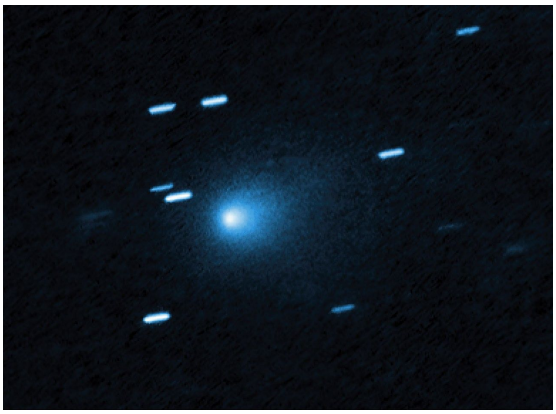
280 *Monatl. Corresp. 1801. SEPTEMBER.*

1801	Mittlere Sonnen-Zeit	Gerade Aufsteig. in Zeit	Gerade Aufsteigung in Graden	Nördl. Abweich.	Geocentrische Länge	Geocentr. Breite	Ort der Sonne + 20″ Aberration	Logar. d. Distanz ☉ ♁
	St ′ ″	St ′ ″	° ′ ″	° ′ ″	Z ° ′ ″	° ′ ″	Z ° ′ ″	
Jan. 1	8 43 17,8	3 27 11,25	51 47 48,8	15 37 43,5	1 23 22 58,3	3 6 42,1	9 11 1 30,9	9,9926156
2	8 39 4,6	3 26 53,85	51 43 27,8	15 41 5,5	1 23 19 44,3	3 2 24,9	9 12 2 28,6	9,9926317
3	8 34 53,3	3 26 38,4::	51 39 36,0	15 44 31,6	1 23 16 58,6	2 58 9,9	9 13 3 26,6	9,9926324
4	8 30 42,1	3 26 23,15	51 35 47,3	15 47 57,6	1 23 14 15,5	2 53 55,6	9 14 4 24,9	9,9926418
10	8 6 15,8	3 25 32,1::	51 23 1,5	16 10 32,0	1 23 7 59,1	2 29 0,6	9 20 10 17,5	9,9927641
11	8 2 17,5	3 25 29,73	51 22 26,0	. . . . .	. . . . .	. . . .	. . . . .	. . . . .
13	7 54 26,2	3 25 30,30	51 22 34,5	16 22 49,5	1 23 10 37,6	2 16 59,7	9 23 12 13,8	9,9928490
14	7 50 31,7	3 25 31,72	51 22 55,8	16 27 5,7	1 23 12 1,2	2 12 56,7	9 24 14 13,5	9,9928809
17	. . . . .	. . . . .	. . . . .	16 40 13,0	. . . . .	. . . .	. . . . .	. . . . .
18	7 35 11,3	3 25 55, ::	51 28 45,0	. . . . .	. . . . .	. . . .	. . . . .	. . . . .
19	7 31 28,5	3 26 8,15	51 32 2,3	16 49 16,1	1 23 25 59,2	1 53 38,2	9 29 19 53,8	9,9930607
21	7 24 2,7	3 26 34,27	51 38 34,1	16 58 35,9	1 23 34 21,3	1 46 6,0	10 1 20 40,3	9,9931434
22	7 20 21,7	3 26 49,42	51 42 21,3	17 3 18,5	1 23 39 1,8	1 42 28,1	10 2 21 32,0	9,9931886
23	7 16 43,5	3 27 6,90	51 46 43,5	17 8 5,5	1 23 44 15,7	1 38 52,1	10 3 22 22,7	9,9932348
28	6 58 51,3	3 28 54,55	52 13 38,3	17 32 54,1	1 24 15 15,7	1 21 6,9	10 8 26 20,1	9,9935062
30	6 51 52,9	3 29 48,14	52 27 2,1	17 43 11,0	1 24 30 9,0	1 14 16,0	10 10 27 46,2	9,9936332
31	6 48 25,4	3 30 17,25	52 34 18,8	17 48 21,5	1 24 38 7,3	1 10 54,6	10 11 28 28,5	9,9937007
Febr. 1	6 44 59,9	3 30 47,2::	52 41 48,0	17 53 36,5	1 24 46 19,3	1 7 30,9	10 12 29 9,6	9,9937703
2	6 41 35,8	3 31 19,06	52 49 45,9	17 58 57,5	1 24 54 57,9	1 4 10,5	10 13 29 49,9	9,9938423
5	6 31 31,5	3 33 2,70	53 15 40,5	18 15 1,0	1 25 22 43,4	0 54 28,9	10 16 31 45,5	9,9940751
8	6 21 39,2	3 34 58,50	53 44 37,5	18 31 23,2	1 25 53 29,5	0 45 5,0	10 19 33 33,3	9,9943276
11	6 11 58,2	3 37 6,54	54 16 36,1	18 47 58,8	1 26 26 40,0	0 36 2,9	10 22 35 11,4	9,9945823

Figure 32: Piazzi's published observations of Ceres from 1801. Gauss used observations from this data set to calculate the dwarf planet's orbit and predict its reappearance, allegedly employing the method of least squares. Image credit: Giuseppe Piazzi; Franz Xaver von Zach, Monatliche Correspondenz



(a) The trajectory of the interstellar comet 3I/ATLAS passing through the Solar System. It is one of three known interstellar comets to pass through the Solar System. Image credit: NASA



(b) Hubble captured 3I/ATLAS on July 25th, 2025 when it was 277 million miles from Earth. Image credit: NASA, ESA, David Jewitt

Problem 4.6.3

Ignoring gravitational effects from other celestial bodies, the orbit of a comet may be modeled as an ellipse, parabola, or hyperbola. In polar coordinates, its position (r, θ) satisfies

$$r = \beta + e\big(r\cos\theta\big),$$

where β is a constant and e is the eccentricity of the orbit. The orbit is classified according to the following table:

Eccentricity	Orbit Type
$0 \le e < 1$	Ellipse
$e = 1$	Parabola
$e > 1$	Hyperbola

Observations of a newly discovered comet produce the following data:

θ (rad)	0.88	1.10	1.42	1.77	2.14
r (AU)	3.00	2.30	1.65	1.25	1.01

Use least squares to estimate β and e. Then classify the comet's orbit and predict its radial distance when $\theta = 4.6$ radians.

Solution

Define $x_i = r_i \cos\theta_i$. The orbital model then becomes

$$r_i = \beta + e x_i.$$

Thus, the model is linear in the unknown parameters β and e. Let

$$\mathbf{c} = \begin{bmatrix} \beta \\ e \end{bmatrix}.$$

The corresponding least-squares system is

$$\underbrace{\begin{bmatrix} 1 & r_1\cos\theta_1 \\ 1 & r_2\cos\theta_2 \\ 1 & r_3\cos\theta_3 \\ 1 & r_4\cos\theta_4 \\ 1 & r_5\cos\theta_5 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \beta \\ e \end{bmatrix}}_{\mathbf{c}} \approx \underbrace{\begin{bmatrix} 3.00 \\ 2.30 \\ 1.65 \\ 1.25 \\ 1.01 \end{bmatrix}}_{\mathbf{r}}.$$

Using MATLAB,

```
>> theta = [0.88; 1.10; 1.42; 1.77; 2.14];

>> r = [3.00; 2.30; 1.65; 1.25; 1.01];

>> x = r.*cos(theta);

>> X = [ones(size(theta)), x];

>> betaHat = X \ r

betaHat =

    1.4509
    0.8111
```

Therefore,

$$\hat{\beta} = 1.4509, \quad \hat{e} = 0.8111.$$

Since $0 \leq \hat{e} < 1$, the fitted orbit is an ellipse.

To predict the radial distance, rearrange the orbital model:

$$r = \beta + er\cos\theta \Rightarrow r\left(1 - e\cos\theta\right) = \beta \Rightarrow r = \frac{\beta}{1 - e\cos\theta}$$

Substituting $\theta = 4.6$ radians into the fitted model gives

```
>> beta = betaHat(1);

>> e = betaHat(2);

>> thetaPred = 4.6;

>> rPred = beta/(1 - e*cos(thetaPred))

betaHat =

    1.4509
    0.8111

rPred =

    1.3300
```

Thus, the predicted radial distance is

$$r(4.6) \approx 1.3300 \text{ AU}.$$

Problem 4.6.4

In an undergraduate experiment, atomic absorption spectroscopy was used to determine the iron content of prune juice. Before analyzing the juice, several standard solutions with known iron concentrations were measured at a wavelength of 248.3 nm. The Week 3 calibration measurements are shown below.*

Iron Concentration, c (ppm)	1.0	2.0	3.0	4.0	5.0	6.0	10.0
Absorbance, A	0.041	0.078	0.110	0.147	0.180	0.216	0.350

(a) According to the Beer–Lambert law, absorbance is approximately linear in concentration. Use the model $A = \beta_0 + \beta_1 c$ to determine the least-squares calibration line in MATLAB.

(b) A prepared prune-juice sample produced a measured absorbance of $A = 0.124$. Use the fitted calibration line to estimate the iron concentration of the prepared sample solution in ppm. Then plot the calibration measurements and the least-squares line on the same figure.

*Data adapted from Grace Doe, Michael Giron, Michael Kulon, and Julia Wisniewski, *Determination of the Iron Content in Prune Juice*, Group R3 final laboratory report, University of Pennsylvania.

Solution

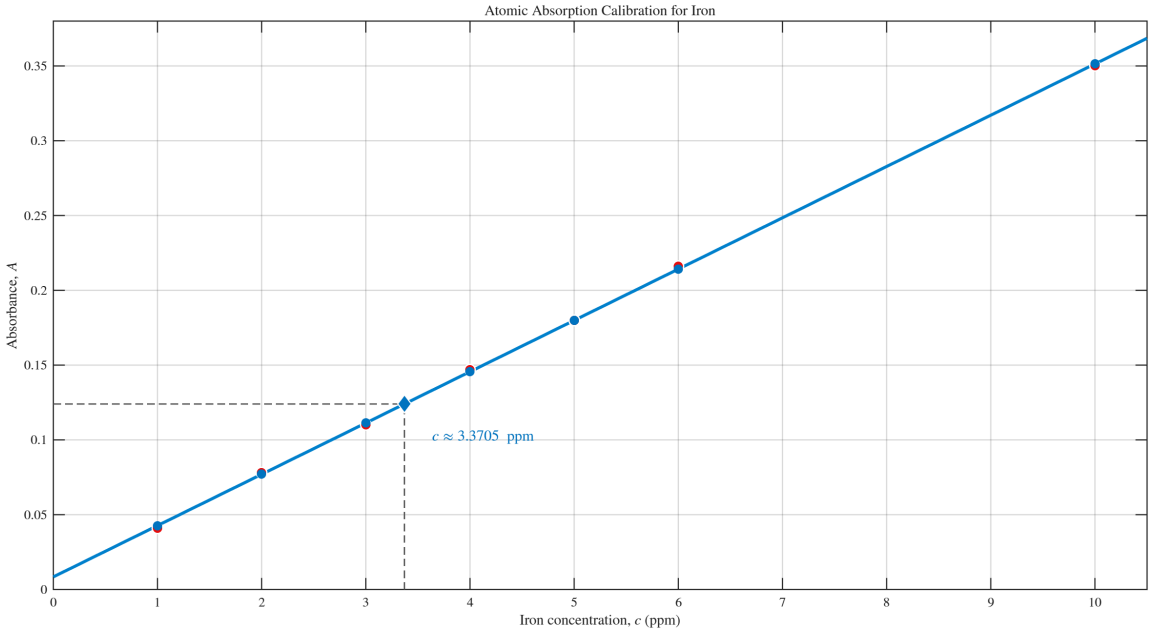
(a) The least-squares system is

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 10 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}}_{\boldsymbol{\beta}} \approx \underbrace{\begin{bmatrix} 0.041 \\ 0.078 \\ 0.110 \\ 0.147 \\ 0.180 \\ 0.216 \\ 0.350 \end{bmatrix}}_{\mathbf{A}}.$$

The least-squares coefficients and the concentration of the prepared sample are computed in MATLAB:

```
>> c = [1; 2; 3; 4; 5; 6; 10];

>> A = [0.041; 0.078; 0.110; 0.147; 0.180; 0.216; 0.350];

>> X = [ones(size(c)), c];

>> betaHat = X \ A

betaHat =

    0.0084
    0.0343

>> ASample = 0.124;

>> cSample = (ASample - betaHat(1))/betaHat(2)

cSample =

    3.3705
```

Therefore,



$$\hat{\beta}_0 = 0.0084, \quad \hat{\beta}_1 = 0.0343 \text{ ppm}^{-1},$$

and the least-squares calibration line is

$$\hat{A}(c) = 0.0084 + 0.0343c.$$

Absorbance is dimensionless, so the intercept is dimensionless and the slope has units of ppm^{-1}.

(b) Plug in $A = 0.124$. Solving the fitted model for c gives

$$c = \frac{A - \hat{\beta}_0}{\hat{\beta}_1} = \frac{0.124 - 0.0084}{0.0343} \approx 3.3705 \text{ ppm}.$$

Problem 4.6.5

Figure 34: A CMM. Image credit: Hexagon

A coordinate-measuring machine (CMM) is used to inspect a nominally circular hole machined into an aluminum plate. The specified diameter of the hole is 50.000 mm. A ruby touch probe records four points along the inner wall:

Point	x (mm)	y (mm)
1	25.02	0.01
2	0.03	24.98
3	−24.97	−0.02
4	−0.01	−25.03

Because of surface roughness, machine vibration, and thermal drift, the measured points do not lie exactly on a common circle. The equation of a circle with center (x_c, y_c) and radius R is

$$(x - x_c)^2 + (y - y_c)^2 = R^2.$$

Construct a least-squares system that estimates x_c, y_c, and R. Determine the fitted center, radius, and diameter of the hole, and graph the least-squares circle fit using MATLAB.

Solution

For each measured point (x_i, y_i),

$$(x_i - x_c)^2 + (y_i - y_c)^2 = R^2 \Rightarrow 2x_i x_c + 2y_i y_c + \left(R^2 - x_c^2 - y_c^2\right) = x_i^2 + y_i^2.$$

Define $\gamma = R^2 - x_c^2 - y_c^2$. The circle-fitting problem is now linear in the unknown parameters x_c,

y_c, and γ:

$$2x_i x_c + 2y_i y_c + \gamma \approx x_i^2 + y_i^2$$

Therefore, the least-squares system is

$$\underbrace{\begin{bmatrix} 50.04 & 0.02 & 1 \\ 0.06 & 49.96 & 1 \\ -49.94 & -0.04 & 1 \\ -0.02 & -50.06 & 1 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} x_c \\ y_c \\ \gamma \end{bmatrix}}_{\mathbf{p}} \approx \underbrace{\begin{bmatrix} 626.0005 \\ 624.0013 \\ 623.5013 \\ 626.5010 \end{bmatrix}}_{\mathbf{b}}.$$

MATLAB computes the least-squares estimate and recovers the radius as follows:

```
>> x = [25.02; 0.03; -24.97; -0.01];

>> y = [0.01; 24.98; -0.02; -25.03];

>> X = [2*x, 2*y, ones(size(x))];

>> b = x.^2 + y.^2;

>> pHat = X \ b

pHat =

    0.0250
   -0.0250
  624.9994

>> xc = pHat(1);

>> yc = pHat(2);

>> gamma = pHat(3);

>> R = sqrt(gamma + xc^2 + yc^2)

R =

   25.0000
```

```
>> D = 2*R

D =

   50.0000
```

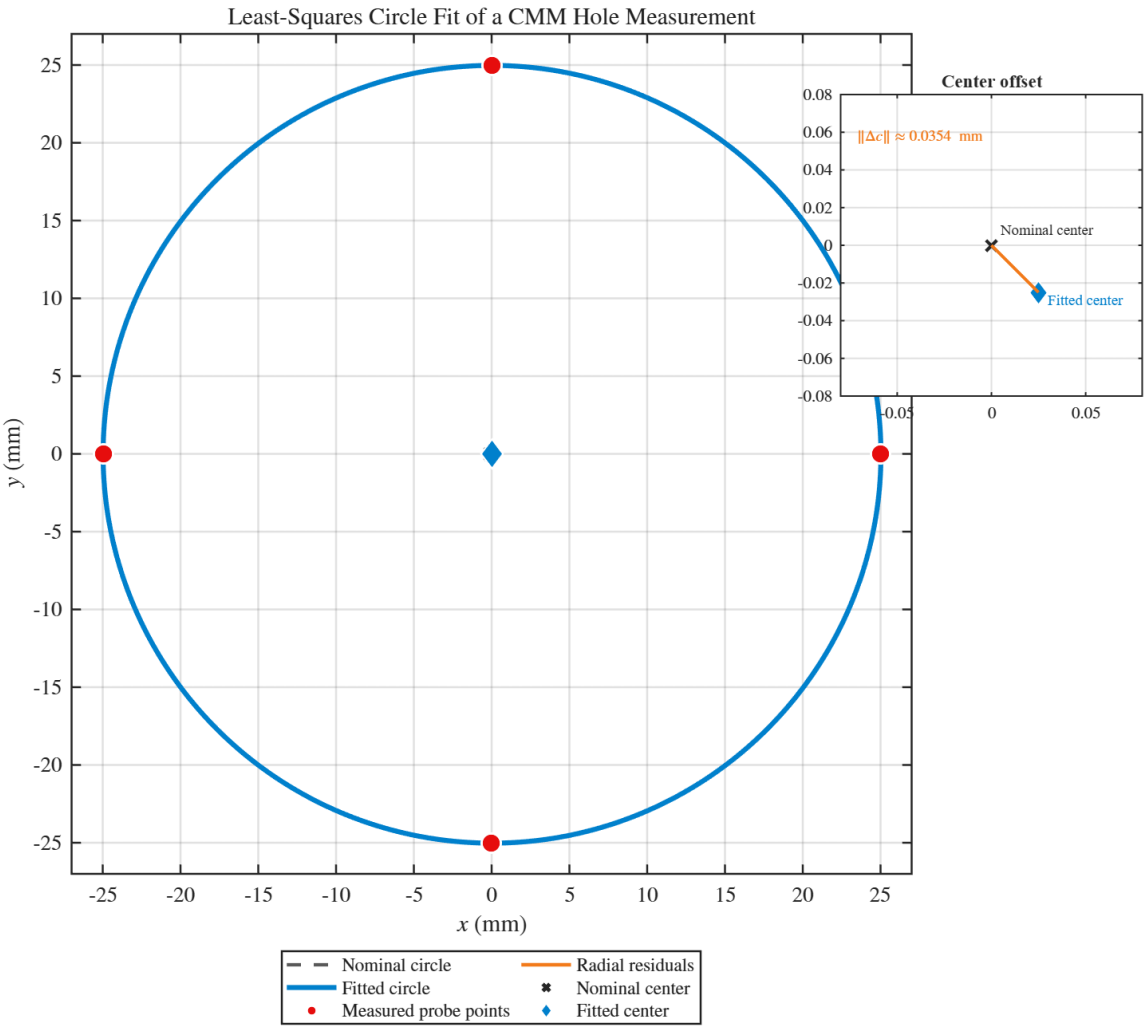
Thus, the estimated center of the hole is

$$(x_c, y_c) \approx (0.0250, -0.0250) \text{ mm}.$$

The fitted radius and diameter are

$$R \approx 25.0000 \text{ mm}, \quad D \approx 50.0000 \text{ mm}.$$

The unrounded calculation gives $D \approx 50.000026$ mm, which agrees with the nominal diameter of 50.000 mm to the reported measurement precision.



The fitted center is displaced from the nominal center $(0, 0)$ by

$$\sqrt{x_c^2 + y_c^2} = \sqrt{(0.0250)^2 + (-0.0250)^2} \approx 0.0354 \text{ mm}.$$

Thus, the hole is centered approximately 35.4 μm away from its nominal position. The radial residuals are only a few thousandths of a millimeter, so they are nearly invisible when the entire 50 mm diameter hole is shown. This visual overlap indicates that the measured points lie extremely close to the fitted circle.

6.5 Inner Product Spaces and Hilbert Spaces

The ideas of length, distance, and orthogonality do not belong only to $\mathbb{R}^n$. They can be extended to polynomials, functions, and many other vector spaces once we define an operation that plays the role of the dot product.

Inner Product Space

An **inner product** on a real vector space V is a function that assigns a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ to every pair of vectors $\mathbf{u}, \mathbf{v} \in V$. For all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and every scalar c, it must satisfy the following:

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$ — Symmetry
2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$ — Additivity
3. $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$ — Homogeneity
4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0, \quad \langle \mathbf{u}, \mathbf{u} \rangle = 0 \iff \mathbf{u} = \mathbf{0}$ — Positive definiteness

A vector space paired with an inner product is called an **inner product space**.

The standard dot product

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \mathbf{v}$$

makes $\mathbb{R}^n$ an inner product space. However, the inner product need not be the ordinary dot product. Different inner products assign different notions of length, angle, and importance to the components of a vector.

Example 6.5.1

Show that $\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1v_1 + 5u_2v_2$ defines an inner product on $\mathbb{R}^2$.

Solution

Let

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}.$$

Symmetry follows because multiplication of real numbers is commutative:

$$\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1v_1 + 5u_2v_2 = 4v_1u_1 + 5v_2u_2 = \langle \mathbf{v}, \mathbf{u} \rangle$$

Additivity and homogeneity follow from distributing through each component:

$$\begin{aligned} \langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle &= 4(u_1 + v_1)w_1 + 5(u_2 + v_2)w_2 \\ &= \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle \end{aligned}$$

and

$$\langle c\mathbf{u}, \mathbf{v} \rangle = 4(cu_1)v_1 + 5(cu_2)v_2 = c\langle \mathbf{u}, \mathbf{v} \rangle$$

Finally,

$$\langle \mathbf{u}, \mathbf{u} \rangle = 4u_1^2 + 5u_2^2 \geq 0,$$

and this quantity equals zero only when $u_1 = u_2 = 0$. Therefore, the four inner-product axioms are satisfied.

The coefficients 4 and 5 weight the two components differently. In applications such as weighted least squares, larger weights may be assigned to measurements considered more reliable or more important.

Polynomials and other functions may also be treated as vectors. Their inner products will not necessarily be computed by multiplying coordinates; instead, they may be defined using sums, integrals, or other operations that satisfy the same four axioms. Note that unbolded notation will be used when the vectors in an inner product space are explicitly polynomials, functions, or similar objects. Although they are not written in bold, they are still vectors if they are being treated as elements of a vector space.

Example 6.5.2

Let $\mathbb{P}_3$ have the inner product defined by evaluation at -3, -1, 1, and 3:

$$\langle p, q \rangle = p(-3)q(-3) + p(-1)q(-1) + p(1)q(1) + p(3)q(3).$$

Let $p_0(t) = 1$, $p_1(t) = t$, and $p_2(t) = t^2$.

(a) Compute the orthogonal projection of p_2 onto $W = \text{Span}\{p_0, p_1\}$.

(b) Find a polynomial q orthogonal to both p_0 and p_1 such that $\{p_0, p_1, q\}$ is an orthogonal basis for $\text{Span}\{p_0, p_1, p_2\}$. Scale q so that its values at -3, -1, 1, and 3 are 1, -1, -1, and 1, respectively.

Solution

(a) First observe that p_0 and p_1 are orthogonal because

$$\begin{aligned} \langle p_0, p_1 \rangle &= 1(-3) + 1(-1) + 1(1) + 1(3) \\ &= 0. \end{aligned}$$

Therefore, $\{p_0, p_1\}$ is an orthogonal basis for W.

The orthogonal projection of p_2 onto W is

$$\text{proj}_W p_2 = \frac{\langle p_2, p_0 \rangle}{\langle p_0, p_0 \rangle} p_0 + \frac{\langle p_2, p_1 \rangle}{\langle p_1, p_1 \rangle} p_1.$$

The required inner products are

$$\begin{aligned} \langle p_2, p_0 \rangle &= 9 + 1 + 1 + 9 = 20, \\ \langle p_0, p_0 \rangle &= 1 + 1 + 1 + 1 = 4, \\ \langle p_2, p_1 \rangle &= 9(-3) + 1(-1) + 1(1) + 9(3) = 0. \end{aligned}$$

Thus,

$$\begin{aligned} \text{proj}_W p_2 &= \frac{20}{4} p_0 + \frac{0}{20} p_1 \\ &= 5. \end{aligned}$$

(b) Subtract the projection from p_2:

$$q(t) = p_2(t) - \text{proj}_W\, p_2(t) = t^2 - 5.$$

Because q is the residual from an orthogonal projection, it is orthogonal to every polynomial in W, including p_0 and p_1. Therefore, $\{1, t, t^2 - 5\}$ is an orthogonal basis for $\text{Span}\{p_0, p_1, p_2\}$.

The values of $q(t) = t^2 - 5$ at the four evaluation points are

$$q(-3) = 4, \quad q(-1) = -4, \quad q(1) = -4, \quad q(3) = 4.$$

Dividing by 4 produces the requested values. Hence, the scaled polynomial is

$$q(t) = \frac{1}{4}(t^2 - 5),$$

and

$$\left\{1,\, t,\, \frac{1}{4}(t^2 - 5)\right\}$$

is the desired orthogonal basis.

Length, Distance, and Orthogonality

Let V be an inner product space.

The **norm** of a vector $\mathbf{v} \in V$ is

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}.$$

The **distance** between $\mathbf{u}$ and $\mathbf{v}$ is

$$\|\mathbf{u} - \mathbf{v}\|.$$

The vectors $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** when

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0.$$

A vector $\mathbf{v}$ is a **unit vector** when

$$\|\mathbf{v}\| = 1.$$

These definitions have exactly the same geometric meaning that they do in $\mathbb{R}^n$. The inner product determines the geometry of the vector space: once the inner product is known, notions of length, distance, and orthogonality follow.

Earlier in this chapter, the Gram–Schmidt process was developed for vectors in $\mathbb{R}^n$ using the standard dot product. However, none of its steps actually depend on coordinates or on the particular formula $\mathbf{u}^T\mathbf{v}$. They require only an inner product, so it applies to any finite-dimensional inner product space. Consequently, orthogonal bases, orthogonal projections, the orthogonal decomposition theorem, and the best approximation theorem all extend beyond $\mathbb{R}^n$.

For an orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ of a subspace W, the projection of $\mathbf{v}$ onto W remains

$$\operatorname{proj}_W \mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle}\mathbf{u}_1 + \cdots + \frac{\langle \mathbf{v}, \mathbf{u}_p \rangle}{\langle \mathbf{u}_p, \mathbf{u}_p \rangle}\mathbf{u}_p.$$

This inequality generalizes the familiar Cauchy–Schwarz inequality from $\mathbb{R}^n$. It guarantees that an inner product cannot be larger in magnitude than the product of the lengths of the two vectors.

Cauchy–Schwarz Inequality

For all vectors $\mathbf{u}$ and $\mathbf{v}$ in an inner product space,

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|.$$

An Inner Product on $C[a, b]$

Let $C[a, b]$ denote the vector space of all continuous real-valued functions on the interval $[a, b]$. For $f, g \in C[a, b]$, define

$$\langle f, g \rangle = \int_a^b f(t)g(t)\, dt.$$

This definition may be viewed as a continuous version of the dot product. A finite dot product multiplies corresponding coordinates and adds the results, while the integral inner product multiplies corresponding function values and accumulates them continuously across an interval:

$$\sum_{j=1}^{n} f(t_j^*)g(t_j^*)\, \Delta t_j \approx \int_a^b f(t)g(t)\, dt.$$

Symmetry, additivity, and homogeneity follow from the usual properties of definite integrals. Moreover,

$$\langle f, f \rangle = \int_a^b [f(t)]^2\, dt \geq 0.$$

Because f is continuous, this integral equals zero only when $f(t) = 0$ for every $t \in [a, b]$. Therefore, positive definiteness is also satisfied.

The induced norm is

$$\|f\| = \sqrt{\int_a^b [f(t)]^2\, dt},$$

and two functions f and g are orthogonal on $[a, b]$ when

$$\int_a^b f(t)g(t)\,dt = 0.$$

Example 6.5.3

Let W be the subspace of $C[0,1]$ spanned by

$$p_1(t) = 1, \quad p_2(t) = 2t - 1, \quad p_3(t) = 12t^2.$$

Use the Gram–Schmidt process to construct an orthogonal basis for W.

Solution

Begin with

$$q_1(t) = p_1(t) = 1.$$

Since

$$\langle p_2, q_1 \rangle = \int_0^1 (2t - 1)\,dt = 0,$$

the first two polynomials are already orthogonal. Thus,

$$q_2(t) = p_2(t) = 2t - 1.$$

To construct q_3, subtract from p_3 its projection onto $\text{Span}\{q_1, q_2\}$. The required inner products are

$$\langle p_3, q_1 \rangle = \int_0^1 12t^2\,dt = 4, \quad \langle q_1, q_1 \rangle = \int_0^1 1\,dt = 1,$$

and

$$\langle p_3, q_2 \rangle = \int_0^1 12t^2(2t - 1)\,dt = 2, \quad \langle q_2, q_2 \rangle = \int_0^1 (2t - 1)^2\,dt = \frac{1}{3}.$$

Therefore,

$$\begin{aligned} q_3 &= p_3 - \frac{\langle p_3, q_1 \rangle}{\langle q_1, q_1 \rangle} q_1 - \frac{\langle p_3, q_2 \rangle}{\langle q_2, q_2 \rangle} q_2 \\ &= 12t^2 - 4 - 6(2t - 1) \\ &= 12t^2 - 12t + 2. \end{aligned}$$

Hence, an orthogonal basis for W is

$$\left\{1,\ 2t - 1,\ 12t^2 - 12t + 2\right\}.$$

These polynomials are orthogonal with respect to the integral inner product on $C[0, 1]$, even though they would not generally be orthogonal under a different inner product.

Problem 6.5.1

Let $\mathbb{P}_4$ have the inner product defined by evaluation at -2, -1, 0, 1, and 2:

$$\langle p, q \rangle = \sum_{k=-2}^{2} p(k)q(k)$$

Suppose

$$p_0(t) = 1, \quad p_1(t) = t, \quad p_2(t) = t^2 - 2$$

form an orthogonal basis for a three-dimensional subspace of $\mathbb{P}_3$. Find a polynomial p_3 such that $\{p_0, p_1, p_2, p_3\}$ is an orthogonal basis for $\mathbb{P}_3$. Scale p_3 so that its values at -2, -1, 0, 1, and 2 are

$$(-1, 2, 0, -2, 1).$$

Solution

Apply the Gram–Schmidt process to t^3. Since t^3 is odd while $p_0(t) = 1$ and $p_2(t) = t^2 - 2$ are even,

$$\langle t^3, p_0 \rangle = 0, \quad \langle t^3, p_2 \rangle = 0.$$

Therefore, only the projection onto $p_1(t) = t$ must be removed:

$$p_3(t) = t^3 - \frac{\langle t^3, t \rangle}{\langle t, t \rangle} t$$

The required inner products are

$$\begin{aligned} \langle t^3, t \rangle &= (-2)^4 + (-1)^4 + 0^4 + 1^4 + 2^4 \\ &= 16 + 1 + 0 + 1 + 16 \\ &= 34, \end{aligned}$$

and

$$\begin{aligned} \langle t, t \rangle &= (-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2 \\ &= 4 + 1 + 0 + 1 + 4 \\ &= 10. \end{aligned}$$

Thus,

$$p_3(t) = t^3 - \frac{17}{5} t.$$

Its values at the five evaluation points are

$$p_3(-2) = -\frac{6}{5}, \quad p_3(-1) = \frac{12}{5}, \quad p_3(0) = 0, \quad p_3(1) = -\frac{12}{5}, \quad p_3(2) = \frac{6}{5}.$$

Multiplying by $\frac{5}{6}$ produces the requested vector of values. Therefore, choose

$$p_3(t) = \frac{5}{6} \left(t^3 - \frac{17}{5} t \right) = \frac{1}{6} \left(5t^3 - 17t \right).$$

Hence,

$$\left\{1,\, t,\, t^2-2,\, \frac{1}{6}\left(5t^3-17t\right)\right\}$$

is an orthogonal basis for P_3.

Problem 6.5.2

Let $V = C[-1, 1]$ have the inner product

$$\langle f, g\rangle = \int_{-1}^{1} f(t)g(t)\,dt.$$

Use the Gram–Schmidt process to find an orthogonal basis for

$$W = \text{Span}\{1, t, t^2\}.$$

Solution

Begin with

$$p_1(t) = 1, \quad p_2(t) = t, \quad p_3(t) = t^2.$$

Set $u_1(t) = p_1(t) = 1$. The second orthogonal polynomial is

$$u_2(t) = p_2(t) - \frac{\langle p_2, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1(t).$$

Since

$$\langle t, 1 \rangle = \int_{-1}^{1} t \, dt = 0,$$

we obtain $u_2(t) = t$.

The third orthogonal polynomial is

$$u_3(t) = p_3(t) - \frac{\langle p_3, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1(t) - \frac{\langle p_3, u_2 \rangle}{\langle u_2, u_2 \rangle} u_2(t).$$

The necessary inner products are

$$\langle t^2, 1 \rangle = \int_{-1}^{1} t^2 \, dt = \frac{2}{3},$$

$$\langle 1, 1 \rangle = \int_{-1}^{1} 1 \, dt = 2,$$

and

$$\langle t^2, t \rangle = \int_{-1}^{1} t^3 \, dt = 0.$$

Therefore,

$$u_3(t) = t^2 - \frac{2/3}{2} - 0$$
$$= t^2 - \frac{1}{3}.$$

Thus, one orthogonal basis for W is

$$\left\{1,\, t,\, t^2 - \frac{1}{3}\right\}.$$

Multiplying the third polynomial by 3 gives the equivalent orthogonal basis

$$\left\{1,\, t,\, 3t^2 - 1\right\}.$$

These polynomials are scalar multiples of the first three Legendre polynomials.

We have learned how an inner product space provides notions of length, distance, and orthogonality. In finite-dimensional spaces, these geometric ideas behave exactly as expected. In infinite-dimensional spaces, however, a new issue may arise: a sequence of vectors can approach a limiting object that does not belong to the original space.

To prevent this, we require the space to be complete.

Hilbert Space

A sequence $\{\mathbf{v}_k\}$ in an inner product space V is called a **Cauchy sequence** if its terms become arbitrarily close to one another. More precisely, for every $\varepsilon > 0$, there exists an integer N such that

$$\|\mathbf{v}_m - \mathbf{v}_n\| < \varepsilon$$

whenever $m, n \geq N$.

The space V is **complete** if every Cauchy sequence in V converges to an element that also belongs to V.

A complete inner product space is called a **Hilbert space**.

Every finite-dimensional inner product space is complete. Therefore, $\mathbb{R}^n$, paired with the standard dot product or any other valid inner product, is a Hilbert space.

Completeness becomes especially important in infinite-dimensional spaces. Two fundamental examples are the sequence space ℓ^2 and the function space $L^2[a, b]$. In both spaces, limits of Cauchy sequences remain inside the space.

The space $C[a, b]$, equipped with the integral inner product

$$\langle f, g \rangle = \int_a^b f(t)g(t)\, dt,$$

is an inner product space, but it is not complete under the induced norm. A sequence of continuous functions may converge in this norm to a function that is not continuous. The larger space $L^2[a, b]$ includes these limits and is therefore a Hilbert space.

Example 6.5.4

Let H be a Hilbert space, and let M be a closed subspace of H. Suppose

$$v = m + z,$$

where $m \in M$ and $z \in M^\perp$. Show that m is the unique vector in M closest to v.

Solution

Let $w \in M$. Since $m, w \in M$, their difference $m - w$ also belongs to M. Because $z \in M^\perp$,

$$\langle z, m - w \rangle = 0.$$

Now,

$$v - w = m + z - w = (m - w) + z.$$

The vectors $m - w$ and z are orthogonal, so the Pythagorean theorem gives

$$\|v - w\|^2 = \|m - w\|^2 + \|z\|^2.$$

Since $\|m - w\|^2 \geq 0$,

$$\|v - w\|^2 \geq \|z\|^2 = \|v - m\|^2.$$

Therefore,

$$\|v - w\| \geq \|v - m\|,$$

so m is a closest vector to v in M.

Equality holds only when

$$\|m - w\|^2 = 0,$$

which implies $w = m$. Therefore, m is the unique vector in M closest to v.

Weighted least squares replaces the standard Euclidean norm with a weighted norm.

Let M be a symmetric positive-definite matrix. The weighted inner product induced by M is

$$\langle \mathbf{u}, \mathbf{v} \rangle_M = \mathbf{u}^T M \mathbf{v},$$

and the corresponding weighted norm is

$$\|\mathbf{v}\|_M = \sqrt{\mathbf{v}^T M \mathbf{v}}.$$

When M is diagonal,

$$M = \operatorname{diag}(\omega_1, \ldots, \omega_m), \qquad \omega_i > 0,$$

the weighted norm becomes

$$\|\mathbf{v}\|_M = \sqrt{\omega_1 v_1^2 + \cdots + \omega_m v_m^2}.$$

Thus, larger values of ω_i cause the corresponding components of $\mathbf{v}$ to have a stronger influence on its length.

Weighted least squares is useful when experimental observations are not equally reliable. Suppose that $A\mathbf{x}$ approximates a vector of observations $\mathbf{b}$, and let

$$\mathbf{r} = \mathbf{b} - A\mathbf{x}$$

denote the residual vector. Ordinary least squares minimizes the Euclidean length $\|\mathbf{r}\|^2$, treating every residual equally. Weighted least squares instead minimizes

$$\|\mathbf{r}\|_M^2 = (\mathbf{b} - A\mathbf{x})^T M(\mathbf{b} - A\mathbf{x}).$$

For a diagonal weight matrix, this objective becomes

$$\|\mathbf{b} - A\mathbf{x}\|_M^2 = \sum_{i=1}^{m} \omega_i \left(b_i - \hat{b}_i\right)^2,$$

where $\hat{\mathbf{b}} = A\mathbf{x}$. Larger weights cause the corresponding observations to influence the fitted model more strongly. In statistical applications, measurements with smaller uncertainty are commonly assigned larger weights.

To convert the weighted problem into an ordinary least-squares problem, define the row-scaling matrix

$$W = M^{1/2} = \operatorname{diag}\left(\sqrt{\omega_1}, \ldots, \sqrt{\omega_m}\right).$$

Because $W^T W = M$,

$$\begin{aligned} \|\mathbf{b} - A\mathbf{x}\|_M^2 &= (\mathbf{b} - A\mathbf{x})^T W^T W(\mathbf{b} - A\mathbf{x}) \\ &= \|W(\mathbf{b} - A\mathbf{x})\|^2. \end{aligned}$$

Therefore, weighted least squares is equivalent to applying ordinary least squares to the transformed system

$$WA\mathbf{x} = W\mathbf{b}.$$

The corresponding normal equations are

$$(WA)^T(WA)\hat{\mathbf{x}} = (WA)^T W\mathbf{b}.$$

Since $W^TW = M$, these simplify to

$$A^TMA\hat{\mathbf{x}} = A^TM\mathbf{b}.$$

Provided A^TMA is invertible, the weighted least-squares solution is

$$\hat{\mathbf{x}} = \left(A^TMA\right)^{-1}A^TM\mathbf{b}.$$

Example 6.5.5

Find the weighted least-squares line $\hat{y} = \beta_0 + \beta_1 x$ that best fits the data (-2,3), (-1,5), (0,5), (1,4), and (2,3).

Suppose the first three observations are more reliable and should contribute four times as much to the weighted sum of squared residuals as the final two observations. Compare the resulting line with the ordinary least-squares line.

Solution

The linear model can be written as $X\boldsymbol{\beta} \approx \mathbf{y}$, where

$$X = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 3 \\ 5 \\ 5 \\ 4 \\ 3 \end{bmatrix}.$$

To give the first three residuals four times as much influence, use the weight matrix

$$M = \operatorname{diag}(4, 4, 4, 1, 1).$$

The corresponding row-scaling matrix is

$$W = M^{1/2} = \operatorname{diag}(2, 2, 2, 1, 1).$$

Thus, weighted least squares is equivalent to applying ordinary least squares to the transformed system $WX\boldsymbol{\beta} = W\mathbf{y}$, where

$$WX = \begin{bmatrix} 2 & -4 \\ 2 & -2 \\ 2 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad W\mathbf{y} = \begin{bmatrix} 6 \\ 10 \\ 10 \\ 4 \\ 3 \end{bmatrix}.$$

The weighted normal equations are

$$(WX)^T(WX)\hat{\boldsymbol{\beta}} = (WX)^T W\mathbf{y}.$$

Computing each side gives

$$(WX)^T(WX) = \begin{bmatrix} 14 & -9 \\ -9 & 25 \end{bmatrix}, \quad (WX)^T W\mathbf{y} = \begin{bmatrix} 59 \\ -34 \end{bmatrix}.$$

Therefore,

$$\begin{bmatrix} 14 & -9 \\ -9 & 25 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 59 \\ -34 \end{bmatrix}.$$

Solving the system gives

$$\hat{\beta}_0 = \frac{1169}{269} \approx 4.35, \quad \hat{\beta}_1 = \frac{55}{269} \approx 0.20.$$

Hence, the weighted least-squares line is

$$\hat{y} = 4.35 + 0.20x.$$

For comparison, ordinary least squares treats all five observations equally and gives

$$\hat{y} = 4.00 - 0.10x.$$

The first three observations generally increase as x increases, while the final two decrease. Giving the first three observations greater weight therefore changes the fitted slope from negative to positive.

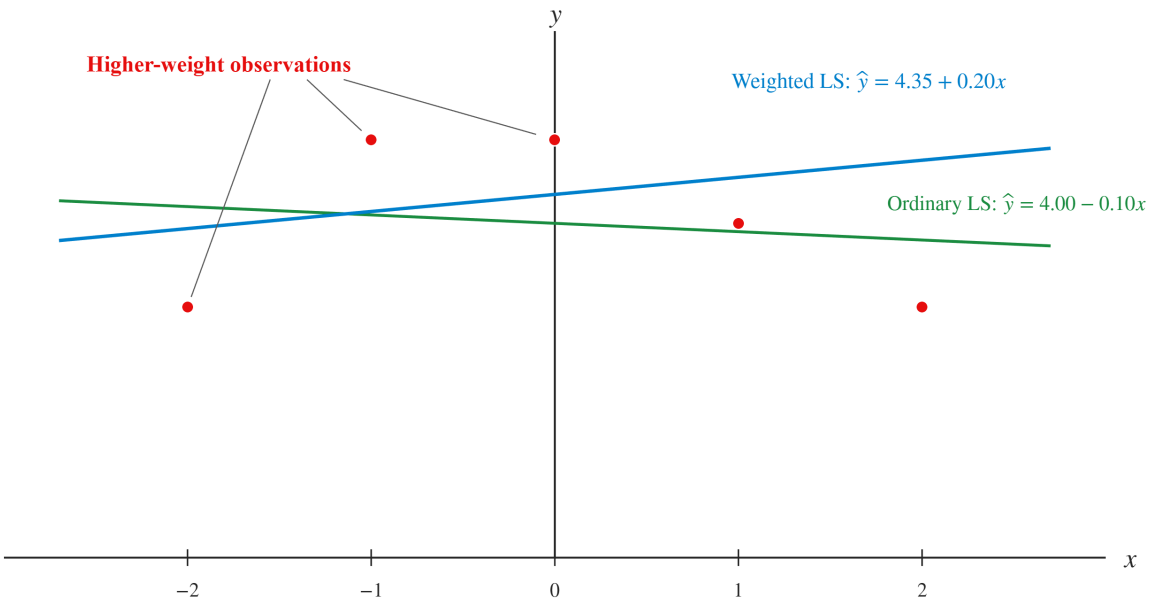


Figure 35: Weighted least squares penalizes errors made at more certain data points more heavily, so the fitted line is pulled closer to them.

Fourier series provide one of the most important applications of inner product spaces. A periodic function may often be approximated by a linear combination of sine and cosine functions. Such functions arise naturally in sound waves, electrical signals, and vibrating mechanical systems.

On the interval $[0, 2\pi]$, a function may be approximated by a **trigonometric polynomial** of the form

$$p_n(t) = \frac{a_0}{2} + \sum_{k=1}^{n} \left(a_k \cos(kt) + b_k \sin(kt)\right).$$

The functions

$$1,\ \cos t,\ \cos(2t),\ \ldots,\ \cos(nt),\ \sin t,\ \sin(2t),\ \ldots,\ \sin(nt)$$

form an orthogonal set in $C[0, 2\pi]$ under the inner product

$$\langle f, g \rangle = \int_0^{2\pi} f(t)g(t)\,dt.$$

This orthogonality allows the coefficients a_k and b_k to be found by projecting the function onto the sine and cosine functions, just as the coordinates of a vector are found by projecting it onto an orthogonal basis. A Fourier series therefore extends the familiar idea of orthogonal decomposition from finite-dimensional vectors to periodic functions.

Example 6.5.6

Let $C[0, 2\pi]$ have with the inner product

$$\langle f, g\rangle = \int_0^{2\pi} f(t)g(t)\,dt.$$

Let m and n be unequal positive integers. Show that $\cos(mt)$ and $\cos(nt)$ are orthogonal.

Solution

The two functions are orthogonal if their inner product equals zero. Using the product-to-sum identity

$$\cos(mt)\cos(nt) = \frac{1}{2}\left[\cos((m+n)t) + \cos((m-n)t)\right],$$

we obtain

$$\begin{aligned}\langle \cos(mt), \cos(nt)\rangle &= \int_0^{2\pi} \cos(mt)\cos(nt)\,dt \\ &= \frac{1}{2}\int_0^{2\pi} \left[\cos((m+n)t) + \cos((m-n)t)\right] dt.\end{aligned}$$

Integrating gives

$$\langle \cos(mt), \cos(nt)\rangle = \frac{1}{2}\left[\frac{\sin((m+n)t)}{m+n} + \frac{\sin((m-n)t)}{m-n}\right]_0^{2\pi}.$$

Because $m+n$ and $m-n$ are nonzero integers,

$$\sin(2\pi(m+n)) = \sin(2\pi(m-n)) = 0.$$

Therefore,

$$\langle \cos(mt), \cos(nt) \rangle = 0,$$

so $\cos(mt)$ and $\cos(nt)$ are orthogonal on $[0, 2\pi]$.

Let W_n be the subspace of $C[0, 2\pi]$ spanned by

$$1,\ \cos t,\ \cos(2t),\ \ldots,\ \cos(nt),\ \sin t,\ \sin(2t),\ \ldots,\ \sin(nt).$$

Because these functions form an orthogonal set under the inner product

$$\langle f, g \rangle = \int_0^{2\pi} f(t) g(t)\, dt,$$

the best approximation to a function $f \in C[0, 2\pi]$ by a function in W_n is its orthogonal projection onto W_n. This projection is called the n**th-order Fourier approximation** of f.

It has the form

$$p_n(t) = \frac{a_0}{2} + \sum_{k=1}^{n} \left[a_k \cos(kt) + b_k \sin(kt) \right],$$

where a_k and b_k are the **Fourier coefficients** of f.

Since

$$\langle \cos(kt), \cos(kt) \rangle = \int_0^{2\pi} \cos^2(kt)\, dt = \pi$$

and

$$\langle \sin(kt), \sin(kt) \rangle = \int_0^{2\pi} \sin^2(kt)\, dt = \pi,$$

the projection formulas give

$$a_k = \frac{\langle f, \cos(kt) \rangle}{\langle \cos(kt), \cos(kt) \rangle} = \frac{1}{\pi} \int_0^{2\pi} f(t) \cos(kt)\, dt, \qquad k \geq 1,$$

and

$$b_k = \frac{\langle f, \sin(kt)\rangle}{\langle \sin(kt), \sin(kt)\rangle} = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin(kt)\,dt, \qquad k \geq 1.$$

The coefficient of the constant function 1 is

$$\frac{\langle f, 1\rangle}{\langle 1, 1\rangle} = \frac{1}{2\pi}\int_0^{2\pi} f(t)\,dt.$$

Defining

$$a_0 = \frac{1}{\pi}\int_0^{2\pi} f(t)\,dt$$

allows this coefficient to be written as $a_0/2$. Thus, the Fourier coefficients are

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos(kt)\,dt, \qquad k \geq 0,$$

and

$$b_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin(kt)\,dt, \qquad k \geq 1.$$

Example 6.5.7

Find the nth-order Fourier approximation to the function

$$f(t) = t$$

on the interval $[0, 2\pi]$.

Solution

The constant term is

$$\begin{aligned}\frac{a_0}{2} &= \frac{1}{2\pi}\int_0^{2\pi} t\,dt \\ &= \frac{1}{2\pi}\left[\frac{t^2}{2}\right]_0^{2\pi} \\ &= \pi.\end{aligned}$$

For $k \geq 1$,

$$a_k = \frac{1}{\pi}\int_0^{2\pi} t\cos(kt)\,dt.$$

Using integration by parts, we get

$$\int t\cos(kt)\,dt = \frac{t}{k}\sin(kt) + \frac{1}{k^2}\cos(kt).$$

Therefore,

$$\begin{aligned}a_k &= \frac{1}{\pi}\left[\frac{t}{k}\sin(kt) + \frac{1}{k^2}\cos(kt)\right]_0^{2\pi} \\ &= 0.\end{aligned}$$

Similarly,

$$b_k = \frac{1}{\pi}\int_0^{2\pi} t\sin(kt)\,dt.$$

Using integration by parts, we get

$$\int t\sin(kt)\,dt = -\frac{t}{k}\cos(kt) + \frac{1}{k^2}\sin(kt).$$

Thus,

$$\begin{aligned}b_k &= \frac{1}{\pi}\left[-\frac{t}{k}\cos(kt) + \frac{1}{k^2}\sin(kt)\right]_0^{2\pi} \\ &= -\frac{2}{k}.\end{aligned}$$

Substituting these coefficients into the Fourier approximation gives

$$\begin{aligned} p_n(t) &= \frac{a_0}{2} + \sum_{k=1}^{n} \left[a_k \cos(kt) + b_k \sin(kt)\right] \\ &= \pi - 2\sum_{k=1}^{n} \frac{\sin(kt)}{k}. \end{aligned}$$

Therefore, the nth-order Fourier approximation to $f(t) = t$ is

$$p_n(t) = \pi - 2\sin t - \sin(2t) - \frac{2}{3}\sin(3t) - \cdots - \frac{2}{n}\sin(nt).$$

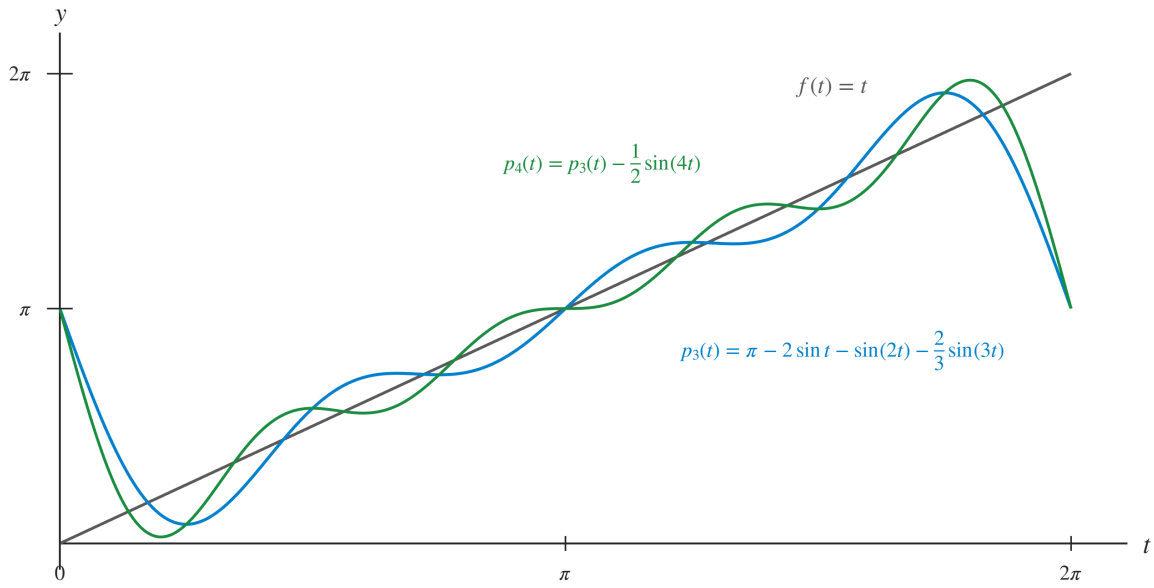


Figure 36: The third-order and fourth-order Fourier approximations of $f(t) = t$

The error between a function f and its nth-order Fourier approximation p_n is measured using the norm induced by the integral inner product:

$$\|f - p_n\|^2 = \int_0^{2\pi} \left[f(t) - p_n(t)\right]^2 dt$$

The corresponding **mean square error** is

$$\frac{1}{2\pi}\int_0^{2\pi}\left[f(t)-p_n(t)\right]^2\,dt.$$

Thus, the mean square error is proportional to the squared distance between f and its Fourier approximation in the inner product space. For functions in $L^2[0,2\pi]$, this error approaches zero as the order n increases.

This motivates the infinite expansion

$$f(t)=\frac{a_0}{2}+\sum_{m=1}^{\infty}\left[a_m\cos(mt)+b_m\sin(mt)\right],$$

which is called the **Fourier series** of f on $[0,2\pi]$.

Each term represents an orthogonal projection of f onto a one-dimensional subspace. For example, $a_m\cos(mt)$ is the projection of f onto the subspace spanned by $\cos(mt)$, while $b_m\sin(mt)$ is the projection onto the subspace spanned by $\sin(mt)$. A Fourier series therefore expresses a function as an infinite orthogonal decomposition into elementary oscillations.

Problem 6.5.3

Find the first-order and third-order Fourier approximations to

$$f(t)=3-2\sin t+5\sin(2t)-6\cos(2t)$$

on the interval $[0,2\pi]$.

Solution

The first-order Fourier approximation is the orthogonal projection of f onto

$$W_1 = \text{Span}\{1, \cos t, \sin t\}.$$

The terms $5\sin(2t)$ and $-6\cos(2t)$ are orthogonal to every function in W_1, so they do not contribute to the projection. Therefore,

$$p_1(t) = 3 - 2\sin t.$$

The third-order Fourier approximation is the orthogonal projection of f onto

$$W_3 = \text{Span}\{1, \cos t, \cos(2t), \cos(3t), \sin t, \sin(2t), \sin(3t)\}.$$

Because $f \in W_3$, its orthogonal projection onto W_3 is the function itself. Thus,

$$p_3(t) = 3 - 2\sin t + 5\sin(2t) - 6\cos(2t).$$

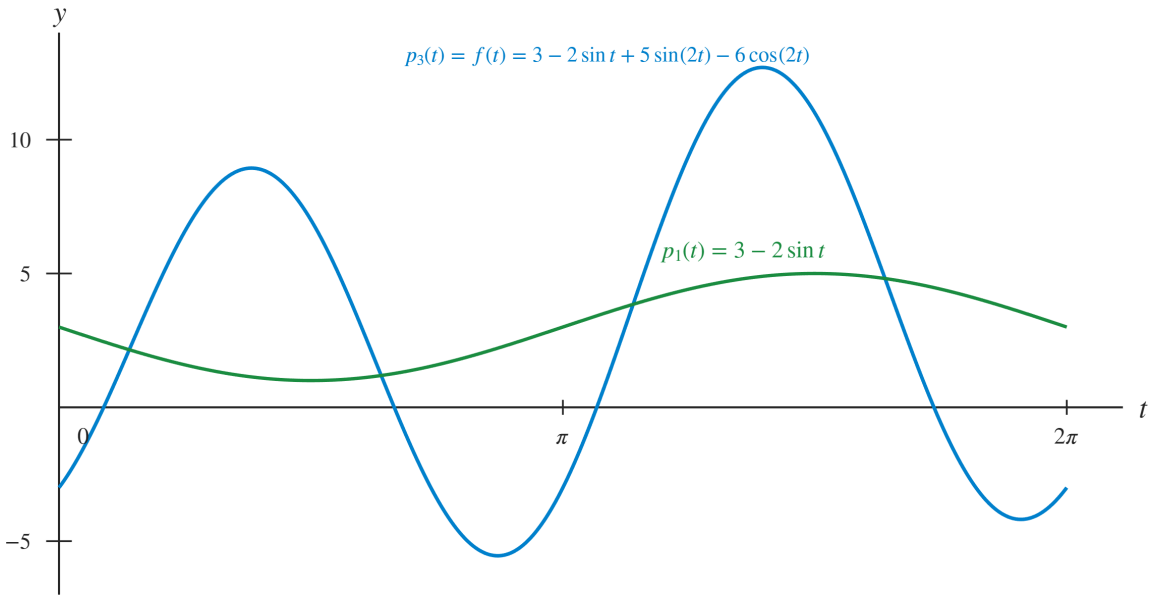


Figure 37: The first-order and third-order Fourier approximations of $f(t) = 3 - 2\sin t + 5\sin(2t) - 6\cos(2t)$

Problem 6.5.4

Find the third-order Fourier approximation to

$$f(t) = \cos^3 t$$

without evaluating any integrals.

Solution

Using the trigonometric identity

$$\cos(3t) = 4\cos^3 t - 3\cos t,$$

solve for $\cos^3 t$:

$$\cos^3 t = \frac{3}{4}\cos t + \frac{1}{4}\cos(3t)$$

This function already belongs to

$$W_3 = \text{Span}\{1, \cos t, \cos(2t), \cos(3t), \sin t, \sin(2t), \sin(3t)\}.$$

Therefore, its orthogonal projection onto W_3 is the function itself. Hence, the third-order Fourier approximation is

$$p_3(t) = \frac{3}{4}\cos t + \frac{1}{4}\cos(3t).$$

Problem 6.5.5

Explain why each Fourier coefficient of the sum of two functions is equal to the sum of the corresponding Fourier coefficients of the individual functions.

Solution

Let $f, g \in C[0, 2\pi]$. For $m \geq 0$, the cosine coefficient of $f + g$ is

$$\begin{aligned} a_m(f+g) &= \frac{1}{\pi}\int_0^{2\pi} [f(t)+g(t)]\cos(mt)\,dt \\ &= \frac{1}{\pi}\int_0^{2\pi} f(t)\cos(mt)\,dt + \frac{1}{\pi}\int_0^{2\pi} g(t)\cos(mt)\,dt \\ &= a_m(f) + a_m(g). \end{aligned}$$

Similarly, for $m \geq 1$,

$$\begin{aligned} b_m(f+g) &= \frac{1}{\pi}\int_0^{2\pi} [f(t)+g(t)]\sin(mt)\,dt \\ &= \frac{1}{\pi}\int_0^{2\pi} f(t)\sin(mt)\,dt + \frac{1}{\pi}\int_0^{2\pi} g(t)\sin(mt)\,dt \\ &= b_m(f) + b_m(g). \end{aligned}$$

Therefore, the Fourier coefficients of $f + g$ are the sums of the corresponding Fourier coefficients of f and g. This follows directly from the linearity of the integral, or equivalently, the linearity of the inner product.

7 Symmetric Matrices and the Singular Value Decomposition

7.1 Diagonalizing Symmetric Matrices

A matrix A is called **symmetric** when

$$A^T = A.$$

A symmetric matrix must be square. Its entries are reflected across the main diagonal, so

$$a_{i,j} = a_{j,i}$$

for every pair of indices i and j. The diagonal entries may be chosen arbitrarily.

Example 7.1.1

Determine which of the following matrices are symmetric:

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & 8 \\ 0 & 8 & -7 \end{bmatrix}, \quad C = \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix},$$

$$D = \begin{bmatrix} 1 & -3 \\ 3 & 0 \end{bmatrix}, \quad E = \begin{bmatrix} 1 & -4 & 0 \\ -6 & 1 & -4 \\ 0 & -6 & 1 \end{bmatrix}.$$

Solution

The matrices A, B, and C are symmetric because their entries agree across the main diagonal. That is, $a_{i,j} = a_{j,i}$.

The matrix D is not symmetric because $d_{1,2} = -3$, but $d_{2,1} = 3$. Similarly, E is not symmetric because $e_{1,2} = -4$, but $e_{2,1} = -6$.

Symmetric matrices possess stronger diagonalization properties than general square matrices. In particular, their eigenspaces interact naturally with the geometry developed in the previous chapter.

Example 7.1.2

Orthogonally diagonalize

$$A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}.$$

Solution

The characteristic polynomial is

$$\begin{aligned} \det(A - \lambda I) &= -\lambda^3 + 17\lambda^2 - 90\lambda + 144 \\ &= -(\lambda - 8)(\lambda - 6)(\lambda - 3). \end{aligned}$$

Therefore, the eigenvalues are

$$\lambda_1 = 8, \quad \lambda_2 = 6, \quad \lambda_3 = 3.$$

Corresponding eigenvectors are

$$\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

These eigenvectors are mutually orthogonal:

$$\mathbf{v}_1^T \mathbf{v}_2 = 0, \quad \mathbf{v}_1^T \mathbf{v}_3 = 0, \quad \mathbf{v}_2^T \mathbf{v}_3 = 0$$

Normalizing them gives

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\1\\0\end{bmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix}-1\\-1\\2\end{bmatrix}, \quad \mathbf{u}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}.$$

Let

$$P = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3\end{bmatrix} = \begin{bmatrix}-\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}}\\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}}\end{bmatrix},$$

and

$$D = \begin{bmatrix}8 & 0 & 0\\0 & 6 & 0\\0 & 0 & 3\end{bmatrix}.$$

Because the columns of P form an orthonormal basis,

$$P^{-1} = P^T.$$

Therefore,

$$A = PDP^T.$$

This is called an **orthogonal diagonalization** of A.

Orthogonality of Eigenvectors

Let A be a symmetric matrix. Eigenvectors of A corresponding to distinct eigenvalues are orthogonal.

Proof

Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be eigenvectors corresponding to distinct eigenvalues λ_1 and λ_2. Then

$$A\mathbf{v}_1 = \lambda_1\mathbf{v}_1, \quad A\mathbf{v}_2 = \lambda_2\mathbf{v}_2.$$

Consider the scalar $\lambda_1\mathbf{v}_1^T\mathbf{v}_2$. Since $A\mathbf{v}_1 = \lambda_1\mathbf{v}_1$,

$$\begin{aligned}\lambda_1\mathbf{v}_1^T\mathbf{v}_2 &= (A\mathbf{v}_1)^T\mathbf{v}_2 \\ &= \mathbf{v}_1^T A^T\mathbf{v}_2.\end{aligned}$$

Because A is symmetric, $A^T = A$. Therefore,

$$\begin{aligned}\lambda_1\mathbf{v}_1^T\mathbf{v}_2 &= \mathbf{v}_1^T A\mathbf{v}_2 \\ &= \mathbf{v}_1^T(\lambda_2\mathbf{v}_2) \\ &= \lambda_2\mathbf{v}_1^T\mathbf{v}_2.\end{aligned}$$

Thus,

$$(\lambda_1 - \lambda_2)\mathbf{v}_1^T\mathbf{v}_2 = 0.$$

Since $\lambda_1 \neq \lambda_2$,

$$\mathbf{v}_1^T\mathbf{v}_2 = 0.$$

Therefore, $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal. □

Orthogonal Diagonalization

An $n \times n$ matrix A is **orthogonally diagonalizable** if there exist an orthogonal matrix P and a diagonal matrix D such that

$$A = PDP^T.$$

The columns of P form an orthonormal basis of eigenvectors of A, while the diagonal entries of D are the corresponding eigenvalues.

If

$$A = PDP^T,$$

then

$$\begin{aligned} A^T &= (PDP^T)^T \\ &= PD^T P^T \\ &= PDP^T \\ &= A, \end{aligned}$$

because $D^T = D$. Therefore, every orthogonally diagonalizable matrix is symmetric.

The converse is one of the most critical results of linear algebra.

Spectral Theorem for Real Symmetric Matrices

Let A be an $n \times n$ real symmetric matrix. Then:

1. A has n real eigenvalues, counting multiplicities.
2. For each eigenvalue λ, the dimension of its eigenspace equals its algebraic multiplicity.
3. Eigenspaces corresponding to distinct eigenvalues are mutually orthogonal.
4. A is orthogonally diagonalizable

Example 7.1.3

Orthogonally diagonalize

$$A = \begin{bmatrix} 5 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}.$$

Solution

The characteristic polynomial is

$$\begin{aligned}\det(A-\lambda I) &= -\lambda^3 + 12\lambda^2 - 21\lambda - 98 \\ &= -(\lambda-7)^2(\lambda+2).\end{aligned}$$

Therefore, the eigenvalues are

$$\lambda_1 = 7, \quad \lambda_2 = -2.$$

A basis for the eigenspace corresponding to $\lambda = 7$ is

$$\mathbf{v}_1 = \begin{bmatrix}1\\0\\1\end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix}-\frac{1}{2}\\1\\0\end{bmatrix}.$$

These vectors are linearly independent but not orthogonal because $\mathbf{v}_1^T\mathbf{v}_2 = -\frac{1}{2}$.

Apply the Gram–Schmidt process. Set $\mathbf{z}_1 = \mathbf{v}_1$, and compute

$$\begin{aligned}\mathbf{z}_2 &= \mathbf{v}_2 - \frac{\mathbf{v}_2^T\mathbf{v}_1}{\mathbf{v}_1^T\mathbf{v}_1}\mathbf{v}_1 \\ &= \begin{bmatrix}-\frac{1}{2}\\1\\0\end{bmatrix} - \frac{-\frac{1}{2}}{2}\begin{bmatrix}1\\0\\1\end{bmatrix} \\ &= \begin{bmatrix}-\frac{1}{4}\\1\\\frac{1}{4}\end{bmatrix}.\end{aligned}$$

Thus, $\{\mathbf{z}_1, \mathbf{z}_2\}$ is an orthogonal basis for the eigenspace corresponding to $\lambda = 7$.

Normalizing gives

$$\mathbf{u}_1 = \frac{\mathbf{z}_1}{\|\mathbf{z}_1\|} = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix},$$

and

$$\mathbf{u}_2 = \frac{\mathbf{z}_2}{\|\mathbf{z}_2\|} = \frac{1}{\sqrt{18}} \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix}.$$

For $\lambda = -2$, one eigenvector is

$$\mathbf{v}_3 = \begin{bmatrix} -1 \\ -\frac{1}{2} \\ 1 \end{bmatrix}.$$

Normalizing gives

$$\mathbf{u}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \frac{1}{3} \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix}.$$

Because A is symmetric, eigenvectors corresponding to distinct eigenvalues are orthogonal. Therefore, $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal basis of eigenvectors for $\mathbb{R}^3$.

Let

$$P = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} & -\frac{2}{3} \\ 0 & \frac{4}{\sqrt{18}} & -\frac{1}{3} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{18}} & \frac{2}{3} \end{bmatrix}, \quad D = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

Since the columns of P are orthonormal, $P^{-1} = P^T$. Hence, $A = PDP^T$.

Suppose A is a real symmetric matrix with the orthogonal diagonalization $A = PDP^T$, where

$$P = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$$

contains an orthonormal basis of eigenvectors of A, and

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

contains the corresponding eigenvalues.

Since

$$P^T = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix},$$

we may write

$$\begin{aligned} A &= PDP^T \\ &= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\mathbf{u}_1 & \lambda_2\mathbf{u}_2 & \cdots & \lambda_n\mathbf{u}_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix}. \end{aligned}$$

Using the column–row expansion of a matrix product gives the following decomposition.

Spectral Decomposition of A

$$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T$$

This representation expresses the matrix as a sum of components determined by its eigenvalues and orthonormal eigenvectors.

Each matrix $\mathbf{u}_j \mathbf{u}_j^T$ has rank one and is the orthogonal projection matrix onto $\text{Span}\{\mathbf{u}_j\}$.

Indeed, for any $\mathbf{x} \in \mathbb{R}^n$,

$$\left(\mathbf{u}_j \mathbf{u}_j^T\right) \mathbf{x} = \left(\mathbf{u}_j^T \mathbf{x}\right) \mathbf{u}_j = \text{proj}_{\mathbf{u}_j} \mathbf{x},$$

because $\|\mathbf{u}_j\| = 1$.

Therefore,

$$A\mathbf{x} = \sum_{j=1}^{n} \lambda_j \left(\mathbf{u}_j^T \mathbf{x}\right) \mathbf{u}_j.$$

Thus, A first decomposes $\mathbf{x}$ into its orthogonal components along the eigenvector directions, scales each component by the corresponding eigenvalue, and then adds the scaled components together.

Example 7.1.4

Construct a spectral decomposition of

$$A = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix},$$

given the orthogonal diagonalization

$$A = \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 8 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}.$$

Solution

The columns of P are the orthonormal eigenvectors

$$\mathbf{u}_1 = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -\frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix},$$

with corresponding eigenvalues

$$\lambda_1 = 8, \quad \lambda_2 = 3.$$

Therefore, the spectral decomposition is

$$A = 8\mathbf{u}_1\mathbf{u}_1^T + 3\mathbf{u}_2\mathbf{u}_2^T.$$

The first rank-one projection matrix is

$$\begin{aligned} \mathbf{u}_1\mathbf{u}_1^T &= \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{4}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{bmatrix}. \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbf{u}_2\mathbf{u}_2^T &= \begin{bmatrix} -\frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{4}{5} \end{bmatrix}. \end{aligned}$$

Thus,

$$
\begin{aligned}
8\mathbf{u}_1\mathbf{u}_1^T + 3\mathbf{u}_2\mathbf{u}_2^T &= \begin{bmatrix} \frac{32}{5} & \frac{16}{5} \\ \frac{16}{5} & \frac{8}{5} \end{bmatrix} + \begin{bmatrix} \frac{3}{5} & -\frac{6}{5} \\ -\frac{6}{5} & \frac{12}{5} \end{bmatrix} \\
&= \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix} \\
&= A.
\end{aligned}
$$

Hence,

$$A = 8\mathbf{u}_1\mathbf{u}_1^T + 3\mathbf{u}_2\mathbf{u}_2^T$$

is the spectral decomposition of A.

Symmetric eigenvalue problems are especially favorable for numerical computation. Modern algorithms first use orthogonal similarity transformations to reduce a real symmetric matrix A to a symmetric tridiagonal matrix. An iterative eigensolver, often based on the QR algorithm, is then applied to compute the eigenvalues.

Because orthogonal transformations preserve lengths and angles, they are numerically stable and help prevent roundoff errors from being amplified. If the transformations are accumulated, they produce an orthogonal matrix whose columns are eigenvectors of A. Thus, the computed factorization has the form

$$A \approx PDP^T,$$

where P is orthogonal and D is diagonal.

A matrix is called **dense** when most of its entries are nonzero, while a matrix is called **sparse** when most of its entries are zero. The distinction is computational rather than purely mathematical; sparse matrices can often be stored and manipulated using only their nonzero entries.

For a dense $n \times n$ matrix, storing all entries requires n^2 values. A sparse matrix may require storage proportional only to the number of nonzero entries, often denoted by $\operatorname{nnz}(A)$. This can reduce both memory usage and computational cost dramatically for very large systems.

Dense eigenvalue problems are commonly handled using routines from the LAPACK numerical library. For large sparse symmetric matrices, methods such as the Lanczos algorithm are often used to compute only selected eigenvalues and eigenvectors. Implementations of Lanczos-type methods are available through the ARPACK numerical package, which is accessed by many popular

programming languages and scientific-computing environments. For example, ANSYS Mechanical provides Block Lanczos methods for extracting natural frequencies and vibration modes, and related eigenvalue methods are used in linear buckling analysis. These approaches are particularly effective for the large sparse matrices produced by finite-element models.

Nonsymmetric matrices are more difficult. They do not need to possess an orthonormal basis of eigenvectors and may even be defective. Numerical algorithms therefore commonly compute a Schur decomposition first, from which eigenvalues and eigenvectors may be obtained.

Problem 7.1.1

Show that if A is orthogonally diagonalizable, then A^2 is also orthogonally diagonalizable.

Solution

Since A is orthogonally diagonalizable, there exist an orthogonal matrix P and a diagonal matrix D such that

$$A = PDP^T.$$

Therefore,

$$\begin{aligned} A^2 &= \left(PDP^T\right)\left(PDP^T\right) \\ &= PD\left(P^TP\right)DP^T \\ &= PD^2P^T, \end{aligned}$$

because $P^TP = I$.

Since D is diagonal, D^2 is also diagonal. Thus, A^2 has an orthogonal diagonalization

$$A^2 = PD^2P^T.$$

Therefore, A^2 is orthogonally diagonalizable. Its eigenvectors are the same as those of A, while its eigenvalues are the squares of the corresponding eigenvalues of A.

Problem 7.1.2

Let $A = PDP^{-1}$, where P is orthogonal and D is diagonal. Suppose λ is an eigenvalue of A with algebraic multiplicity k, so λ appears k times on the diagonal of D. Explain why the eigenspace corresponding to λ has dimension k.

Solution

The columns of P form a linearly independent basis of eigenvectors of A, and each column corresponds to the eigenvalue appearing in the same position on the diagonal of D.

Since λ appears exactly k times on the diagonal of D, exactly k columns of P are eigenvectors corresponding to λ. These eigenvectors are linearly independent because P is invertible.

Moreover, because the columns of P form a basis of $\mathbb{R}^n$, every eigenvector corresponding to λ lies in the span of those k columns. Therefore, those columns form a basis for the eigenspace associated with λ. Thus, $\dim(E_\lambda) = k$.

Problem 7.1.3

Orthogonally diagonalize

$$A = \begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix}.$$

Then use the orthogonal diagonalization to construct a spectral decomposition of A.

Solution

The characteristic polynomial is

$$\begin{aligned}\det(A-\lambda I) &= \begin{vmatrix} 6-\lambda & 2 \\ 2 & 3-\lambda \end{vmatrix} \\ &= (6-\lambda)(3-\lambda)-4 \\ &= \lambda^2 - 9\lambda + 14 \\ &= (\lambda-7)(\lambda-2).\end{aligned}$$

Thus, the eigenvalues are

$$\lambda_1 = 7, \quad \lambda_2 = 2.$$

For $\lambda_1 = 7$,

$$A - 7I = \begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix},$$

so an eigenvector is

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

For $\lambda_2 = 2$,

$$A - 2I = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix},$$

so an eigenvector is

$$\mathbf{v}_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

The eigenvectors are orthogonal because

$$\mathbf{v}_1^T\mathbf{v}_2 = 2(1) + 1(-2) = 0.$$

Since

$$\|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \sqrt{5},$$

the corresponding unit eigenvectors are

$$\mathbf{u}_1 = \frac{1}{\sqrt{5}}\begin{bmatrix}2\\1\end{bmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{5}}\begin{bmatrix}1\\-2\end{bmatrix}.$$

Let

$$P = \begin{bmatrix}\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}}\\ \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}}\end{bmatrix}, \quad D = \begin{bmatrix}7 & 0\\0 & 2\end{bmatrix}.$$

Because the columns of P are orthonormal, $P^{-1} = P^T$. Therefore,

$$A = PDP^T.$$

The spectral decomposition is

$$A = 7\mathbf{u}_1\mathbf{u}_1^T + 2\mathbf{u}_2\mathbf{u}_2^T.$$

The rank-one projection matrices are

$$\mathbf{u}_1\mathbf{u}_1^T = \begin{bmatrix}\frac{4}{5} & \frac{2}{5}\\ \frac{2}{5} & \frac{1}{5}\end{bmatrix}, \quad \mathbf{u}_2\mathbf{u}_2^T = \begin{bmatrix}\frac{1}{5} & -\frac{2}{5}\\ -\frac{2}{5} & \frac{4}{5}\end{bmatrix}.$$

Thus,

$$
\begin{aligned}
A &= 7\begin{bmatrix} \frac{4}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{bmatrix} + 2\begin{bmatrix} \frac{1}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{4}{5} \end{bmatrix} \\
&= \begin{bmatrix} \frac{28}{5} & \frac{14}{5} \\ \frac{14}{5} & \frac{7}{5} \end{bmatrix} + \begin{bmatrix} \frac{2}{5} & -\frac{4}{5} \\ -\frac{4}{5} & \frac{8}{5} \end{bmatrix} \\
&= \begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix}.
\end{aligned}
$$

In MATLAB, the function `eig(A)` returns the eigenvectors and eigenvalues of A. The columns of P are eigenvectors, while the diagonal entries of D are the corresponding eigenvalues, so $AP = PD$. If A is real and symmetric, MATLAB returns an orthonormal set of eigenvectors, giving the orthogonal diagonalization $A = PDP^T$. For a nonsymmetric matrix, P is generally not orthogonal.

The function `diag` does not diagonalize a matrix; it only extracts diagonal entries or constructs a diagonal matrix.

7.2 Quadratic and Hermitian Forms

Most expressions considered so far have been linear in their variables. Quadratic expressions also arise throughout science and engineering, including in kinetic and potential energy, rotational kinematics, process optimization, and orbital mechanics.

A quadratic form packages such an expression into matrix notation using a symmetric matrix.

Quadratic Form

A **quadratic form** on $\mathbb{R}^n$ is a function $Q : \mathbb{R}^n \to \mathbb{R}$ of the form

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x},$$

where A is an $n \times n$ symmetric matrix. The matrix A is called the **matrix of the quadratic form**.

The simplest nonzero quadratic form is

$$Q(\mathbf{x}) = \mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|^2,$$

which represents the squared Euclidean length of $\mathbf{x}$.

For a vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

a general symmetric matrix has the form

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}.$$

Then

$$\begin{aligned} \mathbf{x}^T A\mathbf{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= ax_1^2 + 2bx_1x_2 + cx_2^2. \end{aligned}$$

The off-diagonal entry b appears twice, which is what produces the cross term $2bx_1x_2$.

Example 7.2.1

Let

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Compute $\mathbf{x}^T A\mathbf{x}$ for each matrix.

$$A_1 = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix}.$$

Solution

For A_1,

$$\begin{aligned}\mathbf{x}^T A_1 \mathbf{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 4x_1 \\ 3x_2 \end{bmatrix} \\ &= 4x_1^2 + 3x_2^2.\end{aligned}$$

Because A_1 is diagonal, the quadratic form contains no cross term.

For A_2,

$$\begin{aligned}\mathbf{x}^T A_2 \mathbf{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 3x_1 - 2x_2 \\ -2x_1 + 7x_2 \end{bmatrix} \\ &= x_1(3x_1 - 2x_2) + x_2(-2x_1 + 7x_2) \\ &= 3x_1^2 - 4x_1x_2 + 7x_2^2.\end{aligned}$$

There are two off-diagonal -2s which combine to produce the cross term $-4x_1x_2$.

Example 7.2.2

For $\mathbf{x} \in \mathbb{R}^3$, let

$$Q(\mathbf{x}) = 5x_1^2 + 3x_2^2 + 2x_3^2 - x_1x_2 + 8x_2x_3.$$

Write $Q(\mathbf{x})$ in the form

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x},$$

where A is symmetric.

Solution

The coefficients of the squared terms become the diagonal entries of A:

$$a_{1,1} = 5, \quad a_{2,2} = 3, \quad a_{3,3} = 2.$$

Because A must be symmetric, each cross-term coefficient is divided equally between two corresponding off-diagonal entries.

The term $-x_1x_2$ requires

$$a_{1,2} = a_{2,1} = -\frac{1}{2},$$

while the term $8x_2x_3$

requires

$$a_{2,3} = a_{3,2} = 4.$$

Because there is no x_1x_3 term,

$$a_{1,3} = a_{3,1} = 0.$$

Therefore,

$$A = \begin{bmatrix} 5 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 3 & 4 \\ 0 & 4 & 2 \end{bmatrix}.$$

Hence,

$$Q(\mathbf{x}) = \mathbf{x}^T \begin{bmatrix} 5 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 3 & 4 \\ 0 & 4 & 2 \end{bmatrix} \mathbf{x}.$$

Quadratic forms are often easier to interpret when they contain no cross-product terms. This occurs when the matrix of the quadratic form is diagonal. Because the matrix of a quadratic form is symmetric, the spectral theorem for real symmetric matrices guarantees that an orthogonal change of variables can always eliminate these cross terms.

Let $\mathbf{x} \in \mathbb{R}^n$ be a variable vector. A **change of variables** has the form

$$\mathbf{x} = P\mathbf{y},$$

where P is an invertible matrix and $\mathbf{y} \in \mathbb{R}^n$ is the new variable vector. Equivalently,

$$\mathbf{y} = P^{-1}\mathbf{x}.$$

The entries of $\mathbf{y}$ are the coordinates of $\mathbf{x}$ relative to the basis formed by the columns of P.

Consider the quadratic form

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}.$$

Substituting $\mathbf{x} = P\mathbf{y}$ gives

$$\begin{aligned} Q(\mathbf{x}) &= (P\mathbf{y})^T A(P\mathbf{y}) \\ &= \mathbf{y}^T P^T AP\mathbf{y}. \end{aligned}$$

Therefore, after the change of variables, the matrix of the quadratic form becomes

$$P^T AP.$$

Since A is symmetric, the spectral theorem for real symmetric matrices guarantees that an orthogonal matrix P can be chosen so that

$$P^T AP = D,$$

where D is diagonal. Equivalently,

$$A = PDP^T.$$

The quadratic form then becomes

$$Q(\mathbf{x}) = \mathbf{y}^T D\mathbf{y}.$$

If

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix},$$

then

$$Q(\mathbf{x}) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2.$$

Thus, the orthogonal change of variables replaces the original coordinate axes with the orthonormal eigendirections of A, eliminating every cross-product term.

Example 7.2.3

Make a change of variables that transforms the quadratic form

$$Q(\mathbf{x}) = x_1^2 - 8x_1x_2 - 5x_2^2$$

into a quadratic form with no cross-product term. Then, evaluate the quadratic form at

$$\mathbf{x} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}.$$

Solution

The symmetric matrix associated with the quadratic form is

$$A = \begin{bmatrix} 1 & -4 \\ -4 & -5 \end{bmatrix},$$

since $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$.

To eliminate the cross-product term, orthogonally diagonalize A. Its characteristic polynomial is

$$\begin{aligned}\det(A-\lambda I) &= \begin{vmatrix} 1-\lambda & -4 \\ -4 & -5-\lambda \end{vmatrix} \\ &= (1-\lambda)(-5-\lambda) - 16 \\ &= \lambda^2 + 4\lambda - 21 \\ &= (\lambda-3)(\lambda+7).\end{aligned}$$

Thus, the eigenvalues are

$$\lambda_1 = 3, \quad \lambda_2 = -7.$$

Corresponding unit eigenvectors are

$$\mathbf{u}_1 = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}.$$

Because A is symmetric and the eigenvalues are distinct, these eigenvectors are orthogonal. Therefore, let

$$P = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}, \quad D = \begin{bmatrix} 3 & 0 \\ 0 & -7 \end{bmatrix}.$$

Then $A = PDP^T$ and $P^T AP = D$.

Use the orthogonal change of variables $\mathbf{x} = P\mathbf{y}$:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

Substituting $\mathbf{x} = P\mathbf{y}$ into the quadratic form gives

$$\begin{aligned} Q(\mathbf{x}) &= \mathbf{x}^T A\mathbf{x} \\ &= (P\mathbf{y})^T A(P\mathbf{y}) \\ &= \mathbf{y}^T P^T AP\mathbf{y} \\ &= \mathbf{y}^T D\mathbf{y} \\ &= 3y_1^2 - 7y_2^2. \end{aligned}$$

Therefore, the change of variables $\mathbf{x} = P\mathbf{y}$ transforms the quadratic form into

$$Q(\mathbf{x}) = 3y_1^2 - 7y_2^2,$$

which contains no cross-product term.

Since $\mathbf{x} = P\mathbf{y}$, we have

$$\mathbf{y} = P^{-1}\mathbf{x} = P^T\mathbf{x}.$$

Therefore,

$$\begin{aligned} \mathbf{y} &= \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 2 \\ -2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{6}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{bmatrix}. \end{aligned}$$

Plugging into the diagonal form, we obtain

$$\begin{aligned} Q(\mathbf{x}) &= 3\left(\frac{6}{\sqrt{5}}\right)^2 - 7\left(-\frac{2}{\sqrt{5}}\right)^2 \\ &= 3\left(\frac{36}{5}\right) - 7\left(\frac{4}{5}\right) \\ &= \frac{108}{5} - \frac{28}{5} \\ &= 16. \end{aligned}$$

See Figure 38.

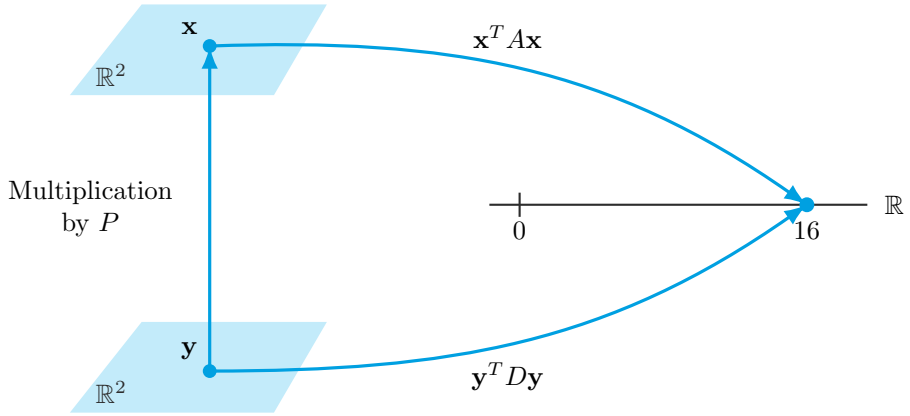


Figure 38: The orthogonal change of variables $\mathbf{x} = P\mathbf{y}$ in $\mathbf{x}^T A\mathbf{x}$ preserves the value of the quadratic form.

Example 7.2.3 illustrates a general consequence of the spectral theorem for real symmetric matrices. Because the matrix of a quadratic form is symmetric, it can be orthogonally diagonalized. The corresponding orthogonal change of variables eliminates every cross-product term.

Principal Axes Theorem

Let A be an $n \times n$ symmetric matrix. Then there exists an orthogonal matrix P such that the change of variables $\mathbf{x} = P\mathbf{y}$ transforms the quadratic form $\mathbf{x}^T A\mathbf{x}$ into the diagonal quadratic form

$$\mathbf{y}^T D\mathbf{y},$$

where

$$D = P^T AP$$

is diagonal. Therefore, the transformed quadratic form contains no cross-product terms.

The columns of P are called the **principal axes** of the quadratic form. They form an orthonormal basis of eigenvectors of A, while $\mathbf{y}$ is the coordinate vector of $\mathbf{x}$ relative to this basis.

Thus, the principal axes theorem replaces the original coordinate directions with mutually orthogonal eigenvector directions in which the quadratic form becomes

$$\mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2.$$

Let

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x},$$

where A is an invertible 2×2 symmetric matrix. For a constant c, the level set

$$\mathbf{x}^T A\mathbf{x} = c$$

describes a curve in $\mathbb{R}^2$. Depending on the eigenvalues of A and the value of c, this curve may be an ellipse, a hyperbola, a pair of intersecting lines, a single point, or the empty set.

When A is diagonal, the quadratic form contains no cross-product term, so the graph is already in standard position. For example,

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$$

describes an ellipse, while

$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1$$

describes a hyperbola.

When A is not diagonal, the graph is generally rotated relative to the original coordinate axes. Because A is symmetric, the principal axes theorem guarantees an orthogonal matrix P such that

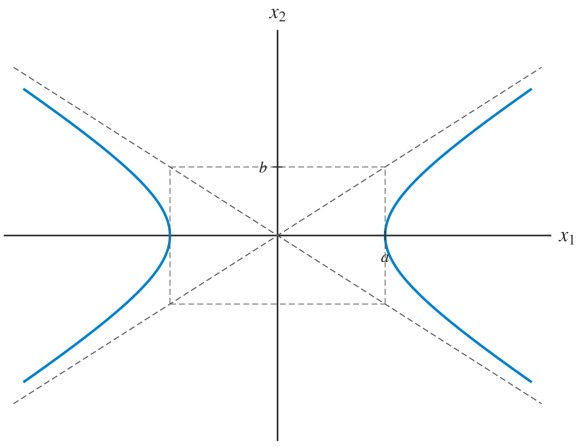
$$A = PDP^T.$$

Using the change of variables $\mathbf{x} = P\mathbf{y}$, the level-set equation becomes

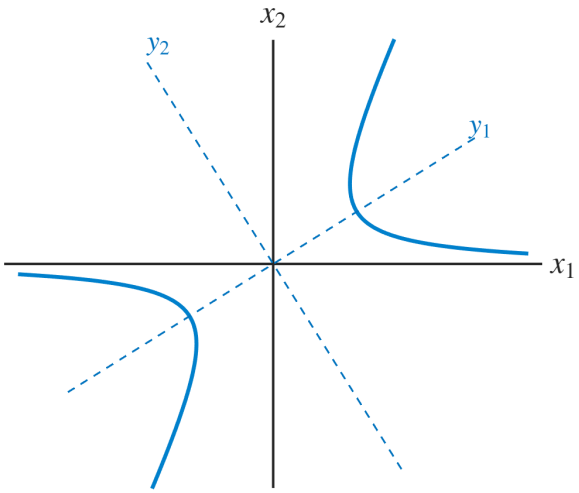
$$\mathbf{y}^T D\mathbf{y} = c.$$

Thus, the eigenvectors of A, which form the columns of P, determine the principal axes of the curve. Relative to these new coordinate axes, the graph is placed in standard position and the cross-product terms disappear.

In two dimensions, the signs of the eigenvalues reveal the basic geometry. If both eigenvalues have the same sign, the level sets are elliptic curves when they are nonempty. If the eigenvalues have opposite signs, the level sets are hyperbolae. See Figure 39.



(a) The hyperbola $\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1$ in standard position.



(b) A hyperbola rotated out of standard position relative to the original x_1- and x_2-axes. The dashed blue lines indicate the principal axes y_1 and y_2, along which the equation takes the standard form $\frac{y_1^2}{a^2} - \frac{y_2^2}{b^2} = 1$.

Figure 39: The same hyperbola in standard position and rotated relative to the original x_1- and x_2-axes

A quadratic form

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$$

is a real-valued function on $\mathbb{R}^n$. When $n = 2$, its graph is the surface

$$x_3 = Q(x_1, x_2).$$

The sign of $Q(\mathbf{x})$ determines the general shape of this surface. Some quadratic forms are positive for every nonzero vector, some are always negative, and others assume both positive and negative values.

Classification of Quadratic Forms

Let $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ be a quadratic form on $\mathbb{R}^n$. Then Q is:

1. **positive definite** if $Q(\mathbf{x}) > 0$ for every $\mathbf{x} \neq \mathbf{0}$;
2. **negative definite** if $Q(\mathbf{x}) < 0$ for every $\mathbf{x} \neq \mathbf{0}$;
3. **indefinite** if $Q(\mathbf{x})$ assumes both positive and negative values;
4. **positive semidefinite** if $Q(\mathbf{x}) \geq 0$ for every $\mathbf{x} \in \mathbb{R}^n$;
5. **negative semidefinite** if $Q(\mathbf{x}) \leq 0$ for every $\mathbf{x} \in \mathbb{R}^n$.

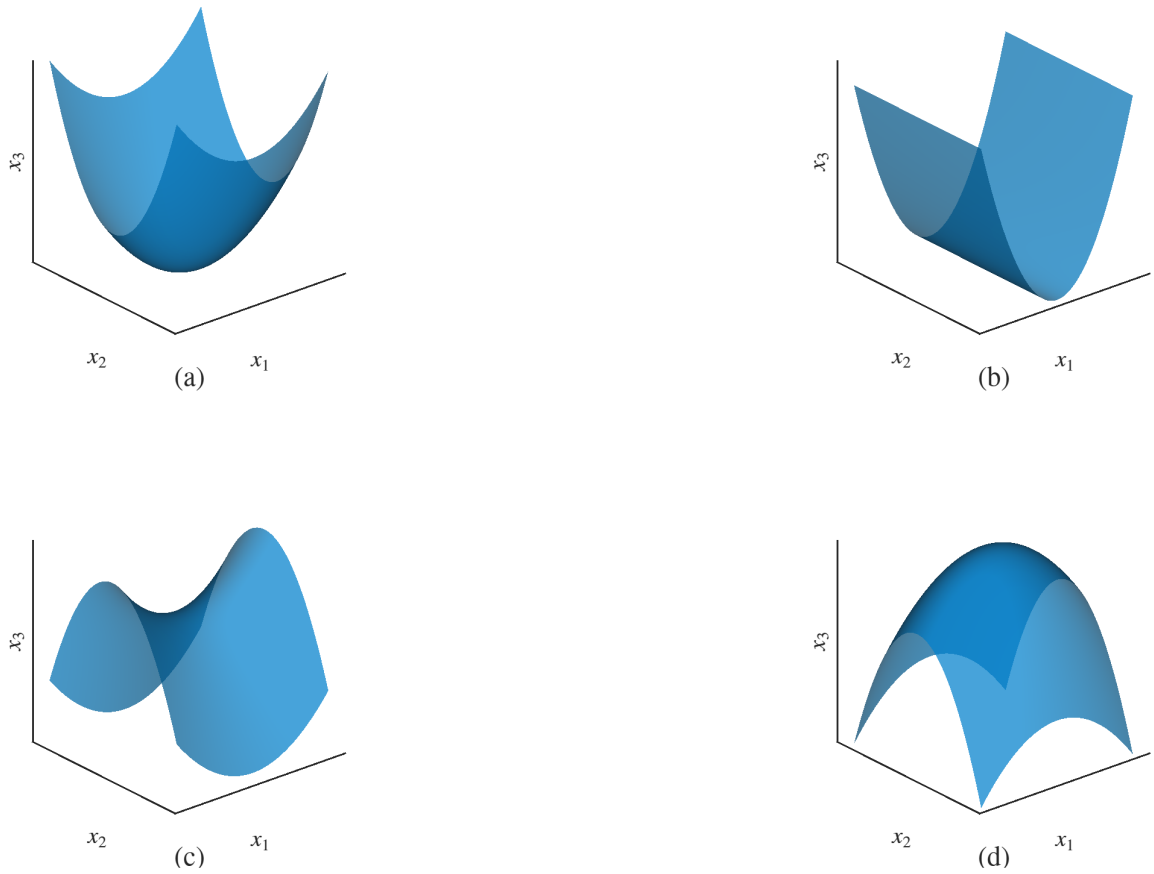


Figure 40: Graphs of the quadratic forms (a) $x_3 = 3x_1^2+7x_2^2$, which is positive definite; (b) $x_3 = 3x_1^2$, which is positive semidefinite; (c) $x_3 = 3x_1^2 - 7x_2^2$, which is indefinite; and (d) $x_3 = -3x_1^2 - 7x_2^2$, which is negative definite.

A positive definite quadratic form produces an upward-opening bowl-shaped surface, while a negative definite form produces a downward-opening bowl. An indefinite form produces a saddle-shaped surface because the quadratic form increases in some directions and decreases in others.

A semidefinite quadratic form may equal zero for some nonzero vectors. For example,

$$Q(\mathbf{x}) = 3x_1^2$$

is positive semidefinite because it is never negative. However, it is not positive definite because

$$Q\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = 0.$$

The eigenvalues of the symmetric matrix A provide a direct method for classifying the quadratic form.

Quadratic Forms and Eigenvalues

Let A be an $n \times n$ symmetric matrix. Then the quadratic form

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$$

is:

1. positive definite if and only if every eigenvalue of A is positive;
2. negative definite if and only if every eigenvalue of A is negative;
3. indefinite if and only if A has both positive and negative eigenvalues;
4. positive semidefinite if and only if every eigenvalue of A is nonnegative;
5. negative semidefinite if and only if every eigenvalue of A is nonpositive.

To understand this result, let $A = PDP^T$ be an orthogonal diagonalization of A, and use the change of variables $\mathbf{x} = P\mathbf{y}$.

Then

$$\begin{aligned} Q(\mathbf{x}) &= \mathbf{x}^T A\mathbf{x} \\ &= \mathbf{y}^T D\mathbf{y} \\ &= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2. \end{aligned}$$

Because each $y_i^2 \geq 0$, the signs of the eigenvalues determine the possible signs of $Q(\mathbf{x})$. If all eigenvalues are positive, every nonzero vector produces a positive value. If all are negative, every nonzero vector produces a negative value. If eigenvalues of both signs occur, the quadratic form assumes both positive and negative values.

Example 7.2.4

Determine whether the quadratic form

$$Q(\mathbf{x}) = 3x_1^2 + 2x_2^2 + x_3^2 + 4x_1x_2 + 4x_2x_3$$

is positive definite.

Solution

Although every term appears with a positive coefficient, the signs of the individual terms alone do not determine whether the quadratic form is positive definite.

The symmetric matrix associated with Q is

$$A = \begin{bmatrix} 3 & 2 & 0 \\ 2 & 2 & 2 \\ 0 & 2 & 1 \end{bmatrix},$$

because the coefficients of the cross-product terms are divided equally between the corresponding off-diagonal entries.

The eigenvalues of A are

$$\lambda_1 = 5, \quad \lambda_2 = 2, \quad \lambda_3 = -1.$$

Since A has both positive and negative eigenvalues, the quadratic form is indefinite. Therefore, Q is not positive definite.

Example 7.2.5

For each symmetric matrix A, write the quadratic form

$$f(x,y) = \begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix} = ax^2 + 2bxy + cy^2.$$

Then complete the square to express f as a sum of one or two squared terms.

$$A_1 = \begin{bmatrix} 1 & 2 \\ 2 & 9 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix}.$$

Solution

For A_1, the quadratic form is

$$\begin{aligned} f_1(x,y) &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= x^2 + 4xy + 9y^2. \end{aligned}$$

Complete the square in x:

$$\begin{aligned} f_1(x,y) &= x^2 + 4xy + 4y^2 + 5y^2 \\ &= (x+2y)^2 + 5y^2 \end{aligned}$$

Thus, f_1 is a sum of two positive squared terms and is therefore positive definite.

For A_2,

$$\begin{aligned} f_2(x,y) &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= x^2 + 6xy + 9y^2 \\ &= (x+3y)^2. \end{aligned}$$

Thus, f_2 is nonnegative for every (x, y), but it equals zero whenever $x = -3y$. Therefore, f_2 is positive semidefinite rather than positive definite.

Completing the square can simplify a quadratic form and help determine its definiteness, but it does not generally produce an orthogonal change of variables. The resulting squared expressions may correspond to directions that are not mutually perpendicular. This procedure is not the same as the orthogonal principal-axes transformation.

Problem 7.2.1

Let $Q(\mathbf{x}) = 5x_1^2 + 4x_1x_2 + 2x_2^2$.

(a) Write $Q(\mathbf{x})$ as $\mathbf{x}^T A\mathbf{x}$.

(b) Orthogonally diagonalize A.

(c) Find an orthogonal change of variables $\mathbf{x} = P\mathbf{y}$ that eliminates the cross-product term.

(d) Write the transformed quadratic form and classify Q.

Solution

(a) The quadratic form is

$$Q(\mathbf{x}) = 5x_1^2 + 4x_1x_2 + 2x_2^2.$$

Because the coefficient of x_1x_2 is 4, the corresponding off-diagonal entries are both 2. Therefore,

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x},$$

where

$$A = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}.$$

(b) To orthogonally diagonalize A, compute its characteristic polynomial:

$$\begin{aligned}\det(A-\lambda I) &= \begin{vmatrix} 5-\lambda & 2 \\ 2 & 2-\lambda \end{vmatrix} \\ &= (5-\lambda)(2-\lambda)-4 \\ &= \lambda^2 - 7\lambda + 6 \\ &= (\lambda-6)(\lambda-1)\end{aligned}$$

Thus, the eigenvalues are

$$\lambda_1 = 6, \quad \lambda_2 = 1.$$

For $\lambda_1 = 6$,

$$A - 6I = \begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix},$$

so one eigenvector is

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

For $\lambda_2 = 1$,

$$A - I = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix},$$

so one eigenvector is

$$\mathbf{v}_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

These eigenvectors are orthogonal because

$$\mathbf{v}_1^T \mathbf{v}_2 = 2(1) + 1(-2) = 0.$$

Since

$$\|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \sqrt{5},$$

the corresponding unit eigenvectors are

$$\mathbf{u}_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

Therefore, $A = PDP^T$ with

$$P = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{bmatrix}, \quad D = \begin{bmatrix} 6 & 0 \\ 0 & 1 \end{bmatrix}.$$

(c) Use the orthogonal change of variables $\mathbf{x} = P\mathbf{y}$:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

(d) Then

$$\begin{aligned} Q(\mathbf{x}) &= \mathbf{x}^T A \mathbf{x} \\ &= (P\mathbf{y})^T A (P\mathbf{y}) \\ &= \mathbf{y}^T P^T A P \mathbf{y} \\ &= \mathbf{y}^T D \mathbf{y} \\ &= 6y_1^2 + y_2^2. \end{aligned}$$

Thus, the transformed quadratic form contains no cross-product term. Since both eigenvalues are positive, Q is positive definite.

Problem 7.2.2

Classify the quadratic form

$$Q(\mathbf{x}) = 4x_1^2 - 4x_1x_2 + x_2^2.$$

Find all nonzero vectors $\mathbf{x}$ for which $Q(\mathbf{x}) = 0$, and explain why the form is not positive definite.

Solution

The quadratic form is

$$Q(\mathbf{x}) = 4x_1^2 - 4x_1x_2 + x_2^2.$$

Completing the square gives

$$Q(\mathbf{x}) = (2x_1 - x_2)^2.$$

Because a square is always nonnegative,

$$Q(\mathbf{x}) \geq 0$$

for every $\mathbf{x} \in \mathbb{R}^2$. Therefore, Q is either positive definite or positive semidefinite.

The quadratic form equals zero when $2x_1 - x_2 = 0$ or equivalently, $x_2 = 2x_1$.

Thus, every nonzero vector of the form

$$\mathbf{x} = t \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad t \neq 0,$$

satisfies $Q(\mathbf{x}) = 0$.

Since Q vanishes for nonzero vectors, it is not positive definite. Therefore, Q is positive semidefinite.

The associated symmetric matrix is

$$A = \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix}.$$

Its characteristic polynomial is

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} 4-\lambda & -2 \\ -2 & 1-\lambda \end{vmatrix} \\ &= (4-\lambda)(1-\lambda) - 4 \\ &= \lambda^2 - 5\lambda \\ &= \lambda(\lambda - 5). \end{aligned}$$

Thus, the eigenvalues are

$$\lambda_1 = 5, \quad \lambda_2 = 0.$$

Because both eigenvalues are nonnegative and one is zero, this confirms that Q is positive semidefinite.

Although a real quadratic form could initially be written as

$$Q(\mathbf{x}) = \mathbf{x}^T M \mathbf{x},$$

where M is any real square matrix, allowing M to be nonsymmetric does not produce any additional quadratic forms. Decompose M as

$$M = \frac{1}{2}\left(M + M^T\right) + \frac{1}{2}\left(M - M^T\right) = A + B,$$

where $A^T = A$ and $B^T = -B$. Since $\mathbf{x}^T B \mathbf{x}$ is a scalar,

$$\begin{aligned} \mathbf{x}^T B \mathbf{x} &= \left(\mathbf{x}^T B \mathbf{x}\right)^T \\ &= \mathbf{x}^T B^T \mathbf{x} \\ &= -\mathbf{x}^T B \mathbf{x}. \end{aligned}$$

Therefore,

$$\mathbf{x}^T B\mathbf{x} = 0,$$

and hence

$$\mathbf{x}^T M\mathbf{x} = \mathbf{x}^T A\mathbf{x}.$$

Thus, every real quadratic form may be represented using a symmetric matrix.

When vectors and matrices contain complex entries, symmetry must be replaced by a closely related condition involving complex conjugation.

Conjugate Transpose

The **complex conjugate** of a matrix A, denoted by $\overline{A}$, is obtained by taking the complex conjugate of every entry of A.

The **conjugate transpose**, or **adjoint**, of A is

$$A^* = \overline{A}^T = \overline{A^T}.$$

Thus, the entries of A are first conjugated and then transposed. Some texts denote the conjugate transpose by A^H or $A^\dagger$.

For a real matrix, $\overline{A} = A$, so

$$A^* = A^T.$$

The conjugate transpose reverses the order of multiplication:

$$(AB)^* = B^* A^*$$

If A is invertible, then

$$\left(A^{-1}\right)^* = \left(A^*\right)^{-1}.$$

Example 7.2.6

Find the conjugate transpose of

$$A = \begin{bmatrix} 1 & 2 & 3i \\ 1+i & 1 & 0 \end{bmatrix}.$$

Solution

First, take the complex conjugate of every entry:

$$\overline{A} = \begin{bmatrix} 1 & 2 & -3i \\ 1-i & 1 & 0 \end{bmatrix}.$$

Transposing gives

$$A^* = \begin{bmatrix} 1 & 1-i \\ 2 & 1 \\ -3i & 0 \end{bmatrix}.$$

Hermitian Matrix

A complex square matrix A is called **Hermitian** when

$$A^* = A.$$

Equivalently, its entries satisfy

$$a_{j,i} = \overline{a_{i,j}}.$$

Therefore, entries on opposite sides of the main diagonal are complex conjugates of one another, and every diagonal entry must be real.

For example,

$$A = \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix}$$

is Hermitian because

$$A^* = \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix} = A.$$

A real matrix is Hermitian if and only if it is symmetric. Hermitian matrices are therefore the natural complex analogue of real symmetric matrices.

Hermitian Form

Let A be an $n \times n$ Hermitian matrix. The function

$$H(\mathbf{z}) = \mathbf{z}^* A \mathbf{z},$$

for $\mathbf{z} \in \mathbb{C}^n$ is called a **Hermitian form**.

The conjugate transpose is essential. The expression $\mathbf{z}^T A \mathbf{z}$ may be complex, but a Hermitian form is always real. Indeed,

$$\begin{aligned} \overline{H(\mathbf{z})} &= (\mathbf{z}^* A \mathbf{z})^* \\ &= \mathbf{z}^* A^* \mathbf{z} \\ &= \mathbf{z}^* A \mathbf{z} \\ &= H(\mathbf{z}). \end{aligned}$$

Since $H(\mathbf{z})$ equals its own complex conjugate, it must be real.

Example 7.2.7

Let

$$A = \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} 1+i \\ 2-i \end{bmatrix}.$$

Evaluate the Hermitian form $H(\mathbf{z}) = \mathbf{z}^* A\mathbf{z}$.

Solution

The conjugate transpose of $\mathbf{z}$ is

$$\mathbf{z}^* = \begin{bmatrix} 1-i & 2+i \end{bmatrix}.$$

First compute

$$A\mathbf{z} = \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix} \begin{bmatrix} 1+i \\ 2-i \end{bmatrix} = \begin{bmatrix} 6+4i \\ 6-2i \end{bmatrix}.$$

Therefore,

$$\begin{aligned} H(\mathbf{z}) &= \begin{bmatrix} 1-i & 2+i \end{bmatrix} \begin{bmatrix} 6+4i \\ 6-2i \end{bmatrix} \\ &= (1-i)(6+4i) + (2+i)(6-2i) \\ &= 24. \end{aligned}$$

As expected, the value of the Hermitian form is real.

Spectral Theorem for Hermitian Matrices

Let A be an $n \times n$ Hermitian matrix. Then:

(i) Every eigenvalue of A is real.

(ii) Eigenspaces corresponding to distinct eigenvalues are orthogonal with respect to the complex inner product.

(iii) There exists an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of A.

(iv) There exist a unitary matrix U and a real diagonal matrix D such that $A = UDU^*$.

A matrix U is **unitary** when $U^*U = UU^* = I$.

Unitary matrices are the complex analogue of orthogonal matrices. If $A = UDU^*$ and $\mathbf{z} = U\mathbf{y}$, then

$$\begin{aligned} H(\mathbf{z}) &= \mathbf{z}^* A \mathbf{z} \\ &= \mathbf{y}^* D \mathbf{y} \\ &= \lambda_1 |y_1|^2 + \lambda_2 |y_2|^2 + \cdots + \lambda_n |y_n|^2. \end{aligned}$$

Because each $|y_i|^2 \geq 0$, the eigenvalues classify Hermitian forms exactly as they classify real quadratic forms.

Hermitian Forms and Eigenvalues

Let A be Hermitian. Then the Hermitian form $H(\mathbf{z}) = \mathbf{z}^* A \mathbf{z}$ is

(a) positive definite if and only if every eigenvalue of A is positive;

(b) negative definite if and only if every eigenvalue of A is negative;

(c) indefinite if and only if A has both positive and negative eigenvalues;

(d) positive semidefinite if and only if every eigenvalue of A is nonnegative;

(e) negative semidefinite if and only if every eigenvalue of A is nonpositive.

Hermitian matrices arise in physics and electrical engineering. In quantum mechanics, observable quantities are represented by Hermitian operators because their eigenvalues and expectation values are real. In signal processing, covariance and spectral-density matrices are Hermitian, while Hermitian matrix formulations also arise in electrical power-system analysis and optimization.

Problem 7.2.3

Let

$$A = \begin{bmatrix} 2 & i \\ -i & 2 \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} 1+i \\ 2-i \end{bmatrix}.$$

(a) Verify that A is Hermitian.

(b) Evaluate the Hermitian form $H(\mathbf{z}) = \mathbf{z}^* A \mathbf{z}$.

(c) Find the eigenvalues of A and classify the Hermitian form.

Solution

(a) The conjugate transpose of A is

$$\begin{aligned} A^* &= \overline{A}^T \\ &= \begin{bmatrix} 2 & -i \\ i & 2 \end{bmatrix}^T \\ &= \begin{bmatrix} 2 & i \\ -i & 2 \end{bmatrix} \\ &= A. \end{aligned}$$

Therefore, A is Hermitian.

(b) The conjugate transpose of $\mathbf{z}$ is

$$\mathbf{z}^* = \begin{bmatrix} 1-i & 2+i \end{bmatrix}.$$

First compute

$$A\mathbf{z} = \begin{bmatrix} 2 & i \\ -i & 2 \end{bmatrix} \begin{bmatrix} 1+i \\ 2-i \end{bmatrix} = \begin{bmatrix} 3+4i \\ 5-3i \end{bmatrix}.$$

Thus,

$$\begin{aligned} H(\mathbf{z}) &= \mathbf{z}^* A\mathbf{z} \\ &= \begin{bmatrix} 1-i & 2+i \end{bmatrix} \begin{bmatrix} 3+4i \\ 5-3i \end{bmatrix} \\ &= (1-i)(3+4i) + (2+i)(5-3i) \\ &= (7+i) + (13-i) \\ &= 20. \end{aligned}$$

As expected, the value of the Hermitian form is real.

(c) The characteristic polynomial of A is

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} 2-\lambda & i \\ -i & 2-\lambda \end{vmatrix} \\ &= (2-\lambda)^2 - i(-i) \\ &= (2-\lambda)^2 - 1 \\ &= (\lambda - 1)(\lambda - 3). \end{aligned}$$

Therefore, the eigenvalues are

$$\lambda_1 = 1, \quad \lambda_2 = 3.$$

Since both eigenvalues are positive, the Hermitian form

$$H(\mathbf{z}) = \mathbf{z}^* A\mathbf{z}$$

is positive definite.

7.3 Constrained Optimization

We will often need to maximize or minimize a quadratic form over a restricted set of vectors. A particularly important problem is to maximize or minimize

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$$

subject to

$$\|\mathbf{x}\| = 1.$$

The unit-vector constraint prevents the magnitude of $\mathbf{x}$ from growing without bound. It may be written equivalently as

$$\|\mathbf{x}\| = 1, \quad \|\mathbf{x}\|^2 = 1, \quad \mathbf{x}^T\mathbf{x} = 1,$$

or, in coordinates,

$$x_1^2 + x_2^2 + \cdots + x_n^2 = 1.$$

Thus, $\mathbf{x}$ is restricted to the unit sphere in $\mathbb{R}^n$. The goal is to determine the largest and smallest values attained by the quadratic form on this sphere.

Example 7.3.1

Find the maximum and minimum values of

$$Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$$

subject to

$$\mathbf{x}^T\mathbf{x} = x_1^2 + x_2^2 + x_3^2 = 1.$$

Solution

Since x_1^2, x_2^2, and x_3^2 are nonnegative,

$$\begin{aligned} Q(\mathbf{x}) &= 9x_1^2 + 4x_2^2 + 3x_3^2 \\ &\leq 9x_1^2 + 9x_2^2 + 9x_3^2 \\ &= 9\left(x_1^2 + x_2^2 + x_3^2\right) \\ &= 9. \end{aligned}$$

This upper bound is attained when

$$\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

for which $Q(\mathbf{x}) = 9$. Therefore, the maximum value is 9.

Similarly,

$$\begin{aligned} Q(\mathbf{x}) &= 9x_1^2 + 4x_2^2 + 3x_3^2 \\ &\geq 3x_1^2 + 3x_2^2 + 3x_3^2 \\ &= 3\left(x_1^2 + x_2^2 + x_3^2\right) \\ &= 3. \end{aligned}$$

This lower bound is attained when

$$\mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

for which $Q(\mathbf{x}) = 3$. Therefore,

$$\min_{\|\mathbf{x}\|=1} Q(\mathbf{x}) = 3, \quad \max_{\|\mathbf{x}\|=1} Q(\mathbf{x}) = 9.$$

These values are precisely the smallest and largest diagonal entries of the matrix

$$A = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix},$$

which are also its smallest and largest eigenvalues.

Example 7.3.2

Let

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 7 \end{bmatrix}, \quad Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}.$$

Find the maximum and minimum values of $Q(\mathbf{x})$ subject to $\|\mathbf{x}\| = 1$.

Solution

The quadratic form is

$$Q(\mathbf{x}) = 3x_1^2 + 7x_2^2,$$

and the constraint is

$$x_1^2 + x_2^2 = 1.$$

Every point on the unit circle may be written as

$$x_1 = \cos t, \quad x_2 = \sin t.$$

Therefore,

$$\begin{aligned} Q(\mathbf{x}) &= 3\cos^2 t + 7\sin^2 t \\ &= 3\left(\cos^2 t + \sin^2 t\right) + 4\sin^2 t \\ &= 3 + 4\sin^2 t. \end{aligned}$$

Because $0 \leq \sin^2 t \leq 1$,

$$3 \leq Q(\mathbf{x}) \leq 7.$$

The minimum value 3 occurs when $\sin t = 0$, corresponding to

$$\mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{or} \quad \mathbf{x} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

These are unit eigenvectors associated with the eigenvalue 3.

The maximum value 7 occurs when $\sin^2 t = 1$, corresponding to

$$\mathbf{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{or} \quad \mathbf{x} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

These are unit eigenvectors associated with the eigenvalue 7.

Geometrically, the surface $x_3 = 3x_1^2 + 7x_2^2$ intersects the vertical cylinder $x_1^2 + x_2^2 = 1$ in a closed curve. The lowest points on this curve have height 3, while the highest points have height 7. See Figure 41.

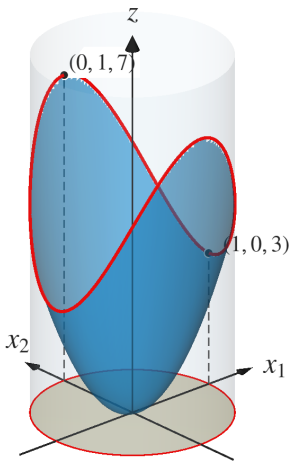


Figure 41: The quadratic surface $z = 3x_1^2 + 7x_2^2$, the unit cylinder $x_1^2 + x_2^2 = 1$, and their intersection curve

The same phenomenon occurs for every real symmetric matrix.

Extreme Values of a Quadratic Form

Let A be an $n \times n$ real symmetric matrix with smallest eigenvalue $\lambda_{\min}$ and largest eigenvalue $\lambda_{\max}$. Then

$$\min_{\|\mathbf{x}\|=1} \mathbf{x}^T A\mathbf{x} = \lambda_{\min},$$

and

$$\max_{\|\mathbf{x}\|=1} \mathbf{x}^T A\mathbf{x} = \lambda_{\max}.$$

The minimum is attained at any unit eigenvector corresponding to $\lambda_{\min}$, while the maximum is attained at any unit eigenvector corresponding to $\lambda_{\max}$.

Proof

Because A is symmetric, the spectral theorem for real symmetric matrices guarantees an orthogonal diagonalization $A = PDP^T$ where

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

and the columns of P are orthonormal eigenvectors of A. Order the eigenvalues so that

$$\lambda_{\min} = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n = \lambda_{\max}.$$

Use the orthogonal change of variables. Since P is orthogonal,

$$\|\mathbf{x}\| = \|P\mathbf{y}\| = \|\mathbf{y}\|.$$

Therefore, $\mathbf{x}$ is a unit vector if and only if $\mathbf{y}$ is a unit vector. Moreover,

$$\begin{aligned}\mathbf{x}^T A\mathbf{x} &= (P\mathbf{y})^T A(P\mathbf{y}) \\ &= \mathbf{y}^T P^T AP\mathbf{y} \\ &= \mathbf{y}^T D\mathbf{y} \\ &= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2.\end{aligned}$$

Because $\|\mathbf{y}\| = 1$,

$$y_1^2 + y_2^2 + \cdots + y_n^2 = 1.$$

Thus, $\mathbf{y}^T D\mathbf{y}$ is a weighted average of the eigenvalues, with nonnegative weights y_i^2 that sum to one. Consequently,

$$\lambda_{\min} \leq \mathbf{y}^T D\mathbf{y} \leq \lambda_{\max}.$$

The upper bound is attained when

$$\mathbf{y} = \mathbf{e}_n,$$

because

$$\mathbf{e}_n^T D\mathbf{e}_n = \lambda_{\max}.$$

The corresponding vector in the original coordinate system is

$$\mathbf{x} = P\mathbf{e}_n,$$

which is the column of P containing a unit eigenvector associated with $\lambda_{\max}$.

Similarly, the lower bound is attained when $\mathbf{y} = \mathbf{e}_1$, so the corresponding $\mathbf{x} = P\mathbf{e}_1$ is a unit eigenvector associated with $\lambda_{\min}$.

Therefore,

$$\min_{\|\mathbf{x}\|=1} \mathbf{x}^T A\mathbf{x} = \lambda_{\min}, \quad \max_{\|\mathbf{x}\|=1} \mathbf{x}^T A\mathbf{x} = \lambda_{\max}.$$

For any nonzero vector $\mathbf{x}$, the expression

$$R_A(\mathbf{x}) = \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

is called the **Rayleigh quotient** of A at $\mathbf{x}$. Dividing by $\mathbf{x}^T\mathbf{x}$ removes the effect of scaling, so the quotient depends only on the direction of $\mathbf{x}$.

Our knowledge about the extreme values of a quadratic form implies

$$\lambda_{\min} \leq R_A(\mathbf{x}) \leq \lambda_{\max}$$

for every nonzero $\mathbf{x}$. Equality occurs when $\mathbf{x}$ is an eigenvector corresponding to the smallest or largest eigenvalue.

This result can also be understood using Lagrange multipliers. To optimize $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ subject to $g(\mathbf{x}) = \mathbf{x}^T\mathbf{x} = 1$, the gradients must satisfy

$$\nabla Q = \lambda \nabla g.$$

Since A is symmetric,

$$\nabla Q = 2A\mathbf{x}, \quad \nabla g = 2\mathbf{x}.$$

Therefore,

$$2A\mathbf{x} = 2\lambda\mathbf{x},$$

or

$$A\mathbf{x} = \lambda\mathbf{x}.$$

Thus, the constrained stationary points occur precisely along eigenvector directions, and the corresponding values of the quadratic form are the eigenvalues. □

The preceding theorem identifies the smallest and largest eigenvalues of a symmetric matrix through constrained optimization. The same idea can be extended to recover the eigenvalues between these two extremes.

Example 7.3.3

Let

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

Find the maximum value of $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ subject to $\mathbf{x}^T\mathbf{x} = 1$, and find a unit vector at which the maximum is attained.

Solution

The maximum value of $\mathbf{x}^T A\mathbf{x}$ over all unit vectors is the largest eigenvalue of A.

The characteristic polynomial is

$$\begin{aligned} \det(A - \lambda I) &= -\lambda^3 + 10\lambda^2 - 27\lambda + 18 \\ &= -(\lambda - 6)(\lambda - 3)(\lambda - 1). \end{aligned}$$

Thus, the eigenvalues are

$$\lambda_1 = 6, \quad \lambda_2 = 3, \quad \lambda_3 = 1.$$

The largest eigenvalue is 6. Solving $(A - 6I)\mathbf{x} = \mathbf{0}$ gives an eigenvector

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Normalizing yields

$$\mathbf{u}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Therefore,

$$\max_{\|\mathbf{x}\|=1} \mathbf{x}^T A\mathbf{x} = 6,$$

and the maximum is attained at $\mathbf{x} = \pm\mathbf{u}_1$.

The largest eigenvalue is obtained by allowing $\mathbf{x}$ to point in any unit direction. To uncover the second-largest eigenvalue, we exclude the direction $\mathbf{u}_1$ by requiring $\mathbf{x}$ to be orthogonal to it.

Second-Largest Eigenvalue

Let A be a real symmetric matrix with eigenvalues ordered as

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n,$$

and let $\mathbf{u}_1$ be a unit eigenvector corresponding to λ_1. Then

$$\max\left\{\mathbf{x}^T A\mathbf{x} : \mathbf{x}^T\mathbf{x} = 1, \quad \mathbf{x}^T\mathbf{u}_1 = 0\right\} = \lambda_2.$$

The maximum is attained at any unit eigenvector corresponding to λ_2.

The additional condition

$$\mathbf{x}^T\mathbf{u}_1 = 0$$

removes every component of $\mathbf{x}$ in the direction of the dominant eigenvector. The optimization is therefore restricted to the orthogonal complement of $\mathbf{u}_1$, where the largest remaining eigenvalue is λ_2.

Example 7.3.4

For the matrix

$$A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix},$$

find the maximum value of $\mathbf{x}^T A\mathbf{x}$ subject to

$$\mathbf{x}^T\mathbf{x} = 1, \quad \mathbf{x}^T\mathbf{u}_1 = 0,$$

where

$$\mathbf{u}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Find a unit vector at which the maximum is attained.

Solution

The eigenvalues of A are

$$\lambda_1 = 6, \quad \lambda_2 = 3, \quad \lambda_3 = 1.$$

Because the direction associated with the largest eigenvalue 6 has been excluded, the constrained maximum is the second-largest eigenvalue:

$$\max\left\{\mathbf{x}^T A\mathbf{x} : \mathbf{x}^T\mathbf{x} = 1, \quad \mathbf{x}^T\mathbf{u}_1 = 0\right\} = 3$$

Solving $(A - 3I)\mathbf{x} = \mathbf{0}$ gives an eigenvector

$$\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}.$$

Its norm is $\|\mathbf{v}_2\| = \sqrt{6}$, so the corresponding unit eigenvector is

$$\mathbf{u}_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}.$$

Moreover, $\mathbf{u}_1^T \mathbf{u}_2 = 0$, as expected because the two vectors correspond to distinct eigenvalues of a symmetric matrix.

Therefore, the constrained maximum is 3, attained at $\mathbf{x} = \pm\mathbf{u}_2$.

The same procedure may be repeated. After excluding the first $k-1$ orthonormal eigenvector directions, the largest value that remains is the kth-largest eigenvalue.

Successive Constrained Characterization of Eigenvalues

Let A be an $n \times n$ real symmetric matrix with eigenvalues ordered as

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n,$$

and let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ be corresponding orthonormal eigenvectors. For $k = 1, 2, \ldots, n$,

$$\max\left\{\mathbf{x}^T A\mathbf{x} : \mathbf{x}^T\mathbf{x} = 1, \quad \mathbf{x}^T\mathbf{u}_1 = 0, \ldots, \mathbf{x}^T\mathbf{u}_{k-1} = 0\right\} = \lambda_k.$$

The maximum is attained at $\mathbf{x} = \pm\mathbf{u}_k$.

When $k = 1$, there are no orthogonality constraints.

To see why this works, write a unit vector in the orthonormal eigenbasis:

$$\mathbf{x} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n,$$

where

$$c_1^2 + c_2^2 + \cdots + c_n^2 = 1.$$

Then

$$\mathbf{x}^T A\mathbf{x} = \lambda_1 c_1^2 + \lambda_2 c_2^2 + \cdots + \lambda_n c_n^2.$$

The constraints

$$\mathbf{x}^T \mathbf{u}_1 = 0, \ldots, \mathbf{x}^T \mathbf{u}_{k-1} = 0$$

force

$$c_1 = \cdots = c_{k-1} = 0.$$

Therefore,

$$\mathbf{x}^T A\mathbf{x} = \lambda_k c_k^2 + \lambda_{k+1} c_{k+1}^2 + \cdots + \lambda_n c_n^2,$$

which is a weighted average of the remaining eigenvalues. Its largest possible value is λ_k, attained when $c_k = \pm 1$ and every other coefficient is zero.

Example 7.3.5

A city government is planning a transit-oriented redevelopment project in a dense urban district. Let x represent one hundred thousand square meters of new mixed-use office and retail space, and let y represent ten hectares of public gardens, rooftop parks, and elevated green space.

Because both projects compete for funding, land, labor, and construction capacity, the feasible developments satisfy

$$4x^2 + 9y^2 \leq 36, \qquad x \geq 0, \quad y \geq 0.$$

Suppose the overall value assigned to a development plan is modeled by the utility function

$$q(x, y) = xy.$$

Find the feasible development plan that maximizes q.

Solution

Because $q(x, y) = xy$ increases as either x or y increases in the first quadrant, the maximum must occur on the boundary of the feasible region:

$$4x^2 + 9y^2 = 36$$

Rewrite the constraint as

$$\left(\frac{x}{3}\right)^2 + \left(\frac{y}{2}\right)^2 = 1.$$

Define the new variables

$$u_1 = \frac{x}{3}, \qquad u_2 = \frac{y}{2},$$

so that

$$x = 3u_1, \qquad y = 2u_2.$$

Let

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}.$$

The constraint becomes

$$u_1^2 + u_2^2 = \mathbf{u}^T \mathbf{u} = 1,$$

while the utility function becomes

$$q(x, y) = (3u_1)(2u_2) = 6u_1 u_2.$$

This expression can be written as the quadratic form

$$q(x, y) = \mathbf{u}^T A \mathbf{u},$$

where

$$A = \begin{bmatrix} 0 & 3 \\ 3 & 0 \end{bmatrix}.$$

The characteristic polynomial is

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} -\lambda & 3 \\ 3 & -\lambda \end{vmatrix} \\ &= \lambda^2 - 9 \\ &= (\lambda - 3)(\lambda + 3). \end{aligned}$$

Thus, the eigenvalues are

$$\lambda_1 = 3, \quad \lambda_2 = -3.$$

The maximum value of $\mathbf{u}^T A \mathbf{u}$ subject to $\|\mathbf{u}\| = 1$ is therefore the largest eigenvalue,

$$\max(q(x, y)) = 3.$$

A unit eigenvector corresponding to $\lambda_1 = 3$ is

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Therefore,

$$u_1 = u_2 = \frac{1}{\sqrt{2}}.$$

Returning to the original variables gives

$$x = 3u_1 = \frac{3}{\sqrt{2}} \approx 2.12,$$

and

$$y = 2u_2 = \sqrt{2} \approx 1.41.$$

Thus, the utility-maximizing development plan contains approximately 2.12×10^5 m^2 of mixed-use commercial space and $1.41(10) = 14.1$ ha of green space.

The curves $xy = c$ represent development plans with equal utility. At the optimum, the highest attainable utility curve is tangent to the boundary $4x^2 + 9y^2 = 36$, giving $q_{\max} = 3$. See Figure 42.

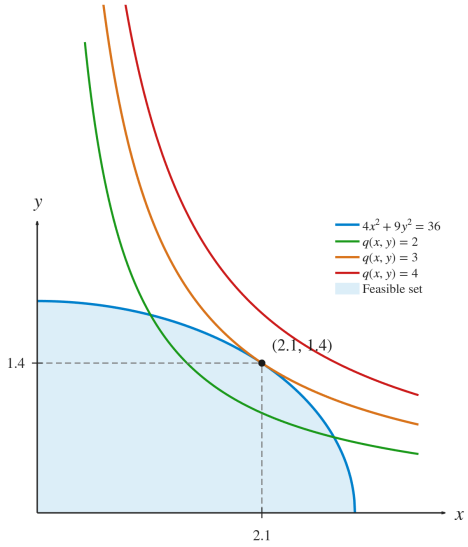


Figure 42: Feasible development plans and utility curves

Figure 43: Proposed commercial space in San Francisco. Image credit: SITELAB

Figure 44: Singapore's Gardens by the Bay is perhaps the most famous urban green space in the world. Image credit: Palani Mohan, National Geographic

Problem 7.3.1

Let $Q(\mathbf{x}) = 3x_1^2 + 2x_1x_2 + 3x_2^2$.

(a) Find an orthogonal change of variables $\mathbf{x} = P\mathbf{y}$ that removes the cross-product term, and write the resulting quadratic form in terms of $\mathbf{y}$.

(b) Find the maximum value of $Q(\mathbf{x})$ subject to $\mathbf{x}^T\mathbf{x} = 1$, and find a unit vector at which the maximum is attained.

Solution

(a) The matrix of the quadratic form is

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix},$$

since $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$.

The characteristic polynomial is

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{vmatrix} \\ &= (3-\lambda)^2 - 1 \\ &= (\lambda - 4)(\lambda - 2). \end{aligned}$$

Thus, the eigenvalues are

$$\lambda_1 = 4, \quad \lambda_2 = 2.$$

For $\lambda_1 = 4$, solving $(A - 4I)\mathbf{v}_1 = \mathbf{0}$ gives

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

For $\lambda_2 = 2$, solving $(A - 2I)\mathbf{v}_2 = \mathbf{0}$ gives

$$\mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Normalizing these eigenvectors gives

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Therefore,

$$P = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}, \quad D = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix},$$

and

$$A = PDP^T.$$

Using the orthogonal change of variables, the quadratic form becomes

$$\begin{aligned} Q(\mathbf{x}) &= \mathbf{x}^T A \mathbf{x} \\ &= \mathbf{y}^T P^T A P \mathbf{y} \\ &= \mathbf{y}^T D \mathbf{y} \\ &= 4y_1^2 + 2y_2^2. \end{aligned}$$

Thus, the transformed quadratic form contains no cross-product term.

(b) Because A is symmetric, the maximum value of $Q(\mathbf{x})$ over all unit vectors is the largest eigenvalue of A. Therefore,

$$\max_{\mathbf{x}^T\mathbf{x}=1} Q(\mathbf{x}) = \lambda_1 = 4.$$

This maximum is attained at any unit eigenvector corresponding to $\lambda_1 = 4$. Hence,

$$\mathbf{x} = \pm \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Indeed,

$$Q\left(\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}\right) = 4.$$

See Figure 43.

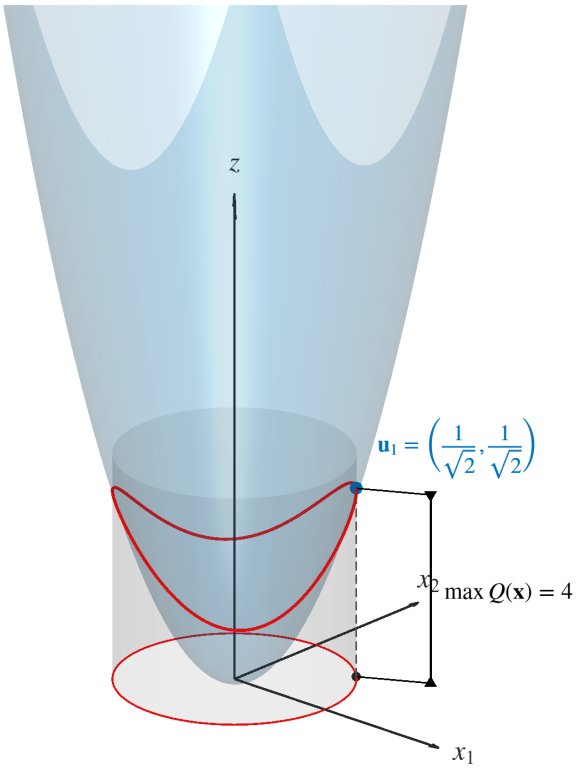


Figure 45: The quadratic surface $Q(\mathbf{x}) = 3x_1^2 + 2x_1x_2 + 3x_2^2$ over the unit-circle constraint, with the maximum value $Q(\mathbf{u}_1) = 4$.

7.4 The Singular Value Decomposition

Diagonalization reveals the structure of a linear transformation by identifying directions that the transformation treats independently. When a square matrix A is diagonalizable, it can be written as

$$A = PDP^{-1}$$

where the columns of P are eigenvectors of A, and the diagonal entries of D are the corresponding eigenvalues. For a real symmetric matrix, the spectral theorem for real symmetric matrices gives the stronger factorization

$$A = PDP^T,$$

where P is orthogonal.

Ordinary diagonalization, however, has two important limitations. First, not every square matrix is diagonalizable. Second, a rectangular matrix does not even have eigenvalues in the usual sense, because it maps between spaces of different dimensions.

The singular value decomposition overcomes both limitations. It provides a factorization for every $m \times n$ matrix, whether square or rectangular, singular or nonsingular, diagonalizable or defective. Instead of searching for directions that remain parallel to themselves under A, the singular value decomposition identifies orthogonal input directions that are mapped to orthogonal output directions.

The geometry behind this factorization begins with the unit sphere

$$\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = 1\}.$$

An $m \times n$ matrix A transforms this sphere into an ellipsoid in $\mathbb{R}^m$, potentially flattened into a lower-dimensional subspace. The principal axes of this ellipsoid reveal the directions in which A stretches, shrinks, or completely collapses vectors.

For a unit vector $\mathbf{x}$, the length of its image is $\|A\mathbf{x}\|$. To understand how this length depends on the direction of $\mathbf{x}$, it is convenient to consider its square:

$$\begin{aligned} \|A\mathbf{x}\|^2 &= (A\mathbf{x})^T(A\mathbf{x}) \\ &= \mathbf{x}^T A^T A\mathbf{x} \end{aligned}$$

Thus, the stretching behavior of A is governed by the quadratic form associated with $A^T A$.

The matrix A^TA is always symmetric because

$$(A^TA)^T = A^T(A^T)^T = A^TA.$$

It is also positive semidefinite. Indeed, for every $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x}^TA^TA\mathbf{x} = \|A\mathbf{x}\|^2 \geq 0.$$

Therefore, every eigenvalue of A^TA is real and nonnegative.

By the spectral theorem, there exists an orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ for $\mathbb{R}^n$ consisting of eigenvectors of A^TA. Let the corresponding eigenvalues be ordered so that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0.$$

For each unit eigenvector $\mathbf{v}_i$,

$$A^TA\mathbf{v}_i = \lambda_i\mathbf{v}_i.$$

The length of $A\mathbf{v}_i$ is determined directly by λ_i:

$$\begin{aligned}\|A\mathbf{v}_i\|^2 &= (A\mathbf{v}_i)^T(A\mathbf{v}_i)\\ &= \mathbf{v}_i^TA^TA\mathbf{v}_i\\ &= \mathbf{v}_i^T(\lambda_i\mathbf{v}_i)\\ &= \lambda_i\mathbf{v}_i^T\mathbf{v}_i\\ &= \lambda_i.\end{aligned}$$

Because $\mathbf{v}_i$ is a unit vector,

$$\|A\mathbf{v}_i\| = \sqrt{\lambda_i}.$$

The eigenvectors of A^TA therefore identify special input directions, while the square roots of the corresponding eigenvalues measure the amount of stretching in those directions.

The connection with constrained optimization is immediate. Since

$$\|A\mathbf{x}\|^2 = \mathbf{x}^TA^TA\mathbf{x},$$

maximizing $\|A\mathbf{x}\|$ over all unit vectors is equivalent to maximizing the quadratic form $\mathbf{x}^T A^T A\mathbf{x}$ subject to $\|\mathbf{x}\| = 1$. Using what we know about the extreme values of a quadratic form,

$$\max_{\|\mathbf{x}\|=1} \mathbf{x}^T A^T A\mathbf{x} = \lambda_1.$$

Consequently,

$$\max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\| = \sqrt{\lambda_1},$$

and the maximum is attained when $\mathbf{x} = \pm\mathbf{v}_1$.

The vector $\mathbf{v}_1$ is the input direction stretched most strongly by A, and $\sqrt{\lambda_1}$ is the corresponding stretching factor. After excluding that direction, the next-largest stretching factor is obtained from $\mathbf{v}_2$, and the process continues through the remaining eigenvectors.

Example 7.4.1

Let

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}.$$

The transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps the unit sphere in $\mathbb{R}^3$ to an ellipse in $\mathbb{R}^2$.
Find a unit vector $\mathbf{x}$ at which $\|A\mathbf{x}\|$ is maximized, and determine the maximum value.

Solution

Because $\|A\mathbf{x}\|^2 = \mathbf{x}^T A^T A\mathbf{x}$, we first compute $A^T A$:

$$A^T A = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$$

$$= \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}.$$

The eigenvalues of A^TA are

$$\lambda_1 = 360, \quad \lambda_2 = 90, \quad \lambda_3 = 0.$$

Corresponding orthonormal eigenvectors are

$$\mathbf{v}_1 = \frac{1}{3}\begin{bmatrix}1\\2\\2\end{bmatrix}, \quad \mathbf{v}_2 = \frac{1}{3}\begin{bmatrix}-2\\-1\\2\end{bmatrix}, \quad \mathbf{v}_3 = \frac{1}{3}\begin{bmatrix}2\\-2\\1\end{bmatrix}.$$

The largest eigenvalue is $\lambda_1 = 360$, so $\|A\mathbf{x}\|^2$ is maximized at $\mathbf{x} = \pm\mathbf{v}_1$.

Using $\mathbf{v}_1$,

$$\begin{aligned} A\mathbf{v}_1 &= \begin{bmatrix}4 & 11 & 14\\ 8 & 7 & -2\end{bmatrix}\frac{1}{3}\begin{bmatrix}1\\2\\2\end{bmatrix} \\ &= \begin{bmatrix}18\\6\end{bmatrix}. \end{aligned}$$

Therefore,

$$\begin{aligned} \|A\mathbf{v}_1\| &= \sqrt{18^2 + 6^2} \\ &= \sqrt{360} \\ &= 6\sqrt{10}. \end{aligned}$$

Hence,

$$\max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\| = 6\sqrt{10},$$

attained at

$$\mathbf{x} = \pm\frac{1}{3}\begin{bmatrix}1\\2\\2\end{bmatrix}.$$

Geometrically, $A\mathbf{v}_1$ points along the major axis of the ellipse obtained by transforming the unit sphere.

Singular Values

Let A be an $m \times n$ matrix, and let

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

be the eigenvalues of A^TA. The **singular values** of A are

$$\sigma_i = \sqrt{\lambda_i}, \qquad i = 1, 2, \ldots, n.$$

They are conventionally ordered as

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0.$$

The relation $\|A\mathbf{v}_i\| = \sigma_i$ gives the singular values a direct geometric meaning. Each σ_i is the length of the image of the corresponding unit eigenvector $\mathbf{v}_i$.

For the matrix in Example 7.4.1, we have

$$\lambda_1 = 360, \quad \lambda_2 = 90, \quad \lambda_3 = 0.$$

Thus the singular values are

$$\sigma_1 = \sqrt{360} = 6\sqrt{10}, \quad \sigma_2 = \sqrt{90} = 3\sqrt{10}, \quad \sigma_3 = 0.$$

The first singular value is the greatest amount by which A stretches a unit vector. The second singular value is the greatest amount of stretching among unit vectors orthogonal to $\mathbf{v}_1$.

Indeed,

$$A\mathbf{v}_2 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \frac{1}{3} \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix}$$
$$= \begin{bmatrix} 3 \\ -9 \end{bmatrix},$$

and therefore

$$\|A\mathbf{v}_2\| = \sqrt{3^2 + (-9)^2} = \sqrt{90} = 3\sqrt{10}.$$

The vectors $A\mathbf{v}_1$ and $A\mathbf{v}_2$ point along the major and minor axes of the ellipse. Their lengths are the corresponding semiaxis lengths.

The fact that these two output vectors are orthogonal is not a coincidence.

Orthogonality of the Transformed Eigenvectors

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be orthonormal eigenvectors of A^TA, with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.
Then $A\mathbf{v}_i$ and $A\mathbf{v}_j$ are orthogonal whenever $i \neq j$.

Proof

For $i \neq j$,

$$\begin{aligned}(A\mathbf{v}_i)^T(A\mathbf{v}_j) &= \mathbf{v}_i^T A^T A\mathbf{v}_j \\ &= \mathbf{v}_i^T(\lambda_j \mathbf{v}_j) \\ &= \lambda_j \mathbf{v}_i^T \mathbf{v}_j \\ &= 0,\end{aligned}$$

because $\mathbf{v}_i$ and $\mathbf{v}_j$ are orthogonal. Therefore,

$$A\mathbf{v}_i \perp A\mathbf{v}_j.$$

□

The eigenvectors of A^TA are therefore orthogonal input directions that A maps to orthogonal output directions.

Some of these output vectors may be zero. Since $\|A\mathbf{v}_i\| = \sigma_i$, we have $A\mathbf{v}_i = \mathbf{0}$ exactly when $\sigma_i = 0$.

Thus, the nonzero singular values identify the directions that survive the transformation, while zero singular values identify directions that are collapsed into the null space.

Singular Values and Rank

Suppose A has exactly r nonzero singular values. Then $\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_r\}$ is an orthogonal basis for $\text{Col}(A)$. Consequently, $\text{rank}(A) = r$.

Proof

Because the first r singular values are nonzero,

$$\|A\mathbf{v}_i\| = \sigma_i > 0, \qquad i = 1, 2, \ldots, r.$$

Therefore, $A\mathbf{v}_i \neq \mathbf{0}$ for $i \leq r$. The preceding theorem shows that these vectors are mutually orthogonal, so they are linearly independent.

For $i > r$, $\sigma_i = 0$, and hence $A\mathbf{v}_i = \mathbf{0}$.

Now let $\mathbf{y} \in \text{Col}(A)$. Then $\mathbf{y} = A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$. Because $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthonormal basis for $\mathbb{R}^n$, we may write

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n.$$

Therefore,

$$\begin{aligned}\mathbf{y} &= A\mathbf{x} \\ &= c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2 + \cdots + c_nA\mathbf{v}_n \\ &= c_1A\mathbf{v}_1 + \cdots + c_rA\mathbf{v}_r,\end{aligned}$$

because $A\mathbf{v}_i = \mathbf{0}$ for $i > r$.

Thus,

$$\text{Col}(A) = \text{Span}\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_r\}.$$

Since these vectors are also linearly independent, they form an orthogonal basis for $\mathrm{Col}(A)$. Hence,

$$\mathrm{rank}(A) = \dim \mathrm{Col}(A) = r.$$

□

The geometry is now almost complete. The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ provide orthonormal directions in the input space. The nonzero vectors $A\mathbf{v}_1, A\mathbf{v}_2 \ldots, A\mathbf{v}_r$ provide orthogonal directions in the output space, and their lengths are $\sigma_1, \sigma_2, \ldots, \sigma_r$.

We are now ready to assemble the orthogonal input directions, orthogonal output directions, and singular values into a single matrix factorization.

Let A be an $m \times n$ matrix with rank r, and let its nonzero singular values be ordered as

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0.$$

Define the $r \times r$ diagonal matrix

$$D = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix}.$$

The singular values are placed inside an $m \times n$ rectangular diagonal matrix

$$\Sigma = \left[\begin{array}{c|c} D & 0 \\ \hline 0 & 0 \end{array}\right].$$

Here, the zero blocks have the dimensions needed to make Σ an $m \times n$ matrix. More precisely,

$$D \in \mathbb{R}^{r \times r},$$

while the remaining blocks have dimensions

$$r \times (n-r), \quad (m-r) \times r, \quad (m-r) \times (n-r).$$

When $r = m$, $r = n$, or both, some of these zero blocks disappear.

Singular Value Decomposition

Let A be an $m \times n$ matrix with rank r. Then there exist an $m \times m$ orthogonal matrix U, an $n \times n$ orthogonal matrix V, and an $m \times n$ rectangular diagonal matrix Σ such that

$$A = U\Sigma V^T.$$

The first r diagonal entries of Σ are the nonzero singular values

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0,$$

and every remaining entry of Σ is zero.

The columns of U are called the **left-singular vectors** of A, while the columns of V are called the **right-singular vectors** of A.

Proof

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A^TA, ordered so that the corresponding eigenvalues satisfy

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n.$$

The singular values are

$$\sigma_i = \sqrt{\lambda_i}, \qquad i = 1, 2, \ldots, r.$$

From the preceding theorem, the vectors $A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_r$ form an orthogonal basis for $\mathrm{Col}(A)$. Normalize each of these vectors by defining

$$\mathbf{u}_i = \frac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|}, \qquad i = 1, 2, \ldots, r.$$

Because $\|A\mathbf{v}_i\| = \sigma_i$, we may write

$$\mathbf{u}_i = \frac{1}{\sigma_i} A\mathbf{v}_i.$$

Equivalently,

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i, \qquad i = 1, 2, \ldots, r.$$

The vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r$ form an orthonormal basis for $\text{Col}(A)$. Extend this set to an orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ for $\mathbb{R}^m$, and define

$$U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \end{bmatrix},$$

and

$$V = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}.$$

Because the columns of U and V are orthonormal,

$$U^T U = I_m, \quad V^T V = I_n.$$

Thus, U and V are orthogonal matrices.

Now consider the product AV. Since the columns of V are the vectors $\mathbf{v}_i$,

$$AV = \begin{bmatrix} A\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_n \end{bmatrix}.$$

For $i \leq r$,

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i.$$

For $i > r$, the corresponding singular value is zero, so

$$\|A\mathbf{v}_i\| = \sigma_i = 0 \Rightarrow A\mathbf{v}_i = \mathbf{0}.$$

Therefore,

$$AV = \begin{bmatrix} \sigma_1 \mathbf{u}_1 & \sigma_2 \mathbf{u}_2 & \cdots & \sigma_r \mathbf{u}_r & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}.$$

By the construction of Σ,

$$U\Sigma = \begin{bmatrix} \sigma_1 \mathbf{u}_1 & \sigma_2 \mathbf{u}_2 & \cdots & \sigma_r \mathbf{u}_r & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}.$$

Hence,

$$AV = U\Sigma.$$

Since V is orthogonal,

$$VV^T = I_n.$$

Multiplying $AV = U\Sigma$ on the right by V^T gives

$$AVV^T = U\Sigma V^T,$$

and therefore

$$A = U\Sigma V^T.$$

□

The three factors describe the transformation geometrically:

$$\mathbf{x} \overset{V^T}{\longmapsto} V^T\mathbf{x} \overset{\Sigma}{\longmapsto} \Sigma V^T\mathbf{x} \overset{U}{\longmapsto} U\Sigma V^T\mathbf{x}.$$

The matrix V^T expresses the input vector in the orthonormal basis of right singular vectors. The matrix Σ stretches each coordinate direction by its corresponding singular value, while collapsing directions associated with zero singular values. Finally, U places the resulting vector into the orthonormal basis of left singular vectors in the output space.

Thus, every matrix transformation can be understood as

orthogonal change of input coordinates
↓
axis-aligned stretching or collapsing
↓
orthogonal change of output coordinates

Example 7.4.2

Construct a singular value decomposition of

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}.$$

Solution

From Example 7.4.1, the eigenvalues of A^TA are

$$\lambda_1 = 360, \quad \lambda_2 = 90, \quad \lambda_3 = 0,$$

with corresponding orthonormal eigenvectors

$$\mathbf{v}_1 = \frac{1}{3}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \quad \mathbf{v}_2 = \frac{1}{3}\begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix}, \quad \mathbf{v}_3 = \frac{1}{3}\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}.$$

1. Construct V.

The right singular vectors are the orthonormal eigenvectors of A^TA. Therefore,

$$V = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

2. Construct Σ.

The singular values are the square roots of the eigenvalues of A^TA:

$$\sigma_1 = 6\sqrt{10}, \quad \sigma_2 = 3\sqrt{10}, \quad \sigma_3 = 0$$

Because A is a 2×3 matrix with rank 2,

$$\Sigma = \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}.$$

3. Construct U.

For each nonzero singular value,

$$\mathbf{u}_i = \frac{1}{\sigma_i} A\mathbf{v}_i.$$

Thus,

$$\mathbf{u}_1 = \frac{1}{6\sqrt{10}} \begin{bmatrix} 18 \\ 6 \end{bmatrix} = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ 1 \end{bmatrix},$$

$$\mathbf{u}_2 = \frac{1}{3\sqrt{10}} \begin{bmatrix} 3 \\ -9 \end{bmatrix} = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ -3 \end{bmatrix}.$$

These vectors already form an orthonormal basis for $\mathbb{R}^2$, so

$$U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \end{bmatrix}.$$

Therefore, a singular value decomposition of A is

$$A = \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ -\frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix},$$

where the three factors are U, Σ, and V^T, respectively.

Example 7.4.3

Find a singular value decomposition of

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

Solution

1. Construct V.

First,

$$A^T A = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}.$$

Its eigenvalues are

$$\lambda_1 = 18, \quad \lambda_2 = 0,$$

with corresponding orthonormal eigenvectors

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Therefore,

$$V = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

2. Construct Σ.

The singular values are

$$\sigma_1 = \sqrt{18} = 3\sqrt{2}, \quad \sigma_2 = 0.$$

Since A is a 3×2 matrix,

$$\Sigma = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Because A has only one nonzero singular value, $\operatorname{rank}(A) = 1$.

3. Construct U.

The left singular vector corresponding to σ_1 is

$$\begin{aligned} \mathbf{u}_1 &= \frac{1}{\sigma_1} A\mathbf{v}_1 \\ &= \frac{1}{3\sqrt{2}} \begin{bmatrix} \sqrt{2} \\ -2\sqrt{2} \\ 2\sqrt{2} \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}. \end{aligned}$$

Since U must be a 3×3 orthogonal matrix, $\mathbf{u}_1$ must be extended to an orthonormal basis for $\mathbb{R}^3$. One possible choice is

$$\mathbf{u}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{u}_3 = \frac{1}{\sqrt{45}} \begin{bmatrix} -2 \\ 4 \\ 5 \end{bmatrix}.$$

Thus,

$$U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{2}{\sqrt{5}} & -\frac{2}{\sqrt{45}} \\ -\frac{2}{3} & \frac{1}{\sqrt{5}} & \frac{4}{\sqrt{45}} \\ \frac{2}{3} & 0 & \frac{5}{\sqrt{45}} \end{bmatrix}.$$

Therefore, a singular value decomposition of A is

$$A = \begin{bmatrix} \frac{1}{3} & \frac{2}{\sqrt{5}} & -\frac{2}{\sqrt{45}} \\ -\frac{2}{3} & \frac{1}{\sqrt{5}} & \frac{4}{\sqrt{45}} \\ \frac{2}{3} & 0 & \frac{5}{\sqrt{45}} \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

The vectors $\mathbf{u}_2$ and $\mathbf{u}_3$ are not uniquely determined. Any orthonormal completion of $\mathbf{u}_1$ produces a valid full SVD.

Example 7.4.4

Let $A = U\Sigma V^T$ be an SVD of an $m \times n$ matrix A with rank r. Let the columns of U be $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$, and let the columns of V be $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.

Use the singular vectors to find orthonormal bases for the four fundamental subspaces of A.

Solution

For each nonzero singular value,

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i, \qquad i = 1, 2, \ldots, r.$$

Therefore, the first r left singular vectors point in the directions produced by the transformation A. They form an orthonormal basis for the column space:

$$\text{Col}(A) = \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$$

Because U is orthogonal, the remaining columns of U span the orthogonal complement of $\text{Col}(A)$. Since

$$\text{Col}(A)^\perp = \text{Nul}(A^T),$$

we obtain

$$\text{Nul}(A^T) = \text{Span}\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\}.$$

For $i > r$, the corresponding singular value is zero. Hence,

$$\|A\mathbf{v}_i\| = \sigma_i = 0,$$

so $A\mathbf{v}_i = \mathbf{0}$.

Thus, the remaining right singular vectors form an orthonormal basis for the null space:

$$\text{Nul}(A) = \text{Span}\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$$

Finally,

$$\text{Row}(A) = \text{Nul}(A)^{\perp}.$$

Because the columns of V form an orthonormal basis for $\mathbb{R}^n$, the first r right singular vectors form an orthonormal basis for the row space:

$$\text{Row}(A) = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$$

The SVD therefore organizes all four fundamental subspaces as such:

$$\begin{aligned}
\text{Col}(A) &= \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_r\} \\
\text{Nul}(A^T) &= \text{Span}\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\} \\
\text{Row}(A) &= \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\} \\
\text{Nul}(A) &= \text{Span}\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}
\end{aligned}$$

The left-singular vectors divide the output space into $\text{Col}(A)$ and $\text{Nul}(A^T)$, while the right-singular vectors divide the input space into $\text{Row}(A)$ and $\text{Nul}(A)$.

We can now complete the invertible matrix theorem.

The Invertible Matrix Theorem (concluded)

Let A be an $n \times n$ matrix. Then the following statements are each equivalent to the statement that A is invertible:

(u) $\text{Col}(A)^{\perp} = \{\mathbf{0}\}$.

(v) $\text{Nul}(A)^{\perp} = \mathbb{R}^n$.

(w) $\text{Row}(A) = \mathbb{R}^n$.

(x) A has n nonzero singular values.

Example 7.4.5

Let $A = U\Sigma V^T$ be an SVD of an $m \times n$ matrix with rank r. Show how the zero rows and columns of Σ may be removed to obtain the reduced SVD, and use this factorization to define the pseudoinverse of A.

Solution

Let

$$U_r = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_r \end{bmatrix}, \quad V_r = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_r \end{bmatrix},$$

and let

$$D = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix}.$$

The columns of U_r and V_r correspond precisely to the nonzero singular values. Partition the full orthogonal matrices as

$$U = \begin{bmatrix} U_r & U_0 \end{bmatrix}, \quad V = \begin{bmatrix} V_r & V_0 \end{bmatrix}.$$

Then

$$\Sigma = \left[\begin{array}{c|c} D & 0 \\ \hline 0 & 0 \end{array}\right].$$

Substituting these partitions into the full SVD gives

$$\begin{aligned} A &= \begin{bmatrix} U_r & U_0 \end{bmatrix} \left[\begin{array}{c|c} D & 0 \\ \hline 0 & 0 \end{array}\right] \begin{bmatrix} V_r^T \\ V_0^T \end{bmatrix} \\ &= U_r D V_r^T. \end{aligned}$$

Thus, the **reduced singular value decomposition** is

$$A = U_r D V_r^T,$$

where

$$U_r \in \mathbb{R}^{m \times r}, \quad D \in \mathbb{R}^{r \times r}, \quad V_r \in \mathbb{R}^{n \times r}.$$

The reduced SVD contains all of the information needed to reconstruct A, while omitting the singular vectors associated with zero singular values.

Because every diagonal entry of D is positive, D is invertible:

$$D^{-1} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_r} \end{bmatrix}$$

The **pseudoinverse**, or **Moore–Penrose inverse**, of A is defined by

$$A^+ = V_r D^{-1} U_r^T.$$

The pseudoinverse reverses the nonzero stretching performed by A. Each nonzero singular value σ_i is replaced by its reciprocal $1/\sigma_i$, while directions collapsed by A remain excluded.

When A is square and invertible, $r = n$, and the pseudoinverse agrees with the ordinary inverse:

$$A^+ = A^{-1}$$

Example 7.4.6

Let A be an $m \times n$ matrix with reduced SVD $A = U_r D V_r^T$.

Show that

$$\widehat{\mathbf{x}} = A^+ \mathbf{b}$$

is a least-squares solution of $A\mathbf{x} = \mathbf{b}$, and that it has the smallest norm among all least-squares solutions.

Solution

Using the pseudoinverse,

$$\widehat{\mathbf{x}} = A^+ \mathbf{b} = V_r D^{-1} U_r^T \mathbf{b}.$$

Multiplying by A gives

$$\begin{aligned} A\widehat{\mathbf{x}} &= \left(U_r D V_r^T\right)\left(V_r D^{-1} U_r^T \mathbf{b}\right) \\ &= U_r D \left(V_r^T V_r\right) D^{-1} U_r^T \mathbf{b} \\ &= U_r D D^{-1} U_r^T \mathbf{b} \\ &= U_r U_r^T \mathbf{b}. \end{aligned}$$

Because the columns of U_r form an orthonormal basis for $\mathrm{Col}(A)$, the matrix $U_r U_r^T$ is the orthogonal projection matrix onto $\mathrm{Col}(A)$. Therefore,

$$A\widehat{\mathbf{x}} = \mathrm{proj}_{\mathrm{Col}(A)} \mathbf{b}.$$

The residual is

$$\mathbf{b} - A\widehat{\mathbf{x}} = \mathbf{b} - \mathrm{proj}_{\mathrm{Col}(A)} \mathbf{b},$$

which is orthogonal to $\mathrm{Col}(A)$. Hence,

$$A^T(\mathbf{b} - A\widehat{\mathbf{x}}) = \mathbf{0},$$

so $\widehat{\mathbf{x}}$ satisfies the normal equations and is therefore a least-squares solution.

It remains to show that this solution has minimum norm. Because

$$\widehat{\mathbf{x}} = V_r D^{-1} U_r^T \mathbf{b},$$

the vector $\widehat{\mathbf{x}}$ belongs to

$$\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\} = \text{Row}(A).$$

Every other least-squares solution has the form

$$\mathbf{x} = \widehat{\mathbf{x}} + \mathbf{z}, \quad \mathbf{z} \in \text{Nul}(A).$$

Since $\text{Row}(A) = \text{Nul}(A)^\perp$, the vectors $\widehat{\mathbf{x}}$ and $\mathbf{z}$ are orthogonal. Therefore,

$$\begin{aligned} \|\mathbf{x}\|^2 &= \|\widehat{\mathbf{x}} + \mathbf{z}\|^2 \\ &= \|\widehat{\mathbf{x}}\|^2 + \|\mathbf{z}\|^2 \\ &\geq \|\widehat{\mathbf{x}}\|^2. \end{aligned}$$

Equality occurs only when $\mathbf{z} = \mathbf{0}$. Consequently,

$$\widehat{\mathbf{x}} = A^+ \mathbf{b}$$

is the unique least-squares solution with minimum norm.

Computing the singular values and singular vectors from $A^T A$ is convenient for hand calculations. In numerical computation, however, explicitly forming $A^T A$ is usually avoided. Its condition number satisfies

$$\kappa_2(A^T A) = \kappa_2(A)^2,$$

so forming $A^T A$ can amplify numerical sensitivity and reduce accuracy.

Modern software instead computes the SVD using more stable algorithms, typically beginning with a reduction of A to bidiagonal form. In MATLAB, the function `[U,S,V] = svd(A)` returns matrices satisfying $A = USV^T$, where the diagonal entries of S are the singular values of A.

Problem 7.4.1

Suppose $A = U\Sigma V^T$ is a singular value decomposition of A. Find an SVD of A^T. How are the singular values of A and A^T related?

Solution

Taking the transpose of the factorization gives

$$\begin{aligned} A^T &= \left(U\Sigma V^T\right)^T \\ &= V\Sigma^T U^T. \end{aligned}$$

Because U and V are orthogonal matrices and Σ^T is again a rectangular diagonal matrix, this is a singular value decomposition of A^T.

The matrices Σ and Σ^T have the same nonzero diagonal entries. Therefore, A and A^T have the same nonzero singular values, with the same multiplicities.

When A is rectangular, the matrices A^TA and AA^T may contain different numbers of zero eigenvalues, but their positive eigenvalues, and hence the nonzero singular values of A and A^T, are identical.

Problem 7.4.2

Let A be an $n \times n$ matrix. Use the singular value decomposition to show that AA^T and A^TA are orthogonally similar.

Solution

Let $A = U\Sigma V^T$ be a singular value decomposition of A. Since A is square, Σ is an $n \times n$ diagonal matrix, so $\Sigma^T = \Sigma$.

First,

$$\begin{aligned} AA^T &= \left(U\Sigma V^T\right)\left(U\Sigma V^T\right)^T \\ &= U\Sigma V^T V\Sigma U^T \\ &= U\Sigma^2 U^T. \end{aligned}$$

Similarly,

$$\begin{aligned} A^T A &= \left(U\Sigma V^T\right)^T\left(U\Sigma V^T\right) \\ &= V\Sigma U^T U\Sigma V^T \\ &= V\Sigma^2 V^T. \end{aligned}$$

Define

$$Q = UV^T.$$

Because U and V are orthogonal,

$$\begin{aligned} Q^T Q &= VU^T UV^T \\ &= VV^T \\ &= I, \end{aligned}$$

so Q is orthogonal. Moreover,

$$\begin{aligned} Q^T\left(AA^T\right)Q &= \left(VU^T\right)\left(U\Sigma^2 U^T\right)\left(UV^T\right) \\ &= V\Sigma^2 V^T \\ &= A^T A. \end{aligned}$$

Therefore,

$$A^T A = Q^T\left(AA^T\right)Q,$$

so AA^T and A^TA are orthogonally similar.

Problem 7.4.3

Find a singular value decomposition of

$$A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}.$$

Hint: It is more efficient to begin with AA^T than with A^TA.

Solution

Because A has only two rows, begin with

$$AA^T = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix}.$$

The eigenvalues of AA^T are

$$\lambda_1 = 25, \quad \lambda_2 = 9,$$

with corresponding orthonormal eigenvectors

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Therefore, the singular values are

$$\sigma_1 = 5, \quad \sigma_2 = 3,$$

and

$$U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

The corresponding right-singular vectors are found from

$$\mathbf{v}_i = \frac{1}{\sigma_i} A^T \mathbf{u}_i.$$

For $\sigma_1 = 5$,

$$\begin{aligned}\mathbf{v}_1 &= \frac{1}{5}A^T\mathbf{u}_1 \\ &= \frac{1}{5}\begin{bmatrix}3 & 2\\ 2 & 3\\ 2 & -2\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}1\\ 1\end{bmatrix} \\ &= \frac{1}{\sqrt{2}}\begin{bmatrix}1\\ 1\\ 0\end{bmatrix}.\end{aligned}$$

For $\sigma_2 = 3$,

$$\begin{aligned}\mathbf{v}_2 &= \frac{1}{3}A^T\mathbf{u}_2 \\ &= \frac{1}{3}\begin{bmatrix}3 & 2\\ 2 & 3\\ 2 & -2\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}-1\\ 1\end{bmatrix} \\ &= \frac{1}{\sqrt{18}}\begin{bmatrix}-1\\ 1\\ -4\end{bmatrix}.\end{aligned}$$

Since V must be a 3×3 orthogonal matrix, choose a unit vector orthogonal to both $\mathbf{v}_1$ and $\mathbf{v}_2$. One such vector is

$$\mathbf{v}_3 = \frac{1}{3}\begin{bmatrix}-2\\ 2\\ 1\end{bmatrix}.$$

Thus,

$$V = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3\end{bmatrix} = \begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} & -\frac{2}{3}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{18}} & \frac{2}{3}\\ 0 & -\frac{4}{\sqrt{18}} & \frac{1}{3}\end{bmatrix},$$

and

$$\Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}.$$

Therefore, a singular value decomposition of A is

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{18}} & \frac{1}{\sqrt{18}} & -\frac{4}{\sqrt{18}} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

The reduced singular value decomposition may be written as

$$A = U_r \Sigma_r V_r^T,$$

where $U_r \in \mathbb{R}^{m \times r}$, $V_r \in \mathbb{R}^{n \times r}$, and

$$\Sigma_r = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r).$$

Writing the matrix product one singular value at a time gives

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T.$$

Each matrix $\sigma_i \mathbf{u}_i \mathbf{v}_i^T$ has rank one. Thus, the SVD expresses A as a sum of rank-one matrices arranged from most significant to least significant according to the magnitudes of their singular values.

Rank-One Expansion

Let A be an $m \times n$ matrix with rank r and nonzero singular values

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0.$$

Then

$$A = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T.$$

To approximate A with a matrix of lower rank, retain only the first k terms:

$$A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \qquad 1 \leq k < r.$$

The discarded terms correspond to the smaller singular values and therefore represent directions in which the transformation has less influence.

Let

$$U_k = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{bmatrix},$$

let

$$\Sigma_k = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k \end{bmatrix},$$

and let

$$V_k = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{bmatrix}.$$

Then the rank-k truncated SVD may be written as

$$A_k = U_k \Sigma_k V_k^T = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T.$$

Thus, A_k is constructed by retaining the first k left singular vectors, the first k singular values, and the first k right singular vectors. Because these singular values are nonzero,

$$\operatorname{rank}(A_k) = k.$$

The problem of finding a best rank-k approximation of A may be written as

$$\min_{\operatorname{rank}(B) \leq k} \|A - B\|_F.$$

Equivalently, a minimizing matrix belongs to the set

$$\arg\min_{\operatorname{rank}(B) \leq k} \|A - B\|_F.$$

The following theorem shows that the truncated SVD A_k solves this optimization problem.

Eckart–Young Theorem

Let A be an $m \times n$ matrix of rank r, and let $1 \leq k < r$.

Among all matrices B satisfying $\operatorname{rank}(B) \leq k$, the truncated SVD A_k is the closest approximation to A. In particular,

$$\|A - A_k\|_2 = \sigma_{k+1},$$

and

$$\|A - A_k\|_F = \sqrt{\sum_{i=k+1}^{r} \sigma_i^2}.$$

The theorem shows that truncating the SVD is not merely a convenient method for producing a low-rank matrix. It gives the best possible rank-k approximation under both the matrix 2-norm and the Frobenius norm.

A grayscale digital image with m rows and n columns may be represented by an $m \times n$ matrix A, where each entry records the brightness of one pixel. A color image may be represented by three matrices,

$$A_R, \quad A_G, \quad A_B, \tag{28}$$

corresponding to its red, green, and blue intensities of each pixel, respectively.

To compress a color image, compute a truncated SVD of each channel:

$$(A_R)_k, \quad (A_G)_k, \quad (A_B)_k,$$

and then recombine the three approximations. When the singular values decrease rapidly, a relatively small value of k can preserve most of the visible structure of the original image.

An uncompressed $m \times n$ grayscale image stores mn numerical values. A rank-k SVD approximation stores $mk + k + nk = k(m + n + 1)$ values: the entries of U_k, the first k singular values, and the entries of V_k^T. Compression is therefore effective when

$$k(m + n + 1) < mn.$$

This reduction is possible because many images have high exact rank but much lower effective rank: most of their important visual information is concentrated in the first several singular values.

To illustrate low-rank approximation, we apply the truncated SVD to Vincent van Gogh's *The Starry Night*. This application was inspired by a low-rank image-compression in [3]. The digital image has 507 rows, 640 columns, and three color channels. It is represented by the matrices in (28).

The Python script first converts the image data to numerical values between 0 and 1. It then computes a separate singular value decomposition for each color channel:

$$A_C = U_C \Sigma_C V_C^T, \quad C \in \{R, G, B\}.$$

For a chosen rank k, only the first k singular values and corresponding singular vectors are retained:

$$(A_C)_k = (U_C)_k (\Sigma_C)_k (V_C)_k^T$$

The three approximated channels are then recombined to produce the rank-k color image.

For each channel, the original matrix requires mn stored values. Its truncated SVD requires

$$mk + k + nk = k(m + n + 1)$$

values. Since the same calculation is applied to all three color channels, the factor of three cancels when comparing storage. Thus, the theoretical storage fraction is

$$\frac{k(m + n + 1)}{mn}.$$

For this 507×640 image, the rank-40 approximation uses approximately

$$\frac{40(507 + 640 + 1)}{507 \cdot 640} \approx 0.142,$$

or 14.2% of the numerical storage required by the original image.

The script also computes the relative Frobenius error

$$\frac{\sqrt{\|A_R - (A_R)_k\|_F^2 + \|A_G - (A_G)_k\|_F^2 + \|A_B - (A_B)_k\|_F^2}}{\sqrt{\|A_R\|_F^2 + \|A_G\|_F^2 + \|A_B\|_F^2}}.$$

As k increases, additional singular-vector terms restore progressively finer visual information. At rank 5, only the broad composition and dominant colors remain. By rank 15, the sky, cypress tree, and village are clearly recognizable. The rank-40 approximation restores much of the painting's structure while requiring relatively little storage, while the rank-80 approximation is visually close to the original.

Although the rank-40 approximation has a relative Frobenius error of approximately 17.8%, it preserves the painting remarkably well. This difference between numerical error and perceived image quality arises because the Frobenius norm weights every pixel error equally, whereas human vision is generally more sensitive to large-scale structure than to small losses of texture and detail. This example demonstrates the meaning of effective rank. Although an image matrix may have high exact rank, much of its visually important information may be concentrated in a comparatively small number of singular values.

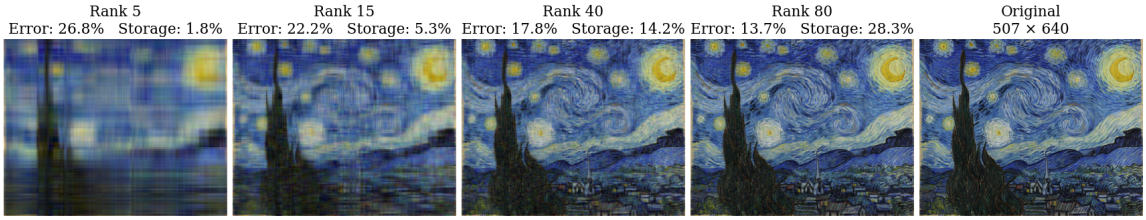


Figure 46: Image compression of *The Starry Night*

View the Python code

8 Chosen Topics

This chapter provides a brief overview of several applications of linear algebra that I hope you will find interesting!

8.1 Markov Chains

Many systems evolve through a sequence of distinct states. The weather may change from sunny to cloudy, customers may switch from one app to another, and populations may move between cities and suburbs. In each case, the future state of the system may depend probabilistically on its present state.

A **Markov chain** models this type of evolution under the assumption that the probability of entering the next state depends only on the current state, not on the complete history of the system.

In Section 2.8, we used discrete dynamical systems to describe populations moving between different regions or states. When the entries of the state vector represent probabilities or proportions and the entries of the transition matrix represent the probabilities of moving between states, the same framework becomes a Markov chain. Linear algebra then allows us to predict how the system evolves over time and, in many cases, determine its long-term behavior.

Let $\mathbf{x}_k$ describe the state of a system after k transitions. The evolution of the system is governed by

$$\mathbf{x}_{k+1} = P\mathbf{x}_k,$$

where P is a stochastic matrix. Each entry of P is nonnegative, and each column of P sums to one. Repeated substitution gives

$$\mathbf{x}_k = P^k\mathbf{x}_0.$$

Thus, predicting the state of the system far into the future reduces to understanding the behavior of P^k as k becomes large.

One interesting application of Markov chains is using them to predict the distant future.

A state vector $\mathbf{q}$ is called a **steady-state vector** if it remains unchanged after another transition:

$$P\mathbf{q} = \mathbf{q}$$

Therefore, a steady-state vector is an eigenvector of P corresponding to the eigenvalue $\lambda = 1$. Because $\mathbf{q}$ represents probabilities or proportions, its entries must also satisfy

$$q_1 + q_2 + \cdots + q_n = 1.$$

Steady-State Theorem

Suppose that P is a regular stochastic matrix, meaning that some power P^m has only positive entries. Then P has a unique steady-state vector $\mathbf{q}$, and

$$\lim_{k\to\infty} P^k \mathbf{x}_0 = \mathbf{q}$$

for every initial probability vector $\mathbf{x}_0$.

Thus, although two systems may begin with different initial distributions, their long-term behavior may approach the same steady state.

Example 8.1.1

Consider the Markov chain

$$\mathbf{x}_{k+1} = P\mathbf{x}_k, \quad P = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0 & 0.4 \end{bmatrix}, \quad \mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Determine the long-term behavior of the system.

Solution

The first several state vectors are obtained by repeatedly multiplying by P:

$$\mathbf{x}_1 = P\mathbf{x}_0 = \begin{bmatrix} 0.5 \\ 0.3 \\ 0.2 \end{bmatrix}, \quad \mathbf{x}_2 = P\mathbf{x}_1 = \begin{bmatrix} 0.37 \\ 0.45 \\ 0.18 \end{bmatrix}, \quad \mathbf{x}_3 = P\mathbf{x}_2 = \begin{bmatrix} 0.329 \\ 0.525 \\ 0.146 \end{bmatrix}$$

Continuing the process gives

$$\mathbf{x}_5 = \begin{bmatrix} 0.30641 \\ 0.58125 \\ 0.11234 \end{bmatrix}, \quad \mathbf{x}_{10} = \begin{bmatrix} 0.30020 \\ 0.59941 \\ 0.10039 \end{bmatrix}, \quad \mathbf{x}_{15} = \begin{bmatrix} 0.30001 \\ 0.59998 \\ 0.10001 \end{bmatrix}.$$

These vectors appear to approach

$$\mathbf{q} = \begin{bmatrix} 0.3 \\ 0.6 \\ 0.1 \end{bmatrix}.$$

To determine the steady state directly, solve $P\mathbf{q} = \mathbf{q}$, or equivalently,

$$(P - I)\mathbf{q} = \mathbf{0},$$

together with the probability condition

$$q_1 + q_2 + q_3 = 1.$$

Thus,

$$\begin{bmatrix} -0.5 & 0.2 & 0.3 & 0 \\ 0.3 & -0.2 & 0.3 & 0 \\ 0.2 & 0 & -0.6 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0.3 \\ 0 & 1 & 0 & 0.6 \\ 0 & 0 & 1 & 0.1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Therefore,

$$\lim_{k \to \infty} \mathbf{x}_k = \begin{bmatrix} 0.3 \\ 0.6 \\ 0.1 \end{bmatrix}.$$

In the long run, approximately 30% of the system occupies the first state, 60% occupies the second state, and 10% occupies the third state. The initial distribution gradually disappears from the prediction, while the steady-state distribution remains.

Example 8.1.2

Suppose the stock market is classified each month as being in one of three states: a bull market, a stable market, or a bear market. The transition matrix

$$P = \begin{bmatrix} 0.70 & 0.30 & 0.15 \\ 0.20 & 0.50 & 0.25 \\ 0.10 & 0.20 & 0.60 \end{bmatrix}$$

describes the probability of moving between these states from one month to the next. The current market distribution is

$$\mathbf{x}_0 = \begin{bmatrix} 0.50 \\ 0.30 \\ 0.20 \end{bmatrix},$$

where the entries correspond to the probabilities of a bull, stable, and bear market, respectively. Predict the market distribution one month from now.

Solution

The predicted distribution is

$$\mathbf{x}_1 = P\mathbf{x}_0 = \begin{bmatrix} 0.70 & 0.30 & 0.15 \\ 0.20 & 0.50 & 0.25 \\ 0.10 & 0.20 & 0.60 \end{bmatrix} \begin{bmatrix} 0.50 \\ 0.30 \\ 0.20 \end{bmatrix} = \begin{bmatrix} 0.47 \\ 0.30 \\ 0.23 \end{bmatrix}.$$

Thus, the model predicts a 47% probability of a bull market, a 30% probability of a stable market, and a 23% probability of a bear market one month from now.

Example 8.1.3

A market consists of three competing water bottle brands, represented by the states B_1, B_2, and B_3. The transition matrix

$$P = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0 & 0.4 \end{bmatrix}$$

describes how customers switch brands from one year to the next. The entry $P_{i,j}$ is the probability that a customer who currently uses brand B_j will use brand B_i the following year.

Consider the two initial distributions

$$\mathbf{x}_0^{(1)} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}_0^{(2)} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

For each initial distribution:

(a) Compute the state vectors $\mathbf{x}_k = P^k \mathbf{x}_0$ for $k = 0, 1, \ldots, 15$.

(b) Plot the proportion of customers using each brand as a function of k.

(c) Determine the steady-state distribution and compare the long-term behavior of the two systems.

Solution

(a) For the first initial distribution,

$$\mathbf{x}_1^{(1)} = P\mathbf{x}_0^{(1)} = \begin{bmatrix} 0.5 \\ 0.3 \\ 0.2 \end{bmatrix}, \quad \mathbf{x}_2^{(1)} = P\mathbf{x}_1^{(1)} = \begin{bmatrix} 0.37 \\ 0.45 \\ 0.18 \end{bmatrix}, \quad \mathbf{x}_3^{(1)} = P\mathbf{x}_2^{(1)} = \begin{bmatrix} 0.329 \\ 0.525 \\ 0.146 \end{bmatrix}.$$

For the second initial distribution,

$$\mathbf{x}_1^{(2)} = P\mathbf{x}_0^{(2)} = \begin{bmatrix} 0.3 \\ 0.3 \\ 0.4 \end{bmatrix}, \quad \mathbf{x}_2^{(2)} = P\mathbf{x}_1^{(2)} = \begin{bmatrix} 0.33 \\ 0.45 \\ 0.22 \end{bmatrix}, \quad \mathbf{x}_3^{(2)} = P\mathbf{x}_2^{(2)} = \begin{bmatrix} 0.321 \\ 0.525 \\ 0.154 \end{bmatrix}.$$

Although the two systems begin in completely different states, their state vectors gradually approach the same distribution.

To determine this distribution directly, let

$$\mathbf{q} = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix}$$

be a steady-state vector. Then

$$P\mathbf{q} = \mathbf{q},$$

so

$$(P - I)\mathbf{q} = \mathbf{0}.$$

Because the entries of $\mathbf{q}$ represent proportions, they must also satisfy

$$q_1 + q_2 + q_3 = 1.$$

Solving the resulting system gives

$$\mathbf{q} = \begin{bmatrix} 0.3 \\ 0.6 \\ 0.1 \end{bmatrix}.$$

Therefore,

$$\lim_{k\to\infty} \mathbf{x}_k^{(1)} = \lim_{k\to\infty} \mathbf{x}_k^{(2)} = \begin{bmatrix} 0.3 \\ 0.6 \\ 0.1 \end{bmatrix}.$$

In the long run, approximately 30% of customers use brand B_1, 60% use brand B_2, and 10% use brand B_3. The plots in Figure 47 show that the initial distribution affects the short-term path of the system, but not its eventual steady state.

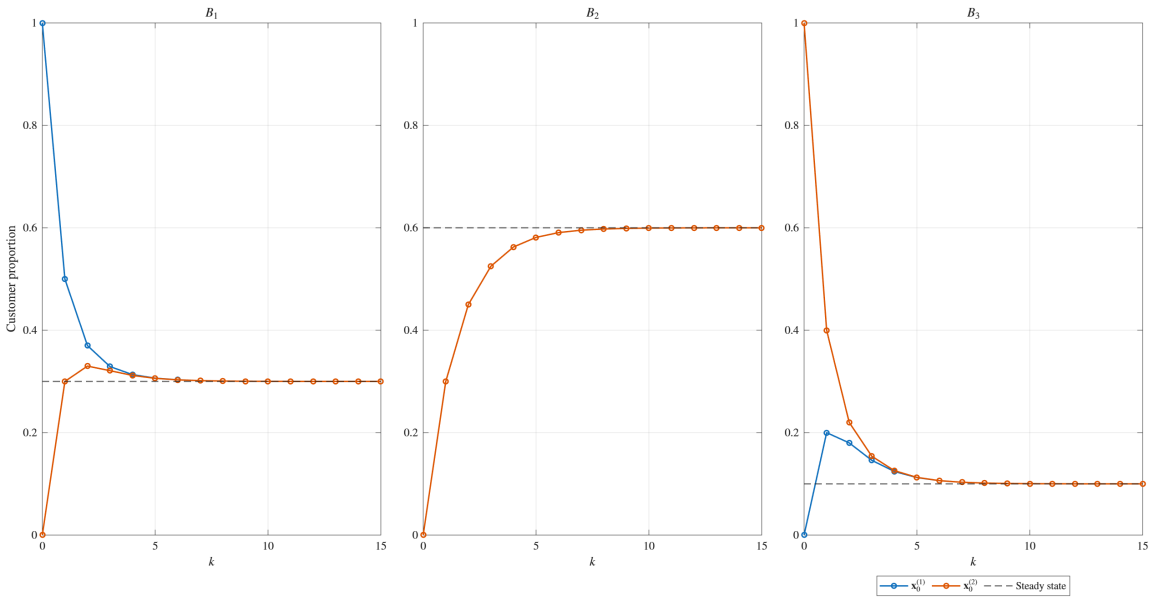


Figure 47: Evolution of the customer proportions for two different initial distributions. Although the systems begin in different states, both converge to the steady-state distribution $\mathbf{q} = \begin{bmatrix} 0.3 & 0.6 & 0.1 \end{bmatrix}^T$.

(a) Image credit: Owala

(b) Image credit: Hydroflask

(c) Image credit: Stanley

Figure 48: Which is which?

8.2 Robotics

A robotic arm consists of a sequence of rigid links connected by joints. Each joint contributes one or more dgrees of freedom, which determine how the links may move relative to one another. The

final link typically carries a gripper, welding tool, camera, or other device known as the end effector.

The position of the end effector depends on the joint variables of the robot. For a robot with n revolute joints, these variables may be collected into the joint configuration vector

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}.$$

The end-effector position is then described by a generally nonlinear function

$$\mathbf{x} = \mathbf{f}(\boldsymbol{\theta}).$$

Although the relationship between joint angles and end-effector position is nonlinear, the relationship between their instantaneous velocities is linear. This local linear transformation is described by the Jacobian.

Differentiating $\mathbf{x} = \mathbf{f}(\boldsymbol{\theta})$ with respect to time gives

$$\dot{\mathbf{x}} = J(\boldsymbol{\theta})\dot{\boldsymbol{\theta}},$$

where

$$J(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial f_1}{\partial \theta_1} & \frac{\partial f_1}{\partial \theta_2} & \cdots & \frac{\partial f_1}{\partial \theta_n} \\ \frac{\partial f_2}{\partial \theta_1} & \frac{\partial f_2}{\partial \theta_2} & \cdots & \frac{\partial f_2}{\partial \theta_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial \theta_1} & \frac{\partial f_m}{\partial \theta_2} & \cdots & \frac{\partial f_m}{\partial \theta_n} \end{bmatrix}$$

is the Jacobian of $\mathbf{f}$.

Jacobian Velocity Map

Let $\dot{\boldsymbol{\theta}} \in \mathbb{R}^n$ be the vector of joint velocities and let $\dot{\mathbf{x}} \in \mathbb{R}^m$ be the velocity of the end effector. The Jacobian matrix $J(\boldsymbol{\theta})$ defines the linear transformation

$$\dot{\mathbf{x}} = J(\boldsymbol{\theta})\dot{\boldsymbol{\theta}}.$$

For a fixed configuration $\boldsymbol{\theta}$, each column of $J(\boldsymbol{\theta})$ describes the end-effector velocity produced by unit motion of one joint while all other joints remain stationary.

Consider a planar robot consisting of two links with lengths L_1 and L_2 as shown in Figure 49.

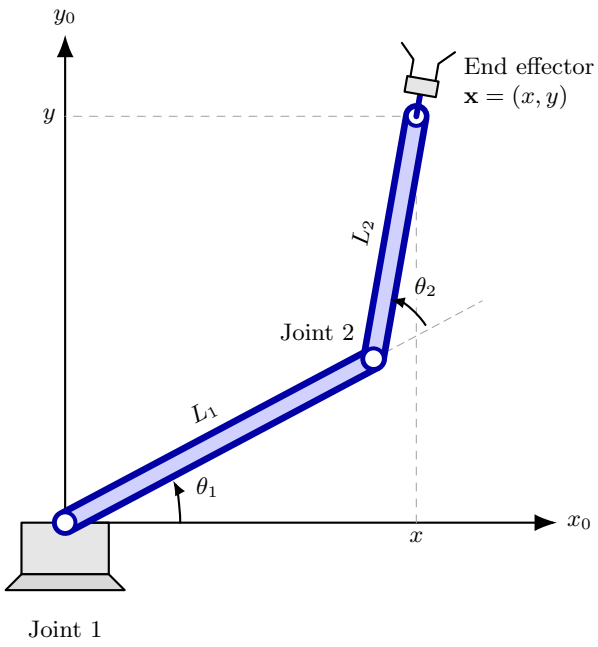


Figure 49: Geometry of a two-link planar robot. The joint angles θ_1 and θ_2 determine the Cartesian position $\mathbf{x} = (x, y)$ of the end effector.

The first joint rotates through the angle θ_1, while the second joint rotates through the angle θ_2 relative to the first link. The end-effector coordinates are

$$x = L_1 \cos\theta_1 + L_2 \cos(\theta_1 + \theta_2)$$

and

$$y = L_1 \sin\theta_1 + L_2 \sin(\theta_1 + \theta_2).$$

The position vector is therefore

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} L_1 \cos\theta_1 + L_2 \cos(\theta_1 + \theta_2) \\ L_1 \sin\theta_1 + L_2 \sin(\theta_1 + \theta_2) \end{bmatrix}.$$

Taking the partial derivatives with respect to θ_1 and θ_2 gives

$$J(\boldsymbol{\theta}) = \begin{bmatrix} -L_1 \sin\theta_1 - L_2 \sin(\theta_1 + \theta_2) & -L_2 \sin(\theta_1 + \theta_2) \\ L_1 \cos\theta_1 + L_2 \cos(\theta_1 + \theta_2) & L_2 \cos(\theta_1 + \theta_2) \end{bmatrix}.$$

Thus,

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} -L_1 \sin\theta_1 - L_2 \sin(\theta_1 + \theta_2) & -L_2 \sin(\theta_1 + \theta_2) \\ L_1 \cos\theta_1 + L_2 \cos(\theta_1 + \theta_2) & L_2 \cos(\theta_1 + \theta_2) \end{bmatrix} \begin{bmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \end{bmatrix}.$$

The first column describes the end-effector velocity generated by rotating the first joint, while the second column describes the velocity generated by rotating the second joint. The total velocity is the linear combination of these two column vectors determined by $\dot{\theta}_1$ and $\dot{\theta}_2$.

Example 8.2.1

A two-link planar robot has link lengths $L_1 = 1.2$ m and $L_2 = 0.8$ m. At a particular instant,

$$\theta_1 = 30^\circ, \quad \theta_2 = 60^\circ, \quad \dot{\boldsymbol{\theta}} = \begin{bmatrix} 0.50 \\ -0.25 \end{bmatrix} \text{ rad/s}.$$

Determine the instantaneous velocity of the end effector.

Solution

Because $\theta_1 + \theta_2 = 90^\circ$,

$$J(\boldsymbol{\theta}) = \begin{bmatrix} -1.2\sin 30^\circ - 0.8\sin 90^\circ & -0.8\sin 90^\circ \\ 1.2\cos 30^\circ + 0.8\cos 90^\circ & 0.8\cos 90^\circ \end{bmatrix}.$$

Therefore,

$$J(\boldsymbol{\theta}) = \begin{bmatrix} -1.4000 & -0.8000 \\ 1.0392 & 0 \end{bmatrix}.$$

The end-effector velocity is

$$\dot{\mathbf{x}} = J(\boldsymbol{\theta})\dot{\boldsymbol{\theta}} = \begin{bmatrix} -1.4000 & -0.8000 \\ 1.0392 & 0 \end{bmatrix} \begin{bmatrix} 0.50 \\ -0.25 \end{bmatrix} = \begin{bmatrix} -0.5000 \\ 0.5196 \end{bmatrix} \text{ m/s}.$$

Thus, the end effector is moving to the left at 0.5000 m/s and upward at approximately 0.5196 m/s.

In many robotic tasks, the desired end-effector velocity is known and the required joint velocities must be determined. This is the inverse problem

$$J(\boldsymbol{\theta})\dot{\boldsymbol{\theta}} = \dot{\mathbf{x}}.$$

When $J(\boldsymbol{\theta})$ is square and invertible,

$$\dot{\boldsymbol{\theta}} = J(\boldsymbol{\theta})^{-1}\dot{\mathbf{x}}.$$

This equation allows a controller to convert a desired motion of the gripper or tool into the joint velocities required to produce that motion. However, the inverse exists only when the Jacobian has full rank.

A configuration at which the Jacobian loses rank is called a *singular configuration*. At such a configuration, the robot loses the ability to produce motion in one or more directions.

For the two-link planar robot,

$$\det J(\boldsymbol{\theta}) = L_1 L_2 \sin\theta_2.$$

Since the link lengths are nonzero, the Jacobian is singular when $\sin\theta_2 = 0$.

Therefore,

$$\theta_2 = 0 \quad \text{or} \quad \theta_2 = \pi$$

within one complete revolution. These correspond to the arm being completely stretched out or folded back upon itself.

When $\theta_2 = 0$, the Jacobian becomes

$$J(\theta_1, 0) = \begin{bmatrix} -(L_1 + L_2)\sin\theta_1 & -L_2\sin\theta_1 \\ (L_1 + L_2)\cos\theta_1 & L_2\cos\theta_1 \end{bmatrix}.$$

The two columns are scalar multiples of the same vector:

$$\begin{bmatrix} -(L_1 + L_2)\sin\theta_1 \\ (L_1 + L_2)\cos\theta_1 \end{bmatrix} = \frac{L_1 + L_2}{L_2} \begin{bmatrix} -L_2\sin\theta_1 \\ L_2\cos\theta_1 \end{bmatrix}.$$

The columns are therefore linearly dependent, so

$$\operatorname{rank} J(\theta_1, 0) = 1.$$

Geometrically, both joint rotations produce instantaneous motion tangent to the circle centered at the robot base. No combination of joint velocities can produce instantaneous motion along the radial direction of the arm. The two-dimensional end-effector velocity space has collapsed to a one-dimensional subspace.

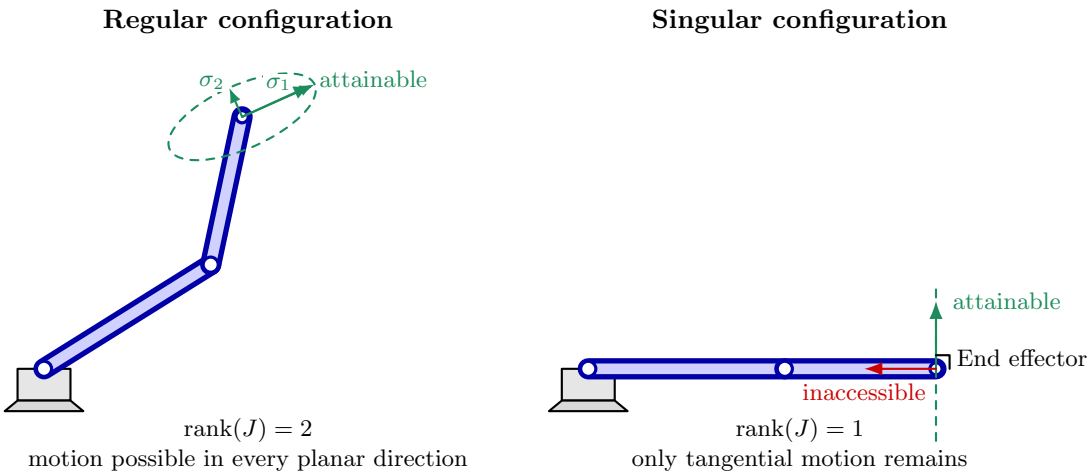


Figure 50: The Jacobian maps joint velocities to end-effector velocities. In a regular configuration, the attainable end-effector velocities form an ellipse, indicating motion in every planar direction. At a singular configuration, this ellipse collapses to a line, so the robot loses one instantaneous degree of freedom and radial motion becomes impossible.

For example, when $\theta_1 = 0$ and $\theta_2 = 0$,

$$J(0,0) = \begin{bmatrix} 0 & 0 \\ L_1 + L_2 & L_2 \end{bmatrix}.$$

Hence,

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 \\ (L_1 + L_2)\dot{\theta}_1 + L_2\dot{\theta}_2 \end{bmatrix}.$$

The horizontal velocity $\dot{x}$ is zero for every possible joint velocity vector. At that instant, the end effector can move vertically but cannot move farther outward or inward.

The singular configuration also creates a nontrivial null space. Solving

$$J(0,0)\dot{\boldsymbol{\theta}} = \mathbf{0}$$

gives

$$(L_1 + L_2)\dot{\theta}_1 + L_2\dot{\theta}_2 = 0.$$

Therefore,

$$\operatorname{Nul} J(0,0) = \operatorname{Span}\left\{\begin{bmatrix} L_2 \\ -(L_1 + L_2) \end{bmatrix}\right\}.$$

A joint velocity vector in this null space produces no instantaneous end-effector motion. The joints may move internally while their effects at the end effector cancel one another.

Example 8.2.2

Let $L_1 = L_2 = 1$ m, $\theta_1 = 0$, and suppose the robot is near its stretched configuration. The desired end-effector velocity is

$$\dot{\mathbf{x}} = \begin{bmatrix} 0.10 \\ 0 \end{bmatrix} \text{ m/s}.$$

Determine how the required joint velocities behave as θ_2 approaches 0.

Solution

With $L_1 = L_2 = 1$ and $\theta_1 = 0$,

$$J(0,\theta_2) = \begin{bmatrix} -\sin\theta_2 & -\sin\theta_2 \\ 1 + \cos\theta_2 & \cos\theta_2 \end{bmatrix}.$$

Since $\det J(0,\theta_2) = \sin\theta_2$,

the inverse exists whenever $\theta_2 \neq k\pi$. Solving

$$J(0,\theta_2)\begin{bmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \end{bmatrix} = \begin{bmatrix} 0.10 \\ 0 \end{bmatrix}$$

gives $\dot{\theta}_1 = 0.10\frac{\cos\theta_2}{\sin\theta_2}$ and $\dot{\theta}_2 = -0.10\frac{1+\cos\theta_2}{\sin\theta_2}$. Several values are shown below.

θ_2	$\dot{\theta}_1$	$\dot{\theta}_2$
10°	0.5671	-1.1430
5°	1.1430	-2.2904
2°	2.8636	-5.7290
1°	5.7290	-11.4589

As θ_2 approaches 0, the required joint velocities increase without bound. At $\theta_2 = 0$, horizontal motion is impossible and the inverse Jacobian no longer exists.

A physical robot cannot generate infinite joint velocity. Motor limits prevent the commanded motion from being followed, while the rapidly increasing joint velocities and accelerations may place severe loads on motors, gears, shafts, and links. Robotic trajectories are therefore commonly designed to avoid singular configurations.

The singular value decomposition provides a general way to measure proximity to a singularity. If

$$J = U\Sigma V^T$$

has singular values

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0, \quad r = \min(m, n),$$

then the Jacobian loses rank when $\sigma_r = 0$. When σ_r is small but nonzero, the robot is near a singular configuration and motion in the corresponding task-space direction requires large joint velocities. The ratio

$$\kappa(J) = \frac{\sigma_1}{\sigma_r}$$

measures the sensitivity of the inverse velocity problem. A large condition number indicates that small changes in the desired end-effector velocity may produce large changes in the required joint velocities.

The Jacobian also relates forces at the end effector to torques at the joints. This is known as force mapping. Let

$$\mathbf{F} = \begin{bmatrix} F_x \\ F_y \end{bmatrix}$$

be a force applied to the end effector, and let

$$\boldsymbol{\tau} = \begin{bmatrix} \tau_1 \\ \\ \tau_2 \end{bmatrix}$$

be the corresponding joint torque vector. Equality of mechanical power gives

$$\mathbf{F}^T \dot{\mathbf{x}} = \boldsymbol{\tau}^T \dot{\boldsymbol{\theta}}.$$

Using $\dot{\mathbf{x}} = J\dot{\boldsymbol{\theta}}$,

$$\mathbf{F}^T J\dot{\boldsymbol{\theta}} = \boldsymbol{\tau}^T \dot{\boldsymbol{\theta}}.$$

Since this must hold for every joint velocity vector,

$$\boldsymbol{\tau} = J(\boldsymbol{\theta})^T \mathbf{F}.$$

At the stretched configuration $\theta_1 = \theta_2 = 0$,

$$J(0,0)^T = \begin{bmatrix} 0 & L_1 + L_2 \\ 0 & L_2 \end{bmatrix}.$$

A radial force

$$\mathbf{F} = \begin{bmatrix} F \\ \\ 0 \end{bmatrix}$$

produces

$$\boldsymbol{\tau} = J(0,0)^T \mathbf{F} = \begin{bmatrix} 0 \\ \\ 0 \end{bmatrix}.$$

The external force produces no torque at either joint. Instead, the load is carried directly through the aligned links of the robot. In linear-algebraic language,

$$\mathbf{F} \in \operatorname{Nul}\left(J^T\right).$$

By contrast, a tangential force produces nonzero joint torques. For $L_1 = 1.2$ m, $L_2 = 0.8$ m, and

$$\mathbf{F} = \begin{bmatrix} 0 \\ 100 \end{bmatrix} \text{ N},$$

the required torques are

$$\boldsymbol{\tau} = \begin{bmatrix} 0 & 2.0 \\ 0 & 0.8 \end{bmatrix} \begin{bmatrix} 0 \\ 100 \end{bmatrix} = \begin{bmatrix} 200 \\ 80 \end{bmatrix} \text{ N} \cdot \text{m}.$$

The same singularity therefore has two complementary interpretations. The column space of J describes the end-effector velocities that the joints can generate, while the null space of J^T describes the external forces that produce no joint torque.

Problem 8.2.1

A two-link planar inspection robot has link lengths $L_1 = 0.6$ m and $L_2 = 0.4$ m. At a particular instant,

$$\theta_1 = 45^\circ, \quad \theta_2 = 90^\circ, \quad \dot{\mathbf{x}} = \begin{bmatrix} 0.10 \\ -0.05 \end{bmatrix} \text{ m/s}.$$

(a) Evaluate the Jacobian matrix at the given configuration.

(b) Determine the joint velocity vector required to produce the desired end-effector velocity.

(c)

(d) Determine all values of θ_2 for which the Jacobian is singular.

(e) At the stretched configuration $\theta_1 = \theta_2 = 0$, determine $\operatorname{Col}(J)$, $\operatorname{Nul}(J)$, and $\operatorname{Nul}(J^T)$. Interpret each subspace physically.

Solution

(a) Since $\theta_1 + \theta_2 = 135^\circ$,

$$J(\boldsymbol{\theta}) = \begin{bmatrix} -0.6\sin 45^\circ - 0.4\sin 135^\circ & -0.4\sin 135^\circ \\ 0.6\cos 45^\circ + 0.4\cos 135^\circ & 0.4\cos 135^\circ \end{bmatrix}.$$

Therefore,

$$J(\boldsymbol{\theta}) = \begin{bmatrix} -0.7071 & -0.2828 \\ 0.1414 & -0.2828 \end{bmatrix}.$$

(b) The joint velocity vector satisfies

$$J(\boldsymbol{\theta})\dot{\boldsymbol{\theta}} = \dot{\mathbf{x}}.$$

Since

$$\det J(\boldsymbol{\theta}) = L_1 L_2 \sin\theta_2 = (0.6)(0.4)\sin 90^\circ = 0.24,$$

the Jacobian is invertible. Thus,

$$\dot{\boldsymbol{\theta}} = J(\boldsymbol{\theta})^{-1}\dot{\mathbf{x}} = \begin{bmatrix} -0.1768 \\ 0.0884 \end{bmatrix} \text{ rad/s}.$$

(c) The Jacobian is singular when

$$L_1 L_2 \sin\theta_2 = 0.$$

Since L_1 and L_2 are nonzero,

$$\sin\theta_2 = 0,$$

so

$$\theta_2 = k\pi, \quad k \in \mathbb{Z}.$$

These configurations occur whenever the two links are collinear.

(d) At $\theta_1 = \theta_2 = 0$,

$$J(0,0) = \begin{bmatrix} 0 & 0 \\ 1.0 & 0.4 \end{bmatrix}.$$

The column space is

$$\mathrm{Col}(J) = \mathrm{Span}\left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}.$$

Thus, the end effector can move only vertically at that instant.

To determine the null space, solve

$$\begin{bmatrix} 0 & 0 \\ 1.0 & 0.4 \end{bmatrix} \begin{bmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \end{bmatrix} = \mathbf{0}.$$

This gives

$$\dot{\theta}_1 + 0.4\dot{\theta}_2 = 0,$$

so

$$\mathrm{Nul}(J) = \mathrm{Span}\left\{ \begin{bmatrix} -0.4 \\ 1 \end{bmatrix} \right\}.$$

Joint velocities in this direction cancel one another and produce no instantaneous end-effector motion.

Finally,

$$J^T = \begin{bmatrix} 0 & 1.0 \\ 0 & 0.4 \end{bmatrix}.$$

Solving $J^T\mathbf{F} = \mathbf{0}$ gives $F_y = 0$, while F_x is free. Therefore,

$$\text{Nul}(J^T) = \text{Span}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}.$$

This is the radial force direction. A horizontal load applied along the stretched links produces no joint torque and is instead carried by the structure of the robot.

Robotic singularities are therefore not merely failures of a matrix inverse. They describe physical configurations in which available motion disappears, internal joint motion becomes invisible at the end effector, required joint speeds become extreme, and certain external forces pass directly through the structure. Rank, column spaces, null spaces, singular values, and matrix transposes all contain physical meaning.

This section was adapted from [14].

8.3 Chemical Reactors

Chemical reactors convert incoming chemical species into new products through careful analysis of heat and mass transfer. One of the most common reactor models is the **continuous stirred-tank reactor**, or CSTR. A CSTR is assumed to be perfectly mixed, so the concentration of each species in the outlet stream is identical to its concentration inside the reactor.

For any species, the material balance takes the form

$$\text{accumulation} = \text{inflow} - \text{outflow} + \text{generation by reaction}.$$

Consider a constant-volume CSTR with reactor volume V, volumetric flow rate Q, feed concentration $c_{j,f}$, and reactor concentration c_j. The material balance for species j is

$$V\frac{dc_j}{dt} = Qc_{j,f} - Qc_j + VR_j,$$

where R_j is the net rate at which species j is produced by chemical reaction.

At steady state, the concentrations no longer change with time, so

$$0 = Qc_{j,f} - Qc_j + VR_j.$$

Defining the residence time

$$\tau = \frac{V}{Q},$$

the steady-state balance becomes

$$0 = c_{j,f} - c_j + \tau R_j.$$

When the reaction rates are first-order, the production rates are linear combinations of the species concentrations. The complete set of reactor balances may therefore be written as a linear system.

Consider the consecutive reactions

$$A \xrightarrow{k_1} C \xrightarrow{k_2} D,$$

with reaction rates

$$r_1 = k_1 c_A, \quad r_2 = k_2 c_C.$$

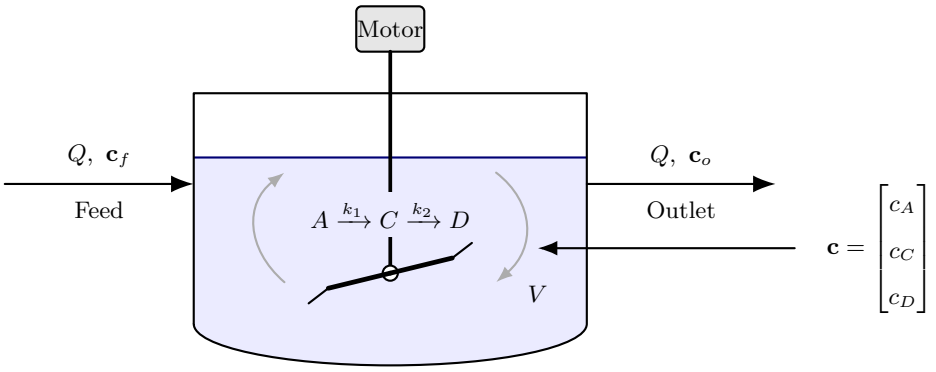


Figure 51: A continuous stirred-tank reactor carrying out the consecutive reactions $A \xrightarrow{k_1} C \xrightarrow{k_2} D$. Perfect mixing makes the outlet concentration vector $\mathbf{c}_o$ identical to the concentration vector $\mathbf{c}$ inside the reactor.

Because the reactor is perfectly mixed, the outlet concentration vector satisfies $\mathbf{c}_o = \mathbf{c}$.

Species A is consumed by the first reaction, species C is produced by the first reaction and consumed by the second, and species D is produced by the second reaction. The species production rates are therefore

$$R_A = -k_1 c_A, \quad R_C = k_1 c_A - k_2 c_C, \quad R_D = k_2 c_C.$$

Collecting the reactor concentrations into the vector

$$\mathbf{c} = \begin{bmatrix} c_A \\ c_C \\ c_D \end{bmatrix},$$

the production-rate vector may be written as

$$\mathbf{R} = \begin{bmatrix} -k_1 & 0 & 0 \\ k_1 & -k_2 & 0 \\ 0 & k_2 & 0 \end{bmatrix} \mathbf{c}.$$

Let

$$K = \begin{bmatrix} -k_1 & 0 & 0 \\ k_1 & -k_2 & 0 \\ 0 & k_2 & 0 \end{bmatrix}.$$

The steady-state reactor balance is then $\mathbf{0} = \mathbf{c}_f - \mathbf{c} + \tau K \mathbf{c}$, so

$$(I - \tau K)\mathbf{c} = \mathbf{c}_f.$$

For this reaction network,

$$\begin{bmatrix} 1 + \tau k_1 & 0 & 0 \\ -\tau k_1 & 1 + \tau k_2 & 0 \\ 0 & -\tau k_2 & 1 \end{bmatrix} \begin{bmatrix} c_A \\ c_C \\ c_D \end{bmatrix} = \begin{bmatrix} c_{A,f} \\ c_{C,f} \\ c_{D,f} \end{bmatrix}.$$

Thus, the reactor converts the feed concentration vector into the reactor concentration vector through a linear transformation. Since $\mathbf{c}_o = \mathbf{c}$, this vector also gives the outlet concentrations. If the coefficient matrix is denoted by M, then

$$M\mathbf{c} = \mathbf{c}_f,$$

and therefore

$$\mathbf{c} = M^{-1}\mathbf{c}_f.$$

In practice, the system would be solved directly rather than by explicitly computing M^{-1}.

Example 8.3.1

A CSTR carries out the consecutive reactions

$$A \xrightarrow{k_1} C \xrightarrow{k_2} D.$$

The residence time and reaction-rate constants are

$$\tau = 2.5 \text{ min}, \quad k_1 = 2.3 \text{ min}^{-1}, \quad k_2 = 4.5 \text{ min}^{-1}.$$

The feed contains only species A:

$$\mathbf{c}_f = \begin{bmatrix} 2.2 \\ 0 \\ 0 \end{bmatrix} \text{ mol/L}.$$

Determine the steady-state outlet concentrations.

Solution

The steady-state system is

$$\begin{bmatrix} 1+\tau k_1 & 0 & 0 \\ -\tau k_1 & 1+\tau k_2 & 0 \\ 0 & -\tau k_2 & 1 \end{bmatrix} \begin{bmatrix} c_A \\ c_C \\ c_D \end{bmatrix} = \begin{bmatrix} 2.2 \\ 0 \\ 0 \end{bmatrix}.$$

Substituting the parameter values gives

$$\begin{bmatrix} 6.75 & 0 & 0 \\ -5.75 & 12.25 & 0 \\ 0 & -11.25 & 1 \end{bmatrix} \begin{bmatrix} c_A \\ c_C \\ c_D \end{bmatrix} = \begin{bmatrix} 2.2 \\ 0 \\ 0 \end{bmatrix}.$$

Because the coefficient matrix is lower triangular, the concentrations may be found sequentially.

The first equation gives

$$6.75c_A = 2.2 \Rightarrow c_A = 0.3259 \text{ mol/L}.$$

The second equation gives

$$-5.75c_A + 12.25c_C = 0 \Rightarrow c_C = \frac{5.75c_A}{12.25} = 0.1530 \text{ mol/L}$$

Finally,

$$-11.25c_C + c_D = 0 \Rightarrow c_D = 11.25c_C = 1.7211 \text{ mol/L}.$$

Thus, the steady-state outlet concentration vector is

$$\mathbf{c} = \begin{bmatrix} 0.3259 \\ 0.1530 \\ 1.7211 \end{bmatrix} \text{ mol/L}.$$

As a physical check,

$$c_A + c_C + c_D = 2.2 \text{ mol/L},$$

which agrees with the total concentration entering the reactor. Each reaction converts one mole of one species into one mole of another, so the total molar concentration is conserved.

Linear algebra is also useful before the reactor balances are even written. A large reaction mechanism may contain reactions that are stoichiometrically redundant. The **stoichiometric matrix** reveals how many independent reaction directions are actually present.

Consider the reactions

$$\begin{aligned} A + B + 2C &\longrightarrow D + E, \\ A &\longrightarrow B + E, \\ 2B + 2C &\longrightarrow D. \end{aligned}$$

Using the species order

$$A, \quad B, \quad C, \quad D, \quad E,$$

each reaction is represented by a stoichiometric vector. Reactants receive negative coefficients and products receive positive coefficients:

$$\boldsymbol{\nu}_1 = \begin{bmatrix} -1 & -1 & -2 & 1 & 1 \end{bmatrix}, \quad \boldsymbol{\nu}_2 = \begin{bmatrix} -1 & 1 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\nu}_3 = \begin{bmatrix} 0 & -2 & -2 & 1 & 0 \end{bmatrix}$$

Placing these vectors into the rows of a matrix gives

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

The first row satisfies $\boldsymbol{\nu}_1 = \boldsymbol{\nu}_2 + \boldsymbol{\nu}_3$. Therefore, the rows of N are linearly dependent and $\operatorname{rank}(N) = 2$.

Although three reactions were listed, only two are stoichiometrically independent. The first reaction is obtained by adding the second and third reactions together.

If the reaction-rate vector is

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix},$$

then the net species-production vector is

$$\mathbf{R} = N^T \mathbf{r}.$$

The rank of N determines the dimension of the space of possible species changes. A dependent reaction does not introduce a new stoichiometric direction, even though it may still be useful when describing the physical reaction mechanism or its kinetics.

Chemical reactor models therefore draw on several central ideas from linear algebra. Matrix equations determine steady-state concentrations, linear transformations map feed compositions to outlet compositions, and the rank of the stoichiometric matrix reveals the number of independent reactions hidden inside a larger mechanism.

This section was adapted from [6, 12, 16].

8.4 Quantum Mechanics

Quantum mechanics is fundamentally a linear theory. The state vectors of a physical system belong to a complex inner product space—more precisely, a Hilbert space—while observables are represented by Hermitian linear operators and reversible dynamics by unitary operators.

In a finite-dimensional system, a quantum state is represented by a unit vector written in Dirac notation as a **ket** $|\psi\rangle$. Relative to an orthonormal basis $\{|1\rangle, \dots, |n\rangle\}$,

$$|\psi\rangle = c_1|1\rangle + \cdots + c_n|n\rangle,$$

where the coefficients $c_1, \dots, c_n$ may be complex. In coordinate form,

$$|\psi\rangle = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

A physical state must be normalized, so

$$\langle \psi \mid \psi \rangle = \sum_{j=1}^{n} |c_j|^2 = 1.$$

The corresponding **bra** is the conjugate transpose of the ket:

$$\langle \psi | = \begin{bmatrix} c_1^* & \cdots & c_n^* \end{bmatrix}.$$

Thus, the inner product between two states is ordinary complex matrix multiplication:

$$\langle \phi \mid \psi \rangle = \begin{bmatrix} \phi_1^* & \cdots & \phi_n^* \end{bmatrix} \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_n \end{bmatrix}.$$

Orthogonal state vectors represent mutually exclusive outcomes of a projective measurement. If $\langle \phi \mid \psi \rangle = 0$, then a system in the state $|\psi\rangle$ has zero probability of being measured in the state $|\phi\rangle$.

The simplest quantum system is a **qubit**, whose state space is two-dimensional. Its standard computational basis is

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

A general qubit state is a linear combination

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \begin{bmatrix} \alpha \\ \beta \end{bmatrix},$$

where $|\alpha|^2 + |\beta|^2 = 1$.

This linear combination is called a **superposition**. Unlike a classical bit, which is either 0 or 1, a qubit may contain components in both basis directions before it is measured.

Suppose $\{|\phi_1\rangle, \ldots, |\phi_n\rangle\}$ is an orthonormal measurement basis. Born's rule states that the probability of measuring the state $|\phi_j\rangle$ is

$$P(\phi_j) = |\langle \phi_j \mid \psi \rangle|^2.$$

The inner product extracts the component of $|\psi\rangle$ in the direction $|\phi_j\rangle$, and its squared magnitude converts that component into a probability. After the measurement, the state collapses to the measured basis vector.

Equivalently, measurement may be described using the projection matrix

$$P_j = |\phi_j\rangle\langle\phi_j|.$$

Applying P_j to the state isolates its component in the direction $|\phi_j\rangle$:

$$P_j|\psi\rangle = |\phi_j\rangle\langle\phi_j \mid \psi\rangle.$$

Physical quantities such as energy, momentum, and spin are represented by Hermitian operators. A matrix A is Hermitian when $A^\dagger = A$.

The eigenvalue equation

$$A|a_j\rangle = a_j|a_j\rangle$$

has a direct physical interpretation. An eigenvector is tied to a measurable state and an eigenvalue is tied to a possible measurement result.

Hermitian matrices are essential because their eigenvalues are real, and eigenvectors corresponding to distinct eigenvalues are orthogonal. By the spectral theorem, every Hermitian matrix admits an orthonormal basis of eigenvectors and is therefore unitarily diagonalizable.

For a normalized state $|\psi\rangle$, the expected value of the observable A is

$$\langle A\rangle_\psi = \langle\psi|A|\psi\rangle.$$

Consider the observable

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Its eigenvectors are $|0\rangle$ and $|1\rangle$, with eigenvalues 1 and -1, respectively. Suppose the system is in the state

$$|\psi\rangle = \frac{\sqrt{3}}{2}|0\rangle + \frac{i}{2}|1\rangle = \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{i}{2} \end{bmatrix}.$$

The state is normalized because

$$\langle \psi \mid \psi \rangle = \left|\frac{\sqrt{3}}{2}\right|^2 + \left|\frac{i}{2}\right|^2 = \frac{3}{4} + \frac{1}{4} = 1.$$

Measuring Z gives the result 1 with probability

$$P(1) = |\langle 0 \mid \psi \rangle|^2 = \frac{3}{4},$$

and the result -1 with probability

$$P(-1) = |\langle 1 \mid \psi \rangle|^2 = \frac{1}{4}.$$

Therefore, the expected measurement is

$$\langle Z \rangle_\psi = (1)\left(\frac{3}{4}\right) + (-1)\left(\frac{1}{4}\right) = \frac{1}{2}.$$

Reversible dynamics of an isolated quantum system, including quantum gates, are represented by unitary matrices, which satisfy $U^\dagger U = I$.

Indeed, for any states $|\phi\rangle$ and $|\psi\rangle$,

$$\langle U\phi \mid U\psi \rangle = \langle \phi | U^\dagger U | \psi \rangle = \langle \phi \mid \psi \rangle.$$

Thus, unitary transformations preserve all inner products and, in particular, preserve the norms of quantum states.

For example, the Hadamard matrix

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

transforms the basis state $|0\rangle$ into an equal superposition:

$$H|0\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}1\\ 0\end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\ 1\end{bmatrix} = \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right).$$

The eigenvectors of a Hermitian operator provide the natural basis for describing a measurement. Suppose A has an orthonormal eigenbasis

$$A|a_n\rangle = a_n|a_n\rangle.$$

Any normalized state may be expanded in this basis as

$$|\psi\rangle = \sum_n c_n|a_n\rangle, \quad c_n = \langle a_n \mid \psi\rangle.$$

Because the basis is orthonormal,

$$\langle\psi \mid \psi\rangle = \sum_n |c_n|^2 = 1.$$

The quantity $|c_n|^2$ is therefore the probability of measuring the eigenvalue a_n. The basis vectors also satisfy the completeness relation

$$I = \sum_n |a_n\rangle\langle a_n|,$$

since

$$I|\psi\rangle = \sum_n |a_n\rangle\langle a_n \mid \psi\rangle = \sum_n c_n|a_n\rangle = |\psi\rangle.$$

Using the same basis, the operator itself may be written in its spectral form:

$$A = \sum_n a_n|a_n\rangle\langle a_n|$$

The expected value of A is then

$$\langle A\rangle_\psi = \langle\psi|A|\psi\rangle = \sum_n |c_n|^2 a_n.$$

Thus, the expectation value is a weighted average of the possible measurement outcomes, with weights determined by the squared magnitudes of the coordinates of $|\psi\rangle$ in the eigenbasis of A.

This is the central linear-algebraic structure of quantum measurement: diagonalizing a Hermitian operator reveals its possible outcomes, while projecting the state onto its eigenvectors determines the probability of each outcome.

Problem 8.4.1

Consider the normalized qubit state

$$|\psi\rangle = \frac{1}{\sqrt{5}}\left(2|0\rangle + i|1\rangle\right) = \frac{1}{\sqrt{5}}\begin{bmatrix}2\\ i\end{bmatrix}.$$

Let

$$Z = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}, \quad H = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}.$$

(a) Verify that $|\psi\rangle$ is normalized.

(b) If the observable Z is measured, determine the probability of obtaining each eigenvalue and compute the expected value $\langle Z\rangle_\psi$.

(c) Compute the transformed state $H|\psi\rangle$ and verify that it remains normalized.

Solution

(a) The corresponding bra is

$$\langle\psi| = \frac{1}{\sqrt{5}}\begin{bmatrix}2 & -i\end{bmatrix}.$$

Therefore,

$$\langle\psi \mid \psi\rangle = \frac{1}{5}\begin{bmatrix}2 & -i\end{bmatrix}\begin{bmatrix}2\\ i\end{bmatrix} = \frac{1}{5}(4+1) = 1.$$

Thus, $|\psi\rangle$ is normalized.

(b) The eigenvectors of Z are $|0\rangle$ and $|1\rangle$, with eigenvalues 1 and -1, respectively. By Born's rule,

$$P(1) = |\langle 0 \mid \psi\rangle|^2 = \left|\frac{2}{\sqrt{5}}\right|^2 = \frac{4}{5},$$

and

$$P(-1) = |\langle 1 \mid \psi\rangle|^2 = \left|\frac{i}{\sqrt{5}}\right|^2 = \frac{1}{5}.$$

The expected value is therefore

$$\langle Z\rangle_\psi = (1)\left(\frac{4}{5}\right) + (-1)\left(\frac{1}{5}\right) = \frac{3}{5}.$$

(c) Applying the Hadamard matrix gives

$$H|\psi\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\frac{1}{\sqrt{5}}\begin{bmatrix}2\\ i\end{bmatrix} = \frac{1}{\sqrt{10}}\begin{bmatrix}2+i\\ 2-i\end{bmatrix}.$$

The squared norm of the transformed state is

$$\|H|\psi\rangle\|^2 = \frac{1}{10}\left(|2+i|^2 + |2-i|^2\right) = \frac{1}{10}(5+5) = 1.$$

Thus, $H|\psi\rangle$ remains normalized, as expected because H is unitary.

Thus, the basic structure of quantum mechanics is built directly from complex vector spaces, orthonormal bases, inner products, projections, eigenvalues, and matrix transformations. Linear algebra provides the language in which quantum mechanics is written.

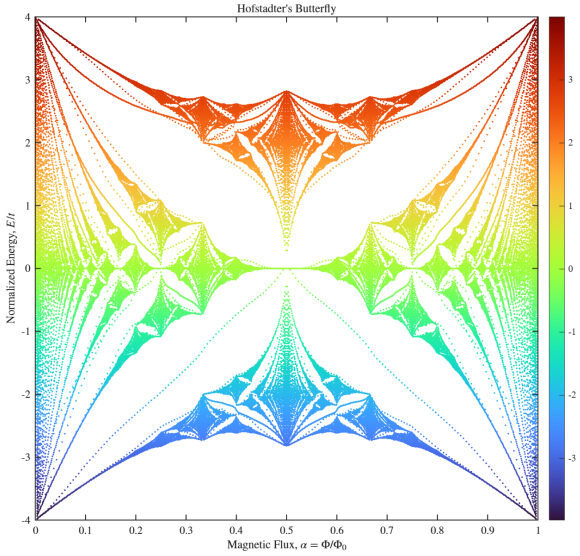


Figure 52: Hofstadter's butterfly, the fractal energy spectrum of the Harper–Hofstadter Hamiltonian. The horizontal axis gives the magnetic flux through each lattice cell in units of the magnetic flux quantum, while the vertical axis gives the normalized energy E/t. Each point corresponds to an allowed energy eigenvalue.

This section was adapted from [4, 8].

8.5 PageRank

Early search engines could identify pages containing a user's search terms, but they struggled to determine which of those pages were most important. Google's PageRank algorithm, created by Larry Page and Sergey Brin in the late 90s, addressed this problem by using the link structure of the web itself. A link from one page to another is treated as an endorsement, although the strength of that endorsement is divided among all pages linked by the original page. To say that PageRank revolutionized the internet forever would be an understatement.

Suppose a collection contains n web pages. If page j contains d_j outgoing links, define the hyperlink

matrix P by

$$P_{i,j} = \begin{cases} \dfrac{1}{d_j}, & \text{if page } j \text{ links to page } i, \\ 0, & \text{otherwise.} \end{cases}$$

Each column describes the probabilities of leaving one page and moving to another. Provided every page contains at least one outgoing link, the entries in each column sum to one, so P is column-stochastic.

PageRank may be interpreted using a **random surfer**. Starting from some page, the surfer repeatedly chooses one of the outgoing links at random. Pages visited more frequently in the long run are considered more important.

A steady-state probability vector $\mathbf{p}$ for this preliminary random-surfer model satisfies

$$P\mathbf{p} = \mathbf{p}, \quad \mathbf{1}^T\mathbf{p} = 1.$$

Thus, page importance is initially represented by an eigenvector of P associated with the eigenvalue 1. However, the hyperlink matrix does not always produce a unique or useful ranking.

First, a page may contain no outgoing links. Such a page is called a **dangling page**, and its corresponding column in P consists entirely of zeros. This column is replaced by the uniform probability vector

$$\frac{1}{n}\mathbf{1} = \frac{1}{n}\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Second, the web graph may contain disconnected groups or cycles that trap the surfer. PageRank resolves this problem by allowing the surfer to occasionally abandon the current link path and jump to any page at random.

Let $\alpha \in (0,1)$ denote the probability of making this random jump. Let $\widetilde{P}$ denote the resulting column-stochastic matrix. The **Google matrix** is

$$G = (1-\alpha)\widetilde{P} + \frac{\alpha}{n}\mathbf{1}\mathbf{1}^T.$$

The first term represents following an outgoing link, while the second represents jumping uniformly to any page. A commonly used interpretation is that the surfer follows a link with probability $1-\alpha$ and becomes bored and jumps elsewhere with probability α.

Every entry of G is at least $\alpha/n > 0$, so G is a positive stochastic matrix. The Perron–Frobenius theorem therefore guarantees a unique positive steady-state vector satisfying

$$G\mathbf{p} = \mathbf{p}, \quad \mathbf{1}^T\mathbf{p} = 1.$$

The components of $\mathbf{p}$ are the PageRank scores.

Example 8.5.1

Consider three pages A, B, and C with the following links:

- Page A links to pages B and C.
- Page B links only to page C.
- Page C links only to page A.

Using the page order A, B, C, construct the Google matrix with teleportation probability $\alpha = 0.15$, and determine the PageRank vector.

Solution

Page A divides its importance equally between B and C, while pages B and C pass all their importance to their single outgoing links. Therefore,

$$P = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 1 & 0 \end{bmatrix}.$$

There are no dangling pages, so $\widetilde{P} = P$. Since $n = 3$ and $\alpha = 0.15$,

$$G = 0.85P + \frac{0.15}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Thus,

$$G = \begin{bmatrix} 0.05 & 0.05 & 0.90 \\ 0.475 & 0.05 & 0.05 \\ 0.475 & 0.90 & 0.05 \end{bmatrix}.$$

The PageRank vector satisfies

$$G\mathbf{p} = \mathbf{p}, \qquad p_A + p_B + p_C = 1.$$

Solving gives

$$\mathbf{p} = \begin{bmatrix} 0.3878 \\ 0.2148 \\ 0.3974 \end{bmatrix}.$$

Therefore, the pages are ranked

$$C > A > B.$$

Although page A receives a link from page C, page C receives importance from both A and B, giving it the largest PageRank score.

For a small network, the PageRank vector may be found by solving the eigenvalue problem directly. For the actual web, however, the matrix is far too large for a full eigenvalue decomposition. Only the eigenvector associated with $\lambda = 1$ is needed, so PageRank is commonly computed using the **power method**.

Starting from a probability vector such as

$$\mathbf{p}_0 = \frac{1}{n}\mathbf{1},$$

repeatedly compute

$$\mathbf{p}_{k+1} = G\mathbf{p}_k.$$

Because G is positive and stochastic, the iterates converge to the unique PageRank vector:

$$\lim_{k\to\infty} \mathbf{p}_k = \mathbf{p}.$$

The iteration may be stopped when consecutive vectors differ by less than a chosen tolerance, such as

$$\|\mathbf{p}_{k+1} - \mathbf{p}_k\|_1 < \varepsilon.$$

Problem 8.5.1

Consider a small internet consisting of four pages A, B, C, and D, with links arranged as shown below.

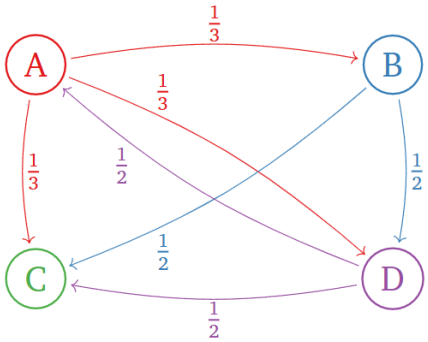


Figure 53: Image credit: [11]

Thus,

- page A links to pages B, C, and D;
- page B links to pages C and D;
- page C contains no outgoing links;
- page D links to pages A and C.

Using the page order A, B, C, D and teleportation probability $\alpha = 0.15$:

(a) Construct the hyperlink matrix P.

(b) Replace the dangling-page column to obtain the modified hyperlink matrix $\widetilde{P}$.

(c) Construct the Google matrix

$$G = (1 - \alpha)\widetilde{P} + \frac{\alpha}{4}\mathbf{1}\mathbf{1}^T.$$

(d) Determine the PageRank vector $\mathbf{p}$, normalized so that its entries sum to 1.

(e) Identify the most important and least important pages.

Solution

(a) Each column records how the page corresponding to that column distributes its importance. Therefore,

$$P = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}.$$

The third column consists entirely of zeros because page C has no outgoing links.

(b) Since C is a dangling page, its zero column is replaced by the uniform probability vector

$$\frac{1}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Thus,

$$\widetilde{P} = \begin{bmatrix} 0 & 0 & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{4} & 0 \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{4} & 0 \end{bmatrix}.$$

(c) Since $\alpha = 0.15$, the probability of following an ordinary link is $1 - \alpha = 0.85$. Therefore,

$$G = 0.85\widetilde{P} + 0.15 \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}.$$

Substituting $\widetilde{P}$ gives

$$G = \begin{bmatrix} 0.0375 & 0.0375 & 0.2500 & 0.4625 \\ 0.3208 & 0.0375 & 0.2500 & 0.0375 \\ 0.3208 & 0.4625 & 0.2500 & 0.4625 \\ 0.3208 & 0.4625 & 0.2500 & 0.0375 \end{bmatrix}.$$

Each column sums to 1, so G is column-stochastic. Moreover, every entry is positive, so the PageRank vector is unique.

(d) The PageRank vector satisfies

$$G\mathbf{p} = \mathbf{p}, \qquad \mathbf{1}^T\mathbf{p} = 1.$$

Equivalently,

$$(G - I)\mathbf{p} = \mathbf{0}, \qquad p_A + p_B + p_C + p_D = 1.$$

Solving this system gives

$$\mathbf{p} = \begin{bmatrix} p_A \\ p_B \\ p_C \\ p_D \end{bmatrix} \approx \begin{bmatrix} 0.2192 \\ 0.1752 \\ 0.3558 \\ 0.2497 \end{bmatrix}.$$

The entries sum to 1, as required.

(e) Ordering the components from largest to smallest gives

$$p_C > p_D > p_A > p_B.$$

Therefore, the page ranking is

$$C > D > A > B.$$

Page C is the most important page, while page B is the least important. Although C has no outgoing links, it receives links from A, B, and D, allowing it to accumulate the largest PageRank score.

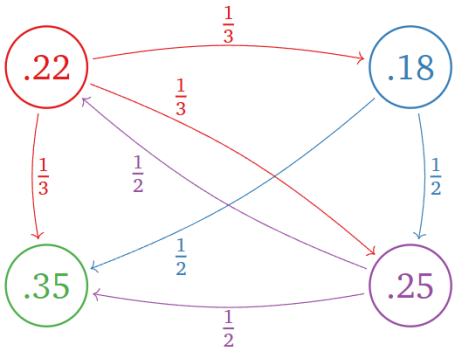


Figure 54: Image credit: [11]

PageRank therefore combines several central ideas from linear algebra: directed graphs are encoded as matrices, web navigation becomes a Markov chain, page importance appears as an eigenvector, and the dominant eigenvector is computed efficiently through repeated matrix–vector multiplication.

This section was adapted from [3, 11].

Language

Suppose we begin with the incomplete sentence

$$\text{The Eiffel Tower is located in}$$

To a human, this is a sequence of familiar words carrying an obvious meaning. To a computer, the sentence must first be divided into **tokens**, which may be complete words or smaller pieces of words. For simplicity, suppose each word shown below forms one token. Each token is assigned an embedding vector:

$$\text{The} \longmapsto \mathbf{x}_1, \quad \text{Eiffel} \longmapsto \mathbf{x}_2, \quad \text{Tower} \longmapsto \mathbf{x}_3, \quad \dots$$

Each vector here is a list of numbers. Combined with information about the position of each token, these vectors form a numerical representation of the sentence.

For centuries, vectors were used to represent physical quantities like forces, velocities, molar flux, electric fields, displacements, and positions. Today, they also represent language: words, phrases, relationships, and ideas.

Once the tokens have become vectors, the model no longer manipulates English directly. It instead manipulates geometry.

Throughout this book, we have studied how matrices transform vectors into new vectors. The same operation lies at the heart of a language model, where it is combined with attention, normalization, and nonlinear functions.

Collect the token vectors into a matrix $\mathbf{X}$. A sequence of transformer layers then produces increasingly contextual representations:

$$\mathbf{X} \longmapsto F_1(\mathbf{X}) \longmapsto F_2\left(F_1(\mathbf{X})\right) \longmapsto \cdots \longmapsto F_n\left(\cdots F_2\left(F_1(\mathbf{X})\right)\right)$$

Each function F_i contains learned matrix transformations, inner products, and weighted combinations of vectors, together with nonlinear operations.

Across these layers, the representations may come to encode grammatical roles, objects, relationships, and contextual meaning. The model may recognize that the Eiffel Tower is a landmark, connect landmarks with cities, and determine that Paris is more appropriate than Seoul.

After the final transformer layer, the context at the last position is represented by a vector $\mathbf{y}$. Every possible next token is associated with an output vector $\mathbf{v}_i$. The model computes a score, called a **logit**, using the inner product

$$z_i = \mathbf{y}^T \mathbf{v}_i.$$

A larger logit produces a greater next-token probability after the scores are passed through the softmax function. Suppose the model computes

$$z_{\text{Paris}} = 18.7, \quad z_{\text{France}} = 13.4, \quad z_{\text{Seoul}} = 6.8.$$

The largest logit belongs to Paris, so it receives the greatest next-token probability. Under greedy decoding, the sentence is completed as

The Eiffel Tower is located in Paris.

Behind that single word lies the full language of linear algebra: vectors, matrices, inner products, projections, and transformations. Indeed, the same mathematical language that helps describe the vibration of a bridge, the orbit of a satellite, the flow of heat, and the evolution of a quantum state can also help a machine complete a sentence.

We thank linear algebra for giving us a language to describe space, but it has become much more. It has become a language for for describing meaning. For representing meaning. For helping us understand meaning.

The mathematical foundation remains the same, but it now finds itself in a much wider world.

Bibliography

[1] 3Blue1Brown. *Essence of Linear Algebra*. https://www.youtube.com/playlist?list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab

[2] J. Anderson. *Applied Linear Algebra*. https://www.appliedlinearalgebra.com

[3] M. Crovella. *Linear Algebra, Geometry, and Computation*. https://www.cs.bu.edu/fac/snyder/cs132-book/landing-page.html

[4] M. Fowler. *Linear Algebra for Quantum Mechanics*. https://galileo.phys.virginia.edu/classes/751.mf1i.fall02/751LinearAlgebra.htm

[5] H. Gavin. *CEE 421L: Matrix Structural Analysis*. https://people.duke.edu/~hpgavin/cee421/

[6] J. Kitchin. *Applications in Linear Algebra*. https://kitchingroup.cheme.cmu.edu/f19-06623/16-linear-algebra.html

[7] O. Knill. *Math 19b: Linear Algebra with Probability*. https://people.math.harvard.edu/~knill/teaching/math19b_2011/handouts/math19b_2011.pdf

[8] M. Laforest. *The Mathematics of Quantum Mechanics*. https://uwaterloo.ca/institute-for-quantum-computing/sites/default/files/uploads/files/mathematics_qm_v23_qsys.pdf

[9] D. C. Lay, S. R. Lay, and J. J. McDonald. *Linear Algebra and Its Applications*.

[10] Loughborough University. *HELM Workbooks*. https://www.lboro.ac.uk/departments/mlsc/student-resources/helm-workbooks/

[11] D. Margalit and J. Rabinoff. *Interactive Linear Algebra*. https://textbooks.math.gatech.edu/ila/index.html

[12] J. B. Rawlings and J. G. Ekerdt. *Extra Exercises for Chemical Reactor Analysis and Design Fundamentals*. https://sites.engineering.ucsb.edu/~jbraw/chemreacfun/exercises/extra_exercises.pdf

[13] K. F. Riley, M. P. Hobson, and S. J. Bence. *Mathematical Methods for Physics and Engineering*.

[14] J. Rosen. *Jacobian–Implications & Applications*. http://bionics.seas.ucla.edu/education/MAE_263D/Robotics_04_Jacobian_07A_Singulairty.pdf

[15] G. Strang. *Introduction to Linear Algebra*.

[16] K. D. Wittrup and W. H. Green, Jr. *10.37 HW 3 Spring 2007*. https://ocw.mit.edu/courses/10-37-chemical-and-biological-reaction-engineering-spring-2007/d611926eb0f90957cae1f97a9d09b461_pset03_soln.pdf

[17] R. Tanase and R. Radu. *The Mathematics of Web Search*. https://pi.math.cornell.edu/~mec/Winter2009/RalucaRemus/index.html