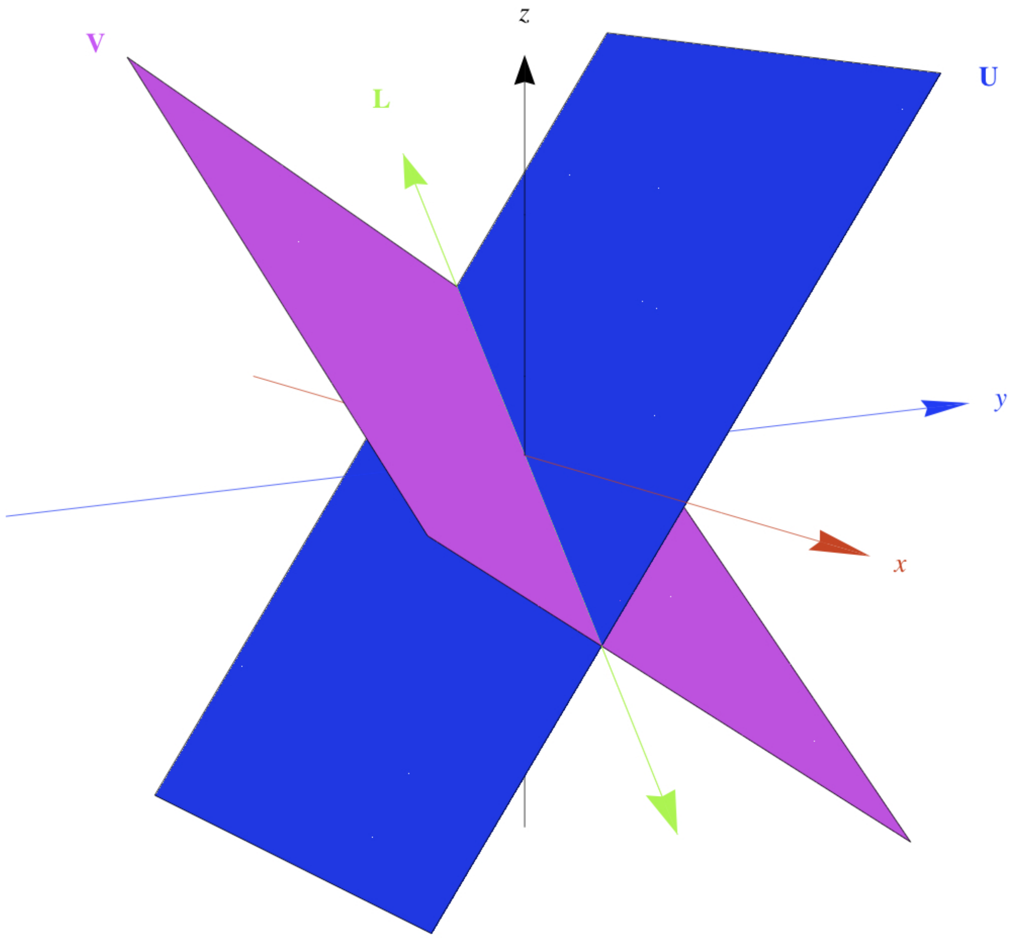


Vector Spaces First

An Introduction to Linear Algebra (Fourth edition)

Thierry Giordano, Barry Jessup and Monica Nevins

Thierry Giordano, email: giordano@uottawa.ca

Barry Jessup, email: bjessup@uottawa.ca

Monica Nevins, email: mnevins@uottawa.ca

THIS DOCUMENT IS AVAILABLE AT THE FOLLOWING LOCATIONS:
https://ruor.uottawa.ca/handle/10393/43955
https://github.com/uottawa-mathstat/linear-algebra-oer

Preface

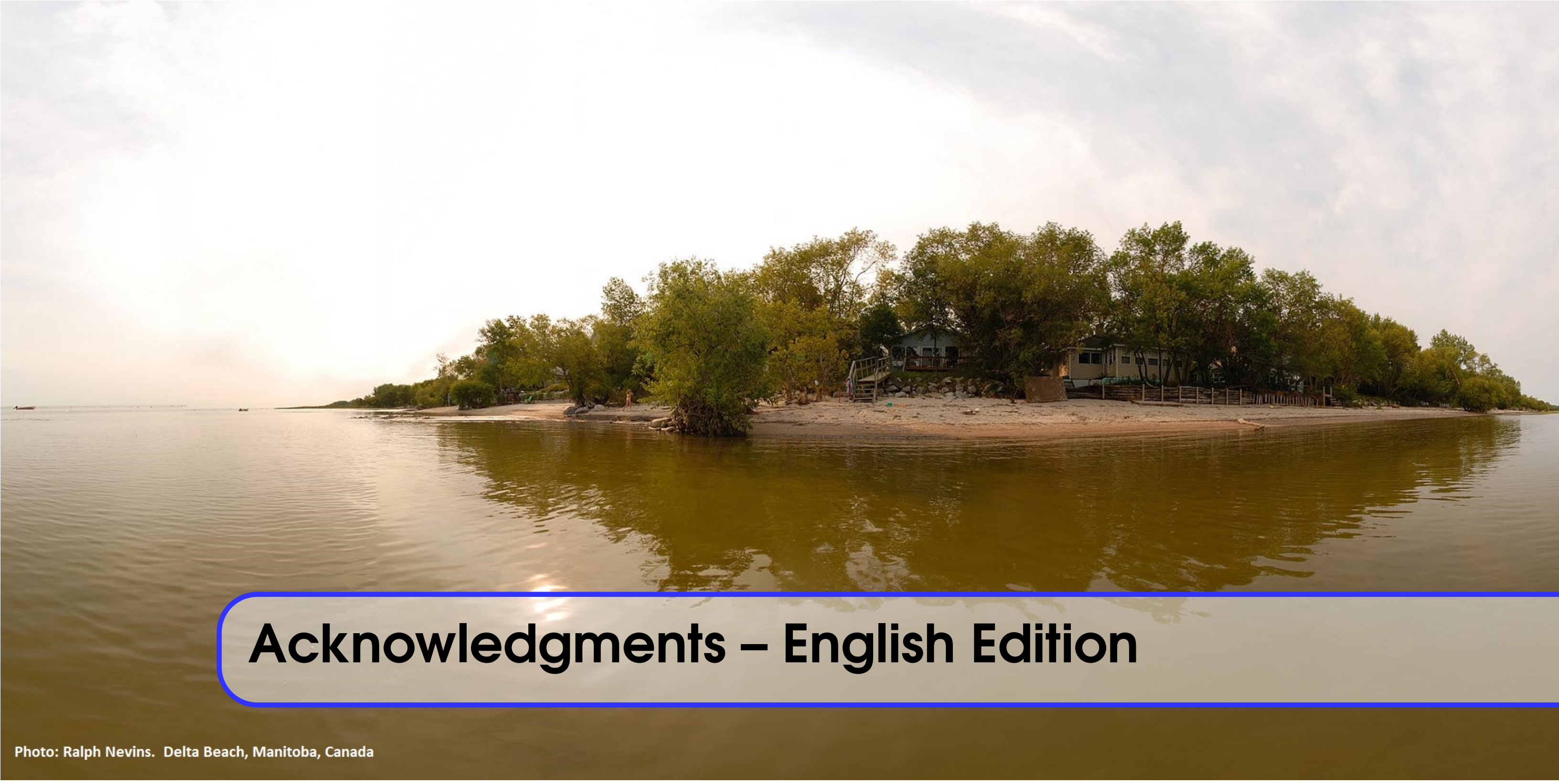
This volume grew from sets of lecture notes by Thierry Giordano, Barry Jessup, and Monica Nevins for teaching the course *Introduction to Linear Algebra* at the University of Ottawa. This book is intended to serve as a text or companion to the course.

The approach we take in this book is not standard: we introduce vector spaces very early and only treat linear systems after a thorough introduction to vector spaces.

We do this for at least two reasons. Our experience in teaching variations of this course over the past 25 years to thousands of students has taught us that the material on vector spaces, usually found toward the end of the course in a traditional textbook, is generally experienced by students as the most difficult part of the course. In a 12 week course, to have the most difficult material near the end of the course does not give most students enough opportunity to come to grips with the (seemingly) new ideas introduced when we meet vector spaces for the first time.

In our experience, starting with vector spaces within the first two weeks allows students much more time to appropriate the 'big' ideas in linear algebra: the notions of the *set of all linear combinations of vectors* (the 'span') and the *'linear independence' of vectors*. These notions lie at the heart of linear algebra and are usually experienced as new and challenging by students who see them for the first time. So, the sooner they see them, the better: we can use them, as well as the notions of *basis* and *dimension* in the rest of the course, in different contexts. In this way, it has been our experience that most students prefer to see the challenging material as early as possible, so that they have time to acquaint themselves with material that is genuinely new and different from what they've seen in high school.

Another reason to tackle vector spaces as soon as possible is to alert students to the fact that there *is* genuinely new and different material in the course! If one begins with material on linear systems many students have seen in high school, (in low dimensions, at least), some can easily fall prey to the idea that there's not much to this course, will and will be at caught later on when vector spaces come along.



Acknowledgments – English Edition

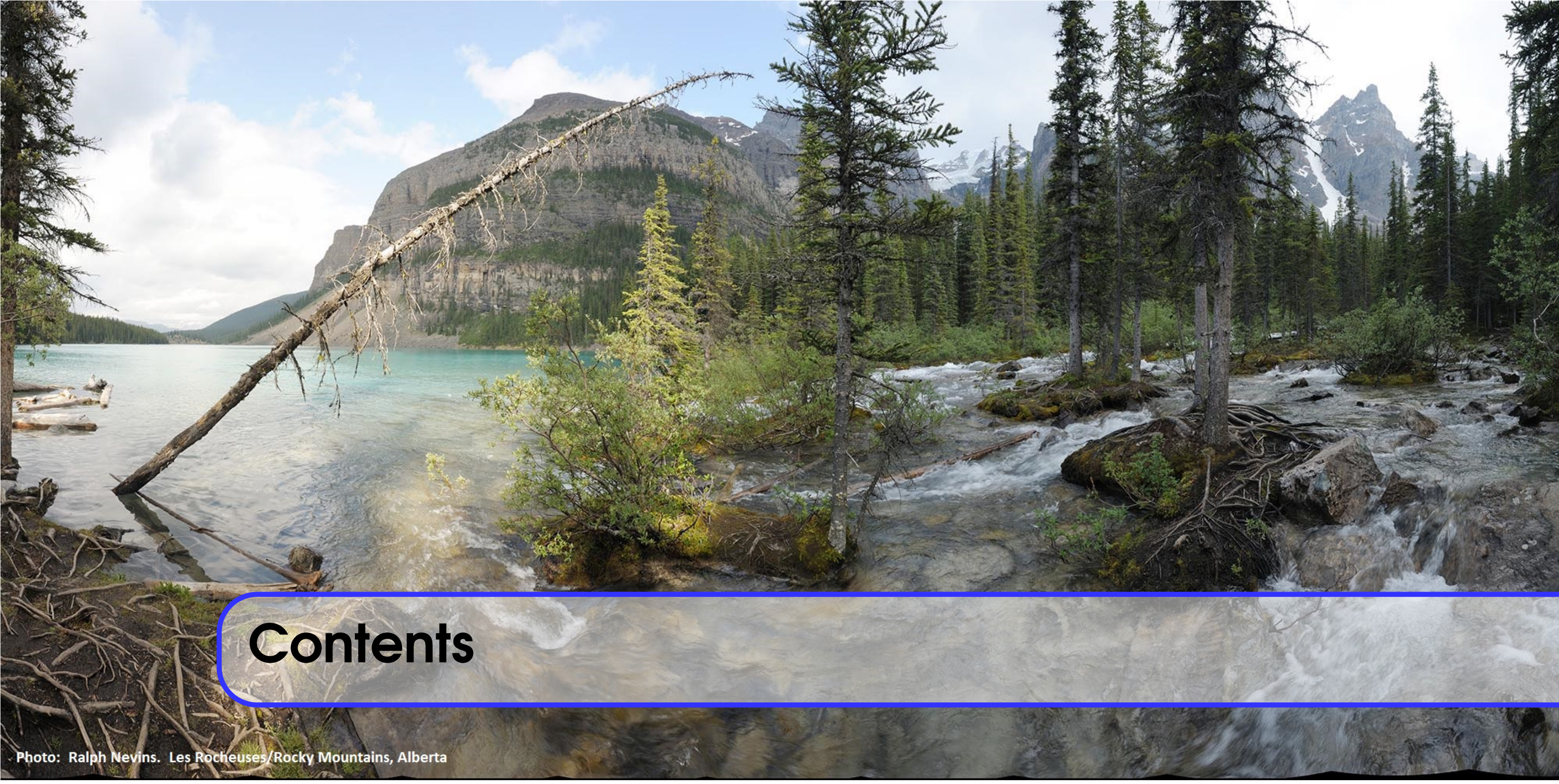
The authors would like to thank those who helped in the making of this book, even though they were completely unaware of it. Our book grew out of our lecture notes for a first year course at the University of Ottawa over a long period, during which time we were guided by several well-known, excellent and comprehensive texts by authors including Howard Anton, W. Keith Nicholson, David C. Lay, and Seymour Lipschutz and Marc Lipson. Over the years, we also consulted texts by Tom M. Apostol, Otto Brestcher, and Gilbert Strang.

We readily extend our gratitude to colleagues who found the inevitable errors and typos in the first edition. Principally, we thank Anne Broadbent, Saeid Molladavoudi, and Charles Starling who sent us embarassingly long lists of typos and errors as well as many excellent suggestions for changes.

The book was first used as a text here in the Fall of 2015, and we offer our thanks to the dozens of students in MAT1341 who read our book, found our mistakes and were kind enough to let us know.

That being said, those errors or typos that remain are entirely due to the authors!

Thierry Giordano, Barry Jessup & Monica Nevins
January 2015
University of Ottawa



Contents

Acronyms

Common abbreviations used in the text

Col (A)	Column space of the matrix A
$\det(A)$	Determinant of the matrix A
im	Image (of a matrix or linear transformation)
ker	Kernel (of a matrix or linear transformation)
LD	Linearly Dependent
LI	Linearly Independent
Null(A)	Nullspace of a matrix A
REF	Row Echelon-Form (of a matrix)
Row (A)	Row space of the matrix A
RREF	Reduced Row Echelon Form (of a matrix)
tr(A)	trace of the matrix A

Common symbols

V	a vector space
U	a subspace of V
$\mathbf{v}$	a vector
$\mathbf{0}$	the zero vector
A^T	transpose of the matrix A
A^{-1}	inverse of the matrix A

I Algebra and Geometry

A major theme of linear algebra is how to generalize ideas in geometry to higher dimensions, by interpreting geometric ideas algebraically. At the same time, one finds that thinking about an algebraic problem geometrically gives insights into the algebra.

We begin our exploration of these themes with a study of the complex numbers, and then a review of some high school geometry (with an eye to extending some of ideas there to higher dimensions).

1. Complex Numbers

In brief, one might say that algebra is the study of solutions of polynomial equations. Linear algebra is then the study of solutions to linear equations. (It gets interesting when you allow multiple variables and multiple equations, but we'll get to that later.)

For today, let's look at an application of Algebra, as a "welcome back to algebra" for everyone after a long summer away, and as a heads up to our Engineers (particularly Electrical Engineers) and Physicists who will soon be using complex numbers all the time.

1.1 Defining the Complex Numbers

The history of the complex numbers is very interesting: it does not begin, as one might think, with the equation

$$x^2 + 1 = 0,$$

but rather with cubic equations! Everyone was certain that $x^2 + 1 = 0$ has no solutions – just look at the graph of $y = x^2 + 1$, they'd argue.[1] However every *cubic* equation (with real coefficients) does indeed have at least one real solution - and imaginary numbers (they were called 'impossible' numbers for some time) were invented to obtain formulae for the real roots of cubics.[2]

Nevertheless, let us denote one solution of $x^2 + 1 = 0$ by i, for "imaginary" (notation thanks to Euler, 1777):

$$i^2 = -1 \qquad \text{or} \qquad i = \sqrt{-1}.$$

Now we define for a *negative* real number a,

$$\sqrt{a} := (\sqrt{|a|})i.$$

[1] Indeed, in ancient times one would have said that $x + 1 = 0$ has no solutions either, since negative numbers don't represent quantities.

[2] Search the web for "the history of complex numbers".

This may seem a bit fussy, but if we simply try to apply the normal rules of algebra, such as in:

$$\sqrt{-9} = \sqrt{9\cdot(-1)} = \sqrt{9}\cdot\sqrt{-1} = 3i,$$

we obtain the correct answer, but much caution is needed at the third equality.[3] Stick with the fussy definition for now!

(Please note that here, as in Calculus, we adhere to the convention that for *real* a, $\sqrt{a^2} = |a|$, that is, the answer is the one positive square root, and not a choice of them.)

Now i alone is not quite enough. Consider the equation

$$x^2 + 4x + 8 = 0.$$

By the quadratic formula, its roots are the two numbers

$$x = \frac{-4 \pm \sqrt{16-32}}{2} = -2 \pm \frac{1}{2}\sqrt{-16} = -2 \pm \frac{1}{2}(4i) = -2 \pm 2i.$$

Let's check that this makes sense: Plug $x = (-2+2i)$ into the quadratic equation and simplify:

$$(-2+2i)^2 + 4(-2+2i) + 8 = (4-4i-4i+4i^2) + (-8+8i) + 8 = 4 + 4i^2 = 0$$

where in the last step we remembered that $i^2 = -1$. Similarly we can check that $-2-2i$ is also a root.

With these thoughts (and the quadratic formula) in mind, we make the following definition.

Definition 1.1.1 The set of *complex numbers* is the set

$$\mathbb{C} = \{a + b\,i \colon a, b \in \mathbb{R}\}.$$

(Read this as: "the set of all things of the form $a+bi$ where a and b are real numbers".)

When we write

$$z = a + b\,i \in \mathbb{C}$$

(read as: z (or $a+b\,i$) belongs to $\mathbb{C}$), then a is called the *real part* of z (denoted $Re(z)$) and b is called the imaginary part of z (denoted $Im(z)$). Note that $Re(z)$ and $Im(z)$ are real numbers!

When $Re(z) = 0$ then $z = bi$ and we say z is *purely imaginary*; when $Im(z) = 0$ then $z = a$ is real. Thus $\mathbb{R} \subset \mathbb{C}$ (read as: $\mathbb{R}$ is a subset of $\mathbb{C}$).

The real numbers were drawn as the real number line; the complex numbers are drawn as the complex *plane* as we shall soon see.

So notice that given any quadratic equation $ax^2 + bx + c = 0$, with real coefficients a, b, c, the roots are

$$x = \frac{-b}{2a} \pm \frac{\sqrt{b^2-4ac}}{2a}.$$

If $b^2 - 4ac \geq 0$, the roots are real, otherwise, they are complex numbers. So now EVERY quadratic equation has two roots (counting a double root as two, now and always).

[3] A real course in Complex Analysis is needed here, otherwise one can get into trouble: e.g. $9 = \sqrt{81} = \sqrt{(-9)(-9)} \neq \sqrt{-9}\,\sqrt{-9} = 3i\,3i = -9$. Indeed, the rule for positive real numbers a, b that (correctly) states: $\sqrt{ab} = \sqrt{a}\sqrt{b}$ *does not hold* in general. This may seem like a nuisance but is actually the source of lots of interesting math: search the web for "Riemann surfaces and the square root".

1.2 Algebra of the Complex Numbers

- Equality: $a+bi = c+di \Leftrightarrow a = c$ and $b = d$;
- Addition: $(a+bi)+(c+di) = (a+c)+(b+d)i$
- Multiplication: $(a+bi)(c+di) = (ac-bd)+(ad+bc)i$

In fact, complex numbers satisfy all the same properties as $\mathbb{R}$ except there is no ordering (that is, "$z > y$" doesn't make sense).

What about division?

Problem 1.2.1 Solve $(4+3i)z = 1$, if possible; that is, what do we mean by

$$z = \frac{1}{4+3i}?$$

(PLEASE PLEASE remember that $\frac{1}{2+3} \neq \frac{1}{2}+\frac{1}{3}$!!!!)

Solution The idea: remember that i is a square root – so use the standard trick of algebra called *rationalizing the denominator*.

So

$$(a+b\,i)(a-b\,i) = a^2-(b\,i)^2 = a^2+b^2$$

As a and b are *real*, $a^2+b^2 \neq 0$, unless both a and b are zero. So for $z = a+bi$, let

- $\bar{z} = a-bi$, the complex conjugate of z
- $|z| = \sqrt{a^2+b^2}$, the absolute value, or modulus, of z

(Note: $z = 0$ if and only if (iff) $|z| = 0$; and we CAN compare moduli of complex numbers: $|i| = |1| > |0|$. Also note that we used complex conjugates in the quadratic formula; this is familiar!)

Now we have

$$z\bar{z} = |z|^2$$

so

$$z \cdot \left(\frac{\bar{z}}{|z|^2}\right) = 1$$

or

$$\frac{1}{z} = \frac{\bar{z}}{|z|^2}.$$

Excellent! So

$$\frac{1}{4+3i} = \left(\frac{1}{4+3i}\right)\left(\frac{4-3i}{4-3i}\right) = \frac{4-3i}{4^2+3^2} = \frac{4-3i}{25} = \frac{4}{25}-\frac{3}{25}i.$$

Problem 1.2.2 Simplify

$$\frac{3+2i}{-2+4i}$$

Solution Multiply by $1 = \frac{-2-4i}{-2-4i}$:

$$\frac{3+2i}{-2+4i} = \frac{3+2i}{-2+4i}\cdot\frac{-2-4i}{-2-4i} = \frac{(3+2i)(-2-4i)}{4+8i-8i-16i^2} = \frac{2-16i}{20} = \frac{1}{10}-\frac{4}{5}i.$$

This is what we wanted: the answer is in a form that we recognize as a complex number.

Some other properties (easy to prove):

Lemma 1.2.3 — Properties of Complex Numbers. Complex conjugation has the following properties. Suppose that $z, w \in \mathbb{C}$ and $c \in \mathbb{R}$.

- $\frac{1}{z} = \frac{\overline{z}}{|z|^2}$
- $\overline{z+w} = \overline{z} + \overline{w}$.
- $\overline{cz} = c\overline{z}$.
- $\overline{zw} = \overline{z}\,\overline{w}$.
- $\overline{z/w} = \overline{z}/\overline{w}$.
- $\overline{\overline{z}} = z$ (eg: $\overline{\overline{3+2i}} = \overline{3-2i} = 3+2i$)
- $\overline{z} = z$ iff $z \in \mathbb{R}$
- $\overline{z} = -z$ iff z is purely imaginary
- $|z| \in \mathbb{R}$ and $|z| \geq 0$
- $|z| = |\overline{z}|$
- $|zw| = |z|\,|w|$
- $|z/w| = |z|/|w|$
- $|z+w| \leq |z| + |w|$ ('Triangle inequality')
- If a is a real number then $|a+0i|$ is the absolute value of a. (So we just write $z = a$ rather than $z = a + 0i$; there's no problem.)

Essentially what this is saying is that in any expression for a complex number z, the expression for $\overline{z}$ is given by replacing each occurence of the symbol "i" with "$-i$"; you don't have to simplify it first.

Proof. (You can discuss these with me, or with the TA in the DGD. Proving such things is important because it shows you WHY something is true, and that in turn makes it hard to forget, or remember falsely.)

(2) Suppose z and w are complex numbers. That means $z = a + bi$ for some real numbers a and b, and $w = c + di$ for some real numbers c and d. Then

$$z + w = (a+c) + (b+d)i \qquad \text{so} \qquad \overline{z+w} = (a+c) - (b+d)i.$$

On the other hand, $\overline{z} = a - bi$ and $\overline{w} = c - di$, so that

$$\overline{z} + \overline{w} = (a+c) + (-b-d)i = (a+c) - (b+d)i.$$

Hence the two sides are equal, which completes the proof.

We prove another one, and leave the rest as exercises. (For the triangle inequality, it's easier to use geometry, below; and for the multiplicative property, it's easier to use polar form.)

(10) Suppose z is a complex number. Then

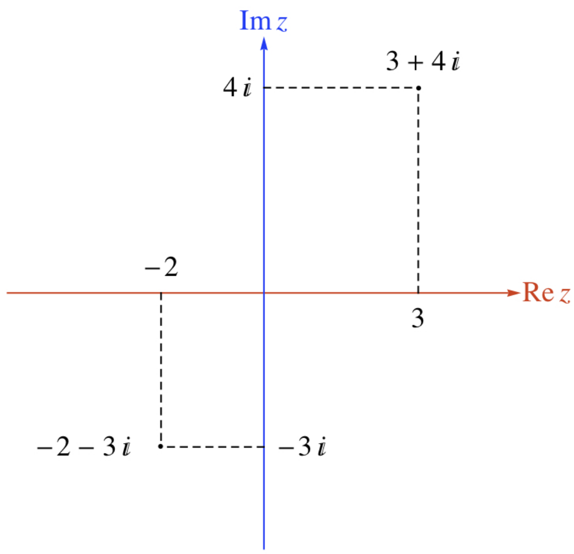
$$|z| = \sqrt{z\overline{z}}$$

whereas

$$|\overline{z}| = \sqrt{\overline{z}\,\overline{\overline{z}}} = \sqrt{\overline{z}z} = \sqrt{z\overline{z}} = |z|$$

by all that we've established before. So the two are equal. ■

1.3 Geometry of the complex numbers



Each complex number is written as $a+bi$ with a and b real numbers. To picture this, think of $a+bi$ as the point (a,b) in the xy-plane. Then $a+bi$ is a point in the plane, which in this context we call the *complex plane*. The horizontal axis is called the *real axis* and the vertical axis is called the *imaginary axis*.

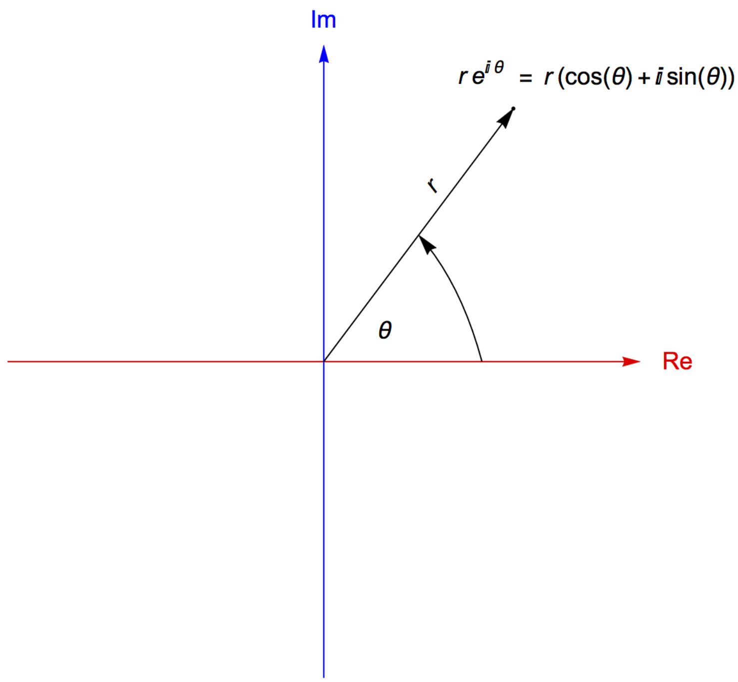
Addition of two complex numbers is the same as the addition of two vectors in the plane. The negation of a complex number is just the negation of the corresponding vector. Multiplication by real numbers is just scalar multiplication, but multiplication by complex numbers is a combination of rotation and scaling (try it out!).

Complex conjugation is just the reflection of the vector through the real axis. So if $z=a+bi$ then $\bar{z}=a-bi$.

The modulus of z is just the length of the vector representing z. (Recall that the length of the vector from the origin to (a,b) is $\sqrt{a^2+b^2}$.)

1.4 Polar Form of Complex Numbers

These are also called “polar coordinates” and show up in Calculus II, for the real plane.



So if $z = x + y i \in \mathbb{C}$ then

$$\frac{x}{r} = \cos(\theta), \quad \frac{y}{r} = \sin(\theta)$$

and

$$r = |z| = \sqrt{x^2 + y^2}$$

Therefore:

$$z = x + yi = r\cos(\theta) + ir\sin(\theta) = r(\cos(\theta) + i\sin(\theta)).$$

(Note that $\theta = \arg(z)$ (argument of z) is not uniquely determined, since $\theta' = \theta + 2n\pi$, $n \in \mathbb{Z}$, also works. We usually pick $-\pi < \theta \leq \pi$ and write $\theta = \operatorname{Arg}(z)$, the principal argument of z.)

In 1748, Euler proved

$$e^z = 1 + z + \frac{z^2}{2} + \cdots = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

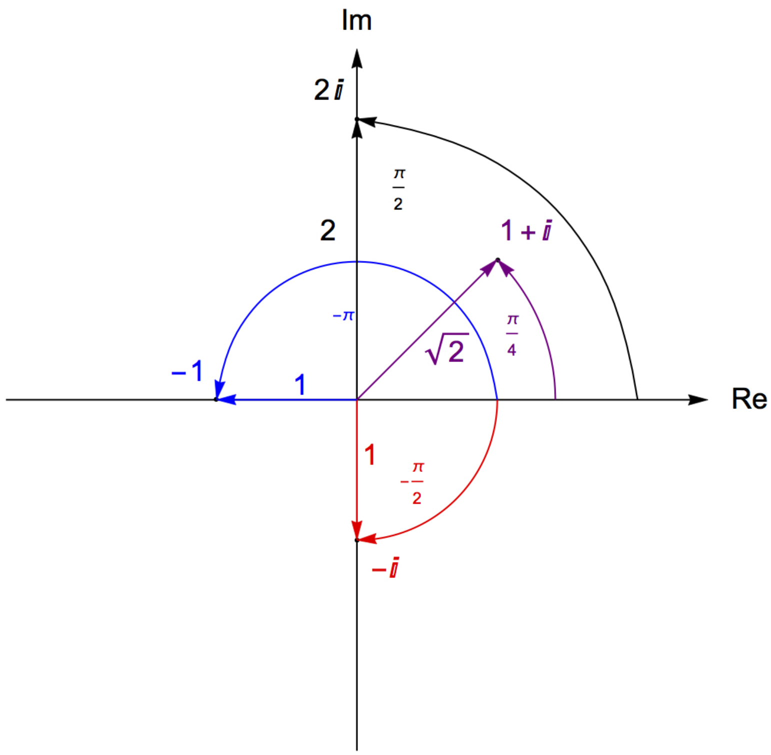
(power series) so in fact by comparing power series with trig functions we get, surprisingly,

$$e^{i\theta} = \cos(\theta) + i\sin(\theta).$$

So modern notation is: the *polar form* of the complex number z is

$$z = re^{i\theta}.$$

■ **Example 1.4.1** See the diagram below and check that $2i = 2e^{i\frac{\pi}{2}}$, $-i = e^{-i\frac{\pi}{2}}$, $-1 = e^{i\pi}$, and $1 + i = \sqrt{2}e^{i\frac{\pi}{4}}$ ■



We have

- $\overline{re^{i\theta}} = r'e^{i\theta'}$ iff $r = r'$ and $\theta = \theta' + 2n\pi$, some $n \in Z$.
- $\overline{re^{i\theta}} = re^{-i\theta}$
- $|e^{i\theta}| = 1$ for any θ

1.5 Multiplying complex numbers in polar form

Note that if $z = re^{i\theta}$ and $w = se^{i\phi}$ then

$$\begin{aligned} zw &= (r(\cos(\theta) + i\sin(\theta)))\,(s(\cos(\phi) + i\sin(\phi)) \\ &= (rs)\,((\cos(\theta)\cos(\phi) - \sin(\theta)\sin(\phi)) + i(\cos(\theta)\sin(\phi) + \sin(\theta)\cos(\phi))) \\ &= rs(\cos(\theta + \phi) + i\sin(\theta + \phi)) \\ &= rse^{i(\theta+\phi)} \end{aligned}$$

That is: it's like the usual multiplication with exponents. (Same for division.)

Problem 1.5.1 Compute

$$\frac{i}{1+i}$$

Solution From before, we have $i = e^{i\frac{\pi}{2}}$ and $1 + i = \sqrt{2}e^{i\frac{\pi}{4}}$ so

$$\frac{i}{1+i} = \frac{e^{i\frac{\pi}{2}}}{\sqrt{2}e^{i\frac{\pi}{4}}} = \frac{1}{\sqrt{2}}e^{i\frac{\pi}{4}}$$

(check, using the long method!)

1.6 The Fundamental Theorem of Algebra

We noted above that every quadratic polynomial with real coefficients has two roots, and in fact if one root was complex and not real, then so was the other, and the two roots are complex conjugates.

Conversely, given a complex number z, one polynomial with roots z and $\overline{z}$ is

$$(x-z)(x-\overline{z}) = x^2 - (z+\overline{z})x + z\overline{z}.$$

Is this a polynomial with real coefficients? YES! Write $z = a+ib$, then $z+\overline{z} = 2a$, which is real; and $z\overline{z}$ is real (as we proved earlier).

So: every real polynomial has complex roots, and every complex number is the root of some real polynomial.

That said, if you take two complex numbers z and w which are not conjugate, then $(x-z)(x-w)$ will just be some random quadratic polynomial with complex coefficients. Which begs the question: if you take all quadratic polynomials with complex coefficients, and use the quadratic formula to solve them, what extra numbers (like i) do you need this time?

Answer: NONE. The complex numbers are the top of the heap, and all you'll ever need:

Theorem 1.6.1 — Fundamental Theorem of Algebra. Every polynomial with coefficients in the complex numbers factors completely into linear factors of the form $ax+b$, with $a, b \in \mathbb{C}$.

1.7 Some musings about the meaning of the Fundamental Theorem of Algebra

This answers a really big nagging problem of algebra: shouldn't every quadratic have 2 roots, and every cubic 3 (allowing multiple roots)? But the quadratic x^2+2 didn't have any roots over the reals.

In Calculus, we accept this by sketching the graph of $y = x^2+2$ and saying, "Look, it doesn't intersect the x-axis. That explains it."

"Explains what, exactly?" an algebraist responds. "You're still missing two roots."

In algebra, we are looking for unifying themes, things that are common across all problems of a particular type. We look for wonderful universal solutions. The complex numbers are one example of a universal solution: we added $\sqrt{-1}$ and suddenly all problems were solved: x^2+2 has two roots, and $ix^7+3x^2-(4+i)$ has 7 roots.

(Well, one might admit, not all problems are completely solved. The Fundamental Theorem of Algebra doesn't say ANYTHING about FINDING the roots. It just says that they exist. We end up going back to Calculus for help in finding them.)

For the rest of this course, we will be considering LINEAR algebra, which is a particular branch of algebra where, it turns out, there are fantastically universal solutions to absolutely everything (not just "existence", but actually ways of finding solutions!).

Problems

Those marked with a $\star$ have solutions at the back of the book. Try them yourself first.

Problem 1.1 *

Express the following complex numbers in Cartesian form: $a+b\,i$, with $a,b\in\mathbb{R}$.

a) $(2+i)(2+2i)$

b) $\dfrac{1}{1+i}$

c) $\dfrac{8+3i}{5-3i}$

d) $\dfrac{5+5\sqrt{3}\,i}{\sqrt{2}-\sqrt{2}\,i}$

e) $\dfrac{(1+2i)(2+5i)}{3+4i}$

f) $\dfrac{1-i}{2-i}+\dfrac{2+i}{1-i}$

g) $\dfrac{1}{(1-i)(3-2i)}$

Problem 1.2 *Find the polar form of the following complex numbers: (i.e., either as $re^{i\theta}$ or as $r(\cos\theta+i\sin\theta)$, with $r\geq 0$ and $-\pi<\theta\leq\pi$)

a) $3\sqrt{3}-3i$

b) $\dfrac{3\sqrt{3}-3i}{\sqrt{2}+i\sqrt{2}}$

c) $\dfrac{1-\sqrt{3}i}{-1+i}$

d) $\dfrac{5+5\sqrt{3}i}{\sqrt{2}-\sqrt{2}i}$

e) $\dfrac{3+3\sqrt{3}i}{-2+2i}$

Problem 1.3 Find the modulus of each of the complex numbers in questions 1 & 2. (Remember that $|zw|=|z|\,|w|$ and that if $w\neq 0$, then $\left|\dfrac{z}{w}\right|=\dfrac{|z|}{|w|}$.)

Problem 1.4 *If z is a complex number,

(i) Is it possible that $z=\bar{z}$?
(ii) Is it possible that $|\bar{z}|>|z|$?
(iii) Is it possible that $\bar{z}=2z$?

Give examples to illustrate your affirmative answers, and explanations if you say the statement is always false.

Problem 1.5 For budding algebraists...
Show that there is no ‘proper’ ordering on the complex numbers, that is, there is no binary relation $>$ with the property that if $x > 0$ and $y > z$ then $xy > xz$.

2. Vector Geometry

Much of the material in this chapter may be review from high school, but it allows us to set the stage for the upcoming chapters.

2.1 Vectors in $\mathbb{R}^n$

Vector comes from the Latin *vehere*, which means to carry, or to convey; abstractly we think of the vector as taking us along the arrow that we represent it with. For example, we use vectors in Physics to indicate the magnitude and direction of a force.

Let's use our understanding of the geometry and algebra of vectors in low dimensions (2 and 3) to develop some ideas about the geometry and algebra in higher dimensions.

Algebra	Geometry
$\mathbb{R}$, real numbers, *scalars*	real line
$\mathbb{R}^2 = \{(x,y) \mid x,y \in \mathbb{R}\}$, vectors $\mathbf{u} = (x,y)$	plane
$\mathbb{R}^3 = \{(x,y,z) \mid x,y,z \in \mathbb{R}\}$, vectors $\mathbf{u} = (x,y,z)$	3-space
$\vdots$	
(why not keep going?)	
$\vdots$	
$\mathbb{R}^4 = \{(x_1,x_2,x_3,x_4) \mid x_i \in \mathbb{R}\}$, vectors $\mathbf{x} = (x_1,x_2,x_3,x_4)$	
Hamilton (1843): extended $\mathbb{C}$ to *hamiltonians*	*space-time*
$n \in \mathbb{Z}$, $n > 0$: $\mathbb{R}^n = \{(x_1,x_2,\cdots,x_n) \mid x_i \in \mathbb{R}\}$	
$\mathbf{x} = (x_1,\cdots,x_n)$	n-space

Notation

We have several different notations which we use for writing vectors, which we use interchangeably:

- $\mathbf{x} = (1, 2, 3, 4)$
- $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$
- $\mathbf{x} = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}^T$, where the exponent T stands for *tranpose*. (*Transposition* means turning rows into columns (and *vice versa*).)

The vertical notation (also referred to as *matrix form*) is the easiest to read but the first one is easier to write.

(The reason for the vertical vector notation comes from matrix multiplication, which we'll get to later.)

2.2 Manipulation of vectors in $\mathbb{R}^n$

The algebraic rules for $\mathbb{R}^n$ extend directly from the algebraic rules for $\mathbb{R}^2$ and $\mathbb{R}^3$:

- Equality: $(x_1, \cdots, x_n) = (y_1, \cdots, y_n) \Leftrightarrow x_i = y_i$ for all $i \in \{1, \cdots, n\}$. (In particular, $(x_1, \cdots, x_n) \neq (y_1, \cdots, y_m)$ if $n \neq m$.)
- Addition: $(x_1, \cdots, x_n) + (y_1, \cdots, y_n) = (x_1 + y_1, \cdots, x_n + y_n)$
- Zero vector: $\mathbf{0} = (0, 0, \cdots, 0) \in \mathbb{R}^n$
- Negative: if $\mathbf{x} = (x_1, \cdots, x_n)$ then $-\mathbf{x} = (-x_1, \cdots, -x_n)$
- Multiplication **by a scalar**: let $c \in \mathbb{R}$ be a scalar, then $c(x_1, \cdots, x_n) = (cx_1, \cdots, cx_n)$

Note that we don't have any way of *multiplying* two vectors; the closest we can get to that is the *dot product*, which gives a scalar as an answer, or the *cross product*, which only works in $\mathbb{R}^3$[1]. There's even a way to have 'multi-products' of $n-1$ vectors in $\mathbb{R}^n$, and get an answer in $\mathbb{R}^n$. Once you've seen the *determinant* later, you'll know exactly what to do if you look again at the definition of the cross product.

There are geometric interpretations of each of the above algebraic rules, which we can draw in 2 (and 3, if you draw well) dimensions.

- Equality: two vectors are equal if they have the same magnitude and direction
- Addition: parallelogram rule, or head-to-tail rule
- Zero vector: the vector with zero length
- Negative: the same arrow with head and tail exchanged
- Scalar multiple: scale the vector by $|c|$, and change direction if $c < 0$; or: two vectors are *parallel* if and only if they are scalar multiples of one another.

2.3 Linear combinations (Important Concept!)

The only operations we have are: vector addition, and scalar multiplication. We are going to be interested in the question: If I already have the vectors $\mathbf{u}_1$, $\mathbf{u}_2$, $\cdots$, $\mathbf{u}_m$ (that is, m vectors in $\mathbb{R}^n$), can I produce another vector by scaling each vector and adding them together, in some way?

By analogy: Say $\mathbf{u}_1$ is parallel to Bank Street pointing north and $\mathbf{u}_2$ is parallel to Laurier Ave, pointing east; then I can get anywhere downtown by going in the direction of $\mathbf{u}_1$ (or its negative) for a

[1] However, there are ways generalize the cross product in $\mathbb{R}^n$ for every n, but the product doesn't live in $\mathbb{R}^n$! Search the web for 'exterior algebra'.

little way, and then taking direction $\mathbf{u}_2$ (or its negative) for a little way. But I can't get underground, or into the air, by following those directions. I'm stuck on the plane which is the ground.

Algebraically, this all comes down to the following definition.

Definition 2.3.1 If $k_1, k_2, \cdots, k_m \in \mathbb{R}$ are scalars, and $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m \in \mathbb{R}^n$ are vectors, then

$$k_1\mathbf{u}_1 + k_2\mathbf{u}_2 + \cdots + k_m\mathbf{u}_m$$

is called a *linear combination* of $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m$.

■ **Example 2.3.2** Let $\mathbf{u}_1 = (1,2,3)$ and $\mathbf{u}_2 = (1,0,0)$.

Then $\mathbf{x} = (17,4,6)$ is a linear combination of $\mathbf{u}_1$ and $\mathbf{u}_2$ because

$$\begin{bmatrix} 17 \\ 4 \\ 6 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 15\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

But $\mathbf{y} = (0,1,0)$ is *not* a linear combination of $\mathbf{u}_1$ and $\mathbf{u}_2$ because the equation

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = a\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + b\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \tag{2.1}$$

would imply

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ 2a \\ 3a \end{bmatrix} + \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} a+b \\ 2a \\ 3a \end{bmatrix}$$

So that

$$0 = a+b, \quad 1 = 2a, \quad \text{and} \quad 0 = 3a;$$

but the second and third equations are contradictory, so there cannot be a solution to (2.1). Thus $\mathbf{y}$ isn't a linear combination of $\mathbf{u}_1$ and $\mathbf{u}_2$. ■

2.4 Properties of vector addition and scalar multiplication

Note that all the usual algebraic properties of addition and scalar multiplication hold, whether our vectors have 2 components or n components. That is, let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and let $c, c' \in \mathbb{R}$; then we have:

- $\mathbf{u} + \mathbf{0} = \mathbf{u}$
- $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
- $(\mathbf{u}+\mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v}+\mathbf{w})$
- $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
- $c(\mathbf{u}+\mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
- $(c+c')\mathbf{u} = c\mathbf{u} + c'\mathbf{u}$
- $(cc')\mathbf{u} = c(c'\mathbf{u})$
- $1\mathbf{u} = \mathbf{u}$

Bear these properties in mind, as they are the key to generalizing to vector spaces beyond $\mathbb{R}^n$.

2.5 More geometry: the Dot Product in $\mathbb{R}^n$

The *dot product* (also called an *inner product*) gives us an algebraic way to describe some interesting geometric properties: length of a vector, and angle between two vectors.

Recall the dot product from $\mathbb{R}^3$:

Let $\mathbf{u} = (x_1, x_2, x_3)$ and $\mathbf{v} = (y_1, y_2, y_3)$ be two vectors in $\mathbb{R}^3$. Then their dot product is

$$\mathbf{u} \cdot \mathbf{v} = x_1 y_1 + x_2 y_2 + x_3 y_3.$$

Note that this is a scalar.

Once we have the dot product, we define:

- $\|\mathbf{u}\| = \sqrt{x_1^2 + x_2^2 + x_3^2} = \sqrt{\mathbf{u} \cdot \mathbf{u}}$, the *length* or *norm* of $\mathbf{u}$;
- $\|\mathbf{u} - \mathbf{v}\|$ is the *distance between* $\mathbf{u}$ *and* $\mathbf{v}$.

We can see that this is correct, by drawing a cube with main diagonal $\mathbf{u}$ and noting that the length of the main diagonal is given by the square root of the sum of the squares of the lengths of the sides by repeated applications of the Pythagorean theorem.

So, let's generalize this to $\mathbb{R}^n$:

Definition 2.5.1 Let $\mathbf{u} = (x_1, \ldots, x_n)$ and $\mathbf{v} = (y_1, \ldots, y_n)$ be vectors in $\mathbb{R}^n$. Then their *dot product* is defined to be

$$\mathbf{u} \cdot \mathbf{v} = x_1 y_1 + \cdots + x_n y_n$$

and the *norm* of $\mathbf{u}$ is defined to be

$$\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}} = \sqrt{x_1^2 + \cdots + x_n^2}$$

We sometimes call $\mathbb{R}^n$, equipped with the dot product, *Euclidean n-space*.[a]

[a] As opposed to, for example, 'Minkowski spacetime', where the 'lengths' of vectors can be imaginary! Search the web for 'minkowski space wiki'.

■ **Example 2.5.2** Let $\mathbf{u} = (1, 2, -1, 0, 1)$, $\mathbf{v} = (1, 3, 2, 1, 1)$. Then

$$\mathbf{u} \cdot \mathbf{v} = 1 + 6 - 2 + 0 + 1 = 6$$

and $\|\mathbf{v}\| = \sqrt{1 + 9 + 4 + 1 + 1} = \sqrt{16} = 4$. ■

Note that for any vector $\mathbf{u} \in \mathbb{R}^n$:

$$\|\mathbf{u}\| = 0 \Leftrightarrow \mathbf{u} = \mathbf{0}$$

(because the norm is still the sum of real squares, and so is never zero unless each component is).

2.6 Orthogonality

Recall that in $\mathbb{R}^2$ and $\mathbb{R}^3$, we have that

$$\mathbf{u} \cdot \mathbf{v} = 0 \Leftrightarrow \mathbf{u} \text{ and } \mathbf{v} \text{ are orthogonal (or 'perpendicular')}.$$

Definition 2.6.1 Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Then $\mathbf{u}$ and $\mathbf{v}$ are said to be *orthogonal* if $\mathbf{u} \cdot \mathbf{v} = 0$.

■ **Example 2.6.2** The vectors $(1,2,-2,1)$ and $(4,1,3,0)$ of $\mathbb{R}^4$ are orthogonal (perpendicular), since

$$\begin{bmatrix} 1 \\ 2 \\ -2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 1 \\ 3 \\ 0 \end{bmatrix} = 4+2-6+0 = 0.$$

■

2.7 Angles between vectors in $\mathbb{R}^n$

Now saying that two vectors are orthogonal is another way of saying that they meet at a 90° angle. We can determine the angle between vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$; can that be generalized as well? We need to know one fact:

Theorem 2.7.1 — Cauchy-Schwarz Inequality. If $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, then

$$|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\| \, \|\mathbf{v}\|$$

The proof[2] is straightforward.

Applying this to $\|\mathbf{u}+\mathbf{v}\|^2 = (\mathbf{u}+\mathbf{v}) \cdot (\mathbf{u}+\mathbf{v})$ yields

$$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

which is the *triangle inequality* (which looks the same as the one we noted for $\mathbb{C}$ in our first class).

Definition 2.7.2 If $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$, then the angle θ between $\mathbf{u}$ and $\mathbf{v}$ is defined to be the number θ which satisfies:

- $\cos\theta = \dfrac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \, \|\mathbf{v}\|}$
- $0 \leq \theta \leq \pi$

(The first condition makes sense because of the Cauchy-Schwarz inequality, since this inequality implies that the number on the right hand side is always between -1 and 1. The second condition guarantees uniqueness of θ.)

■ **Example 2.7.3** The angle between $\mathbf{u} = (0,0,3,4,5)$ and $\mathbf{v} = (-1,1,-1,1,2)$ is θ, where

$$\cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \, \|\mathbf{v}\|} = \frac{11}{\sqrt{50}\sqrt{8}} = \frac{11}{20}.$$

We find $\theta = \arccos(11/20)$.[3] ■

Of particular interest:

- Two non-zero vectors $\mathbf{u}$, $\mathbf{v}$ are orthogonal if the angle between them is $\frac{\pi}{2}$ (or 90°). Now $\cos(\frac{\pi}{2}) = 0$, so by our formula, the numerator $\mathbf{u} \cdot \mathbf{v}$ has to be zero. That's where the orthogonality condition came from.

[2]For $x \in \mathbb{R}$, consider the quadratic function $q(x) = \|u+xv\|^2$. Expand the right hand side as $(u+xv)\cdot(u+xv)$, compute the discriminant 'b^2-4ac', which (as $q(x) \geq 0$ for all x) must satisfy $b^2-4ac \leq 0$. Simplifying this, one obtains the desired inequality.

[3]With the help of a calculator –which you'll never need in this course –, $\theta \simeq 0.988432 \simeq 56.6°$.

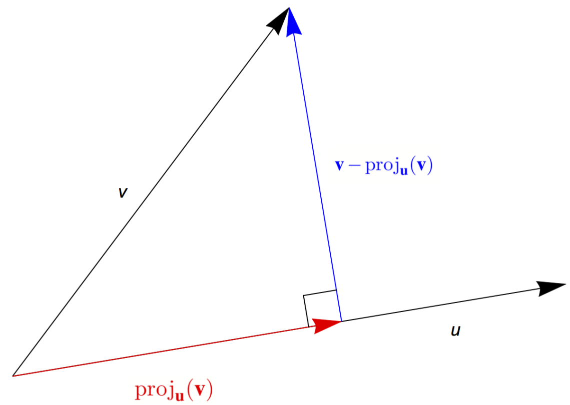


Figure 2.1: Orthogonal projection of **v** on **u**.

- Two vectors **u** and **v** are parallel if the angle between them is either 0° or π. Now $\cos(0) = 1$ and $\cos(\pi) = -1$, so in this case $\mathbf{u} \cdot \mathbf{v} = \pm\|\mathbf{u}\|\,\|\mathbf{v}\|$ attains its maximum value (in absolute terms).

2.8 Orthogonal Projections onto lines in $\mathbb{R}^n$

The idea of the orthogonal projection is: given two nonzero vectors **u** and **v**, then the *projection of* **v** *onto* **u**, denoted

$$\text{proj}_{\mathbf{u}}(\mathbf{v})$$

is the unique vector which satisfies

- $\text{proj}_{\mathbf{u}}(\mathbf{v})$ is parallel to **u** (so a scalar multiple of **u**)
- $\mathbf{v} - \text{proj}_{\mathbf{u}}(\mathbf{v})$ is orthogonal to **u** (so gives dot product zero).

As a result, we have *decomposed* **v** as a sum

$$\mathbf{v} = (\text{proj}_{\mathbf{u}}(\mathbf{v})) + (\mathbf{v} - \text{proj}_{\mathbf{u}}(\mathbf{v}))$$

of something parallel to **u** and something perpendicular to **u**, as in our picture.

So how do we find $\text{proj}_{\mathbf{u}}(\mathbf{v})$? It turns out to be easy. Using either trigonometry, or just solving directly from the above two conditions, you get:

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\mathbf{v} \cdot \mathbf{u}}{\|\mathbf{u}\|^2}\mathbf{u}.$$

(Careful: the quotient is just a scalar. To remember this: if you are projecting onto **u** your answer is always a multiple of **u**, which is also the term that appears in the denominator.)

■ **Example 2.8.1** Let $\mathbf{u} = (1,0,0)$ and $\mathbf{v} = (2,3,4)$. Write **v** as a sum of two vectors, one parallel to **u** and one orthogonal to **u**. (Presumably you could guess the answer to this one, but let's see what the formula does.) We calculate:

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\mathbf{v} \cdot \mathbf{u}}{\|\mathbf{u}\|^2}\mathbf{u} = \frac{2+0+0}{1^2+0^2+0^2}\begin{bmatrix}1\\0\\0\end{bmatrix} = \begin{bmatrix}2\\0\\0\end{bmatrix}$$

and $\mathbf{v} - \text{proj}_{\mathbf{u}}(\mathbf{v}) = (2,3,4) - (2,0,0) = (0,3,4)$, so our decomposition is

$$\mathbf{v} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \\ 4 \end{bmatrix}$$

and it's clear that the first is parallel to $\mathbf{u}$ while the second is orthogonal to $\mathbf{u}$ — and that this is the only possible pair of vectors that satisfy this. ■

■ **Example 2.8.2** Let $\mathbf{u} = (1,1)$ and $\mathbf{v} = (2,3)$. Then

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\mathbf{v} \cdot \mathbf{u}}{\|\mathbf{u}\|^2}\mathbf{u} = \frac{2+3}{1^2+1^2}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5/2 \\ 5/2 \end{bmatrix}$$

is the projection of $(2,3)$ onto $(1,1)$. ■

Orthogonal projection is used in UMTS (Universal Mobile Telecommunications System) to correct fuzziness and errors in signals and produce more reliable communications.

Another way of talking about orthogonal projection is to say that we are finding the closest point to $\mathbf{v}$ on the line from the origin in direction $\mathbf{u}$. So our next goal is to talk about lines and planes.

Problems

Those marked with a $\star$ have solutions at the back of the book. Try them yourself first.

Problem 2.1 Write down the zero vector in $\mathbb{R}^2$, $\mathbb{R}^3$ and $\mathbb{R}^5$.

Problem 2.2 Prove that $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ for any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$.

Problem 2.3 *If $A = (1,\ 2,\ 3)$, $B = (-5,\ -2,\ 5)$, $C = (-2,\ 8,\ -10)$ and D is the midpoint of $\overline{AB}$, find the coordinates of the midpoint of $\overline{CD}$.

Problem 2.4 Solve the following problems using the dot product.

a) Find all values of k such that $(k,\ k,\ 1)$ and $(k,\ 5,\ 6)$ are orthogonal.

b) *Find the angle between the vectors $(0,\ 3,\ -3)$ and $(-2,\ 2,\ -1)$.

c) If $A = (2,\ 4,\ 1)$, $B = (3,\ 0,\ 9)$ and $C = (1,\ 4,\ 0)$, find the angle $\angle BAC$.

Problem 2.5 *Solve the following problems.

a) If $\mathbf{u} = (2,\ 1,\ 3)$ and $\mathbf{v} = (3,\ 3,\ 3)$ find $\text{proj}_{\mathbf{v}}\,\mathbf{u}$.

b) If $\mathbf{u} = (3,\ 3,\ 6)$ and $\mathbf{v} = (2,\ -1,\ 1)$ find $\|\text{proj}_{\mathbf{v}}\,\mathbf{u}\|$.

3. Lines and Planes

In the last chapter we discussed the algebra and geometry of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$, and extended all the notions of vector addition, scalar multiplication, the dot product, angles and orthogonality to $\mathbb{R}^n$.

Now, let's consider lines and planes in $\mathbb{R}^2$ and $\mathbb{R}^3$. We'll see that while some ideas generalize easily to $\mathbb{R}^n$, others take more work; in fact, working out what, exactly, a reasonable analogue of a line or a plane in $\mathbb{R}^n$ is one of our goals in this course.

3.1 Describing Lines

A line in $\mathbb{R}^2$ or $\mathbb{R}^3$ is completely determined by specifying its direction and a point on the line. Since we prefer working with vectors, replace the point by its position vector.

So a line L going through the tip of $\mathbf{v}_0$ and such that $\mathbf{d}$ is a vector parallel to the direction of L can be described as the set

$$L = \{\mathbf{v}_0 + t\mathbf{d} | t \in \mathbb{R}\}.$$

That is, any point $\mathbf{v}$ on the line L can be written as $\mathbf{v} = \mathbf{v}_0 + t\mathbf{d}$ for some t.

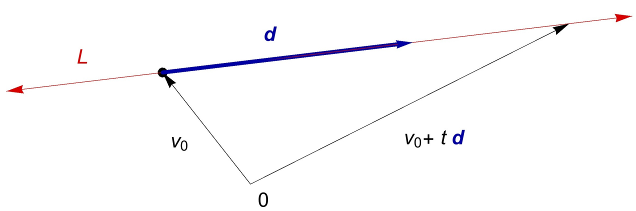


Figure 3.1: Vector parametric form for L

■ **Example 3.1.1** Consider the line $y = 3x+2$ in $\mathbb{R}^2$. We can get a parametric equation for this line by letting $x = t$ be the parameter and solving for y; this gives $x = t$ and $y = 3t+2$. In vector form, this is

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} t \\ 3t+2 \end{bmatrix} = \begin{bmatrix} 0+1t \\ 2+3t \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} + t \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

so the line is

$$L = \left\{ \begin{bmatrix} 0 \\ 2 \end{bmatrix} + t \begin{bmatrix} 1 \\ 3 \end{bmatrix} \middle| t \in \mathbb{R} \right\}.$$

Note: Another way to get a parametric equation for this line: it goes through the points $(0,2)$ and $(-\frac{2}{3},0)$, for example. So a direction vector is

$$\mathbf{d} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} - \begin{bmatrix} -2/3 \\ 0 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 2 \end{bmatrix}.$$

We can take $\mathbf{v}_0 = (0,2)$. So we get

$$L = \left\{ \begin{bmatrix} 0 \\ 2 \end{bmatrix} + t \begin{bmatrix} 2/3 \\ 2 \end{bmatrix} \middle| t \in \mathbb{R} \right\}.$$

NOTICE that our answer is not unique! It depends on our choices. BOTH of our answers are completely correct (check with a sketch!). ■

CAUTION!! We often use the variable t for the parameter. But if you are comparing two different lines, you must use different letters to represent parameters on different lines!

■ **Example 3.1.2** Find the point of intersection of $L = \{t(1,2) | t \in \mathbb{R}\}$ and $L' = \{(0,1)+t(3,0) | t \in \mathbb{R}\}$.

WRONG METHOD: Set $t(1,2) = (0,1)+t(3,0)$ and solve for t.

This doesn't give an answer for t; but that's only because the two lines don't arrive at the point of intersection at the same time t. The lines still intersect!

CORRECT METHOD: Find parameters s and t such that $t(1,2) = (0,1)+s(3,0)$. This gives $t = \frac{1}{2}$ and $s = \frac{1}{6}$, and the point of intersection is thus $(\frac{1}{2},1)$. ■

The form

$$L = \{\mathbf{v}_0 + t\mathbf{d} | t \in \mathbb{R}\}$$

for a line in $\mathbb{R}^n$ is called the *vector form* or *parametric form*. In $\mathbb{R}^2$ (but NOT $\mathbb{R}^3$), you can describe a line by an equation like $ax+by = c$; this is called the *point-normal* or *Cartesian form*.

You can also expand the parametric form in coordinates: if $\mathbf{v}_0 = (a,b,c)$ and $\mathbf{d} = (d_1,d_2,d_3)$ then our line is the set of all $\mathbf{v} = (x_1,x_2,x_3)$ such that

$$\begin{aligned} x_1 &= a+td_1 \\ x_2 &= b+td_2 \\ x_3 &= c+td_3 \end{aligned}$$

for some $t \in \mathbb{R}$.

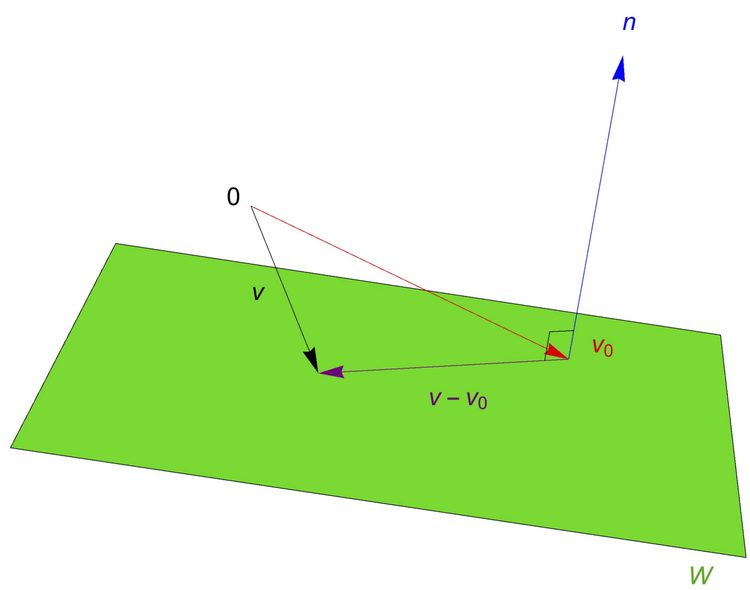


Figure 3.2: The cartesian equation of a plane: point-normal form

3.2 About the Geometry of Lines

There is only one line in $\mathbb{R}$: all of $\mathbb{R}$.

Given two distinct lines in $\mathbb{R}^2$, they are either parallel or they intersect.

Given two distinct lines in $\mathbb{R}^3$, they could be parallel, or they could intersect, or they could be skew. In the first two cases, they are contained in a unique plane; in the third case there is no plane containing both of them (but you can find two parallel planes such that each contains one of the lines).

3.3 Describing Planes in $\mathbb{R}^3$

Recall that planes in $\mathbb{R}^3$ are described by an equation in point-normal or Cartesian form. That is, a plane is described as the set of all points (x,y,z) such that

$$ax+by+cz=d$$

where $\mathbf{n}=(a,b,c)$ is a *normal vector* to the plane and $d\in\mathbb{R}$.

How did we get this equation? Look below. If $\mathbf{v}_0$ is some point on the plane, then $\mathbf{v}$ is on the plane iff $(\mathbf{v}-\mathbf{v}_0)\cdot\mathbf{n}=0$:

So the plane through $\mathbf{v}_0$ with normal vector $\mathbf{n}$ is

$$W=\{\mathbf{v}\in\mathbb{R}^3|(\mathbf{v}-\mathbf{v}_0)\cdot\mathbf{n}=0\}.$$

Example 3.3.1 The plane through $\mathbf{v}_0 = (1,0,3)$ with normal vector $\mathbf{n} = (-1,1,2)$ is

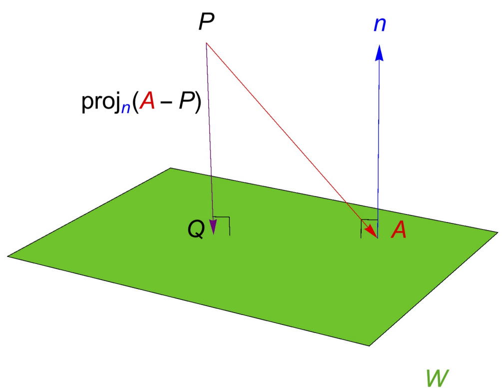
$$\begin{aligned}W &= \{\mathbf{v} \in \mathbb{R}^3 | (\mathbf{v} - \mathbf{v}_0) \cdot \mathbf{n} = 0\} \\ &= \{(x,y,z) | ((x,y,z) - (1,0,3)) \cdot (-1,1,2) = 0\} \\ &= \{(x,y,z) | -(x-1) + (y-0) + 2(z-3) = 0\} \\ &= \{(x,y,z) | -x + y + 2z = 5\}\end{aligned}$$

(Note that the coefficients in the Cartesian equation give you a normal vector.) ■

The normal vector is handy for many things.

Problem 3.3.2 Find the distance from the point $P = (1,2,3)$ to the plane W with Cartesian equation $3x - 4z = -1$.

Solution Let A be any point on the plane, say $A = (1,0,1)$, and Q be the (unknown) closest point on the plane to P. As the diagram suggests[a], we want the length of the projection of $A - P$ onto the normal vector.



Finding the distance of P to the plane W.

So

$$\begin{aligned}\|P - Q\| &= \|\text{proj}_{\mathbf{n}}(P - A)\| \\ &= \|\text{proj}_{(3,0,-4)}(0,2,2)\| \\ &= \|\frac{0+0-8}{3^2+(-4)^2}(3,0,-4)\| \\ &= \frac{8}{25}\|(3,0,-4)\| \\ &= \frac{8}{25}\sqrt{25} = \frac{8}{5}.\end{aligned}$$

[a] And here's a proof, for the properly skeptical reader: if Q' is any other point on W, $\|P - Q'\|^2 = \|P - Q + Q - Q'\|^2 = \|P - Q\|^2 + \|Q - Q'\|^2$ (because $P - Q$ and $Q - Q'$ are perpendicular, so the Pythagorean theorem applies). Hence $\|P - Q'\|^2 \geq \|P - Q\|^2$.

Let's tackle a more straightforward problem: the intersection of two planes.

Problem 3.3.3 Find the intersection of the planes $x+y+z=3$ and $x-y-z=2$.
Solution We need to find all (x,y,z) which satisfy both equations. Subtracting the second equation from the first yields $2y+2z=1$ or $y=\frac{1}{2}-z$; then from the first we have $x=3-(\frac{1}{2}-z)-z=\frac{5}{2}$. But z can be anything; in fact we can take z to our parameter (and so now call it t):

$$x=\frac{5}{2},\quad y=\frac{1}{2}-t,\quad z=t$$

which in vector form is the line

$$L=\left\{\begin{bmatrix}5/2\\1/2\\0\end{bmatrix}+t\begin{bmatrix}0\\-1\\1\end{bmatrix}\middle|t\in\mathbb{R}\right\}.$$

3.4 Geometry of Planes in $\mathbb{R}^3$

We define the "angle" between two planes to be the angle between their normal vectors. It equals the angle of the "wedge" that they make (although proving that takes some thought).

Now, to compare with the case of lines:

The only plane in $\mathbb{R}^2$ is all of $\mathbb{R}^2$.

Given two distinct planes in $\mathbb{R}^3$, they are either parallel or they intersect.

Makes you wonder about $\mathbb{R}^4$, doesn't it?

That said, we don't have a good way to describe a plane in $\mathbb{R}^4$ (yet). A parametric equation in one variable gives a line, in any $\mathbb{R}^n$. But something fishy is happening with the normal forms:

n	Equation in $\mathbb{R}^n$	Resulting Geometric Object
1	$ax=b$	point
2	$ax+by=c$	line
3	$ax+by+cz=d$	plane
4	$ax_1+bx_2+cx_3+dx_4=e$	??

Idea: one equation is cutting down one *degree of freedom* (later: *dimension*), so the result is always an object of *dimension* $n-1$ (which is called a *hyperplane* in dimensions bigger than 3).

Our answer will eventually be: Since intersecting two planes in $\mathbb{R}^3$ (usually) gives you a line, intersecting two hyperplanes in $\mathbb{R}^4$ should give you a plane.

This is something we'll be coming back to over the next few weeks.

But for now: let's get back to solid ground (or rather, $\mathbb{R}^3$) and discuss how to produce normal vectors to planes fairly easily.

3.5 Cross products in $\mathbb{R}^3$

Let's use the notation:

$$\hat{\imath}=\begin{bmatrix}1\\0\\0\end{bmatrix},\quad \hat{\jmath}=\begin{bmatrix}0\\1\\0\end{bmatrix},\quad \hat{k}=\begin{bmatrix}0\\0\\1\end{bmatrix}.$$

Then $(x,y,z)=x\hat{\imath}+y\hat{\jmath}+z\hat{k}$.

In the following, we also use notation from the *determinant* which for now is just very convenient.

The *cross product* of $\mathbf{u} = (x,y,z)$ and $\mathbf{v} = (x',y',z')$ is a new vector, denoted $\mathbf{u} \times \mathbf{v}$, which is calculated as follows:

$$\begin{aligned} \mathbf{u} \times \mathbf{v} &= \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ x & y & z \\ x' & y' & z' \end{vmatrix} \\ &= \left(yz' - y'z, -(xz' - x'z), xy' - x'y\right) \\ &= \left(\begin{vmatrix} y & z \\ y' & z' \end{vmatrix}, - \begin{vmatrix} x & z \\ x' & z' \end{vmatrix}, \begin{vmatrix} x & y \\ x' & y' \end{vmatrix} \right) \end{aligned}$$

Problem 3.5.1 Find $(1,2,3) \times (4,5,6)$.
Solution Write this as a determinant (or at least on top of each other) and then solve:

$$\begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{vmatrix} = (12-15, -(6-12), 5-8) = (-3,6,-3).$$

Problem 3.5.2 Find $(4,5,6) \times (1,2,3)$.
Solution Same process:

$$\begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{vmatrix} = (15-12, -(12-6), 8-5) = (3,-6,3)$$

HA! We got the NEGATIVE.

Notice, though, that

$$(1,2,3) \cdot (3,-6,3) = 3 - 12 + 9 = 0, \quad (4,5,6) \cdot (3,-6,3) = 12 - 30 + 18 = 0.$$

None of this was just luck.

Theorem 3.5.3 — Properties of the Cross Product. Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$. Then

- $\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$
- $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{u} = 0$
- $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{v} = 0$
- $(\mathbf{u} + \mathbf{v}) \times \mathbf{w} = \mathbf{u} \times \mathbf{w} + \mathbf{v} \times \mathbf{w}$
- $\|\mathbf{u} \times \mathbf{v}\| = \|\mathbf{u}\| \, \|\mathbf{v}\| \sin(\theta)$ where $0 \leq \theta \leq \pi$ is the angle between $\mathbf{u}$ and $\mathbf{v}$. This is in fact the area of the parallelogram with sides $\mathbf{u}$ and $\mathbf{v}$.

BUT, usually: $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) \neq (\mathbf{u} \times \mathbf{v}) \times \mathbf{w}$; the cross product is neither commutative nor associative. (Eg: $\hat{\imath} \times (\hat{k} \times \hat{k}) = \mathbf{0}$ but $(\hat{\imath} \times \hat{k}) \times \hat{k} = -\hat{\imath}$.)

Consequently: if two vectors are parallel, then their cross product is zero. Otherwise, their cross product is one of the two vectors orthogonal to both $\mathbf{u}$ and $\mathbf{v}$ and of length $\|\mathbf{u}\| \, \|\mathbf{v}\| \sin\theta$, where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$. The cross product is used in Physics for measuring torque, for example, and figuring out the direction of the answer uses the right-hand rule. We also often memorize $\hat{\imath} \times \hat{\jmath} = \hat{k}$ and its cyclic permutations.

3.6 First application of cross product: finding normal vectors

The cross product gives a normal vector to the plane parallel to two vectors $\mathbf{u}$ and $\mathbf{v}$.

Problem 3.6.1 Find an equation of the plane containing the three points $A = (1,2,3)$, $B = (1,0,0)$ and $C = (0,1,1)$.

Solution The vectors $\vec{BA} = (0,2,3)$ and $\vec{BC} = (-1,1,1)$ are both parallel to the plane, so a normal vector to the plane must be orthogonal to each of these. So take their cross product:

$$\begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ 0 & 2 & 3 \\ -1 & 1 & 1 \end{vmatrix} = (-1,-3,2)$$

(Check your answer! This should be orthogonal to $\vec{BA}$ and $\vec{BC}$.)

So an equation for the plane has the form

$$-x - 3y + 2z = d$$

and plugging in the point B, for example, yields $d = -1$.

3.7 Second Application: volumes of parallelepipeds (Scalar Triple Product)

Theorem 3.7.1 — Volume of a parallelepiped. The volume of the parallelepiped with sides $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^3$ is

$$|(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}|.$$

In particular, the order of the vectors doesn't matter.

We can prove this using more trigonometry: the volume of the parallelepiped is the area of the base (a parallelogram) times the height; the area of the base is the norm of the cross product of two of the vectors ($\|\mathbf{u} \times \mathbf{v}\|$), and the height is going to be $\|\mathbf{w}\| \cos(\theta)$ where θ is the angle between $\mathbf{w}$ and a normal vector to the base.

Problem 3.7.2 Find the volume of the parallelepiped with sides $\mathbf{u} = (1,0,1)$, $\mathbf{v} = (1,2,2)$ and $\mathbf{w} = (10,0,0)$.

Solution We calculate:

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ 1 & 0 & 1 \\ 1 & 2 & 2 \end{vmatrix} = (-2,-1,2)$$

so

$$(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} = (-2,-1,2) \cdot (10,0,0) = -20$$

so the volume is 20.

So what does it mean if $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} = 0$? Zero volume of a parallelepiped means it wasn't really a parallelepiped at all — the three vectors must all lie in a plane. We say such vectors are *coplanar* or *linearly dependent* (key phrase for later).

3.8 Final remarks about lines, planes, and higher-dimensional objects in $\mathbb{R}^n$

These are just some thoughts to preview some of the ideas we want to explore over the next several weeks. We have great ways to describe lines and planes in $\mathbb{R}^2$ and $\mathbb{R}^3$, but we have an inkling that there must be "2-dimensional" objects in $\mathbb{R}^4$ which we can't describe by either of the two methods described so far (parametric equations or normal equations).

Or can we?

To get a line, we gave ourselves one parameter (one degree of freedom). If we give ourselves two parameters, for two different direction vectors, this describes a plane. That is

$$W = \{\mathbf{v}_0 + s\mathbf{d}_1 + t\mathbf{d}_2 | s, t \in \mathbb{R}\}$$

describes the plane with normal vector $\mathbf{d}_1 \times \mathbf{d}_2$ and going through the point $\mathbf{v}_0$.

We will explore statements like this in the course of understanding subspaces of general vector spaces.

Problems

Problem 3.1 Solve the following problems using the cross and/or dot products.

a) If $\mathbf{u} = (3,\ -1,\ 4)$ and $\mathbf{u} = (-1,\ 6,\ -5)$, find $\mathbf{u} \times \mathbf{v}$.

b) *Find all vectors in $\mathbb{R}^3$ which are orthogonal to both $(-1,1,5)$ and $(2,1,2)$.

c) If $\mathbf{u} = (4,\ -1,\ 7)$, $\mathbf{v} = (2,\ 1,\ 2)$ and $\mathbf{w} = (-1,\ -2,\ 3)$, find $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w}$.

d) *If $\mathbf{u} = (-4,\ 2,\ 7)$, $\mathbf{v} = (2,\ 1,\ 2)$ and $\mathbf{w} = (1,\ 2,\ 3)$, find $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$.

Problem 3.2 Solve the following problems using the appropriate products.

a) Find the area of the parallelogram determined by the vectors $\mathbf{u} = (1,\ -1,\ 0)$ and $\mathbf{v} = (2,\ -3,\ 1)$.

b) *Find the area of the triangle with vertices $A = (-1,\ 5,\ 0)$, $B = (1,\ 0,\ 4)$ and $C = (1,\ 4,\ 0)$.

c) Find the area of the triangle whose vertices are $P = (1,\ 1,\ -1)$, $Q = (2,\ 0,\ 1)$ and $R = (1,\ -1,\ 3)$.

d) *Find the volume of the parallelepiped determined by $\mathbf{u} = (1,\ 1,\ 0)$, $\mathbf{v} = (1,\ 0,\ -1)$ and $\mathbf{w} = (1,\ 1,\ 1)$.

e) Find the volume of the parallelepiped determined by $\mathbf{u} = (1,-2,3)$, $\mathbf{v} = (1,3,1)$ and $\mathbf{w} = (2,1,2)$.

Problem 3.3 Solve the following problems.

a) *Find the point of intersection of the plane with Cartesian equation $2x + 2y - z = 5$, and the line with parametric equations $x = 4 - t,\ y = 13 - 6t,\ z = -7 + 4t$.

b) *If L is the line passing through $(1,\ 1,\ 0)$ and $(2,\ 3,\ 1)$, find the point of intersection of L with the plane with Cartesian equation $x + y - z = 1$.

c) Find the point where the lines with parametric equations $x = t - 1$, $y = 6 - t$, $z = -4 + 3t$ and $x = -3 - 4t$, $y = 6 - 2t$, $z = -5 + 3t$ intersect.

d) *Do the planes with Cartesian equations $2x - 3y + 4z = 6$ and $4x - 6y + 8z = 11$ intersect?

e) Find the line of intersection of the planes with Cartesian equations $5x + 7y - 4z = 8$ and $x - y = -8$.

f) *Find the line of intersection of the planes with Cartesian equations $x + 11y - 4z = 40$ and $x - y = -8$.

Problem 3.4 Solve the following problems.

a) Find the distance from the point $(0,\ -5,\ 2)$ to the plane with Cartesian equation $2x + 3y + 5z = 2$.

b) *Find the distance from the point $(-2,\ 5,\ 9)$ to the plane with Cartesian equation $6x + 2y - 3z = -8$

c) Find the distance from the point $(5,\ 4,\ 7)$ to the line containing the points $(3,\ -1,\ 2)$ and $(3,\ 1,\ 1)$.

d) *Find the distance from the point $(8,\ 6,\ 11)$ to the line containing the points $(0,\ 1,\ 3)$ and $(3,\ 5,\ 4)$.

e) Find angle between the planes with Cartesian equations $x - z = 7$ and $y - z = 234$

Problem 3.5 Find the scalar *and* vector parametric forms for the following lines:

a) The line containing $(3,\ -1,\ 4)$ and $(-1,\ 5,\ 1)$.

b) *The line containing $(-5,\ 0,\ 1)$ and which is parallel to the two planes with Cartesian equations $2x - 4y + z = 0$ and $x - 3y - 2z = 1$

c) The line passing through $(1,\ 1,\ -1)$ and which is perpendicular to the plane with Cartesian equation $2x - y + 3z = 4$

Problem 3.6 Find a Cartesian equation for each of the following planes:

a) The plane containing $(3,-1,4)$, $(-1,5,1)$ and $(0,2,-2)$.

b) *The plane parallel to the vector $(1,1,-2)$ and containing the points $(1,5,18)$ and $(4,2,-6)$

c) The plane passing through the points $(2,1,-1)$ and $(3,2,1)$, and parallel to the x–axis.

d) *The plane containing the two lines $\{(t-1, 6-t, -4+3t) \mid t \in \mathbb{R}\}$ and $\{(-3-4t, 6+2t, 7+5t) \mid \in \mathbb{R}\}$

e) The plane which contains the point (-1, 0, 2) and the line of intersection of the two planes $3x + 2y - z = 5$ and $2x + y + 2z = 1$.

f) ⋆The plane containing the point $(1,-1,2)$ and the line $\{(4,-1+2t,2+t) \mid t \in \mathbb{R}\}$.

g) The plane through the origin parallel to the two vectors $(1,1,-1)$ and $(2,3,5)$.

h) ⋆The plane containing the point $(1,\ -7,\ 8)$ which is perpendicular to the line $\{(2+2t,7-4t,-3+t \mid t \in \mathbb{R}\}$.

i) The plane containing the point $(2,\ 4,\ 3)$ and which is perpendicular to the planes with Cartesian equations $x+2y-z=1$ and $3x-4y=2$.

Problem 3.7 Find a vector parametric form for the planes with Cartesian equations given as follows. (i.e. find some $a \in H$ and two non-zero, non-parallel vectors $\mathbf{u},\mathbf{v} \in \mathbb{R}^3$, parallel to the plane H. Then $H = \{a+s\mathbf{u}+t\mathbf{v} \mid s,t \in \mathbb{R}\}$.)

a) $x-y-z=3$

b) ⋆$x-y-2z=4$

c) $2x-y+z=5$

d) $y+2x=-3$

e) $x-y+2z=0$

f) $x+y+z=-1$

Problem 3.8 ⋆Let $\mathbf{u},\mathbf{v}$ and $\mathbf{w}$ be any vectors in $\mathbb{R}^3$. Determine which of the following statements could be false, and give an example to illustrate each of your answers.

(1) $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$
(2) $\mathbf{u}\times\mathbf{v} = \mathbf{v}\times\mathbf{u}$
(3) $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w}) = \mathbf{v}\cdot\mathbf{u}+\mathbf{w}\cdot\mathbf{u}$
(4) $(\mathbf{u}+2\mathbf{v})\times\mathbf{v} = \mathbf{u}\times\mathbf{v}$
(5) $(\mathbf{u}\times\mathbf{v})\times\mathbf{w} = \mathbf{u}\times(\mathbf{v}\times\mathbf{w})$

Problem 3.9 ⋆Let $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ be vectors in $\mathbb{R}^3$. Which of the following statements are (always) true? Explain your answers, including giving examples to illustrate statements which could be false.

(i) $(\mathbf{u}\times\mathbf{v})\cdot\mathbf{v} = 0$.
(ii) $(\mathbf{v}\times\mathbf{u})\cdot\mathbf{v} = -1$.
(iii) $(\mathbf{u}\times\mathbf{v})\cdot\mathbf{w}$ is the volume of the of the parallelepiped determined by $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$.
(iv) $\|\mathbf{u}\times\mathbf{v}\| = \|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta$ where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$.
(v) $|\mathbf{u}\cdot\mathbf{v}| = \|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta$ where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$.

Problem 3.10 Prove the identity $\mathbf{u}\times(\mathbf{v}\times\mathbf{w}) = (\mathbf{u}\cdot\mathbf{w})\mathbf{v}-(\mathbf{u}\cdot\mathbf{v})\mathbf{w}$ for all $\mathbf{u},\mathbf{v},\mathbf{w} \in \mathbb{R}^3$, as follows: Denote the difference between the left hand side of the identity and the right hand side by $D(\mathbf{u},\mathbf{v},\mathbf{w})$.

First note that, by properties of the dot and cross products, for all $k \in \mathbb{R}$, $\mathbf{u},\mathbf{v},\mathbf{w},\mathbf{u}',\mathbf{v}',\mathbf{w}' \in \mathbb{R}^3$, it's easy to see that the following identities hold.

(i) $D(k\mathbf{u}+\mathbf{u}',\mathbf{v},\mathbf{w}) = kD(\mathbf{u},\mathbf{v},\mathbf{w})+D(\mathbf{u}',\mathbf{v},\mathbf{w})$

(ii) $D(\mathbf{u},k\mathbf{v}+\mathbf{v}',\mathbf{w}) = kD(\mathbf{u},\mathbf{v},\mathbf{w})+D(\mathbf{u},\mathbf{v}',\mathbf{w})$

(iii) $D(\mathbf{u},\mathbf{v},k\mathbf{w}+\mathbf{w}') = kD(\mathbf{u},\mathbf{v},\mathbf{w})+D(\mathbf{u},\mathbf{v},\mathbf{w}')$

(iv) Also: $D(\mathbf{u},\mathbf{v},\mathbf{w}) = -D(\mathbf{u},\mathbf{w},\mathbf{v})$

(Properties (i)–(iii) are summarized by saying that "*D is linear in each argument*". More on this when we get to linear transformations.)

Now since every vector in $\mathbb{R}^3$ is a linear combination of $\hat{\imath} = (1,0,0)$, $\hat{\jmath} = (0,1,0)$ and $\hat{k} = (0,0,1)$, by the identities above, it suffices to check that $D(\mathbf{u},\mathbf{v},\mathbf{w}) = 0$ when $\mathbf{u}$ is $\hat{\imath}$, $\hat{\jmath}$ and $\hat{k}$, and $\mathbf{v}$ and $\mathbf{w}$ are one the 6 pairs of two distinct choices from $\{\hat{\imath},\hat{\jmath},\hat{k}\}$. Restricted to these (18) choices, it's easy to see $D(\mathbf{u},\mathbf{v},\mathbf{w})$ is zero unless $\mathbf{u}$ is $\mathbf{v}$ or $\mathbf{w}$, so we really only need to check $D(\mathbf{u},\mathbf{u},\mathbf{w}) = 0$ in the 6 cases where $\mathbf{u} \in \{\hat{\imath},\hat{\jmath},\hat{k}\}$ and $\mathbf{w} \in \{\hat{\imath},\hat{\jmath},\hat{k}\} \setminus \{\mathbf{u}\}$. This amounts to 6 easy computations. Having checked those, you're done!

II

Vector Spaces

Given that the algebra of $\mathbb{R}^2$ and $\mathbb{R}^3$ extended so easily to $\mathbb{R}^n$, we ask ourselves: what other kinds of mathematical objects behave (algebraically speaking) just like $\mathbb{R}^n$? We formulate this question precisely in the first chapter, and proceed to uncover many very familiar examples.

Photo: Ralph Nevins. Montpellier, France

4. Vector Spaces

So far, we have established that it isn't too hard to generalize our notion of *vectors* to include elements of $\mathbb{R}^n$. We still call them *vectors* even though we can no longer quite imagine them as arrows in space.

As we worked through vector geometry, we also saw a number of problems coming up, principally among them: what are the higher-dimensional analogues of lines and planes? How can you describe them?

We will tackle this problem next, but it turns out that the best way to do this is to step back and see just how far we can generalize this notion of *vector spaces* — was $\mathbb{R}^n$ the only set that behaves substantially like $\mathbb{R}^2$ and $\mathbb{R}^3$, or are there many more mathematical objects out there that, if we look at them in the right way, are just like vectors, too?

4.1 A first example

So far: we agree that the elements of $\mathbb{R}$, $\mathbb{R}^2$ and $\mathbb{R}^3$ are *geometric vectors*.

We also agree that we can call elements of $\mathbb{R}^n$, for $n \geq 4$, *vectors*.

What we are looking for next: non-geometric, non-$\mathbb{R}^n$ vectors.

■ **Example 4.1.1 Spaces of Equations**

Consider three equations, which we name E_1, E_2 and E_3:

$$\begin{aligned} E_1 : \quad & x - y - z = -1 \\ E_2 : \quad & 2x - y + z = 1 \\ E_3 : \quad & -x + 2y + 4z = 4 \end{aligned}$$

We can create new equations from these ones; for example,

$$E_4 = E_2 - 2E_1 \quad : \quad y + 3z = 3$$

or

$$E_1 + E_3 \quad : \quad y + 3z = 3$$

So we can even say

$$E_2 - 2E_1 = E_1 + E_3$$

and it is legitimate to rewrite this as:

$$3E_1 - E_2 + E_3 = 0$$

where "0" stands for the equation "$0 = 0$".

Well, what are we really saying here?

- We can **add** two equations to get another equation.
- We can **multiply** an equation **by a scalar** to get another equation.
- There's a **zero equation** given by $0 = 0$. Let's call it E_0.
- Equations have **negatives**:

$$-E_1 : -x + y + z = 1$$

  Why is this the negative? Because now $(E_1) + (-E_1) = E_0$.
- The **usual rules of arithmetic** hold:
  - $E_1 + E_2 = E_2 + E_1$
  - $E_1 + (E_2 + E_3) = (E_1 + E_2) + E_3$
  - $k(E_1 + E_2) = kE_1 + kE_2$
  - $(k + l)E_1 = kE_1 + lE_1$
  - $(kl)E_1 = k(lE_1)$
  - $1E_1 = E_1$

In other words: these are **exactly** the properties of $\mathbb{R}^n$ that we identified last week! So even though we have no reasonable way (yet) of writing "equations" as n-tuples of numbers, algebraically we recognize that they act just like vectors do. ■

! Major idea #1 (this chapter): there are lots of different vector spaces besides $\mathbb{R}^n$.

We can specialize this a bit more (which also helps us start to see the value in this approach):

Consider the space $\mathscr{E}$ of all equations "obtainable from" (also say: *generated by* or *spanned by*) E_1, E_2, and E_3:

$$\mathscr{E} = \{k_1E_1 + k_2E_2 + k_3E_3 \mid k_i \in \mathbb{R}\}.$$

(This is the set of *all* linear combinations of those three equations!) In fact, $\mathscr{E}$ itself acts like a space of vectors (in the sense above, and to be made precise below).

The kinds of questions you'd want to ask: Is there an equation of the form $y = y_0$ in $\mathscr{E}$? That is, can we solve for y?[1] (Or for x, for that matter.)

■ **Example 4.1.2** Can we find $a, b, y_0 \in \mathbb{R}$ so that the equation $y = y_0$ equals $aE_1 + bE_2$? Answer: NO. Why? The coefficient of x in $aE_1 + bE_2$ is $a + 2b$ and the coefficient of z in $aE_1 + bE_2$ is $-a + b$. The only way these can both be zero is if a and b are zero (check); but then $aE_1 + bE_2 = E_0$, the zero equation. So we can't get an equation like $y = y_0$. ■

[1] The connection is that: to solve for y, what you're honestly doing is taking linear combinations of equations.

(We of course want to answer tougher questions than that, but our technique (Gaussian elimination, Chapter 11) will be based entirely on this idea of taking linear combinations of equations.)

Stoichiometry is the science of figuring out how to produce a given chemical as a result of reactions of other, easier to obtain, chemicals. In stoichiometry, you let $\mathscr{E}$ be the set of all known chemical reactions and want to choose a best (linear) combination of reactions which will result in what you wanted.

! Major idea #2: (next two chapters): Subspaces and spanning sets. To answer particular problems, we typically need to understand the subspace spanned by some vectors.

4.2 So what do we really need?

Idea: *vectors* don't have to be geometric, or even n-tuples. But we want things that act like our known examples of vectors *algebraically*. (Please note: we're going to ignore the geometry (like the dot product) for quite a while; we're now just focussing on the algebra.)

So we need:

- V : a set of "vectors" of some sort
- a rule for "adding" two "vectors" (denoted with $+$)
- a rule for "scalar multiplication" of a "vector" by an element $c \in \mathbb{R}$

such that all 10 of the following *axioms* hold:

Closure (These two axioms guarantee that V is "big enough")

(1) The sum of two vectors should again be a vector. (That is, if $\mathbf{u}, \mathbf{v} \in V$ then $\mathbf{u}+\mathbf{v} \in V$.)

(2) Any scalar multiple of a vector should again be a vector. (That is, if $\mathbf{u} \in V$ and $c \in \mathbb{R}$ then $c\mathbf{u} \in V$.)

Existence (These two axioms guarantee that V has the 'basics')

(3) There should be a zero vector $\mathbf{0}$ in V; it must satisfy $\mathbf{0}+\mathbf{u} = \mathbf{u}$ for all $\mathbf{u} \in V$.

(4) Every vector in V should have a negative in V. That is, given $\mathbf{u} \in V$, there should exist another vector $-\mathbf{u} \in V$ such that $\mathbf{u}+(-\mathbf{u}) = \mathbf{0}$.

Arithmetic properties (These axioms guarantee that operations behave the way they did in $\mathbb{R}^n$)

For any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and any $c, d \in \mathbb{R}$:

(5) $\mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}$

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

(7) $c(\mathbf{u}+\mathbf{v}) = c\mathbf{u}+c\mathbf{v}$

(8) $(c+d)\mathbf{u} = c\mathbf{u}+d\mathbf{u}$

(9) $c(d\mathbf{u}) = (cd)\mathbf{u}$

(10) $1\mathbf{u} = \mathbf{u}$.

Definition 4.2.1 Any set V with two operations as above, satisfying these 10 axioms is called a *vector space*.

Note that even though they are not axioms, it is indeed true that the zero vector is the result of scaling any vector by $0 \in \mathbb{R}$, and the negative of a vector is given by multiplying it by $-1 \in \mathbb{R}$. See problem 4.14.

4.3 Examples of Vector Spaces

Example 4.3.1 $\mathbb{R}$, $\mathbb{R}^2$, …, $\mathbb{R}^n$ are all vector spaces, with the usual operations. ■

Example 4.3.2 $\mathscr{E}$ is a vector space, as is the set of *all* linear equations in n variables, with the usual operations. ■

Example 4.3.3 The set $V = \{\mathbf{0}\}$, with operations given by the rule $\mathbf{0}+\mathbf{0}=\mathbf{0}$, and $c\cdot\mathbf{0}=\mathbf{0}$, is a vector space : the *zero vector space*. (But the empty set is not a vector space because it doesn't contain $\mathbf{0}$!) ■

Example 4.3.4 The set $V = \{(x,2x)|x\in\mathbb{R}\}$, with the standard operations from $\mathbb{R}^2$ is a vector space because:

- CLOSURE (1) Let $\mathbf{u},\mathbf{v}\in V$. Then $\mathbf{u}=(x,2x)$ for some $x\in\mathbb{R}$ and $\mathbf{v}=(y,2y)$ for some $y\in\mathbb{R}$. So $\mathbf{u}+\mathbf{v}=(x+y,2x+2y)=(x+y,2(x+y))$ and this is in V because it is of the form $(z,2z)$ with $z=x+y\in\mathbb{R}$.
- CLOSURE (2) Let $\mathbf{u}=(x,2x)$ and $c\in\mathbb{R}$. Then $c\mathbf{u}=(cx,c(2x))=(cx,2(cx))$ which is again in V because it has the form $(z,2z)$ with $z=cx\in\mathbb{R}$.
- EXISTENCE (3) The zero vector of $\mathbb{R}^2$ is $(0,0)$; this lies in V and moreover since it satisfies $\mathbf{0}+\mathbf{u}=\mathbf{u}$ for any $\mathbf{u}\in\mathbb{R}^2$, it certainly satisfies that for the $\mathbf{u}\in V$. So it is a zero vector and it lies in V.
- EXISTENCE (4) The negative of $\mathbf{u}=(x,2x)$ is $(-x,-2x)=(-x,2(-x))$ since $(x,2x)+(-x,2(-x))=(0,0)$ and we see that $(-x,2(-x))\in V$. So it's the negative and it lies in V.
- ALGEBRAIC PROPERTIES (5)-(10): these are all ok because they all work for ANY $\mathbf{u},\mathbf{v},\mathbf{w}$ in $\mathbb{R}^2$, so they certainly work for any vectors $\mathbf{u},\mathbf{v},\mathbf{w}$ in V.

So this is a vector space. ■

Example 4.3.5 The set $V=\{(x,x+2)|x\in\mathbb{R}\}$, with the usual operations of $\mathbb{R}^2$ is NOT A VECTOR SPACE.

To show that V is not a vector space, it is enough to give just ONE example of just ONE axiom that fails even ONCE. (Because being a vector space means that all those axioms are always true.)

But let's look at ALL the axioms, just to see how this fails in lots of different ways.

- CLOSURE (1) Take $(x,x+2)$ and $(y,y+2)$. Then their sum is $(x+y,x+y+4)$ which does NOT have the correct form $(z,z+2)$, for any z. That is, when you add two elements in this set, you end up OUTSIDE the set. It is NOT CLOSED UNDER ADDITION.
- CLOSURE (2) Take $c\in\mathbb{R}$ and $\mathbf{u}=(x,x+2)\in V$. Then $c(x,x+2)=(cx,cx+2c)$ so whenever $c\neq 1$, $c\mathbf{u}\notin V$. So V is NOT CLOSED UNDER SCALAR MULTIPLICATION.
- EXISTENCE (3) The zero vector $(0,0)$ is not in V, since you can't have $(0,0)=(z,z+2)$ for any $z\in\mathbb{R}$. (See exercise 4.1 (h).)
- EXISTENCE (4) The negative of $\mathbf{u}=(x,x+2)$ is $-\mathbf{u}=(-x,-x-2)$, which doesn't lie in V.
- ALGEBRAIC PROPERTIES (5)-(10): these are all fine, because the actual operations are fine; it's the subset of $\mathbb{R}^2$ we chose that was "bad".

■

! Remember this: a subset of a vector space can't be a vector space unless it contains the zero vector. So lines and planes that don't pass through the origin are *not vector spaces*.

Definition 4.3.6 A *matrix* is a table of numbers, usually enclosed in square brackets. It's size is $m \times n$ if it has m rows and n columns. For instance,

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

is a 2×3 matrix. Two matrices of the same size can be added componentwise, like

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 6 & 8 \\ 10 & 12 \end{bmatrix}$$

and we can scalar multiply them componentwise as well, like

$$2\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix}.$$

■ **Example 4.3.7** $V = M_{22}(\mathbb{R}) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} | a,b,c,d \in \mathbb{R} \right\}$, with the above operations.

We check the axioms: closure is fine (sum and scalar multiples give you 2×2 matrices again); existence: (3) $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \in M_{22}(\mathbb{R})$ and clearly adding this to a matrix A gives you A back again, and (4) $\begin{bmatrix} -a & -b \\ -c & -d \end{bmatrix} \in M_{22}(\mathbb{R})$ and adding this to $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ gives you the zero matrix, so indeed every matrix in $M_{22}(\mathbb{R})$ has a negative in $M_{22}(\mathbb{R})$.

The algebraic properties (5)-(10) are just like in $\mathbb{R}^4$; checking them is straightforward. I include the proof below, using a few different (acceptable) techniques.

Begin by letting $\mathbf{u} = \begin{bmatrix} u_1 & u_2 \\ u_3 & u_4 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} v_1 & v_2 \\ v_3 & v_4 \end{bmatrix}$, etc, and let $c,d \in \mathbb{R}$. Then

- (5) $\mathbf{u}+\mathbf{v} = \begin{bmatrix} u_1+v_1 & u_2+v_2 \\ u_3+v_3 & u_4+v_4 \end{bmatrix}$ and $\mathbf{v}+\mathbf{u} = \begin{bmatrix} v_1+u_1 & v_2+u_2 \\ v_3+u_3 & v_4+u_4 \end{bmatrix}$; these are equal. So (5) holds.
- (6) $\mathbf{u}+(\mathbf{v}+\mathbf{w}) = \begin{bmatrix} u_1 & u_2 \\ u_3 & u_4 \end{bmatrix} + \begin{bmatrix} v_1+w_1 & v_2+w_2 \\ v_3+w_3 & v_4+w_4 \end{bmatrix} = \begin{bmatrix} u_1+(v_1+w_1) & u_2+(v_2+w_2) \\ u_3+(v_3+w_3) & u_4+(v_4+w_4) \end{bmatrix}$ and similarly $(\mathbf{u}+\mathbf{v})+\mathbf{w} = \begin{bmatrix} (u_1+v_1)+w_1 & (u_2+v_2)+w_2 \\ (u_3+v_3)+w_3 & (u_4+v_4)+w_4 \end{bmatrix}$; these are equal. So (6) holds.
- (7)

$$\begin{aligned} c(\mathbf{u}+\mathbf{v}) &= c\begin{bmatrix} u_1+v_1 & u_2+v_2 \\ u_3+v_3 & u_4+v_4 \end{bmatrix} \\ &= \begin{bmatrix} c(u_1+v_1) & c(u_2+v_2) \\ c(u_3+v_3) & c(u_4+v_4) \end{bmatrix} \\ &= \begin{bmatrix} cu_1+cv_1 & cu_2+cv_2 \\ cu_3+cv_3 & cu_4+cv_4 \end{bmatrix} \\ &= \begin{bmatrix} cu_1 & cu_2 \\ cu_3 & cu_4 \end{bmatrix} + \begin{bmatrix} cv_1 & cv_2 \\ cv_3 & cv_4 \end{bmatrix} \\ &= c\mathbf{u}+c\mathbf{v} \end{aligned}$$

so this axiom holds.

- (8)

$$\begin{aligned}(c+d)\mathbf{u} &= (c+d)\begin{bmatrix} u_1 & u_2 \\ u_3 & u_4 \end{bmatrix} \\ &= \begin{bmatrix} (c+d)u_1 & (c+d)u_2 \\ (c+d)u_3 & (c+d)u_4 \end{bmatrix} \\ &= \begin{bmatrix} cu_1+du_1 & cu_2+du_2 \\ cu_3+du_3 & cu_4+du_4 \end{bmatrix} \\ &= \begin{bmatrix} cu_1 & cu_2 \\ cu_3 & cu_4 \end{bmatrix} + \begin{bmatrix} du_1 & du_2 \\ du_3 & du_4 \end{bmatrix} \\ &= c\mathbf{u}+d\mathbf{u}\end{aligned}$$

  as required, so this axiom holds.
- (9) Compare the ith component of each side, $1 \le i \le 4$. The ith entry of $a(b\mathbf{u})$ is $a(bu_i)$ and the ith entry of $(ab)\mathbf{u}$ is $(ab)u_i$, and these are equal in $\mathbb{R}$. So $a(b\mathbf{u}) = (ab)\mathbf{u}$ since each of their entries are equal.
- (10) The i entry of $1\mathbf{u}$ is $1u_i = u_i$, which is the ith entry of $\mathbf{u}$. So $1\mathbf{u} = \mathbf{u}$.

So $M_{22}(\mathbb{R})$ is a vector space. ■

! In fact, $M_{mn}(\mathbb{R})$, the set of all $m \times n$ matrices, when given operations analogous to those described above, is a vector space, for any $m,n \ge 1$.

Our next example is a space of functions. We first establish some notation.

Definition 4.3.8 Let $[a,b]$ denote the interval $\{x \in \mathbb{R} | a \le x \le b\}$, and let

$$F[a,b] = \{f | f\colon [a,b] \to \mathbb{R}\}$$

be the set of all functions with domain $[a,b]$ with values in $\mathbb{R}$. For $f,g \in F[a,b]$, we have that $f = g$ if and only if $f(x) = g(x)$ for all $x \in [a,b]$. Moreover, we define $f+g$ to be the function which sends x to $f(x)+g(x)$, that is:

$$(f+g)(x) = f(x)+g(x)$$

and for any scalar $c \in \mathbb{R}$, cf is the function taking x to $cf(x)$, that is:

$$(cf)(x) = c(f(x)).$$

Geometrically, addition corresponds to vertically adding the graphs of f and g; and scalar multiplication corresponds to scaling the graph of f by c.

■ **Example 4.3.9 Space of functions**

$F[a,b]$ is a vector space with these operations.

Check: closure is good; the zero vector is the zero function which sends every x to 0; the negative of f is the function $-f$ which sends x to $-f(x)$; and the algebraic operation axioms all hold. ■

■ **Example 4.3.10** We could also consider $F(\mathbb{R})$, the set of all functions from $\mathbb{R}$ to $\mathbb{R}$, with the same operations; this is also a vector space. So for example $\cos(x) \in F(\mathbb{R})$ and $x + x^2 \in F(\mathbb{R})$; in fact all polynomial functions are contained in $F(\mathbb{R})$. But $\frac{1}{x}, \tan(x) \notin F(\mathbb{R})$ because they are not defined on all of $\mathbb{R}$. ■

In several of our examples, axioms (5)-(10) were obviously true because they were true for a larger superset of vectors. What this means is that we will have a shortcut for checking if a subset of a vector space (with the *same operations*) is itself a vector space.

Problems

Remark: In the following examples, *explain your answer*, which means: if you say a set closed under some operation, give an explanation ('proof') which works in *all* cases: don't choose examples and simply verify it works for your chosen examples. On the other hand, if you say the set isn't closed under some operation, *you must* give an example to illustrate your answer.

Problem 4.1 Determine whether the following sets are closed under the indicated rule for addition.

a) $\{(x, x+2) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$; standard addition of vectors in $\mathbb{R}^2$

b) $^\star\{(x,y) \in \mathbb{R}^2 \mid x - 3y = 0\}$; standard addition of vectors in $\mathbb{R}^2$

c) $\{(x,y) \in \mathbb{R}^2 \mid x - 3y = 1\}$; standard addition of vectors in $\mathbb{R}^2$

d) $^\star\{(x,y) \in \mathbb{R}^2 \mid xy \geq 0\}$; standard addition of vectors in $\mathbb{R}^2$

e) $\{(x,y,z) \in \mathbb{R}^3 \mid x + 2y + z = 0\}$; standard addition of vectors in $\mathbb{R}^3$

f) $^\star\{(x,y,z) \in \mathbb{R}^3 \mid x + 2y + z = 1\}$; standard addition of vectors in $\mathbb{R}^3$

g) $\{(x,y,z,w) \in \mathbb{R}^4 \mid x - y + z - w = 0\}$; standard addition of vectors in $\mathbb{R}^4$.

h) $^\star\{(x, x+2) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$; *Non-standard addition*: $(x,y)\tilde{+}(x',y') = (x+x', y+y'-2)$.

i) $\{(x,y,z) \in \mathbb{R}^3 \mid x + 2y + z = 1\}$; *Non-standard addition*: $(x,y,z)\tilde{+}(x',y',z') = (x+x', y+y', z+z'-1)$.

Problem 4.2 Determine whether each of the following sets is closed under the indicated rule for multiplication of vectors by scalars.

a) $\{(x, x+2) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$; standard rule for multiplication of vectors in $\mathbb{R}^2$ by scalars.

b) $^\star\{(x,y) \in \mathbb{R}^2 \mid x - 3y = 0\}$; standard rule for multiplication of vectors in $\mathbb{R}^2$ by scalars.

c) $\{(x,y) \in \mathbb{R}^2 \mid x - 3y = 1\}$; standard rule for multiplication of vectors in $\mathbb{R}^2$ by scalars.

d) $^\star\{(x,y) \in \mathbb{R}^2 \mid xy \geq 0\}$; standard rule for multiplication of vectors in $\mathbb{R}^2$ by scalars.

e) $\{(x,y,z) \in \mathbb{R}^3 \mid x + 2y + z = 0\}$; standard rule for multiplication of vectors in $\mathbb{R}^3$ by scalars.

f) $^\star\{(x,y,z) \in \mathbb{R}^3 \mid x + 2y + z = 1\}$; standard rule for multiplication of vectors in $\mathbb{R}^3$ by scalars.

g) $\{(x,y,z,w) \in \mathbb{R}^4 \mid x-y+z-w=0\}$; standard rule for multiplication of vectors in $\mathbb{R}^4$ by scalars.

h) $^\star\{(x,x+2) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$; *Non-standard multiplication of vectors by scalars* $k \in \mathbb{R}$:

$$k \circledast (x,y) = (kx, ky-2k+2).$$

i) $\{(x,y,z) \in \mathbb{R}^3 \mid x+2y+z=1\}$; *Non-standard multiplication of vectors by scalars* $k \in \mathbb{R}$:

$$k \circledast (x,y,z) = (kx, ky, kz-k+1).$$

Problem 4.3 Determine whether the following subsets of $\mathbf{F}(\mathbb{R}) = \{f \mid f : \mathbb{R} \to \mathbb{R}\}$ are closed under the standard addition of functions in $\mathbf{F}(\mathbb{R})$. (Recall that $\mathbf{F}(\mathbb{R})$ consists of all real-valued functions of a real variable; i.e., all functions with domain $\mathbb{R}$, taking values in $\mathbb{R}$).

a) $\{f \in \mathbf{F}(\mathbb{R}) \mid f(2)=0\}$;

b) $^\star\{f \in \mathbf{F}(\mathbb{R}) \mid f(2)=1\}$.

c) $\{f \in \mathbf{F}(\mathbb{R}) \mid f(1)=2\}$.

d) $^\star\{f \in \mathbf{F}(\mathbb{R}) \mid$ for all $x \in \mathbb{R}$, $f(x) \leq 0\}$.

e) $\{f \in \mathbf{F}(\mathbb{R}) \mid$ For all $x \in \mathbb{R}$, $f(-x) = f(x)\}$

f) $^\star\{f \in \mathbf{F}(\mathbb{R}) \mid$ For all $x \in \mathbb{R}$, $f(-x) = -f(x)\}$

g) $\{f \in \mathbf{F}(\mathbb{R}) \mid f$ is twice-differentiable, and for all $x \in \mathbb{R}$, $f''(x)+f(x)=0\}$

Problem 4.4 Determine whether the following sets are closed under the standard rule for multiplication of functions by scalars in $\mathbf{F}(\mathbb{R})$.

a) $\{f \in \mathbf{F}(\mathbb{R}) \mid f(2)=0\}$.

b) $^\star\{f \in \mathbf{F}(\mathbb{R}) \mid f(2)=1\}$.

c) $\{f \in \mathbf{F}(\mathbb{R}) \mid f(1)=2\}$.

d) $^\star\{f \in \mathbf{F}(\mathbb{R}) \mid$ For all $x \in \mathbb{R}$, $f(x) \leq 0\}$

e) $\{f \in \mathbf{F}(\mathbb{R}) \mid$ For all $x \in \mathbb{R}$, $f(-x) = f(x)\}$

f) $^\star\{f \in \mathbf{F}(\mathbb{R}) \mid$ For all $x \in \mathbb{R}$, $f(-x) = -f(x)\}$

g) $\{f \in \mathbf{F}(\mathbb{R}) \mid f$ is twice-differentiable, and for all $x \in \mathbb{R}$, $f''(x)+f(x)=0\}$

Problem 4.5 Determine whether the following sets are closed under the standard operation of addition of matrices in $\mathbf{M}_{22}(\mathbb{R})$.

a) $\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; b = c \right\}$.

b) $^\star\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; a + d = 0 \right\}$.

c) $\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; ad - bc = 0 \right\}$.

d) $^\star\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; ad = 0 \right\}$.

e) $\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; bc = 1 \right\}$.

Problem 4.6 Determine whether the following sets are closed under the standard rule for multiplication of matrices by scalars in $\mathbf{M}_{22}(\mathbb{R})$.

a) $\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; b = c \right\}$.

b) $^\star\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; a + d = 0 \right\}$.

c) $\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; ad - bc = 0 \right\}$.

d) $^\star\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; ad = 0 \right\}$.

e) $\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; bc = 1 \right\}$.

Problem 4.7 The following sets have been given the indicated rules for addition of vectors, and multiplication of objects by real scalars (the so-called *'vector operations'*). If possible, check if there is a zero vector in the subset in each case. If it is possible, show your choice works in *all* cases, and if it is not possible, give an example to illustrate your answer.

(Note: in the last two parts, since the vector operations are not the standard ones, the zero vector will probably not be the one you're accustomed to.)

a) $\{(x, x+2) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$; standard vectors operations in $\mathbb{R}^2$.

b) $^\star\{(x, y) \in \mathbb{R}^2 \mid x - 3y = 0\}$; standard vectors operations in $\mathbb{R}^2$.

c) $\{(x,y)\in\mathbb{R}^2 \mid x-3y=1\}$; standard vectors operations in $\mathbb{R}^2$.

d) $^\star\{(x,y)\in\mathbb{R}^2 \mid xy\geq 0\}$; standard vectors operations in $\mathbb{R}^2$.

e) $\{(x,y,z)\in\mathbb{R}^3 \mid x+2y+z=0\}$; standard vectors operations in $\mathbb{R}^3$.

f) $^\star\{(x,y,z)\in\mathbb{R}^3 \mid x+2y+z=1\}$; standard vectors operations in $\mathbb{R}^3$.

g) $\{(x,y,z,w)\in\mathbb{R}^4 \mid x-y+z-w=0\}$; ; standard vectors operations in $\mathbb{R}^4$.

h) $^\star\{(x,x+2)\in\mathbb{R}^2 \mid x\in\mathbb{R}\}$; <u>*Non-standard operations:–*</u> Addition: $(x,y)\tilde{+}(x',y')=(x+x',y+y'-2)$. Multiplication of vectors by scalars $k\in\mathbb{R}$: $k\circledast(x,y)=(kx,ky-2k+2)$.

i) $\{(x,y,z)\in\mathbb{R}^3 \mid x+2y+z=1\}$; <u>*Non-standard operations:–*</u> Addition: $(x,y)\tilde{+}(x',y')=(x+x',y+y',z+z'-1)$. Multiplication of vectors by scalars $k\in\mathbb{R}$: $k\circledast(x,y,z)=(kx,ky,kz-k+1)$.

Problem 4.8 Explain your answers to the following:

a) *Determine whether the zero function of $\mathbf{F}(\mathbb{R})$ belongs to each of the subsets in question 3.

b) Determine whether the zero matrix of $\mathbf{M}_{22}(\mathbb{R})$ belongs to each of the subsets in question 5.

Problem 4.9 The following sets have been given the indicated rules for addition of vectors, and multiplication of objects by real scalars. In each case, If possible, check if vector in the subset has a 'negative' in the subset.

Again, since the vector operations are not the standard ones, the negative of a vector will probably not be the one you're accustomed to seeing.

a) $\{(x,x+2)\in\mathbb{R}^2 \mid x\in\mathbb{R}\}$; <u>*Non-standard Operations:–*</u> Addition: $(x,y)\tilde{+}(x',y')=(x+x',y+y'-2)$. Multiplication of vectors by scalars $k\in\mathbb{R}$: $k\circledast(x,y)=(kx,ky-2k+2)$.

b) $^\star\{(x,y,z)\in\mathbb{R}^3 \mid x+2y+z=1\}$; <u>*Non-standard Operations:–*</u> Addition:

$$(x,y)\tilde{+}(x',y')=(x+x',y+y',z+z'-1).$$

Multiplication of vectors by scalars $k\in\mathbb{R}$: $k\circledast(x,y,z)=(kx,ky,kz-k+1)$.

Problem 4.10 Explain your answers to the following:

a) *Determine whether the subsets in question 1 (given the operations described in questions 1 and 2) are vector spaces.

b) Determine whether the subsets of $\mathbf{F}(\mathbb{R})$ in question 3, equipped with the standard vector operations of $\mathbf{F}(\mathbb{R})$ are vector spaces.

c) Determine whether the subsets of $\mathbf{M}_{22}(\mathbb{R})$ in question 5, equipped with the standard vector operations of $\mathbf{M}_{22}(\mathbb{R})$ are vector spaces.

Problem 4.11 Explain your answers to the following:

a) *Determine whether the subsets of $\mathbf{F}(\mathbb{R})$ in question 3 are vector spaces.

b) Determine whether the zero function of $\mathbf{F}(\mathbb{R})$ belongs to each of the subsets in question 3.

c) Determine whether the zero matrix of $\mathbf{M}_{22}(\mathbb{R})$ belongs to each of the subsets in question 5.

Problem 4.12 Justify your answers to the following:

a) Equip the set $V = \mathbb{R}^2$ with the *non-standard operations:–* Addition:

$$(x,y)\tilde{+}(x',y') = (x+x', y+y'-2).$$

Multiplication of vectors by scalars $k \in \mathbb{R}$:

$$k \circledast (x,y) = (kx, ky-2k+2).$$

Check that $\mathbb{R}^2$, with these new operations, is indeed a vector space.

b) Equip the set $V = \mathbb{R}^3$ *non-standard operations:–* Addition:

$$(x,y,z)\tilde{+}(x',y',z') = (x+x', y+y', +y, z+z'-1).$$

Multiplication of vectors by scalars $k \in \mathbb{R}$:

$$k \circledast (x,y,z) = (kx, ky, kz-k+1).$$

Check that $\mathbb{R}^3$, with these new operations, is indeed a vector space.

In these cases, you will even need to check the arithmetic axioms, as the operations are weird. For example, the distributive axiom: $k \circledast (u\tilde{+}v) = k \circledast u \tilde{+} k \circledast v$ is true, but definitely not *obviously* so!

Problem 4.13 Let $\mathbf{E} = \{\text{“}ax+by+cz=d\text{”} \mid a,b,c,d \in \mathbb{R}\}$ be the set of linear equations with real coefficients in the variables x, y and z. Equip $\mathbf{E}$ with the usual operations on equations that you learned in high school: addition of equations, denoted here by “$\oplus$" and multiplication by scalars, denoted here by “$\circledast$", as follows:

$$\text{“}ax+by+cz=d\text{”} \oplus \text{“}ex+fy+gz=h\text{”} = \text{“}(a+e)x+(b+f)y+(c+g)z=d+h\text{”}$$

and

$$\forall k \in \mathbb{R}, \quad k \circledast \text{“}ax+by+cz=d\text{”} = \text{“}kax+kby+kcz=kd\text{”}.$$

Prove that $\mathbf{E}$ is a vector space.

Problem 4.14 (For the mathematically curious)

a) It is *not* amongst the axioms for a vector space V that $0\mathbf{v} = \mathbf{0}$ for all vectors $\mathbf{v} \in V$. (Here the zero on the left hand side of the equation is the *scalar* zero, while the zero on the right hand side of the equation is the zero *vector*.)
Nevertheless, it is indeed true in every vector space that $0\mathbf{v} = \mathbf{0}$ for all vectors $\mathbf{v} \in V$.
Prove this, using a few of the axioms for a vector space.

b) Neither it is amongst the axioms for a vector space V that $(-1)\mathbf{v} = -\mathbf{v}$ for all vectors $\mathbf{v} \in V$, where the '$(-1)\mathbf{v}$' on the left hand side of the equation indicates the result of multiplication of $\mathbf{v}$ by the scalar -1, and the '$-\mathbf{v}$' on the right hand side of the equation indicates the negative of $\mathbf{v}$ – whose existence is guaranteed by one of the axioms.
Nonetheless, it is indeed true in every vector space that $(-1)\mathbf{v} = -\mathbf{v}$ for all vectors $\mathbf{v} \in V$.
Prove this, using a few of the axioms for a vector space. You might find part (a) useful.

c) Let $\emptyset = \{\}$ denote the empty set. Could it be made into a vector space?

d) And here's another interesting vector space: Define V to be the set of *formal power series*. A formal power series is an expression of the form

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + \cdots$$

where the coefficients a_n are real numbers. Define addition by the formula

$$\sum_{n=0}^{\infty} a_n x^n + \sum_{n=0}^{\infty} b_n x^n = \sum_{n=0}^{\infty} (a_n + b_n) x^n$$

and scalar multiplication by

$$c \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} c a_n x^n.$$

Show that this is a vector space.[a]

[a] Note that in general these are not functions on $\mathbb{R}$, since most of the time, if you plug in any non-zero value for x, the sum makes no sense (technical term: ' the series diverges'). The convergent ones (that is, ones where you can plug in some values of x and it actually makes sense) include the power series for e^x, $\cos(x)$ and $\sin(x)$, which we glimpsed in our first class. Power series are terrifically useful tools in Calculus. If you're lucky, you'll learn what 'convergence' really means for an infinite series in MAT1325 or MAT2125.


Photo: Ralph Nevins. Montpellier, France

5. Subspaces and Spanning Sets

In the last chapter, we established the concept of a *vector space*, which is a set V on which we can perform two operations (addition and scalar multiplication) such that 10 axioms are satisfied. The result is: a vector space is algebraically indistinguishable from things like $\mathbb{R}^n$ (as far as addition and scalar multiplication go).

We met some particularly nice vector spaces:

- $\mathbb{R}^n$, $n \geq 1$
- $\mathscr{E}$, a space of linear equations in 3 variables
- $F[a,b]$, functions on the interval $[a,b]$
- $F(\mathbb{R})$, functions on the real line
- $M_{m\times n}(\mathbb{R})$, $m \times n$ real matrices

And for each of these, we had to check 10 axioms. But we noticed that sometimes, the last 6 axioms (the ones dealing purely with algebraic properties) were "obvious".

Let's consider this more carefully.

5.1 Subsets of Vectors Spaces

Suppose V is a vector space, and $W \subseteq V$ is a *subset* of V.

■ **Example 5.1.1** Let $V = \mathbb{R}^2$ (we know it's a vector space), and let $W = \{(x,2x) | x \in \mathbb{R}\}$. Use the *same* operations of addition and scalar multiplication on W as we use in $\mathbb{R}^2$. What do we really need to verify to decide if it's a vector space?

closure: (1) closure under addition: if $(x,2x)$ and $(y,2y)$ are in W, then $(x,2x)+(y,2y) = (x+y, 2(x+y))$, which is in W because it has the correct form $(z,2z)$ with $z = x+y$. YES.

(2) closure under scalar multiplication: if $(x,2x) \in W$ and $c \in \mathbb{R}$ then $c(x,2x) = (cx, 2(cx)) \in W$ as well. YES.

existence: (3) the zero vector is $(0,0)$, which is in W. But we don't need to bother checking if $\mathbf{0}+\mathbf{u}=\mathbf{u}$ for each $\mathbf{u}\in W$, because we know this is true for all $\mathbf{u}\in\mathbb{R}^2$ already (since V is a vector space and so satisfies axiom (3)). So it's enough to know that $\mathbf{0}\in W$.

(4) if $\mathbf{u}\in W$, then we know that $-\mathbf{u}=(-1)\mathbf{u}\in W$ because (2) is true. And again, we know that the negative has the correct property because we checked it when we decided that V was a vector space.

algebraic properties: (5)-(10) All of these properties are known to be true for any vectors in V; so in particular they are true for any vectors in the smaller set $W\subset V$. So we don't need to check them again!

We conclude: W is a vector space. ■

Definition 5.1.2 A subset W of a vector space V is called a *subspace* (or *subspace of* V) if it is a vector space when given the *same operations* of addition and scalar multiplication as V.

■ **Example 5.1.3** $W=\{(x,2x)|x\in\mathbb{R}\}$ is a subspace of $\mathbb{R}^2$. ■

Theorem 5.1.4 — Subspace Test. : If V is a vector space and $W\subseteq V$, then W is a subspace of V if and only if the following 3 conditions hold:
1. $\mathbf{0}\in W$.
2. W is closed under addition: for every $\mathbf{u},\mathbf{v}\in W$, $\mathbf{u}+\mathbf{v}$ is again in W.
3. W is closed under multiplication by scalars: for every $\mathbf{u}\in W$ and $r\in\mathbb{R}$, $r\mathbf{u}$ is again in W.

Think of this theorem as giving you a shortcut to working out if certain sets are vector spaces. We saw in the example above why it comes down to just these three axioms. [1]

5.2 Many examples

Now let us apply the subspace test to acquire many more examples of vector spaces. On the homework and in the suggested exercises, you'll see many examples of subsets to which the subspace test does not apply (because the operations aren't the same) or where the subspace test fails. Do many examples to develop your intution about what does and does not constitute a vector space!

Problem 5.2.1 Let $T=\{\mathbf{u}\in\mathbb{R}^3|\mathbf{u}\cdot(1,2,3)=0\}$, with the usual operations on $\mathbb{R}^3$. Is this a subspace?

Solution Notice that $T\subset\mathbb{R}^3$ and $T\neq\mathbb{R}^3$. In fact, T is the plane with equation

$$x+2y+3z=0.$$

Since we're using the usual operations on $\mathbb{R}^3$, we apply the subspace test:
1. Is $\mathbf{0}\in T$? Yes, since $\mathbf{0}=(0,0,0)$ satisfies the condition $\mathbf{u}\cdot(1,2,3)=0$.
2. Is T closed under addition?
   Well, suppose $\mathbf{u}$ and $\mathbf{v}$ are in T.
   That means $\mathbf{u}\cdot(1,2,3)=0$ and $\mathbf{v}\cdot(1,2,3)=0$.

[1] You might wonder why we have to include the first one, though, since $0\mathbf{u}=\mathbf{0}$ so it ought to follow from closure under scalar multiplication. It does, but only if the set W is not the empty set. So if W isn't empty, then it's enough to check axioms (2) and (3). BUT as a general rule, (1) is super-easy to check; and if it fails you know W isn't a subspace. So it's a quick way to exclude certain sets and we always use it.

We need to decide if $\mathbf{u}+\mathbf{v} \in T$.
That means we need to decide if $(\mathbf{u}+\mathbf{v}) \cdot (1,2,3) = 0$.
We calculate: $(\mathbf{u}+\mathbf{v}) \cdot (1,2,3) = (\mathbf{u} \cdot (1,2,3)) + (\mathbf{v} \cdot (1,2,3)) = 0+0 = 0$.
So $\mathbf{u}+\mathbf{v} \in T$, and we conclude T is closed under addition.

3. Is T closed under multiplication by scalars?
Let $k \in \mathbb{R}$ and $\mathbf{u} \in T$.
So we have that $\mathbf{u} \cdot (1,2,3) = 0$.
We want to decide if $k\mathbf{u} \in T$.
That means we need to see if $(k\mathbf{u}) \cdot (1,2,3) = 0$.
We calculate: $(k\mathbf{u}) \cdot (1,2,3) = k(\mathbf{u} \cdot (1,2,3)) = k(0) = 0$.
So $k\mathbf{u} \in T$ and so T is closed under scalar multiplication.

So YES, T is a subspace of $\mathbb{R}^3$.

Notice that we could have replaced the vector $(1,2,3)$ with any vector $\mathbf{n}$, and the result would have been the same. In fact:

! Any plane through the origin in $\mathbb{R}^3$ is a subspace.

Furthermore, if a plane in $\mathbb{R}^3$ *doesn't* go through the origin, then it fails the first condition of the subspace test. So we have:

! Any plane in $\mathbb{R}^3$ which doesn't go through the origin is not a subspace.

Problem 5.2.2 Let $\mathbf{v} \in \mathbb{R}^n$ and set $L = \{t\mathbf{v} | t \in \mathbb{R}\}$ to be the line in $\mathbb{R}^n$ through the origin with direction vector $\mathbf{v}$ (with the usual operations from $\mathbb{R}^n$). Is L a subspace?

Solution We apply the subspace test.

1. Yes, $\mathbf{0} \in L$: take $t = 0$.
2. If $t\mathbf{v}$ and $s\mathbf{v}$ are two points in L, then $t\mathbf{v}+s\mathbf{v} = (t+s)\mathbf{v}$ which is again a multiple of $\mathbf{v}$, so lies in L. So L is closed under addition.
3. If $t\mathbf{v} \in L$ and $k \in \mathbb{R}$ then $k(t\mathbf{v}) = (kt)\mathbf{v}$, which is a multiple of $\mathbf{v}$, so it lies in L. Thus L is closed under scalar multiplication.

We conclude that L is a subspace of $\mathbb{R}^n$.

Similarly to above, we deduce:

! Any line through the origin in $\mathbb{R}^n$ is a subspace. Any line in $\mathbb{R}^n$ which does not go through the origin is not a subspace.

Problem 5.2.3 Let $V = F(\mathbb{R})$, the vector space of all functions with domain $\mathbb{R}$. Let $\mathbb{P}$ be the set of all *polynomial functions*, that is, $\mathbb{P}$ consists of all functions that can be written as

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

for some $n \geq 0$ and $a_i \in \mathbb{R}$. Is $\mathbb{P}$ a vector space?

Solution We first note (it may not seem obvious) that the natural operations of adding two polynomials, and multiplying a polynomial by a scalar, are exactly the usual operations on functions $F(\mathbb{R})$. So $\mathbb{P} \subset F(\mathbb{R})$ with the same operations, so we can apply the subspace test. (Phew!)

1. Is $\mathbf{0} \in \mathbb{P}$? Remember that $\mathbf{0}$, here, is the function which when you plug in any x you get 0 as an answer. Well, that's the zero polynomial: $p(x) = 0 + 0x + 0x^2$. (Or just $p(x) = 0$, for short.) It's a polynomial, so $\mathbf{0} \in \mathbb{P}$.
2. Is $\mathbb{P}$ closed under addition? Yes: the sum of two polynomials is again a polynomial (it's not as though you could get an exponential function or something like that).
3. Is $\mathbb{P}$ closed under scalar multiplication? Yes: multiplying a polynomial by a scalar gives another polynomial.

So yes, $\mathbb{P}$ is a subspace, so a vector space.

Definition 5.2.4 The *transpose* of an $m \times n$ matrix A is the $n \times m$ matrix A^T whose rows are the columns of A. For example,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}.$$

Problem 5.2.5 Let $S = \{A \in M_{22}(\mathbb{R}) | A^T = A\}$ be the set of *symmetric* 2×2 matrices (with the usual operations). Is this a vector space?

Solution Since $S \subset M_{22}(\mathbb{R})$, and we're using the same operations, we can apply the subspace test.

But it's helpful to get more comfortable with this set S, first. Rewrite it as:

$$\begin{aligned} S &= \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| \begin{bmatrix} a & b \\ c & d \end{bmatrix}^T = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right\} \\ &= \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right\} \\ &= \left\{ \begin{bmatrix} a & b \\ b & d \end{bmatrix} \middle| a, b, d \in \mathbb{R} \right\} \end{aligned}$$

Oh, that's more clear!

1. By taking $a = 0$, $b = 0$ and $d = 0$, we get the zero matrix; so this is in S.
2. Take two arbitary matrices in S, and add them:

$$\begin{bmatrix} a & b \\ b & d \end{bmatrix} + \begin{bmatrix} a' & b' \\ b' & d' \end{bmatrix} = \begin{bmatrix} a+a' & b+b' \\ b+b' & d+d' \end{bmatrix}$$

   This is again in S, since it matches the required pattern to be in S (that the (1,2) and (2,1) entries of the matrix are equal). So S is closed under addition.
3. Let $k \in \mathbb{R}$; then

$$k \begin{bmatrix} a & b \\ b & d \end{bmatrix} = \begin{bmatrix} ka & kb \\ kb & kd \end{bmatrix}$$

   is again in S, so S is closed under scalar multiplication.

So S is a subspace of $M_{22}(\mathbb{R})$.

Problems

Problem 5.1 Determine whether each of the following is a subspace of the indicated vector space. Assume the vector space has the standard operations unless otherwise indicated.

Remember some useful shortcuts we established in class: (1) every line or plane through the origin in $\mathbb{R}^2$ or $\mathbb{R}^3$ is a subspace; (2) See the Theorem 6.4.1 in the next chapter : Every 'span' is a subspace: i.e. if $W = \text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ for some vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in a vector space V, then W is a subspace of V. Once you've seen Theorem 6.4.1, you only use the subspace test as a last resort!

a) $\{(2x, x) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$; $\mathbb{R}^2$.

b) $^\star\{(x, x-3) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$; $\mathbb{R}^2$.

c) $\{(x, y) \in \mathbb{R}^2 \mid xy = 0\}$; $\mathbb{R}^2$.

d) $\{(x, x+2) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$; $\mathbb{R}^2$.

e) $^\star\{(x, y) \in \mathbb{R}^2 \mid x - 3y = 0\}$; $\mathbb{R}^2$.

f) $\{(x, y) \in \mathbb{R}^2 \mid x - 3y = 1\}$; $\mathbb{R}^2$.

g) $^\star\{(x, y) \in \mathbb{R}^2 \mid xy \geq 0\}$; $\mathbb{R}^2$.

h) $\{(x, y, z) \in \mathbb{R}^3 \mid x + 2y + z = 0\}$; $\mathbb{R}^3$.

i) $^\star\{(x, y, z) \in \mathbb{R}^3 \mid x + 2y + z = 1\}$; $\mathbb{R}^3$.

j) $\{(x, y, z) \in \mathbb{R}^3 \mid xyz = 0\}$; $\mathbb{R}^3$.

k) $^\star\{(x, y, z, w) \in \mathbb{R}^4 \mid x - y + z - w = 0\}$; $\mathbb{R}^4$.

l) $\{(x, y, z, w) \in \mathbb{R}^4 \mid xy = zw\}$; $\mathbb{R}^4$.

m) * $\{(x, x+2) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$; <u>*Non-standard operations:–*</u> Addition:

$$(x, y) \tilde{+} (x', y') = (x + x', y + y - 2).$$

Multiplication of vectors by scalars $k \in \mathbb{R}$: $k \circledast (x, y) = (kx, ky - 2k + 2)$.

n) * $\{(x, y, z) \in \mathbb{R}^3 \mid x + 2y + z = 1\}$; <u>*Non-standard operations:–*</u> Addition:

$$(x, y, z) \tilde{+} (x', y', z') = (x + x', y + y', z + z' - 1)$$

Multiplication of vectors by scalars $k \in \mathbb{R}$: $k \circledast (x, y, z) = (kx, ky, kz - k + 1)$.

Problem 5.2 Determine whether each of the following is a subspace of of $\mathbf{F}(\mathbb{R}) = \{f \mid f : \mathbb{R} \to \mathbb{R}\}$ with its standard operations. (Here, you'll need to use the subspace test.)

a) $\{f \in \mathbf{F}(\mathbb{R}) \mid f(2) = 0\}$;

b) $^\star\{f \in \mathbf{F}(\mathbb{R}) \mid f(2) = 1\}$.

c) $\{f \in \mathbf{F}(\mathbb{R}) \mid f(1) = 2\}$.

d) $^\star\{f \in \mathbf{F}(\mathbb{R}) \mid \text{ for all } x \in \mathbb{R}, f(x) \leq 0\}$.

e) $\{f \in \mathbf{F}(\mathbb{R}) \mid \text{ For all } x \in \mathbb{R}, f(-x) = f(x)\}$

f) $^\star\{f \in \mathbf{F}(\mathbb{R}) \mid \text{ For all } x \in \mathbb{R}, f(-x) = -f(x)\}$

g) $\{f \in \mathbf{F}(\mathbb{R}) \mid f \text{ is twice-differentiable, and for all } x \in \mathbb{R}, f''(x) + f(x) = 0\}$

h) $^\star\mathbb{P} = \{p \in \mathbf{F}(\mathbb{R}) \mid p \text{ is a polynomial function in the variable } x\}$

Problem 5.3 Determine whether the following are subspaces of

$$\mathbb{P} = \{p \in \mathbf{F}(\mathbb{R}) \mid p \text{ is a polynomial function in the variable } x\},$$

with its standard operations. (In some parts, you'll be able to use the fact that everything of the form $\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a subspace.)

a) $\{p \in \mathbb{P} \mid \deg(p) = 2\}$[a]

b) $^\star\{p \in \mathbb{P} \mid \deg(p) \leq 2\}$

c) $\mathbb{P}_n = \{p \in \mathbb{P} \mid \deg(p) \leq n\}$

d) $^\star\{p \in \mathbb{P}_2 \mid p(1) = 0\}$.

e) $\{p \in \mathbb{P}_2 \mid p(1) = 2\}$.

f) $^\star\{p \in \mathbb{P}_3 \mid p(2)\,p(3) = 0\}$.

g) $\{p \in \mathbb{P}_3 \mid p(2) = p(3) = 0\}$.

h) $^\star\{p \in \mathbb{P}_2 \mid p(1) + p(-1) = 0\}$.

[a] Recall that the *degree* of a polynomial $p(x) = a_n x^n + \cdots + a_1 x + a_0$, is $\deg(p) = \max\{k \mid a_k \neq 0\}$

Problem 5.4 Determine whether the following are subspaces of $\mathbf{M}_{22}(\mathbb{R})$, with its standard operations. (If you've read ahead, in some parts, you'll be able to us the fact that everything of the form $\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a subspace.)

a) $\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; b = c \right\}$.

b) $^\star\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; a = d = 0 \quad \& \quad b = -c \right\}$.

c) $\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; a + d = 0 \right\}$.

d) $^\star\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; bc = 1 \right\}$.

e) $\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; ad = 0 \right\}$.

f) $\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; ad - bc = 0 \right\}$.

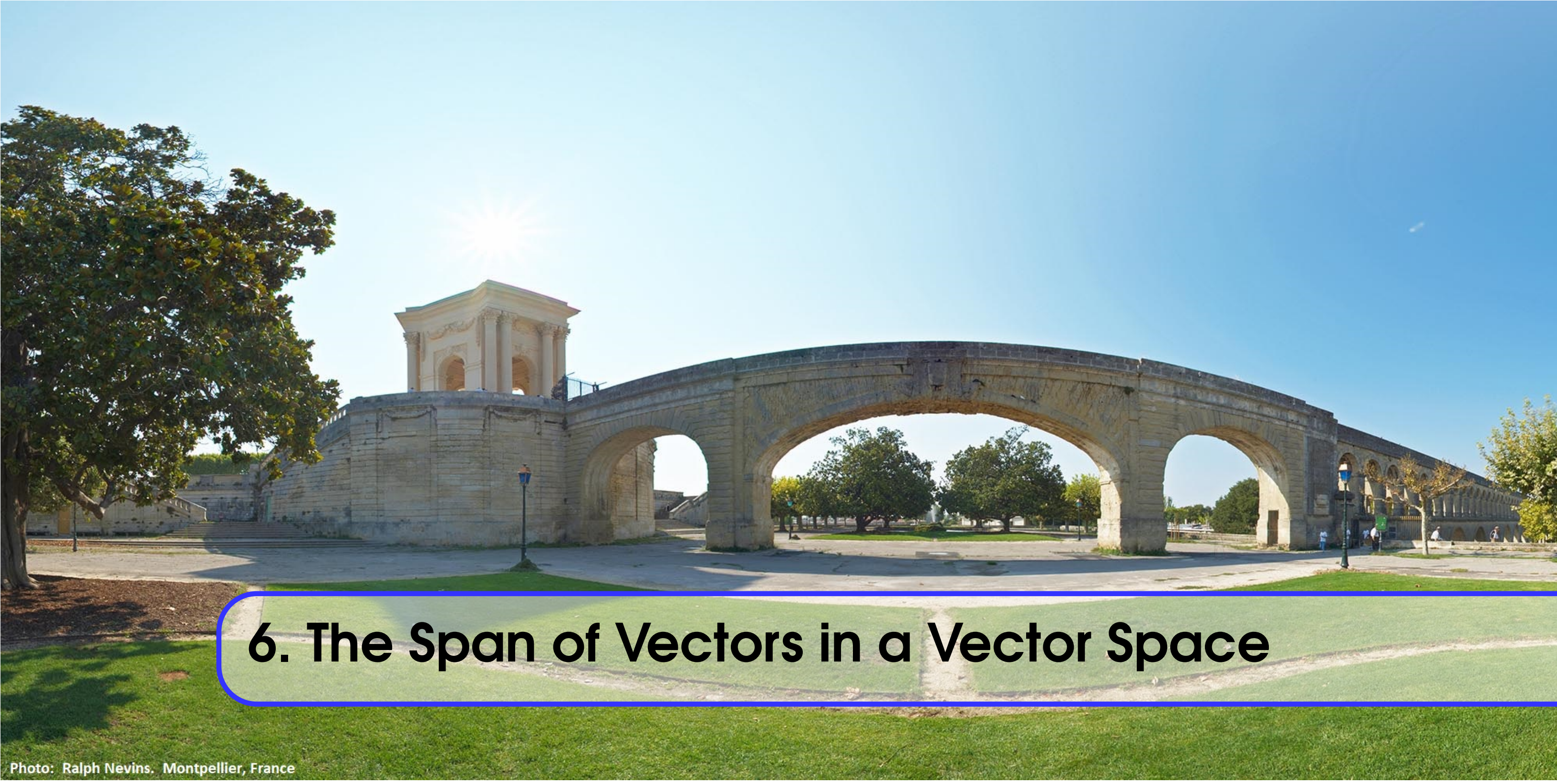


Photo: Ralph Nevins. Montpellier, France

6. The Span of Vectors in a Vector Space

Our goal here is to render finite the infinite! All[1] subspaces are infinite sets of vectors. We have a couple of different ways of describing subspaces, for example:

1. $W = \{\text{things} \mid \text{conditions on things}\}$ like

$$W = \{(x,y,z) \in \mathbb{R}^3 \mid x - 2y + z = 0\}.$$

2. $U = \{\text{things with parameters} \mid \text{parameters are real}\}$, like

$$S = \{ \begin{bmatrix} a & b \\ b & d \end{bmatrix} \mid a,b,d \in \mathbb{R}\}.$$

Each of (1) and (2) has it own advantages, but here, we'll see that you can completely identify a subspace given in the form of (2) by just giving a *finite* list of vectors.

This idea is crucial when it comes time to apply this in the real world: when you need to communicate a subspace (such as an error-correcting code), it is WAY easier to just send n vectors and then say: "take their span", rather than to try to digitally communicate an infinite set defined by equations and/or parameters.

As an added bonus: we'll prove another shortcut to testing if something is a vector space (and this one is fantastically easy).

[1] but one - $\{\mathbf{0}\}$

6.1 Converting sets of form (1) to form (2)

Example 6.1.1 Consider $W = \{(x,y,z) \mid x - 2y + z = 0\}$, of the first kind. Then

$$\begin{aligned} W &= \{(x,y,z) \mid x - 2y + z = 0\} \\ &= \{(x,y,z) \mid x = 2y - z\} \\ &= \{(2y - z, y, z) \mid y, z \in \mathbb{R}\} \end{aligned}$$

which is of the second kind. ■

The idea is just: the "condition" can always be rephrased as one or more equations; solve the equations in terms of one or more variables. (This is something we'll come back to in Chapter 11.)

6.2 Describing (infinite) sets of the form (2) with a finite number of vectors

Example 6.2.1 With W as above:

$$\begin{aligned} W &= \{(2y - z, y, z) \mid y, z \in \mathbb{R}\} \\ &= \{(2y, y, 0) + (-z, 0, z) \mid y, z \in \mathbb{R}\} \\ &= \{y(2,1,0) + z(-1,0,1) \mid y, z \in \mathbb{R}\} \end{aligned}$$

This shows that W is the *set of all linear combinations of* $(2,1,0)$ *and* $(-1,0,1)$. We give it a special name and notation:

$$W = \operatorname{span}\{(2,1,0), (-1,0,1)\}$$

which we read out loud as "W is the span of $(2,1,0)$ and $(-1,0,1)$", or alternatively "W is the span of (the set) $\{(2,1,0), (-1,0,1)\}$". ■

Example 6.2.2 Here's another:

$$\begin{aligned} S &= \left\{ \begin{bmatrix} a & b \\ b & d \end{bmatrix} \middle| a, b, d \in \mathbb{R} \right\} \\ &= \left\{ a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \middle| a, b, d \in \mathbb{R} \right\} \\ &= \operatorname{span} \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}. \end{aligned}$$

■

6.3 The definition of span

Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$ be vectors in a vector space V.

Definition 6.3.1
1. If $a_1, a_2, \cdots, a_m$ are scalars, then the vector

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_m\mathbf{v}_m$$

is called a *linear combination* of $\mathbf{v}_1, \ldots, \mathbf{v}_m$.
2. The set of *all* linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is called the *span of* $\mathbf{v}_1, \ldots, \mathbf{v}_m$. We have

$$\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\} = \{a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_m\mathbf{v}_m \mid a_1, \cdots, a_m \in \mathbb{R}\}.$$

In this case, $\{\mathbf{v}_1,\ldots,\mathbf{v}_m\}$ is called the *spanning set* for $\text{span}\{\mathbf{v}_1,\ldots,\mathbf{v}_m\}$; or we say that $\{\mathbf{v}_1,\ldots,\mathbf{v}_m\}$ *spans* $\text{span}\{\mathbf{v}_1,\ldots,\mathbf{v}_m\}$.

3. A vector space (or subspace) W is *spanned by* $\mathbf{v}_1,\ldots,\mathbf{v}_m \in W$ if $W = \text{span}\{\mathbf{v}_1,\ldots,\mathbf{v}_m\}$. Then we say $\{\mathbf{v}_1,\ldots,\mathbf{v}_m\}$ *spans* W.

■ **Example 6.3.2** Above, we had $W = \text{span}\{(2,1,0),(-1,0,1)\}$. Call these two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$; then W is spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$; or $\{\mathbf{v}_1,\mathbf{v}_2\}$ spans W. ■

■ **Example 6.3.3** We also found that S was the span of three matrices; call them M_1, M_2 and M_3. So S is spanned by M_1, M_2 and M_3; or the set $\{M_1,M_2,M_3\}$ spans S. ■

■ **Example 6.3.4** A line in $\mathbb{R}^n$ through the origin has the form $L = \{t\mathbf{v} \mid t \in \mathbb{R}\}$ So in fact $L = \text{span}\{\mathbf{v}\}$; L is the subspace spanned by the vector $\mathbf{v}$. ■

■ **Example 6.3.5** The vector $(1,2) \in \mathbb{R}^2$ spans the line in direction $(1,2)$. The vector $(1,2)$ is an element of $\mathbb{R}^2$ but it *does not span all of* $\mathbb{R}^2$ because, for example, the vector $(1,3)$ can't be obtained from $(1,2)$ using vector operations (linear combinations). ■

! Caution: If a set of vectors lies in $\mathbb{R}^n$, that doesn't mean they span $\mathbb{R}^n$. The subspace they span is the (usually smaller) subspace of all vectors obtainable as linear combinations from your finite set.

6.4 The BIG THEOREM about spans

Theorem 6.4.1 — Spanned sets are subspaces. Let V be a vector space.
If $\{\mathbf{v}_1,\ldots,\mathbf{v}_m\} \subset V$, define $U = \text{span}\{\mathbf{v}_1,\ldots,\mathbf{v}_m\}$. Then

1. U is *always* a subspace of V.
2. If W is *any* subspace of V which contains all the vectors $\mathbf{v}_1,\ldots,\mathbf{v}_m$, then in fact $U \subseteq W$ (that is, W must also contain every vector in their span). So U is the *smallest* subspace that contains $\mathbf{v}_1,\ldots,\mathbf{v}_m$.

Proof. To prove (1), we apply the subspace test.

1. We have $\mathbf{0} = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_m \in U$
2. If $\mathbf{u} = a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m$ and $\mathbf{w} = b_1\mathbf{v}_1 + \cdots + b_m\mathbf{v}_m$, then $\mathbf{u}+\mathbf{w} = (a_1+b_1)\mathbf{v}_1 + \cdots + (a_m + b_m)\mathbf{v}_m$; and since this is a linear combination of $\mathbf{v}_1,\cdots,\mathbf{v}_m$, it is again in U. So U is closed under addition.
3. If $\mathbf{u} \in U$ is as above, and $k \in \mathbb{R}$, then $k\mathbf{u} = (ka_1)\mathbf{v}_1 + \cdots + (ka_m)\mathbf{v}_m$, which is again in U. So U is closed under scalar multiplication.

Thus U is a subspace of V.

The proof of (2) is a straightfoward[2] exercise using the closure properties of subspaces. ■

■ **Example 6.4.2**

$$\begin{aligned} W &= \{(x,y,x-y) \mid x,y \in \mathbb{R}\} \\ &= \{x(1,0,1) + y(0,1,-1) \mid x,y \in \mathbb{R}\} \\ &= \text{span}\{(1,0,1),(0,1,-1)\} \end{aligned}$$

[2] If you don't think so, please see the professor!

Therefore, W is a subspace of $\mathbb{R}^3$. ■

! If you notice that your set is the *span* of some vectors, then (by this theorem) you automatically know that it is a vector space. You don't even have to use the subspace test! Theorem 6.4.1 guarantees this. (Theorems are great for this kind of time-saving.)

6.5 Applying Theorem 6.4.1 to identify more subspaces

Let's see how this theorem can help us.

Problem 6.5.1 Let $V = F(\mathbb{R})$ and let $f(x) = \cos(x)$, $g(x) = \sin(x)$. Set $W = \text{span}\{f, g\}$. This is a subspace of $F(\mathbb{R})$, by the theorem.

Let's ask ourselves two questions:

(a) Is $\sin(x+1) \in W$?

(b) Is the constant function $h(x) = 1$ in W?

Solution For (a), we recall some nice trig identities for sums of angles:

$$\sin(x+1) = \sin(x)\cos(1) + \sin(1)\cos(x) = \sin(1)f(x) + \cos(1)g(x).$$

Since $\sin(1)$ and $\cos(1)$ are just numbers, this is just a linear combination of f and g. So YES, $\sin(x+1) \in W$.

For (b), we might have no inkling of what to do. The best way to proceed is to get some clues: plug in a few values of x, and see what happens. That is, *suppose* $h(x) = af(x) + bg(x)$ for some scalars a and b. Then we'd have:

$$\text{For } x = 0 : h(0) = af(0) + bg(0) \Rightarrow 1 = a$$
$$\text{For } x = \pi/2 : 1 = a(0) + b(1) \Rightarrow 1 = b$$

Excellent so far! So IF it is true that $h(x) = af(x) + bg(x)$ THEN necessarily we must have $a = 1$ and $b = 1$, so that $h(x) = \cos(x) + \sin(x)$. But wait a minute – if you plug in $x = \pi$, the left hand side is 1, but the right hand side is -1. So this isn't true; it's nonsense, a contradiction. Something went wrong. The substitutions are fine, so it must be that our supposition, our *hypothesis* (that h is a linear combination of f and g) must be false. We can then be sure, as we are that $1 \neq -1$, that h is not in W.

■ **Example 6.5.2** The set $W = \{a(1,0,1) + b(2,1,1) \mid a, b \geq 0\}$ is NOT a subspace[3]. It is also NOT the span of the vectors $(1,0,1)$ and $(2,1,1)$, because the span is the set of *all* linear combinations (not just some). Moral: the parameters have to be allowed to take on all real numbers. ■

For the next example, we need a definition.

Definition 6.5.3 Given an $n \times n$ ("square") matrix A, define the *trace of A* to be the sum of the

[3] Why not?

elements on its main diagonal, denoted $\operatorname{tr}(A)$. Then $\operatorname{tr}(A) \in \mathbb{R}$. For example,

$$\operatorname{tr}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = a + d.$$

Problem 6.5.4 Show that the set $\mathfrak{sl}_2 = \{A \in M_{22}(\mathbb{R}) \mid \operatorname{tr}(A) = 0\}$ is a subspace of $M_{22}(\mathbb{R})$. (This is in fact an example of a *Lie algebra* [a]

Solution Using the formula in the definition above, we can rewrite this set as

$$\mathfrak{sl}_2 = \left\{ \begin{bmatrix} a & b \\ c & -a \end{bmatrix} \mid a, b, c \in \mathbb{R} \right\}.$$

so

$$\mathfrak{sl}_2 = \operatorname{span}\left\{ \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}.$$

Since $\mathfrak{sl}_2$ is the span of some vectors, it is a subspace.

[a] “Lie" is pronounced “lee", named after Sophus Lie, a Norwegian mathematician. Lie Algberas, and Lie Groups — they’re great stuff. No, really — Sophus Lie’s idea of ‘continous symmetry’, now called Lie Groups was a huge boost to mathematics and the ‘Lie Algebra’ idea is a way of rendering the infinite to the finite, just as we have done with the idea of a spanning set for a subpace. See the wiki page for Sophus Lie.

There’s a funny but apocryphal story about a reseacher who was studying Lie Algebras, and who applied for, and received, public funding—after proper peer review and also quite proper government review. Also quite properly, a member of Canada’s Parliament at the time read part of the proposal. The Honourable Member then asked in the House of Commons why Canada should support research into ‘lying’. The meaning of ‘Lie’ Algebras was afterwards explained to him.

■ **Example 6.5.5** Consider the set of 2×2 *diagonal matrices*, that is

$$D_2 = \left\{ \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} \mid a, d \in \mathbb{R} \right\}.$$

Since

$$D_2 = \left\{ a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \mid a, d \in \mathbb{R} \right\} = \operatorname{span}\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

we deduce that D_2 is a subspace of $M_{22}(\mathbb{R})$. ■

6.6 So what are all the subspaces of $\mathbb{R}^n$?

We can answer this geometrically for $n = 1, 2, 3$; and this certainly gives us a sense of the answer for all vector spaces.

6.6.1 Subspaces of $\mathbb{R}$

Well, $\{\mathbf{0}\}$ is a subspace of $\mathbb{R}^n$, for any n; here, $\{0\}$ is a subspace of $\mathbb{R}$.

If W is a subspace of $\mathbb{R}$ which is not the zero subspace, then it contains some nonzero element, call it $\mathbf{x}$. But then, being a subspace, it contains all multiples of $\mathbf{x}$—and that’s all of $\mathbb{R}$!

Conclusion: $\mathbb{R}$ has only two subspaces: the zero space and the whole thing.

6.6.2 Subspaces of $\mathbb{R}^2$

OK, again we get that $\{\mathbf{0}\}$ and $\mathbb{R}^2$ are subspaces of $\mathbb{R}^2$. But we also know that lines through the origin are subspaces of $\mathbb{R}^2$.

Theorem 6.6.1 — Subspaces of $\mathbb{R}^2$. The *only* subspaces of $\mathbb{R}^2$ are

- the zero subspace,
- lines through the origin, and
- all of $\mathbb{R}^2$.

Proof. To prove this, start by letting W be an arbitrary subspace of $\mathbb{R}^2$. We want to argue that it has no choice but to be one of the above subspaces.

Suppose W is a subspace of $\mathbb{R}^2$.
If it's not the zero subspace, then it contains at least one nonzero vector, call it $\mathbf{v}$.
Since W is a subspace, it must contain span$\{\mathbf{v}\}$. (Theorem 6.4.1, part (2))
Now span$\{\mathbf{v}\}$ is the line through the origin with direction vector $\mathbf{v}$.
So if $W \neq$ span$\{\mathbf{v}\}$, it must be bigger; it must contain a vector not on that line, call it $\mathbf{w}$.
Then $\mathbf{v}$ and $\mathbf{w}$ are not parallel; they are not *collinear*.
By (Theorem 6.4.1, part (2)): W contains span$\{\mathbf{v},\mathbf{w}\}$.
So we are done if we can show that span$\{\mathbf{v},\mathbf{w}\} = \mathbb{R}^2$.

The last step of the proof, as done in class, was entirely geometric. Here's an algebraic proof.

Write $\mathbf{v} = (v_1, v_2)$ and $\mathbf{w} = (w_1, w_2)$. Let (x, y) be some arbitrary element of $\mathbb{R}^2$. To show that $(x, y) \in$ span$\{\mathbf{v},\mathbf{w}\}$, we need to show that we can solve the equation

$$\begin{bmatrix} x \\ y \end{bmatrix} = a \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + b \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

for some $a, b \in \mathbb{R}$. This is the same as solving the linear system for a and b:

$$\begin{aligned} av_1 + bw_1 &= x \\ av_2 + bw_2 &= y \end{aligned}$$

Since $\mathbf{v}$ and $\mathbf{w}$ aren't multiples of each other, the area of the parallelogram they generate – $|v_1w_2 - v_2w_1|$– isn't zero. So $v_1w_2 - v_2w_1 \neq 0$. Then it is easy to check that:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{xw_2 - yw_1}{v_1w_2 - v_2w_1}\mathbf{v} + \frac{v_1y - v_2x}{v_1w_2 - v_2w_1}\mathbf{w}.$$

That is, these nasty looking fractions are the values of a and b you'd find by solving the equation above.

(For example, the first coordinate of the right hand side is

$$\frac{xw_2 - yw_1}{v_1w_2 - v_2w_1}v_1 + \frac{v_1y - v_2x}{v_1w_2 - v_2w_1}w_1 = \frac{(xw_2v_1 - yw_1v_1 + v_1yw_1 - v_2xw_1)}{v_1w_2 - v_2w_1} = x$$

which equals the first coordinate of the left hand side.) ∎

! We learned: if two nonzero vectors $\mathbf{v},\mathbf{w} \in \mathbb{R}^2$ are not collinear (not parallel) then span$\{\mathbf{v},\mathbf{w}\} = \mathbb{R}^2$!

6.6.3 Subspaces of $\mathbb{R}^3$

We know several kinds of subspaces of $\mathbb{R}^3$; but in fact, once again, we know them all!

Theorem 6.6.2 — Subspaces of $\mathbb{R}^3$. The only subspaces of $\mathbb{R}^3$ are:
- the zero subspace,
- lines through the origin,
- planes through the origin, and
- all of $\mathbb{R}^3$.

This is something we'll be able to prove algebraically later, when we have more tools at our disposal. For now, we can be convinced by a geometric argument, along the following lines:

Let W be some arbitrary subspace. If it's not the zero space, then it contains $\operatorname{span}\{\mathbf{v}\}$ for some nonzero vector $\mathbf{v} \in W$. If it's not that line, then it contains $\operatorname{span}\{\mathbf{v},\mathbf{w}\}$ for some $\mathbf{w} \in W$ which is not collinear with $\mathbf{v}$. Arguing as in $\mathbb{R}^2$, we deduce that this span is a plane through the origin. If W is not equal to this plane, then it must contain another vector $\mathbf{u}$ such that $\mathbf{u} \notin \operatorname{span}\{\mathbf{v},\mathbf{w}\}$. With the help of another picture, we realize that every vector in all of $\mathbb{R}^3$ lies in $\operatorname{span}\{\mathbf{v},\mathbf{w},\mathbf{u}\}$, and so $W = \mathbb{R}^3$.

6.7 Final thoughts on spans, and some difficulties

Problem 6.7.1 Show that

$$\operatorname{span}\{(0,1,1),(1,0,1)\} = \operatorname{span}\{(1,1,2),(-1,1,0)\} \tag{6.1}$$

Solution Here are two ways of solving this problem. The first one only works in $\mathbb{R}^3$, but it's nice and short. The second one works in any vector space.

(1) We saw above that the span of two non-collinear vectors is a plane through the origin. We can get a normal vector to the plane using the cross product.

So: the plane spanned by $\{(0,1,1),(1,0,1)\}$ has normal vector $(0,1,1)\times(1,0,1) = (1,1,-1)$; so it is given by the normal equation $x+y-z=0$. The plane spanned by $\{(1,1,2),(-1,1,0)\}$ has normal vector $(1,1,2)\times(-1,1,0) = (-2,-2,2)$ so has equation $-2x-2y+2z=0$. But these equations describe the same plane! So the two sides of (6.1) are equal.

(2) What if we didn't have some insight into what these subspaces were (geometrically speaking)? Well, we can use our big theorem on span. Namely:

Since

$$(1,1,2) = 1(0,1,1)+1(1,0,1) \qquad \text{and} \qquad (-1,1,0) = 1(0,1,1)-1(1,0,1)$$

we have that $(1,1,2),(-1,1,0) \in \operatorname{span}\{(0,1,1),(1,0,1)\}$. So by the theorem 6.4.1(2),

$$\operatorname{span}\{(1,1,2),(-1,1,0)\} \subseteq \operatorname{span}\{(0,1,1),(1,0,1)\}.$$

Conversely, since

$$(0,1,1) = \frac{1}{2}(1,1,2)+\frac{1}{2}(-1,1,0) \qquad \text{and} \qquad (1,0,1) = \frac{1}{2}(1,1,2)-\frac{1}{2}(-1,1,0)$$

we have that $(0,1,1),(1,0,1) \in \text{span}\{(1,1,2),(-1,1,0)\}$, so again by Theorem 6.4.1(2),

$$\text{span}\{(0,1,1),(1,0,1)\} \subseteq \text{span}\{(1,1,2),(-1,1,0)\}.$$

But if you have two sets W and U and you know $W \subseteq U$ and $U \subseteq W$, then you must have $W = U$. Hence we know the spans are equal.

! A first issue with spans: it's not easy to tell if two subspaces are equal, just based on the spanning sets you're given.

Problem 6.7.2 Show that $\text{span}\{(0,1,1),(1,0,1)\} = \text{span}\{(0,1,1),(1,0,1),(1,1,2),(-1,1,0)\}$.
Solution The short answer: from the previous problem we know that all the vectors on the right hand side lie in the plane $x+y-z=0$, so their span can't be any bigger than that (since the span is the SMALLEST subspace that contains the given vectors). And it can't be any smaller than the plane, since you can already get every vector of that plane via linear combinations of just two of them. So they are equal.

The more complete answer: use the method (2) above. Clearly

$$(0,1,1),(1,0,1) \in \text{span}\{(0,1,1),(1,0,1),(1,1,2),(-1,1,0)\}$$

(since they're in the spanning set!), so

$$\text{span}\{(0,1,1),(1,0,1)\} \subseteq \text{span}\{(0,1,1),(1,0,1),(1,1,2),(-1,1,0)\}.$$

On the other hand, using the previous example, we have that every one of the 4 vectors on the right hand side lie in the span of $(0,1,1)$ and $(1,0,1)$, so we may similarly conclude that

$$\text{span}\{(0,1,1),(1,0,1),(1,1,2),(-1,1,0)\} \subseteq \text{span}\{(0,1,1),(1,0,1)\}.$$

Thus, the subspaces are equal.

! A second issue: Having more vectors in a spanning set DOESN'T imply that the subspace they span is any bigger. You can't judge the size of a subspace by just looking at the spanning set.

(Or can you? That's coming up soon.)

Problems

Problem 6.1 Justify your answers to the following:

a) Is the vector $(1,2)$ a linear combination of $(1,0)$ and $(1,1)$?

b) *Is the vector $(1,2)$ a linear combination of $(1,1)$ and $(2,2)$?

c) Is the vector $(1,2)$ a linear combination of $(1,0)$ and $(2,2)$?

d) *Is the vector $(1,2,2,3)$ a linear combination of $(1,0,1,2)$ and $(0,0,1,1)$?

e) Is the vector $(1,2,2,4)$ a linear combination of $(1,0,1,2)$ and $(0,0,1,1)$?

f) *Is the matrix $\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$ a linear combination of $\begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$?

g) Is the matrix $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ a linear combination of $\begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$?

h) *Is the polynomial $1+x^2$ a linear combination of $1+x-x^2$ and x?

i) Is the polynomial $1+x^2$ a linear combination of $1+x$ and $1-x$?

j) *Is the function $\sin x$ a linear combination of the constant function 1 and $\cos x$?

k) Is the function $\sin^2 x$ a linear combination of the constant function 1 and $\cos^2 x$?

l) *If u, v and w are any vectors in a vector space V, is $u-v$ a linear combination of u, v and w?

m) If u, v and w are any vectors in a vector space V, is w always a linear combination of u, v ?

Problem 6.2 Justify your answers to the following:

a) Does the vector $(3,4)$ belong to $\operatorname{span}\{(1,2)\}$?

b) *Is '$(3,4) \in \operatorname{span}\{(1,2)\}$' true? (Note that this is the same as the previous part, written using mathematical notation.)

c) Is '$(2,4) \in \operatorname{span}\{(1,2)\}$' true?

d) *How many vectors belong to $\operatorname{span}\{(1,2)\}$?

e) Are the subsets $\{(1,2)\}$ and $\operatorname{span}\{(1,2)\}$ equal?

f) *Is $\{(1,2)\}$ a subset of $\operatorname{span}\{(1,2)\}$?

g) If $\mathbf{v}$ is a non-zero vector in a vector space, show that there are infinitely many vectors in $\operatorname{span}\{\mathbf{v}\}$. (Is this true if $\mathbf{v}=\mathbf{0}$?)

h) *Suppose S is a subset of a vector space V. If $S = \operatorname{span} S$, explain why S must be a subspace of V.

i) * Suppose S is a subset of a vector space V. If S is actually a *subspace* of V, show that $S = \operatorname{span} S$.

Note that to show that $S = \operatorname{span} S$, you must establish two facts:
(i) If $\mathbf{w} \in S$, then $\mathbf{w} \in \operatorname{span} S$, and
(ii) If $\mathbf{w} \in \operatorname{span} S$, then $\mathbf{w} \in S$.

Problem 6.3 Give two distinct *finite* spanning sets for each of the following subspaces.

a) $\{(2x,x) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$

b) $^\star\{(x,y) \in \mathbb{R}^2 \mid 3x - y = 0\}$

c) $\{(x,y,z) \in \mathbb{R}^3 \mid x + y - 2z = 0\}$

d) $^\star\{(x,y,z,w) \in \mathbb{R}^4 \mid x - y + z - w = 0\}$

e) $\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \,\middle|\, b = c \right\}.$

f) $^\star\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \,\middle|\, a = d = 0 \quad \text{and} \quad b = -c \right\}.$

g) $\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \,\middle|\, a + d = 0 \right\}.$

h) $^\star\left\{ \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \,\middle|\, a, b \in \mathbb{R} \right\}.$

i) $\left\{ \begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \,\middle|\, b \in \mathbb{R} \right\}.$

j) $^\star\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \,\middle|\, a + b + c + d = 0 \right\}.$

k) $\mathbb{P}_2 = \{p \mid p \text{ is a polynomial function with} \deg(p) \leq 2\}.$

l) $^\star\mathbb{P}_n = \{p \mid p \text{ is a polynomial function with} \deg(p) \leq n\}.$

m) $\{p \in \mathbb{P}_2 \mid p(2) = 0\}.$

n) $^\star\{p \in \mathbb{P}_3 \mid p(2) = p(3) = 0\}.$

o) $\{p \in \mathbb{P}_2 \mid p(1) + p(-1) = 0\}.$

p) $^\star\text{span}\{\sin x, \cos x\}.$

q) $\text{span}\{1, \sin x, \cos x\}.$

r) $^\star\text{span}\{1, \sin^2 x, \cos^2 x\}.$

s) * $\{(x, x-3) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$, equipped with the *non-standard operations:–* Addition:

$$(x,y)\tilde{+}(x',y') = (x + x', y + y + 3).$$

Multiplication of vectors by scalars $k \in \mathbb{R}$:

$$k \odot (x,y) = (kx, ky + 3k - 3).$$

t) * $\{(x,y,z) \in \mathbb{R}^3 \mid x+2y+z=2\}$; $V=\mathbb{R}^3$, <u>*Non-standard operations:–*</u> Addition:

$$(x,y,z)\tilde{+}(x',y',z') = (x+x',y+y,z+z'-2).$$

Multiplication of vectors by scalars $k \in \mathbb{R}$: $k \odot (x,y,z) = (kx,ky,kz-2k+2)$.

(Hint for parts (m)&(n): Recall the Factor/Remainder theorem from high school: if p is a polynomial in the variable x of degree at least 1, and $p(a)=0$ for some $a \in \mathbb{R}$, then p has a factor of $x-a$, i.e., $p(x)=(x-a)q(x)$, where q is a polynomial with $\deg(q)=\deg(p)-1$.*)*

Problem 6.4 Justify your answers to the following:

a) Suppose $\mathbf{u}$ and $\mathbf{v}$ belong to a vector space V. Show carefully that $\text{span}\{\mathbf{u},\mathbf{v}\} = \text{span}\{\mathbf{u},\mathbf{u}+\mathbf{v}\}$. That is , you must show two things:
  (i) If $\mathbf{w} \in \text{span}\{\mathbf{u},\mathbf{v}\}$, then $\mathbf{w} \in \text{span}\{\mathbf{u},\mathbf{u}+\mathbf{v}\}$, and
  (ii) If $\mathbf{w} \in \text{span}\{\mathbf{u},\mathbf{u}+\mathbf{v}\}$, then $\mathbf{w} \in \text{span}\{\mathbf{u},\mathbf{v}\}$.

b) *Suppose $\mathbf{u}$ and $\mathbf{v}$ belong to a vector space V. Show carefully that $\text{span}\{\mathbf{u},\mathbf{v}\} = \text{span}\{\mathbf{u}-\mathbf{v},\mathbf{u}+\mathbf{v}\}$. That is , you must show two things:
  (i) If $\mathbf{w} \in \text{span}\{\mathbf{u},\mathbf{v}\}$, then $\mathbf{w} \in \text{span}\{\mathbf{u}-\mathbf{v},\mathbf{u}+\mathbf{v}\}$, and
  (ii) If $\mathbf{w} \in \text{span}\{\mathbf{u}-\mathbf{v},\mathbf{u}+\mathbf{v}\}$, then $\mathbf{w} \in \text{span}\{\mathbf{u},\mathbf{v}\}$.

c) Suppose $\mathbf{u} \in \text{span}\{\mathbf{v},\mathbf{w}\}$. Show carefully that $\text{span}\{\mathbf{v},\mathbf{w}\} = \text{span}\{\mathbf{u},\mathbf{v},\mathbf{w}\}$.

d) *Suppose $\text{span}\{\mathbf{v},\mathbf{w}\} = \text{span}\{\mathbf{u},\mathbf{v},\mathbf{w}\}$. Show carefully that $\mathbf{u} \in \text{span}\{\mathbf{v},\mathbf{w}\}$.

e) Show that $x^2 \notin \text{span}\{1,x\}$.[a]

f) *Show that $x^{n+1} \notin \mathbb{P}_n$. [b]

g) * Show that $\mathbb{P}$ does not have a finite spanning set. [c]

h) *Assume for the moment (we'll prove it later) that if W is a subspace of V, and V has a finite spanning set, then so does W.
Use this fact and the previous part to prove that $\mathbf{F}(\mathbb{R})$ does not have a finite spanning set.

---

[a] *Hint: Proceed by contradiction: Suppose $x^2 \in$ span$\{1,x\}$. Write down explicitly what this means, re-write the equation as $q(x)=0$, where $q(x)$ is some quadratic polynomial. Now, remember: every non-zero polynomial of degree 2 has at most 2 distinct roots. Look again at the equation $q(x)=0$, and ask yourself how many roots this equation tells you that q has. Now find your contradiction.*

[b] *Hint: Generalize the idea in the previous hint, recalling that every non-zero polynomial of degree $n+1$ has at most $n+1$ distinct roots.*

[c] *Hint: Proceed by contradiction again. Suppose it did, say, $\mathbb{P} =$ span$\{p_1,\ldots,p_k\}$, for some polynomials $p_1,\ldots,p_k$. Now define $n=\max\{\deg(p_1),\ldots,\deg(p_k)\}$. Now show that $x^{n+1} \notin$ span$\{p_1,\ldots,p_k\}$, using the same argument as in the previous part. But of course $x^{n+1} \in \mathbb{P}$, so there's your contradiction.*

7. Linear Dependence and Independence

In the last chapter, we introduced the concept of the *span* of some vectors. Given finitely many vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$, we define their *span* to be the set of *all linear combinations* of these vectors. We write

$$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\} = \{a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_m\mathbf{v}_m \mid a_i \in \mathbb{R}\}.$$

Thus far, we should agree that:

- $\text{span}\{\mathbf{0}\} = \{\mathbf{0}\}$
- If $\mathbf{v} \neq \mathbf{0}$, then $\text{span}\{\mathbf{v}\}$ is the line through the origin in direction $\mathbf{v}$.
- If $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$ and $\mathbf{u}$ and $\mathbf{v}$ are not parallel, then $\text{span}\{\mathbf{u}, \mathbf{v}\}$ is a plane (at least: in $\mathbb{R}^2$ and $\mathbb{R}^3$).

(We also argued that three non-coplanar vectors should span all of $\mathbb{R}^3$, but this was only done verbally, and I don't expect you to feel convinced of that one, yet.)

Notice the (geometric!) conditions we had to place on our statements. For now, let's agree that they are necessary, and then work out what the correct algebraic version of these conditions should be, so that we can apply it to *any* vector space.

(We'll call the condition *linear independence* and one thing it will guarantee, as we'll see, is that if $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ is a linearly independent set, then the vector space $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ is strictly *bigger* than the vector space $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{m-1}\}$ (or the span of any proper subset).)

7.1 Difficulties with span: Nonuniqueness of spanning sets

The first issue to mention here is not one we're going to be able to solve; but we need to discuss it to be aware of the issue.

■ **Example 7.1.1** $\text{span}\{(1,2)\} = \text{span}\{(2,4)\}$ since these vectors are parallel and so span the same line. ■

■ **Example 7.1.2** If $W = \text{span}\{(1,0,1),(0,1,0)\}$ and $U = \text{span}\{(1,1,1),(1,-1,1)\}$ then in fact $W = U$. Why?

First note that $(1,1,1) = (1,0,1)+(0,1,0)$ and $(1,-1,1) = (1,0,1)-(0,1,0)$. Now apply Theorem 6.4.1:

The set $W = \text{span}\{(1,0,1),(0,1,0)\}$ is a subspace of $\mathbb{R}^3$ by Theorem 6.4.1(1).
We've just shown that $(1,1,1)$ and $(1,-1,1)$ are in W (because they are linear combinations of a spanning set for W).
Thus by part (2) of the theorem, $\text{span}\{(1,1,1),(1,-1,1)\} \subseteq W$.
So $U \subseteq W$.

Next, write $(1,0,1) = \frac{1}{2}(1,1,1)+\frac{1}{2}(1,-1,1)$ and $(0,1,0) = \frac{1}{2}(1,1,1)-\frac{1}{2}(1,-1,1)$, and apply the same argument as above, with the roles of U and W reversed, to deduce that $W \subseteq U$.

Hence $W = U$. ■

! Moral: You can't usually tell if two subspaces are equal just by looking at their spanning sets! Every vector space (except the zero vector space) has *infinitely many* spanning sets.

7.2 Difficulties with span: more vectors in the spanning set doesn't mean more vectors in their span

The second issue is more serious, but it is one that we can (and will) repair.

Problem 7.2.1 Show that

$$\text{span}\{(1,2)\} = \text{span}\{(0,0),(1,2)\} = \text{span}\{(1,2),(2,4),(3,6)\}.$$

Solution For the first equality, note that:

$$\text{span}\{(0,0),(1,2)\} = \{a(0,0)+b(1,2) \mid a,b \in \mathbb{R}\} = \{b(1,2) \mid b \in \mathbb{R}\} = \text{span}\{(1,2)\}.$$

Also, note that

$$\begin{aligned}\text{span}\{(1,2),(2,4),(3,6)\} &= \{a(1,2)+b(2,4)+c(3,6) \mid a,b,c \in \mathbb{R}\}\\ &= \{a(1,2)+2b(1,2)+3c(1,2) \mid a,b,c \in \mathbb{R}\}\\ &= \{(a+2b+3c)(1,2) \mid a,b,c \in \mathbb{R}\}\end{aligned}$$

This set is certainly contained in $\text{span}\{(1,2)\}$. Now, for any $t \in \mathbb{R}$, you could set $a = t, b = 0, c = 0$, to get $t(1,2)$ in this set. So we have

$$\text{span}\{(1,2),(2,4),(3,6)\} = \text{span}\{(1,2)\}$$

! Moral: The number of elements in the spanning set for a subspace W does not necessarily tell you how "big" W is.

Geometrically, we see that the problem is that all the vectors are collinear (meaning, parallel, or all lying on one line). Similar problems would occur in $\mathbb{R}^3$ if we had three *coplanar* vectors, that is, all lying in a plane. So what is the algebraic analogue of these statements?

7.3 Algebraic version of "2 vectors are collinear"

Two vectors $\mathbf{u}$ and $\mathbf{v}$ being *collinear*, or parallel, means that either they are multiples of one another or one of them is $\mathbf{0}$.

So it's not enough to just say $\mathbf{u} = k\mathbf{v}$ for some $k \in \mathbb{R}$ (because maybe $\mathbf{v} = \mathbf{0}$ and then this equation wouldn't work).

You could go with:

- $\mathbf{u} = k\mathbf{v}$ for some $k \in \mathbb{R}$, or
- $\mathbf{v} = k\mathbf{u}$ for some $k \in \mathbb{R}$.

But here's a nice compact way to get all the cases at once:

! Two vectors $\mathbf{u}$ and $\mathbf{v}$ are collinear if there exist scalars $a, b \in \mathbb{R}$, *not both zero*, such that

$$a\mathbf{u} + b\mathbf{v} = \mathbf{0}$$

Since the coefficients aren't both zero, you can rearrange this as above.

Example 7.3.1 The vectors $(1,2,1)$ and $(2,4,2)$ are collinear since

$$2(1,2,1) + (-1)(2,4,2) = (0,0,0).$$

■

Example 7.3.2 The vectors $(1,2,1)$ and $(0,0,0)$ are collinear since

$$0(1,2,1) + 3(0,0,0) = (0,0,0)$$

(note that at least one of the coefficients isn't zero). ■

7.4 Algebraic version of "3 vectors are coplanar"

Again, we could think of this as: one of the three vectors is in the span of the other two. So if the vectors are $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$, they are coplanar if

- $\mathbf{u} = a\mathbf{v} + b\mathbf{w}$ for some $a, b \in \mathbb{R}$ or
- $\mathbf{v} = a\mathbf{u} + b\mathbf{w}$ for some $a, b \in \mathbb{R}$ or
- $\mathbf{w} = a\mathbf{u} + b\mathbf{v}$ for some $a, b \in \mathbb{R}$.

(Where again we had to include all three cases just in case one of the vectors is the zero vector.)

By the same method as above, this can be simplified to:

! Three vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ are coplanar (lie in a plane) if there exists scalars $a, b, c \in \mathbb{R}$, *not all zero* such that

$$a\mathbf{u} + b\mathbf{v} + c\mathbf{w} = \mathbf{0}$$

■ **Example 7.4.1** $(1,0,1),(0,1,0),(1,1,1)$ are coplanar because

$$(1,0,1)+(0,1,0)-(1,1,1)=(0,0,0).$$

■

■ **Example 7.4.2** $(1,1,1),(2,2,2),(0,0,0)$ are coplanar because

$$2(1,1,1)+(-1)(2,2,2)+0(0,0,0)=(0,0,0)$$

or because

$$0(1,1,1)+0(2,2,2)+17(0,0,0)=(0,0,0)$$

(in fact these vectors were even collinear, but let's be happy with coplanar for now). ■

7.5 Linear DEPENDENCE: the algebraic generalization of "collinear" and "coplanar"

Definition 7.5.1 Let V be a vector space and $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_m \in V$. Then the set $\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_m\}$ is *linearly dependent* (or say: $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_m$ *are* linearly dependent) if and only if there are scalars $a_1,a_2,\ldots,a_m \in \mathbb{R}$, *not all zero* such that

$$a_1\mathbf{v}_1+\cdots+a_m\mathbf{v}_m=\mathbf{0}.$$

This (and the definition of linear independence, coming up next) are *key concepts* for this course. Take the time to note what the definition does *not* say: it doesn't just say that there's a solution to the *dependence equation* $a_1\mathbf{v}_1+\cdots+a_m\mathbf{v}_m=\mathbf{0}$; it says there is a *nontrivial solution* to it. That is a huge difference.

7.6 The contrapositive and linear INDEPENDENCE

So linear dependence is a concept (valid for any number of vectors, in any vector space, of any dimension) that captures the idea that our set of vectors has some "redundancy", such as being collinear or coplanar. The definition is quite complicated, so it is worth asking: so, what is the *opposite* of a set being linearly dependent?

Rephrasing what we have above, we can say:

- Two vectors $\mathbf{u}$ and $\mathbf{v}$ are not collinear (equivalently, $\mathbf{u}$ and $\mathbf{v}$ are not parallel) if and only if

$$a\mathbf{u}+b\mathbf{v}=\mathbf{0} \Leftrightarrow a=b=0$$

  In words: they are not parallel if and only the ONLY solution to the equation $a\mathbf{u}+b\mathbf{v}=\mathbf{0}$ is the *trivial* solution $a=b=0$ (which of course is always a solution).
- Three vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ are not coplanar if and only if

$$a\mathbf{u}+b\mathbf{v}+c\mathbf{w}=\mathbf{0} \Leftrightarrow a=b=c=0$$

  In words: they are not coplanar if and only if the ONLY solution to the equation $a\mathbf{u}+b\mathbf{v}+c\mathbf{w}=\mathbf{0}$ is the trivial solution $a=b=c=0$.

Definition 7.6.1 Let V be a vector space and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m \in V$. Then the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ is *linearly independent* (or say: $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ *are* linearly independent) if and only if the *only solution to*

$$a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m = \mathbf{0}$$

is the trivial solution $a_1 = 0, \cdots, a_m = 0$.

! Note that if some vectors are not linearly dependent, then they are linearly independent, and vice versa.

We will abbreviate LI = linearly independent and LD = linearly dependent.

7.7 Examples

Let's work with this definition in several examples.

Example 7.7.1 $\{(1,0),(0,1)\}$ is LI because when you try to solve

$$a(1,0)+b(0,1) = (0,0)$$

you get

$$(a,b) = (0,0)$$

which means $a = 0, b = 0$. No choice! ■

Example 7.7.2 $\{(1,1),(1,-1)\}$ is LI because when you try to solve

$$a(1,1)+b(1,-1) = (0,0)$$

you get

$$(a+b, a-b) = (0,0)$$

which gives $a = b$ and $a = -b$, which forces $a = 0$ and $b = 0$, No choice! ■

Example 7.7.3 $\{(1,0),(0,1),(1,1)\}$ is LD since when you try to solve

$$a(1,0)+b(0,1)+c(1,1) = (0,0)$$

you notice that $a = 1$, $b = 1$ and $c = -1$ works. (In fact, you have infinitely many nontrivial solutions!) Since we found a nontrivial solution, they are dependent. (Note that $(1,1) \in \text{span}\{(1,0),(0,1)\}$.) ■

Example 7.7.4 $\{(1,1,1)\}$ is LI since solving $k(1,1,1) = \mathbf{0}$ forces $k = 0$. ■

Example 7.7.5 $\left\{\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right\}$ is LI since when you try to solve

$$a_1\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + a_2\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} + a_3\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix} + a_4\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix} = \begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}$$

you get, by simplifying the left side,

$$\begin{bmatrix}a_1 & a_2\\a_3 & a_4\end{bmatrix} = \begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}$$

which means all your coefficients a_i are forced to be zero. ■

Example 7.7.6 $\left\{\begin{bmatrix}1&0\\0&0\end{bmatrix},\begin{bmatrix}0&0\\0&1\end{bmatrix},\begin{bmatrix}1&0\\0&-1\end{bmatrix},\begin{bmatrix}0&1\\0&0\end{bmatrix}\right\}$ is LD since we notice that

$$\begin{bmatrix}1&0\\0&0\end{bmatrix}-\begin{bmatrix}0&0\\0&1\end{bmatrix}-\begin{bmatrix}1&0\\0&-1\end{bmatrix}+0\begin{bmatrix}0&1\\0&0\end{bmatrix}=\begin{bmatrix}0&0\\0&0\end{bmatrix}.$$

Note that the third matrix was in the span of the first two. ■

Example 7.7.7 $\{1,x,x^2\}$ is LI in $\mathbb{P}_2$ since

$$a(1)+b(x)+c(x^2)=\mathbf{0}$$

means that

$$a+bx+cx^2=0$$

for *all real values of* x. If $c\neq 0$, the left hand side is a is a quadratic polynomial which can at at most two distinct real roots! So $c=0$. Then we have

$$a+bx=0$$

for *all real values of* x. But if $b\neq 0$, this equation has the single solution $x=-\dfrac{a}{b}$! So b must be zero. Now we see that $a=0$. Hence $a=b=c=0$.[1] ■

Example 7.7.8 $\{4+4x+x^2,1+x,x^2\}$ is LD in $\mathbb{P}_2$ since

$$1(4+4x+x^2)+(-4)(1+x)+(-1)x^2=0$$

is a nontrivial dependence equation. In fact, $4+4x+x^2\in\text{span}\{1+x,x^2\}$. ■

Example 7.7.9 $\{1,\sin x,\cos x\}$ is LI in $F(\mathbb{R})$ because suppose we have the equation

$$a(1)+b\sin(x)+c\cos(x)=0$$

then, plugging in some values of x gives you:

- $x=0: a+c=0$
- $x=\pi/2: a+b=0$
- $x=\pi: a-c=0$

And we see from this (with a little bit of work) that $a=b=c=0$. ■

Example 7.7.10 $\{1,\sin^2 x,\cos^2 x\}$ is LD. Why? Since

$$(-1)1+(1)\sin^2 x+(1)\cos^2 x=0,\quad \text{for all } x\in\mathbb{R},$$

this is a nontrivial dependence relation. Note that $\cos^2 x\in\text{span}\{1,\sin^2 x\}$. ■

7.8 Facts (theorems) about linear independence and linear dependence

Let's see how many general facts we can figure out about LI and LD sets. Let V be a vector space.

[1] Compare this argument with the hints given for problems 6.4 parts (e) and (f).

FACT 1: If $\mathbf{v} \in V$, then $\{\mathbf{v}\}$ is linearly independent if and only if $\mathbf{v} \neq \mathbf{0}$.

Proof. if $\mathbf{v} \neq \mathbf{0}$, then $k\mathbf{v} = \mathbf{0}$ only has solution $k = 0$ (LI); but if $\mathbf{v} = \mathbf{0}$, then we have, for example, $3\mathbf{v} = \mathbf{0}$ (LD). ■

FACT 2: If $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is LD, then any set containing $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is also LD.

Point: you can't "fix" a dependent set by adding more vectors to it. In $\mathbb{R}^3$, if 2 vectors are collinear, then taking a third vector gives you at most a coplanar set.

Proof. Your set is LD so you have a nontrivial dependence relation

$$a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m = \mathbf{0}$$

with not all $a_i = 0$. Now consider the bigger set $\{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{u}_1, \ldots, \mathbf{u}_k\}$. Then you have the equation

$$a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m + 0\mathbf{u}_1 + \cdots + 0\mathbf{u}_k = \mathbf{0}$$

and not all coefficients are zero. So that means the big set is LD, too. ■

FACT 3: If $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ is LI, then any subset is also LI.

■ **Example 7.8.1** If three vectors are not coplanar, then for sure, no pair of them is collinear. ■

Proof. Fact 3 is in fact logically equivalent to Fact 2. Fact 2 says that if you contain an LD subset, then you are LD. So you can't be LI and contain an LD subset. ■

FACT 4: $\{\mathbf{0}\}$ is LD.

FACT 5: Any set containing the zero vector is LD.

! NOTICE: this means that subspaces are LD! And that's perfectly OK. We are only usually interested in whether or not *spanning sets* are LD or LI.

Proof. Use Fact 4 with Fact 2. ■

FACT 6: A set $\{\mathbf{u}, \mathbf{v}\}$ is LD if and only if one of the vectors is a multiple of the other.

Proof. If $a\mathbf{u} + b\mathbf{v} = \mathbf{0}$ is a nontrivial dependence relation and $a \neq 0$, then $\mathbf{u} = -\frac{b}{a}\mathbf{v}$. If $b \neq 0$ then $\mathbf{v} = -\frac{a}{b}\mathbf{u}$. (And if both are zero, it's not a nontrivial dependence relation!) Conversely, if $\mathbf{u} = c\mathbf{v}$ (or $\mathbf{v} = a\mathbf{u}$) for some $c \in \mathbb{R}$, then $\mathbf{u} - c\mathbf{v} = 0$ (or $-c\mathbf{u} + \mathbf{v} = 0$) is a nontrivial dependence relation. ■

FACT 7: A set with three or more vectors can be LD *even though* no two vectors are multiples of one another.

Example 7.8.2 $\{(1,0),(0,1),(1,1)\}$ are coplanar but no two vectors are collinear. ■

FACT 8: (See Theorem 8.1.1 in the next chapter) A set $\{\mathbf{v}_1,\ldots,\mathbf{v}_m\}$ is LD if and only if there is at least one vector $\mathbf{v}_k \in \{\mathbf{v}_1,\ldots,\mathbf{v}_m\}$ which is in the span of $\{\mathbf{v}_1,\ldots,\mathbf{v}_{k-1},\mathbf{v}_{k+1},\ldots,\mathbf{v}_m\}$.

Example 7.8.3 (using the above example) $(1,1) \in \text{span}\{(1,0),(0,1)\}$. ■

! Caution: Fact 8 *doesn't mean* that *every* vector is a linear combination of the others. For example, $\{(1,1),(2,2),(1,3)\}$ is LD BUT $(1,3) \notin \text{span}\{(1,1),(2,2)\}$.

Proof. Suppose $\{\mathbf{v}_1,\cdots,\mathbf{v}_m\}$ is LD. So there is some nontrivial dependence equation

$$a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m = \mathbf{0}$$

with not all $a_i = 0$. Let's say that $a_1 \neq 0$ (we could always renumber the vectors so that this is the case). Then we can solve for $\mathbf{v}_1$:

$$\mathbf{v}_1 = -\frac{a_2}{a_1}\mathbf{v}_2 - \cdots - \frac{a_m}{a_1}\mathbf{v}_m$$

which just says $\mathbf{v}_1 \in \text{span}\{\mathbf{v}_2,\cdots,\mathbf{v}_m\}$. Done.

Now the other direction. Suppose $\mathbf{v}_m \in \text{span}\{\mathbf{v}_1,\cdots,\mathbf{v}_{m-1}\}$. That means $\mathbf{v}_m = b_1\mathbf{v}_1 + \cdots + b_{n-1}\mathbf{v}_{m-1}$. So we have

$$b_1\mathbf{v}_1 + \cdots + b_{n-1}\mathbf{v}_{m-1} + (-1)\mathbf{v}_m = \mathbf{0}$$

and the coefficient of $\mathbf{v}_n$ is -1, which is nonzero. Hence this is a nontrivial dependence relation (it doesn't even matter if all the b_i were zero). Thus the set is LD. ■

Next chapter, we'll put these ideas together with spanning sets.

Problems

Remarks:

1. A question with an asterisk '*' (or two) indicates a bonus-level question.
2. You must justify all your responses.

Problem 7.1 Which of the following sets are linearly independent in the indicated vector space? (If you say they are, you must prove it using the definition; if you say the set is dependent, you must give a non-trivial linear dependence relation that supports your answer. For example, if you say $\{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\}$ is dependent, you must write something like $\mathbf{v}_1 - 2\mathbf{v}_2 + \mathbf{v}_3 = 0$, or $\mathbf{v}_1 = 2\mathbf{v}_2 - \mathbf{v}_3$.)

a) $\{(1,1),(1,2)\}$; $\mathbb{R}^2$

b) $^\star\{(1,1),(2,2)\}$; $\mathbb{R}^2$.

c) $\{(0,0),(1,1)\}$; $\mathbb{R}^2$

d) $^\star\{(1,1),(1,2),(1,0)\}$; $\mathbb{R}^2$.

e) $\{(1,1,1),(1,0,3)\}$; $\mathbb{R}^3$

f) $^\star\{(1,1,1),(1,0,3),(0,0,0)\}$; $\mathbb{R}^3$.

g) $\{(1,1,1),(1,2,3),(2,3,4)\}$; $\mathbb{R}^3$

h) $^\star\{(1,1,1),(1,0,3),(0,3,4)\}$; $\mathbb{R}^3$.

i) $\{(1,0,0),(0,1,0),(0,0,1),(1,2,3)\}$; $\mathbb{R}^3$

j) $^\star\{(0,-3),(3,0)\}$; $\mathbb{R}^2$.

k) $*$[a] $\{(0,-3),(3,0)\}$; $V=\mathbb{R}^2$, but equipped with the <u>*non-standard operations:–*</u>

$$\text{Addition:} \quad (x,y)\tilde{+}(x',y') = (x+x', y+y+3).$$

$$\text{Multiplication of vectors by scalars } k\in\mathbb{R}: \quad k\odot(x,y) = (kx, ky+3k-3).$$

l) $^\star\{(1,0,0),(2,0,-2)\}$; $\mathbb{R}^3$.

m) $*$[b] $\{(1,0,0),(2,0,-2)\}$; $V=\mathbb{R}^3$, but equipped with <u>*non-standard operations:–*</u>

$$\text{Addition:} \quad (x,y,z)\tilde{+}(x',y',z') = (x+x', y+y, z+z'-2).$$

$$\text{Multiplication of vectors by scalars } k\in\mathbb{R}: \quad k\odot(x,y,z) = (kx, ky, kz-2k+2)$$

[a] You may be surprised by the correct answer in this example. See the professor for details.

[b] You may not be so surprised by the correct answer in this example if you've already done part (k). See the professor for details.

Problem 7.2 Which of the following sets are linearly independent in $\mathbf{M}_{22}$? (If you say they are, you must prove it using the definition; if you say set is dependent, you must give a non-trivial linear dependence relation that supports your answer. For example, if you say $\{A_1,A_2,A_3\}$ is dependent, you must write something like $A_1-2A_2+A_3=0$, or $A_1=2A_2-A_3$.)

a) $\left\{\begin{bmatrix}1&0\\1&2\end{bmatrix},\begin{bmatrix}0&1\\0&1\end{bmatrix}\right\}$

b) $^\star\left\{\begin{bmatrix}1&0\\1&2\end{bmatrix},\begin{bmatrix}0&1\\0&1\end{bmatrix},\begin{bmatrix}1&-2\\-1&0\end{bmatrix}\right\}$

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

d) $^\star\left\{\begin{bmatrix}1&0\\0&0\end{bmatrix},\begin{bmatrix}0&1\\0&0\end{bmatrix},\begin{bmatrix}0&0\\1&0\end{bmatrix},\begin{bmatrix}0&0\\0&1\end{bmatrix}\right\}$

Problem 7.3 Which of the following sets are linearly independent in the indicated vector space ? (If you say they are, you must prove it using the definition; if you say set is dependent, you must give a non-trivial linear dependence relation that supports your answer. For example, if you say $\{f_1, f_2, f_3\}$ is dependent, you must write something like $f_1 - 2f_2 + f_3 = 0$, or $f_1 = 2f_2 - f_3$.)

a) $\{1, 1-x, 1-2x\}$; $\mathbb{P}_2$.

b) $^\star\{1, 1+x, x^2\}$; $\mathbb{P}_2$.

c) $\{\sin x, \cos x\}$; $\mathbf{F}(\mathbb{R})$.

d) $^\star\{1, \sin x, 2\cos x\}$; $\mathbf{F}(\mathbb{R})$.

e) $\{2, 2\sin^2 x, 3\cos^2 x\}$; $\mathbf{F}(\mathbb{R})$.

f) $^\star\{\cos 2x, \sin^2 x, \cos^2 x\}$; $\mathbf{F}(\mathbb{R})$.

g) $\{\cos 2x, 1, \sin^2 x\}$; $\mathbf{F}(\mathbb{R})$.

h) $^\star\{\sin 2x, \sin x \cos x\}$; $\mathbf{F}(\mathbb{R})$.

i) $\{\sin(x+1), \sin x, \cos x\}$; $\mathbf{F}(\mathbb{R})$.

Problem 7.4 Justify your answers to the following:

a) Suppose V is a vector space, and a subset $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} \subset V$ is known to be linearly independent. Show carefully that $\{\mathbf{v}_2, \mathbf{v}_3\}$ is also linearly independent.

b) *Suppose V is a vector space, and a subset $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset V$ is known to be linearly independent. Show carefully that $\{\mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is also linearly independent.

c) Suppose V is a vector space, and a subset $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset V$ is known to be linearly independent. If S is *any* subset of V with $S \subseteq \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$, show carefully that S is also linearly independent.

d) *Give an example of a linearly independent subset $\{\mathbf{v}_1, \mathbf{v}_2\}$ in $\mathbb{R}^3$, and a vector $\mathbf{v} \in \mathbb{R}^3$ such that $\{\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2\}$ is linearly *dependent*.

e) Give an example of a linearly independent subset $\{A, B\}$ in $\mathbf{M}_{22}(\mathbb{R})$, and a matrix $C \in \mathbf{M}_{22}(\mathbb{R})$ such that $\{A, B, C\}$ is linearly *dependent*.

f) *Give an example of a linearly independent subset $\{p, q\}$ in $\mathbb{P}_2$, and a polynomial $r \in \mathbb{P}_2$ such that $\{p, q, r\}$ is linearly *dependent*.

g) Give an example of a linearly independent subset $\{f, g\}$ in $\mathbf{F}(\mathbb{R})$, and a function $h \in \mathbf{F}(\mathbb{R})$ such that $\{f, g, h\}$ is linearly *dependent*.

Problem 7.5 Give an example of 3 vectors in $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ such that no two of the vectors are multiples of each other, but $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent. Is this possible with just two vectors?


Photo: Ralph Nevins. Montpellier, France

8. Linear independence and spanning sets

Last time, we defined linear independence and linear dependence. A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m\}$ is *linearly independent* (LI) if the *only* solution to the dependence equation

$$a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m = \mathbf{0}$$

is the trivial solution ($a_1 = 0, \cdots, a_m = 0$). The opposite of this is: the set of vectors is *linearly dependent* (LD) if there *is* a nontrivial solution to

$$a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m = \mathbf{0}$$

meaning, there is a solution to this dependence equation in which *not all coefficients are zero* (although *some* might be). In this case, such an equation (with some nonzero coefficients) is called a *dependence relation* on that set.

8.1 Important Results about Linear Independence and Linear Dependence

We deduced many useful things about LI and LD sets:

1. A set $\{\mathbf{v}\}$ consisting of just one vector is LI if and only if $\mathbf{v} \neq \mathbf{0}$.
2. If a set S is LD, then any set containing S is also LD.
3. If a set S is LI, then any subset of S is also LI.
4. $\{\mathbf{0}\}$ is LD.
5. Any set containing the zero vector is LD.
6. A set with two vectors is LD if and only if one of the vectors is a multiple of the other.
7. A set with three or more vectors *could be LD even if* no two vectors are multiples of one another.

We concluded last time with the following very important statement:

Theorem 8.1.1 — Relation between linear dependence and spanning. A set $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ is LD if and only if there is at least one vector $\mathbf{v}_k$ which is in the span of the rest.

Caution: This theorem *doesn't mean* that *every* vector is a linear combination of the others. For example, $\{(1,1),(2,2),(1,3)\}$ is LD *but* $(1,3) \notin \text{span}\{(1,1),(2,2)\}$.

Proof. We have to show both directions of the "if and only if".

First, suppose $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ is LD. So there is some nontrivial dependence equation

$$a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m = \mathbf{0}$$

with not all $a_i = 0$. Let's say that $a_1 \neq 0$ (we could always renumber the vectors so that this is the case). Then we can solve for $\mathbf{v}_1$:

$$\mathbf{v}_1 = \frac{-a_2}{a_1}\mathbf{v}_2 + \cdots + \frac{-a_m}{a_1}\mathbf{v}_m$$

which just says $\mathbf{v}_1 \in \text{span}\{\mathbf{v}_2, \cdots, \mathbf{v}_m\}$. Done.

Now the other direction. Suppose $\mathbf{v}_n \in \text{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_{n-1}\}$. That means $\mathbf{v}_n = b_1\mathbf{v}_1 + \cdots + b_{n-1}\mathbf{v}_{n-1}$. So we have

$$b_1\mathbf{v}_1 + \cdots + b_{n-1}\mathbf{v}_{n-1} + (-1)\mathbf{v}_n = \mathbf{0}$$

and the coefficient of $\mathbf{v}_n$ is -1, which is nonzero. Hence this is a nontrivial dependence relation (it doesn't even matter if all the b_i were zero). Thus the set is LD. ■

8.2 Consequence: Any linearly dependent spanning set can be reduced

This is how we're going to solve our problem!

■ **Example 8.2.1** Let $W = \text{span}\{(1,0,0),(0,1,0),(1,1,0)\}$. Since

$$(1,1,0) = (1,0,0) + (0,1,0) \in \text{span}\{(1,0,0),(0,1,0)\},$$

and obviously $(1,0,0)$ and $(0,1,0)$ are also in $\text{span}\{(1,0,0),(0,1,0)\}$, it follows (from Theorem 6.4.1) that the span of these three vectors is contained in $\text{span}\{(1,0,0),(0,1,0)\}$. So we deduce that

$$W = \text{span}\{(1,0,0),(0,1,0)\}.$$

(Notice the skipped step: we have

$$W \subset \text{span}\{(1,0,0),(0,1,0)\}$$

by our argument but we had

$$\text{span}\{(1,0,0),(0,1,0)\} \subset W$$

as well (since the spanning set of the left side is a subset of the spanning set on the right side), and that's why we have equality.) ■

This example illustrates the following general theorem:

Theorem 8.2.2 — Reducing spanning sets. Suppose $W = \text{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$. If $\mathbf{v}_1 \in \text{span}\{\mathbf{v}_2, \cdots, \mathbf{v}_m\}$ then

$$W = \text{span}\{\mathbf{v}_2, \cdots, \mathbf{v}_m\}.$$

Proof. It is clear that $\text{span}\{\mathbf{v}_2, \cdots, \mathbf{v}_m\} \subseteq W$.

Now suppose $\mathbf{v}_1 = b_2\mathbf{v}_2 + \cdots + b_m\mathbf{v}_m$. If $w = a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m \in W$, then

$$w = (a_1b_2 + a_2)\mathbf{v}_2 + \cdots + (a_1b_m + a_m)\mathbf{v}_m \in \text{span}\{\mathbf{v}_2, \cdots, \mathbf{v}_m\}.$$

Hence, $W \subseteq \text{span}\{\mathbf{v}_2, \cdots, \mathbf{v}_m\}$, and so equality holds. ■

Using the notation of the theorem above, note that $\mathbf{v}_1 \in \text{span}\{\mathbf{v}_2, \cdots, \mathbf{v}_m\}$ implies (by Theorem 8.1.1) that $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m\}$ is LD.

! In other words: We can DECREASE the size of any LINEARLY DEPENDENT spanning set.

Further, if $\{\mathbf{v}_2, \cdots, \mathbf{v}_m\}$ is dependent (say, $\mathbf{v}_2 \in \text{span}\{\mathbf{v}_3, \cdots, \mathbf{v}_m\}$), then

$$W = \text{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_m\} = \text{span}\{\mathbf{v}_2, \cdots, \mathbf{v}_m\} = \text{span}\{\mathbf{v}_3, \cdots, \mathbf{v}_m\}.$$

And you can keep doing this until what you're left with is a *linearly independent* spanning set for W.

8.3 Another consequence: finding bigger linearly independent sets

We certainly have an inkling, now, that if you want a linearly independent set, you can't take too many vectors. (But how many can you take? Good question; we'll come back to this.) For now we have:

Theorem 8.3.1 — Enlarging linearly independent sets. Suppose $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ is a LI subset of a subspace W. For any $\mathbf{v} \in W$, we have

$$\{\mathbf{v}, \mathbf{v}_1, \cdots, \mathbf{v}_m\} \text{ is LI} \quad \Leftrightarrow \quad \mathbf{v} \notin \text{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}.$$

Proof. This is almost like Theorem 8.1.1 (contrast and compare!). So it has a similar proof.

Suppose the new larger set is LI. Then by Theorem 8.2.2, NO element of $\{\mathbf{v}, \mathbf{v}_1, \cdots, \mathbf{v}_m\}$ can be a linear combination of the rest. In particular, $\mathbf{v}$ can't be a linear combination of the rest. That's the same as saying that $\mathbf{v} \notin \text{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$.

Conversely, suppose $\mathbf{v} \notin \text{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$. Let's show that $\{\mathbf{v}, \mathbf{v}_1, \cdots, \mathbf{v}_m\}$ is LI. Namely, if we consider the dependence equation:

$$a_0\mathbf{v} + a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m = \mathbf{0}.$$

We want to decide if this could have any nontrivial solutions.

Well, if $a_0 \neq 0$, then we know (previous proof) that we could solve for $\mathbf{v}$ in terms of the other vectors, but our hypothesis is that this is not the case. So for sure we know that $a_0 = 0$.

But then what we have left is

$$a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m = \mathbf{0}$$

which is a dependence equation for the set $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$. This set is LI, so all these a_i must be zero.

Hence the only solution is the trivial one, which shows that $\{\mathbf{v}, \mathbf{v}_1, \cdots, \mathbf{v}_m\}$ is LI. ■

In other words: we can INCREASE the size of any LINEARLY INDEPENDENT set so long as it does NOT already span our vector space.

8.4 Examples

Let's see how to apply Theorem 8.3.1.

Example 8.4.1 The set $\{x^2, 1+2x\} \subset \mathbb{P}_3$ is LI, and $x^3 \notin \text{span}\{x^2, 1+2x\}$ – see the hints given for problems 6.4 parts (e) and (f), or example 7.7.7 on P. 81. Thus, $\{x^3, x^2, 1+2x\}$ is LI. ■

Example 8.4.2 The set $\left\{ \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\}$ is LI (since we showed that there is a bigger set containing this one that is LI, last time). Moreover,

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \notin \text{span} \left\{ \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\}$$

since you can't solve $\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix}$. Thus

$$\left\{ \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}$$

is LI. ■

It can be hard work to find an element $\mathbf{v}$ that is not in the span of your LI set! Next chapter we'll start with some more thoughts on this.

Problems

Problem 8.1 Justify your answers to the following: (The setting is a general vector space V.)

a) Suppose $\mathbf{u} \in \text{span}\{\mathbf{v}, \mathbf{w}\}$. Show carefully that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent.

b) *Suppose $\mathbf{u} \in \text{span}\{\mathbf{v}, \mathbf{w}\}$. Show carefully that $\text{span}\{\mathbf{v}, \mathbf{w}\} = \text{span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$.

c) Suppose $\text{span}\{\mathbf{v}, \mathbf{w}\} = \text{span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$. Show carefully that $\mathbf{u} \in \text{span}\{\mathbf{v}, \mathbf{w}\}$.

d) *Suppose $\text{span}\{\mathbf{v}, \mathbf{w}\} = \text{span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$. Show carefully that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent.

e) Suppose $\text{span}\{\mathbf{u}, \mathbf{v}\} = \text{span}\{\mathbf{w}, \mathbf{x}\}$. Show carefully that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ and $\{\mathbf{u}, \mathbf{v}, \mathbf{x}\}$ are both linearly dependent.

f) *Suppose $\{\mathbf{v}, \mathbf{w}\}$ is linearly independent, and that $\mathbf{u} \notin \text{span}\{\mathbf{v}, \mathbf{w}\}$. Show carefully that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly independent.

g) Suppose $\{\mathbf{v}, \mathbf{w}\}$ is linearly independent, and that $\mathbf{u} \notin \text{span}\{\mathbf{v}, \mathbf{w}\}$. Show carefully that $\text{span}\{\mathbf{v}, \mathbf{w}\} \neq \text{span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$.

h) *Suppose $\text{span}\{\mathbf{v}, \mathbf{w}\} \neq \text{span}\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$. Show carefully that $\mathbf{u} \notin \text{span}\{\mathbf{v}, \mathbf{w}\}$.

Problem 8.2 Justify your answers to the following:

a) Suppose that $\mathbf{u}, \mathbf{v}, \mathbf{w}$ are non-zero vectors in $\mathbb{R}^4$ such that $\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot \mathbf{w} = \mathbf{v} \cdot \mathbf{w} = 0$. Prove that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly independent.

b) *Suppose two polynomials $\{p, q\}$ satisfy $p \neq 0$ and $\deg(p) < \deg(q)$. Show carefully that $\{p, q\}$ is linearly independent.

c) Suppose a set of polynomials $\{p_1, \ldots, p_k\}$ satisfies $0 \neq p_1$ and $\deg(p_1) < \deg(p_2) < \cdots < \deg(p_k)$. Show carefully that $\{p_1, \ldots, p_k\}$ is linearly independent.

d) * Suppose $f, g \in \mathbf{F}(\mathbb{R})$ are differentiable functions and that $fg' - f'g$ is not the zero function. Prove carefully that $\{f, g\}$ is linearly independent.[a]

e) Use the previous part of this question to give another proof (i.e., other the one seen in class) that $\{\sin x, \cos x\}$ is linearly independent.

f) ** Suppose $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ are three vectors in $\mathbb{R}^3$ such that $\mathbf{u} \cdot \mathbf{v} \times \mathbf{w} \neq 0$. Prove carefully that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly independent.[b]

[a]Hint: Proceed by contradiction.

[b]*A geometric argument involving 'volume' is not sufficient.* Hint: Instead, first recall that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly independent iff none of the vectors is a linear combination of the others. Now proceed by contradiction, and reduce the number of cases to check from 3 to 1 by recalling from high school that for any three vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in \mathbb{R}^3$ we know that $\mathbf{v}_1 \cdot \mathbf{v}_2 \times \mathbf{v}_3 = \mathbf{v}_3 \cdot \mathbf{v}_1 \times \mathbf{v}_2 = \mathbf{v}_2 \cdot \mathbf{v}_3 \times \mathbf{v}_1$.

9. Basis and Dimension

Last time, we showed that:

- Any spanning set which is linearly dependent can be reduced (without changing its span), by removing a vector which is in the span of the rest.
- Any linearly independent set in W which doesn't span W can be made into a larger linearly independent set in W, by throwing in a vector which is not in the span of the set.

■ **Example 9.0.1** The set $\{(1,2,1,1),(1,3,5,6)\}$ is LI (since there are two vectors and neither is a multiple of the other). Let's find a bigger LI set containing these two vectors.

To find something not in their span, write it out:

$$\text{span}\left\{\begin{bmatrix}1\\2\\1\\1\end{bmatrix},\begin{bmatrix}1\\3\\5\\6\end{bmatrix}\right\}=\left\{a\begin{bmatrix}1\\2\\1\\1\end{bmatrix}+b\begin{bmatrix}1\\3\\5\\6\end{bmatrix}\mid a,b\in\mathbb{R}\right\}=\left\{\begin{bmatrix}a+b\\2a+3b\\a+5b\\a+6b\end{bmatrix}\mid a,b\in\mathbb{R}\right\}$$

So we want to choose a vector (x_1,x_2,x_3,x_4) which can't be expressed in this way. Well, we don't need to try hard, actually: what about $(1,0,0,0)$? Then we'd have

$$a+b=1, 2a+3b=0, a+5b=0, a+6b=0$$

the last two equations give $b=0$, so the first equation gives $a=1$, but then the second equation fails.

Conclusion: $\{(1,2,1,1),(1,3,5,6),(1,0,0,0)\}$ is LI. ■

Remark 9.0.2. *Generally speaking, if your set doesn't span the whole space, then if you pick a vector at random, chances are it won't be in the span of your set. (Think of a line or plane in $\mathbb{R}^3$: there are FAR MORE points NOT on the line or plane than are actually on the line or plane.)*

Problem 9.0.3 The set $\{(1,0),(0,1)\}$ is LI, as we saw before in example 7.7.1. Can we find $\mathbf{v} \in \mathbb{R}^2$ so that $\{(1,0),(0,1),\mathbf{v}\}$ is LI?
Solution NO! We know that $\text{span}\{(1,0),(0,1)\} = \mathbb{R}^2$ since every $(x,y) \in \mathbb{R}^2$ can be expressed as $x(1,0)+y(0,1)$. So there are no vectors $\mathbf{v}$ that satisfy the required condition of not being in the span of our set.

In fact, we know that any two non-zero non-collinear (in other words, LI) vectors in $\mathbb{R}^2$ span all of $\mathbb{R}^2$, so we can *never* make such a set bigger. We can rephrase this as:

FACT Any set of 3 or more vectors in $\mathbb{R}^2$ is LD.

Proof. Suppose S is a subset of $\mathbb{R}^2$ with 3 or more vectors. Let $\mathbf{v}_1, \mathbf{v}_2 \in S$.
If these are LD, we're done (by our facts).
Otherwise, we know they span $\mathbb{R}^2$, as we saw in the proof of Theorem 6.6.1 – and noted afterwards.
So if $\mathbf{v}_3 \in S$ is a third vector, it satisfies $\mathbf{v}_3 \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$.
But that means $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a LD subset of S.
So S is LD. ■

! In other words: any LI set in $\mathbb{R}^2$ has at most two vectors; and any spanning set of $\mathbb{R}^2$ has at least two vectors. So EVERY linearly independent spanning set has exactly two vectors. WOW!

9.1 The BIG theorem relating LI sets to spanning sets

So what we know so far is: if S is a spanning set, then any bigger set (containing S) has to be LD. And if instead S is a linearly independent set, then no proper subset could span the whole space. In fact, a MUCH stronger result is true:

Theorem 9.1.1 — LI sets are never bigger than spanning sets. If a vector space V can be spanned by n vectors, then any linearly independent subset has *at most* n vectors.

Equivalently: if V has a subset of m linearly independent vectors, then any spanning set has *at least* m vectors.

! In other words: the size of any linearly independent set in V $\leq$ the size of any spanning set of V.

The proof of this theorem is interesting; we'll prove a special case later on, when V is a subspace of $\mathbb{R}^n$, and this can be used to see it in general. See me, or take MAT2141!

Let's apply this theorem in some examples.

■ **Example 9.1.2** We know that $\mathbb{R}^3$ is spanned by 3 vectors (eg: $(1,0,0),(0,1,0),(0,0,1)$). So *any set in $\mathbb{R}^3$ with 4 or more vectors is LD!* ■

■ **Example 9.1.3** We saw that $M_{2,2}(\mathbb{R})$ is spanned by 4 vectors. So any set of 5 or more 2×2 matrices is LD! ■

■ **Example 9.1.4** The set of diagonal 2×2 matrices is spanned by two vectors, so any set of 3 or more diagonal 2×2 matrices is LD. ■

■ **Example 9.1.5** let W be a plane through the origin in $\mathbb{R}^3$. Then it is spanned by 2 vectors, so any 3 or more vectors in W are LD. (But of course you can always find a set of 3 vectors in $\mathbb{R}^3$ that are LI — you just can't find a set of 3 vectors that all lie in the subspace W that are LI.) ■

Well, this last example wasn't really news (it amounts to saying that any three coplanar vectors are linearly dependent, which is where we started before), *except* to say:

! The theorem applies to *any vector space*, including SUBSPACES; it talks about the maximum number of linearly independent vectors IN THE SUBSPACE.

9.2 The critical balance: a basis of a vector space

Definition 9.2.1 A set $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m\}$ of vectors in V is called a *basis* of V if:

1. $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m\}$ in linearly independent, AND
2. $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m\}$ spans V.

Different ways to think about a basis:

- It's a linearly independent spanning set of V.
- It's a biggest possible linearly independent set in V.
- It's a smallest possible spanning set of V.

■ **Example 9.2.2** $\{(1,0),(0,1)\}$ is a basis of $\mathbb{R}^2$. ■

■ **Example 9.2.3** $\{1,x,x^2\}$ is a basis of $\mathbb{P}_2$. ■

■ **Example 9.2.4** $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ is a basis of $M_{2,2}(\mathbb{R})$. ■

■ **Example 9.2.5** $\{(1,0)\}$ is not a basis for $\mathbb{R}^2$, because it does not span $\mathbb{R}^2$. (It is a basis for U, the line which is the x-axis in $\mathbb{R}^2$.) ■

■ **Example 9.2.6** $\{(1,0),(0,1),(1,1)\}$ is not a basis for $\mathbb{R}^2$, because it is LD. ■

Theorem 9.2.7 — All bases have the same size. If $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ and $\{\mathbf{w}_1, \cdots, \mathbf{w}_k\}$ are two bases for a vector space V, then $m = k$.

Proof. Since $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ spans V, and $\{\mathbf{w}_1, \cdots, \mathbf{w}_k\}$ is LI, we know that $m \geq k$ (big theorem).

Since $\{\mathbf{w}_1, \cdots, \mathbf{w}_k\}$ spans V and $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ is LI, we also know that $k \geq m$ (big theorem).

So $m = k$. ■

! In other words: all bases of V have the SAME number of vectors.

9.3 Dimension of a vector space

Definition 9.3.1 If V has a finite basis $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$, then the *dimension* of V is n, the number of vectors in this basis. We write

$$\dim(V) = n$$

and also can say that V is *finite-dimensional*. If V doesn't have a finite basis, then V is *infinite dimensional*.

Remark 9.3.2. *We will mainly focus on finite dimensional vector spaces in this course. To do interesting things with infinite-dimensional spaces, take the next step after Calculus:* analysis *(MAT2125).*

■ **Example 9.3.3** $\dim(\mathbb{R}^2) = 2$, because we found a basis with 2 elements. ■

■ **Example 9.3.4** $\dim(\mathbb{P}_2) = 3$, because we found a basis with 3 elements. ■

■ **Example 9.3.5** $\dim(M_{2,2}(\mathbb{R})) = 4$, because we found a basis with 4 elements. ■

■ **Example 9.3.6** The set $\{(1,0,\cdots,0),(0,1,\cdots,0),\cdots,(0,0,\cdots,1)\}$ of vectors in $\mathbb{R}^n$ is linearly independent and spans $\mathbb{R}^n$ (check!) and so $\dim(\mathbb{R}^n) = n$. ■

■ **Example 9.3.7** The set $\{1,x,x^2,\cdots,x^n\}$ is linearly independent – generalize the proof (found in Example 7.7.7) that $\{1,x,x^2\}$ is linearly independent. It also spans $\mathbb{P}_n$ (check!) and so $\dim(\mathbb{P}_n) = n+1$. ■

■ **Example 9.3.8** Consider the set of $m \times n$ matrices $\{E_{ij} \mid 1 \leq i \leq m, 1 \leq j \leq n\}$ where E_{ij} is the matrix with zeros everywhere except for a 1 in the (i,j)th position. These are linearly independent since

$$\sum_{i,j} a_{ij} E_{ij}$$

is the matrix with (i,j)th entry a_{ij}. Hence if this sum is zero, each entry is zero, and so the dependence equation has only the trivial solution. They span $M_{m,n}(\mathbb{R})$ since an arbitrary matrix can be written in the above form, with a_{ij} equal to its (i,j)th coefficient for each i,j.

So this is a basis; and counting its elements we deduce that $\dim(M_{m,n}(\mathbb{R})) = mn$. ■

■ **Example 9.3.9** The vector space $\mathbb{P}$ is infinite dimensional. Why? We saw in Example 9.3.7 that for any n, the set $\{1,x,x^2,\cdots,x^n\}$ is linearly independent. Since a basis must be larger than any linearly independent set, this shows that you couldn't possibly find a finite basis for all of $\mathbb{P}$. ■

■ **Example 9.3.10** The vector space $F(\mathbb{R})$ is also infinite dimensional, by the same argument. ■

We can also consider bases and dimensions of subspaces.

■ **Example 9.3.11** Consider $L = \{A \in M_{2,2}(\mathbb{R}) \mid \operatorname{tr}(A) = 0\}$. We saw this is equal to

$$L = \left\{ \begin{bmatrix} a & b \\ c & -a \end{bmatrix} \mid a,b,c \in \mathbb{R} \right\} = \operatorname{span}\left\{ \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}$$

so it is a subspace and we have a spanning set $\{M_1, M_2, M_3\}$. Is this spanning set linearly independent? We check

$$aM_1 + bM_2 + cM_3 = 0 \Leftrightarrow \begin{bmatrix} a & b \\ c & -a \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

which means $a = b = c = 0$. So $\{M_1, M_2, M_3\}$ is linearly independent and also spans L. This means it's a basis of L. So $\dim(L) = 3$. ■

Problem 9.3.12 Find a basis for $W = \text{span}\{1, \sin(x), \cos(x)\}$, a subspace of $F(\mathbb{R})$.
Solution We note right away that $\{1, \sin(x), \cos(x)\}$ is a spanning set for W. Is it linearly independent? YES, as we verified a while ago. Therefore it is a basis for W and $\dim(W) = 3$.

Problem 9.3.13 Find a basis for $U = \{(x,y,z) \mid x+z=0\}$
Solution First we find a spanning set.

$$U = \{(x,y,-x) \mid x,y \in \mathbb{R}\} = \text{span}\{(1,0,-1),(0,1,0)\}$$

So $\{(1,0,-1),(0,1,0)\}$ spans U and it is linearly independent (since it consists of two vectors which are not multiples of one another) so it is a basis. Thus $\dim(U) = 2$.

Problems

1. A question with an asterisk '*' (or two) indicates a bonus-level question.
2. You must justify all your responses.

Problem 9.1 Give two distinct bases for each of the following subspaces, and hence give the dimension of each subspace.

a) $\{(2x,x) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$

b) $^\star\{(x,y) \in \mathbb{R}^2 \mid 3x - y = 0\}$

c) $\{(x,y,z) \in \mathbb{R}^3 \mid x+y-2z=0\}$

d) $^\star\{(x,y,z,w) \in \mathbb{R}^4 \mid x-y+z-w=0\}$

e) $\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \,\middle|\, b = c \right\}$.

f) $^\star\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \,\middle|\, a+d=0 \right\}$.

g) $\left\{ \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \,\middle|\, a,b \in \mathbb{R} \right\}$.

h) $^\star\left\{ \begin{bmatrix} 0 & -b \\ -b & 0 \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \,\middle|\, b \in \mathbb{R} \right\}$.

i) $\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \,\middle|\, a+b+c+d=0 \right\}$.

j) $^\star\mathbb{P}_n$.

k) $\{p \in \mathbb{P}_2 \mid p(2) = 0\}$.

l) $^\star\{p \in \mathbb{P}_3 \mid p(2) = p(3) = 0\}$.

m) $\{p \in \mathbb{P}_2 \mid p(1) + p(-1) = 0\}$.

n) *span$\{\sin x, \cos x\}$.

o) span$\{1, \sin x, \cos x\}$.

p) *span$\{1, \sin^2 x, \cos^2 x\}$.

q) * $\{(x, x-3) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$, equipped with the *non-standard operations:–* Addition:

$$(x,y)\tilde{+}(x',y') = (x+x', y+y+3).$$

Multiplication of vectors by scalars $k \in \mathbb{R}$:

$$k \odot (x,y) = (kx, ky+3k-3).$$

r) * $\{(x,y,z) \in \mathbb{R}^3 \mid x+2y+z=2\}$; $V = \mathbb{R}^3$, *Non-standard operations:–* Addition:

$$(x,y,z)\tilde{+}(x',y',z') = (x+x', y+y, z+z'-2).$$

Multiplication of vectors by scalars $k \in \mathbb{R}$:

$$k \odot (x,y,z) = (kx, ky, kz-2k+2).$$

Hint for parts (k)&(l): Recall the Factor/Remainder theorem from high school: if p is a polynomial in the variable x of degree at least 1, and $p(a) = 0$ for some $a \in \mathbb{R}$, then p has a factor of $x-a$, i.e., $p(x) = (x-a)q(x)$, where q is a polynomial with $\deg(q) = \deg(p) - 1$.)

Problem 9.2 Determine whether the following sets are bases of the indicated vector spaces.

a) $\{(1,2)\}$; $(\mathbb{R}^2)$

b) $^\star\{(1,2), (-2,-4)\}$; $(\mathbb{R}^2)$.

c) $\{(1,2), (3,4)\}$; $(\mathbb{R}^2)$.

d) $^\star\{(1,2), (3,4), (0,0)\}$; $(\mathbb{R}^2)$.

e) $\{(1,2,3), (4,8,6)\}$; $(\mathbb{R}^3)$.

f) $^\star\{(1,2,3), (4,8,7)\}$; $(\mathbb{R}^3)$.

g) $\{(1,2,3), (4,8,7), (3,6,4)\}$; $(\mathbb{R}^3)$.

h) $^\star\{(1,0,1,0), (0,1,0,1)\}$; $(\mathbb{R}^4)$.

i) $\left\{ \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \right\}$; $(\mathbf{M}_{22})$

j) $^\star\left\{ \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -2 \\ 1 & 0 \end{bmatrix} \right\}$; $(\mathbf{M}_{22})$

k) $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$; ($\mathbf{M}_{22}$)

l) $\{1, 1-x, 1-2x\}$; ($\mathbb{P}_2$).

m) $^\star\{1, 1+x, x^2\}$; ($\mathbb{P}_2$).

n) $\{\sin x, \cos x\}$; ($\mathbf{F}(\mathbb{R})$).

o) $^\star\{1, \sin x, 2\cos x\}$; ($\mathbf{F}(\mathbb{R})$).

p) $\{2, 2\sin^2 x, 3\cos^2 x\}$; ($\mathbf{F}(\mathbb{R})$).

q) ** $\{(1,0),(0,0)\}$; $V = \mathbb{R}^2$, _Non-standard operations:–_ Addition:

$$(x,y)\tilde{+}(x',y') = (x+x', y+y-2).$$

Multiplication of vectors by scalars $k \in \mathbb{R}$:

$$k \circledast (x,y) = (kx, ky-2k+2).$$

r) ** $\{(1,3),(2,4)\}$; $V = \mathbb{R}^2$, _Non-standard operations:–_ Addition:

$$(x,y)\tilde{+}(x',y') = (x+x', y+y-2).$$

Multiplication of vectors by scalars $k \in \mathbb{R}$:

$$k \circledast (x,y) = (kx, ky-2k+2).$$

Problem 9.3 Let $f(x) = 1+x$, $g(x) = x+x^2$ and $h(x) = x+x^2+x^3$ be polynomials in $\mathbb{P}$ and define $W = \text{span}\{f,g,h\}$.

a) Show that f, g and h are linearly independent.
b) Find a basis for W and the dimension of W.
c) If $j(x) = 1-x^2+x^3$ show that $j \in W$.
d) What is $\dim \text{span}\{f,g,h,j\}$?

Problem 9.4 Let $U = \{(x,y,z,w) \in \mathbb{R}^4 \mid x-y+z-w = 0\}$.

a) Find a basis for U and hence determine $\dim U$.
b) Extend your basis in (a) to a basis of $\mathbb{R}^4$.

Problem 9.5 Consider the vector space $\mathbf{F}[0,2] = \{f \mid f : [0,2] \to \mathbb{R}\}$. (Give it the same operations as $\mathbf{F}(\mathbb{R})$.) Suppose $f(x) = x, \quad g(x) = \frac{1}{x+1}$ and let

$$W = \text{span}\{f, g\}.$$

a) Show that $\{f, g\}$ is linearly independent.
b) Find $\dim W$.
c) If $h(x) = x^2$, show that $h \notin W$.
d) What is the dimension of $\text{span}\{f, g, h\}$?

Problem 9.6 Let $\mathbf{E} = \{\text{“}ax + by + cz = d\text{”} \mid a, b, c, d \in \mathbb{R}\}$ be the set of linear equations with real coefficients in the variables x, y and z. Equip $\mathbf{E}$ with the usual operations on equations that you learned in high school: addition of equations, denoted here by “$\oplus$" and multiplication by scalars, denoted here by “$\circledast$", as follows:

$$\text{“}ax + by + cz = d\text{”} \oplus \text{“}ex + fy + gz = h\text{”} = \text{“}(a+e)x + (b+f)y + (c+g)z = d + h\text{”}$$

and

$$\forall k \in \mathbb{R}, \quad k \circledast \text{“}ax + by + cz = d\text{”} = \text{“}kax + kby + kcz = kd\text{”}.$$

In an earlier exercise, you showed that with these operations, E is a vector space. Find a basis for E and hence find $\dim E$.


Photo: Ralph Nevins. Montpellier, France

10. Dimension Theorems

In the previous chapter, we saw the important inequality:

> ! the size of any LI set in V $\leq$ the size of any spanning set of V

We used this result to deduce that any *linearly independent spanning set* of a vectors space V (which we called a *basis of* V) has the same number of elements. We called the number n of elements in a basis for V the *dimension of* V, denoted $\dim(V)$.

We also found bases for many of our favourite vector spaces, and started to see how to find bases of other, more complicated, subspaces.

From the definition of dimension, we can now improve our inequality:

> ! the size of any LI set in V $\leq \dim V \leq$ the size of any spanning set of V

10.1 Every subspace of a finite-dimensional space has a finite basis

In the past chapters, we have shown that:

- If $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ is a LI subset of V, then there are vectors $\mathbf{v}_{m+1}, \cdots, \mathbf{v}_n$ in V such that $\{\mathbf{v}_1, \cdots, \mathbf{v}_m, \mathbf{v}_{m+1}, \cdots, \mathbf{v}_n\}$ is a basis for V *(that is, every linearly independent subset of V can be extended to a basis of V);*
- Conversely, we saw that if $S = \{\mathbf{v}_1, \cdots, \mathbf{v}_l\}$ is a spanning set of V, then there is a subset of S which is a basis of V *(that is, every spanning set can be reduced to a basis of V).*

Remark 10.1.1. *If you were in an infinite-dimensional space, the process of adding more vectors to your LI set might never stop, since it could happen that you can't span your subspace with a finite number of vectors. Hence our restriction. See also MAT2141.*

This means that in theory at least, you can always find a basis for any subspace: either start with a spanning set and cut it down, or else start taking nonzero vectors from your set and forming larger and larger LI sets.

A problem: this is computationally intensive. (We'll come back to it soon!)

10.2 A shortcut to checking if a set is a basis

Theorem 10.2.1 — Shortcut to deciding if a set is a basis. Let V be a vector space and suppose that we know that $\dim(V) = n$ (and $n < \infty$). Then:

1. Any LI set $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ of n vectors in V is a basis of V *(that is, it also necessarily spans V!);*
2. Any spanning set $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ of V consisting of exactly n vectors is a basis of V *(that is, it is also necessarily LI!);*

So this theorem says that if you already know the dimension of V, you have a shortcut to finding a basis: just find an LI *or* a spanning set with the *right number of elements*.

Proof.

1. Suppose that $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\} \subseteq V$ is LI and suppose instead that $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ didn't span V. Then that would mean we could find $\mathbf{v} \in V$, such that $\{\mathbf{v}, \mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is LI (*cf* results last week about extending LI sets). But then we'd have a set with $n+1$ LI elements in V, even though $\dim(V) = n$ means V can be spanned by just n elements. This contradicts our important inequality, so can't be true. Thus our set must already span V, and hence be a basis.
2. Suppose now that $\text{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_n\} = V$ but that $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is NOT LI. Since it is LD, we can remove at least one "redundant" vector without changing the span. That is, there is a subset $S \subset \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ with less than n elements such that $\text{span}(S) = \text{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_n\} = V$. But then we'd have a spanning set for V with fewer than n elements, contradicting our important inequality. So that can't be true; the spanning set must have been LI, and hence a basis.

■

Problem 10.2.2 Show that $\{(2,2,2), (7,1,-11)\}$ is a basis for $U = \{(x,y,z) | 2x - 3y + z = 0\}$.

Solution We see this is a plane through the origin, hence a 2-dimensional space. We verify that both of the vectors given lie in U, since

$$2(2) - 3(2) + 2 = 0 \qquad \textit{and} \qquad 2(7) - 3(1) + (-11) = 0.$$

Moreover, since there are just two vectors, and they are not multiples of one another, we see they are linearly independent. So:

- we have 2 vectors
- they lie in a 2-dimensional space
- they are linearly independent

So by the theorem, $\{(2,2,2), (7,1,-11)\}$ is in fact a basis for U.

Problem 10.2.3 Extend $\{(2,2,2), (7,1,-11)\}$ to a basis for $\mathbb{R}^3$.

Solution We know that $\dim(\mathbb{R}^3) = 3$; so if we can find one more vector $\mathbf{v}$ so that $\{\mathbf{v}, (2,2,2), (7,1,-11)\}$ is LI, we can deduce by the theorem that this new set is a basis for $\mathbb{R}^3$.

Now recall that since $\{(2,2,2), (7,1,-11)\}$ is LI, the larger set $\{\mathbf{v}, (2,2,2), (7,1,-11)\}$ is LI if and only if $\mathbf{v} \notin \operatorname{span}\{(2,2,2),(7,1,-11)\}$. That is, if and only if $\mathbf{v} \notin U$. So pick any $\mathbf{v} = (x,y,z)$ which does not satisfy the condition for being in U, such as $\mathbf{v} = (1,0,0)$.

Then $\{(1,0,0),(2,2,2),(7,1,-11)\}$ is a basis for $\mathbb{R}^3$.

! Caution: we get this "shortcut" for checking that a set is a basis for V *only if* you know the dimension of V.

So what kinds of clues help us to figure out the dimension of a space?

10.3 Dimension of subspaces of V

Theorem 10.3.1 — Dimensions of subspaces. Suppose that $\dim(V) = n$ and that W is a subspace of V. Then

1. $0 \leq \dim(W) \leq \dim(V)$
2. $\dim(W) = \dim(V)$ if and only if $W = V$.
3. $\dim(W) = 0$ if and only if $W = \{\mathbf{0}\}$.

Proof.

1. Start with a basis of W; it has $\dim(W)$ elements (which is ≥ 0!). Then this is a LI set in $W \subset V$, so by our important inequality, $\dim(W) \leq \dim(V)$ since V is spanned by $\dim(V)$ elements.
2. Suppose $\dim(W) = \dim(V) = n$. If $W \neq V$, then that means there is a vector $\mathbf{v} \in V$, such that $\mathbf{v} \notin W$. But then if $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is a basis for W, we would deduce that $\{\mathbf{v}, \mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is a LI set in V with $n+1$ elements, which contradicts the important inequality for V. So this is impossible; we must have $V \subset W$. (And we started with $W \subset V$ so we conclude $W = V$.)
3. If $\dim(W) = 0$, and W contains a non-zero vector then by the important inequality we have a contradiction, since then $\dim(W) \geq 1$. Thus $\dim W = 0 \implies W = \{\mathbf{0}\}$. If on the other hand $W = \{\mathbf{0}\}$, then we shall take it as a convention that $\dim(W) = 0$.[1]

■

This theorem has some great consequences!

■ **Example 10.3.2** Any subspace of $\mathbb{R}^3$ has dimension 0, 1, 2 or 3, by Theorem part (1). These correspond to: the zero space (Theorem part (3)), lines (we proved this long ago), planes (ditto), and all of $\mathbb{R}^3$ (Theorem part (2)). ■

■ **Example 10.3.3** Any 4-dimensional subspace of $M_{2\times 2}(\mathbb{R})$ is all of $M_{2\times 2}(\mathbb{R})$. ■

■ **Example 10.3.4** Any 2-dimensional subspace of $U = \{(x,y,z) | 2x - 3y + z = 0\}$ is all of U. ■

This last example illustrates an important idea:

! If U is a subspace of V such that $\dim(U) = m$, then any subspace of V which is contained in U is a subspace of U and so has dimension at most m.

[1] There are two other ways to define the dimension of a vector space: (a) take it to be the size of the largest linearly independent set, or (b) the size of the smallest spanning set. These are equivalent to our definition for $\dim(W) \geq 1$, and if we use (a), for $W = \{\vec{0}\}$, it's clear that the largest subset of W that is linearly independent is the empty set, which of course has size 0. One can reconcile this with (b) if we all agree that that the span of the empty set is $\{\vec{0}\}$. If this sounds too weird to you, stick with (a).

That is, we don't have to apply the theorem to just our favourite "big" vector spaces V.

10.4 What subspaces and dimension have given us

At the beginning of the course, we talked about lines and planes in $\mathbb{R}^2$ and $\mathbb{R}^3$ and wondered what their generalizations to $\mathbb{R}^n$ should be.

We proposed the general concept of vector spaces, of which subspaces of $\mathbb{R}^n$ are an example, and deduced that lines and planes through the origin were subspaces of $\mathbb{R}^2$ and $\mathbb{R}^3$, and (besides the zero space and the whole space) these were *all* subspaces of $\mathbb{R}^2$ and $\mathbb{R}^3$.

So we agreed that "subspaces" provide the correct higher-dimensional analogue of "lines and planes" (at least, the ones that go through the origin).

Now we have learned that every subspace has a basis, and that if S is a basis for U, then $U = \text{span}(S)$. In other words, we can describe every subspace of every vector space using parametric equations which are like "higher-dimensional analogues" of the parametric equation for a line.

Furthermore, we have learned how to measure the size of a subspace (namely, by its dimension), and thereby can classify all the possible subspaces of any vector space by their dimension. (For example, in $\mathbb{R}^n$ we have subspaces of every dimension up to and including n, in $\mathbb{P}_n$ we have subspaces of every dimension up to and including $n+1$, and in $M_{m\times n}(\mathbb{R})$ we have subspaces of every dimension up to and including mn.)

So the first step in taking geometry to higher dimensions is behind us.

10.5 What bases of subspaces give us, Part I

So what is a basis good for? Last chapter, we alluded to one very important application: namely, you know that any plane through the origin in $\mathbb{R}^3$ is "basically the same" as $\mathbb{R}^2$ — it looks the same, and geometrically it is the same kind of object, but algebraically, it's quite a challenge!

Problem 10.5.1 Consider the plane $U = \{(x,y,z) | x - 4y + z = 0\}$ in $\mathbb{R}^3$. Show that

$$\{(2/3, 1/3, 2/3), (1/\sqrt{2}, 0, -1/\sqrt{2})\}$$

is a basis for U, and that moreover these basis vectors are orthogonal and have norm one. (We will call such a special basis an *orthonormal basis* when we get to Chapter 19.)

Solution As before, we see that it suffices to verify that these two vectors (a) lie in U and (b) are LI, since $\dim(U) = 2$. So they form a basis. Next, we calculate their dot product (and get zero) so they are orthogonal. Finally, we calculate their norms, and deduce that each has norm equal to 1.

(Recall that $\|(x,y,z)\| = \sqrt{x^2+y^2+z^2}$.)

What makes such a basis special is that these are *exactly* the properties that make the standard basis special! So this is "like" a standard basis of U.

So as an application: Suppose you have an image in $\mathbb{R}^2$ (say, a photograph whose lower left corner is at the origin, and whose upper right corner is at the point $(640, 480)$, so that each integer coordinate pair corresponds to a pixel.). Then you can map this image onto the plane U by thinking of

$$(a,b) \in \mathbb{R}^2 \Leftrightarrow a\left[2/3, 1/3, 2/3\right] + b\left[1/\sqrt{2}, 0, -1/\sqrt{2}\right]$$

So for example if you want to colour the picture, then you have a colour for each integer coordinate pair (a,b) and this tells you which (x,y,z) you should colour in $\mathbb{R}^3$ to make it look like your picture is on this plane. ■

This opens up more questions, though: how did I get such a nice basis for U? Can I do that in different vector spaces and higher dimensions? We'll see part of the answer in this course, and you'll do more of it later on.

10.6 What bases of subspaces are good for, Part II

An even more startling and wonderful application of bases is as follows.

First we need a definition.

Definition 10.6.1 An *ordered basis* $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is the set $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ with the given order of the vectors.

■ **Example 10.6.2** The ordered basis $\{(1,0),(0,1)\}$ of $\mathbb{R}^2$ is not the same ordered basis as $\{(0,1),(1,0)\}$ — even though, *as sets*, $\{(1,0),(0,1)\} = \{(0,1),(1,0)\}$, since both sides have exactly the same vectors in them. The apparent order matters when we talk about ordered bases. ■

Theorem 10.6.3 — Coordinates. Suppose $B = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is an *ordered* basis for a vector space V. Then for every $\mathbf{v} \in V$, there are *unique* scalars $x_1, \cdots, x_n \in \mathbb{R}$ so that

$$\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n.$$

We call the n-tuple $(x_1, x_2, \ldots, x_n)$ the *coordinates of* $\mathbf{v}$ *relative to the ordered basis* B.

Proof. This isn't hard to prove: Since B spans V, you can certainly write any vector as a linear combination of elements of B. For uniqueness, we suppose we have two such expressions and prove they are the same, as follows:

Suppose $\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$ and $\mathbf{v} = y_1\mathbf{v}_1 + \cdots + y_n\mathbf{v}_n$. Subtract these two expressions; the answer is $\mathbf{v} - \mathbf{v} = \mathbf{0}$. So we have

$$(x_1 - y_1)\mathbf{v}_1 + \cdots + (x_n - y_n)\mathbf{v}_n = \mathbf{0}$$

but B is LI, so each of these coefficients must be zero, which says $x_i = y_i$ for all i. Uniqueness follows. ■

We can use this to *identify* n-dimensional vector spaces with $\mathbb{R}^n$, just like we identified that plane U with $\mathbb{R}^2$, above.

The idea is:

$$\begin{array}{ccc} V & \longleftrightarrow & \mathbb{R}^n \\ \mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n & \longleftrightarrow & (x_1, x_2, \ldots, x_n) \end{array}$$

For example, choosing the ordered basis $B = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ (the so-called 'standard ordered' basis) of $M_{22}(\mathbb{R})$, we have:

$$\begin{array}{ccc} M_{22}(\mathbb{R}) & \longleftrightarrow & \mathbb{R}^4 \\ \begin{bmatrix} a & b \\ c & d \end{bmatrix} & \longleftrightarrow & (a, b, c, d) \end{array}$$

Or, choosing $B = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}$ as an ordered basis for $\mathfrak{sl}_2$, we have

$$\begin{array}{ccc} \mathfrak{sl}_2 & \longleftrightarrow & \mathbb{R}^3 \\ \begin{bmatrix} a & b \\ c & -a \end{bmatrix} & \longleftrightarrow & (a,b,c) \end{array}$$

Another example: choosing the standard ordered basis $\{1,x,x^2\}$ for $\mathbb{P}_2$ we have

$$\begin{array}{ccc} \mathbb{P}_2 & \longleftrightarrow & \mathbb{R}^3 \\ a+bx+cx^2 & \longleftrightarrow & (a,b,c) \end{array}$$

! Caution: note that the order of your basis matters! If we'd picked $B = \{x^2, x, 1\}$ as our ordered basis for $\mathbb{P}_2$, we'd get coordinates (c,b,a) instead of (a,b,c), which has the potential for a great deal of confusion. Hence: always be a bit cautious to keep the order in mind.

Problems

Problem 10.1 Find the coordinates of the following vectors v with respect to the given ordered bases $\mathscr{B}$ of the indicated vector space V:

a) $v = (1,0)$; $\mathscr{B} = \{(1,2),(2,-1)\}$; $V = \mathbb{R}^2$

b) $^\star v = (1,0,1)$; $\mathscr{B} = \{(1,1,0),(1,-1,0),(0,0,1)\}$; $V = \mathbb{R}^3$

c) $v = (1,3,-3)$; $\mathscr{B} = \{(1,0,0),(0,-1,1),\}$; $V = \{\mathbf{w} \in \mathbb{R}^3 \mid \mathbf{w} \cdot (0,1,1) = 0\}$

d) $^\star v = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}$; $\mathscr{B} = \left\{ \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right\}$; $V = \{A \in \mathbf{M}_{22} \mid \operatorname{tr} A = 0\}$

e) $v = 1 - x + 3x^2$; $\mathscr{B} = \{1-x, 1+x, 1+x^2\}$; $V = \mathbb{P}_2$

f) $^\star v = \sin(x+2)$; $\mathscr{B} = \{\sin x, \cos x\}$; $V = \operatorname{span}\{\sin x, \cos x\}$

Problem 10.2 a) * Suppose V is a vector space, with $\dim V = 3$, and that W and U are both two-dimensional subspaces of V. Prove carefully that $U \cap W \neq \{0\}$.

b) * Suppose V is a vector space, with $\dim V = n > 0$, and that W and U are subspaces of V with $\dim U + \dim W > n$. Prove carefully that $U \cap W \neq \{0\}$.

III Solving Linear Systems

We have seen how to solve a number of important problems, at least theoretically; but the calculations needed to get to the solution — solving linear systems — seems overwhelming. What we need is a practical, efficient, reliable means to solve these linear systems, one that will minimize the calculations we are required to carry out and the amount of writing we need to do; one that will allow us to deduce not just whether or not there is a solution, but also if there are lots of solutions (*e.g.* nontrivial solutions).

This is the goal and purpose of Gauss-Jordan elimination (commonly called: row reduction).

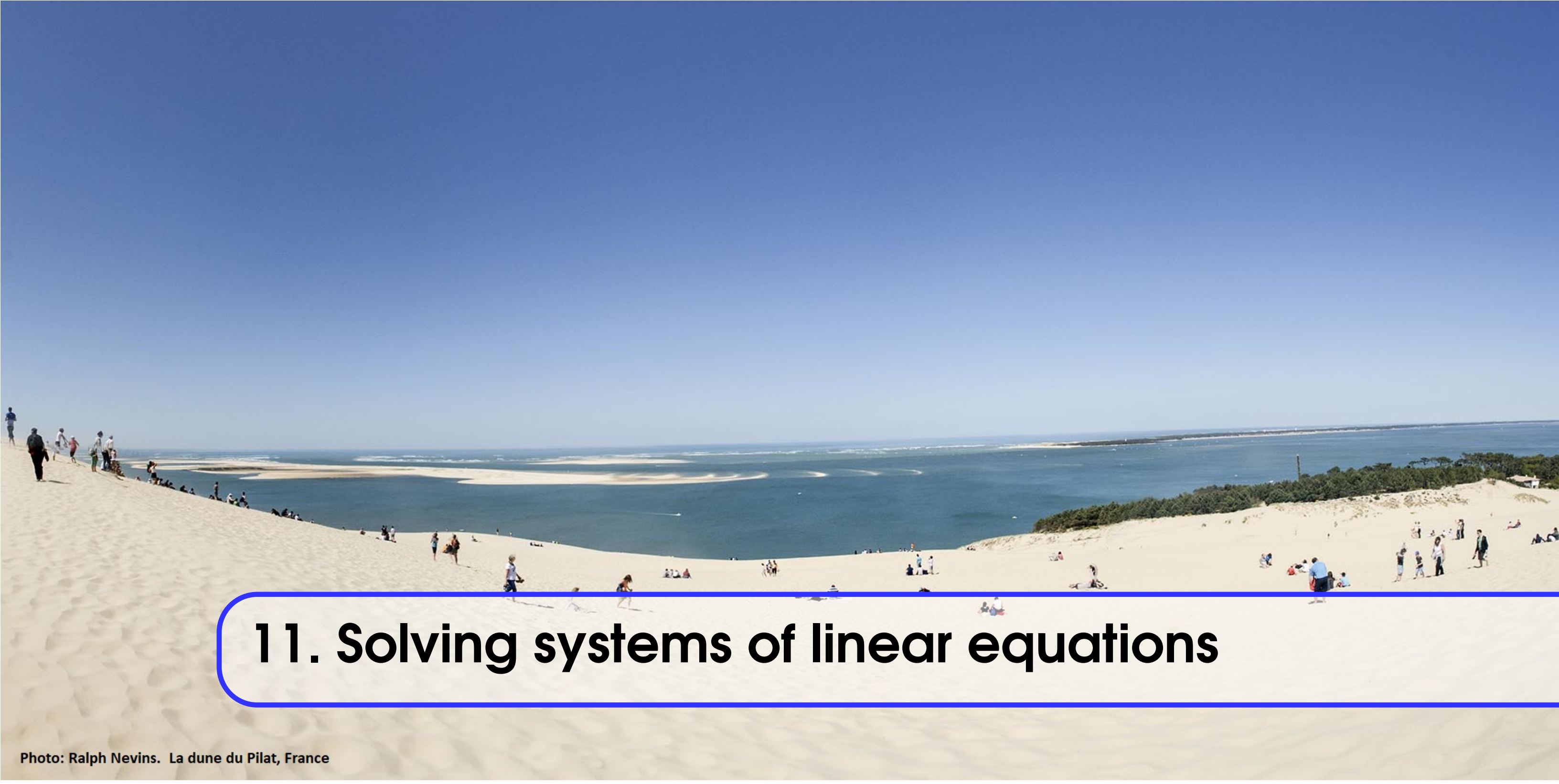


Photo: Ralph Nevins. La dune du Pilat, France

11. Solving systems of linear equations

We have seen many different kinds of linear systems showing up in different contexts in our course. They arise in problems like: finding intersections of planes in $\mathbb{R}^3$; solving for a vector as a linear combination of other vectors; or checking if a set of vectors is linearly independent. But they also occur in a million other contexts, from balancing chemical equations to modelling physical systems. And while the linear systems we've been looking at in this course tend to be fairly small, in practice there are physical models that include thousands of variables and thousands of equations — systems that would be IMPOSSIBLE to solve were it not for the fantastic fact that they are linear. Linear algebra gives incredibly powerful tools for solving linear systems.

Here, we want to (a) decide what the goal of a good method for solving linear systems should be, (b) establish the notation of augmented matrices which we use in row reduction and (c) start to learn the process of row reduction.

11.1 Linear systems

By a *linear system*, we mean a collection of, say, m linear equations in, say, n variables, that we want to solve simultaneously. By "simultaneously", we mean: a *solution* to a linear system is an assignment of values to each of the variables which makes ALL of the equations hold.

■ Example 11.1.1

$$\begin{array}{ccccccccc} x_1 & + & 2x_2 & & & + & x_4 & = & 1 \\ & & & & x_3 & - & x_4 & = & -1 \end{array} \tag{11.1}$$

is a linear system with $m = 2$ equations and $n = 4$ *unknowns* (or *variables*). We identify also the *coefficients* of the system, and the *right hand side (RHS)*, or *constant terms*.

Then $(1,0,0,0)$ is NOT a solution to this linear system, because it doesn't satisfy the second equation.

And $(0,0,0,1)$ IS a solution to this linear system, because it satisfies BOTH equations. ■

Definition 11.1.2 The *general solution* to a linear system is the set of *all* solutions.

■ **Example 11.1.3** We claim that

$$S = \{(1-2s-t, s, t-1, t) \mid s,t \in \mathbb{R}\}$$

is the general solution to the linear system (11.1).

To show this, we check two things:

I: Check that $(1-2s-t, s, t-1, t)$ is a solution to (11.1) for every $s,t \in \mathbb{R}$:

$$\begin{aligned} (1-2s-t)+2(s)+t &= 1 \qquad \checkmark \\ t-1-(t) &= -1 \qquad \checkmark \end{aligned}$$

II: Check that EVERY solution to (11.1) is in S: Well, use the first equation to write

$$x_1 = 1-2x_2-x_4$$

and the second equation to write

$$x_3 = -1+x_4.$$

Thus, (x_1,x_2,x_3,x_4) is a solution to (11.1) if and only if

$$(x_1,x_2,x_3,x_4) = (1-2x_2-x_4, x_2, -1+x_4, x_4).$$

Changing the names of x_2 to s and x_4 to t, we recognize that this is exactly the set S above.

We say that s,t are the *parameters* of the general solution. Note that since the solution set S has parameters, the system (11.1) has *infinitely many* solutions, since every different value of $s \in \mathbb{R}$ gives a different solution.[1] ■

Let's look at some other examples, to get a sense of (a) what the possible kinds of general solutions are, and (b) what makes a system "easy" to solve.

■ **Example 11.1.4**

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

This is a 3×3 system (meaning, 3 equations and 3 unknowns); it could correspond to the intersection of 3 planes. This system has a *unique solution*: $(-7,4,1)$ (as you can check by back-substitution). ■

■ **Example 11.1.5**

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

is a 2×2 system with *no solutions* at all. We could say the general solution is the empty set $S=\emptyset$. ■

[1] The same holds for t, of course. So there is a 'doubly-infinite' family of solutions. We have been more precise about this 'doubly-infinite' business in a similar context before when we spoke of 'dimension' in Chapter 9. But beware: the set of solutions here is *not* a subspace and so we can't—in this course—speak of its dimension.

Definition 11.1.6
- A linear system that has NO SOLUTIONS is called *inconsistent*.
- A linear system that has AT LEAST ONE solution is called *consistent*.

Sometimes, it's easy to see that a system is consistent. For example, consider

$$x_1 + x_2 - x_3 = 0$$
$$2x_1 - x_2 + x_3 = 0.$$

It's clear that $(0,0,0)$ is a solution, because all the constant terms are zero.

Definition 11.1.7
- A linear system in which all the constants on the right hand side are zeros is called *homogeneous*.
- A linear system in which at least one of the constants on the right hand side is nonzero is called *inhomogeneous*.

! Homogeneous linear systems are ALWAYS consistent, since $\mathbf{0} = (0,0,\cdots,0)$ is always a solution (called the *trivial solution*).

So homogeneous systems have all zeros on the RHS; what about when there are all zeros on the LHS?

Definition 11.1.8 A linear equation in which all the *coefficients* are zero is called *degenerate*, that is, it looks like

$$0x_1 + 0x_2 + \cdots + 0x_n = b$$

for some $b \in \mathbb{R}$. Note that if $b \neq 0$, then this equation has no solution! (If $b = 0$, this equation has every vector in $\mathbb{R}^n$ as a solution.)

! Any linear system containing a degenerate inhomogeneous equation is inconsistent.

Now the amazing thing is: the three examples above list all possible types of behaviours of the solution set.

Theorem 11.1.9 — Types of general solutions. Any linear system with real (or complex) coefficients has either

1. a unique solution
2. no solution, or
3. infinitely many solutions.

To see how remarkable that is, consider how very NOT TRUE it is for NON-LINEAR systems: $x^2 = 64$ has exactly two solutions; $(x-2)(x-1)(x+1)(x-2) = 0$ has exactly 4 solutions.

11.2 Solving Linear Systems

The idea of our method for solving linear systems is: we start with a linear system, whose solution is "hard to see"; then we apply an algorithm (called *row reduction* or *Gaussian elimination* or *Gauss-Jordan elimination*) to change the linear system into a new one which has *exactly the same general solution*, but where that solution set is really "easy to see."

Also, we want a practical solution: an algorithm where you're less likely to lose or forget an equation, or where you mis-write some variables, or make mistakes in recopying things.

So let's begin with an example of solving a system very methodically, being careful about not losing equations:

■ **Example 11.2.1** Solve the linear system

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

Let's add $(-1)\times$ Eq(1) to Eq(2), to eliminate the x; but we'll recopy the other two equations to keep track:

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

Next, we tackle the y variable. I can't use the first equation to solve for y, because that's the only equation with an x in it. So in fact, let me swap the second and third equations:

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

and now add $2\times$ Eq(2) to Eq(3):

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

Finally, multiply Eq(3) by $-\frac{1}{3}$:

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

Wonderful! At this point, I can see that $z=-1/3$, and I know that I can plug this into Eq(2) to deduce that $y=2/3$, and then plug both into Eq(1) to deduce that $x=3$. (Check that this is a solution!) ■

First reality check: what operations did we perform?

- Add a multiple of one row to *another* row.

- Interchange two rows.
- Multiply a row by a nonzero scalar.

Each of these steps can be reversed, and that ensures that we are not changing the general solution. (In contrast, if we ever multiplied a row/equation by zero, that's an irreversible process, and it amounts to deleting one of the equations, which we agree would change your general solution.)

! These are called *elementary row operations* and they are exactly the three operations you can do to a linear system without changing the general solution. (What you're doing instead is turning the system itself into a simpler system.)

Second reality check: Did we really need all those variables and $=$ signs?
Let's replace the linear system above with its *augmented matrix*:

$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & 3 \\ 1 & -1 & 1 & 2 \\ 0 & 1 & -1 & 1 \end{array}\right]$$

So the part on the left is called the *coefficient matrix* and consists of all the coefficients of the original linear system. If there were m equations and n unknowns then the coefficient matrix is of size $m \times n$. Each column of the coefficient matrix corresponds to one variable of the linear system (in this case, x, y or z).

The line represents the equals sign, and helps to separate the last column. The last column contains all the constant terms from all the equations in the system.

Now let's perform the elementary row operations as before, but this time to the augmented matrix:

$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & 3 \\ 1 & -1 & 1 & 2 \\ 0 & 1 & -1 & 1 \end{array}\right] \underset{-R_1+R_2 \to R_2}{\sim} \left[\begin{array}{ccc|c} 1 & 1 & 2 & 3 \\ 0 & -2 & -1 & -1 \\ 0 & 1 & -1 & 1 \end{array}\right] \underset{R_2 \leftrightarrow R_3}{\sim} \left[\begin{array}{ccc|c} 1 & 1 & 2 & 3 \\ 0 & 1 & -1 & 1 \\ 0 & -2 & -1 & -1 \end{array}\right]$$

$$\underset{2R_2+R_3 \to R_3}{\sim} \left[\begin{array}{ccc|c} 1 & 1 & 2 & 3 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & -3 & 1 \end{array}\right] \underset{-\frac{1}{3}R_3 \to R_3}{\sim} \left[\begin{array}{ccc|c} 1 & 1 & 2 & 3 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1/3 \end{array}\right]$$

Notice:

- At each step, the augmented matrix here corresponded exactly to the modified linear system we had created before.
- This last matrix corresponds to the last step of our calculation, at the moment that we realized we could solve our linear system for a unique solution. And this last matrix has a special "shape" which we will call the REF (row echelon form), below.
- We write $\sim$ rather than $=$ because the matrices are definitely not EQUAL to each other. Rather, we use the symbol $\sim$ for "equivalent" (later: " row equivalent") — meaning, the two linear systems corresponding to these matrices have the same general solution.
- It's helpful to write down the operation next to the row you will modify, particularly if you need to "debug" a calculation.

■

Now, we still have some questions to answer:

- Not every system will end up with a unique solution (as we saw); what is REF in general? How do I know what I'm aiming for when I solve using row reduction?
- How do I read my solution off from REF?
- Are there any more shortcuts?

11.3 Row Echelon Form (REF) and Reduced Row Echelon Form (RREF)

Definition 11.3.1 A matrix (augmented or not) is in *row echelon form* or *REF* if
(1) All zero rows are at the bottom.
(2) The first nonzero entry in each row is a 1 (called a *leading one* or a *pivot*).
(3) Each leading 1 is to the right of the leading 1s in the rows above.
If, in addition, the matrix satisfies
(4) Each leading 1 is the only nonzero entry in its column
then the matrix is said to be in *reduced row echelon form* or *RREF*.

■ **Example 11.3.2** The matrix

$$\left[\begin{array}{ccccc|c} 1 & 2 & 3 & 4 & 5 & 5 \\ 0 & 0 & 1 & 1 & 2 & 0 \end{array}\right]$$

is in REF because it satisfies (1)-(3). It is not in RREF because the entry above the leading 1 in the second row is a 3 instead of a zero. ■

■ **Example 11.3.3** The matrix

$$\left[\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & 1 & 8 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

is also in REF but not RREF. ■

Basically: from the REF we'll always be able to tell if the system is inconsistent, or has a unique solution, or has infinitely many solutions. From the RREF we will be able to just read off the solution directly.

■ **Example 11.3.4** Suppose you have row reduced and obtain the following RREF:

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & a \\ 0 & 1 & 0 & b \\ 0 & 0 & 1 & c \end{array}\right]$$

Then the linear system corresponding to this augmented matrix is $x = a$, $y = b$ and $z = c$. In other words, the solution to this linear system is unique and is given by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

■

■ **Example 11.3.5** Suppose you have row reduced and you got to the following REF:

$$\begin{bmatrix} 1 & a & 0 & b & | & d \\ 0 & 0 & 1 & e & | & f \\ 0 & 0 & 0 & 0 & | & g \end{bmatrix}$$

Then there are two cases:

If $g \neq 0$, then the last row corresponds to a degenerate equation, and so the system is inconsistent.

If $g = 0$, then this system is in RREF, and to get the general solution, we set the *non-leading variables*, that is, the variables corresponding to columns of the coefficient matrix which don't have a leading 1, equal to parameters. Here, set $x_2 = s$ and $x_4 = t$; then we deduce (by writing out the equations corresponding to this matrix):

$$x_1 = -as - bt + d, \quad x_2 = s, \quad x_3 = f - et, \quad x_4 = t$$

So our general solution is

$$\{(-as - bt + d, s, f - et, t) \mid s, t \in \mathbb{R}\}.$$

■

! Rows of zeros are completely accidental; they happen whenever you started off with one or more completely redundant equations. In particular, please note that having infinitely many solutions is related to having non-leading variables, not to having rows of zeros.

■ **Example 11.3.6** Suppose our RREF is

$$\begin{bmatrix} 1 & a & 0 & 0 & | & c \\ 0 & 0 & 1 & 0 & | & d \\ 0 & 0 & 0 & 1 & | & e \end{bmatrix}$$

Then we have only one nonleading variable: $x_2 = s$, so our solution is

$$\{(c - as, s, d, e) \mid s \in \mathbb{R}\}$$

■

Problems

Problem 11.1 Find the augmented matrix of the following linear systems.

a)

$$\begin{array}{rcrcrcl} x & + & y & + & z & = & 0 \\ -9x & - & 2y & + & 5z & = & 0 \\ -x & + & y & + & 3z & = & 0 \\ -7x & - & 2y & + & 3z & = & 0 \end{array}$$

b) ⋆

$$\begin{array}{rcrcrcrcr} x & & & & & + & w & = & 1 \\ x & & & + & z & + & w & = & 0 \\ x & + & y & + & z & & & = & -3 \\ x & + & y & & & - & 2w & = & 2 \end{array}$$

c)

$$\begin{array}{rcrcrcrcrl} & & & - & 2x_3 & & & + & 7x_5 & = 12 \\ 2x_1 & + & 4x_2 & - & 10x_3 & + & 6x_4 & + & 12x_5 & = 28 \\ 2x_1 & + & 4x_2 & - & 5x_3 & + & 6x_4 & - & 5x_5 & = -1 \end{array}$$

Problem 11.2 a) Find x and y so that the matrix $\begin{bmatrix} 1 & 2 \\ x & y \end{bmatrix}$ is in reduced row-echelon form.

b) ⋆Find all (x,y) so that the matrix $\begin{bmatrix} 1 & 0 & 1 \\ x & y & 0 \end{bmatrix}$ is in reduced row-echelon form.

c) Find all (a,b,c) so that the matrix $\begin{bmatrix} a & 1 & b & b & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & c \end{bmatrix}$ is in reduced row-echelon form.

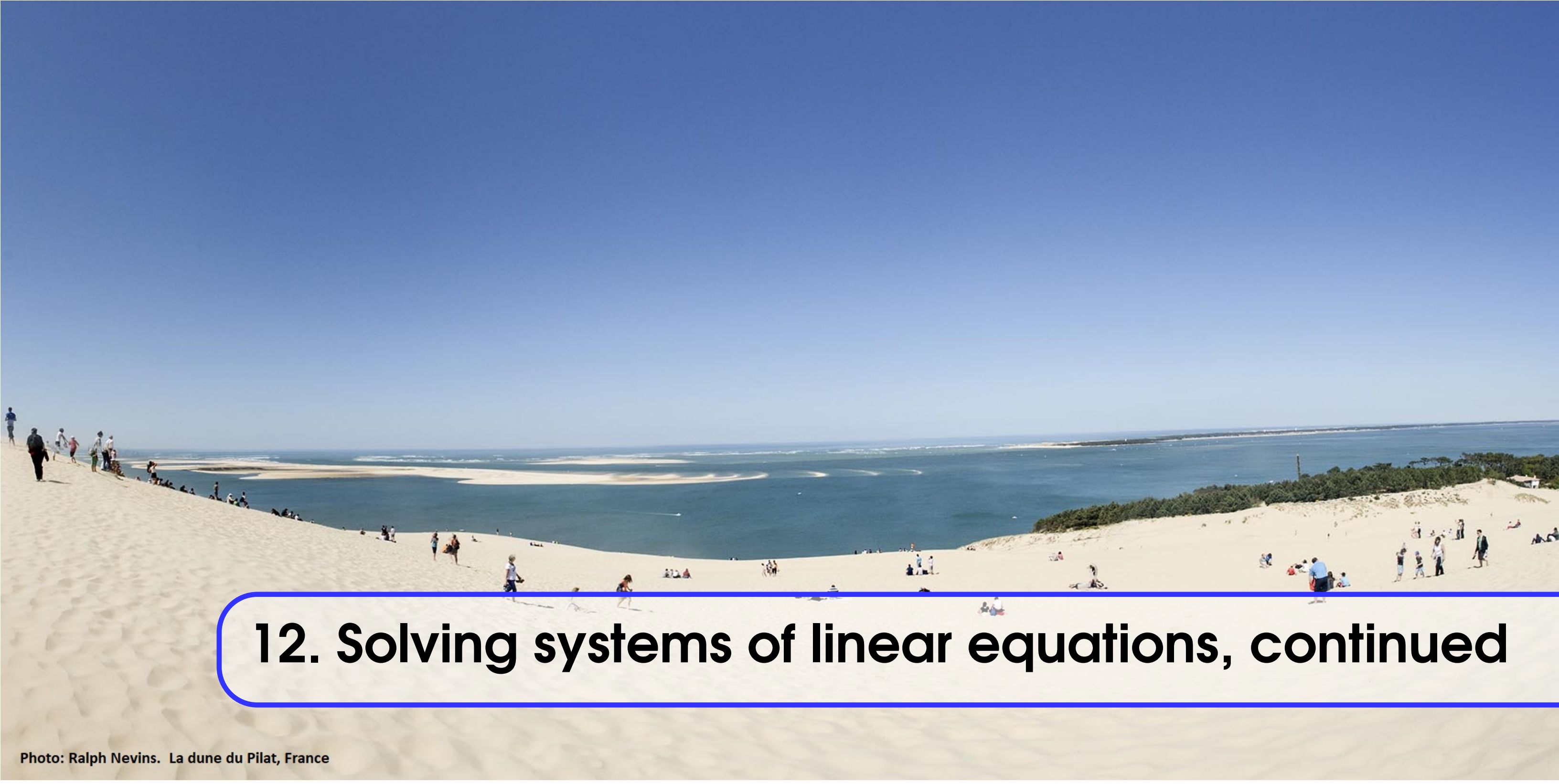


Photo: Ralph Nevins. La dune du Pilat, France

12. Solving systems of linear equations, continued

Last time, we introduced the notion of the augmented matrix of a linear system, and defined the three *elementary row operations* that we may perform:

- Add a multiple of one row *to* another row.
- Interchange two rows.
- Multiply a row by a nonzero scalar.

Definition 12.0.1 We say that two linear systems are *equivalent* if they have the same general solution.

Theorem 12.0.2 — Equivalence of linear systems under row reduction. If an elementary row operation is performed on the augmented matrix of a linear system, the resulting linear system is equivalent to the original one.

Hence we make the following definition:

Definition 12.0.3 Two matrices A is *row-equivalent* to B, written $A \sim B$, if B can be obtained from A by a finite sequence of elementary row operations.

(See problem 12.3 for interesting properties of this relation.)

So how does this help? Here, we'll show how to *row reduce* ANY linear system to RREF, and also show how to read the general solution from the RREF.

Recall: A matrix (augmented or not) is in *row echelon form* or REF if

(1) All zero rows are at the bottom.
(2) The first nonzero entry in each row is a 1 (called a *leading one* or a *pivot*).
(3) Each leading 1 is to the right of the leading 1s in the rows above.

If, in addition, the matrix satisfies
(4) Each leading 1 is the only nonzero entry in its column
then the matrix is said to be in *reduced row echelon form* or RREF.

Theorem 12.0.4 — Uniqueness of the RREF. Every matrix is row equivalent to a *unique* matrix in RREF.

(But matrices in just REF are not unique.)

12.1 Reading the solution from RREF

In the last chapter, we worked through several general examples of reading the solution from the RREF:

Example 12.1.1 From Example 11.3.4:

$$\text{RREF:} \quad \left[\begin{array}{ccc|c} 1 & 0 & 0 & a \\ 0 & 1 & 0 & b \\ 0 & 0 & 1 & c \end{array}\right] \qquad \text{Solution:} \quad \left\{ \begin{bmatrix} a \\ b \\ c \end{bmatrix} \right\}.$$

■

Example 12.1.2 From Example 11.3.5:

$$\text{RREF:} \quad \left[\begin{array}{cccc|c} 1 & a & 0 & b & d \\ 0 & 0 & 1 & e & f \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \qquad \text{Solution:} \quad \left\{ \begin{bmatrix} d \\ 0 \\ f \\ 0 \end{bmatrix} + s \begin{bmatrix} -a \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -b \\ 0 \\ -e \\ 1 \end{bmatrix} \middle| s,t \in \mathbb{R} \right\},$$

whereas

$$\text{REF:} \quad \left[\begin{array}{cccc|c} 1 & a & 0 & b & d \\ 0 & 0 & 1 & e & f \\ 0 & 0 & 0 & 0 & g \end{array}\right], g \neq 0 \qquad \text{Solution:} \quad \emptyset.$$

■

Example 12.1.3 From Example 11.3.6:

$$\text{RREF:} \quad \left[\begin{array}{cccc|c} 1 & a & 0 & 0 & c \\ 0 & 0 & 1 & 0 & d \\ 0 & 0 & 0 & 1 & e \end{array}\right] \qquad \text{Solution:} \quad \left\{ \begin{bmatrix} c \\ 0 \\ d \\ e \end{bmatrix} + s \begin{bmatrix} -a \\ 1 \\ 0 \\ 0 \end{bmatrix} \middle| s \in \mathbb{R} \right\}.$$

■

Now we ask ourselves: can we infer some general patterns, applicable to *any* system whose augmented matrix is in RREF?

The general rule for reading off the *type* of general solution from the REF:

- If your system contains a degenerate equation *with a non-zero right hand side*, then it is inconsistent. So if the augmented matrix contains a row like

$$\left[\begin{array}{cccc|c} 0 & 0 & \cdots & 0 & b \end{array}\right]$$

  with $b \neq 0$, then STOP! The system is inconsistent; the general solution is the empty set.

- Otherwise, look at the columns of the coefficient matrix.
  - If every column has a leading 1, then there is a unique solution.
  - If there is a column which does not have a leading 1, then you have infinitely many solutions.

The general rule for writing down the general solution of a consistent system from the RREF:

- If there is a unique solution, then this is the vector in the augmented column.
- Otherwise, identify all your variables as leading or non-leading.
  - Each leading varible corresponds to one row of the augmented matrix; write down the equation for this row. Solve for the leading variable in terms of the non-leading variables (by putting them all on the right hand side).
  - Set each non-leading variable equal to a different parameter, and substitute these into the equations for the leading variables as well.
  - Write down the general solution; do NOT forget to include ALL of your variables. (Eg: $x_1 = 2 - s$ and $x_3 = 3$ is not a general solution because you haven't said what x_2 is.)

■ **Example 12.1.4** Suppose the RREF of our system is

$$\begin{bmatrix} 1 & 0 & 0 & 2 & 0 & | & 3 \\ 0 & 0 & 1 & 1 & 0 & | & 4 \\ 0 & 0 & 0 & 0 & 1 & | & 0 \end{bmatrix}$$

This system is

$$x_1 + 2x_4 = 3, \quad x_3 + x_4 = 4, \quad x_5 = 0;$$

it is consistent. The leading variables are x_1, x_3 and x_5 and the non-leading variables are x_2 and x_4. So we set

$$x_2 = s, \qquad x_4 = t$$

and thus conclude that

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 3 - 2x_4 \\ x_2 \\ 4 - x_4 \\ x_4 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 - 2t \\ s \\ 4 - t \\ t \\ 0 \end{bmatrix}$$

So the general solution (written in parametric vector form) is

$$\left\{ \begin{bmatrix} 3 \\ 0 \\ 4 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} \mid s, t \in \mathbb{R} \right\}.$$

■

12.2 Reducing systems to REF and RREF: Gaussian elimination

Let's write down the process for Gaussian elimination. It can be applied to any matrix C, and stops at a matrix $\tilde{C}$ which is in RREF.

Step 1 If the matrix C is zero, stop.

Step 2 Locate the left-most nonzero column, and interchange the top row with another, if necessary, to bring a non-zero entry to the first row of this column.
Step 3 Scale the first row, as necessary, to get a leading 1.
Step 4 If necessary, annihilate the rest of the column BELOW using this leading 1 as a pivot. That is, if a_i is the entry in this column of Row i, then add $-a_iR_1$ to R_i, and put the result back in the i^{th} row.
Step 5 This completes the operations (for now) with the first row. If there was only one row in your matrix, at this stage, stop. Otherwise, *ignore the first row* (but don't lose it!) and go back to step 1.

When this stops, the matrix you have will be in REF. Now proceed with the following steps to put the matrix in RREF:
Step 6 If the *right most* leading 1 is in row 1, stop.
Step 7 Start with the *right most* leading 1 – this will be in the last non-zero row. Use it to annihilate every entry *above* it in its column. That is, if $a_i \neq 0$ is the entry in this column in row i, then add $-a_i$ times this row to R_i, and put the result back in the i^{th} row.
Step 8 Cover up the row you used and go to step 6.

■ **Example 12.2.1** Let's run this on $C = \begin{bmatrix} 0 & 0 & -2 & 2 \\ 1 & 1 & 3 & -1 \\ 1 & 1 & 2 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix}$.

Step 1 The matrix is not zero, so we proceed.
Step 2 The left-most nonzero column is column 1, but we need to get a non-zero entry in row 1, so let's interchange rows 1 and 2:

$$\begin{bmatrix} 0 & 0 & -2 & 2 \\ 1 & 1 & 3 & -1 \\ 1 & 1 & 2 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix} \begin{matrix} R_1 \leftrightarrow R_2 \\ \sim \end{matrix} \begin{bmatrix} \textcircled{1} & 1 & 3 & -1 \\ 0 & 0 & -2 & 2 \\ 1 & 1 & 2 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix}$$

Step 3 There's no need to rescale: we already have a leading 1 in row 1.
Step 4 We need to clear the column below our leading 1: we subtract the first row from rows 3 and 4:

$$\begin{bmatrix} 1 & 1 & 3 & -1 \\ 0 & 0 & -2 & 2 \\ 1 & 1 & 2 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix} \begin{matrix} -R_1 + R_3 \to R_3 \\ \sim \\ -R_1 + R_4 \to R_4 \end{matrix} \begin{bmatrix} 1 & 1 & 3 & -1 \\ 0 & 0 & -2 & 2 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & -3 & 3 \end{bmatrix}$$

Step 5 We ignore the first row and go to Step 1.
Step 1 Even ignoring the original first row, the matrix is not zero.
Step 2 The left-most non zero column now (remember: we ignore row 1) is column 3, and there is a non-zero entry (2) in the second row (which is the *first* row of the matrix left when we ignore the first row of the whole matrix), so there's nothing to do now.
Step 3 Let's divide row 2 by -2 to get a leading one in the second row.

$$\begin{bmatrix} 1 & 1 & 3 & -1 \\ 0 & 0 & -2 & 2 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & -3 & 3 \end{bmatrix} \begin{matrix} -\frac{1}{2}R_2 \to R_2 \\ \sim \end{matrix} \begin{bmatrix} 1 & 1 & 3 & -1 \\ 0 & 0 & \textcircled{1} & -1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & -3 & 3 \end{bmatrix}$$

Step 4 Now we need to clear column 3 below our new leading one:

$$\begin{bmatrix} 1 & 1 & 3 & -1 \\ 0 & 0 & \textcircled{1} & -1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & -3 & 3 \end{bmatrix} \begin{matrix} R_2+R_3 \to R_3 \\ 3R_2+R_4 \to R_4 \\ \sim \end{matrix} \begin{bmatrix} 1 & 1 & 3 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Step 5 Now we ignore the first two rows, and go to Step 1

Step 1 Well if we we ignore the first two rows, the matrix we see is zero, so we stop this part and proceed to Step 6

Step 6 The *right most* leading 1 is in row 2, so we proceed to Step 7

Step 7 We use the leading one in row two to clear its column above it:

$$\begin{bmatrix} 1 & 1 & 3 & -1 \\ 0 & 0 & \textcircled{1} & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} -3R_2+R_1 \to R_1 \\ \sim \end{matrix} \begin{bmatrix} 1 & 1 & 0 & 2 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Step 8 We cover up row 2 and go back to Step 6

Step 6 Ignoring row 2, now the right-most leading 1 *is* in row 1! So we stop. The matrix is now in RREF.

We're not machines, of course, and we could see at the end of Step 7 that the matrix was in RREF, so there was no need to have done Steps 8 and 6, but we did it above to illustrate the algorithm.

■

12.3 Using the Gaussian algorithm to solve a linear system

Now let's see how we use this to solve a linear system with augmented matrix $[A|b]$. The idea is that we apply the row operations in the algorithm above to the rows of the whole augmented matrix, but *with the aim of getting the coefficient matrix into RREF* - then we'll stop. The coefficient matrix will be in RREF – the augmented matrix might not be, but it doesn't matter.

Once the coefficient matrix is in RREF, the general solution can be found as follows:

1. Decide if the system is consistent or not. If consistent:
2. Assign parameters to non-leading variables.
3. Solve for leading variables in terms of parameters.

Let's illustrate this with an example. Remember: our goal is to get the coefficient matrix into RREF, but we apply each and every row operation to the rows of the whole augmented matrix. All decisions in the algorithm will only depend on the coefficient side.

■ **Example 12.3.1** We begin with the following augmented matrix:

$$\begin{bmatrix} 0 & 1 & 2 & 3 & | & 4 \\ 1 & 2 & 3 & 4 & | & 5 \\ 2 & 3 & 4 & 5 & | & 6 \end{bmatrix}$$

Step 1 It's not the zero matrix!

Step 2 It's the first column (but sometimes it isn't!). The top row has a zero, which can't be a pivot. So interchange R_1 and R_2, for example:

$$\begin{bmatrix} 0 & 1 & 2 & 3 & | & 4 \\ 1 & 2 & 3 & 4 & | & 5 \\ 2 & 3 & 4 & 5 & | & 6 \end{bmatrix} \begin{matrix} R_1 \leftrightarrow R_2 \\ \sim \\ \end{matrix} \begin{bmatrix} 1 & 2 & 3 & 4 & | & 5 \\ 0 & 1 & 2 & 3 & | & 4 \\ 2 & 3 & 4 & 5 & | & 6 \end{bmatrix}$$

Step 3 There's already a leading 1. Move on.
Step 4 We just need to zero off the 2 in R_3:

$$\begin{bmatrix} \textcircled{1} & 2 & 3 & 4 & | & 5 \\ 0 & 1 & 2 & 3 & | & 4 \\ 2 & 3 & 4 & 5 & | & 6 \end{bmatrix} \begin{matrix} \sim \\ \\ -2R_1 + R_3 \to R_3 \end{matrix} \begin{bmatrix} 1 & 2 & 3 & 4 & | & 5 \\ 0 & 1 & 2 & 3 & | & 4 \\ 0 & -1 & -2 & -3 & | & -4 \end{bmatrix}$$

Step 5 We're done with the first row. Back to Step 1, just considering R_2 and R_3:
Step 1 It's not the zero matrix. On we go.
Step 2 The first non-zero column (ignoring row 1) is in column 2. And the entry in the "top row" (which is R_2 this time) is nonzero, so no row interchanges needed.
Step 3 It's already a leading 1.
Step 4 We just need to zero off the -1 in R_3:

$$\begin{bmatrix} 1 & 2 & 3 & 4 & | & 5 \\ 0 & \textcircled{1} & 2 & 3 & | & 4 \\ 0 & -1 & -2 & -3 & | & -4 \end{bmatrix} \begin{matrix} \sim \\ \\ R_2 + R_3 \to R_3 \end{matrix} \begin{bmatrix} 1 & 2 & 3 & 4 & | & 5 \\ 0 & 1 & 2 & 3 & | & 4 \\ 0 & 0 & 0 & 0 & | & 0 \end{bmatrix}$$

Step 5 We're done using the second row. Back to Step 1, just considering R_3:
Step 1 It's the zero matrix (ignoring rows 1 and 2), so we proceed to Step 6.
And yes, indeed, the coefficient matrix is in REF (coincidentally, so is the whole augmented matrix). We see that we will have infinitely many solutions (since the system is consistent *and* we have non-leading variables).
To write down the solution, we continue to RREF:
Step 6 The *right most* leading 1 is not in row 1, so on we go.
Step 7 The right-most leading 1 is in column 2. Annihilate the 2 above it:

$$\begin{bmatrix} 1 & 2 & 3 & 4 & | & 5 \\ 0 & \textcircled{1} & 2 & 3 & | & 4 \\ 0 & 0 & 0 & 0 & | & 0 \end{bmatrix} \begin{matrix} -2R_2 + R_1 \to R_1 \\ \sim \\ \end{matrix} \begin{bmatrix} 1 & 0 & -1 & -2 & | & -3 \\ 0 & 1 & 2 & 3 & | & 4 \\ 0 & 0 & 0 & 0 & | & 0 \end{bmatrix}$$

Step 8 We cover up the second row, and go to step 6.
Step 6 Stop! – the *right most* leading 1 is in row 1.
The coefficient matrix is in RREF.[1] After some experience, you'll notice after step 7 that RREF had been reached.

The nonleading variables are x_3 and x_4, so let $x_3 = s$, $x_4 = t$. Then $x_1 = x_3 + 2x_4 - 3 = -3 + s + 2t$ and $x_2 = 4 - 2x_3 - 3x_4 = 4 - 2s - 3t$. Thus the general solution is

$$\begin{aligned} x_1 &= -3 + s + 2t \\ x_2 &= 4 - 2s - 3t \\ x_3 &= s \\ x_4 &= t \end{aligned}$$

[1]Coincidentally, so is the augmented matrix. This will always occur for consistent systems.

with $s,t \in \mathbb{R}$. ■

12.4 Key concept: the rank of a matrix

So we've stated the theorem that says that the RREF of a matrix exists and is unique. The existence is obvious from our algorithm; uniqueness takes some extra work to see, but also follows from our algorithm. Since the RREF is unique, this means in particular that the number of leading 1s in the RREF of a matrix doesn't depend on any choices made in the row reduction. This number is very important to us!

Definition 12.4.1 The *rank* of a matrix A, denoted $\operatorname{rank}(A)$, is the number of leading ones ('pivots') in any REF of A.

Remark: In the Gaussian Algorithm, the passage from the REF to the RREF does not change the number of leading ones.

■ **Example 12.4.2** The rank of $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 3 \end{bmatrix}$ is 2. ■

Problems

Problem 12.1 Find the reduced row-echelon form of the following matrices:

a) $\begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 0 & 1 & 5 \\ 0 & 1 & 0 & 4 \end{bmatrix}$

b) * $\begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 2 \end{bmatrix}$

c) $\begin{bmatrix} 0 & 2 & 3 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

d) * $\begin{bmatrix} 1 & 2 & -1 & -1 \\ 2 & 4 & -1 & 3 \\ -3 & -6 & 1 & -7 \end{bmatrix}$

e) $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & 2 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix}$

f) * $\begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 4 & 3 \\ 1 & 0 & 2 & 1 \end{bmatrix}$

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

Problem 12.2 Find the general solutions to the linear systems whose augmented matrices are given below.

a) $\begin{bmatrix} 0 & 1 & | & 0 \\ 0 & 0 & | & 0 \end{bmatrix}$

b) * $\begin{bmatrix} 1 & 0 & -1 & | & 0 \\ 0 & 1 & 2 & | & 0 \\ 0 & 0 & 0 & | & 1 \\ 0 & 0 & 0 & | & 0 \end{bmatrix}$

c) $\begin{bmatrix} 1 & 0 & 0 & 3 & | & 3 \\ 0 & 1 & 0 & 2 & | & -5 \\ 0 & 0 & 1 & 1 & | & -1 \\ 0 & 0 & 0 & 0 & | & 0 \end{bmatrix}$

d) * $\begin{bmatrix} 1 & 2 & 0 & 3 & 0 & | & 7 \\ 0 & 0 & 1 & 0 & 0 & | & 1 \\ 0 & 0 & 0 & 0 & 1 & | & 2 \end{bmatrix}$

e) $\begin{bmatrix} 1 & 0 & 0 & -1 & 0 & | & 10 \\ 0 & 1 & 0 & 1 & 1 & | & 1 \\ 0 & 0 & 1 & 1 & 1 & | & 4 \\ 0 & 0 & 0 & 0 & 0 & | & 0 \end{bmatrix}$

Problem 12.3 If A, and C are matrices of the same size, explain why the following statements are true:

a) $A \sim A$
b) If $A \sim B$ then $B \sim A$.
c) If $A \sim B$ and $B \sim C$, then $A \sim C$

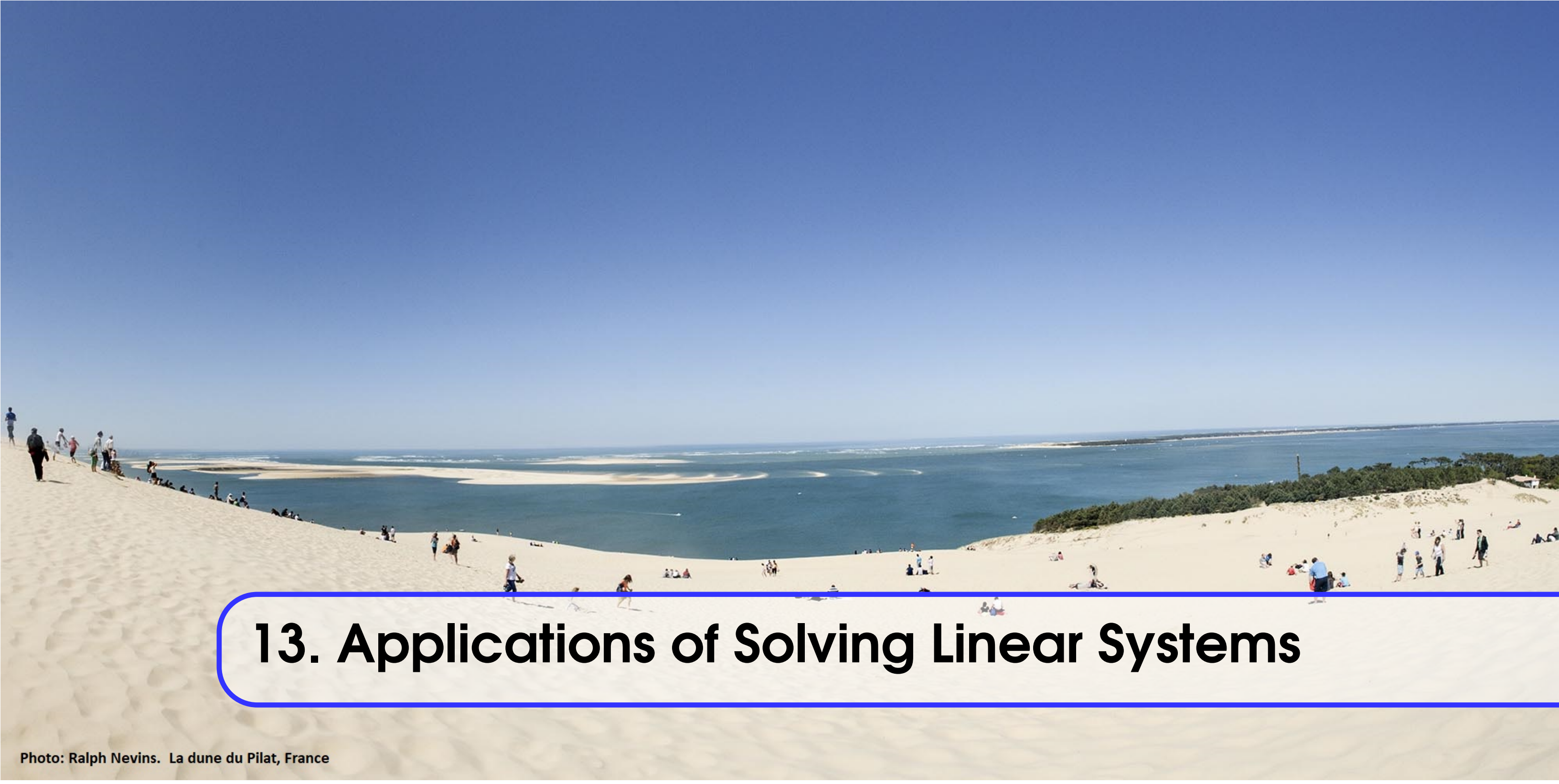


Photo: Ralph Nevins. La dune du Pilat, France

13. Applications of Solving Linear Systems

13.1 The rank of a matrix, and its importance

In the last chapter, we defined the rank of a matrix.

Definition 13.1.1 The *rank* of a matrix A, denoted $\text{rank}(A)$, is the number of leading ones ('pivots') in any REF of A.

■ **Example 13.1.2** Since $\begin{bmatrix} 1 & 4 & 1 \\ 2 & 3 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 4 & 1 \\ 0 & -5 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & 4 & 1 \\ 0 & 1 & \frac{2}{5} \end{bmatrix}$, *each* of these matrices has rank 2. ■

> ! The rank doesn't change when one does elementary row operations; if you're using the Gaussian algorithm, it just becomes easier to see.

■ **Example 13.1.3** The matrix $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ has rank 3. ■

Note that the rank of a matrix can never exceed the number of rows or columns of the matrix, since each leading 1 lies in a different row and column from all the others.

We also mentioned last time that you could see whether a system is consistent or inconsistent from the rank of A versus the rank of $[A|\mathbf{b}]$. Namely, suppose you reduce the coefficient matrix A in $[A|\mathbf{b}]$ to

RREF. Then either you get

$$\begin{bmatrix} * & \cdots & * & | & * \\ 0 & \cdots & 0 & | & 0 \\ \vdots & \cdots & \vdots & | & \vdots \\ 0 & \cdots & 0 & | & 0 \end{bmatrix}$$

(where the top $(*)$ part means rows with leading ones in the coefficient matrix part) (a consistent system) or you get

$$\begin{bmatrix} * & \cdots & * & | & * \\ 0 & \cdots & 0 & | & a_1 \\ \vdots & \cdots & \vdots & | & \vdots \\ 0 & \cdots & 0 & | & a_k \\ 0 & \cdots & 0 & | & 0 \\ \vdots & \cdots & \vdots & | & \vdots \\ 0 & \cdots & 0 & | & 0 \end{bmatrix}$$

where $a_i \neq 0$ (which tells you you have an inconsistent system). You can spot the difference by comparing the ranks of the coefficient matrix and of the augmented matrix: in the first case, $\text{rank}(A) = \text{rank}([A|\mathbf{b}])$ since all the leading 1s occured in the coefficient matrix part, and none in the augmented column. In the second case, $\text{rank}([A|\mathbf{b}]) = \text{rank}(A) + 1$, since the non-zero entry a_1 in the last column will lead to a (single) new leading 1.

! So:

$$\text{rank}(A) \leq \text{rank}([A|\mathbf{b}]) \leq \text{rank}(A) + 1$$

... and where the equality holds tells you whether or not your system is consistent.

■ **Example 13.1.4** In a homogeneous linear system, you can never get a nonzero entry in the augmented column (it's all zeros) so certainly you can never have a leading 1 there. Thus $\text{rank}(A) = \text{rank}([A|0])$, and the system is consistent. (We knew that already.) ■

In summary: suppose $[A|\mathbf{b}]$ is the augmented matrix of a linear system. Then:

- The system is inconsistent if and only if $\text{rank}(A) < \text{rank}([A|\mathbf{b}])$.
- The system has a unique solution if and only if $\text{rank}(A) = \text{rank}([A|\mathbf{b}])$ **and** $\text{rank}(A) =$# columns of A.
- The system has infinitely many solutions if and only if $\text{rank}(A) = \text{rank}([A|\mathbf{b}])$ **and** $\text{rank}(A) <$# columns of A.

13.2 Application I: Solving network and traffic flow problems

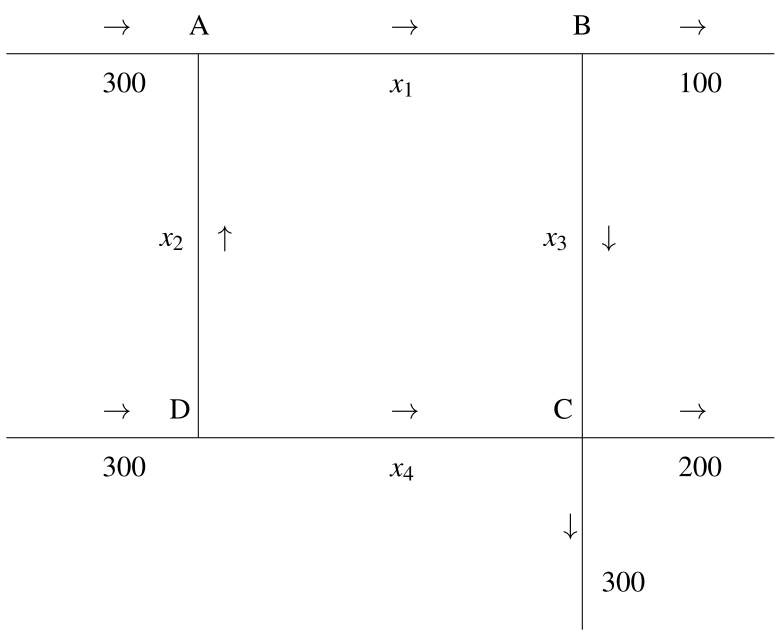
Let's consider several different applications of linear systems and our Gaussian elimination method. The first is a common application of linear systems to solving traffic flow/network problems. The idea is that you can model the internal flow of a network just by understanding its inputs and outputs and how traffic is restricted in between. You don't expect a unique solution or even finitely many (in theory, of course) because there may be loops – see problem 13.4, for example.

Problem 13.2.1 The diagram in the figure below represents a network of one-way streets. The numbers on the figure represent the flow of traffic (in cars per hour) along each street, and the intersections are labeled A, B, C and D. The arrows indicate the direction of the flow of traffic. The variables x_1, x_2, x_3, x_4 represent the (unknown) level of traffic on certain streets.

Analyse this system to answer the following questions:

(a) What is the maximum flow along AB?
(b) If DA is blocked for roadwork, will there be a traffic jam?
(c) What if BC is blocked for roadwork?

Solution



The total flow in is 600 (cars per hour, say) while the total flow out is 600. Good; let us take that as our hypothesis for all the intersections as well: flow in = flow out (Kirchhoff's first law).

We have labeled with x_1, x_2, x_3, x_4 the street segments on which we don't know the flow of traffic. So let us set up the flows by writing down the equation for each intersection (starting at the northwest and working clockwise):

Intersection	**Flow in**	$=$	**Flow out**
A	$300+x_2$	$=$	x_1
B	x_1	$=$	x_3+100
C	x_3+x_4	$=$	$200+300$
D	300	$=$	$x_2+x_4.$

This is the linear system

$$\begin{aligned} x_1 - x_2 &= 300 \\ x_1 - x_3 &= 100 \\ x_3 + x_4 &= 500 \\ x_2 + x_4 &= 300 \end{aligned}$$

with augmented matrix

$$\left[\begin{array}{cccc|c} 1 & -1 & 0 & 0 & 300 \\ 1 & 0 & -1 & 0 & 100 \\ 0 & 0 & 1 & 1 & 500 \\ 0 & 1 & 0 & 1 & 300 \end{array}\right]$$

This reduces to

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 600 \\ 0 & 1 & 0 & 1 & 300 \\ 0 & 0 & 1 & 1 & 500 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

So our solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 600 - s \\ 300 - s \\ 500 - s \\ s \end{bmatrix}$$

for any $s \in \mathbb{R}$.

Now that we have found a solution to our linear system, let us consider the questions we wanted to answer.

(a) Note that since the streets are one-way, each $x_i \geq 0$, for $1 \leq i \leq 4$. When you implement these constraints, you find that $0 \leq s \leq 300$. So for example, the maximum flow along *AB* is 300 cars per hour (when the flow along *DA* is zero).

(b) If *DA* is blocked for roadwork, then we are setting $x_2 = 0$. This means $s = 300$; we see that the solution is $(x_1, x_2, x_3, x_4) = (300, 0, 200, 300)$, which is fine. There are no parameters, meaning that cars now have no choices about how to navigate the network, but they make it.

(c) If *BC* is blocked for roadwork, however, then $x_3 = 0$. This means $s = 500$, which gives $(x_1, x_2, x_3, x_4) = (100, -200, 0, 500)$ — meaning we would have to allow two-way traffic on the street *DA* to accommodate the traffic.

In summary: we applied the general solution to the model to first establish some additional constraints on our parameters (here, positive flow) and then used this constrained general solution to easily work out various scenarios in traffic flow (without having to solve the system all over again!).

13.3 Application II: testing scenarios

Problem 13.3.1 Find all values of k for which the following linear system has (a) no solution, (b) a unique solution, and (c) infinitely many solutions.

$$\begin{aligned} kx+y+z&=1\\ x+ky+z&=1\\ x+y+kz&=1. \end{aligned}$$

Solution This is a linear system in the variables x,y,z. (It is *not linear in* k,x,y,z. So this is one way to tackle a special kind of nonlinear system with linear algebra.)

Let's use row reduction. The augmented matrix is

$$\left[\begin{array}{ccc|c} k & 1 & 1 & 1\\ 1 & k & 1 & 1\\ 1 & 1 & k & 1 \end{array}\right]$$

Now be careful! We can't divide by k, because we don't know if it's zero or not. So either we split into two cases right now ($k=0$ and $k\neq 0$) or else we use a different row to give us our leading 1.

$$R_1 \leftrightarrow R_2 \left[\begin{array}{ccc|c} 1 & k & 1 & 1\\ k & 1 & 1 & 1\\ 1 & 1 & k & 1 \end{array}\right] \begin{array}{c} \sim \\ -kR_1+R_2\to R_2 \\ -R_1+R_3\to R_3 \end{array} \left[\begin{array}{ccc|c} 1 & k & 1 & 1\\ 0 & 1-k^2 & 1-k & 1-k\\ 0 & 1-k & k-1 & 0 \end{array}\right]$$

Now we have no choice: there are variables in both entries from which we need to choose our leading 1.

Case 1: $k=1$. Then our matrix is

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 1\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 \end{array}\right]$$

which is in RREF and has infinitely many solutions. So (c) includes the case $k=1$.

Case 2: $k\neq 1$. That means $k-1\neq 0$ so we can divide by it:

$$\begin{array}{c} \frac{1}{1-k}R_2\to R_2 \\ \frac{1}{1-k}R_3\to R_3 \end{array} \left[\begin{array}{ccc|c} 1 & k & 1 & 1\\ 0 & 1+k & 1 & 1\\ 0 & 1 & -1 & 0 \end{array}\right] R_2 \leftrightarrow R_3 \left[\begin{array}{ccc|c} 1 & k & 1 & 1\\ 0 & 1 & -1 & 0\\ 0 & 1+k & 1 & 1 \end{array}\right]$$

And now continue row reducing:

$$\begin{array}{c} \sim \\ -(1+k)R_2+R_3\to R_3 \end{array} \left[\begin{array}{ccc|c} 1 & k & 1 & 1\\ 0 & 1 & -1 & 0\\ 0 & 0 & 2+k & 1 \end{array}\right]$$

For the third column: if $k=-2$, then we cannot get a leading 1 in the third column; in fact, if $k=-2$ then the system is inconsistent.

Otherwise, we can divide R_3 by $2+k$ to get a leading 1 in the third column; and then we deduce that we have a unique solution.

Conclusion: if $k=1$, infinitely many solutions. If $k=-2$, there is no solution. If $k\neq 1$ and $k\neq -2$ then there is a unique solution.

Notice how you get a unique solution "most of the time" — this coincides with the idea that if you take three random planes in $\mathbb{R}^3$ then most of the time they should intersect in a single point.

13.4 Application III: solving vector equations

Problem 13.4.1 Does the set $\{(1,2,3),(4,5,6),(7,8,9)\}$ span $\mathbb{R}^3$?

Solution We need to answer the question: given an arbitrary vector $(x,y,z) \in \mathbb{R}^3$, do there exist scalars a_1, a_2, a_3 such that

$$a_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + a_2 \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} + a_3 \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} ?$$

When you equate components on each side, we obtain

$$\begin{aligned} a_1 + 4a_2 + 7a_3 &= x \\ 2a_1 + 5a_2 + 8a_3 &= y \\ 3a_1 + 6a_2 + 9a_3 &= z \end{aligned}$$

which corresponds to the augmented matrix

$$\begin{bmatrix} 1 & 4 & 7 & | & x \\ 2 & 5 & 8 & | & y \\ 3 & 6 & 9 & | & z \end{bmatrix}$$

Notice how the columns of this matrix correspond to the vectors in our linear system!

We row reduce this linear system, and get

$$\begin{bmatrix} 1 & 4 & 7 & | & x \\ 0 & 1 & 2 & | & -(y-2x)/3 \\ 0 & 0 & 0 & | & x-2y+z \end{bmatrix}$$

So: the linear system is consistent if and only if $x - 2y + z = 0$. But this says that the span of the given three vectors is only a plane in $\mathbb{R}^3$ and not all of $\mathbb{R}^3$ (and this method even gave us the equation of that plane!).

Notice also that by setting $(x,y,z) = (0,0,0)$ in the preceding example, you could deduce that the dependence equation has infinitely many solutions, and thus that these three vectors are linearly dependent.

Problems

Problem 13.1 a) Find the rank of each matrix in Problem 12.1.

b) *Find the ranks of the coefficient matrices and the ranks of the augmented matrices corresponding to the linear systems in Problem 12.2.

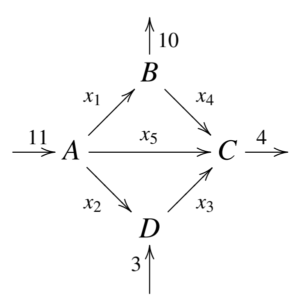
Problem 13.2 Suppose $a, c \in \mathbb{R}$ and consider the following linear system in the variables x, y and z:

$$\begin{array}{ccccccc} x & + & y & + & az & = & 2 \\ 2x & + & y & + & 2az & = & 3. \\ 3x & + & y & + & 3az & = & c \end{array}$$

Note that the general solution of this system may depend on the values of a and c. Let $[A\,|\,b]$ denote the augmented matrix of the system above.

a) For <u>all</u> values of a and c, find
   (i) $\operatorname{rank} A$
   (ii) $\operatorname{rank}[A\,|\,b]$
b) *Find <u>all</u> values of a and c so that this system has
   (i) a unique solution,
   (ii) infinitely many solutions, or
   (iii) no solutions.
c) In case b (ii) above, give the general solution, and its complete geometric description.

Problem 13.3 Consider the network of streets with intersections A, B, C and D below. The arrows indicate the direction of traffic flow along the one-way streets, and the numbers refer to the exact number of cars observed to enter or leave A B,C and D during one minute. Each x_i denotes the unknown number of cars which passed along the indicated streets during the same period.

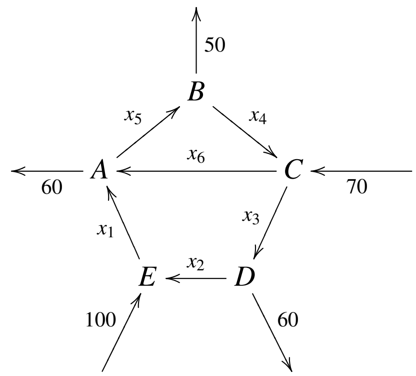


a) Write down the linear system which describes the the traffic flow, **together with all the constraints** on the variables x_i, $i = 1, \ldots, 5$. (Do not perform any operations on your equations: this is done for you in (b).)
b) The reduced row-echelon form of the augmented matrix from part (a) is

$$\left[\begin{array}{ccccc|c} 1 & 0 & 0 & -1 & 0 & 10 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

   Give the general solution. (Ignore the constraints at this point.)
c) If AC were closed due to roadwork, find all possible traffic flows, <u>using your results from (b).</u>

Problem 13.4 Consider the network of streets with intersections A, B, C, D and E below. The arrows indicate the direction of traffic flow along the **one-way streets**, and the numbers refer to the **exact** number of cars observed to enter or leave A, B, C, D and E during one minute. Each x_i denotes the unknown number of cars which passed along the indicated streets during the same period.



a) Write down a system of linear equations which describes the the traffic flow, together with all the constraints on the variables x_i, $i = 1,\dots,6$.

b) *The reduced row-echelon form of the augmented matrix of the system in part (a) is

$$\left[\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & -1 & 1 & 60 \\ 0 & 1 & 0 & 0 & -1 & 1 & -40 \\ 0 & 0 & 1 & 0 & -1 & 1 & 20 \\ 0 & 0 & 0 & 1 & -1 & 0 & -50 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Give the general solution. (Ignore the constraints from (a) at this point.)

c) If $\overline{ED}$ were closed due to roadwork, find the minimum flow along $\overline{AC}$, **using your results from (b).**

Problem 13.5 State whether each of the following is (always) true, or is (possibly) false.

- If you say the statement may be false, you must give an explicit example.
- If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

a) Every non-homogeneous system of 3 equations in 2 unknowns is inconsistent.

b) *Every non-homogeneous system of 3 equations in 2 unknowns is consistent.

c) Every non-homogeneous system of 2 equations in 3 unknowns has infinitely many solutions.

d) *Every system of 2 equations in 2 unknowns has a unique solution.

e) If there is a row of zeroes in the augmented matrix of a linear system, the system will have infinitely many solutions.

f) *If there is a column of zeroes in the coefficient matrix of a consistent linear system, the system will have infinitely many solutions.

g) If a consistent linear system has infinitely many solutions, there must be a row of zeroes in the reduced augmented matrix (i.e. when the augmented matrix has been reduced so that the coefficient matrix is in RREF.)

h) *If a consistent linear system has infinitely many solutions, there must be a column of zeroes in the reduced coefficient matrix (i.e. when the augmented matrix has been reduced so that the coefficient matrix is in RREF.)

i) If the augmented matrix of a homogeneous linear system of 3 equations in 5 unknowns has rank 2, there are 4 parameters in the general solution.

j) *If a homogeneous linear system has a unique solution, it must have the same number of equations as unknowns.

14. Matrix Multiplication

Photo: Ralph Nevins. La dune du Pilat, France

We have seen how to use matrices to solve linear systems. We have also seen essentially two completely different views of an augmented matrix: in the first, the columns represented the coefficients of the variables of a linear system; in the second (our last application) the columns corresponded to vectors in $\mathbb{R}^m$. To really establish the link between these two points of view (which will in turn yield powerful tools to answer the questions we raised about vector spaces in the first part of this course), we need to introduce the notion of matrix multiplication.

That said, matrix multiplication shows up as a key tool in a number of other very concrete applications: not just in linear systems. You see them showing up:

- in probability, as transition matrices of Markov processes;
- in economics, as part of the Leontief input/output model;
- in geometry
- in quantum theory
- for solving linear systems
- for vector spaces, for keeping track of linear combinations

A large part of why they are so versatile is that there are many different ways of thinking of a matrix:

- as a table of numbers
- as a collection of row vectors
- as a collection of column vectors
- as "generalized numbers": things you can add and multiply!

14.1 How to multiply matrices

Matrix multiplication is a generalization of the usual multiplication of numbers; the idea being to multiply across many variables at the same time.

■ **Example 14.1.1** To calculate your weekly expenditures on salaries, you calculate, for each employee,

$$\text{(dollars/hour)} \times \text{(hours/week)} = \text{dollars/week}$$

and then add up over all employees.

Suppose now you have much more information to keep track of:

- The hourly wages of several employees at different jobs. Rows correspond to cashier and stockroom; columns correspond to Ali, Bob and Cho; entries represent the hourly wage of that person in that job:
$$W = \begin{bmatrix} 12 & 10 & 14 \\ 8 & 8 & 10 \end{bmatrix}$$
- The hours worked each week. Rows correspond to Ali, Bob and Cho; columns correspond to Week 1 and Week 2; entries are hours worked by that person that week:
$$H = \begin{bmatrix} 10 & 0 \\ 10 & 10 \\ 5 & 10 \end{bmatrix}$$

Then we could calculate total salary spent at each job each week. Rows are jobs (cashier, stockroom), columns are weeks (Week 1, Week 2) and entries are total cost for that job that week:

$$\begin{aligned} C = WH &= \begin{bmatrix} 12 & 10 & 14 \\ 8 & 8 & 10 \end{bmatrix} \begin{bmatrix} 10 & 0 \\ 10 & 10 \\ 5 & 10 \end{bmatrix} \\ &= \begin{bmatrix} 12\times 10 + 10\times 10 + 14\times 5 & 12\times 0 + 10\times 10 + 14\times 10 \\ 8\times 10 + 8\times 10 + 10\times 5 & 8\times 0 + 8\times 10 + 10\times 10 \end{bmatrix} \\ &= \begin{bmatrix} 290 & 240 \\ 210 & 180 \end{bmatrix} \end{aligned}$$

This says that \$290 were spend on cashiers in Week 1, and \$240 were spent on cashiers in Week 2, for example.

Remarks:

- We organized things so that the rows of the answer correspond to the rows of the first matrix; the columns of the answer correspond to the columns of the second matrix; and the variable over which we summed (here, the employees) corresponded to both the columns of the first and the rows of the second, in the same order.
- Note that we calculated the entry in row i and column j of the product by taking the *dot product* of row i of the first matrix and column j of the second matrix.

■

This example helps to illustrate why we choose to define matrix multiplication in a way that at first seems quite strange.

Definition 14.1.2 If A is an $m\times n$ matrix and B is an $n\times p$ matrix, then their *product* AB is the $m\times p$ matrix whose (i, j) entry is the *dot product* of the ith row of A with the jth column of B.

So, if $A = [a_{ij}]$, and $B = [b_{ij}]$ then $AB = [c_{ij}]$ where

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = [a_{i1}\ a_{i2}\ \cdots\ a_{in}] \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix}$$

In particular, "AB" only makes sense if the number of columns of A equals the number of rows of B.

Example 14.1.3

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x+2y+3z \\ 4x+5y+6z \end{bmatrix} = x \begin{bmatrix} 1 \\ 4 \end{bmatrix} + y \begin{bmatrix} 2 \\ 5 \end{bmatrix} + z \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

That is, this kind of matrix product can be thought of as a shorthand for a linear system *or* as an expression of a linear combination. ■

Example 14.1.4

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} = \begin{bmatrix} a+2c+3e & b+2d+3f \end{bmatrix} = 1 \begin{bmatrix} a & b \end{bmatrix} + 2 \begin{bmatrix} c & d \end{bmatrix} + 3 \begin{bmatrix} e & f \end{bmatrix}$$

So in this way, this matrix product gives a linear combination of the rows of the matrix. ■

14.2 Some strange properties of matrix multiplication

First, note something odd:

Example 14.2.1 Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$. Then A is of size 2×2, B is of size 2×3, so the product is of size 2×3 and equals

$$AB = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 9 & 12 & 15 \\ 19 & 26 & 33 \end{bmatrix}$$

but BA is *not defined*, because the number of columns of B is not the same as the number of rows of A. ■

! SO sometimes you can calculate AB but BA is *not defined*!

Even worse:

Example 14.2.2 Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Then

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

but

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

Note that $AB \neq BA$. ■

> **!** SO even when you ARE allowed to calculate both AB and BA, it can happen (it usually happens) that $AB \neq BA$!

In other words: the product is not commutative.

■ **Example 14.2.3** Let $A = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $B = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$ be two column matrices (or call them vectors). Remember the transpose operation (definition 5.2.4), which swaps rows and columns. Then

- AB is not defined.
- BA is not defined.
- $A^T\,B = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 32 \end{bmatrix} = 32$[1], which is just the dot product of the two vectors.
- $B^T\,A = \begin{bmatrix} 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 32$ is the same (Which is good, since we already knew that the dot product doesn't depend on the order of the matrices.)
- $AB^T = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 4 & 5 & 6 \\ 8 & 10 & 12 \\ 12 & 15 & 18 \end{bmatrix}$ is not at all anything to do with the dot product! We call it the *tensor product*.

■

There are even stranger things about matrix multiplication!

■ **Example 14.2.4** Let $C = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ and $D = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$. Then

$$CD = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

So $CD = 0$ but neither C nor D was the zero matrix. ■

> **!** It can happen that $AB = 0$ but neither A nor B is zero.

■ **Example 14.2.5** Let $A = \begin{bmatrix} 4 & -1 \\ 1 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 \\ 7 & -1 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$. Then

$$AC = \begin{bmatrix} 4 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 4 & 8 \end{bmatrix}$$

[1] We always identify the 1×1 matrix $\begin{bmatrix} a \end{bmatrix}$ with its single (scalar) entry a.

and

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

That is: $AC = BC$ but $A \neq B$! ■

! It can happen that $AC = BC$ but $C \neq 0$ and $A \neq B$. In other words, we can't cancel out C, even if $C \neq 0$.

14.3 Some good properties of matrix multiplication

First note that the transpose operation on matrices satisfies

- $(A+B)^T = A^T + B^T$
- $(kA)^T = kA^T$ for a scalar k
- $(A^T)^T = A$

We also define a special matrix, called the *identity matrix* (although there's actually one for every n):

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

And we still have the *zero matrix* (one for each possible size):

$$0_{2\times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad 0_{2\times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad 0_{3\times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Let's list some properties of the matrix product.

Theorem 14.3.1 — Properties of the matrix product. Let A, B and C be matrices and let k be a scalar. Then, whenever defined, we have

1. $(AB)C = A(BC)$ (Associativity)
2. $A(B+C) = AB + AC$ (Distributivity on the right)
3. $(B+C)A = BA + CA$ (Distributivity on the left)
4. $k(AB) = (kA)B = A(kB)$
5. $(AB)^T = B^T A^T$ (NOTE the reversal of order!)
6. $AI = A$ and $IB = B$.
7. If A is $m \times n$, then $A0_{n\times p} = 0_{m\times p}$ and $0_{q\times m}A = 0_{q\times n}$.

Proof. (1) Write $A = [a_{ij}]$, $B = [b_{ij}]$, and $C = [c_{ij}]$. Suppose A is $m \times n$, B is $n \times p$ and C is of size $p \times q$. Then the ijth entry of $(AB)C$ is

$$\sum_{l=1}^{p} \left(\sum_{k=1}^{n} a_{ik} b_{kl} \right) c_{lj}$$

whereas the ijth entry of $A(BC)$ is

$$\sum_{k=1}^{n} a_{ik}(\sum_{l=1}^{p} b_{kl}c_{lj})$$

and these two sums are equal.

(2)-(4): exercise.

(5) Note first that the change of order was necessary: if A is $m \times n$ and B is $n \times p$, then AB is defined but $A^T B^T$ probably isn't. Now the ijth entry of AB is the dot product of the ith row of A with the jth column of B; so this is the jith entry of $(AB)^T$. Now the jith entry of $B^T A^T$ is the dot product of the jth row of B^T (which is the jth column of B) with the ith column of A^T (which is the ith row of A). Since the order of vectors in a *dot product* doesn't matter, it follows that these are equal.

(6)-(7) exercise. ■

■ Example 14.3.2

$$\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$$

So the size of the identity matrix you need to play the role of 1 in a matrix product depends on the size of A *and* the side you are multiplying on. ■

These properties let us do lots of algebraic manipulations with matrices which are very similar to what we do with real numbers.

Problem 14.3.3 Simplify the expression $(A+B)(C+D)$.
Solution We have $(A+B)(C+D) = (A+B)C + (A+B)D$ by distributivity on the right; and then this equals $AC+BC+AD+BD$ by distributivity on the left.

But we have to be careful!

Problem 14.3.4 Simplify $(A+B)(A-B)$.
Solution This becomes $A^2+BA-AB-B^2$. It is NOT A^2-B^2 since it is very likely that $AB \neq BA$.

Problem 14.3.5 Simplify $AC = BC$.
Solution There's nothing to simplify; we could write this as $AC-BC=0$ or $(A-B)C=0$, but we cannot conclude that $A=B$ or $C=0$, since we've seen already that we can get the zero matrix as the product of two nonzero matrices.

! BEWARE: these operations let you do most simple algebraic operations, but they DON'T let you CANCEL (multiplicatively) or DIVIDE. Look ahead to matrix inversion in chapter 18 to solve this problem.

14.4 Some really amazing things about matrix multiplication

There are some really amazing things about matrix multiplication as well. The following is an application of the Cayley-Hamilton Theorem, for example.

Suppose A is a 2×2 matrix; such a matrix is called square and has the property that you can define $A^2 = AA$ and $A^3 = AAA$ etc. (If A is not square, A^2 is not defined.) So for example if

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

then

$$A^2 = \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 2 & 4 \end{bmatrix}$$

and

$$A^3 = \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 5 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 9 & 10 \\ 10 & 4 \end{bmatrix}$$

etc. Now recall $\mathrm{tr}(A)$ was the trace of A, which is the sum of the diagonal entries. So here, $\mathrm{tr}(A) = 1$. Also $\det(A)$ is $ad - bc$, which here equals -4. The *characteristic polynomial of* A is

$$x^2 - \mathrm{tr}(A)x + \det(A).$$

If we plug in A for x (and put in an identity matrix at the end so that everything becomes a matrix), we get

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

This always happens for 2×2 matrices, and there's a generalization for all square matrices. See definition 22.2.1 for the characteristic polynomial of an general square matrix.

There is also *block multiplication* which can sometimes simplify your calculations by letting you treat submatrices as numbers.

■ **Example 14.4.1** Suppose

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

Break this up this into "blocks" by setting

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad C = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

and then thinking of A as

$$A = \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}$$

where the zero matrix in the $(1,2)$ position is 3×2, and that in the $(2,1)$ position is a 2×3.

Then $A^2 = \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix} \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix} = \begin{bmatrix} B^2 & 0 \\ 0 & C^2 \end{bmatrix}$. In fact,

$$A^{100} = \begin{bmatrix} B^{100} & 0 \\ 0 & C^{100} \end{bmatrix}$$

which is nice, because

$$B^{100} = I_3^{100} = I_3$$

since I_3 is just the identity matrix; and

$$C^2 = 0$$

so $C^{100} = 0$ as well. So in fact $A^{100} = \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix}$. ■

14.5 Back to linear systems

Thinking in terms of blocks –which in the following are just columns – helps us switch between different points of view of linear systems.

The following equations are all equivalent:

- The linear system:

$$\begin{aligned} x + 2y + 3z &= 4 \\ x - y + z &= 2 \\ y - 3z &= 0 \end{aligned}$$

- The matrix equation $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & -1 & 1 \\ 0 & 1 & -3 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}$$

  that is,

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & -1 & 1 \\ 0 & 1 & -3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}$$

- The vector equation

$$x \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} + z \begin{bmatrix} 3 \\ 1 \\ -3 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}$$

Furthermore, all of these can be solved by row reducing the augmented matrix

$$[A|\mathbf{b}] = \begin{bmatrix} 1 & 2 & 3 & | & 4 \\ 1 & -1 & 1 & | & 2 \\ 0 & 1 & -3 & | & 0 \end{bmatrix}.$$

In terms of block multiplication, we can understand this by writing

$$A = \begin{bmatrix} c_1 & c_2 & c_3 \end{bmatrix}$$

where c_i is the ith column of A, and then multiplying

$$A\mathbf{x} = \begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} = x\mathbf{c}_1 + y\mathbf{c}_2 + z\mathbf{c}_3$$

(where we can switch the order because x, y, z are just scalars). This is true for any size of matrix A: if $A = [\mathbf{c}_1\ \mathbf{c}_2\ \cdots \mathbf{c}_n]$ then $Ax = x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \cdots + x_n\mathbf{c}_n$, which is indeed a linear combination of the columns of A.

FACT $A\mathbf{x}$ is a linear combination of the columns of A (with coefficients given by the column vector $\mathbf{x}$).

This gives us a new perspective on our criteria for solving linear systems!

FACT $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ is a linear combination of the columns of A.

FACT $A\mathbf{x} = \mathbf{0}$ has a unique solution if and only if the columns of A are linearly independent.

One proves these facts by going between the various interpretations of the matrix product. For instance, if the system corresponding to $A\mathbf{x} = \mathbf{b}$ is consistent, then this means there is an $\mathbf{x}$ for which the corresponding linear combination of columns of A yields $\mathbf{b}$, which is simply saying that $\mathbf{b}$ is in the span of the columns of A.

Definition 14.5.1 Suppose $A = [\mathbf{c}_1\ \mathbf{c}_2 \ldots \mathbf{c}_n]$ is an $m \times n$ matrix with columns $\mathbf{c}_i$. Set $\mathrm{Col}(A) = \mathrm{span}\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$. This is a subspace of $\mathbb{R}^m$, called the *column space of* A.

FACT $\mathrm{Col}(A) = \mathbb{R}^m$ if and only if the columns of A span $\mathbb{R}^m$, if and only if $A\mathbf{x} = \mathbf{b}$ is consistent for all $\mathbf{b} \in \mathbb{R}^m$, if and only if there is no zero row in the RREF of A.

These things all follow from the definitions, although the last part of the equivalence is not entirely obvious, though one direction is easy.[2]

FACT The columns of A are linearly independent if and only if $A\mathbf{x} = \mathbf{0}$ has a unique solution, if and only if there is a leading one in every column of an REF of A, if and only if $\mathrm{rank}(A)$ equals the number of columns of A.

This again is just putting everything together.

Finally, we deduce something we proved in greater generality before:

FACT If A is an $n \times n$ matrix, then the columns of A form a basis for $\mathbb{R}^n$ if and only if they are linearly independent, if and only if $\mathrm{rank}(A) = n$, if and only if $\mathrm{Col}(A) = \mathbb{R}^n$, if and only if the columns of A

[2]Suppose that there are no zero rows in $\tilde{A}$, the RREF of A. Then $A\mathbf{x} = \mathbf{b}$ will of course be consistent for all $\mathbf{b} \in \mathbb{R}^m$. To prove the converse, we prove that if the the last row in the RREF $\tilde{A}$ of A is zero, then there is a $\mathbf{b} \in \mathbb{R}^m$ with $A\mathbf{x} = \mathbf{b}$ inconsistent: so suppose last row in the RREF $\tilde{A}$ of A is zero. Construct a $\tilde{\mathbf{b}} \in \mathbb{R}^n$ with a 1 in that row – so the system $[\tilde{A}|\tilde{\mathbf{b}}]$ is inconsistent. Now reverse the row operations that took A to $\tilde{A}$ – do them on $[\tilde{A}|\tilde{\mathbf{b}}]$ and you'll obtain a system $[A|\mathbf{b}]$ that is inconsistent; that is, you'll have found a $\mathbf{b} \in \mathbb{R}^n$ with $A\mathbf{x} = \mathbf{b}$ inconsistent.

span $\mathbb{R}^n$ if and only if no REF of A has a zero row.

Problems

Remarks:

1. A question with an asterisk '*' (or two) indicates a bonus-level question.
2. You must justify all your responses.

Matrix Multiplication

Problem 14.1 a) Find the matrix product $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

b) *Write the matrix product $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$ as a linear combination of the columns of $A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$.

c) Write the matrix product $\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ as a linear combination of the rows of $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

d) *Find the matrix product $\begin{bmatrix} a \\ b \end{bmatrix} \begin{bmatrix} c & d \end{bmatrix}$.

e) Find the matrix product $\begin{bmatrix} c & d \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$.

f) *If $A = \begin{bmatrix} c_1 & c_2 & c_3 \end{bmatrix}$ is an $m \times 3$ matrix written in block column form, and $x = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}$ is a column vector in $\mathbb{R}^3$, express Ax as a linear combination of c_1, c_2 and c_3.

g) Show that $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

h) *Show that if $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, then $A^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

i) Show that if $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, and $B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ then $AB \neq BA$.

j) *If C is a $m \times 4$ matrix and $D = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, express the columns of CD in terms of the columns of C.

k) If $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and B is a $3 \times n$ matrix , express the rows of AB in terms of the rows of B

l) *Find all (a, b, c) so that $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} a & b \\ c & a \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

m) Compute $\begin{bmatrix} 1 & 0 & 0 & -3 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{2015}$

Problem 14.2 Show the formula $A^2 - \operatorname{tr}(A)A + \det(A)I = 0$ holds for any 2×2 matrix A.

Problem 14.3 State whether each of the following is (always) true, or is (possibly) false. In this question, A and B are matrices for which the indicated products exist.

- If you say the statement may be false, you must give an explicit example.
- If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

a) $(A+B)^2 = A^2 + 2AB + B^2$

b) *$C(A+B) = CA + CB$

c) $(A+B)I_6 = A + B$

d) *$AB = BA$

e) $(AB)C = A(BC)$

f) *If $A^2 = 0$ for a square matrix A, then $A = 0$.

Problem 14.4 * Suppose A is a *symmetric* $n \times n$ matrix, i.e., $A = A^t$. Show that if v and w are any vectors in $\mathbb{R}^n$, then $(Av) \cdot w = v \cdot (Aw)$.

3

Applications to Linear Systems

Problem 14.5 Write the matrix equation which is equivalent to each of the following linear systems.

a)

$$\begin{array}{rcrcrcl} x & + & y & + & z & = & 0 \\ -9x & - & 2y & + & 5z & = & 0 \\ -x & + & y & + & 3z & = & 0 \\ -7x & - & 2y & + & 3z & = & 0 \end{array}$$

[3]Hint: Remember that if we write vectors as columns, then the dot product $x \cdot y$ is the same as the matrix product $x^t y$. Write the dot products as matrix products and expand.

b) *

$$\begin{array}{ccccccccc} x & & & & & + & w & = & 1 \\ x & & & + & z & + & w & = & 0 \\ x & + & y & + & z & & & = & -3 \\ x & + & y & & & - & 2w & = & 2 \end{array}$$

c)

$$\begin{array}{ccccccccccc} & & & - & 2x_3 & & & + & 7x_5 & = & 12 \\ 2x_1 & + & 4x_2 & - & 10x_3 & + & 6x_4 & + & 12x_5 & = & 28 \\ 2x_1 & + & 4x_2 & - & 5x_3 & + & 6x_4 & - & 5x_5 & = & -1 \end{array}$$

Problem 14.6 Write the matrix equation of the linear system corresponding to each of the augmented matrices given below.

a) $\left[\begin{array}{cc|c} 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right]$

b) * $\left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$

c) $\left[\begin{array}{cccc|c} 1 & 0 & 0 & 3 & 3 \\ 0 & 1 & 0 & 2 & -5 \\ 0 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$

d) * $\left[\begin{array}{ccccc|c} 1 & 2 & 0 & 3 & 0 & 7 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{array}\right]$

e) $\left[\begin{array}{ccccc|c} 1 & 0 & 0 & -1 & 0 & 10 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$

Problem 14.7 State whether each of the following is (always) true, or is (possibly) false. The matrices are assumed to be square.

- If you say the statement may be false, you must give an explicit example.
- If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

a) If $[A\,|\,\mathbf{b}]$ is the augmented matrix of a linear system, then $\operatorname{rank} A > \operatorname{rank}[A\,|\,\mathbf{b}]$ is possible.

b) *If $[A\,|\,\mathbf{b}]$ is the augmented matrix of a linear system, then $\operatorname{rank} A < \operatorname{rank}[A\,|\,\mathbf{b}]$ is possible.

c) If $[A \mid \mathbf{b}]$ is the augmented matrix of a linear system, and $\operatorname{rank} A = \operatorname{rank}[A \mid \mathbf{b}]$, then the system is inconsistent.

d) *If $[A \mid \mathbf{b}]$ is the augmented matrix of a linear system, and $\operatorname{rank} A = \operatorname{rank}[A \mid \mathbf{b}]$, then the system is consistent.

e) If $[A \mid \mathbf{b}]$ is the augmented matrix of a linear system, and $\operatorname{rank} A = \operatorname{rank}[A \mid \mathbf{b}]$, then the vector $\mathbf{b}$ is a linear combination of the columns of A.

f) *If A is an $m \times n$ matrix and $A\mathbf{x} = \mathbf{0}$ has a unique solution for $\mathbf{x} \in \mathbb{R}^n$, then the columns of A are linearly independent.

g) If A is an $m \times n$ matrix and $A\mathbf{x} = \mathbf{0}$ has a unique solution for $\mathbf{x} \in \mathbb{R}^n$, then the rows of A are linearly independent.

h) *If A is an $m \times n$ matrix and $A\mathbf{x} = \mathbf{0}$ has infinitely many solutions for $\mathbf{x} \in \mathbb{R}^n$, then the columns of A are linearly dependent.

i) If A is an $m \times n$ matrix and $A\mathbf{x} = \mathbf{0}$ has infinitely many solutions for $\mathbf{x} \in \mathbb{R}^n$, then the rows of A are linearly dependent.

j) *If A is a 6×5 matrix and $\operatorname{rank} A = 5$, then $A\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$ for $\mathbf{x} \in \mathbb{R}^5$

k) If A is a 6×5 matrix and $\operatorname{rank} A = 5$, then $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b} \in \mathbb{R}^6$.

l) *If A is a 5×6 matrix and $\operatorname{rank} A = 5$, then $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b} \in \mathbb{R}^5$.

m) If A is a 5×6 matrix and $\operatorname{rank} A = 5$, then $A\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$ for $\mathbf{x} \in \mathbb{R}^6$.

n) *If A is a 3×2 matrix and $\operatorname{rank} A = 1$, then $A\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$ for $\mathbf{x} \in \mathbb{R}^2$.

o) The columns of a 19×24 matrix are always linearly dependent.

p) *The rows of a 19×24 matrix are always linearly dependent.

15. Vector spaces associated to Matrices

Last time, we established that the matrix equation $A\mathbf{x} = \mathbf{b}$ could be equally viewed as a system of linear equations as an expression of $\mathbf{b}$ as a linear combination of the columns of A. This allows us to translate and relate concrete facts about systems of linear equations to the abstract world of vector spaces.

Here we introduce and discuss three very useful vector spaces associated to any matrix.

15.1 Column space, row space and nullspace

Let A be an $m \times n$ matrix.

Definition 15.1.1 The *column space* of A (also called the *image* of A, in which case it is denoted $\mathrm{im}(A)$) is

$$\mathrm{Col}(A) = \mathrm{span}\{\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_n\}$$

where $\{\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_n\}$ are the columns of A, viewed as vectors in $\mathbb{R}^m$.

Recall that if $\mathbf{x} \in \mathbb{R}^n$, then $A\mathbf{x}$ is a linear combination of the columns of A. So we may write

$$\mathrm{Col}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$$

Since $\mathrm{Col}(A)$ is given as the span of some vectors in $\mathbb{R}^m$, it is a subspace of $\mathbb{R}^m$.

Although for most applications, column vectors are the key, there is no reason we couldn't also consider row vectors.

Definition 15.1.2 The *row space* of A is

$$\mathrm{Row}(A) = \mathrm{span}\{\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_m\}$$

where $\{\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_m\}$ are the rows of A (typically transposed to make them into $n \times 1$ matrices), viewed as vectors in $\mathbb{R}^n$.

Again, $\text{Row}(A)$ is a subspace, but this time of $\mathbb{R}^n$.

Looking at matrices in a completely different way — that is, as the coefficient matrix of a linear system — yields an equally important subspace.

Definition 15.1.3 The *nullspace* of A (also known as the *kernel* of A, and in that case denoted $\ker(A)$) is

$$\text{Null}(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\};$$

that is, the nullspace is the general solution to the homogeneous linear system given by $A\mathbf{x} = \mathbf{0}$.

Lemma 15.1.4 $\text{Null}(A)$ is a subspace of $\mathbb{R}^n$.

Proof. First note that A is the coefficient matrix of the linear system, and it has size $m \times n$. Thus the linear system has m equations and n variables, meaning that the solution consists of vectors with n components. Hence $\text{Null}(A) \subset \mathbb{R}^n$.

We now need to verify that $\text{Null}(A)$ is a subspace of $\mathbb{R}^n$, and we use the subspace test.

1. First: is $\mathbf{0} \in \text{Null}(A)$? Well, a vector $\mathbf{x}$ is in $\text{Null}(A)$ if and only if $A\mathbf{x} = \mathbf{0}$. Thus since $A\mathbf{0} = \mathbf{0}$, we deduce that $\mathbf{0} \in \text{Null}(A)$.
2. Is $\text{Null}(A)$ closed under vector addition? Well, let $\mathbf{x}$ and $\mathbf{y}$ be two elements of $\text{Null}(A)$. That means $A\mathbf{x} = \mathbf{0}$ and $A\mathbf{y} = \mathbf{0}$. We need to decide if their sum $\mathbf{x}+\mathbf{y}$ lies in $\text{Null}(A)$, meaning, we need to compute $A(\mathbf{x}+\mathbf{y})$.
   Now by distributivity of the matrix product, $A(\mathbf{x}+\mathbf{y}) = A\mathbf{x}+A\mathbf{y}$, and each of these is $\mathbf{0}$ by the above. So we deduce
   $$A(\mathbf{x}+\mathbf{y}) = A\mathbf{x}+A\mathbf{y} = \mathbf{0}+\mathbf{0} = \mathbf{0},$$
   and so $\text{Null}(A)$ is closed under addition.
3. Is $\text{Null}(A)$ closed under scalar multiplication? Let $\mathbf{x} \in \text{Null}(A)$; this means $A\mathbf{x} = \mathbf{0}$. Let k be any scalar. Is $k\mathbf{x} \in \text{Null}(A)$? That is, is it true that $A(k\mathbf{x}) = \mathbf{0}$? We compute, using the axioms of matrix multiplication:
   $$A(k\mathbf{x}) = k(A\mathbf{x}) = k(\mathbf{0}) = \mathbf{0}$$
   so indeed, $A(k\mathbf{x}) = \mathbf{0}$ and thus $\text{Null}(A)$ is closed under scalar multiplication.

Since $\text{Null}(A)$ passes the subspace test, it is a subspace (of $\mathbb{R}^n$). ■

There is a fourth vector space we could naturally associate to A: the nullspace of A^T. But it has no name of its own; we'll use it when we talk about orthogonal complements of subspaces, later on.

! Please note that these vector spaces are generally distinct from one another!

So we have some interesting vector spaces: our next step is to determine their dimension. We begin with $\text{Null}(A)$ and will return to $\text{Col}(A)$ in a few chapters.

15.2 Finding a basis for $\text{Null}(A)$

Problem 15.2.1 Find a basis for $\text{Null}(A)$, where $A = \begin{bmatrix} 1 & 2 & 3 & -3 \\ 0 & 1 & 1 & -2 \\ 1 & 0 & 1 & 1 \end{bmatrix}$.

Solution Since $\text{Null}(A)$ is the general solution to the linear system $A\mathbf{x} = \mathbf{0}$, our first step is to solve that linear system. So we write down the augmented matrix and row reduce to RREF:

$$\left[\begin{array}{cccc|c} 1 & 2 & 3 & -3 & 0 \\ 0 & 1 & 1 & -2 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{array}\right] \underset{-R_1+R_3 \to R_3}{\sim} \left[\begin{array}{cccc|c} 1 & 2 & 3 & -3 & 0 \\ 0 & 1 & 1 & -2 & 0 \\ 0 & -2 & -2 & 4 & 0 \end{array}\right]$$

$$\begin{array}{c} 2R_2+R_3 \to R_3 \\ \sim \\ -2R_2+R_1 \to R_1 \end{array} \left[\begin{array}{cccc|c} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

which is in RREF.

The general solution comes by setting the nonleading variables equal to a parameter ($x_3 = r$, $x_4 = t$) and solving for the leading variables in terms of the nonleading variables (so $x_1 = -x_3 - x_4 = -r - t$ and $x_2 = -x_3 + 2x_4 = -r + 2t$). Written in vector form (since we're interested in vectors) this is

$$\begin{aligned} \text{Null}(A) &= \left\{ \begin{bmatrix} -r-t \\ -r+2t \\ r \\ t \end{bmatrix} \mid r,t \in \mathbb{R} \right\} \\ &= \left\{ r \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 2 \\ 0 \\ 1 \end{bmatrix} \mid r,t \in \mathbb{R} \right\} = \text{span} \left\{ \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 0 \\ 1 \end{bmatrix} \right\} \end{aligned}$$

Since these vectors span $\text{Null}(A)$, and since the set

$$\left\{ \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is linearly independent (just look at the two entries in these vectors – the ones corresponding to our nonleading variables!), we deduce that this is a basis for $\text{Null}(A)$.

The vectors in the spanning set that we obtain by writing down the solution from the RREF form has a special name: they are the *basic solutions*. This name certainly suggests they ought to form a basis.

Theorem 15.2.2 — Basic solutions form a basis for $\text{Null}(A)$. The spanning set of $\text{Null}(A)$ obtained from the RREF of $[A|0]$ (in other words, the set of basic solutions of $A\mathbf{x} = \mathbf{0}$) is a basis for $\text{Null}(A)$.

Proof. By definition, this is a spanning set, so it only remains to prove linear independence. Let $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ be the set of basic solutions. By construction, the number of basic solutions k equals the number of parameters in the general solution, and this equals the number of non-leading variables. Now note that for each non-leading variable x_{n_i}, there is a unique vector $\mathbf{v}_i$ which has a nonzero entry in row n_i. (This is by construction, since row n_i corresponds in our general solution to the equation $x_{n_i} = r_i$, a parameter; in particular only one parameter and thus only one basic solution contributes to this variable.)

Suppose that $\mathbf{v}_i$ is the basic vector corresponding to setting the parameter for x_{j_i} equal to 1 and the rest equal to zero.

Consider now the dependence equation

$$r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_k\mathbf{v}_k = \mathbf{0}$$

By what we said above, for each $i = 1, 2\cdots, k$, the n_ith row on the LHS has value r_i, which must equal the n_ith element on the RHS, which is zero. So we deduce that $r_1 = r_2 = \cdots = r_k = 0$. Thus the set of basic solutions is linearly independent, and hence a basis. ■

Corollary 15.2.3 — Rank-Nullity Theorem. The dimension of the nullspace of A is equal to the number of nonleading variables of A. That is,

$$\dim \mathrm{Null}(A) + \mathrm{rank}(A) = n$$

where n is the number of columns of A.

Proof. The dimension of a space is the number of vectors in a basis for that space, so $\dim \mathrm{Null}(A)$ is equal to the number of nonleading variables of A. Recall that the rank of A equals the number of leading variables, and n is the number of variables in total; so the number of nonleading variables equals $n - \mathrm{rank}(A)$. ■

15.3 What all this tells us about solutions to inhomogeneous systems

Problem 15.3.1 Solve $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & 2 & 3 & -3 \\ 0 & 1 & 1 & -2 \\ 1 & 0 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 10 \\ 3 \\ 4 \end{bmatrix}$$

Solution We row reduce the resulting augmented matrix:

$$\left[\begin{array}{cccc|c} 1 & 2 & 3 & -3 & 10 \\ 0 & 1 & 1 & -2 & 3 \\ 1 & 0 & 1 & 1 & 4 \end{array}\right] \underset{-R_1+R_3\to R_3}{\sim} \left[\begin{array}{cccc|c} 1 & 2 & 3 & -3 & 10 \\ 0 & 1 & 1 & -2 & 3 \\ 0 & -2 & -2 & 4 & -6 \end{array}\right]$$

$$\begin{array}{c} 2R_2+R_3\to R_3 \\ \sim \\ -2R_2+R_1\to R_1 \end{array} \left[\begin{array}{cccc|c} 1 & 0 & 1 & 1 & 4 \\ 0 & 1 & 1 & -2 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

which is in RREF. *Note that we used the SAME operations as before, when we solved the corresponding homogeneous system!*

So our general solution is

$$\left\{ \begin{bmatrix} 4-r-t \\ 3-r+2t \\ r \\ t \end{bmatrix} \mid r,t \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} 4 \\ 3 \\ 0 \\ 0 \end{bmatrix} + r \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 2 \\ 0 \\ 1 \end{bmatrix} \mid r,t \in \mathbb{R} \right\}$$

The thing to notice is that the general solution to the inhomogeneous equation is given by adding a particular solution of the inhomogeneous system to the general solution of the homogeneous system. Geometrically, this means we take the subspace $\text{Null}(A)$ and translate it away from the origin to produce the solution set to the inhomogeneous system. (And the vector by which we translate it is just a (any) particular solution to the system.)

We can state this as a theorem.

Theorem 15.3.2 — Inhomogeneous systems and nullspace. Suppose that $A\mathbf{x} = \mathbf{b}$ is a consistent linear system.

1. If $\mathbf{x} = \mathbf{v}$ is a solution to the system $A\mathbf{x} = \mathbf{b}$, and $\mathbf{x} = \mathbf{u}$ is any solution to the associated homogeneous system $A\mathbf{x} = \mathbf{0}$, then $\mathbf{v} + \mathbf{u}$ is a solution to $A\mathbf{x} = \mathbf{b}$.
2. If $\mathbf{v}$ and $\mathbf{w}$ are two solutions to the system $A\mathbf{x} = \mathbf{b}$, then $\mathbf{x} = \mathbf{v} - \mathbf{w}$ is a solution to the homogeneous system $A\mathbf{x} = \mathbf{0}$.

These two statements together ensure that the general solution to the inhomogeneous system is given exactly by $\mathbf{v}$ plus the solutions to the homogeneous system.

Proof. (1) We have $A\mathbf{v} = \mathbf{b}$ and $A\mathbf{u} = \mathbf{0}$ so $A(\mathbf{v}+\mathbf{u}) = A\mathbf{v} + A\mathbf{u} = \mathbf{b} + \mathbf{0} = \mathbf{b}$.

(2) We have $A\mathbf{v} = \mathbf{b}$ and $A\mathbf{w} = \mathbf{b}$ so $A(\mathbf{v}-\mathbf{w}) = A\mathbf{v} - A\mathbf{w} = \mathbf{b} - \mathbf{b} = \mathbf{0}$

■

The value of the theorem is in giving us a picture of the solution set of an inhomogeneous system — in particular, if a homogeneous system has a unique solution, so does any consistent inhomogeneous system with the same coefficient matrix; and if a homogeneous system has infinitely many solutions, indexed by k basic solutions, then any consistent inhomogeneous system with the same coefficient matrix will also have k parameters in its solution.

But unless you happen to stumble across a particular solution (by guessing, or advance knowledge) of your inhomogeneous system, this theorem doesn't help you solve the system. In particular, if $\mathbf{b} \notin \text{Col}(A)$, there won't be a solution at all, and knowing the nullspace is of no help whatsoever.

15.4 Summary of facts relating to consistency of a linear system

Let A be an $m \times n$ matrix.

A linear system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if:

- $\mathbf{b}$ is a linear combination of the columns of A, iff
- $\mathbf{b} \in \text{Col}(A)$ iff
- $\text{rank}[A \mid \mathbf{b}] = \text{rank} A$.

Theorem 15.4.1 Let A be an $m \times n$ matrix and $\mathbf{x} \in \mathbb{R}^n$. The following statements are equivalent for a system with matrix equation $A\mathbf{x} = \mathbf{b}$:
(1) The equation $A\mathbf{x} = \mathbf{b}$ is be consistent *for all choices of* $\mathbf{b} \in \mathbb{R}^m$;
(2) $\text{rank}(A) = m$;
(3) There are no zero rows in the RREF of A;
(4) Every $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of A;
(5) $\text{Col}(A) = \mathbb{R}^m$;
(6) $\dim(\text{Col}(A)) = m$.

Proof. There are several ways to do this. We will prove (1) $\implies$ (2) $\implies$ (3) $\implies$ (4) $\implies$ (5) $\iff$ (6) and (5) $\implies$ (1). This will achieve a proof of the equivalence of the statements (1) – (6).

(1) $\implies$ (2): Suppose $A\mathbf{x} = \mathbf{b}$ is consistent for all $b \in \mathbb{R}^m$, but that $\text{rank} A < m$. So the last row in the RREF $\tilde{A}$ of A must be zero.

Choose a $\tilde{\mathbf{b}} \in \mathbb{R}^n$ with a 1 as its last entry – so the system $[\tilde{A}|\tilde{\mathbf{b}}]$ is inconsistent. Now reverse the row operations that took A to $\tilde{A}$ – do them on $[\tilde{A}|\tilde{\mathbf{b}}]$ – and we obtain a system $[A|b]$ that we know is inconsistent; that is, we've found a $\mathbf{b} \in \mathbb{R}^n$ with $A\mathbf{x} = \mathbf{b}$ inconsistent. This is a contradiction, so $\text{rank} A = m$.

(2) $\implies$ (3): As A is an $m \times n$ matrix, this is obvious.

(3) $\implies$ (4): If there are no zero rows in the RREF of A, $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b} \in \mathbb{R}^m$, and so every $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of A.

(4) $\implies$ (5): If every $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of A then the columns of A span all of $\mathbb{R}^m$, so $\text{Col}(A) = \mathbb{R}^m$ (Of course we always have $\text{Col}(A) \subset \mathbb{R}^m$ — it's the equality here that is interesting.)

(5) $\iff$ (6): We know that a subspace of $\mathbb{R}^m$ has dimension m iff it's the whole space.

(5) $\implies$ (1): Since every $\mathbf{b} \in \mathbb{R}^m$ is in $\text{Col}(A)$, every $\mathbf{b} \in \mathbb{R}^m$is a linear combination of the columns of A, and so $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b} \in \mathbb{R}^m$. ■

15.5 Summary of facts relating to the number of solutions of a consistent linear system

Theorem 15.5.1 Let A be an $m \times n$ matrix, and $\mathbf{b} \in \mathbb{R}^n$. The following statements are equivalent for a consistent system with matrix equation $A\mathbf{x} = \mathbf{b}$:
(1) The system $A\mathbf{x} = \mathbf{b}$ has a unique solution;
(2) Every variable is a leading variable;
(3) There is a leading 1 in every column of the RREF of A;
(4) The associated homogeneous sysytem $A\mathbf{x} = \mathbf{0}$ has a unique solution;
(5) The columns of A are linearly independent;
(6) $\text{Null}(A) = \{\mathbf{0}\}$;
(7) $\dim(\text{Null}(A)) = 0$;
(8) $\text{rank}(A) = n$.

Proof. (1) $\implies$ (2): If the solution is unique, there cannot be any parameters and so every variable is a leading variable.

(2) $\implies$ (3): Leading variables are the variables with a leading 1 in their column in an RREF of A, so this is clear.

(3) $\implies$ (4): If there's a leading one in very column of the RREF of A, there are no parameters in the general solution to $A\mathbf{x} = \mathbf{0}$, so the solution is unique.

(4) $\implies$ (5): The linear system $A\mathbf{x} = \mathbf{0}$ is the same as the vector equation

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

where the $\mathbf{c}_i$ are the columns of A. This vector equation has a unique solution if and only if the vectors $\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ are linearly independent.

(5) $\implies$ (6): Null(A) is the general solution to $A\mathbf{x} = \mathbf{0}$, so this consists of the unique solution which is the trivial solution.

(6) $\implies$ (7): Recall again our theorem about the dimensions of subspaces; Null(A) is the zero subspace if and only if its dimension is zero.

(7) $\implies$ (8): This is a direct consequence of the Rank-Nullity Theorem, theorem 15.2.3

(8) $\implies$ (1): If rank$A = n$, then there are no parameters in the general solution to $A\mathbf{x} = \mathbf{b}$, so the solution is unique. ■

15.6 Application

We've seen subspaces of $\mathbb{R}^n$ defined in two kinds of ways:

1. It is given as the nullspace of some matrix A, *i.e.* $W = \text{Null}(A)$ or
2. It is given as the span of some vectors, *i.e.* $W = \text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$

We already know how to convert a description of type (1) into type (2): simply find a basis of Null(A). But how can we convert a description of type (2) into type (1)? Hang on – why would we want to?

Well, suppose we're given some vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ and we're asked "Are the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ in W?" If we have a description of W in form (1), all we need to do is compute $A\mathbf{u}_1, \ldots, A\mathbf{u}_k$ and see if we get zero in each case. That's pretty easy. If, on the other hand, we only know W in the second form, we'd have to *solve the k linear systems* $\begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_m & | & \mathbf{u}_j\end{bmatrix}$ *for* $1 \le j \le k$*!* That's more effort — especially if $m > 1$.

Let's look at a couple of examples first.

■ **Example 15.6.1** Let $W = \text{span}\{(1,1,2)\}$. Let's find a matrix A such that $W = \text{Null}A$.

Well, $(x,y,z) \in W$ iff the system with augmented matrix

$$\begin{bmatrix} 1 & | & x \\ 1 & | & y \\ 2 & | & z \end{bmatrix}$$

is consistent. But we know this system is equivalent (after row reduction) to the system whose augmented matrix is

$$\begin{bmatrix} 1 & | & x \\ 0 & | & y-x \\ 0 & | & z-2x \end{bmatrix}$$

Now we know that this system is consistent iff $x - y = 0$ and $2x - z = 0$. So

$$W = \{(x,y,z) \in \mathbb{R}^3 \mid x - y = 0 \text{ and } 2x - z = 0\}$$

Thus if $A = \begin{bmatrix} 1 & -1 & 0 \\ 2 & 0 & -1 \end{bmatrix}$, then Null$A = W$. ■

That was a pretty simple example, and it wouldn't have been difficult to decide if a vector is a multiple of $(1,1,2)$. Nevertheless, we did end up expressing the line W as the intersection of two planes!

Let's try an example that's a bit more interesting.

■ **Example 15.6.2** Define a subspace W of $\mathbb{R}^4$ by

$$W = \text{span}\{(1,0,0,1), (1,1,1,0), (2,1,-1,1)\}$$

We'll proceed as before: make W the column space of a matrix, *i.e.*

$$W = \text{Col}\begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}.$$

So $(x,y,z,w) \in W$ iff the system with augmented matrix

$$\begin{bmatrix} 1 & 1 & 2 & | & x \\ 0 & 1 & 1 & | & y \\ 0 & 1 & -1 & | & z \\ 1 & 0 & 1 & | & w \end{bmatrix}$$

is consistent. So we row reduce (and we don't need to go all the way) and find that this system is equivalent to

$$\begin{bmatrix} 1 & 1 & 2 & | & x \\ 0 & 1 & 1 & | & y \\ 0 & 0 & -2 & | & z-y \\ 0 & 0 & 0 & | & -x+y+w \end{bmatrix}.$$

Now we can see that this is consistent iff $x - y - w = 0$, *i.e.*

$$W = \{(x,y,z,w) \mid x-y-w = 0\}.$$

So $W = \text{Null}\begin{bmatrix} 1 & -1 & 0 & -1 \end{bmatrix}$. Figuring out if a vector is in W is now a piece of cake — just one equation to check!

■

Problems

Problem 15.1 Find a basis for the kernel (also called nullspace) of the following matrices:

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

b) * $\begin{bmatrix} 1 & 2 & -1 & 3 \end{bmatrix}$

c) $\begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 2 \end{bmatrix}$

d)* $\begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

e) $\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}$

f)* $\begin{bmatrix} -4 & 2 & 2 \\ 2 & -4 & 2 \\ 2 & 2 & -4 \end{bmatrix}$

g) $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

Problem 15.2 For each of the following subspaces W, find a matrix A such that $W = \text{Null}(A)$.

1. $W = \text{span}\{(1,2,3),(1,0,1)\}$
2. $W = \text{span}\{(1,0,1,0),\ (1,1,1,0)\}$
3. $W = \text{span}\{(1,0,1,0),\ (1,1,1,0),\ (1,1,1,1)\}$

Photo: Ralph Nevins. La dune du Pilat, France

16. The row and column space algorithms

Here we aim to solve the following three problems about finding bases for subspaces of $\mathbb{R}^n$. The solutions to each will involve the Gaussian algorithm at some stage.

1. Given a spanning set $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ for a subspace W, find *any* basis for W.
2. Given a spanning set $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ for a subspace W, find a basis for W which *is a subset of* $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$, i.e., the basis should only contain vectors from $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$.
3. Given a linearly independent set $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ in $\mathbb{R}^n$, *extend it to a basis of* $\mathbb{R}^n$, i.e., find vectors $\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n$ such that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is a basis of $\mathbb{R}^n$.

The key ideas in the solutions to these problems will be to find bases for the row and column space of a matrix. Let's start with the row space, as it's pretty straightforward.

16.1 Finding a basis for the row space: the row space algorithm

When we row reduce a matrix, you may have noticed that the new rows are linear combinations of the old ones. In Chapter 11 we mentioned that each elementary row operation can be reversed, and indeed by an elementary row operation of the same kind. So this means that the old rows are linear combinations of the new rows as well. This has a very beautiful and useful consequence:

Proposition 16.1.1 — The row space is invariant under row equivalence. If A is row equivalent to B then $\operatorname{Row}(A) = \operatorname{Row}(B)$. That is, the spans of their rows are exactly the same subspace of $\mathbb{R}^n$.

Proof. Remember that "row equivalent" means that you can get from A to B by a sequence of elementary row operations. The three elementary row operations are:

1. Add a multiple of one row to another.
2. Multiply a row by a nonzero scalar.

3. Interchange two rows.

Replace the word “row” with “row vector”; it is clear that (2) and (3) do not change the span of the row vectors involved. To see that (1) doesn’t change the span amounts to showing that

$$\operatorname{span}\{\mathbf{u},\mathbf{v}\} = \operatorname{span}\{\mathbf{u}, c\mathbf{u}+\mathbf{v}\}$$

for any two vectors **u** and **v** and a scalar c. (Compare this with Problem 6.4). Since none of the individual operations change the span of the row vectors, the row spaces of A and B are the same. ■

What’s so great about this? Well, for one thing a matrix in RREF has lots of zeros about, so it might be easier to spot a basis. Let A be any matrix and $\tilde{A}$ its RREF. Given that $\operatorname{Row}(A) = \operatorname{Row}(\tilde{A})$, a basis for $\operatorname{Row}(\tilde{A})$ will also be a basis for $\operatorname{Row}(A)$!

Problem 16.1.2 Find a basis for $\operatorname{Row}(A)$, where A is the matrix

$$A = \begin{bmatrix} 1 & 2 & 2 & 3 \\ 2 & -2 & -8 & 4 \\ 1 & 1 & 0 & 1 \\ 0 & 2 & 4 & 1 \end{bmatrix}$$

Solution By definition,

$$\operatorname{Row}(A) = \operatorname{span}\{(1,2,2,3),(2,-2,-8,4),(1,1,0,1),(0,2,4,1)\}$$

But again, we don’t know if this spanning set is linearly independent or not.

Let’s apply Proposition 16.1.1.

We row reduce A to RREF and get:

$$\tilde{A} = \begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The proposition says that $\operatorname{Row}(A) = \operatorname{Row}(\tilde{A})$. Well, $\operatorname{Row}(\tilde{A})$ is spanned by the rows of $\tilde{A}$ — but it’s perfectly obvious we can ignore the zero row, so we have

$$\operatorname{Row}(\tilde{A}) = \operatorname{span}\{(1,0,-2,0),(0,1,2,0),(0,0,0,1)\}$$

Even better: just like the basic solutions to $\operatorname{Null}(A)$, these vectors are obviously linearly independent because of the positions of the leading ones.

Conclusion : The nonzero rows of $\tilde{A}$ form a basis for $\operatorname{Row}(A)$.

And what a fantastic basis! Lots of zeros! It’s easy to see that, for example, $(3,2,-2,5)$ is in the $\operatorname{Row}(A)$ and that $(1,1,1,1)$ is not, because of the nice ones and zeros. (Try it!)

Having found a basis for the row space here, let’s take a look: did this help with finding a basis for the column space? Unfortunately, not: look at the column spaces of A and $\tilde{A}$. They are not the same at all: for example, the vector $c_2 = \begin{bmatrix} 2 \\ -2 \\ 1 \\ 2 \end{bmatrix}$ (the second column of A) is certainly in $\operatorname{Col}(A)$, but it isn’t in

$\mathrm{Col}(\tilde{A})$, since —take a look— the fourth component of *every vector* in $\mathrm{Col}(\tilde{A})$ will have to be 0. And that's simply not the case for c_2.

! If $A \sim \tilde{A}$ then $\mathrm{Row}(A) = \mathrm{Row}(\tilde{A})$ but $\mathrm{Col}(A) \neq \mathrm{Col}(\tilde{A})$ in general!

Let's record our result for the row space:

Theorem 16.1.3 — Basis for $\mathrm{Row}(A)$: The row space algorithm. The nonzero rows in any REF of A form a basis for $\mathrm{Row}(A)$, and so $\dim \mathrm{Row}(A) = \mathrm{rank}\, A$.

If you take the RREF, you get a "standard basis" for your subspace $\mathrm{Row}(A)$.

! You have to take the rows of an REF of A, not the rows of A.

■ **Example 16.1.4** Let $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 0 & 1 \end{bmatrix}$; its RREF is $\tilde{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$. The first two rows of $\tilde{A}$ give a basis for $\mathrm{Row}(\tilde{A}) = \mathrm{Row}(A)$. But the first two rows of A do not. And the nonzero rows of A also do not. (check). The problem with going back to A is that you would need to keep track of row interchanges. ■

This solves our first problem on page 165! Given a spanning set for a subspace W, assemble a matrix A with the vectors as the *rows* of A, so that $W = \mathrm{Row}(A)$. Now apply the row space algorithm to obtain a basis for $\mathrm{Row}(A) = W$.

■ **Example 16.1.5** Find any basis for

$$W = \mathrm{span}\{(1,2,2,3),(2,-2,-8,4),(1,1,0,1),(0,2,4,1)\}$$

Well, let's just assemble a matrix A for which $W = \mathrm{Row}(A)$, apply the row space algorithm and get a basis for $\mathrm{Row}(A)$, which will of course be a basis for W. This was done in Problem 16.1.2. So

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

is a basis for W.

N.B. Note that none of these vectors is in our original spanning set, so this doesn't help with the second problem on page 165. We'll need another tool for that. ■

The row space algorithm also helps us solve the third problem on page 165: Given a linearly independent set $\{\mathbf{u}_1,\ldots,\mathbf{u}_k\}$ in $\mathbb{R}^n$, to extend it to a basis of $\mathbb{R}^n$, i.e., find vectors $\mathbf{u}_{k+1},\ldots,\mathbf{u}_n$ such that $\{\mathbf{u}_1,\ldots,\mathbf{u}_k,\mathbf{u}_{k+1},\ldots,\mathbf{u}_n\}$ is a basis of $\mathbb{R}^n$, we proceed as follows: Assemble an $n \times n$ matrix A whose first k *rows* are $\mathbf{u}_1,\ldots,\mathbf{u}_k$, and leave the $n-k$ unknown vectors as symbols i.e. $A = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_k \\ \mathbf{u}_{k+1} \\ \vdots \\ \mathbf{u}_n \end{bmatrix}$. We know that the rows of A will be a basis of $\mathbb{R}^n$ iff $\mathrm{rank}\, A = n$. So we row reduce just the part of A we know:

rows 1 to k, and see which columns the leading ones are in. We then choose vectors $\mathbf{u}_{k+1},\ldots,\mathbf{u}_n$ from the standard basis of $\mathbb{R}^n$ to make sure there's a leading one in every column. Let's see this in two examples:

■ **Example 16.1.6** Extend the linearly independent set $\{(0,1,1)\}$ to a basis of $\mathbb{R}^3$. Since $\dim\mathbb{R}^3 = 3$, we'll need two more vectors.

Set $A = \begin{bmatrix} 0 & 1 & 1 \\ & \mathbf{u}_2 & \\ & \mathbf{u}_3 & \end{bmatrix} = \begin{bmatrix} 0 & ① & 1 \\ & \mathbf{u}_2 & \\ & \mathbf{u}_3 & \end{bmatrix}$. Since we have one leading one in the second column, if we set $\mathbf{u}_2 = (1,0,0)$ and $\mathbf{u}_3 = (0,0,1)$, then

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

Then it is clear that with these choices, $\operatorname{rank} A = 3$, so $\dim \operatorname{Row}(A) = 3$, and a basis for $\operatorname{Row}(A) = \mathbb{R}^3$ will be $\{(0,1,1),(1,0,0),(0,0,1)\}$.

■

Let's try an example in $\mathbb{R}^4$.

■ **Example 16.1.7** Extend $\{(1,2,3,4),(2,4,7,8)\}$ to a basis of $\mathbb{R}^4$.

So we set $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 7 & 8 \\ & & \mathbf{u}_3 & \\ & & \mathbf{u}_4 & \end{bmatrix}$. Now row reduce as much as we can:

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 7 & 8 \\ & & \mathbf{u}_3 & \\ & & \mathbf{u}_4 & \end{bmatrix} \overset{-2R_1+R_2\to R_2}{\sim} \begin{bmatrix} ① & 2 & 3 & 4 \\ 0 & 0 & ① & 0 \\ & & \mathbf{u}_3 & \\ & & \mathbf{u}_4 & \end{bmatrix}$$

So we can see two leading ones – in columns 1 and 3, so if we set $\mathbf{u}_3 = (0,1,0,0)$ and $\mathbf{u}_4 = (0,0,0,1)$, then

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 7 & 8 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} ① & 2 & 3 & 4 \\ 0 & 0 & ① & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \overset{R_2\leftrightarrow R_3}{\sim} \begin{bmatrix} ① & 2 & 3 & 4 \\ 0 & ① & 0 & 0 \\ 0 & 0 & ① & 0 \\ 0 & 0 & 0 & ① \end{bmatrix}$$

It is now clear that $\operatorname{rank} A = 4$, so the desired extension is

$$\{(1,2,3,4),(2,4,7,8),(0,1,0,0),(0,0,0,1)\}$$

■

Finally, let's tackle the second problem on page 165. For this we'll need a new tool – which we can fashion out of things we already know.

16.2 Finding a basis for the column space: the column space algorithm

We noted before that row reduction really messes up the column space of a matrix: if A row reduces to $\tilde{A}$, then it's generally[1] the case that $\mathrm{Col}(A) \neq \mathrm{Col}(\tilde{A})$.[2]

We could always find a basis for $\mathrm{Col}(A)$ by finding a basis for $\mathrm{Row}(A^t)$, but here's another way that obtains a very special kind of basis: one that is a subset of the columns!

Here's the idea: given a matrix $A = \begin{bmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_n \end{bmatrix}$ in block column form, we'll try to collect, from left to right, the greatest number of those columns that are linearly independent, discarding any that are not. As you will see, we will be able to decide which columns to keep and which to discard simply by looking at the RREF of A!

Let's start with an example. Before we do, we recall two important facts that will turn out to be extremely useful:

(1) Theorem 8.3.1: If $\{\mathbf{v}_1, \dots \mathbf{v}_k\}$ is linearly in dependent, then $\mathbf{v}_{k+1} \notin \mathrm{span}\{\mathbf{v}_1, \dots \mathbf{v}_k\}$ iff $\{\mathbf{v}_1, \dots \mathbf{v}_k, \mathbf{v}_{k+1}\}$ is still linearly independent; and

(2) From the beginning of Section 15.4: If $\begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_k & | & \mathbf{v} \end{bmatrix}$ is the augmented matrix of a linear system (in block column form), then $\mathbf{v} \in \mathrm{span}\{\mathbf{v}_1, \dots \mathbf{v}_k\}$ iff

$$\mathrm{rank}\begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_k & | & \mathbf{v} \end{bmatrix} = \mathrm{rank}\begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_k \end{bmatrix}.$$

Problem 16.2.1 Find a basis for $\mathrm{Col}(A)$, where

$$A = \begin{bmatrix} 1 & 2 & 2 & 3 \\ 2 & -2 & -8 & 4 \\ 1 & 1 & 0 & 1 \\ 0 & 2 & 4 & 1 \end{bmatrix}$$

Solution We'll try to collect, from left to right, the greatest number of those columns that are linearly independent, discarding any that are not.

Write $A = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & \mathbf{c}_4 \end{bmatrix}$ in block column form. Now it's clear that as $\mathbf{c}_1 \neq 0$, the set $\{\mathbf{c}_1\}$ is linearly independent. But will it span $\mathrm{Col}(A)$? Well, let's see if $\mathbf{c}_2 \in \mathrm{span}\{\mathbf{c}_1\}$. It's obvious that it isn't, but let's look at this in another way that will turn out to be helpful: We know that $\mathbf{c}_2 \in \mathrm{span}\{\mathbf{c}_1\}$ iff the linear system with augmented matrix

$$\begin{bmatrix} \mathbf{c}_1 & | & \mathbf{c}_2 \end{bmatrix} = \begin{bmatrix} 1 & | & 2 \\ 2 & | & -2 \\ 1 & | & 1 \\ 0 & | & 2 \end{bmatrix}$$

is consistent, iff $\mathrm{rank}\begin{bmatrix} \mathbf{c}_1 & | & \mathbf{c}_2 \end{bmatrix} = \mathrm{rank}\begin{bmatrix} \mathbf{c}_1 \end{bmatrix}$. When we row reduce this, it reduces to the first two columns of $\tilde{A}$, of course:

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

[1]We'll see later that they are the same for *invertible* square matrices. See Chapter 18.
[2]Although it will turn out that they *will* have the same dimension for all matrices.

The presence of the second leading one in column 2 of $\tilde{A}$ prevents this system from being consistent. That is, rank $\begin{bmatrix}\mathbf{c}_1 & | & \mathbf{c}_2\end{bmatrix} > \text{rank}\begin{bmatrix}\mathbf{c}_1\end{bmatrix}$. So,

$$\mathbf{c}_2 \notin \text{span}\{\mathbf{c}_1\},$$

and hence $\{\mathbf{c}_1, \mathbf{c}_2\}$ is linearly independent: span$\{\mathbf{c}_1, \mathbf{c}_2\}$ is genuinely larger than span$\{\mathbf{c}_1\}$. So we keep $\mathbf{c}_1$ and $\mathbf{c}_2$.

Ok, so let's see if $\{\mathbf{c}_1, \mathbf{c}_2\}$ spans Col(A). Well if it does, then $\mathbf{c}_3 \in \text{span}\{\mathbf{c}_1, \mathbf{c}_2\}$. We know that this is true iff $\begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & | & \mathbf{c}_3\end{bmatrix}$ is the augmented matrix of a consistent system iff rank $\begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & | & \mathbf{c}_3\end{bmatrix} =$ rank $\begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2\end{bmatrix}$. Now

$$\begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & | & \mathbf{c}_3\end{bmatrix} = \begin{bmatrix}1 & 2 & | & 2\\ 2 & -2 & | & -8\\ 1 & 1 & | & 0\\ 0 & 2 & | & 4\end{bmatrix} \sim \begin{bmatrix}1 & 0 & | & -2\\ 0 & 1 & | & 2\\ 0 & 0 & | & 0\\ 0 & 0 & | & 0\end{bmatrix}$$

This *is* the augmented matrix of a consistent system: rank $\begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & | & \mathbf{c}_3\end{bmatrix} = \text{rank}\begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2\end{bmatrix}$ this time —no new leading one in column three of $\tilde{A}$ — so it *is true* that

$$\mathbf{c}_3 \in \text{span}\{\mathbf{c}_1, \mathbf{c}_2\}.$$

Hence span$\{\mathbf{c}_1, \mathbf{c}_2\} = \text{span}\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\}$, so we won't keep $\mathbf{c}_3$, as $\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\}$ has the same span as $\{\mathbf{c}_1, \mathbf{c}_2\}$ but $\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\}$ is *not* independent.

So now we ask: is $\mathbf{c}_4 \in \text{span}\{\mathbf{c}_1, \mathbf{c}_2\}$? Since span$\{\mathbf{c}_1, \mathbf{c}_2\} = \text{span}\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\}$, this is the same question as: "is $\mathbf{c}_4 \in \text{span}\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\}$?"[a] It will be iff the linear system corresponding to $\begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & | & \mathbf{c}_4\end{bmatrix}$ is consistent iff rank $\begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & | & \mathbf{c}_4\end{bmatrix} = \text{rank}\begin{bmatrix}\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\end{bmatrix}$. Well,

$$\begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & | & \mathbf{c}_4\end{bmatrix} = \begin{bmatrix}1 & 2 & 2 & | & 3\\ 2 & -2 & -8 & | & 4\\ 1 & 1 & 0 & | & 1\\ 0 & 2 & 4 & | & 1\end{bmatrix} \sim \begin{bmatrix}1 & 0 & -2 & | & 0\\ 0 & 1 & 2 & | & 0\\ 0 & 0 & 0 & | & \textcircled{1}\\ 0 & 0 & 0 & | & 0\end{bmatrix}.$$

This is the augmented matrix of an *inconsistent* system: rank $\begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & | & \mathbf{c}_4\end{bmatrix} > \text{rank}\begin{bmatrix}\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\end{bmatrix}$ because of the third leading one in column 4 of $\tilde{A}$. So indeed,

$$\mathbf{c}_4 \notin \text{span}\{\mathbf{c}_1, \mathbf{c}_2\}.$$

So we know that $\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_4\}$ is independent, and that span$\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_4\} = \text{span}\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4\} = \text{Col}(A)$. So we're done! $\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_4\}$ is a basis for Col(A).

[a] You may ask: "Why do this?" – "why put $\mathbf{c}_3$ back in?" Because we can read the answer directly from $\tilde{A}$. Read on.

! Look carefully at this example: we kept only those columns of A where there *were leading ones in* $\tilde{A}$.

This always works. Keeping in mind that the columns where the leading ones sit don't change when we go from REF to RREF, we can state the theorem as:

Theorem 16.2.2 — Basis for $\mathrm{Col}(A)$: the column space algorithm. Let A be an $m \times n$ matrix. Then a basis for $\mathrm{Col}(A)$ consists of those columns of A which give rise to leading ones in an REF of A. Hence $\dim \mathrm{Col}(A) = \mathrm{rank} A$.

Proof. Suppose $A = [\mathbf{u}_1\ \mathbf{u}_2 \cdots \mathbf{u}_n]$ is written in block column form. To choose a subset of the columns of A which are a basis for A, we proceed as follows.

- Start with $\{\mathbf{u}_1\}$. If it's zero, then throw it away; it can't be part of a basis. Otherwise, keep it.
- Next consider $\{\mathbf{u}_1, \mathbf{u}_2\}$. If $\mathbf{u}_2 \in \mathrm{span}\{\mathbf{u}_1\}$, then the set $\{\mathbf{u}_1, \mathbf{u}_2\}$ is LD and $\mathrm{span}\{\mathbf{u}_1, \mathbf{u}_2\}$ is just $\mathrm{span}\{\mathbf{u}_1\}$ — so discard $\mathbf{u}_2$. Otherwise, keep it.
- Next consider $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$. If $\mathbf{u}_3 \in \mathrm{span}\{\mathbf{u}_1, \mathbf{u}_2\}$, then as above it is redundant and we throw it away; otherwise keep it.
- Continue in this way, deciding at each step whether to keep the vector or throw it out, until you've covered all the vectors.

Now: what we're left with will be linearly independent and will span the column space. Why? We checked linear independence at each step; and we only removed vectors which were linear combinations of ones we kept, so the span didn't change.

Finally: how do you decide which to keep and which to discard? Say we want to decide if $\mathbf{u}_2$ is in $\mathrm{span}\{\mathbf{u}_1\}$. We do this by row reducing the augmented matrix $[\mathbf{u}_1|\mathbf{u}_2]$. But these are just the first two columns of A; so we don't actually have to row reduce again — just look at the first two columns of the RREF of A. If the second column has a leading 1, then the system is inconsistent, and so we deduce NO, $\mathbf{u}_2 \notin \mathrm{span}\{\mathbf{u}_1\}$. Otherwise, there is no leading 1, so the system is consistent, which means YES $\mathbf{u}_2 \in \mathrm{span}\{\mathbf{u}_1\}$ and should be thrown out.

This works at every step: we want to decide if $\mathbf{u}_k \in \mathrm{span}\{\mathbf{u}_1, \cdots, \mathbf{u}_{k-1}\}$, so consider the RREF of the matrix

$$[\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_{k-1} \mid \mathbf{u}_k]$$

But since this is just the first k columns of A, we already know its RREF (the first k columns of the RREF of A). The system is consistent (meaning: we should throw out $\mathbf{u}_k$) exactly when the rank of the coefficient matrix equals the rank of the augmented matrix; in other words, exactly when there is NO leading 1 in the kth column. The system is inconsistent (meaning: we should keep $\mathbf{u}_k$) exactly when the rank of the coefficient matrix is smaller than the rank of the augmented matrix, which happens exactly when there IS a leading 1 in the kth column.

We conclude: we will throw out (eliminate, discard) $\mathbf{u}_i$ exactly if there is NO LEADING 1 in column i of the RREF of A. In other words: the vectors we keep are the ones which give rise to leading ones. ■

Corollary 16.2.3 For any matrix A with transpose matrix A^T:

$$\dim \mathrm{Row}(A) = \dim \mathrm{Col}(A) = \dim \mathrm{Col}(A^T) = \mathrm{rank}(A^T) = \dim \mathrm{Row}(A^T),$$

and all are equal to $\mathrm{rank}(A)$.

Proof. Remember: the rank is the number of leading ones in the RREF of A, and $\mathrm{Row}(A) = \mathrm{Col}(A^T)$. Now use Theorems 16.1.3 and 16.2.2 and you have it! ■

This is **amazing fact number one** for first year linear algebra: just think about it. No matter what size the matrix, the maximum number of independent rows — $\dim \mathrm{Row}(A)$ — is *exactly the same as* the maximum number of independent columns — $\dim \mathrm{Col}(A)$! This is not obvious at all ... but it is true.

■ **Example 16.2.4** Given that the matrix

$$A = \begin{bmatrix} 2 & 1 & 9 & 3 & 7 \\ 3 & -1 & 11 & 1 & 2 \\ 1 & 1 & 5 & 1 & 4 \end{bmatrix}$$

has RREF

$$\tilde{A} = \begin{bmatrix} \textcircled{1} & 0 & 4 & 0 & 1 \\ 0 & \textcircled{1} & 1 & 0 & 2 \\ 0 & 0 & 0 & \textcircled{1} & 1 \end{bmatrix}$$

we deduce that a basis for $\mathrm{Col}(A)$[3] is

$$\left\{ \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

■

! WARNING: The leading columns of the RREF of A usually do NOT form a basis for $\mathrm{Col}(A)$ – you have to go back to the original matrix A.

■ **Example 16.2.5** If $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \end{bmatrix}$ then the RREF of A is $\tilde{A} = \begin{bmatrix} \textcircled{1} & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. Clearly[4] $\{(1,2,1)\}$ is a basis for $\mathrm{Col}(A)$, whereas the first column of $\tilde{A}$, $(1,0,0)$, is not even *in* $\mathrm{Col}(A)$. (But $\{(1,0,0)\}$ is a basis for $\mathrm{Col}(\tilde{A})$.) ■

We can now solve our second problem from the beginning of the chapter. Given a spanning set for a subspace U from which we wish to extract a basis, assemble a matrix A with the given vectors as *columns*, so $U = \mathrm{Col}(A)$, and then use the column space algorithm to find a basis of $\mathrm{Col}(A) = U$: it will consist of some of the columns of A, and so it will indeed consist of some of the given spanning vectors.

Problem 16.2.6 Let $U = \mathrm{span}\{(1,2,1,0),(2,-2,1,2),(2,-8,0,4),(3,4,1,1)\}$. Find a basis for U which is a subset of the given spanning set.

Solution Let's assemble a matrix A whose *columns* are the given vectors. So

$$A = \begin{bmatrix} 1 & 2 & 2 & 3 \\ 2 & -2 & -8 & 4 \\ 1 & 1 & 0 & 1 \\ 0 & 2 & 4 & 1 \end{bmatrix}.$$

[3] This means that $\mathrm{Col}(A) = \mathbb{R}^3$!

[4] If you prefer to write this vector as a column vector, feel free to do so.

We know from Problem 16.1.2 that the RREF of A is

$$\tilde{A} = \begin{bmatrix} \textcircled{1} & 0 & -2 & 0 \\ 0 & \textcircled{1} & 2 & 0 \\ 0 & 0 & 0 & \textcircled{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The leading ones occur in columns 1, 2 and 4, so a basis for $\operatorname{Col}(A) = U$ is

$$\{(1,2,1,0),(2,-2,1,2),(3,4,1,1)\}.$$

16.3 Summary: two methods to obtain a basis

So we have two distinct ways to get a basis from a spanning set. Say $W = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$. Then either:

1. Write the vectors as the rows of a matrix $B = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \cdots \\ \mathbf{v}_k \end{bmatrix}$. So $W = \operatorname{Row}(B)$. Row reduce, and take the nonzero rows of the RREF for your basis.
2. Write the vectors as the columns of a matrix $A = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_k]$. So $W = \operatorname{Col}(A)$. Row reduce, and take the column vectors in A which give rise to leading ones in the RREF.

Use the first if you want a nice basis for W that's easy to work with. Use the second if the vectors you have in hand are special to you and you just want to throw out the redundant ones.

The point: this works even if the subspace of which you're trying to get a basis didn't start off its life as the column space or row space of a matrix. Anytime you have a spanning set, you can apply this method. (And if your vectors are things like polynomials or matrices or functions: write them in coordinates relative to a 'standard' basis, then apply these methods.)

Remark 16.3.1. *Before we leave this for now, let's make a final observation regarding the nullspace of a matrix A and its rowspace. It's a kind of 'geometric interpretation' of the rank-nullity theorem, Corollary 15.2.3:*

$$\dim \operatorname{Null}(A) + \operatorname{rank}(A) = n$$

We can now write this as

$$\dim \operatorname{Null}(A) + \dim \operatorname{Row}(A) = n$$

Write $A = \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_m \end{bmatrix}$ *in block row form. Then*

$$\begin{aligned} \mathbf{v} \in \operatorname{Null} A &\iff A\mathbf{v} = 0 \\ &\iff \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_m \end{bmatrix} \mathbf{v} = 0 \\ &\iff \begin{bmatrix} \mathbf{r}_1 \cdot \mathbf{v} \\ \vdots \\ \mathbf{r}_m \cdot \mathbf{v} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \end{aligned}$$

That is, $\mathbf{v} \in \text{Null}(A)$ *iff* $\mathbf{v}$ *is orthogonal to each and every row of A. So* $\mathbf{v}$ *will be orthogonal to any and all linear combinations of rows of A, and so indeed* $\mathbf{v} \in \text{Null}(A)$ *iff* $\mathbf{v}$ *is orthogonal to each and every element of* $\text{Row}(A)$*! We'll come back to this idea in chapter 20.*

To end the chapter, we now present another way to see that the column space algorithm works.

Problem 16.3.2 Find a basis for $\text{Col}(A)$, where

$$A = \begin{bmatrix} 1 & 2 & 2 & 3 \\ 2 & -2 & -8 & 4 \\ 1 & 1 & 0 & 1 \\ 0 & 2 & 4 & 1 \end{bmatrix}$$

Solution By definition,

$$\text{Col}(A) = \text{span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ -8 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 1 \\ 1 \end{bmatrix} \right\}$$

The point of this problem is: we don't know if this spanning set is linearly independent or not.

But we just saw that to check for linear independence, we need to row reduce A. So let's do that; we get that the RREF of A is

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

By Theorem 15.5.1, the fact that $A\mathbf{x} = 0$ does not have a unique solution implies that columns are not linearly independent, and do not form a basis for $\text{Col}(A)$. But we have even more information than that: we know *all* the possible dependence relations on the set of columns!

Let's reason this out. First:

$$\begin{aligned} \text{Null}(A) &= \{\mathbf{x} = (a_1, a_2, a_3, a_4) \mid A\mathbf{x} = \mathbf{0}\} \\ &= \{(a_1, a_2, a_3, a_4) \mid a_1\mathbf{c}_1 + a_2\mathbf{c}_2 + a_3\mathbf{c}_3 + a_4\mathbf{c}_4 = \mathbf{0}\} \end{aligned}$$

Now given the above RREF, we know that

$$\text{Null}(A) = \{(2r, -2r, r, 0) \mid r \in \mathbb{R}\}$$

In other words, ALL solutions to the dependence equation on the columns of $A = [\mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_3\ \mathbf{c}_4]$ have the form

$$(2r)\mathbf{c}_1 + (-2r)\mathbf{c}_2 + r\mathbf{c}_3 + 0\mathbf{c}_4 = \mathbf{0}.$$

In particular, set $r = 1$ and solve for $\mathbf{c}_3$ to get

$$\mathbf{c}_3 = -2\mathbf{c}_1 + 2\mathbf{c}_2$$

which implies that $\mathbf{c}_3 \in \text{span}\{\mathbf{c}_1, \mathbf{c}_2\}$. (Check! $-2\begin{bmatrix}1\\2\\1\\0\end{bmatrix} + 2\begin{bmatrix}2\\-2\\1\\2\end{bmatrix} = \begin{bmatrix}2\\-8\\0\\4\end{bmatrix}$ – yes!)

By what we talked about earlier in the course, this means:

$$\text{Col}\,(A) = \text{span}\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_4\}$$

Are these linearly independent? Yes! Two ways to see this:

1. We know ALL dependence relations on these 4 vectors. If these 3 remaining vectors were linearly dependent, we could find a nontrivial equation like

$$b_1\mathbf{c}_1 + b_2\mathbf{c}_2 + b_4\mathbf{c}_4 = \mathbf{0}$$

   but then $(b_1, b_2, 0, b_4)$ would lie in Null(A), and the only vector in Null(A) with third component 0 is the zero vector. Hence no nontrivial relation exists.
2. Alternately: imagine checking that these are linearly independent. You put them as the columns of a matrix and row reduce. But all we're doing is covering up column 3 of the matrix A as we row reduce, and all the steps would be the same — because the elements in a non-leading column *are never involved in steps that change the matrix during row reduction*. That means row reducing the matrix with just these 3 columns gives exactly the same answer as row reducing A and just crossing out or ignoring column 3 at every step. The result: each of these 3 columns gave a leading 1, and thus they are linearly independent.

So we have found a basis for Col$\,(A)$:

$$\left\{\begin{bmatrix}1\\2\\1\\0\end{bmatrix}, \begin{bmatrix}2\\-2\\1\\2\end{bmatrix}, \begin{bmatrix}3\\4\\1\\1\end{bmatrix}\right\}$$

Theorem 16.3.3 — Basis for Col$\,(A)$- a reprise. Let A be an $m \times n$ matrix. Then a basis for Col$\,(A)$ consists of those columns of A which give rise to leading ones in an REF of A.

Idea of proof. One has to see that the example generalizes to any number of non-leading variables.

Let the columns of A be $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$. Any nontrivial solution to

$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$$

is given by a vector $(a_1, \cdots, a_n) \in$ Null(A). In particular, each basic solution gives one dependence relation on these vectors. Now recall that in each basic solution, all but the last nonzero entry comes from the equation for a leading variable, and the last nonzero entry is a 1 and corresponds to a non-leading variable.

That means: the nontrivial dependence relation has a column vector corresponding to a non-leading variable with coefficient 1, so we can solve for it in terms of the other column vectors occuring in this basic solution.

In other words: each basic solution lets us solve for one vector (corresponding to a non-leading variable) in terms of vectors which correspond to leading variables. Hence every vector which corresponds to a non-leading variable is in the span of those corresponding to leading variables, which means the vectors corresponding to non-leading variables can be discarded.

Finally, we note that the remaining vectors cannot have any nontrivial dependence relation on them, since every nontrivial dependence relation involves one of the non-leading variables. So they are linearly independent and span $\text{Col}(A)$, so form a basis. ■

Problems

Remarks:

1. A question with an asterisk '*' (or two) indicates a bonus-level question.
2. You must justify all your responses.

Problem 16.1 Find a basis for the row space and column space for each of the matrices below, and check that $\dim \text{Row}\, A = \dim \text{Col}\, A$ in each case.

a) $\begin{bmatrix} 1 & 1 & 0 & 7 \\ 0 & 0 & 1 & 5 \\ 0 & 1 & 1 & 9 \end{bmatrix}$

b)* $\begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 2 \end{bmatrix}$

c) $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 0 & 3 & 3 \end{bmatrix}$

d)* $\begin{bmatrix} 1 & 2 & -1 & -1 \\ 2 & 4 & -1 & 3 \\ -3 & -6 & 1 & -7 \end{bmatrix}$

e) $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 0 \\ 1 & 1 & 2 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix}$

f)* $\begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 4 & 3 \\ 1 & 0 & 2 & 1 \end{bmatrix}$

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

Problem 16.2 Find a basis of the desired type for the given subspace in each case. (Use your work from the previous question where useful.)

a) $U = \text{span}\{(1,1,0),(2,0,3),(3,1,3)\}$: the basis must be a subset of $\{(1,1,0),(2,0,3),(3,1,3)\}$.

b) $^\star W = \text{span}\{(1,2,-1,-1),(2,4,-1,3),(-3,-6,1,-7)\}$: any basis suffices.

c) $X = \text{span}\{(1,0,1,1),(1,1,1,1),(1,2,2,0),(1,0,0,2)\}$: the basis must be a subset of the given spanning set.

d) $^\star Y = \text{span}\{(1,0,1,1),(-1,1,2,0),(1,1,4,2),(0,1,3,1)\}$: the basis must be a subset of the given spanning set.

e) $V = \text{span}\{(1,0,1,3),(1,2,-1,-1),(0,1,-2,5)\}$: any basis suffices.

Problem 16.3 Extend each of the bases you obtained in Problem 16.2 to a basis of $\mathbb{R}^n$, where the vectors are from $n=3$ for part (a) and $n=4$ for each of the other parts. (Solutions are given for parts (b)*and (d)*.)

Problem 16.4 State whether each of the following is (always) true, or is (possibly) false.

- If you say the statement may be false, you must give an explicit example.
- If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

a) For all matrices A, $\dim \text{Row}(A) + \dim \ker(A) = \dim \text{Col}(A)$.

b) *For some matrices A, $\dim \text{Row}(A) + \dim \ker(A) = \dim \text{Col}(A)$

c) For all matrices A, $\dim \text{Row}(A) + \dim \ker(A) = m$, where m is the number of rows of A.

d) *For all matrices A, $\dim \text{Row}(A) + \dim \ker(A) = n$, where n is the number of columns of A.

e) For all $m \times n$ matrices A, $\dim\{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\} + \dim\{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\} = \text{rank} A$.

f) *For all $m \times n$ matrices A, $\dim\{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\} + \dim\{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = 0\} = m$.

g) For all $m \times n$ matrices A, $\dim\{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\} + \dim\{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\} = n$.

h) *For all $m \times n$ matrices A, the dot product of every vector in $\ker(A)$ with any of the rows of A is zero.

i) * If A and B are two matrices such that AB is defined, then $\dim \text{Row}(AB) \leq \dim \text{Row}(B)$.

j) * If A and B are two matrices such that AB is defined, then $\dim \text{Col}(AB) \leq \dim \text{Col}(A)$.

k) * If A and B are two matrices such that AB is defined, then $\text{rank}(AB) \leq \text{rank} A$, and $\text{rank}(AB) \leq \text{rank} B$.

17. Bases for finite dimensional vector spaces

Photo: Ralph Nevins. La dune du Pilat, France

Let A be an $m \times n$ matrix. We have now established techniques for finding bases of

- $\text{Null}(A)$, the nullspace of A, which is a subspace of $\mathbb{R}^n$;
- $\text{Col}\,(A)$, the column space of A, which is a subspace of $\mathbb{R}^m$;
- $\text{Row}\,(A)$, the row space of A, which is a subspace of $\mathbb{R}^n$.

Along the way, we determined that

$$\dim(\text{Null}(A)) + \text{rank}(A) = n;$$

this is called the rank-nullity theorem. But since

$$\text{rank}(A) = \dim(\text{Col}\,(A)) = \dim(\text{Row}\,(A)),$$

we could equally rephrase this as

$$\dim(\text{Null}(A)) + \dim(\text{Col}\,(A)) = n,$$

which expresses a "conservation of dimension" of matrix multiplication, since $\text{Null}(A) = \{\mathbf{x} \mid A\mathbf{x} = \mathbf{0}\}$ and $\text{Col}\,(A) = \text{im}\,(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$; We can therefor write this equation as

$$\dim(\text{Null}(A)) + \dim(\text{Row}\,(A)) = n.$$

which is particularly interesting in light of the fact (check!) that every vector in the rowspace of A is orthogonal to every vector in the nullspace of A — so these spaces are *orthogonal complements*, coming up soon in Chapter 20.

17.1 Finding bases for general vector spaces

We mentioned last time that the algorithms for finding bases for $\mathrm{Col}(A)$ and $\mathrm{Row}(A)$ can be used to find a basis for any subspace W of $\mathbb{R}^n$, if we start with a spanning set of W. But in fact it works for any subspace of any vector space.

Problem 17.1.1 Find a basis for the subspace W of $\mathbb{P}_3$ spanned by

$$\{3+x+4x^2+2x^3, 2+4x+6x^2+8x^3, 1+3x+4x^2+6x^3, -1+2x+x^2+4x^3\}$$

Solution Begin by writing down these vectors in coordinates relative to the standard basis $\{1,x,x^2,x^3\}$:

$$\mathbf{u}_1 = \begin{bmatrix}3\\1\\4\\2\end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix}2\\4\\6\\8\end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix}1\\3\\4\\6\end{bmatrix}, \mathbf{u}_4 = \begin{bmatrix}-1\\2\\1\\4\end{bmatrix}.$$

By our previous discussion about coordinates, we see that if we are able to find a basis for

$$U = \mathrm{span}\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3,\mathbf{u}_4\}$$

(which is a subspace of $\mathbb{R}^4$), then this basis will be the set of coordinate vectors of a basis for W.

Let's use the rowspace algorithm, so that our result is a nice basis.

$$A = \begin{bmatrix}3 & 1 & 4 & 2\\ 2 & 4 & 6 & 8\\ 1 & 3 & 4 & 6\\ -1 & 2 & 1 & 4\end{bmatrix} \sim \cdots \sim R = \begin{bmatrix}1 & 0 & 1 & 0\\ 0 & 1 & 1 & 2\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}$$

By our algorithm last time, we conclude that $\{(1,0,1,0),(0,1,1,2)\}$ is a basis for U.

So, going back to the polynomial expressions, we deduce that $\{1+x^2, x+x^2+2x^3\}$ is a basis for W.

Check! These polynomials are certainly linearly independent; we next need to check that each of the four polynomials above is a linear combination of these two. But this isn't hard: look at the coefficients of 1 and x:

$$a(1+x^2)+b(x+x^2+2x^3) = a+bx+(a+b)x^2+2bx^3.$$

So it's easy to see when something is (or isn't) a linear combination of these; we don't need to do any work to find a and b. (Now do the check.)

17.2 Enlarging linearly independent sets to bases

Another problem that we have considered before: what if you have a linearly independent set, and you want to extend it to a basis for your space? Where this arises most often is the following type of example, so we give a solution in that case.

Problem 17.2.1 The set $\{(1,2,3,1),(1,2,3,2)\}$ is a basis for a subspace of $\mathbb{R}^4$. Extend this to a basis for $\mathbb{R}^4$.

Solution We need to find vectors that are not linear combinations of these. There are lots of techniques we could try (recall that even random guessing works well!) but here's a nice useful one.

Put the vectors in the rows of matrix.

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

and reduce to REF:

$$R = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

We see that we are missing pivots (leading ones) in the second and third columns. So the two vectors we will add to our set are

$$(0,1,0,0) \quad \text{and} \quad (0,0,1,0).$$

Let's see if $\{(1,2,3,1),(1,2,3,2),(0,1,0,0),(0,0,1,0)\}$ spans $\mathbb{R}^4$ (and thus a basis for $\mathbb{R}^4$ since these are 4 vectors and $\dim(\mathbb{R}^4) = 4$):

$$B = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 1 & 2 & 3 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 3 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

So there's a leading 1 in every row, meaning $\dim(\text{Row}(B)) = 4$. Hence the rows of B are a basis for $\mathbb{R}^4$.

The point: we are just exactly adding the missing leading 1s!

So: row reduction (surprisingly!) gave us some easy techniques for getting bases of subspaces — better techniques than what we'd decided on theoretically a few weeks ago! Moreover, by working in coordinates relative to a standard basis, all these techniques apply to any subspace of any vector space (by turning all vectors into elements of $\mathbb{R}^n$ for some n).

17.3 More about bases

We've spent a lot of time talking about the row space and column space of a matrix, and have characterizations of those matrices that are great in the sense that every consistent system has a unique solution (equivalently, its columns are linearly independent); and characterizations of matrices that are great in the sense that every system is consistent (equivalently, that its columns span $\mathbb{R}^m$). What can we say about a matrix which has BOTH of these great properties?

Well, right off the bat: it has to be square. (Look back at the characterizations to convince yourself of this — compare the ranks.) So the following theorem is about square matrices.

Theorem 17.3.1 Let A be an $n \times n$ matrix. Then the following are all equivalent:
(1) $\text{rank}(A) = n$.
(2) $\text{rank}(A^T) = n$.
(3) Every linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution.
(4) The RREF of A is I.
(5) $\text{Null}(A) = \{\mathbf{0}\}$.
(6) $\text{Col}(A) = \mathbb{R}^n$.

(7) $\text{Row}(A) = \mathbb{R}^n$.
(8) The columns of A are linearly independent.
(9) The rows of A are linearly independent.
(10) The columns of A form a basis for $\mathbb{R}^n$.
(11) The rows of A form a basis for $\mathbb{R}^n$.

Proof. We just have to put our two theorems from before together. The nice statement (3) about getting consistency and uniqueness together. Statement (4) is just pointing out that the RREF of a square matrix with a leading one in every row and column is just exactly the identity matrix. The last two are just using the definition of a basis. ■

So matrices which satisfy the conditions of this theorem are quite special: our next goal is to show that they are beyond special, they are fantastic: they are *invertible*.

Problems

Problem 17.1 Extend the given linearly independent set of $\mathbb{R}^n$ to a basis of $\mathbb{R}^n$. (Use your work from Problem 16.1 where useful.)

a) $\{(1,1,0,7),(0,0,1,5),(0,1,1,9)\}$ $(\mathbb{R}^4)$

b) $^\star\{(1,0,1),(0,1,1)\}$ $(\mathbb{R}^3)$

c) $\{(1,0,1,3),(1,2,-1,-1),(0,1,-2,5)\}$ $(\mathbb{R}^4)$

d) $^\star\{(1,0,1,3)\}$ $(\mathbb{R}^4)$

18. Matrix Inverses

Photo: Ralph Nevins. La dune du Pilat, France

Throughout this chapter, A is assumed to be a square $n \times n$ matrix.

18.1 Matrix multiplication: definition of an inverse

We've seen previously that there is profit in looking at an equation like $A\mathbf{x} = \mathbf{b}$ in several different ways. By looking at the entries, it becomes a linear system; by looking at the column vectors, it becomes a vector equation. But what about looking at it just as an algebraic expression, that is, comparing

$$A\mathbf{x} = \mathbf{b} \quad \text{with} \quad ax = b?$$

The nice thing about a scalar equation like $ax = b$ ($a, x, b \in \mathbb{R}$) is that if $a \neq 0$, then $x = b/a$. Can we "divide" by a matrix A to get the same kind of thing?

Hmmm: a major obstacle here: it is hard enough to multiply by matrices; how can you "undo" that operation?

Well, let's rephrase things a bit: we can think of $x = b/a$ as $x = a^{-1}b$; that is, multiply by the inverse, instead of divide. So now our goal is just to find out what "A^{-1}" is. That is, we are looking for a matrix A^{-1} so that whenever we have

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

we can deduce

$$\mathbf{x} = A^{-1}\mathbf{b}.$$

For real numbers, the property of the inverse is that $a^{-1}a = 1$; so that's what we're looking for here. So our first definition is just going to say what we're looking for in an inverse — but it won't tell us how to find it! We still have to work that part out.

Recall that I_n is the $n \times n$ identity matrix; the square matrix with 1s down the diagonal and 0s everywhere else. The key property of the identity matrix is that it acts just like the number 1 in terms of matrix multiplication: $BI = B$ and $IB = B$.

Definition 18.1.1 If A is an $n \times n$ matrix and B is an $n \times n$ matrix such that

$$AB = BA = I$$

then B is called an *inverse* of A and we write $B = A^{-1}$. In this case, A is called *invertible*.

■ **Example 18.1.2** Consider I_n. Then $I_n^{-1} = I_n$ since $I_n I_n = I_n$. (Good: $1^{-1} = 1$ in $\mathbb{R}$.) ■

■ **Example 18.1.3** Let $A = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & -1 \\ -2 & 1 \end{bmatrix}$. Then

$$AB = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$$

and

$$BA = \begin{bmatrix} 3 & -1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} = I_2$$

so this proves that $B = A^{-1}$ — and that $A = B^{-1}$. (Good: if $b = 1/a$ then $a = 1/b$ in $\mathbb{R}$.) ■

18.2 Finding the inverse of a 2×2 matrix

Lemma 18.2.1 — Inverse of a 2×2 matrix. Suppose $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then if $ad - bc \neq 0$,

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

If $ad - bc = 0$, then A is not invertible.

Idea of proof. Write B for this alleged inverse of A. By multiplying out, you can see that $AB = BA = I$, so in fact yes, $B = A^{-1}$.

Showing that if $ad - bc = 0$ then A is not invertible is more work (exercise); you can generalize the following example to get the proof.

(You could have found this formula from that surprise equation we had a while back (Cayley-Hamilton theorem): $A^2 - tr(A)A + \det(A)I = 0$.) ■

■ **Example 18.2.2** The matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ is NOT INVERTIBLE. We can show this in a couple of ways. (Certainly $ad - bc = 1 - 1 = 0$, so the lemma says that A should not be invertible.)

1. First way: directly. Suppose $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ were an inverse of A. Then

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a+c & b+d \\ a+c & b+d \end{bmatrix} \neq \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

for any choice of $a, b, c, d \in \mathbb{R}$. So no inverse exists.

2. Second way: let $\mathbf{x} = (1, -1)$. Then

$$A\mathbf{x} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathbf{0}.$$

Now, suppose to the contrary that A^{-1} were to exist. Then we could multiply the equation $A\mathbf{x} = \mathbf{0}$ on the left by A^{-1} to get

$$A^{-1}A\mathbf{x} = A^{-1}\mathbf{0}.$$

The left side simplifies to $I_2\mathbf{x}$, which in turn simplifies to $\mathbf{x}$; whereas the right side simplifies to $\mathbf{0}$. But then that says $\mathbf{x} = \mathbf{0}$, which is FALSE. So our assumption (that A^{-1} existed) must have been wrong; we deduce that A is not invertible.

■

This example gives us our first glimpse of the connection with those extra-wonderful matrices from before: If A is invertible, then $A\mathbf{x} = \mathbf{0}$ has a unique solution.

We have deduced:

Lemma 18.2.3 — Inverses and solutions of linear systems. Suppose A is an invertible $n \times n$ matrix. Then any linear system $A\mathbf{x} = \mathbf{b}$

(a) is consistent, and

(b) has a unique solution.

Proof. Note that $A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I_n\mathbf{b} = \mathbf{b}$, so $\mathbf{x} = A^{-1}\mathbf{b}$ is a solution to the system; thus the system is consistent.

If $\mathbf{y}$ is some other solution, meaning $A\mathbf{y} = \mathbf{b}$. Multiply on the left by A^{-1}; this gives

$$A^{-1}A\mathbf{y} = A^{-1}\mathbf{b}$$

and again we deduce (from $A^{-1}A = I_n$ this time) that $\mathbf{y} = A^{-1}\mathbf{b}$. So the solution is unique. ■

In particular, this means A satisfies all the criteria of the theorem of last time, because this shows that every linear system with coefficient matrix A has a unique solution.

However, do remember that all this is contingent on A being invertible; and certainly not all square matrices are invertible.

18.3 Algebraic properties of inverses

Proposition 18.3.1 — Properties of the inverse. If $k \neq 0$ is a scalar, p is an integer and A and C are invertible $n \times n$ matrices, then so are A^{-1}, A^p, A^T, kA, and AC. In fact, we have

(1) $(A^{-1})^{-1} = A$;

(2) $(A^p)^{-1} = (A^{-1})^p$;

(3) $(A^T)^{-1} = (A^{-1})^T$;

(4) $(kA)^{-1} = \frac{1}{k}A^{-1}$; and

(5) $(AC)^{-1} = C^{-1}A^{-1}$ — NOTE ORDER.

Furthermore, if AC is invertible, then so are A and C.

Partial proof. Remember that to show a matrix is invertible, you need to find its inverse. Here, we've already given you the inverse, so all that you have to do is show that the matrix times the alleged inverse equals the identity.

So for example, $(A^t)(A^{-1})^T = (A^{-1}A)^T$ (by properties of products and transposes) and this is I^T (since A is invertible) which is I. Same argument for $(A^{-1})^T(A^T) = I$. So A^T is invertible and its inverse is the transpose of the inverse of A.

Also: note that $(AC)(C^{-1}A^{-1}) = A(CC^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$; so we have the correct formula for the inverse of AC. You see why we had to reverse the order to make this work!

For the last statement: just note that $A^{-1} = C(AC)^{-1}$ (by multiplying by A) and $C^{-1} = (AC)^{-1}A$. ■

Problem 18.3.2 Simplify $(A^TB)^{-1}A^T$.
Solution $(A^TB)^{-1}A^T = (B^{-1}(A^T)^{-1})A^T = B^{-1}((A^T)^{-1})A^T) = B^{-1}$

Problem 18.3.3 Simplify $(A+B)^{-1}$.
Solution TOUGH LUCK. This can't be simplified, even when A and B are numbers. (Think of $(2+3)^{-1} \neq \frac{1}{2} + \frac{1}{3}$.) You can't even expect $A+B$ to be invertible; for instance, suppose $A = -B$!

18.4 Finding the inverse: general case

We saw last time that if $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ then A is invertible if and only if $\det(A) = ad - bc \neq 0$, in which case

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

(And if A is a 1×1 matrix, then A is just a real number like $A = [a]$, so $A^{-1} = [\frac{1}{a}]$, if $a \neq 0$.)

There exist formulae for $n \times n$ inverses, but they're not very efficient (as we'll eventually see); instead of looking for a formula, let's see how what we know about inverses has actually given us all we need to be able to calculate them.

Remember that to find the inverse of A, if it exists, we want to solve

$$AB = I_n$$

for the unknown matrix B (which will be A^{-1}, if all goes well).

Write B and I as collections of column vectors: $B = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n]$ and $I_n = [\mathbf{e}_1\ \mathbf{e}_2\ \cdots\ \mathbf{e}_n]$. Note that $\mathbf{e}_i$ is the ith standard basis vector of $\mathbb{R}^n$.

Now multiply:

$$AB = [A\mathbf{v}_1\ A\mathbf{v}_2\ \cdots\ A\mathbf{v}_n] = [\mathbf{e}_1\ \mathbf{e}_2\ \cdots\ \mathbf{e}_n] = I_n$$

which gives a bunch of matrix equations to solve:

$$A\mathbf{v}_1 = \mathbf{e}_1,\ A\mathbf{v}_2 = \mathbf{e}_2,\ \cdots,\ A\mathbf{v}_n = \mathbf{e}_n.$$

The theorem says that if A is invertible, then each of these matrix equations will be consistent and have a unique solution. And those solutions are exactly the columns of B!

Great! To get the inverse of A, we have to row reduce each of the augmented matrices

$$[A|\mathbf{e}_1], [A|\mathbf{e}_2] \cdots, [A|\mathbf{e}_n];$$

and since each of these will give a unique solution, the RREFs will look like

$$[I|\mathbf{v}_1], [I|\mathbf{v}_2], \cdots, [I|\mathbf{v}_n].$$

The final moment of brilliance: We would do EXACTLY THE SAME row operations for each of these linear systems, so why don't we just augment $[A|\mathbf{e}_1\mathbf{e}_2\cdots\mathbf{e}_n] = [A|I]$? Then the RREF will be $[I|B] = [I|A^{-1}]$.

Problem 18.4.1 Use this method to find the inverse of

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 5 \\ 2 & 2 & 5 \end{bmatrix}$$

Solution So we're trying to solve $AB = I_3$ for B. Write $B = [\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3]$. Then

$$AB = [A\mathbf{v}_1\ A\mathbf{v}_2\ A\mathbf{v}_3]$$

and we want this to be equal to

$$I_3 = [\mathbf{e}_1\ \mathbf{e}_2\ \mathbf{e}_3]$$

To solve the three linear systems $A\mathbf{v}_i = \mathbf{e}_i$ at once, we set up a super-augmented matrix

$$[A|\mathbf{e}_1\ \mathbf{e}_2\ \mathbf{e}_3] = \begin{bmatrix} 1 & 1 & 2 & | & 1 & 0 & 0 \\ 1 & 2 & 5 & | & 0 & 1 & 0 \\ 2 & 2 & 5 & | & 0 & 0 & 1 \end{bmatrix}$$

We will row reduce this and read off the matrix B from the RREF (if $A \sim I_3$).

$$[A|I_3] \sim \begin{bmatrix} 1 & 1 & 2 & | & 1 & 0 & 0 \\ 0 & 1 & 3 & | & -1 & 1 & 0 \\ 0 & 0 & 1 & | & -2 & 0 & 1 \end{bmatrix}$$

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

Oh, wonderful! The coefficient matrix reduced to I_3, which means that every linear system has a unique solution, and furthermore, that solution is now in the corresponding augmented column. That is to say,

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 5 \\ -2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 1 \\ -3 \\ 1 \end{bmatrix}.$$

Let's check! Set

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

and multiply:

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

and

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

YES! This is an inverse of A — and in fact, we've shown it's the only inverse of A. So

$$A^{-1} = \begin{bmatrix} 0 & -1 & 1 \\ 5 & 1 & -3 \\ -2 & 0 & 1 \end{bmatrix}$$

Now note that this method tells you exactly when A is invertible! For suppose A is $n \times n$.

If A does NOT row reduce to the identity matrix in RREF, then $\text{rank}(A) < n$ and then we know that $A\mathbf{x} = \mathbf{b}$ won't have a unique solution, so in particular, A can't be invertible.

On the other hand, if A DOES row reduce to I_n, then this method gives a matrix B such that $AB = I_n$. BUT WAIT! How do we know that this matrix also satisfies $BA = I_n$?

Lemma 18.4.2 — One-sided inverses are two-sided inverses. If A and B are $n \times n$ matrices such that $AB = I_n$, then $BA = I_n$.

Proof. First let's show that $\text{rank}(B) = n$. Namely, suppose $B\mathbf{x} = \mathbf{0}$. Then

$$A(B\mathbf{x}) = A\mathbf{0} = \mathbf{0}$$
$$(AB)\mathbf{x} = I\mathbf{x} = \mathbf{x}$$

but these must be equal, so $\mathbf{x} = \mathbf{0}$. Thus we deduce that $\text{Null}(B) = \{\mathbf{0}\}$ and so $\text{rank}(B) = n$.

OK, that means we could apply our algorithm to B: $[B|I] \sim \cdots \sim [I|C]$ and get a matrix C such that $BC = I$.

So we have

$$A(BC) = AI = A$$
$$(AB)C = IC = C$$

and these are equal ($A = C$). So in fact $BA = I$, as required. ■

We have proven the following theorem.

Theorem 18.4.3 — Finding matrix inverses. Suppose A is an $n \times n$ matrix. If the rank of A is n, then A is invertible and A^{-1} can be computed by the algorithm above, that is, by row reducing

$$[A|I] \sim \cdots \sim [I|A^{-1}].$$

Furthermore, if $\text{rank}(A) < n$, then A is *not invertible*.

18.5 Examples

Problem 18.5.1 Find A^{-1}, if it exists, where

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

Solution We see directly that A does not have full rank; there can be no leading one in the second column. Hence A is not invertible.

Problem 18.5.2 Find A^{-1}, if it exists, where

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Solution This is already in REF and we can see that $\text{rank}(A) = 3$. So we proceed with our algorithm

$$[A|I] = \begin{bmatrix} 1 & 2 & 3 & | & 1 & 0 & 0 \\ 0 & 1 & 1 & | & 0 & 1 & 0 \\ 0 & 0 & 1 & | & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & | & 1 & -2 & -3 \\ 0 & 1 & 0 & | & 0 & 1 & -1 \\ 0 & 0 & 1 & | & 0 & 0 & 1 \end{bmatrix}$$

And so $A^{-1} = \begin{bmatrix} 1 & -2 & -3 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$ (CHECK!!).

Problem 18.5.3 Find A^{-1}, if it exists, where

$$A = \begin{bmatrix} 2 & 2 & 0 \\ 1 & 1 & 2 \\ 1 & -1 & 3 \end{bmatrix}$$

Solution We can't tell what $\text{rank}(A)$ is, off-hand, so we simply proceed to the algorithm. (If at any point we were to see than the rank is too low, we would stop and announce that the matrix isn't invertible.)

$$[A|I] = \begin{bmatrix} 2 & 2 & 0 & | & 1 & 0 & 0 \\ 1 & 1 & 2 & | & 0 & 1 & 0 \\ 1 & -1 & 3 & | & 0 & 0 & 1 \end{bmatrix}$$

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

$$\sim \begin{bmatrix} 1 & 1 & 2 & | & 0 & 1 & 0 \\ 0 & -2 & 1 & | & 0 & -1 & 1 \\ 0 & 0 & -4 & | & 1 & -2 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & \frac{5}{2} & | & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & -\frac{1}{2} & | & 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 1 & | & -\frac{1}{4} & \frac{1}{2} & 0 \end{bmatrix}$$

$$\sim \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{5}{8} & -\frac{3}{4} & \frac{1}{2} \\ 0 & 1 & 0 & -\frac{1}{8} & \frac{3}{4} & -\frac{1}{2} \\ 0 & 0 & 1 & -\frac{1}{4} & \frac{1}{2} & 0 \end{array}\right]$$

So

$$A^{-1} = \begin{bmatrix} \frac{5}{8} & -\frac{3}{4} & \frac{1}{2} \\ -\frac{1}{8} & \frac{3}{4} & -\frac{1}{2} \\ -\frac{1}{4} & \frac{1}{2} & 0 \end{bmatrix} = \frac{1}{8}\begin{bmatrix} 5 & -6 & 4 \\ -1 & 6 & -4 \\ -2 & 4 & 0 \end{bmatrix}$$

(check!)

18.6 Summary: another condition for our big theorem

We knew last time that being invertible meant that A had full rank; but this method shows the converse is also true. So our big theorem has become even bigger:

Theorem 18.6.1 — Invertible Matrix Theorem. Let A be an $n \times n$ matrix. Then the following statements are all equivalent (that is, either all true about A or all false about A):

(1) $\operatorname{rank}(A) = n$.
(2) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
(3) $A\mathbf{x} = \mathbf{b}$ is consistent for all $\mathbf{b} \in \mathbb{R}^n$.
(4) Every linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution.
(5) The RREF of A is I.
(6) $\operatorname{Null}(A) = \{0\}$.
(7) $\operatorname{Col}(A) = \mathbb{R}^n$.
(8) $\operatorname{Row}(A) = \mathbb{R}^n$.
(9) $\operatorname{rank}(A^T) = n$
(10) The columns of A are linearly independent.
(11) The rows of A are linearly independent.
(12) The columns of A span $\mathbb{R}^n$.
(13) The rows of A span $\mathbb{R}^n$.
(14) The columns of A form a basis for $\mathbb{R}^n$.
(15) The rows of A form a basis for $\mathbb{R}^n$.
(16) A is invertible.
(17) A^T is invertible.

■ **Example 18.6.2** Every statement of the theorem is true for

$$A = \begin{bmatrix} 2 & 2 & 0 \\ 1 & 1 & 2 \\ 1 & -1 & 3 \end{bmatrix}$$

because we have just seen that it is invertible. ■

■ **Example 18.6.3** Every statement of the theorem is false for

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

(CHECK!). ■

Problems

Problem 18.1 Find the inverse of each of the following matrices, or give reasons if it not invertible.

a) $\begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

b) * $\begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & -2 \\ 0 & 2 & 3 \end{bmatrix}$

c) $\begin{bmatrix} 1 & 2 & 2 \\ 1 & 3 & 1 \\ 1 & 3 & 2 \end{bmatrix}$

d) * $\begin{bmatrix} 1 & x \\ -x & 1 \end{bmatrix}$

e) $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 2 & 3 & 2 \end{bmatrix}$

Problem 18.2 State whether each of the following is (always) true, or is (possibly) false.

- If you say the statement may be false, you must give an explicit example.
- If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

a) If A is invertible and $AB = 0$ then $B = 0$.

b) *If $A^2 = 0$ for a square matrix A, then A is not invertible.

c) If $A^2 = 0$ for an $n \times n$ matrix A, then $\operatorname{rank} A < n$.

d) *If A is invertible then the RREF form of A has a row of zeros.

e) If A is an invertible $n \times n$ matrix then $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b} \in \mathbb{R}^n$.

f) *If A is a non-invertible $n \times n$ matrix then $A\mathbf{x} = \mathbf{b}$ is inconsistent for every $\mathbf{b} \in \mathbb{R}^n$.

g) If A is a non-invertible $n \times n$ matrix then $A\mathbf{x} = \mathbf{0}$ has a unique solution.

h) *If an $n \times n$ matrix A satisfies $A^3 - 3A^2 + I_n = 0$, then A is invertible and $A^{-1} = 3A - A^2$.

IV Orthogonality

We now return to some geometric ideas that we haven't touched on since defining abstract vector spaces. Although we only specifically work with the dot product of vectors in $\mathbb{R}^n$ in the following chapters, everything that is covered is applicable to any finite dimensional vector space equipped with an inner product.

19. Orthogonality and the Gram-Schmidt Algorithm

So far, in all our generalizations of vectors spaces from $\mathbb{R}^2$ and $\mathbb{R}^3$, we have not used the geometry of the dot product. In this section, we want to see what that extra geometry gives us, in the case of $\mathbb{R}^n$.

In fact, one can add the extra geometry of an inner product to any vector space; on spaces of (continuous) functions with domain $[a,b]$, for example, the "correct" replacement to the dot product of vectors is the following *inner product* of functions:

$$\langle f,g\rangle = \int_a^b f(x)g(x)\,dx$$

This is a topic that you get to explore in second year analysis. One can also give a natural inner product to spaces of matrices $\mathbf{M}_{mn}$ by the formula

$$\langle A,B\rangle = \operatorname{tr}(AB^T)$$

The dot product and the above inner products have a number of extremely nice properties; but there are other *bilinear forms* that encode different geometries (such as space-time) more accurately. You can explore this topic in more detail in 3rd year applied linear algebra.

19.1 Orthogonality

Recall that two vectors $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$ are said to be *orthogonal* if their dot product $\mathbf{u}\cdot\mathbf{v}$ is zero. The dot product satisfies:

- $\mathbf{u}\cdot\mathbf{u}\geq 0$ and is equal to zero iff $\mathbf{u}=\mathbf{0}$
- $\mathbf{u}\cdot\mathbf{v}=\mathbf{v}\cdot\mathbf{u}$ (symmetry)
- $(a\mathbf{u}+b\mathbf{v})\cdot(c\mathbf{w}) = ac\mathbf{u}\cdot\mathbf{w}+bc\mathbf{v}\cdot\mathbf{w}$ for all $a,b,c\in\mathbb{R}$ and $\mathbf{u},\mathbf{v},\mathbf{w}\in\mathbb{R}^n$ and similarly
- $(a\mathbf{u})\cdot(b\mathbf{v}+c\mathbf{w}) = ab\mathbf{u}\cdot\mathbf{v}+ac\mathbf{u}\cdot\mathbf{w}$ (bilinearity).

Example 19.1.1 The vectors $(1,-1,0,1)$ and $(1,1,1,0)$ in $\mathbb{R}^4$ are orthogonal, since

$$(1,-1,0,1)\cdot(1,1,1,0) = 1\cdot 1+(-1)\cdot 1+0\cdot 1+1\cdot 0 = 1-1 = 0.$$

Note that the standard basis of $\mathbb{R}^3$, namely $\{(1,0,0),(0,1,0),(0,0,1)\}$ (or $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$) consists of vectors that are all pairwise orthogonal: that is, $\mathbf{e}_i\cdot\mathbf{e}_j = 0$ for $i\neq j$.

We'd like to generalize this idea, since this basis has (at least) one very convenient property: for any $\mathbf{v}\in\mathbb{R}^3$, it's easy to find a,b and c such that $\mathbf{v} = a\,\mathbf{e}_1 + b\,\mathbf{e}_2 + c\,\mathbf{e}_3$: we all know that $a = \mathbf{v}\cdot\mathbf{e}_1, b = \mathbf{v}\cdot\mathbf{e}_2$ and $c = \mathbf{v}\cdot\mathbf{e}_3$.

19.2 Orthogonal sets of vectors

Definition 19.2.1 A set of vectors $\{\mathbf{v}_1,\cdots,\mathbf{v}_m\}$ in $\mathbb{R}^n$ is called *orthogonal* if

$$\mathbf{v}_i\cdot\mathbf{v}_j = 0$$

for all $1\leq i<j\leq m$, and $\mathbf{v}_j\neq 0$ for all $1\leq i\leq m$. That is, every pair of vectors is orthogonal and no vector is zero.

Note: If an orthogonal set consists of vectors all of which have length 1, we call the set *orthonormal*. Indeed, if $\{\mathbf{v}_1,\cdots,\mathbf{v}_m\}$ is orthogonal, then $\{\frac{\mathbf{v}_1}{\|\mathbf{v}_1\|},\ldots,\frac{\mathbf{v}_m}{\|\mathbf{v}_m\|}\}$ is orthonormal.

Example 19.2.2 The standard basis of $\mathbb{R}^n$ is an orthogonal set. (It's also orthonormal.)

Example 19.2.3 $\{(1,0,0),(0,1,0),(1,0,1)\}$ is NOT ORTHOGONAL even though two of the products are zero — since the first and third vectors aren't orthogonal, this is not an orthogonal set.

Orthogonal sets of vectors also enjoy a general Pythagorean property: if $\{\mathbf{v}_1,\cdots,\mathbf{v}_m\}$ is orthogonal, then

$$\|\mathbf{v}_1+\mathbf{v}_2+\cdots+\mathbf{v}_m\|^2 = \|\mathbf{v}_1\|^2+\|\mathbf{v}_2\|^2+\cdots+\|\mathbf{v}_m\|^2 \qquad (\star)$$

Prove this for yourself, remembering that $\|\mathbf{v}\|^2 = \mathbf{v}\cdot\mathbf{v}$.

Even better, (and just like the standard basis of $\mathbb{R}^n$) every orthogonal set is linearly independent:

Theorem 19.2.4 — Linear independence of orthogonal sets. Any orthogonal set of vectors is linearly independent.

Proof. Let $\{\mathbf{v}_1,\cdots,\mathbf{v}_m\}$ be an orthogonal set and suppose there are scalars $a_1,\ldots a_m$ such that

$$a_1\mathbf{v}_1+\cdots+a_m\mathbf{v}_m = \mathbf{0}.$$

Take the dot product of both sides with $\mathbf{v}_1$; the answer has to be zero because the right hand side is $\mathbf{v}_1\cdot\mathbf{0} = 0$. But on the left hand side, we get

$$\mathbf{v}_1\cdot(a_1\mathbf{v}_1+\cdots+a_m\mathbf{v}_m) = a_1\mathbf{v}_1\cdot\mathbf{v}_1 + a_2 0+\cdots+a_m 0$$

since $\mathbf{v}_1\cdot\mathbf{v}_j = 0$ for all $j>1$. Since $\mathbf{v}_1\cdot\mathbf{v}_1\neq 0$ (since we assumed $\mathbf{v}_1\neq\mathbf{0}$) we deduce that $a_1 = 0$. Continuing in this way with each of the vectors gives $a_i = 0$ for all i, so this set is indeed linearly independent. ■

This is a marvelous and wonderful fact! It gives us

- Any orthogonal set of vectors in $\mathbb{R}^n$ has at most n elements.
- Any orthogonal set of n vectors in $\mathbb{R}^n$ is a basis for $\mathbb{R}^n$, called an *orthogonal basis*.

Furthermore, the proof gave us a *very* easy way to calculate the coordinates of a vector relative to an orthogonal basis.

Theorem 19.2.5 — The Expansion Theorem: Coordinates relative to an orthogonal basis. Suppose $\{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ is an orthogonal basis for a subspace W of $\mathbb{R}^n$. Then any vector $\mathbf{w} \in W$ can be written as

$$\mathbf{w} = \left(\frac{\mathbf{w}\cdot\mathbf{w}_1}{\|\mathbf{w}_1\|^2}\right)\mathbf{w}_1 + \left(\frac{\mathbf{w}\cdot\mathbf{w}_2}{\|\mathbf{w}_2\|^2}\right)\mathbf{w}_2 + \cdots + \left(\frac{\mathbf{w}\cdot\mathbf{w}_m}{\|\mathbf{w}_m\|^2}\right)\mathbf{w}_m \qquad (\star\star)$$

The coefficients of $\mathbf{w}_1, \ldots, \mathbf{w}_m$ obtained above – the coordinates of w relative the the ordered basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ – are sometimes also called the *Fourier coefficients* of $\mathbf{w}$ relative to that orthogonal basis.

The really exciting fact here: normally, to write $\mathbf{w}$ as a linear combination of the elements of a basis, you have to row reduce a matrix. Here, there's a simple formula that gives you the answer!

Proof. Since $\{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ is a basis of W, we know that there are scalars $a_1, \cdots, a_n$ such that

$$\mathbf{w} = a_1\mathbf{w}_1 + \cdots + a_n\mathbf{w}_n.$$

Now take the dot product of both sides with $\mathbf{w}_i$; then as in the proof of Theorem 19.2.4 above, this gives zero on every term except the ith on the right, so we deduce

$$\mathbf{w}_i\cdot\mathbf{w} = a_i\mathbf{w}_i\cdot\mathbf{w}_i$$

or

$$a_i = \frac{\mathbf{w}_i\cdot\mathbf{w}}{\mathbf{w}_i\cdot\mathbf{w}_i} = \frac{\mathbf{w}_i\cdot\mathbf{w}}{\|\mathbf{w}_i\|^2}$$

as required. ■

Problem 19.2.6 Express $(1,6,7)$ as a linear combination of $(1,2,1)$, $(1,0,-1)$ and $(1,-1,1)$.
Solution We notice that those last three vectors are orthogonal, and hence form a basis for $\mathbb{R}^3$. Therefore the theorem applies, and we deduce

$$\begin{bmatrix}1\\6\\7\end{bmatrix} = \frac{20}{6}\begin{bmatrix}1\\2\\1\end{bmatrix} + \frac{-6}{2}\begin{bmatrix}1\\0\\-1\end{bmatrix} + \frac{2}{3}\begin{bmatrix}1\\-1\\1\end{bmatrix}$$

which (you'll notice) is true.

■ **Example 19.2.7** Let $W = \{(x,y,z,w) \in \mathbb{R}^4 \mid x-y-z+w=0\}$. We know W is a subspace of $\mathbb{R}^4$, and that $\dim W = 3$, because $W = \operatorname{Null}(\begin{bmatrix}1 & -1 & -1 & 1\end{bmatrix})$, and $\begin{bmatrix}1 & -1 & -1 & 1\end{bmatrix}$ has rank 1. Remember the Rank-Nullity Theorem (Corollary 15.2.3):

$$\dim W = \dim\operatorname{Null}(\begin{bmatrix}1 & -1 & -1 & 1\end{bmatrix}) = 4 - \operatorname{rank}(\begin{bmatrix}1 & -1 & -1 & 1\end{bmatrix}) = 3.$$

Now consider the 3 vectors[1] $\mathbf{w}_1 = (1,1,1,1), \mathbf{w}_2 = (1,-1,1,-1)$ and $\mathbf{w}_3 = (1,1,-1,-1)$. It's easy to check they are in W, and it's also easy to check that $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ is an orthogonal set, and hence is

[1] Don't worry about how we found these. They are not the *basic* solutions we would find using the standard method for finding a basis of a nullspace. We'll discuss that later.

linearly independent. Since we have 3 linearly independent vectors in the subspace W of dimension 3, $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ is an orthogonal basis for W.

Now $\mathbf{w} = (1,2,3,4) \in W$ (you can check this). Let's use our theorem to write it as a linear combination of $\mathbf{w}_1, \mathbf{w}_2$ and $\mathbf{w}_3$.

It's a simple calculation!

$$\frac{\mathbf{w}\cdot\mathbf{w}_1}{\|\mathbf{w}_1\|^2} = \frac{10}{4} = \frac{5}{2}$$
$$\frac{\mathbf{w}\cdot\mathbf{w}_2}{\|\mathbf{w}_2\|^2} = \frac{-2}{4} = -\frac{1}{2}$$
$$\frac{\mathbf{w}\cdot\mathbf{w}_3}{\|\mathbf{w}_3\|^2} = \frac{-4}{4} = -1$$

So we can conclude (don't you just *love* theorems!):

$$\mathbf{w} = \frac{5}{2}\mathbf{w}_1 - \frac{1}{2}\mathbf{w}_2 - \mathbf{w}_3.$$

(You can check this is true directly if you wish.)

Compare and contrast this with the way you solve this problem if you didn't happen to notice we have an orthogonal basis:

$$\begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2 & \mathbf{w}_3 & | & \mathbf{w}\end{bmatrix} = \begin{bmatrix}1 & 1 & 1 & | & 1\\ 1 & -1 & 1 & | & 2\\ 1 & 1 & -1 & | & 3\\ 1 & -1 & -1 & | & 4\end{bmatrix} \sim \begin{bmatrix}1 & 1 & 1 & | & 1\\ 0 & -2 & 0 & | & 1\\ 0 & 0 & -2 & | & 2\\ 0 & -2 & -2 & | & 3\end{bmatrix} \sim \begin{bmatrix}1 & 0 & 1 & | & \frac{3}{2}\\ 0 & 1 & 0 & | & -\frac{1}{2}\\ 0 & 0 & -2 & | & 2\\ 0 & 0 & -2 & | & 2\end{bmatrix}$$

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

which gives the same answer, but took more effort. ■

! Point: *if* you have an orthogonal basis, some calculations are much easier. *But if you don't* then you have to use the "old" techniques.

19.3 Orthogonal projection: a useful formula may have more than one use...

Now let's look again at the right hand side of the equation ($\star\star$) in the Expansion Theorem 19.2.5. It makes perfect sense *even if* $\mathbf{w} \notin W$!

"So what?", you might say. Well, look at the very first term in that sum:

$$\left(\frac{\mathbf{w}\cdot\mathbf{w}_1}{\|\mathbf{w}_1\|^2}\right)\mathbf{w}_1$$

We recognize this as the same expression for the projection of the vector $\mathbf{w}$ onto $\mathbf{w}_1$ that we saw in the context of $\mathbb{R}^3$ way back in Section 2.8 on page 34. The other terms are also projections of the same

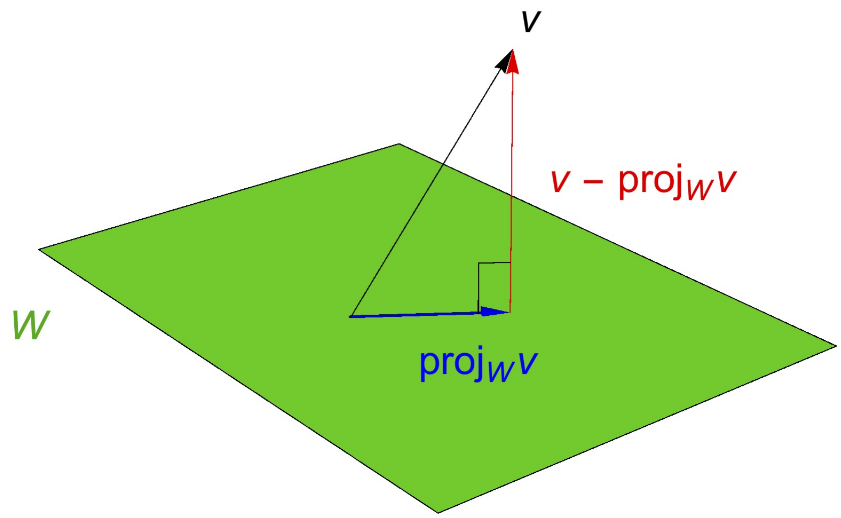


Figure 19.1: The orthogonal projection of $\mathbf{v}$ onto the subspace W.

vector $\mathbf{w}$ onto the other vectors $\mathbf{w}_2, \ldots, \mathbf{w}_m$. So the expression on the right hand side of the equation $(\star\star)$ in Theorem 19.2.5 is a sum of the projections of $\mathbf{w}$ onto the (orthogonal) vectors $\mathbf{w}_1, \ldots, \mathbf{w}_m$, which are a basis of the subspace W. So it's not *completely* outrageous to think of this expression as the orthogonal projection of $\mathbf{w}$ onto the *subspace* W.

So let's run with the ball!

Definition 19.3.1 — Orthogonal projection onto a subspace. Let W be a subspace of $\mathbb{R}^n$ and $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ an *orthogonal* basis for W. Then for any $\mathbf{v} \in \mathbb{R}^n$, the *orthogonal projection of* $\mathbf{v}$ *onto* W is defined by

$$\text{proj}_W(\mathbf{v}) = \left(\frac{\mathbf{v}\cdot\mathbf{w}_1}{\|\mathbf{w}_1\|^2}\right)\mathbf{w}_1 + \cdots + \left(\frac{\mathbf{v}\cdot\mathbf{w}_m}{\|\mathbf{w}_m\|^2}\right)\mathbf{w}_m$$

In Figure 19.1 there's a diagram (since most of us are low dimensional beings) illustrating this idea in 'three' dimensions– seen in just two, here, of course.

Now your enquiring mind may have at least two thoughts at this point. In no particular order, they may be:

1. *Does every subspace have an orthogonal basis?* The formula needs one!
2. *What if I compute this projection with two different orthogonal bases? Surely I'll get different answers! All the numbers in the sum will be different...*

We'll answer the second question first – it doesn't matter[2] which orthogonal basis you use! But not before an example.

■ **Example 19.3.2** Let's look again at $W = \{(x,y,z,w) \in \mathbb{R}^4 \mid x - y - z + w = 0\}$. We know the three vectors $\mathbf{w}_1 = (1,1,1,1), \mathbf{w}_2 = (1,-1,1,-1)$ and $\mathbf{w}_3 = (1,1,-1,-1)$ form an orthogonal basis for W.[3]

[2] Amazing fact from linear algebra number 3

[3] Don't worry about how we found these. We'll discuss that later, in answer to the first question above.

Now $\mathbf{v} = (1,2,3,5) \notin W$: check it! Let's use the projection formula to find the orthogonal projection of $\mathbf{v}$ onto W, and do a bit of checking.

Solution We begin by calculating:

$$\frac{\mathbf{v}\cdot\mathbf{w}_1}{\|\mathbf{w}_1\|^2} = \frac{11}{4}$$
$$\frac{\mathbf{v}\cdot\mathbf{w}_2}{\|\mathbf{w}_2\|^2} = \frac{-3}{4}$$
$$\frac{\mathbf{v}\cdot\mathbf{w}_3}{\|\mathbf{w}_3\|^2} = \frac{-5}{4}$$

and conclude

$$\text{proj}_W(\mathbf{v}) = \frac{11}{4}\mathbf{w}_1 - \frac{3}{4}\mathbf{w}_2 - \frac{5}{4}\mathbf{w}_3 = \frac{1}{4}(3,9,13,19)$$

If you check, you'll see that $\frac{1}{4}(3,9,13,19) \in W$: simply show it satsifies $x - y - z + w = 0$.

N.B. <u>You must not rescale this answer!</u> This is a very, very common error at this stage. Rescaling an orthogonal basis is fine, but the projection is a fixed answer.

Moreover, as Figure 19.1, and indeed the properties of $\mathbf{v} - \text{proj}_W(\mathbf{v})$ in $\mathbb{R}^3$ (see Section 2.8 and Figure 2.1) all suggest, $\mathbf{v} - \text{proj}_W(\mathbf{v}) = \frac{1}{4}(1,-1,-1,1)$ is indeed orthogonal to every vector in W: look again at the equation defining W: it says $(x,y,z,w) \in W$ iff

$$0 = x - y - z + w = (x,y,z,w)\cdot(1,-1,-1,1)!$$

So every vector in W *is* orthogonal to $\mathbf{v} - \text{proj}_W(\mathbf{v})$.

■

! This *always* happens. And as both Figures 2.1 and 19.1 also suggest, this makes $\text{proj}_W(\mathbf{v})$ *the closest vector in W to $\mathbf{v}$.*

Another way to say this is that $\text{proj}_W(\mathbf{v})$ is the *best approximation to $\mathbf{v}$ by vectors in W.*

Let's show that this is true:

Theorem 19.3.3 — The Best Approximation Theorem. Let W be a subspace of $\mathbb{R}^n$ and let $\mathbf{v} \in \mathbb{R}^n$. Then,

(1) $\text{proj}_W(\mathbf{v}) \in W$,
(2) $\mathbf{v} - \text{proj}_W(\mathbf{v})$ is orthogonal to every vector in W,
(3) $\text{proj}_W(\mathbf{v})$ is best approximation to $\mathbf{v}$ by vectors in W, and
(4) The vector $\text{proj}_W(\mathbf{v})$ is the <u>only</u> vector in $\mathbb{R}^n$ which satisfies (1) and (2).

(Property 4 says that the orthogonal projection is uniqely characterized by the two properties (1) and (2).)

Proof. Part (1) is easy: look at the formula in Definition 19.3.1. The vector $\text{proj}_W(\mathbf{v})$ is a linear combination of vectors in the (subspace) W and so must belong to W.

For (2), let's think: how *could* we show a vector is orthogonal to every vector in W? There are infinitely many vectors in (most) subspaces! But that's one thing bases are very good for – rendering finite the infinite !

So suppose $\{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ is the orthogonal basis for W we used to define the projection. Every $\mathbf{u} \in W$ is a linear combination of $\mathbf{w}_1, \cdots, \mathbf{w}_m$, *i.e.*,

$$\mathbf{u} = a_1\mathbf{w}_1 + \cdots + a_m\mathbf{w}_m$$

for some scalars $a_, \ldots, a_m$. Set $\mathbf{w} = \text{proj}_W(\mathbf{v})$. So if $(\mathbf{v} - \text{proj}_W(\mathbf{v})) \cdot \mathbf{w}_i = 0$ for $1 \leq i \leq m$, then $(\mathbf{v} - \mathbf{w}) \cdot \mathbf{w}_i = 0$ so

$$\begin{aligned}(\mathbf{v}-\mathbf{w})\cdot\mathbf{u} &= (\mathbf{v}-\mathbf{w})\cdot(a_1\mathbf{w}_1+\cdots+a_m\mathbf{w}_m)\\ &= a_1((\mathbf{v}-\mathbf{w})\cdot\mathbf{w}_1)+\cdots+a_1((\mathbf{v}-\mathbf{w})\cdot\mathbf{w}_m)\\ &= a_1(0)+\cdots+a_1(0)\\ &= 0.\end{aligned}$$

So, it suffices[4] to show that $(\mathbf{v} - \mathbf{w}) \cdot \mathbf{w}_i = 0$ for $1 \leq i \leq m$. So we calculate:

$$\begin{aligned}(\mathbf{v}-\mathbf{w})\cdot\mathbf{w}_i &= (\mathbf{v}-(\frac{\mathbf{v}\cdot\mathbf{w}_1}{\|\mathbf{w}_1\|^2}\mathbf{w}_1+\cdots+\frac{\mathbf{v}\cdot\mathbf{w}_m}{\|\mathbf{w}_m\|^2}\mathbf{w}_m))\cdot\mathbf{w}_i\\ &= \mathbf{v}\cdot\mathbf{w}_i-\frac{\mathbf{v}\cdot\mathbf{w}_1}{\|\mathbf{w}_1\|^2}\mathbf{w}_1\cdot\mathbf{w}_i-\cdots-\frac{\mathbf{v}\cdot\mathbf{w}_m}{\|\mathbf{w}_m\|^2}\mathbf{w}_m\cdot\mathbf{w}_i\\ &= \mathbf{v}\cdot\mathbf{w}_i-\frac{\mathbf{v}\cdot\mathbf{w}_i}{\|\mathbf{w}_i\|^2}\mathbf{w}_i\cdot\mathbf{w}_i\end{aligned}$$

(since all the other dot products between $\mathbf{w}_i$ and $\mathbf{w}_j$ are zero); and this last term gives

$$= \mathbf{v}\cdot\mathbf{w}_i-\frac{\mathbf{v}\cdot\mathbf{w}_i}{\|\mathbf{w}_i\|^2}\|\mathbf{w}_i\|^2 = 0.$$

This is true for every $i = 1, 2, \cdots, m$ so we conclude that $\mathbf{v} - \mathbf{w}$ is orthogonal to every vector in W, as required. This shows that (2) is true.

To show that (3) holds, denote, as before $\mathbf{w} = \text{proj}_W(\mathbf{v})$. To prove this is the closest vector in W to $\mathbf{v}$, let $\mathbf{w}' \in W$ be any other vector in W. Then

$$\|\mathbf{v}-\mathbf{w}'\|^2 = \|(\mathbf{v}-\mathbf{w})+(\mathbf{w}-\mathbf{w}')\|^2$$

But since $\mathbf{v} - \mathbf{w}$ is orthogonal to every vector in W, and $\mathbf{w} - \mathbf{w}' \in W$ (W is a subspace and both $\mathbf{w}$ and $\mathbf{w}'$ belong to W), the two vectors $\mathbf{v} - \mathbf{w}$ and $\mathbf{w} - \mathbf{w}'$ are orthogonal! So by the Pythagorean theorem $\star$ on P. 196, we have

$$\|\mathbf{v}-\mathbf{w}'\|^2 = \|(\mathbf{v}-\mathbf{w})\|^2+\|(\mathbf{w}-\mathbf{w}')\|^2 \geq \|(\mathbf{v}-\mathbf{w})\|^2,$$

since $0 \leq \|(\mathbf{w} - \mathbf{w}')\|^2$ is always true. Thus we've shown (3) is true.

The last part (4) is also straightforward. Now suppose two vectors $\mathbf{w}, \mathbf{w}'$ satisfy the properties (1) and (2) above, namely:

- $\mathbf{w}, \mathbf{w}' \in W$
- $\mathbf{v} - \mathbf{w}$ and $\mathbf{v} - \mathbf{w}'$ are both orthogonal to every vector in W.

[4] Note that we did not use any property of $\{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ except that it spanned W, so the same computation shows that a vector is orthogonal to every vector in a subspace W iff it is orthogonal to any *spanning set* of W.

Then, $\mathbf{w}-\mathbf{w}' \in W$ since W, being a subspace, is closed under subtraction. But look:

$$\mathbf{w}-\mathbf{w}' = (\mathbf{v}-\mathbf{w}') - (\mathbf{v}-\mathbf{w}).$$

Now the sum – or difference – of two vectors which are both orthogonal to every vector in W will again be orthogonal to every vector in W.[5] This means that the vector $\mathbf{w}-\mathbf{w}'$ is orthogonal to every vector in W, including $\mathbf{w}-\mathbf{w}'$ itself! But the only vector orthogonal to itself is the zero vector: so $\mathbf{w}-\mathbf{w}'=0$, i.e., $\mathbf{w}=\mathbf{w}'$.

This shows that there is only one vector with the properties (1) and (2) above, namely $\text{proj}_W(\mathbf{v})$.

■

Note that property (3) above of the orthogonal projection is extremely useful, in many areas. It is a huge part of the mathematical foundation for modern signal processing. It is used in almost all quantitative methods for 'best' curve fitting for data. We present an example of this in Section 20.4.

The basic idea it that we have often a 'favourite' subspace of vectors W (in a vector space of signals, for example), and another vector that we can't be sure belongs to our subspace, but we only wish to deal with vectors in W, so we project our vector onto W and deal with that instead.

For example, in a simplified view of signal processing, on your cellphone, for example, when you take a picture and want to send it to a friend, the picture is captured and stored and treated as a vector in a large dimensional space ($\mathbb{R}^n$ for some very large n). But instead of sending all the data involved, it will be (orthogonally) projected onto a much smaller subspace – with an orthogonal basis peculiar to the protocol being used – and then only the relatively small number (the dimension of W) of Fourier coefficients will be transmitted. At the other end, your friend will receive the Fourier coefficients, and knowing which protocol you used, will use the same orthogonal basis, and rebuild the projection using the the Fourier coefficients, getting an approximation (better or worse, depending on $\dim W$) of the picture you sent.

Part (4) above seems innocuous until you recall that Definition 19.3.1 *looked like* it depended on a choice of orthogonal basis. So what Part (4) is saying is that the *apparently* different formulae one gets when using different orthogonal bases to compute the projection all give exactly the same answer!

19.4 Finding orthogonal bases: The Gram-Schmidt algorithm

So far, we've implicitly assumed that every subspace W of $\mathbb{R}^n$ *has* an orthogonal basis. Is is true? Yes, and even better, there's a simple algorithm that will take any basis of W as input and output an orthogonal basis of W.

Suppose $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ is any basis for W. We want to find an orthogonal basis for W.

The geometric idea is quite simple: start with your first vector $\mathbf{u}_1$. Then take $\mathbf{u}_2 - \text{proj}_{\mathbf{u}_1}(\mathbf{u}_2)$ to get something which is orthogonal to $\mathbf{u}_1$ but — crucially — lies in the span of $\mathbf{u}_1$ and $\mathbf{u}_2$. We then keep subtracting off projections onto the span of the preceding vectors we obtain and the result is an orthogonal basis for W.

[5] I don't know about you, but I'm getting tired of writing 'is orthogonal to every vector in W', so soon we will save ourslves some bother and give a name to the collection of all vectors orthogonal to every vector in W. We'll call it $W^\perp$ — see Chapter 20.

Theorem 19.4.1 — Gram-Schmidt Algorithm. Let $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ be *any* basis for U. Define

- $\mathbf{w}_1 = \mathbf{u}_1$ and $V_1 = \text{span}\{\mathbf{w}_1\}$
- $\mathbf{w}_2 = \mathbf{u}_2 - \text{proj}_{V_1} \mathbf{u}_2$ and $V_2 = \text{span}\{\mathbf{w}_1, \mathbf{w}_2\}$
- $\mathbf{w}_3 = \mathbf{u}_3 - \text{proj}_{V_2} \mathbf{u}_3$ and $V_3 = \text{span}\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$
- $\cdots$
- $\mathbf{w}_m = \mathbf{u}_m - \text{proj}_{V_{m-1}} \mathbf{u}_m$ and $V_m = \text{span}\{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$

Then $W = V_m$ and $\{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ is an orthogonal basis for W.

Written out in more detail: define

- $\mathbf{w}_1 = \mathbf{u}_1$,
- $\mathbf{w}_2 = \mathbf{u}_2 - \text{proj}_{\mathbf{w}_1} \mathbf{w}_2$,
- $\mathbf{w}_3 = \mathbf{u}_3 - \text{proj}_{\mathbf{w}_1} \mathbf{u}_3 - \text{proj}_{\mathbf{w}_2} \mathbf{u}_3$,
- $\cdots$
- $\mathbf{w}_m = \mathbf{u}_m - \sum_{i=1}^{m-1} \text{proj}_{\mathbf{w}_i} \mathbf{u}_m$.

Then $\{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ is an orthogonal basis for U.

Finally, one could scale each of the vectors $\mathbf{w}_i$ by dividing by their norm to produce an orthonormal basis for W.

Notice that at each step, the vectors in the spanning set for V_i are orthogonal, so that means it's easy to calculate the projection onto V_k. And in fact, $V_k = \text{span}\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$; that is, the span of the first k vectors of the set is the same as the span of the first k vectors of the original spanning set.

Indeed, one could begin with just a *spanning* set $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ for U, and apply the Gram-Schmidt Algorithm as above: then the *non-zero* vectors in $\{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$ would be an orthogonal basis for U - see problem 19.6 in the exercises.

■ **Example 19.4.2** Let's perform the Gram-Schmidt algorithm on the the set

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

(You can check that this is a basis for the subspace $W = \text{span}\{(1,0,0,1), (1,1,1,0), (2,1,-1,1)\} = \{(x,y,z,w) \in \mathbb{R}^4 \mid x - y - w = 0\}$.[6])

Well, we start with

$$\mathbf{w}_1 = (1,0,0,1)$$

and then

$$\begin{aligned} \mathbf{w}_2 &= \mathbf{u}_2 - \text{proj}_{\mathbf{w}_1} \mathbf{u}_2 \\ &= \mathbf{u}_2 - \frac{\mathbf{u}_2 \cdot \mathbf{w}_1}{\|\mathbf{w}_1\|^2} \mathbf{w}_1 \\ &= (1,1,1,0) - \frac{1}{2}(1,0,0,1) = (\frac{1}{2}, 1, 1, -\frac{1}{2}). \end{aligned}$$

(Before going on, it's always a good idea to check that $\mathbf{w}_1 \cdot \mathbf{w}_2 = 0$! It's OK here.)

So in this example, there's just one more step:

[6] We saw how to do this at the end of chapter 15.

$$\begin{aligned}
\mathbf{w}_3 &= \mathbf{u}_3 - \frac{\mathbf{u}_3 \cdot \mathbf{w}_1}{\|\mathbf{w}_1\|^2}\mathbf{w}_1 - \frac{\mathbf{u}_3 \cdot \mathbf{w}_2}{\|\mathbf{w}_2\|^2}\mathbf{w}_2 \\
&= (2,1,-1,1) - \frac{3}{2}(1,0,0,1) - \frac{\frac{1}{2}}{\frac{5}{2}}(\frac{1}{2},1,1,-\frac{1}{2}) \\
&= (2,1,-1,1) - (\frac{3}{2},0,0,\frac{3}{2}) - (\frac{1}{10},\frac{1}{5},\frac{1}{5},-\frac{1}{10}) \\
&= \frac{2}{5}(1,2,-3,-1).
\end{aligned}$$

Then we can see that $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ is indeed orthogonal – and it's always a good idea to check this.[7] Also, by looking at our calculations more carefully, we can see that indeed, each successive element is in the span of the original vectors. The resulting orthogonal basis is

$$\{(1,0,0,1), \frac{1}{2}(1,2,2,-1), \frac{2}{5}(1,2,-3,-1)\}.$$

■ **Remark 19.4.3.** *We often choose to simplify our calculations when using the formula in Theorem 19.4.1 by scaling the* $\mathbf{w}_i$ *by something convenient to avoid fractions: since*

$$\{(1,0,0,1), (1,2,2,-1), (1,2,-3,-1)\}$$

is just as good an orthogonal set as the one above, and might be easier to work with if you had to calculate another projection.

Finally, if one really wants an orthonormal basis, we have to divide each of these vectors by their norm. Since $\|(1,0,0,1)\| = \sqrt{2}$, $\|(1,2,2,-1)\| = \sqrt{10}$, and $\|(1,2,-3,-1)\| = \sqrt{15}$, we obtain the orthonormal basis:

$$\{\frac{\sqrt{2}}{2}(1,0,0,1), \frac{\sqrt{10}}{10}(1,2,2,-1), \frac{\sqrt{15}}{15}(1,2,-3,-1)\}.$$

(Check this!) ■

■ **Example 19.4.4** Calculate the best approximation to $\mathbf{v} = (1,2,3,4)$ in W, where

$$W = \text{span}\{(1,0,0,1), (1,1,1,0), (2,1,-1,1)\} = \{(x,y,z,w) \in \mathbb{R}^4 \mid x - y - w = 0\}$$

Let's use the preceding exercise. We already know that an orthogonal basis for W is

$$\{(1,0,0,1), (1,2,2,-1), (1,2,-3,-1)\}$$

(call these $\mathbf{n}_1$, $\mathbf{n}_2$ and $\mathbf{n}_3$, say) so therefore

$$\begin{aligned}
\text{proj}_W(\mathbf{v}) &= \frac{\mathbf{v} \cdot \mathbf{n}_1}{\|\mathbf{n}_1\|^2}\mathbf{n}_1 + \frac{\mathbf{v} \cdot \mathbf{n}_2}{\|\mathbf{n}_2\|^2}\mathbf{n}_2 + \frac{\mathbf{v} \cdot \mathbf{n}_3}{\|\mathbf{n}_3\|^2}\mathbf{n}_3 \\
&= \frac{5}{2}(1,0,0,1) + \frac{7}{10}(1,2,2,-1) - \frac{8}{15}(1,2,-3,-1) \\
&= \frac{1}{3}(8,1,3,7).
\end{aligned}$$

[7] If you can, it's also a very good idea to check that your vectors $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ belong to W– which of course they should! In this case that's easy, as we have a simple description of W – namely, $W = \{(x,y,z,w) \in \mathbb{R}^4 \mid x - y - w = 0\}$.

If you can, check that your answer actually belongs to W: here it's easy since we have a simple description of W.[8] ■

Problems

Remarks:

1. A question with an asterisk '*' (or two) indicates a bonus-level question.
2. You must justify all your responses.

Problem 19.1 In each case, find the Fourier coefficients of the vector $\mathbf{v}$ with respect to the given orthogonal basis $\mathscr{B}$ of the indicated vector space W.

a) $\mathbf{v} = (1,2,3)$, $\mathscr{B} = \{(1,0,1),(-1,0,1),(0,1,0)\}$, $W = \mathbb{R}^3$.

b) $^\star\mathbf{v} = (1,2,3)$, $\mathscr{B} = \{(1,2,3),(-5,4,-1),(1,1,-1)\}$, $W = \mathbb{R}^3$.

c) $\mathbf{v} = (1,2,3)$, $\mathscr{B} = \{(1,0,1),(-1,2,1)\}$,

$$W = \{(x,y,z) \in \mathbb{R}^3 \mid x+y-z=0\}.$$

d) $^\star\mathbf{v} = (4,-5,0)$, $\mathscr{B} = \{(-1,0,5),(10,13,2)\}$,

$$W = \{(x,y,z) \in \mathbb{R}^3 \mid 5x-4y+z=0\}.$$

e) $\mathbf{v} = (1,1,1,1)$, $\mathscr{B} = \{(1,0,1,1),(0,1,0,0),(0,0,1,-1)\}$,

$$W = \{(x,y,z,w) \in \mathbb{R}^4 \mid x-w=0\}.$$

f) $^\star\mathbf{v} = (1,0,1,2)$, $\mathscr{B} = \{(1,0,1,1),(0,1,0,0),(0,0,1,-1),(1,0,0,-1)\}$, $W = \mathbb{R}^4$.

Problem 19.2 Find the formula for the orthogonal projection onto the subspaces in parts c), d)*and e) above.

Problem 19.3 Apply the Gram-Schmidt algorithm to each of the following linearly independent sets (to obtain an orthogonal set), and check that your resulting set of vectors is orthogonal.

a) $\{(1,1,0),(2,0,3)\}$

b) $^\star\{(1,0,0,1),(0,1,0,-1),(0,0,1,-1)\}$

c) $\{(1,1,1,1),(0,1,0,0),(0,0,1,-1)\}$

d) $^\star\{(1,1,0),(1,0,2),(1,2,1)\}$

[8] If you have the time, you should also check that $(1,2,3,4)-\frac{1}{3}(8,1,3,7)=\frac{1}{3}(-5,5,0,5)$ is orthogonal to everything in W, and it is, since $\frac{1}{3}(-5,5,0,5).(x,y,z,w)=-\frac{5}{3}(x-y-w)=0$ whenever $(x,y,z,w)\in W$!

Problem 19.4 Find an orthogonal basis for each of the following subspaces, and check that your basis is orthogonal. (First, find a basis in the standard way, and then apply the Gram-Schmidt algorithm.)

a) $W = \{(x,y,z) \in \mathbb{R}^3 \mid x+y+z=0\}$

b) ⋆$U = \{(x,y,z,w) \in \mathbb{R}^4 \mid x+y-w=0\}$

c) $X = \{(x,y,z,w) \in \mathbb{R}^4 \mid x+y-w=0 \text{ and } z+y=0\}$

d) ⋆$V = \ker \begin{bmatrix} 1 & 2 & -1 & -1 \\ 2 & 4 & -1 & 3 \\ -3 & -6 & 1 & -7 \end{bmatrix}$

Problem 19.5 Find the best approximation to each of the given vectors $\mathbf{v}$ from the given subspace W.

a) $\mathbf{v} = (1,1,1)$, $W = \{(x,y,z) \in \mathbb{R}^3 \mid x+y-z=0\}$

b) ⋆$\mathbf{v} = (1,1,1)$, $W = \{(x,y,z) \in \mathbb{R}^3 \mid 5x-4y+z=0\}$

c) $\mathbf{v} = (1,1,1,2)$, $W = \{(x,y,z,w) \in \mathbb{R}^4 \mid x-w=0\}$

Problem 19.6 State whether each of the following is (always) true, or is (possibly) false.

- If you say the statement may be false, you must give an explicit example.
- If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

a) Every orthogonal set is linearly independent.

b) ⋆Every linearly independent set is orthogonal.

c) When finding the orthogonal projection of a vector $\mathbf{v}$ onto a subspace W, plugging *any* basis of W into the formula in Definition 19.3.1 will work.

d) ⋆When finding the orthogonal projection of a vector $\mathbf{v}$ onto a subspace W, once the answer is obtained, it's OK to re-scale the answer to eliminate fractions.

e) When applying the Gram-Schmidt algorithm to a basis, at each step, it's OK to re-scale the vector obtained to eliminate fractions.

f) ⋆When finding the orthogonal projection of a vector $\mathbf{v}$ onto a subspace W, using different orthogonal bases of W in the formula in Definition 19.3.1 can give different answers.

g) If $\text{proj}_W(\mathbf{v})$ denotes the orthogonal projection of a vector $\mathbf{v}$ onto a subspace W, then $\mathbf{v} - \text{proj}_W(\mathbf{v})$ is always orthogonal to every vector in W.

h) ⋆To check that a vector, say $\mathbf{u}$, is orthogonal to every vector in W, it suffices to check that $\mathbf{u}$ is orthogonal to every vector in *any* basis of W.

i) ** If you apply the the Gram-Schmidt algorithm to a *spanning set* $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ of a subspace W, rather than a basis of W, and if you obtain a zero vector at step k, that means $\mathbf{w}_k \in \operatorname{span}\{\mathbf{w}_1, \ldots, \mathbf{w}_{k-1}\}$.

j) ** If you apply the the Gram-Schmidt algorithm to a *spanning set* $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ of a subspace W, rather than a basis of W, and discard any zero vectors that appear in the process, you will in the end still obtain an orthogonal basis for W.

20. Orthogonal Complements and Applications

20.1 Orthogonal Complements

You will recall that one of two important properties of the orthogonal projection $\text{proj}_W(\mathbf{v})$ onto a subspace W was that $\mathbf{v} - \text{proj}_W(\mathbf{v})$ was *orthogonal to every vector in that subspace.* This suggests the following definition.

Definition 20.1.1 Let U be a subspace of $\mathbb{R}^n$. The *orthogonal complement* of U is the set, denoted $U^\perp$ (pronounced: 'U-perp') defined by

$$U^\perp = \{\mathbf{v} \in \mathbb{R}^n \mid \mathbf{u} \cdot \mathbf{v} = 0 \text{ for all } \mathbf{u} \in U\}.$$

■ **Example 20.1.2** Let $U = \text{span}\{(1,0,3),(0,1,2)\}$. We know that U is the plane through the origin with normal $(1,0,3) \times (0,1,2) = (-3,-2,1)$.

So a vector is orthogonal to U iff it is parallel to $(-3,-2,1)$. Thus,

$$U^\perp = \text{span}\{(-3,-2,1)\}.$$

Of course $U^\perp$ is the line through the origin with direction vector $(-3,-2,1)$.

■

■ **Example 20.1.3** Let U be the subspace of $\mathbb{R}^4$ defined by

$$U = \text{span}\{(1,-1,0,-1)\}$$

Since we already noted that a vector is orthogonal to every vector in a subspace iff it is orthogonal to a basis of the subspace (see the proof of Theorem 19.3.3) a vector $(x,y,z,w) \in U^\perp$ iff

$$(x,y,z,w) \cdot (1,-1,0,-1) = x - y - w = 0.$$

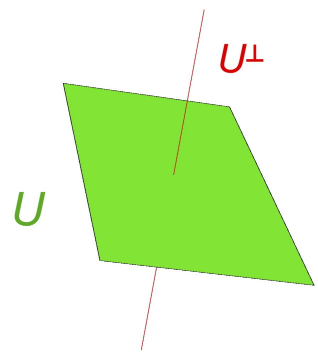


Figure 20.1: The plane U and its orthogonal complement.

That is, in this case,

$$U^\perp = \{(x,y,z,w) \in \mathbb{R}^4 \mid x - y - w = 0\}.$$

You will recognize this subspace as W of Example 19.4.2 from the last two examples in the previous chapter. ■

What about $W^\perp$ from the example just above?

■ **Example 20.1.4** Let $W = \{(x,y,z,w) \in \mathbb{R}^4 \mid x - y - w = 0\}$.

We know that $\mathbf{v} \in W^\perp$ iff $\mathbf{v}$ is orthogonal to any basis of W. Well, we've lots of those: *e.g.* from Example 19.4.4: we know that

$$\{\mathbf{v}_1 = (1,0,0,1), \mathbf{v}_2 = (1,1,1,0), \mathbf{v}_3 = (2,1,-1,1)\}$$

is a basis of W, so we just need to find all $\mathbf{v}$ such that

$$\begin{aligned} \mathbf{v}_1 \cdot \mathbf{v} &= 0 \\ \mathbf{v}_2 \cdot \mathbf{v} &= 0 \\ \mathbf{v}_3 \cdot \mathbf{v} &= 0. \end{aligned} \tag{20.1}$$

We've seen this sort of thing before! Remember Remark 16.3.1? It said, in part, that for a matrix A, $\mathbf{v} \in \text{Null}(A)$ iff $\mathbf{v}$ is orthogonal to every row of A.

That's what we're staring at above in (20.1). If we set (in block row form)

$$A = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{bmatrix}$$

So solving the system in (20.1) is equivalent to finding $\text{Null}(A)$! Let's do it:

$$\text{Null}(A) = \text{Null} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 2 & 1 & -1 & 1 \end{bmatrix} = \text{Null} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \{(-s,s,0,s) \mid s \in \mathbb{R}\}.$$

So the set of basic solutions – just one in this case – namely $\{(-1,1,0,1)\}$ is a basis for $\text{Null}(A) = W^\perp$. But this subspace is our original U!

So $U^\perp = W$ and $W^\perp = U$, or said differently, $(U^\perp)^\perp = U$! Is this always true? ■

Indeed it is, and more is true as well:

Theorem 20.1.5 — Properties of the orthogonal complement. Let U be a subspace of $\mathbb{R}^n$. Then:
(1) $U^\perp$ is a subspace of $\mathbb{R}^n$
(2) $(U^\perp)^\perp = U$
(3) $\dim(U) + \dim(U^\perp) = n$.
When two subspaces U and V satisfy $U = V^\perp$ (or equivalently, $V = U^\perp$) then we say they are *orthogonal complements*.

Proof. (1) We apply the subspace test. We have

$$U^\perp = \{\mathbf{v} \in \mathbb{R}^n \mid \mathbf{u} \cdot \mathbf{v} = 0 \text{ for all } \mathbf{u} \in U.\}$$

(a) Since the zero vector is orthogonal to every vector in U (it's orthogonal to everything!), $\mathbf{0} \in U^\perp$. Good.

(b) Suppose $\mathbf{v}, \mathbf{w} \in U^\perp$. That means $\mathbf{u} \cdot \mathbf{v} = 0$ and $\mathbf{u} \cdot \mathbf{w} = 0$. Is $\mathbf{v} + \mathbf{w} \in U^\perp$? We calculate

$$\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w} = 0 + 0 = 0.$$

OK!

(c) Suppose $\mathbf{v} \in U^\perp$ and $k \in \mathbb{R}$. That means $\mathbf{u} \cdot \mathbf{v} = 0$. We calculate

$$\mathbf{u} \cdot (k\mathbf{v}) = k\mathbf{u} \cdot \mathbf{v} = k(0) = 0$$

so indeed $k\mathbf{v} \in U^\perp$.

Thus $U^\perp$ is a subspace.

Let's prove (3) next. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be a basis for U. Create the matrix A whose rows are these vectors (transposed!). Then the set of all vectors $\mathbf{v}$ which are orthogonal to U is exactly the set of vectors $\mathbf{v}$ such that $A\mathbf{v} = \mathbf{0}$ (using the argument of the example above). Hence $U^\perp = \text{Null}(A)$. By the rank-nullity theorem,

$$\dim(\text{Row}\,(A)) + \dim(\text{Null}(A)) = n \Leftrightarrow \dim(U) + \dim(U^\perp) = n.$$

Finally, to prove (2): Let $V = U^\perp$; this is a subspace by (1) and its dimension is $n - \dim(U)$ by (3). We want to conclude that $U = V^\perp$. Since every vector in U is indeed orthogonal to every vector in V, it is true that $U \subseteq V^\perp$. Furthermore, by (3), we have that $\dim(V^\perp) = n - (n - \dim(U)) = \dim(U)$. So U is a subspace of $V^\perp$ but has the same dimension as $V^\perp$: it must be all of $V^\perp$. (This is the theorem about how a subspace that isn't the whole space must have a strictly smaller dimension.)

■

Problem 20.1.6 Let A be any matrix. Show that $\text{Row}\,(A)$ and $\text{Null}(A)$ are orthogonal complements, as are $\text{Col}\,(A)$ and $\text{Null}(A^T)$.

Solution This is a reprise of our example above, in general. Suppose A is an $m \times n$ matrix. Recall that $\text{Null}A = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = 0\}$. But by definition of the multiplication of matrices, this is exactly saying that that dot product of each row of A with $\mathbf{x}$ gives 0. So

$$\mathrm{Null}(A) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \text{ is orthogonal to every row of } A\}$$

Since the rows of A span Row (A), by the footnote 4 on page 201, $\mathbf{x} \in \mathrm{Null}(A)$ iff $\mathbf{x}$ is orthogonal to every vector in Row (A), i.e. $\mathrm{Null}(A) = \mathrm{Row}\,(A)^{\perp}$, as required.

To see that $\mathrm{Null}(A^T) = \mathrm{Col}\,(A)^{\perp}$, substitute A^T for A in $\mathrm{Null}(A) = \mathrm{Row}\,(A)^{\perp}$ to obtain $\mathrm{Null}(A^T) = \mathrm{Row}\,(A^T)^{\perp}$ and note that $\mathrm{Col}\,(A) = \mathrm{Row}\,(A^T)$.

20.2 Orthogonal projection — an encore

Let's use what we know about orthogonal complements to find another way to compute the orthogonal projection. Don't be put off by the first example. It is easy — we could have done it in high school, but the method we use will be quite different, and will lead to the new way to compute the orthogonal projection.

Problem 20.2.1 Find the projection of $\mathbf{b} = (1,2,3)$ on the subspace U spanned by $\{(1,1,1)\}$.

Solution We're looking for a point $\mathbf{v} \in U$ such that $\mathbf{b} - \mathbf{v} \in U^{\perp}$. Think of $U = \mathrm{Col}\,(A)$ where $A = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Then $U^{\perp} = \mathrm{Null}(A^T)$.

An element of U is the same thing as an element of Col (A), which can be written as $\mathbf{v} = Ax$ for some x (in this silly case, $x \in \mathbb{R}$). To say that $\mathbf{b} - \mathbf{v} = \mathbf{b} - Ax$ is in $U^{\perp} = \mathrm{Null}(A^T)$ means that

$$A^T(\mathbf{b} - Ax) = 0$$

or

$$A^T Ax = A^T \mathbf{b}.$$

Now $A^T A = (1,1,1) \cdot (1,1,1) = 3$; and $A^T \mathbf{b} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 6$ so $x = 2$. Our answer is

$$\mathrm{proj}_U(\mathbf{b}) = \mathbf{v} = Ax = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$$

which coincides with our known formula from before:

$$\mathrm{proj}_{(1,1,1)}((1,2,3)) = \frac{(1,2,3) \cdot (1,1,1)}{\|(1,1,1)\|^2} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}.$$

But we've just found another way to calculate the projection onto *any* subspace of $\mathbb{R}^n$:

Theorem 20.2.2 — Second method to calculate the projection. Let U be a subspace of $\mathbb{R}^n$ and let A be a matrix such that $\mathrm{Col}\,(A) = U$. Then $\mathrm{proj}_U(\mathbf{b})$ is the vector $A\mathbf{x}$ where $\mathbf{x}$ is any solution to

the system
$$A^T A\mathbf{x} = A^T\mathbf{b}.$$

To prove it, just apply the argument in the example above.

Note that because we know that $\text{proj}_U(\mathbf{b})$ always exists, the system $A^T A\mathbf{x} = A^T\mathbf{b}$ will *always* be consistent!

20.3 Application: Best approximations to solutions of inconsistent systems

Corollary 20.3.1 — Best Approximation to a Solution. Suppose that $A\mathbf{x} = \mathbf{b}$ is inconsistent. Then the *best approximation* to a solution of this system is any vector $\mathbf{z}$ which solves

$$A^T A\mathbf{z} = A^T\mathbf{b}$$

That is, such a $\mathbf{z}$ minimizes $\|A\mathbf{z} - \mathbf{b}\|$.

Proof. Minimizing $\|A\mathbf{z} - \mathbf{b}\|$ over all $\mathbf{z} \in \mathbb{R}^n$ means finding the closest vector in $\text{Col}(A)$ to $\mathbf{b}$, since $\text{Col}(A)$ is exactly the set of vectors of the form $A\mathbf{z}$. So $A\mathbf{z}$ will be have to be $\text{proj}_{\text{Col}\,A}(\mathbf{b})$. This is found exactly as in the theorem above. ■

Problem 20.3.2 Find the best approximation to a solution for the inconsistent system $A\mathbf{x} = \mathbf{b}$ where $\mathbf{b} = (1,2,1,2)$ and

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

Solution Let's apply the corollary. We calculate

$$A^T A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

and

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

So we need to solve

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

which we do by row reducing the augmented matrix

$$\begin{bmatrix} 1 & 0 & 1 & | & 1 \\ 0 & 1 & 0 & | & 2 \\ 1 & 0 & 2 & | & 2 \end{bmatrix} \sim \cdots \sim \begin{bmatrix} 1 & 0 & 0 & | & 0 \\ 0 & 1 & 0 & | & 2 \\ 0 & 0 & 1 & | & 1 \end{bmatrix}$$

which has unique solution $(0,2,1)$. So we claim that

$$\mathbf{x} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$$

gives the best possible approximation to a solution to the system $A\mathbf{x} = \mathbf{b}$; we have

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

which isn't quite $\mathbf{b}$ but is fairly close.

To be really convinced this is the best, consider any other $\mathbf{v} = (x,y,z)$; then

$$A\mathbf{v} = (x+z, y, z, 0)$$

so

$$\|(1,2,1,2) - (x+z,y,z,0)\|^2 = (1-(x+z))^2 + (2-y)^2 + (1-z)^2 + 2^2$$

and the answer we found makes the first three terms 0, which clearly gives the minimum possible value (4) for the total distance squared.

20.4 Application: Least-Squares Best Fitting Line or Curve

Problem 20.4.1 Suppose you collect the following data

x	y
-2	18
-1	8.1
0	3.8
1	3.1
2	6.1

These data points don't exactly lie on a parabola, but you think that's experimental error; what is the best-fitting quadratic function through these points?

Solution This doesn't seem to be a linear problem, at first. Let's rephrase this as a problem in linear algebra!

We want to find a quadratic polynomial

$$y = a + bx + cx^2$$

so that when we plug in one of the x-values above, we get "close" to the y value we measured.

Well, what we're really saying is: solve for a, b, c (or at least, try to) in the following linear system:

$$\begin{aligned} a+b(-2)+c(-2)^2 &= 18 \\ a+b(-1)+c(-1)^2 &= 8.1 \\ a+b(0)+c(0)^2 &= 3.8 \\ a+b(1)+c(1)^2 &= 3.1 \\ a+b(2)+c(2)^2 &= 6.1\,. \end{aligned}$$

This is the matrix equation $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 18 \\ 8.1 \\ 3.8 \\ 3.1 \\ 6.1 \end{bmatrix}.$$

You can check that this is inconsistent.

So: we know from above that we should solve instead

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

where

$$A^T A = \begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix} \qquad \text{and} \qquad A^T\mathbf{b} = \begin{bmatrix} 39.1 \\ -28.8 \\ 107.6 \end{bmatrix}$$

which gives (after a bit of row reduction, and maybe the help of a calculator)

$$\mathbf{x} = \begin{bmatrix} 3.62 \\ -2.88 \\ 2.1 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

meaning our quadratic polynomial is

$$y = (3.62) - (2.88)x + (2.1)x^2.$$

(You'd need a graphing calculator to see if this is really the best possible fit.)

In what sense was this the "best answer"? We minimized the distance between $A\mathbf{x}$ and $\mathbf{b}$. Recall that if $\mathbf{x} = (a, b, c)$ then $A\mathbf{x}$ is the y-values for the points on the curve $a + bx + cx^2$ at the given x-values in the table (or, in the second column of A). Call these $(y_{-2}, y_{-1}, y_0, y_1, y_2)$. And $\mathbf{b}$ is the measured values in the table; call these $(y'_{-2}, y'_{-1}, y'_0, y'_1, y'_2)$. So we've minimized

$$\|A\mathbf{x} - \mathbf{b}\|^2 = (y_{-2} - y'_{-2})^2 + (y_{-1} - y'_{-1})^2 + (y_0 - y'_0)^2 + (y_1 - y'_1)^2 + (y_2 - y'_2)^2$$

that is, we've found the *least squares best fit*: the curve that minimizes the sum of the squares of the differences in y-values.

! This is the de facto standard for fitting curves to experimental data.

Summary of method:

1. Given: data points (x_i, y_i), $i = 1, \cdots, n$.
2. Goal: to find the best polynomial of degree m fitting these data points, say

$$p(x) = a_0 + a_1 x + \cdots + a_m x^m$$

3. Set

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \cdots & \cdots & \cdots & \ddots & \cdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}$$

4. Solve $(A^T A)\mathbf{x} = (A^T \mathbf{b})$ for the vector $\mathbf{x} = (a_0, a_1, \cdots, a_m)$ of coefficients of your best-fitting polynomial.

Problems

Problem 20.1 Suppose U is a subspace of $\mathbb{R}^n$.

a) Show that $U \cap U^\perp = \{\mathbf{v} \in \mathbb{R}^n \mid \mathbf{v} \in U \text{and } \mathbf{v} \in U^\perp\} = \{0\}$.

b) [a] Show that if $\mathbf{v} \in \mathbb{R}^n$ is any vector, then there are vectors $\mathbf{u} \in U$ and $\mathbf{w} \in U^\perp$ such that $\mathbf{v} = \mathbf{u} + \mathbf{w}$.

c) Show that if $\mathbf{v} = \mathbf{u} + \mathbf{w}$ and $\mathbf{v} = \mathbf{u}' + \mathbf{w}'$, with $\mathbf{u}, \mathbf{u}' \in U$ and $\mathbf{w}, \mathbf{w}' \in U^\perp$, then $\mathbf{u} = \mathbf{u}'$ and $\mathbf{w} = \mathbf{w}'$.

---

[a] Hint: look carefully at the seemingly useless equation $\mathbf{v} = \text{proj}_U(\mathbf{v}) + \mathbf{v} - \text{proj}_U(\mathbf{v})$.

V

Determinants, Eigen-everything and Diagonalization

We are almost ready to see one of the most useful and applicable parts of linear algebra: the eigenvalues and eigenvectors of square matrices. To this end, we need to first develop an important tool: the determinant of a square matrix.

21. Determinants

Photo: Ralph Nevins. Montpellier, France

We are about to change gears, and rather dramatically at that. Our goal: to uncover the basis for $\mathbb{R}^n$ — when it exists — that is best or ideally adapted to multiplication by a given $n \times n$ matrix. This is called a basis of *eigenvectors* (of the matrix), from the German prefix “eigen” meaning “my own". Eigen-something means “my own something", so eigenvectors of a matrix are the vectors that are special for that matrix. We will see that multiplication of an eigenvector of a matrix by that matrix gives[1] another eigenvector in the same direction.

To get there, though, we first need to master the calculation of the determinant of a square matrix — and see why it’s such an interesting number to calculate.[2]

21.1 The determinant of a square matrix

Let A be an $n \times n$ matrix.

When $n = 1$, $\det[a] = a$.

[1]Except in one case: when the *eigenvalue* is zero. See section 22.1.1.

[2]There is a formula which generalizes the formula we saw in 18.2.1 for a 2×2 matrix. For the moment, search for ‘Invertible matrix’ and ‘Adjugate matrix’ for the inverse of an invertible $n \times n$ square matrix you might see in a later course that uses the matrix of cofactors (where the *cofactors* are the terms $(-1)^{i+j}\det(A_{ij})$ that appeared in our formulae for the determinant). As such, it’s really inefficient for large matrices unless they are quite *sparse* (meaning: have lots of zeros).

There’s also a formula called *Cramer’s rule*, for the solution to $A\mathbf{x} = \mathbf{b}$ when A is invertible, and that was historically how mathematicians solved systems of linear equations. In practice one uses Gaussian elimination for any problem involving systems larger that 3×3, but even for 3×3 systems, using Cramer’s rule involves computing *four* 3×3 determinants. You’ll see later in this chapter why that’s best avoided!

It’s interesting that determinants were discovered long before row reduction (in the western world) and determinants were the main route to results in linear algebra until around 1900.

When $n = 2$, we have previously discussed the determinant; recall that

$$\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$$

and it's important for two things:

- algebraically: it's the denominator in the formula for the inverse of $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ — in fact A is invertible if and only if $\det(A) \neq 0$; and
- geometrically: its absolute value is the area of the parallelogram with sides defined by the vectors which are the rows of A.

When $n = 3$, we use the same type of formula as for the cross product:

$$\begin{aligned}\det\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} &= a\det\begin{bmatrix} e & f \\ h & i \end{bmatrix} - b\det\begin{bmatrix} d & f \\ g & i \end{bmatrix} + c\det\begin{bmatrix} d & e \\ g & h \end{bmatrix} \\ &= a(ei - fh) - b(di - fg) + c(dh - eg)\end{aligned}$$

Note that the end result is again simply a number, a scalar. The above formula is called the *Laplace or cofactor expansion along the first row*. Its importance includes:

- in Algebra: we'll soon show that $\det(A) \neq 0$ if and only if A is invertible (and, although we won't do it this term, there does exist a formula for the inverse of a matrix with denominator $\det(A)$); and
- in Geometry: recall the scalar triple product of vectors $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$, whose absolute value is the volume of the parallelipiped with sides defined by $\mathbf{u}, \mathbf{v}, \mathbf{w}$. Note that $\det(A)$ equals the scalar triple product of the row vectors of A.

For $n \geq 4$, the formula is recursive, meaning, you need to calculate determinants of smaller matrices, which in turn are calculated in terms of even smaller matrices, until finally the matrices are 2×2 and we can write down the answer.

Definition 21.1.1 Let $A = (a_{ij})$ (meaning: the (i, j) entry of A is denoted a_{ij}). Then

$$\det A = a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + \cdots + (-1)^{1+n} a_{1n}\det(A_{1n})$$

where A_{ij} is the $(n-1) \times (n-1)$ matrix obtained from A by deleting row i and column j.

■ **Example 21.1.2** Let's calculate $\det(A)$ where

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

Well, the formula says this is

$$\det A = 2\det\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} - 3\det\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} + 4\det\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} - 5\det\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Now to calculate each of the smaller ones, we use the formula again (although to save time and effort, let's not write down the subdeterminants when we're just going to multiply the term by 0):

$$\det\begin{bmatrix}0&0&1\\1&0&1\\0&1&0\end{bmatrix} = 0(\cdot)-0(\cdot)+1\det\begin{bmatrix}1&0\\0&1\end{bmatrix} = 1(1-0)=1$$

$$\det\begin{bmatrix}1&0&1\\0&0&1\\0&1&0\end{bmatrix} = 1\det\begin{bmatrix}0&1\\1&0\end{bmatrix}-0(\cdot)+1\det\begin{bmatrix}0&0\\0&1\end{bmatrix} = 1(0-1)+1(0-0)=-1$$

$$\det\begin{bmatrix}1&0&1\\0&1&1\\0&0&0\end{bmatrix} = 1\det\begin{bmatrix}1&1\\0&0\end{bmatrix}-0(\cdot)+1\det\begin{bmatrix}0&1\\0&0\end{bmatrix} = 0$$

$$\det\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix} = 1\det\begin{bmatrix}1&0\\0&1\end{bmatrix}-0+0=1$$

so our final answer is

$$\det(A) = 2(1)-3(-1)+4(0)-5(1)=0$$

which is rather interesting. Do notice that the matrix A is not invertible: equivalently, A has linearly dependent rows ($R_1 = 2R_2+3R_3+4R_4$); has linearly dependent columns ($C_4 = C_1+C_2+C_3$); has nontrivial nullspace; has rank less than 4, *etc.*). ■

Now, this formula is all well and good, but utterly torturous. To calculate the determinant of a 3×3 matrix there are $6 = 3!$ products of 3 terms, so $2\times 3!$ products (just counting the hard part, multiplying). For a 4×4 matrix, it's $3\times 4\times 6 = 3\times 4! = 72$ products that you have to add together. So for a 10×10 matrix, there are

$$9\times 10! = 32,659,200$$

terms. For a 100×100 matrix, you're looking at around 9×10^{159} terms. Have fun seeing how long this would take to calculate if you used the current[3] fastest computer (Fugaku 415-PFLOPS), which can do an absolutely astonishing 415.5×10^{15} calculations per second. (Hint: there are only about $\pi\times 10^7$ seconds in a year....)

Luckily (and remarkably) there are some wonderful shortcuts.

21.2 First shortcut: expansion along ANY column or row

Theorem 21.2.1 — Cofactor expansion along any row or column. Let $A = (a_{ij})$. Then the determinant can be calculated via the *cofactor expansion along row* i, for any i:

$$\det(A) = (-1)^{i+1}a_{i1}\det(A_{i1}) + (-1)^{i+2}a_{i2}\det(A_{i2}) + \cdots + (-1)^{i+n}a_{in}\det(A_{in})$$

[3] As of (about) 2020: see See https://en.wikipedia.org/wiki/Fugaku_(supercomputer)

and similarly, the determinant equals the *cofactor expansion along column j*, for any j:

$$\det(A) = (-1)^{1+j} a_{1j} \det(A_{1j}) + (-1)^{2+j} a_{2j} \det(A_{2j}) + \cdots + (-1)^{n+j} a_{nj} \det(A_{nj}).$$

! Tip: the signs you need to use (that is, the factors $(-1)^{i+j}$ in the cofactor expansion) are given by:

$$\begin{bmatrix} + & - & + & - & + & \cdots \\ - & + & - & + & - & \cdots \\ + & - & + & - & + & \cdots \\ - & + & - & + & - & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

■ **Example 21.2.2** Calculate

$$\det \begin{bmatrix} 2 & 3 & 4 \\ 1 & 0 & 3 \\ 2 & 0 & 4 \end{bmatrix}$$

We could do this on the first row:

$$\det(A) = 2\det \begin{bmatrix} 0 & 3 \\ 0 & 4 \end{bmatrix} - 3\det \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} + 4\det \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix} = 2(0) - 3(4-6) + 4(0) = 6$$

or, how about we pick the row or column with the most zeros? Like column 2:

$$\det(A) = -3\det \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} + 0 - 0 = 6$$

which is certainly faster. ■

! Moral: choose the row or column with the most zeros to do your cofactor expansion.

This theorem gives us a couple of immediate consequences:

Proposition 21.2.3 — Quick properties of the determinant. Let A be an $n \times n$ matrix.

(1) If A has a row or column of zeros, then $\det(A) = 0$.
(2) $\det(A) = \det(A^T)$.

Proof. (1) Do the cofactor expansion along that row or column; all the terms are zero.

(2) Doing the expansion along row 1 of A is the same calculation as doing the expansion along column 1 of A^T, and by the theorem you get the same answer. ■

■ **Example 21.2.4** Calculate

$$\det \begin{bmatrix} 2 & 2 & 4 & 7 & 6 \\ 0 & -3 & 7 & 1 & 3 \\ 0 & 0 & 1 & 12 & -8 \\ 0 & 0 & 0 & -2 & 21 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix}.$$

This is easy if you do cofactor expansion on the first column:

$$\begin{aligned} \det A &= 2\det \begin{bmatrix} -3 & 7 & 1 & 3 \\ 0 & 1 & 12 & -8 \\ 0 & 0 & -2 & 21 \\ 0 & 0 & 0 & 3 \end{bmatrix} \\ &= 2(-3)\det \begin{bmatrix} 1 & 12 & -8 \\ 0 & -2 & 21 \\ 0 & 0 & 3 \end{bmatrix} \\ &= 2(-3)(1)\det \begin{bmatrix} -2 & 21 \\ 0 & 3 \end{bmatrix} \\ &= 2(-3)(1)(-2)(3) \end{aligned}$$

which is simply the product of the diagonal entries! ■

Proposition 21.2.5 — Determinant of triangular matrices. The determinant of a triangular matrix is the product of the diagonal entries.

Hmm... notice that every matrix in RREF is triangular, so this proposition applies. Thinking about the possibilities, we conclude: if A is in RREF then either

- A has rank n and the determinant is one;
- A has rank less than n and the determinant is zero.

Very nice. Except: does this dash our hopes of being able to calculate the determinant using row reduction?

21.3 Second Shortcut: Using row reduction to calculate the determinant

It turns out you *can* use row reduction, as long as you keep track of your steps. The miracle is that our favourite, most useful, row reduction step doesn't change the determinant at all!

Theorem 21.3.1 — Effect of row reduction on the determinant. Let A be an $n \times n$ matrix and suppose you do an elementary row operation and the answer is the matrix B. Then:

(1) If the row operation was *interchange two rows* then $\det(B) = -\det(A)$.
(2) If the row operation was *multiply a row by r* then $\det(B) = r\det(A)$.
(3) If the row operation was *add a multiple of one row to another* then $\det(B) = \det(A)$.

Remark 21.3.2. *The theorem above remains true if 'row' is replaced by 'column'.*

Problem 21.3.3 Calculate $\det(A)$ where

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

Solution Let's row reduce A and keep track of our operations:

$$\begin{bmatrix} 2 & 1 & 3 & 5 \\ 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 3 & 1 & 4 \end{bmatrix} \begin{array}{c} R_1 \leftrightarrow R_2 \\ \sim \end{array} \begin{bmatrix} 1 & 2 & 3 & 1 \\ 2 & 1 & 3 & 5 \\ 0 & 1 & 2 & 3 \\ 0 & 3 & 1 & 4 \end{bmatrix} \begin{array}{c} -2R_1 + R_2 \\ \sim \\ -3R_3 + R_4 \end{array} \begin{bmatrix} 1 & 2 & 3 & 1 \\ 0 & -3 & -3 & 3 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & -5 & -5 \end{bmatrix}$$

$$\begin{array}{c} R_2 \leftrightarrow R_3 \\ \sim \\ -\frac{1}{3}R_3 \\ -\frac{1}{5}R_4 \end{array} \begin{bmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{array}{c} -R_2 + R_3 \\ \sim \end{array} \begin{bmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & -1 & -4 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{array}{c} \sim \\ R_3 + R_4 \end{array} \begin{bmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & -1 & -4 \\ 0 & 0 & 0 & -3 \end{bmatrix} = R.$$

OK: now let's see what our operations did. The last matrix R has determinant 3 (by multiplying the diagonal entries). The only operations that changed this value are row interchanges (each at a cost of -1) and scaling rows (each at a cost of the factor you multiplied by). So:

$$\det(R) = (-1)(-\frac{1}{3})(-\frac{1}{5})(-1)\det(A)$$

which gives

$$\det(A) = 15\det(R) = 45.$$

You can check this by cofactor expansion.

! Point: if you've row reduced and kept track of your operations, then the determinant can be computed with no extra work.

21.4 Properties of determinants

Theorem 21.4.1 — Properties of the determinant. Let A and B be $n \times n$ matrices. Then

(1) $\det(rA) = r^n \det(A)$;
(2) $\det(AB) = \det(A)\det(B)$;
(3) $\det(A) = 0$ if and only if A is not invertible;
(4) if A is invertible then $\det(A^{-1}) = \frac{1}{\det(A)}$.

Proof. (1) Multiplying the whole matrix by r means multiplying each of the n rows of A by r; and each of these costs a factor of r. (Geometrically: the n-dimensional volume of a box goes up by a factor of r^n if you scale every edge by a factor of r.)

(2) This takes some effort to prove, but isn't beyond us at all[4].

[4] For a reference, see Brescther, '*Linear Algebra with Applications.*'

(3) We've just seen that row reducing can change the value of the determinant by a *nonzero* factor. So $\det(A) = 0$ if and only if $\det(R) = 0$, where R is the RREF of A. And as we've remarked previously: $\det(R) = 0$ if and only if the rank of A is less than n, which happens if and only if A is not invertible.

(4) If A is invertible then $AA^{-1} = I$ so by part (2), we have $\det(A)\det(A^{-1}) = \det(I) = 1$. ■

Problems

Remarks:

1. A question with an asterisk '*' (or two) indicates a bonus-level question.
2. You must justify all your responses.

Problem 21.1 Find the determinant of each of the following matrices. In the following, λ represents a variable. Use appropriate row and/or column operations where useful: the definition is the tool of last resort!

a) $\begin{bmatrix} 2 & -1 \\ 3 & 0 \end{bmatrix}$

b) * $\begin{bmatrix} 2 & -1 & 3 \\ 3 & 0 & -5 \\ 1 & 1 & 2 \end{bmatrix}$

c) $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 0 & 1 & 1 \end{bmatrix}$

d) * $\begin{bmatrix} 3 & 4 & -1 \\ 1 & 0 & 3 \\ 2 & 5 & -4 \end{bmatrix}$

e) $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$

f) * $\begin{bmatrix} \lambda - 6 & 0 & 0 \\ 0 & \lambda & 3 \\ 0 & 4 & \lambda + 4 \end{bmatrix}$

g) $\begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \end{bmatrix}$

h) * $\begin{bmatrix} -\lambda & 2 & 2 \\ 2 & -\lambda & 2 \\ 2 & 2 & -\lambda \end{bmatrix}$

Problem 21.2 If $\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = 3$, find

a) $\begin{vmatrix} d & e & f \\ a & b & c \\ g & h & i \end{vmatrix}$

b) * $\begin{vmatrix} b & a & c \\ e & d & f \\ h & g & i \end{vmatrix}$

c) $\begin{vmatrix} b & 3a & c \\ e & 3d & f \\ h & 3g & i \end{vmatrix}$

d) * $\begin{vmatrix} b & 3a & c-4b \\ e & 3d & f-4e \\ h & 3g & i-4h \end{vmatrix}$

e) $\begin{vmatrix} 4g & a & d-2a \\ 4h & b & e-2b \\ 4i & c & f-2c \end{vmatrix}$

Problem 21.3 a) If Q is a 3×3 matrix and $\det Q = 2$, find $\det((3Q)^{-1})$.

b) *If B is a 4×4 matrix and $\det(2BB^T) = 64$, find $|\det(3B^2B^T)|$.

c) If A and B are 4×4 matrices with $\det(A) = 2$ and $\det(B) = -1$, find $\det(3AB^TA^{-2}BA^TB^{-1})$.

d) *Compute the determinant of $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 9 & 10 \\ 11 & 12 \end{bmatrix} \begin{bmatrix} 13 & 14 \\ 15 & 16 \end{bmatrix}$.

e) Find all x so that the matrix $\begin{bmatrix} 0 & x & -4 \\ 2 & 3 & -2 \\ 1 & 4 & 1 \end{bmatrix}$ is not invertible.

Problem 21.4 State whether each of the following is (always) true, or is (possibly) false.

- If you say the statement may be false, you must give an explicit example.
- If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

In the following A and B are $n \times n$ matrices (with $n > 1$) and k is a scalar.

a) $\det(AB) = \det(A)\det(B)$

b) $^\star\det(A+B) = \det(A) + \det(B)$

c) $\det(kA) = k\det(A)$

d) $^\star\det(kA) = k^n\det(A)$

e) $\det A^T = \det(A)$

f) *If A and B are the same except the first row of A is twice the first row of B, then $\det(A) = 2\det(B)$

g) ** If $A = \begin{bmatrix} \mathbf{c}_1 + \mathbf{b}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{bmatrix}$ is written in block column form, meaning that the first column of A is $\mathbf{c}_1 + \mathbf{b}_1$ and $\mathbf{c}_2, \mathbf{c}_3, \ldots, \mathbf{c}_n$ are the other columns of A, then $\det(A) = \det\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{bmatrix} + \det\begin{bmatrix} \mathbf{b}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{bmatrix}$

Problem 21.5 a) If A is any 2 by 2 matrix, and $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ and $\mathbf{v}_4$ are column vectors in $\mathbb{R}^3$ satisfying

$$\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = A\begin{bmatrix} \mathbf{v}_3 & \mathbf{v}_4 \end{bmatrix},$$

show that $\mathbf{v}_1 \times \mathbf{v}_2 = (\det(A))\,\mathbf{v}_3 \times \mathbf{v}_4$.

b) *If $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$ are vectors in $\mathbb{R}^3$, use properties of 3 by 3 determinants to show that

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

c) Suppose B is an $1 \times n$, matrix D is an $n \times n$ matrix, $a \in \mathbb{R}$. Show that $\det\begin{bmatrix} a & B \\ 0 & D \end{bmatrix} = a\det(D)$. (The matrix is expressed here in block form.)

d) ^{*a} Suppose D is an $n \times n$ matrix and B an $m \times n$ matrix. Show that $\det\begin{bmatrix} I_m & B \\ 0 & D \end{bmatrix} = \det(D)$. (The matrix $\begin{bmatrix} I_m & 0 \\ 0 & B \end{bmatrix}$, of size $(m+n) \times (m+n)$, is expressed here in block form.)

e) ^{*b} Suppose A is an $m \times m$ matrix. Show that $\det\begin{bmatrix} A & B \\ 0 & I_n \end{bmatrix} = \det(A)$.

f) Suppose A, B and D are respectively of sizes $m \times m$, $m \times n$ and $n \times n$. Noting that $\begin{bmatrix} A & B \\ 0 & D \end{bmatrix} = \begin{bmatrix} I_m & 0 \\ 0 & D \end{bmatrix}\begin{bmatrix} A & B \\ 0 & I_n \end{bmatrix}$, show that $\det\begin{bmatrix} A & B \\ 0 & D \end{bmatrix} = \det(A)\det(D)$.

g) *Suppose A, B, C and D are respectively of sizes $m \times m$, $m \times n$, $n \times m$ and $n \times n$. Suppose that D is invertible. Noting that $\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} I_m & 0 \\ -D^{-1}C & I_n \end{bmatrix} = \begin{bmatrix} A - BD^{-1}C & B \\ 0 & D \end{bmatrix}$, show that $\det\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A - BD^{-1}C)\det(D)$.

h) Suppose A, B, C and D are $n \times n$ matrices with D invertible and $CD = DC$. Let $E = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$, an $(2n) \times (2n)$ matrix, expressed here in block form. Use the last part and properties of the determinant to show that $\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(AD - BC)$.

i) $^{***\ c}$ Suppose A, B, C and D are $n \times n$ matrices with $CD = DC$. Show that $\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(AD - BC)$.

j) ** Suppose A is a 3 by 3 matrix that satisfies $AA^T = I_3$.

   i) Show that $A^T A = I_3$ as well.

   ii) Suppose $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is the standard (orthogonal and orthonormal) ordered basis of $\mathbb{R}^3$. Using the fact that the dot product $\mathbf{v} \cdot \mathbf{w}$ is equal to the matrix product $\mathbf{v}^T \mathbf{w}$ (writing vectors as 3×1 matrices), show that $\{A\mathbf{e}_1, A\mathbf{e}_2, A\mathbf{e}_3\}$ is also an orthogonal set, which is indeed *orthonormal.*

   iii) If $\mathbf{u}$ and $\mathbf{v}$ are any vectors in $\mathbb{R}^3$, use the Expansion theorem 19.2.5 to show that

$$\mathbf{u} \times \mathbf{v} = \sum_{i=1}^{3} (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{e}_i$$

   and

$$A\mathbf{u} \times A\mathbf{v} = \sum_{i=1}^{3} (A\mathbf{u} \times A\mathbf{v}) \cdot A\mathbf{e}_i.$$

   iv) Recalling the definition of the 3 by 3 determinant, show that

$$(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{e}_i = \det \begin{bmatrix} \mathbf{u} & \mathbf{v} & \mathbf{e}_i \end{bmatrix}$$

   and

$$(A\mathbf{u} \times A\mathbf{v}) \cdot A\mathbf{e}_i = \det \begin{bmatrix} A\mathbf{u} & A\mathbf{v} & A\mathbf{e}_i \end{bmatrix},$$

   where the matrices $\begin{bmatrix} \mathbf{u} & \mathbf{v} & \mathbf{e}_i \end{bmatrix}$ and $\begin{bmatrix} A\mathbf{u} & A\mathbf{v} & A\mathbf{e}_i \end{bmatrix}$ on the right have been written in block column form.

   v) Recalling what you know about block multiplication, and properties of determinants, show that

$$\det \begin{bmatrix} A\mathbf{u} & A\mathbf{v} & A\mathbf{e}_i \end{bmatrix} = \det(A) \det \begin{bmatrix} \mathbf{u} & \mathbf{v} & \mathbf{e}_i \end{bmatrix}.$$

   vi) Now prove that for any vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$, (when $AA^T = I_3$)

$$A\mathbf{u} \times A\mathbf{v} = \det(A) (\mathbf{u} \times \mathbf{v})$$

vii) Give an example of a matrix A, which does not satisfy $AA^T = I_3$ and two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ for which $A\mathbf{u} \times A\mathbf{v} \neq \det(A)\,(\mathbf{u} \times \mathbf{v})$

[a] This is for those of you who know about *proofs by induction*. Search for 'Mathematical induction', for example.

[b] Use the same technique as in the previous part.

[c] This is for those of you who know about *continuity*, and the fact that invertible $n \times n$ matrices are *dense* in all $n \times n$ matrices. Search for 'Invertible_matrix'. There is another proof of this identity that doesn't use the density argument, by J.R. Silvester, in *Determinants of block matrices*, Math. Gaz., 84(501) (2000), pp. 460-467.

22. Eigenvalues and Eigenvectors

This chapter and the next are amongst two of the most useful and applicable parts of all of linear algebra! Get ready for a great ride.

22.1 Eigenvalues and Eigenvectors of a (square) matrix

Definition 22.1.1 Let A be an $n \times n$ matrix. If $\lambda \in \mathbb{R}$ is a scalar and $\mathbf{x} \in \mathbb{R}^n$ is a *nonzero* vector such that

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

then $\mathbf{x}$ is called an *eigenvector of* A and λ is its corresponding *eigenvalue*.

■ **Example 22.1.2** Let $A = \begin{bmatrix} 3 & -1 \\ -2 & 2 \end{bmatrix}$. Then:

- 1 is an eigenvalue and $(1,2)$ is an eigenvector corresponding to $\lambda = 1$ because

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

- 4 is an eigenvalue and $(-1,1)$ is an eigenvector corresponding to $\lambda = 4$ because

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

Note that $(10,20)$ and $(17,-17)$ are also eigenvectors of A; but they are just multiples of what we already found. ■

■ **Example 22.1.3** Let $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$. Then 0 is an eigenvalue, and $(1,-1)$ is a corresponding eigenvector, since $A \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \mathbf{0} = 0 \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. ■

! The matrix A can have 0 as an *eigenvalue* but the vector $\mathbf{0}$ is NEVER an *eigenvector*.

The essential application: suppose you can find a basis $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ for $\mathbb{R}^n$ consisting of eigenvectors of A, with corresponding eigenvalues $\lambda_1, \cdots, \lambda_n$. Then if $\mathbf{v} = a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n$ then

$$\begin{aligned} A\mathbf{v} &= A(a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n) \\ &= A(a_1\mathbf{x}_1) + \cdots + A(a_n\mathbf{x}_n) \\ &= a_1 A\mathbf{x}_1 + \cdots + a_n A\mathbf{x}_n \\ &= a_1\lambda_1\mathbf{x}_1 + \cdots + a_n\lambda_n\mathbf{x}_n \\ &= \lambda_1 a_1\mathbf{x}_n + \cdots + \lambda_n a_n\mathbf{x}_n. \end{aligned}$$

So although the answer won't be a scalar multiple of $\mathbf{v}$ (since the λ_i may be different), the point is: given a vector $\mathbf{v}$ in coordinates relative to an eigenbasis, calculating $A\mathbf{v}$ doesn't require us to multiply matrices at all — we just have to scale each of the coordinates (coefficients) by the appropriate λ_i.

(For those of you thinking in terms of computarional complexity: this means going from $O(n^2)$ to $O(n)$ — fairly significant!)

22.2 Finding eigenvalues of A

Since both λ and $\mathbf{x}$ are unknown, the equation $A\mathbf{x} = \lambda\mathbf{x}$ is in fact nonlinear. As we've seen previously, however, we can sometimes turn nonlinear equations into linear ones by focussing on solving for one set of variables at a time.

The *key* is the following chain of equalities, which turns $A\mathbf{x} = \lambda\mathbf{x}$ into the equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$:

$$\begin{aligned} & A\mathbf{x} = \lambda\mathbf{x} \\ \iff & A\mathbf{x} = \lambda I\mathbf{x} \\ \iff & A\mathbf{x} - \lambda I\mathbf{x} = \mathbf{0} \\ \iff & (A - \lambda I)\mathbf{x} = \mathbf{0} \end{aligned}$$

This point of view is the key to letting us solve the problem using our methods from before, as follows.

First think of λ as being a fixed, known quantity. Then this equation is a homogeneous system of linear equations, and any *nontrivial* solution is an eigenvector of A. In other words:

- if this system has a unique solution, then there are no eigenvectors corresponding to λ, which means λ is not an eigenvalue.
- if this system has infinitely many solutions, then any nontrivial one of these is an eigenvector, and so λ is an eigenvalue.

We can rephrase this as follows, by remembering that when A is a square matrix, having infinitely many solutions to a linear system is the same as being not invertible:

! The number λ is an eigenvalue of A if and only if the matrix $A - \lambda I$ is NOT invertible.

We could use many different methods to check if the matrix is not invertible, but since the matrix has a variable in it (namely, λ) the nicest one is the determinant condition:

! The number λ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$.

Definition 22.2.1 The polynomial $\det(A - \lambda I)$ is called the *characteristic polynomial* of A[a].

[a] Some authors prefer $\det(\lambda I - A)$, but for our computational purposes, using $\det(A - \lambda I)$ is less prone to error, and gives exactly the same answer, up to a sign.

Problem 22.2.2 Find all the eigenvalues of $A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$.

Solution We need to find all values of λ so that $A - \lambda I$ is not invertible. Begin by writing down $A - \lambda I$:

$$A - \lambda I = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 1-\lambda & 3 \\ 3 & 1-\lambda \end{bmatrix}$$

(watch those negative signs on the two λs!).

We calculate the characteristic polynomial:

$$\det(A - \lambda I) = (1-\lambda)(1-\lambda) - 9 = \lambda^2 - 2\lambda - 8 = (\lambda - 4)(\lambda + 2),$$

and deduce that this is zero iff λ is either 4 or -2. Hence, the eigenvalues of A are 4 and -2.

To check, let's write down $A - 4I$ and $A + 2I$ and make sure they are in fact non-invertible:

$$A - 4I = \begin{bmatrix} -3 & 3 \\ 3 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}, \qquad A + 2I = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

Good! Because both these matrices are not invertible, it's now clear that 4 and -2 are eigenvalues of A. To see that these are ALL the eigenvalues of A, note that $\det(A - \lambda I)$ is a quadratic polynomial and so has *at most* two roots.

Key for calculations: The eigenvalues of A are the roots of the characteristic polynomial $\det(A - \lambda I)$.

Definition 22.2.3 The multiplicity of λ as a root of the characteristic polynomial is called its *algebraic multiplicity*.

■ **Example 22.2.4** The characteristic polynomial of $\begin{bmatrix} 16 & 2 & 17 \\ 0 & -43 & 14 \\ 0 & 0 & 16 \end{bmatrix}$ is $-(\lambda - 16)^2(\lambda + 43)$. (Check!) Hence the eigenvalues of this matrix are $\lambda = -43$ (with algebraic multiplicity 1) and $\lambda = 16$ (with algebraic multiplicity 2). ■

22.3 Finding the eigenvectors of A

Suppose now that you have found some eigenvalues of the matrix A. To find the eigenvectors of A associated to that eigenvalue, recall that they are all the nontrivial solutions $\mathbf{x}$ of

$$(A-\lambda I)\mathbf{x}=\mathbf{0}.$$

In other words: the eigenvectors of A asssociated to λ are the nonzero vectors in $\ker(A-\lambda I) = \mathrm{Null}(A-\lambda I)$.

Definition 22.3.1 Let λ be an eigenvalue of A. Then the subspace

$$\begin{aligned} E_\lambda &= \mathrm{Null}(A-\lambda I) \\ &= \{\mathbf{x}\in\mathbb{R}^n \mid (A-\lambda I)\mathbf{x}=\mathbf{0}\} \\ &= \{\mathbf{x}\in\mathbb{R}^n \mid A\mathbf{x}=\lambda\mathbf{x}\} \end{aligned}$$

is called the λ-eigenspace of A. Its *nonzero* elements are the eigenvectors of A corresponding to the eigenvalue λ.

So rather than ask: "What are the eigenvectors of A?" it's simpler to ask: "What is a basis for each eigenspace of A?".

Problem 22.3.2 Find a basis for each eigenspace of

$$A=\begin{bmatrix}1 & 3\\ 3 & 1\end{bmatrix}$$

(recalling that its eigenvalues were 4 and -2).

Solution For the eigenvalue $\lambda=4$:

$$A-\lambda I = A-4I = \begin{bmatrix}-3 & 3\\ 3 & -3\end{bmatrix} \sim \begin{bmatrix}1 & -1\\ 0 & 0\end{bmatrix}$$

so the general solution to $(A-4I)\mathbf{x}=\mathbf{0}$ is

$$E_4=\mathrm{Null}(A-4I)=\{(r,r)\mid r\in\mathbb{R}\}$$

and a basis for E_4 is

$$\{(1,1)\}.$$

(Notice that with $\mathbf{x}=(1,1)$, $A\mathbf{x}=4\mathbf{x}$, as required.)

For the eigenvalue $\lambda=-2$:

$$A-\lambda I = A+2I = \begin{bmatrix}3 & 3\\ 3 & 3\end{bmatrix} \sim \begin{bmatrix}1 & 1\\ 0 & 0\end{bmatrix}$$

so the general solution to $(A-2I)\mathbf{x}=\mathbf{0}$ is

$$E_{-2}=\mathrm{Null}(A+2I)=\{(-r,r)\mid r\in\mathbb{R}\}$$

and a basis for E_{-2} is
$$\{(-1,1)\}.$$
(Notice that with $\mathbf{x} = (-1,1)$, $A\mathbf{x} = -2\mathbf{x}$, as required.)

The dimension of the eigenspace associated to the eigenvalue λ is another important number we want to keep track of.

Definition 22.3.3 The dimension of the eigenspace E_λ is called the *geometric multiplicity* of λ.

22.4 Towards an eigenvector basis of $\mathbb{R}^n$

We notice that so far, the eigenvectors we are getting for different eigenvalues are linearly independent. That's not too hard to see for two vectors, but takes a little more work for more.[1]

Theorem 22.4.1 — Independence of Eigenvectors — Distinct Eigenvalues. Let A be an $n \times n$ matrix. Then any set consisting of eigenvectors of A corresponding to distinct eigenvalues is linearly independent.

Let's put this together:

- The characteristic polynomial of an $n \times n$ matrix A is a polynomial of degree n, so it has at most n distinct roots.
- Each root of the characteristic polynomial is an eigenvalue of A.
- Each eigenvalue of A gives an eigenspace of dimension at least 1.
- So IF the characteristic polynomial has exactly n distinct roots, then we can choose one eigenvector from each eigenspace and thus produce a basis of $\mathbb{R}^n$ consisting entirely of eigenvectors of A.

Definition 22.4.2 The $n \times n$ matrix A is said to be *diagonalizable over the reals* if there is a basis of $\mathbb{R}^n$ consisting entirely of eigenvectors of A.

(Why "diagonalizable", you say? We'll come back to this in a bit.)

FACT If an $n \times n$ matrix A has n distinct *real* eigenvalues then A is diagonalizable.

Can a matrix ever fail to be diagonalizable?

22.5 Problematic cases: Case I (not enough real roots)

■ **Example 22.5.1** What are the eigenvalues of
$$A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}?$$

[1] The best way is to use *Induction*.

Well: the characteristic polynomial is

$$\begin{aligned}\det(A-\lambda I) &= \det\begin{bmatrix}\cos\theta-\lambda & \sin\theta \\ -\sin\theta & cos\theta-\lambda\end{bmatrix} \\ &= (\lambda-\cos\theta)^2+\sin^2\theta \\ &= \lambda^2-2\cos(\theta)\lambda+\cos^2\theta+\sin^2\theta \\ &= \lambda^2-2\cos(\theta)\lambda+1.\end{aligned}$$

Its roots, using the quadratic formula, are:

$$\lambda = \frac{1}{2}\left(2\cos\theta \pm \sqrt{4\cos^2\theta-4}\right) = \cos\theta \pm \sqrt{-\sin^2\theta} = \cos\theta \pm i\sin\theta$$

which are complex numbers unless $\sin\theta = 0$, *i.e.*, $\theta = 0$ or π.

If we were to calculate the eigenvectors associated to these eigenvalues for $\theta \neq 0,\pi$, they would have complex coordinates — meaning they aren't in $\mathbb{R}^2$, but in $\mathbb{C}^2$. That's fine for certain types of applications but not for others; mathematically we usually work over $\mathbb{C}$ (because the fundamental theorem of algebra tells us that's where all the roots will lie)[2].

If you happen to know that $A = \begin{bmatrix}\cos\theta & \sin\theta \\ -\sin\theta & \cos\theta\end{bmatrix}$ is a rotation matrix, it's pretty clear why (when $\theta \neq 0,\pi$) there are no real values of λ for which rotation of a vector $\mathbf{x} \in \mathbb{R}^2$ by θ, namely $A\mathbf{x}$, will result in a real multiple $\lambda\mathbf{x}$ of $\mathbf{x}$! ■

! Moral: it can happen that some eigenvalues are complex numbers. In this case, you will not get an eigenvector basis of $\mathbb{R}^n$; A is not diagonalizable over $\mathbb{R}$.

22.6 Problematic cases: Case II (not "enough" eigenvectors)

By definition, if λ is an eigenvalue then $\dim(E_\lambda) \geq 1$. But it λ is a repeated root of the characteristic equation, say of multiplicity k, then we'd really need $\dim(E_\lambda) = k$, and that doesn't always happen:

Problem 22.6.1 Find the eigenvalues of $A = \begin{bmatrix}3 & 1 \\ 0 & 3\end{bmatrix}$ and a basis for each eigenspace.

Solution The characteristic polynomial is $(\lambda-3)^2$ so the only eigenvalue is 3 (with algebraic multiplicity 2). But $\text{rank}(A-3I) = 1$ so $\dim E_2 = \dim \text{Null}(A-3I) = 2-1 = 1$. That is, we only got a 1-dimensional subspace of eigenvectors, so it's impossible to find an eigenvector basis of $\mathbb{R}^n$. A is not diagonalizable over the reals (or even over $\mathbb{C}$).

! If the characteristic polynomial has a repeated root, then it *can happen* that we do not have enough linearly independent eigenvectors of A to form a basis.

[2]Indeed, if we allow complex scalars, and so complex eigenvalues and complex eigenvectors in $\mathbb{C}^2$, $\mathbb{C}^2$ does have a basis of eigenvectors of the matrix A. In this case, we say A is *diagonalizable over* $\mathbb{C}$.

■ **Example 22.6.2** Consider

$$A = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{bmatrix}$$

Find all eigenvalues and a basis for each eigenspace and decide if A is diagonalizable.[3]

You'll find that A is in fact diagonalizable, even though its characteristic polynomial has a repeated root. ■

These examples illustrate a fundamental fact.

Theorem 22.6.3 — Limits of geometric multiplicity. Let λ be an eigenvalue of A. Then the geometric multiplicity of λ is at least 1, and at most equal to the algebraic multiplicity of λ. That is,

$$1 \leq \dim E_\lambda \leq \text{algebraic multiplicity of } \lambda$$

Note that the first inequality follows directly from the definition: λ being an eigenvalue implies that the eigenspace is nonzero (and so must have dimension at least 1). (Showing the other inequality is more work and we will not prove it here.)

Problems

Remarks:

1. A question with an asterisk '*' (or two) indicates a bonus-level question.
2. You must justify all your responses.

Problem 22.1 Find the eigenvalues of the following matrices.

a) $\begin{bmatrix} 2 & 1 \\ 3 & 0 \end{bmatrix}$

b) * $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 2 \end{bmatrix}$

c) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & 1 & 2 \end{bmatrix}$

d) * $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$

e) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 2 & 1 & 2 \end{bmatrix}$

[3] This example is done in detail in Section 23.1.

f) * $\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$

g) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & 1 & 2 \end{bmatrix}$

h) * $\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

i) $\begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

Problem 22.2 For each of the matrices in the previous question, find a basis for each eigenspace. (Solutions to parts b, d, f and h are available.)

23. Diagonalizability

In the previous chapter, we defined the eigenvalues and eigenvectors of a square matrix A. To each eigenvalue we associate its *algebraic multiplicity*, which is its multiplicity as a root of the characteristic polynomial $\det(A-\lambda I)$ of A. We also associate to each eigenvalue its *geometric multiplicity*, which is the dimension of the associated eigenspace E_λ. The key result (Theorem 22.6.3) is that for each eigenvalue λ of A, we have

$$1 \leq \text{geometric multiplicity of } \lambda \leq \text{algebraic multiplicity of } \lambda.$$

In particular, the dimension of the eigenspace E_λ can *never* be larger than the algebraic multiplicity of λ.

23.1 About diagonalizability

Recall Defintion 22.4.2: The $n \times n$ matrix A is said to be *diagonalizable over the reals* if there is a basis of $\mathbb{R}^n$ consisting entirely of eigenvectors of A.

We saw last time that if A has n distinct real eigenvalues, then A is diagonalizable. But A can be diagonalizable even if it has repeated eigenvalues (it's just not *guaranteed* to work):

Problem 23.1.1 Consider

$$A = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{bmatrix}$$

Find all eigenvalues and a basis for each eigenspace and decide if A is diagonalizable.

Solution We need to calculate the eigenvalues of A first. So we begin by computing the characteristic polynomial:

$$\det(A-\lambda I) = \det\begin{bmatrix} 3-\lambda & 2 & 4 \\ 2 & -\lambda & 2 \\ 4 & 2 & 3-\lambda \end{bmatrix}$$

Although we could compute this directly, it's easier with some zeros, so try the row operation $-2R_2 + R_3 \to R_3$, which doesn't change the value of the determinant:

$$= \det\begin{bmatrix} 3-\lambda & 2 & 4 \\ 2 & -\lambda & 2 \\ 0 & 2+2\lambda & -1-\lambda \end{bmatrix}$$

and now we notice that the $(3,2)$ entry is exactly -2 times the $(3,3)$ entry. So, remembering that the determinant of a matrix is the same as that of its transpose, we can perform the column operation $2C_3 + C_2 \to C_2$ and this won't change the determinant either:

$$= \det\begin{bmatrix} 3-\lambda & 10 & 4 \\ 2 & 4-\lambda & 2 \\ 0 & 0 & -1-\lambda \end{bmatrix}$$

and now if we do the cofactor expansion on the third row, we'll get a partially factored answer:

$$-(\lambda+1)\left((3-\lambda)(4-\lambda)-20\right) = -(\lambda+1)(\lambda^2-7\lambda-8) = -(\lambda+1)^2(\lambda-8)$$

This is the characteristic polynomial of A.

Now the eigenvalues of A are the roots: -1 is an eigenvalue, of algebraic multiplicity 2; and 8 is an eigenvalue, of algebraic multiplicity 1.

Now we need to find a basis for each eigenspace.

For $\lambda = 8$, we find $E_8 = \text{Null}(A-8I)$ by row reducing

$$\begin{bmatrix} -5 & 2 & 4 \\ 2 & -8 & 2 \\ 4 & 2 & -5 \end{bmatrix} \sim \begin{bmatrix} 1 & -4 & 1 \\ 5 & -2 & -4 \\ -4 & -2 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & -4 & 1 \\ 0 & 18 & -9 \\ 0 & -18 & 9 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & -4 & 1 \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}$$

and we are pleased: $\dim(E_8) = \dim\text{Null}(A-8I)$ (which is just the number of nonleading variables) is 1, as it should be. The geometric multiplicty of 8 is 1.

A basis for E_8 is $\{(1,\frac{1}{2},1)\}$, or $\{(2,1,2)\}$, if you prefer.

Next, to solve for $E_{-1} = \text{Null}(A-(-I)) = \text{Null}(A+I)$ we row reduce:

$$\begin{bmatrix} 4 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & \frac{1}{2} & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

which implies the geometric multiplicity of -1 is $\dim(E_{-1}) = 2$, which coincides with the algebraic multiplicity of -1. A basis for the -1-eigenspace is

$$\{(-\frac{1}{2}, 1, 0), (-1, 0, 1)\}$$

(or multiply the first by 2 if you prefer, to get $\{(-1,2,0),(-1,0,1)\}$.)

Finally, notice that the set

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

is linearly independent (check this!) and hence a basis for $\mathbb{R}^3$. Since $\mathbb{R}^3$ has a basis consisting entirely of eigenvectors of A, we deduce that A is diagonalizable.

So what can you do with a diagonalizable matrix?

Let's continue with the above example.

Given $A = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{bmatrix}$ as above, construct the matrix P whose columns are an eigenvector basis corresponding to A:

$$P = \begin{bmatrix} 2 & -1 & -1 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{bmatrix}$$

Since its columns are a basis for $\mathbb{R}^3$, P is invertible. Call those columns $\mathbf{v}_i$ for short. Then

$$\begin{aligned} AP &= A\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3\end{bmatrix} \\ &= \begin{bmatrix}A\mathbf{v}_1 & A\mathbf{v}_2 & A\mathbf{v}_3\end{bmatrix} \\ &= \begin{bmatrix}8\mathbf{v}_1 & (-1)\mathbf{v}_2 & (-1)\mathbf{v}_3\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3\end{bmatrix} \begin{bmatrix} 8 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \\ &= PD \end{aligned}$$

where D is the diagonal matrix consisting of the eigenvalues of A (given in the same order as the eigenvectors in P).

Since P is invertible, this yields

$$P^{-1}AP = D \qquad \text{or } A = PDP^{-1}$$

Proposition 23.1.2 — Diagonalizing a matrix. If P is a matrix whose columns are an eigenvector basis of $\mathbb{R}^n$ corresponding to A, and D is the diagonal matrix whose diagonal entries are the corresponding eigenvalues, then

$$P^{-1}AP = D \qquad \text{or} \qquad A = PDP^{-1}.$$

Problem 23.1.3 Calculate A^n, for any integer n.
Solution We have $A = PDP^{-1}$; so

$$A^5 = AAAAA = (PDP^{-1})(PDP^{-1})(PDP^{-1})(PDP^{-1})(PDP^{-1}) = PD^5P^{-1}$$

and, since D is diagonal, calculating its powers is easy:

$$D^5 = \begin{bmatrix} 8 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}^5 = \begin{bmatrix} 8^5 & 0 & 0 \\ 0 & (-1)^5 & 0 \\ 0 & 0 & (-1)^5 \end{bmatrix} = \begin{bmatrix} 32,768 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

So in general, $A^n = PD^nP^{-1}$ and D^n is the diagonal matrix with diagonal entries 8^n, $(-1)^n$ and $(-1)^n$.

A little calculation produces P^{-1}:

$$P^{-1} = \frac{1}{9}\begin{bmatrix} 2 & 1 & 2 \\ -1 & 4 & -1 \\ -4 & -2 & 5 \end{bmatrix}$$

so

$$A^n = \begin{bmatrix} 2 & -1 & -1 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 8^5 & 0 & 0 \\ 0 & (-1)^5 & 0 \\ 0 & 0 & (-1)^5 \end{bmatrix} \frac{1}{9}\begin{bmatrix} 2 & 1 & 2 \\ -1 & 4 & -1 \\ -4 & -2 & 5 \end{bmatrix}$$

which multiplies out to give a (slightly nasty, but what do you expect) formula for A^n, for any integer n.

Even $n = -1$ (!).

The equation $A = PDP^{-1}$ is why we call this process "diagonalization;" and applications like the above are among the main reasons for our interest in diagonalization.

For your interest: in your course on differential equations, one of the classes of equations that you come across are systems of linear differential equations. Eigenvalues are eigenvectors are the key to solving the systems; they allow you to "uncouple" the variables. (See Problem 23.4.)

23.2 Failure of diagonalization

We have already pointed out that real matrices could have complex eigenvalues. In that case, the matrix could be diagonalizable over $\mathbb{C}$ without being diagonalizable over $\mathbb{R}$.

More serious, however, is when there is a deficiency in the geometric multiplicity of one or more eigenvalues.

■ **Example 23.2.1** Is $A = \begin{bmatrix} 2 & -4 & -1 \\ 0 & -18 & -4 \\ 0 & 80 & 18 \end{bmatrix}$ diagonalizable?

We compute the eigenvalues:

$$\det\begin{bmatrix} 2-\lambda & -4 & -1 \\ 0 & -18-\lambda & -4 \\ 0 & 80 & 18-\lambda \end{bmatrix} = (2-\lambda)\,((\lambda+18)(\lambda-18)+320)$$
$$= -(\lambda-2)(\lambda^2-4)$$
$$= -(\lambda-2)^2(\lambda+2)$$

and thus deduce that the eigenvalues are 2 (with algebraic multiplicity 2) and -2 (with algebraic multiplicity 1).

For the eigenspace E_2, we row reduce $A-2I$:

$$\begin{bmatrix} 0 & -4 & -1 \\ 0 & -20 & -4 \\ 0 & 80 & 16 \end{bmatrix} \sim \begin{bmatrix} 0 & 1 & \frac{1}{4} \\ 0 & 1 & \frac{1}{5} \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

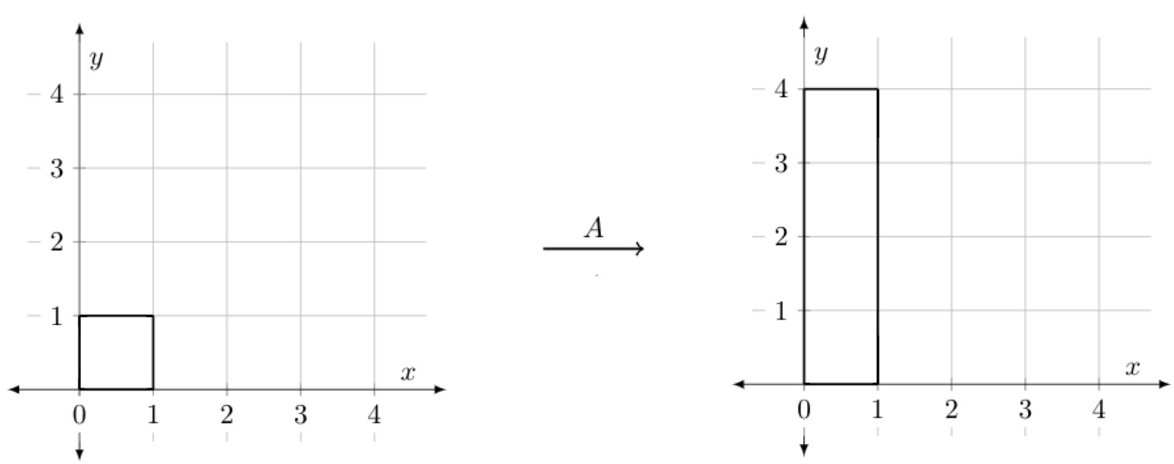
So x_1 is the only nonleading variable; this gives only one basic solution $(1,0,0)$, so $\dim(E_2)=1$. Since the geometric multiplicity of 2 is not equal to its algebraic multiplicity, we deduce that NO, A is not diagonalizable.

(Note that we are using the fact that -2 will give rise to exactly a 1-dimensional eigenspace; so there is nowhere to get a third eigenvector from, and that's the problem.) ■

23.3 Interpreting eigenvectors geometrically

So think of multiplication by A as *transforming* $\mathbb{R}^n$.

■ **Example 23.3.1** Take $A = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$. Then multiplication by A stretches the y-axis by a factor of 4. That is, if we start with a square with corners $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$, then after multiplying by A, these 4 corners go to the points $(0,0)$, $(0,4)$, $(1,0)$ and $(1,4)$, which are the corners of a stretched square. This is illustrated in the following diagram:

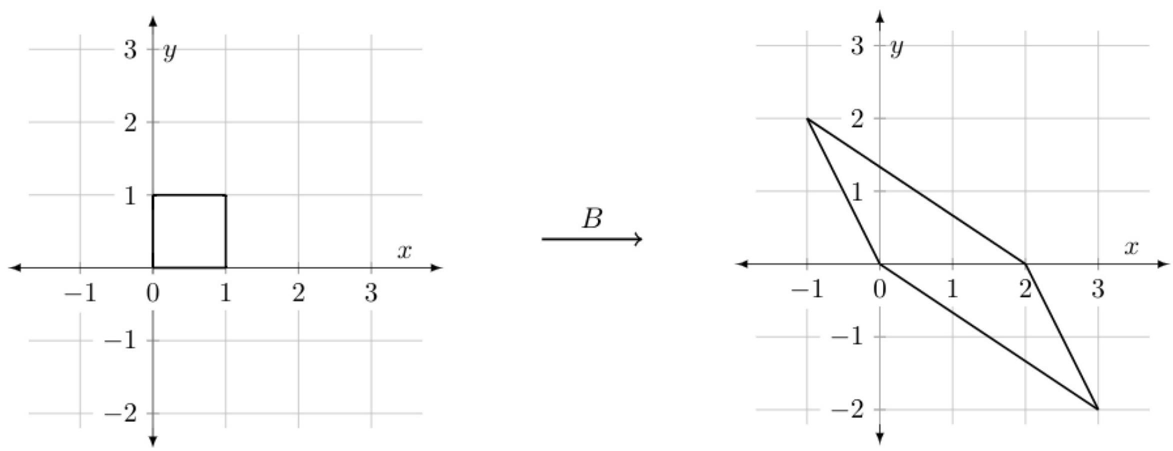


So: multiplication by a diagonal matrix is easy to understand.

What about $B = \begin{bmatrix} 3 & -1 \\ -2 & 2 \end{bmatrix}$? Applying this to the same four points of the unit square $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$ yields

$$(0,0),(-1,2),(3,-2),(2,0)$$

which is a completely different parallelogram, as in the figure below:



But what does it mean?

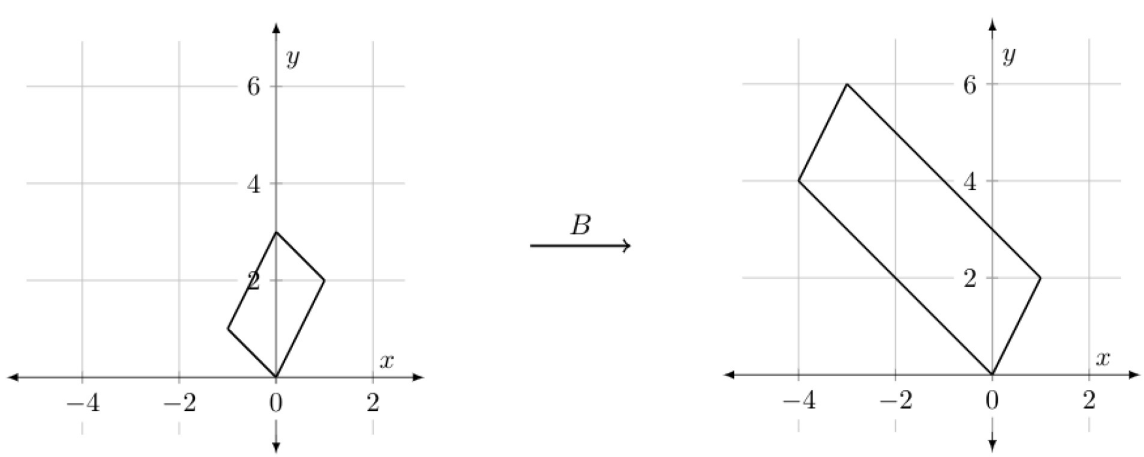
The secret is in the eigenvectors. They are $(1,2)$ (corresponding to eigenvalue 1) and $(-1,1)$ (corresponding to eigenvalue 4). So if we were to draw a parallelogram with vertices

$$(0,0),(1,2),(-1,1),(0,3)$$

then after multiplication by B, you'd get

$$(0,0),(1,2),(-4,4),(-3,6)$$

which makes sense when you sketch it, as in the figure below:



Multiplication by B stretch the parallelogram by a factor of 4 in the direction $(-1,1)$ (and by a factor of 1 in the direction $(1,2)$). Looking back, this also explains what happened to our standard unit square: it was stretched along the axis $\text{span}\{(-1,1)\}$.

The point: the basis of eigenvectors is the secret to understanding the geometry of multiplication by A. ■

Viewing A as a transformation in this way — as a *linear transformation* — is fundamental to discrete time dynamical systems see 'Markov processes in probability'[1].

What we want to do next: explore this idea of linear transformations in greater detail, and get back to that picture we once used in class, illustrating the nullspace and column space of a matrix via a map from $\mathbb{R}^n$ to $\mathbb{R}^m$ (the map being: multiplication by A).

Problems

Remarks:

1. A question with an asterisk '*' (or two) indicates a bonus-level question.
2. You must justify all your responses.

Problem 23.1 a) For each of the matrices A in Problem 22.1, if possible, find an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$. If this is not possible, explain why. (Solutions to parts b, d, f and h are available.)

b) Use the fact that $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ is diagonalizable to compute $A^{10^{1000}}$ before the sun becomes a red giant and (possibly) engulfs the earth.[a]

[a] You have between 5 and 6 billion years, but it should only take you less than 5 minutes.

Problem 23.2 State whether each of the following is (always) true, or is (possibly) false.

- If you say the statement may be false, you must give an explicit example.
- If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

a) If 3 is an eigenvalue of an $n \times n$ matrix A, there must be a non-zero vector $\mathbf{v} \in \mathbb{R}^n$ with $Av = 3v$.

b) *The matrix $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ has no real eigenvalues.

c) The matrix $\begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$ is diagonalizable.

d) *If 0 is an eigenvalue of $n \times n$ matrix A, then A is not invertible.

e) If an $n \times n$ matrix A is not invertible, then 0 is an eigenvalue of A.

f) *Every invertible matrix is diagonalizable.

[1] Search the web for 'Markov chain'.

g) Every diagonalizable matrix is invertible.

h) *If an $n \times n$ matrix has n distinct eigenvalues, then the matrix is diagonalizable.

i) If an $n \times n$ matrix is diagonalizable then it must have n distinct eigenvalues.

j) *[a] If an $n \times n$ matrix A has eigenvalues $\lambda_1, \dots, \lambda_n$, then $\det A = \lambda_1 \dots \lambda_n$.

k) * If $\mathbf{v}$ and $\mathbf{w}$ are eigenvectors of a symmetric matrix A (i.e. $A = A^T$) which correspond to different eigenvalues, then $\mathbf{v} \cdot \mathbf{w} = 0$.[b]

---

[a] Hint: Use the fact that we know $\det(A - \lambda I_n) = (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2)\dots(\lambda - \lambda_n)$.

[b] Hint: Remember that because A is symmetric, $A\mathbf{v} \cdot \mathbf{w} = \mathbf{v} \cdot A\mathbf{w}$ (by Problem 14.4). Now simplify both sides using the fact that $\mathbf{v}$ and $\mathbf{w}$ are eigenvectors and see what you obtain.

Problem 23.3 *Let $A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$.

a) Compute $\det(A - \lambda I_3)$ and hence show that the eigenvalues of A are 2 and -1.
b) Find a basis of $E_2 = \{\mathbf{x} \in \mathbb{R}^3 \mid A\mathbf{x} = 2\mathbf{x}\}$.
c) Find a basis of $E_{-1} = \{\mathbf{x} \in \mathbb{R}^3 \mid A\mathbf{x} = -\mathbf{x}\}$.
d) Find an invertible matrix P such that $P^{-1}AP = D$ is diagonal, and give this diagonal matrix D. Explain why your choice of P is invertible.
e) Find an invertible matrix $Q \neq P$ such that $Q^{-1}AQ = \tilde{D}$ is also diagonal, and give this diagonal matrix $\tilde{D}$.

Problem 23.4 *[a] Consider the coupled system of second order differential equations for the the two functions f and g:

$$\ddot{f} = -2f + g$$
$$\ddot{g} = f - 2g$$

(Here $\ddot{f}$ and $\ddot{g}$ denote $\dfrac{d^2 f}{dt^2}$ and $\dfrac{d^2 g}{dt^2}$ respectively.) This can be also written in matrix form as: $\begin{bmatrix} \ddot{f} \\ \ddot{g} \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} f \\ g \end{bmatrix}$. Let $A = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$. Diagonalize A to write $A = PDP^{-1}$ for some invertible matix P and some diagonal matrix $D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$. Now define two new functions h and k by $\begin{bmatrix} h \\ k \end{bmatrix} = P^{-1} \begin{bmatrix} f \\ g \end{bmatrix}$. Show that $\begin{bmatrix} \ddot{f} \\ \ddot{g} \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} f \\ g \end{bmatrix}$ is equivalent to $\begin{bmatrix} \ddot{h} \\ \ddot{k} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} h \\ k \end{bmatrix}$, or, written out fully as

$$\ddot{h} = \lambda_1 h$$
$$\ddot{k} = \lambda_2 k$$

You will find that both λ_1 and λ_2 are negative, so the solutions to these two (linear) second order differential equations[b] are $h(t) = a\sin(\sqrt{|\lambda_1|}t) + b\cos(\sqrt{|\lambda_1|}t)$ and $k(t) = c\sin(\sqrt{|\lambda_2|}t) + c\cos(\sqrt{|\lambda_2|}t)$, where a, b, c, d are real constants.

Use this to find f and g, and solve for a, b, c and d in terms of the so-called *'initial conditions'* $f(0), \dot{f}(0), g(0), \dot{g}(0)$.

[a] This example is a simplified version of the equations of motion for 2 masses connected by springs. Search the web for 'Normal mode' for an example.

[b] Search the web for 'Linear differential equation' for some details.

VI Linear Transformations

In mathematics, one might begin by studying the objects under consideration — in this case, vector spaces — but then equally important is to understand the maps between them, that is, the maps that are "compatible" with the vector space structure. These maps are called *linear transformations*. We have seen and used several already: the coordinate map, the projection map.

24. Linear transformations

Photo: Ralph Nevins. Parc Provincial Pinery Provincial Park, Ontario

Last time, we considered the geometric interpretation of "multiplication by A": it is a transformation of $\mathbb{R}^n$. It's a special kind of transformation: it takes squares to parallelograms (as opposed to anything else your imagination could provide!).

The key properties that made this work:

- $A\mathbf{0} = \mathbf{0}$
- $A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v}$ for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$
- $A(r\mathbf{u}) = r(A\mathbf{u})$, for any $\mathbf{u} \in \mathbb{R}^n$, $r \in \mathbb{R}$

These properties imply that the four points of a parallelogram with vertices

$$\mathbf{0}, \quad \mathbf{u}, \quad \mathbf{v}, \quad \mathbf{u}+\mathbf{v}$$

are sent to the four points

$$\mathbf{0}, \quad A\mathbf{u}, \quad A\mathbf{v}, \quad A\mathbf{u}+A\mathbf{v}$$

(which again define the corners of a parallelogram, although it could be degenerate (a line segment) if $A\mathbf{u}$ and $A\mathbf{v}$ are linearly dependent); and moreover that the straight line $r\mathbf{u}$ is sent onto the straight line $r(A\mathbf{u})$ (and similarly for the line through $\mathbf{v}$).

Definition 24.0.1 Let U and V be vector spaces. A *linear transformation* T is a map from U to V satisfying

1. for all $\mathbf{u}, \mathbf{v} \in U$, $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$
2. for all $\mathbf{u} \in U$, $r \in \mathbb{R}$, $T(r\mathbf{u}) = rT(\mathbf{u})$

We use the word *transformation* here, and *map*; we could use *function* but that's often reserved for a map whose range is the real numbers. Whatever the terminology, the point is that T is a black box or formula or rule which accepts a vector of U as input, and produces a (uniquely determined) vector in V as output; and furthermore we stipulate that it takes sums to sums and scalar multiples to scalar multiples.

So: multiplication by a square matrix A is a linear transformation.

We must now ask ourselves: what other kinds of linear transformations do we know?

24.1 Examples of linear transformations

Example 24.1.1 Let A be an $m \times n$ matrix. Define the map

$$T : \mathbb{R}^n \to \mathbb{R}^m$$

by $T_A(\mathbf{u}) = A\mathbf{u}$. We claim that T_A is a linear tranformation.

To prove this, we need to verify that:

1. for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, $T_A(\mathbf{u}+\mathbf{v}) = T_A(\mathbf{u}) + T_A(\mathbf{v})$:

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

   as required.

2. for any $\mathbf{u} \in \mathbb{R}^n$, $r \in \mathbb{R}$, $T_A(r\mathbf{u}) = rT_A(\mathbf{u})$:

$$T_A(r\mathbf{u}) = A(r\mathbf{u}) = r(A\mathbf{u}) = rT_A(\mathbf{u})$$

   as required.

So this is a linear transformation, for any m, n and A. ■

Example 24.1.2 Consider the projection onto the plane W given by

$$W = \{(x,y,z) \mid x - z = 0\}$$

Let's first find a formula for this projection. The easy way is to note that $W^{\perp} = \text{span}\{(1,0,-1)\}$ so for $\mathbf{u} = (u_1, u_2, u_3) \in \mathbb{R}^3$, we have

$$\text{proj}_{W^{\perp}}(\mathbf{u}) = \frac{\mathbf{u}\cdot(1,0,-1)}{(1,0,-1)\cdot(1,0,-1)}\begin{bmatrix}1\\0\\-1\end{bmatrix} = \frac{u_1-u_3}{2}\begin{bmatrix}1\\0\\-1\end{bmatrix}$$

and therefore

$$\text{proj}_W(\mathbf{u}) = \mathbf{u} - \text{proj}_{W^{\perp}}(\mathbf{u}) = \begin{bmatrix}u_1\\u_2\\u_3\end{bmatrix} - \frac{u_1-u_3}{2}\begin{bmatrix}1\\0\\-1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}u_1+u_3\\2u_2\\u_1+u_3\end{bmatrix}.$$

Now is this a linear transformation? We verify the two properties of linearity (with $T = \text{proj}_W$):

1. Is $\text{proj}_W(\mathbf{u}+\mathbf{v}) = \text{proj}_W(\mathbf{u}) + \text{proj}_W(\mathbf{v})$? Let $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$; then

$$\begin{aligned}
\text{proj}_W(\mathbf{u}+\mathbf{v}) &= \text{proj}_U(u_1+v_1, u_2+v_2, u_3+v_3) \\
&= \frac{1}{2}\begin{bmatrix}(u_1+v_1)+(u_3+v_3) \\ 2(u_2+v_2) \\ (u_1+v_1)+(u_3+v_3)\end{bmatrix} \\
&= \frac{1}{2}\left(\begin{bmatrix}u_1+u_3 \\ 2u_2 \\ u_1+u_3\end{bmatrix} + \begin{bmatrix}v_1+v_3 \\ 2v_2 \\ v_1+v_3\end{bmatrix}\right) \\
&= \frac{1}{2}\begin{bmatrix}u_1+u_3 \\ 2u_2 \\ u_1+u_3\end{bmatrix} + \frac{1}{2}\begin{bmatrix}v_1+v_3 \\ 2v_2 \\ v_1+v_3\end{bmatrix} \\
&= \text{proj}_W(\mathbf{u}) + \text{proj}_W(\mathbf{v})
\end{aligned}$$

as required

2. Is $\text{proj}_W(r\mathbf{u}) = r\,\text{proj}_W(\mathbf{u})$? Let $r \in \mathbb{R}$ and $\mathbf{u}$ as above, then

$$\begin{aligned}
\text{proj}_W(r\mathbf{u}) &= \text{proj}_W(ru_1, ru_3, ru_3) \\
&= \frac{1}{2}\begin{bmatrix}ru_1+ru_3 \\ 2ru_2 \\ ru_1+ru_3\end{bmatrix} \\
&= r\left(\frac{1}{2}\begin{bmatrix}u_1+u_3 \\ 2u_2 \\ u_1+u_3\end{bmatrix}\right) \\
&= r\,\text{proj}_W(\mathbf{u})
\end{aligned}$$

Since both axioms are satisfied, this is a linear transformation. ■

■ **Example 24.1.3** Show that the map $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x,y) = (x+1, xy)$ is not a linear transformation.

It suffices to show that there exists even one pair of vectors $\mathbf{u}, \mathbf{v}$ for which $T(\mathbf{u}+\mathbf{v}) \neq T(\mathbf{u}) + T(\mathbf{v})$; OR to show that there is even one vector and one scalar for which $T(r\mathbf{u}) \neq rT(\mathbf{u})$. Let's show that in fact these axioms fail for almost all vectors and scalars!

1. Note that

$$T(\mathbf{u}+\mathbf{v}) = T(u_1+v_1, u_2+v_2) = (u_1+v_1+1, (u_1+v_1)(u_2+v_2))$$

whereas

$$T(\mathbf{u}) + T(\mathbf{v}) = (u_1+1, u_1u_2) + (v_1+1, v_1v_2) = (u_1+v_1+2, u_1u_2+v_1v_2)$$

but the first components can NEVER be equal (and the second components are only equal if the "cross terms" are zero).

So, for example, $T(1,0) = (2,0)$, $T(0,1) = (1,0)$ and

$$T(1,1) = (2,1) \neq (2,0) + (1,0) = T(1,0) + T(0,1).$$

We could stop now: we've found an example that shows T is not linear. But let's show that T doesn't satisfy the second condition either.

2. Note that
$$T(r\mathbf{u}) = T(ru_1, ru_2) = (ru_1 + 1, (ru_1)(ru_2)) = (ru_1 + 1, r^2 u_1 u_2)$$
whereas
$$rT(\mathbf{u}) = r(u_1 + 1, u_1 u_2) = (ru_1 + r, ru_1 u_2)$$
but the first component is equal only when $r = 1$ and the second only when either $r = 1$ or $r = 0$ or $u_1 u_2 = 0$.
So, for example, $T(1,1) = (2,1)$, but
$$T(2(1,1)) = T(2,2) = (3,4) \neq 2(2,1) = 2T(1,1).$$

So this map is defintely not linear. ■

24.2 Constructing and describing linear transformations

Checking the linearity of a map is eerily similar to the subspace test. Is it the same thing? NO: the subspace test is for sets, but linear transformations are maps between vector spaces (or subspaces).

The phrase "linear transformation" comes about in part because the second property implies that the image under T of a line through the origin is again a line through the origin (or just $\{0\}$).

In particular, taking $r = 0$, we deduce that for any linear transformation, $T(\mathbf{0}) = \mathbf{0}$.

But the two properties imply much more: they say that linear combinations are sent to linear combinations, in the following sense.

Theorem 24.2.1 — Determination of Linear Transformations on a Basis.

1. Suppose $T : U \to V$ is a linear transformation and $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is a basis for U. Then T is completely determined by the vectors $T(\mathbf{u}_1)$, $\cdots$, $T(\mathbf{u}_n)$.
2. Suppose $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is a basis for U and $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ are ANY n vectors in V (even possibly dependent or $\mathbf{0}$). Then there is a unique linear transformation T which satisfies $T(\mathbf{u}_i) = \mathbf{v}_i$ for all i.

Proof. 1. What we mean by "completely determined" is: we can determine $T(\mathbf{u})$, without having a formula for T, IF we know the vectors $T(\mathbf{u}_1)$, $\cdots$, $T(\mathbf{u}_n)$. Namely, let $\mathbf{u} \in U$ be arbitrary. Since $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is a basis for U, we can write
$$\mathbf{u} = a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n.$$
Then
$$T(\mathbf{u}) = T(a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n) = a_1T(\mathbf{u}_1) + \cdots + a_nT(\mathbf{u}_n),$$
as we wanted to show.

2. We need to give a rule for a map T which sends $\mathbf{u}_i$ to $\mathbf{v}_i$. We use the idea above: for any $\mathbf{u} \in U$, write $\mathbf{u} = a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n$. Then define
$$T(\mathbf{u}) = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n.$$
This gives a function from U to V and you can check that it is linear (try it!).

■

This theorem is very powerful. Much in the same way that proving the existence of a basis of any vector space simplified things by showing that every n-dimensional vector space is really just $\mathbb{R}^n$ in disguise, this theorem implies that every linear transformation is just matrix multiplication in disguise!

Theorem 24.2.2 — The standard matrix of a linear transformation. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there is an $m \times n$ matrix A such that

$$T(\mathbf{x}) = A\mathbf{x}$$

for all $\mathbf{x} \in \mathbb{R}^n$. More precisely, if $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ is the standard basis for $\mathbb{R}^n$, then the matrix A is given by:

$$A = [T(\mathbf{e}_1)\ T(\mathbf{e}_2)\ \cdots\ T(\mathbf{e}_n)].$$

The matrix A is called the *standard matrix* of T.

■ **Example 24.2.3** Consider the projection $T = \text{proj}_W$ onto the plane $W = \{(x,y,z) \mid x - z = 0\}$; we saw earlier that $T(u_1,u_2,u_3) = \frac{1}{2}(u_1+u_3, 2u_2, u_1+u_3)$.

Construct the matrix as in the theorem: $T(1,0,0) = (\frac{1}{2},0,\frac{1}{2})$, $T(0,1,0) = (0,1,0)$, $T(0,0,1) = (\frac{1}{2},0,\frac{1}{2})$, so

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

and we see directly that

$$A\mathbf{u} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} u_1+u_3 \\ 2u_2 \\ u_1+u_3 \end{bmatrix} = T(\mathbf{u})$$

as required. ■

Proof. To see why it works, recall that if you have a vector

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

then in fact

$$\mathbf{u} = u_1\mathbf{e}_1 + \cdots + u_n\mathbf{e}_n.$$

Thus $T(\mathbf{u})$, by Theorem 24.2.1, is

$$T(\mathbf{u}) = u_1 T(\mathbf{e}_1) + \cdots + u_n T(\mathbf{e}_n).$$

But taking linear combinations is just matrix multiplication, so we have

$$T(\mathbf{u}) = [T(\mathbf{e}_1)\ T(\mathbf{e}_2)\ \cdots\ T(\mathbf{e}_n)] \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = A\mathbf{u}.$$

■

24.3 Kernels and images

This new interpretation of the projection map as a linear transformation gives us more geometric ideas about what the operation of projection really does. For instance, the projection map annihilates (sends to $\mathbf{0}$) any vector of $U^{\perp}$. But it completely covers U in the sense that ever point of U is the image of some element under T.

To state this more clearly, we need some definitions.

Definition 24.3.1 Let $T: U \to V$ be a linear transformation. Then

- the *kernel* of T, denoted $\ker(T)$, is the set of all elements of U which are sent to $\mathbf{0}$ by T, that is,
$$\ker(T) = \{\mathbf{u} \in U \mid T(\mathbf{u}) = \mathbf{0}\}$$
- the *image* of T, denoted $\operatorname{im}(T)$, is the set of all elements of V which are equal to $T(\mathbf{u})$ for some $\mathbf{u} \in U$, that is,
$$\operatorname{im}(T) = \{\mathbf{v} \in V \mid \mathbf{v} = T(\mathbf{u}) \text{ for some } \mathbf{u} \in U.\}$$

Now both $\ker(T)$ and $\operatorname{im}(T)$ are subsets of vector spaces, and the natural first question is: are they subspaces? Answer: YES, and they're even familiar ones!

Theorem 24.3.2 — Kernels and Images of the Standard Matrix. Let $T: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with standard matrix A. Then
$$\ker(T) = \operatorname{Null}(A) \qquad \text{and} \qquad \operatorname{im}(T) = \operatorname{Col}(A).$$

The first equality is clear once you recall that $T(\mathbf{u}) = A\mathbf{u}$; and for the second, one should go back and look up the original definition of $\operatorname{Col}(A) = \operatorname{im}(A)$.

24.4 A return to the rank-nullity theorem

Note that the rank-nullity theorem[1] can be restated, for a linear transformation $T: U \to V$, as
$$\dim(\ker(T)) + \dim(\operatorname{im}(T)) = n$$
where $n = \dim(U)$. We interpret this as giving an accounting of what T does to the vectors in U. If T sends U onto a subspace of V of dimension equal to U, then the kernel must be $\{\mathbf{0}\}$. On the other hand, if the dimension of the image is smaller than $\dim U$, then those 'missing' dimensions had to go somewhere; in fact, they ended up in the kernel of T, as the equation above suggests.

For example. the projection onto the plane W we saw earlier has kernel equal to the 1-dimensional $W^{\perp}$ and image is all of (2-dimensional) W: and $2+1=3=\dim(\mathbb{R}^3)$.

Conversely, knowing that we can think about T as a matrix multiplication means that we know that a basis for $\operatorname{im}(T)$ is any basis for $\operatorname{Col}(A)$ (and this is something that we find easy to answer).

[1] Some like to call this theorem *the conservation of dimension*: since the dimension of the subspace sent to 0 ($\dim\ker T$) plus the dimension of the image of T ("what's left"), is the same the dimension you began with: $n = \dim U$. "Total dimension is preserved."

24.5 A remark about the projection matrix

When we calculated the projection onto a subspace before, one of our methods was:

- Create a matrix B such that $\mathrm{Col}(B) = W$
- Solve $(B^T B)\mathbf{x} = B^T\mathbf{b}$.
- $\mathrm{proj}_W(\mathbf{b}) = B\mathbf{x}$

Putting this all together: Suppose B has linearly independent columns, so that $B^T B$ is invertible. (This was a homework exercise.) Then

$$\mathrm{proj}_W(\mathbf{b}) = B(B^T B)^{-1}B^T\mathbf{b}$$

so the projection is given by multiplication by the matrix

$$B(B^T B)^{-1}B^T.$$

Therefore, this must be the standard matrix of T! You can try this out for the example above.

Problems

Remarks:

1. A question with an asterisk '*' (or two) indicates a bonus-level question.
2. You must justify all your responses.

Problem 24.1 State whether each of the following defines a linear transformation.

- If you say it isn't linear, you must give an explicit example to illustrate.
- If you say it is linear, you must give a clear explanation - by quoting a theorem presented in class, or by verifying the conditions in the definition *in every case*.

a) $T : \mathbb{R}^2 \to \mathbb{R}^3$ defined by $T(x,y) = (x,y,x+y)$

b) $^\star T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x,y,z) = (2z+x,y)$

c) $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(x,y) = (x,xy)$

d) $^\star T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(\mathbf{v}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\mathbf{v}$

e) $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(\mathbf{v}) = \mathbf{v} \times (1,2,3)$, where '$\times$' denotes the cross product.

f) $^\star T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(\mathbf{v}) = \mathrm{proj}_{(1,1,-1)}(\mathbf{v})$.

g) $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(\mathbf{v}) = v - \mathrm{proj}_{(1,1,-1)}(\mathbf{v})$.

h) $^\star T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(\mathbf{v}) = \mathrm{proj}_{\mathbf{v}}(1,1,-1)$.

i) $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(\mathbf{v}) = \big(\mathbf{v}\cdot(1,1,-1)\big)(1,0,1)$.

j) $^\star T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(\mathbf{v}) = 2\mathbf{v}$.

k) $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(\mathbf{v}) = \mathrm{proj}_H(\mathbf{v})$, where H is the plane through the origin with normal $(1,1,0)$.

l) $^\star T:\mathbb{R}^3\to\mathbb{R}^2$ defined by $T(\mathbf{v})=A\mathbf{v}$, where $A=\begin{bmatrix}1&0&1\\1&2&3\end{bmatrix}$.

Problem 24.2 In each of the following, find the standard matrix of T and use it to give a basis for $\ker T$ and $\operatorname{im} T$ and verify the conservation of dimension.

a) $T:\mathbb{R}^2\to\mathbb{R}^3$ defined by $T(x,y)=(x,y,x+y)$

b) $^\star T:\mathbb{R}^3\to\mathbb{R}^2$ defined by $T(x,y,z)=(2z+x,y)$

c) $T:\mathbb{R}^3\to\mathbb{R}^3$ defined by $T(\mathbf{v})=\mathbf{v}\times(1,2,3)$, where '$\times$' denotes the cross product.

d) $^\star T:\mathbb{R}^3\to\mathbb{R}^3$ defined by $T(\mathbf{v})=\operatorname{proj}_{(1,1,-1)}(\mathbf{v})$.

e) $T:\mathbb{R}^3\to\mathbb{R}^3$ defined by $T(\mathbf{v})=\mathbf{v}-\operatorname{proj}_{(1,1,-1)}(\mathbf{v})$.

f) $^\star T:\mathbb{R}^3\to\mathbb{R}^3$ defined by $T(\mathbf{v})=\operatorname{proj}_H(\mathbf{v})$, where H is the plane through the origin with normal $(1,1,0)$.

Problem 24.3 State whether each of the following is (always) true, or is (possibly) false.

- If you say the statement may be false, you must give an explicit example.
- If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

a) If $T:\mathbb{R}^3\to\mathbb{R}^2$ is linear, then $\ker T\neq\{0\}$.

b) *If $T:\mathbb{R}^4\to\mathbb{R}^2$ is linear, then $\dim\ker T\geq 2$.

c) If $T:\mathbb{R}^4\to\mathbb{R}^5$ is linear, then $\dim\operatorname{im} T\leq 4$.

d) *If $T:\mathbb{R}^3\to\mathbb{R}^2$ is linear, and $\{\mathbf{v}_1,\mathbf{v}_2\}\subset\mathbb{R}^3$ is linearly independent, then $\{T(\mathbf{v}_1),T(\mathbf{v}_2)\}\subset\mathbb{R}^2$ is linearly independent.

e) If $T:\mathbb{R}^3\to\mathbb{R}^2$ is linear, $\ker T=\{\mathbf{0}\}$, and $\{\mathbf{v}_1,\mathbf{v}_2\}\subset\mathbb{R}^3$ is linearly independent, then $\{T(\mathbf{v}_1),T(\mathbf{v}_2)\}\subset\mathbb{R}^2$ is linearly independent.

f) *If $T:\mathbb{R}^3\to\mathbb{R}^3$ is linear, and $\ker T=\{\mathbf{0}\}$, then $\operatorname{im} T=\mathbb{R}^3$.

g) If $T:\mathbb{R}^2\to\mathbb{R}^3$ is linear, and $\ker T=\{\mathbf{0}\}$, then $\operatorname{im} T=\mathbb{R}^3$.

Problem 24.4 * State whether each of the following defines a linear transformation.

- If you say it isn't linear, you must give an explicit example to illustrate.
- If you say it is linear, you must give a clear explanation - by quoting a theorem presented in class, or by verifying the conditions in the definition *in every case*.

a) $T:\mathbb{P}\to\mathbb{P}$ defined by $T(p)=p'$, where p' denotes the derivative of p.

b) $^\star T:\mathbb{P}\to\mathbb{P}$ defined by $T(p)(t)=\int_0^t p(s)ds$.

c) $\mathrm{tr}:\mathbf{M}_{22}\to\mathbb{R}$ defined by $\mathrm{tr}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = a+d$.

d) $^\star\det:\mathbf{M}_{22}\to\mathbb{R}$ defined by $\det\begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad-bc$.

e) $T:\mathbf{F}(\mathbb{R})\to\mathbb{R}$ defined by $T(f)=f(1)$.

VII Solutions

Try every question *before* looking at the solution — once you've seen a solution, you'll need a fresh new problem to learn from.

25. Solutions to selected exercises

Problems of Chapter 1

1.1 Express the following complex numbers in Cartesian form: $a+bi$, with $a,b \in \mathbb{R}$.

a) $(2+i)(2+2i) = 2+6i$

b) $\dfrac{1}{1+i} = \dfrac{1}{2} - \dfrac{1}{2}i$

c) $\dfrac{8+3i}{5-3i} = \dfrac{31}{34} - \dfrac{39}{34}i$

d) $\dfrac{5+5i}{1-i} = 5i$

e) $\dfrac{(1+2i)(2+5i)}{3+4i} = \dfrac{12}{25} + \dfrac{59}{25}i$

f) $\dfrac{1-i}{2-i} + \dfrac{2+i}{1-i} = \dfrac{11}{10} + \dfrac{13}{10}i$

g) $\dfrac{1}{(1-i)(3-2i)} = \dfrac{1}{26} + \dfrac{5}{26}i$

1.2 Find the polar form of the following complex numbers: (i.e. either as $re^{i\theta}$ or as $r(\cos\theta + i\sin\theta)$, with $r \geq 0$ and $-\pi < \theta \leq \pi$)

a) $3\sqrt{3} - 3i = 6(\cos(-\pi/6) + i\sin(-\pi/6)) = 6e^{-\frac{\pi}{6}i}$

b) $\dfrac{3\sqrt{3}-3i}{\sqrt{2}+i\sqrt{2}} = 3(\cos(-5\pi/12) + i\sin(-5\pi/12)) = 3e^{-\frac{5\pi}{12}i}$

c) $\dfrac{1-\sqrt{3}i}{-1+i} = \sqrt{2}(\cos(11\pi/12) + i\sin(11\pi/12)) = \sqrt{2}e^{\frac{11\pi}{12}i}$

d) $\dfrac{5+5\sqrt{3}i}{\sqrt{2}-\sqrt{2}i} = 5(\cos(7\pi/12) + i\sin(7\pi/12)) = 5e^{\frac{7\pi}{12}i}$

e) $\dfrac{3+3\sqrt{3}i}{-2+2i} = \dfrac{3\sqrt{2}}{2}(\cos(5\pi/12) - i\sin(5\pi/12)) = \dfrac{3\sqrt{2}}{2}e^{\frac{5\pi}{12}i}$

1.4 If z is a complex number,

(i) Is it possible that $z = \bar{z}$? **Solution:** (Yes, iff $z \in \mathbb{R}$:if $z = a+bi$, then $z = \bar{z}$ iff $a+bi = a-bi$) iff $b = -b$ iff $z = a \in \mathbb{R}$.

(ii) Is it possible that $|\bar{z}| > |z|$? **Solution:** (No: $|\bar{z}| = |z|$, always: if $z = a+bi$, then $|z| = \sqrt{a^2+b^2} = \sqrt{a^2+(-b)^2} = |\bar{z}|$.)

(iii) Is it possible that $\bar{z} = 2z$? **Solution:** (Yes, but only if $z = 0$: If $\bar{z} = 2z$, then by the previous part, $|z| = 2|z|$, so $|z| = 0$. So $z = 0$ too.)

Give examples to illustrate your affirmative answers, and explanations if you say the statement is always false.

Problems of Chapter 2

2.3 If $A = (1,\ 2,\ 3)$, $B = (-5,\ -2,\ 5)$, $C = (-2,\ 8,\ -10)$ and D is the midpoint of $\overline{AB}$, find the coordinates of the midpoint of $\overline{CD}$.

Solution: The position vector v of the midpoint of $\overline{CD}$ is $v = (C+D)/2$. But as $D = (A+B)/2$, $v = (C+(A+B)/2)/2 = (-2,\ 4,\ -3)$

2.4 Solve the following problems using the dot product.

(b) Find the angle between the vectors $(0,\ 3,\ -3)$ and $(-2,\ 2,\ -1)$.

Solution: If θ is the angle between these two vectors, then

$$\cos\theta = \frac{(0,\ 3,\ -3)\cdot(-2,\ 2,\ -1)}{\|(0,\ 3,\ -3)\|\|(-2,\ 2,\ -1)\|} = \frac{9}{\sqrt{18}\sqrt{9}} = \frac{\sqrt{2}}{2}.$$

Hence $\theta = \dfrac{\pi}{4}$.

2.5 Solve the following problems.

(a) If $u = (2,\ 1,\ 3)$ and $v = (3,\ 3,\ 3)$ find $\operatorname{proj}_v u$.

Solution:

$$\operatorname{proj}_v u = \frac{u\cdot v}{\|v\|^2}v = \frac{(2,\ 1,\ 3)\cdot(3,\ 3,\ 3)}{\|(3,\ 3,\ 3)\|^2}(3,\ 3,\ 3) = \frac{18}{27}(3,\ 3,\ 3) = (2,2,2).$$

(b) If $u = (3,\ 3,\ 6)$ and $v = (2,\ -1,\ 1)$ find the length of the projection of u along v.

Solution:

$$\|\operatorname{proj}_v u\| = \frac{|u\cdot v|}{\|v\|^2}\|v\| = \frac{|u\cdot v|}{\|v\|} = \frac{(3,\ 3,\ 6)\cdot(2,\ -1,\ 1)}{\|((2,\ -1,\ 1))\|} = \frac{3}{2}\sqrt{6}.$$

(c) Find angle between the planes with Cartesian equations $x - z = 7$ and $y - z = 234$

Solution: The angle between two planes is always the acute angle between their normal lines. So we find the angle φ between their normals $(1,0,1$ and $(0,1,-1)$ and adjust if necessary. The angle φ is $\dfrac{2\pi}{3}$, so the answer is $\pi - \dfrac{2\pi}{3} = \dfrac{\pi}{3}$

Problems of Chapter 3

3.1 Solve the following problems using the cross and/or dot products.

(b) Find all vectors in $\mathbb{R}^3$ which are orthogonal to both $(-1,1,5)$ and $(2,1,2)$.

Solution: Such vectors will be parallel to the cross product of these two vectors, which is found to be $(-3,12,-3)$. Hence $\{(t,\ -4t,\ t)|\ t \in \mathbb{R}\}$ is the solution.

(d) If $u = (-4,\ 2,\ 7),\ v = (2,\ 1,\ 2)$ and $w = (1,\ 2,\ 3)$, find $u \cdot (v \times w)$.

Solution: This is $(-4,\ 2,\ 7) \cdot (2,\ 1,\ 2) \times (1,\ 2,\ 3) = (-4,\ 2,\ 7) \cdot (-1,-4,3) = 17$.
3.2 Solve the following problems using the appropriate products.

(b) Find the area of the triangle with vertices $A = (-1,\ 5,\ 0)$, $B = (1,\ 0,\ 4)$ and $C = (1,\ 4,\ 0)$

Solution: This will be $\frac{1}{2}$ the length of the cross product of the two vectors $B-A$ and $C-A$, and so is $\frac{1}{2}\|(4,8,8)\| = 6$.

(d) Find the volume of the parallelepiped determined by $u = (1,\ 1,\ 0),\ v = (1,\ 0,\ -1)$ and $w = (1,\ 1,\ 1)$.

Solution: This is simply the absolute value of $u \cdot v \times w$ and so is 1.
3.3 Solve the following problems.

(a) Find the point of intersection of the plane with Cartesian equation $2x+2y-z=5$, and the line with parametric equations $x = 4-t,\ y = 13-6t,\ z = -7+4t$.

Solution: Once we substitute $x = 4-t,\ y = 13-6t$ and $z = -7+4t$ into $2x+2y-z=5$, we can solve for t, which we then substitute backinto $x = 4-t,\ y = 13-6t$ and $z = -7+4t$ to obtain $(2,\ 1,\ 1)$.

(b) If L is the line passing through $(1,\ 1,\ 0)$ and $(2,\ 3,\ 1)$, find the point of intersection of L with the plane with Cartesian equation $x+y-z=1$.

Solution: We find the (scalar) parametric equations for the L, and then proceed as in part (a) to obtain $(1/2,0,\ -1/2)$

(d) Do the planes with Cartesian equations $2x-3y+4z=6$ and $4x-6y+8z=11$ intersect?

Solution: No, their normals are parallel, so the planes are, but their equations are not multiples of the other.

(f) Find the line of intersection of the planes with Cartesian equations $x+11y-4z=40$ and $x-y=-8$.

Solution: A direction vector for this line will be perpendicular to both normal vectors, and so can be obtained as the cross product of these normals, namely $(-4,-4,-12)$. (We will choose $(1,1,3)$ instead.) Now all we need to do is find one point on this line. This is done by substituting $x = y-8$ into $x+11y-4z=40$, and simplifying to obtain $3y-z=12$. Now set $z=0$, to obtain the point $(-4,\ 4,\ 0)$. Hence the line is $\{(-4,\ 4,\ 0)+t(1,\ 1,\ 3),\ |\ t \in \mathbb{R}\}$

3.4 Solve the following problems.

(b) Find the distance from the point $Q = (-2,\ 5,\ 9)$ to the plane with Cartesian equation $6x+2y-3z=-8$

Solution: Choose any point P on the plane - say $P=(0,-4,0)$. The the distance between Q and the plane is the length of the projection of QP in the direction of the normal $(6,2,-3)$. This length is $\left|\frac{(Q-P)\cdot(6,2,-3)}{\|(6,2,-3)\|}\right| = 3.$

(d) Find the distance from the point $P = (8,\ 6,\ 11)$ to the line containing the points $Q=(0,\ 1,\ 3)$ and $R=(3,\ 5,\ 4)$.

Solution: Draw a picture: this will be the smaller of the lengths of the vectors $(Q-P) \pm \text{proj}_{R-Q}(Q-P)$. Alternatively, we could solve for s in $0=(R-Q)\cdot(P-(Q+s(R-Q))$ to find the point $S = Q+s(R-Q)$ on the line closest to P, and then compute $\|P-S\|$. In either case, the answer is 7.

3.5 Find the scalar *and* vector parametric forms for the following lines:

(b) The line containing $(-5,\ 0,\ 1)$ and which is parallel to the two planes with Cartesian equations $2x-4y+z=0$ and $x-3y-2z=1$.

Solution: This line will be perpendicular to both normals, and so parallel to the cross product of the normals. Hence the scalar parametric form is $x=-5+11t$, $y=5t$, $z=1-2t$, $t\in\mathbb{R}$, and the vector parametric form is $(-5,\ 0,\ 1)+t(11,5,-2)$.

3.6 Find a Cartesian equation for each of the following planes:

(b) The plane parallel to the vector $(1,1,-2)$ and containing the points $P=(1,5,18)$ and $Q=(4,2,-6)$ $5x-3y+z=8$.

Solution: This will be the plane through P with normal parallel to the cross product of $(1,1,-2)$ and $P-Q$. An easy computation then yields the equation $5x-3y+z=8$ (after dividing by 6).

(d) The plane containing the two lines $\{(t-1,6-t,-4+3t) \mid t\in\mathbb{R}\}$ and $\{(-3-4t,6+2t,7+5t) \mid \in\mathbb{R}\}$

Solution: Pick a point on either line, say $P=(-1,6,-4)$. Then this will be the plane through P with normal parallel to the cross product of the direction vectors of each line. An easy computation then yields the equation $11x+17y+2z=83$.

(f) The plane containing the point $P=(1,-1,2)$ and the line $\{(4,-1+2t,2+t) \mid t\in\mathbb{R}\}$.**Solution:**

Pick a point on the line, say $Q=(4,-1,2)$. This will be the plane through P with normal parallel to the cross product of $P-Q$ and a direction vector for the line (say) $(0,2,1)$. One obtains (after division by ±3) the equation $y-2z=-5$.

(h) The plane containing the point $P=(1,\ -7,\ 8)$ which is perpendicular to the line $\{(2+2t,7-4t,-3+t \mid t\in\mathbb{R}\}$. **Solution:** This will be the plane through P with normal paralle to a direction vector for the line. One obtains the equation $2x-4y+z=38$.

3.7 Find a vector parametric form for the planes with Cartesian equations given as follows. (i.e. find some $a\in H$ and two non-zero, non-parallel vectors $u,v\in\mathbb{R}^3$, parallel to the plane H. Then $H=\{a+su+tv \mid s,t\in\mathbb{R}\}$.)

(b) $x-y-2z=4$.

Solution: Take $a=(4,0,0)\in H$. To pick u and v, simply choose two (non-zero, non-parallel) vectors perpendicular to a normal vector $(1,-1,-2)$. So $u=(1,1,0)$ and $v=(2,0,1)$ will do. Then $H=\{(4,0,0)+s(1,1,0)+t(2,0,1)\mid s,t\in\mathbb{R}\}$. (There are of course infinitely many correct answers.)

3.8 Let u,v and w be any vectors in $\mathbb{R}^3$. Which of the following statements could be false, and give an example to illustrate each of your answers.

(1) $u\cdot v=v\cdot u$. **Solution:** This is always true.

(2) $u\times v=v\times u$. **Solution:** Since $u\times v=-v\times u$ always holds, this is only true if u and v are parallel or either is zero. So for a counterexample, take $u=(1,0,0)$ and $v=(0,1,0)$. Then $u\times v=(0,0,1)\neq(0,0,-1)=v\times u$

(3) $u\cdot(v+w)=v\cdot u+w\cdot u$. **Solution:** This is always true.

(4) $(u+2v)\times v=u\times v$. **Solution:** This is always true, since $v\times v=0$ always holds.

(5) $(u\times v)\times w=u\times(v\times w)$. **Solution:** This is almost always false. Indeed by the last question in this set of exercises, it is true only if $(u\cdot w)v-(u\cdot v)w=(w\cdot u)v-(w\cdot v)u \iff (u\cdot v)w=(w\cdot v)u$. So for a counterexample, take $u=(1,0,0)=v$ and $w=(0,1,0)$. Then $(u\times v)\times w=(0,0,0)\neq(0,-1,0)=u\times(v\times w)$.

3.9 Let u,v and w be vectors in $\mathbb{R}^3$. Which of the following statements are (always) true? Explain your answers, including giving examples to illustrate statements which could be false.

(i) $(u\times v)\cdot v=0$. **Solution:** This is always true: it is a property of the cross product which is easliy checked.

(ii) $(v\times u)\cdot v=-1$ **Solution:** This is always false, and the left hand side is always zero. So take $u=v=0$.

(iii) $(u\times v)\cdot w$ is the volume of the of the parallelepiped determined by u, v and w. **Solution:** Volumes are always positive, so this is only true if $(u\times v)\cdot w=|(u\times v)\cdot w|$; so if $u=(1,0,0),v=(0,1,0)$ and $w=(0,0,-1)$, $(u\times v)\cdot w=-1$, which is not the volume of the unit cube.

(iv) $||u\times v||=||u||\,||v||\cos\theta$ where θ is the angle between u and v. **Solution:** This is only true if $\cos\theta=\sin\theta$, since it is always true that $||u\times v||=||u||\,||v||\sin\theta$, where θ is the angle between u and v. So take $u=(1,0,0)$ and $v=(0,1,0)$ for a counterexample.

(v) $|u\cdot v|=||u||\,||v||\cos\theta$ where θ is the angle between u and v. **Solution:** This is always true.

Problems of Chapter 4

4.1 Determine whether the following sets are closed under the indicated rule for addition.

(b) $L=\{(x,y)\in\mathbb{R}^2\mid x-3y=0\}$; standard addition of vectors in $\mathbb{R}^2$.

Solution: This is closed under addition: Suppose $(x,y),(x',y')\in L$. Then $x-3y=0$ and $x'-3y'=0$. Since $(x,y)+x',y')=(x+x',y+y')$ satisfies $x+x'-3(y+y')=(x-3y)+(x'-3y')=0+0=0$, we see that $(x,y)+(x',y')\in L$.

(d) $S = \{(x,y) \in \mathbb{R}^2 \mid xy \geq 0\}$; standard addition of vectors in $\mathbb{R}^2$.

Solution: This is not closed under addition. For example, $(1,2)$ and $(-2,-1)$ both belong to S, but their sum, $(-1,1)$ does not.

(f) $K = \{(x,y,z) \in \mathbb{R}^3 \mid x+2y+z = 1\}$; standard addition of vectors in $\mathbb{R}^3$. **Solution:** This is not closed under addition. For example, both $(1,0,0)$ and $(0,0,1)$ belong to K but their sum, $(1,0,1)$ does not.

(h) $M = \{(x,x+2) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$; *Non-standard addition*: $(x,y)\tilde{+}(x',y') = (x+x',y+y-2)$.

Solution: This *is* closed under the weird addition rule (but not under the standard one- see part (a)): Let $u = (x,x+2)$ and $v = (x',x'+2)$ be any two points in M. Then $u\tilde{+}v = (x+x',(x+2)+(x'+2)-2) = (x+x',x+x'+2) \in M$.

4.2 Determine whether each of the following sets is closed under the indicated rule for multiplication of vectors by scalars.

(b) $L = \{(x,y) \in \mathbb{R}^2 \mid x-3y = 0\}$; standard rule for multiplication of vectors in $\mathbb{R}^2$ by scalars.

Solution: This is closed under multiplication by scalars: Let $u = (x,y) \in L$ (so $x-3y = 0$) and $k \in \mathbb{R}$ be any scalar. Then $ku = (kx,ky)$ satisfies $kx-3(ky) = k(x-3y) = k0 = 0$, and so $ku \in L$.

(d) $S = \{(x,y) \in \mathbb{R}^2 \mid xy \geq 0\}$; standard rule for multiplication of vectors in $\mathbb{R}^2$ by scalars.

Solution: This is closed under multiplication by scalars (despite not being closed under addition- see (d) in Q.1): Let $u = (x,y) \in S$ (so $xy \geq 0$) and $k \in \mathbb{R}$ be any scalar. Then $ku = (kx,ky)$ satisfies $kx(ky) = k^2xy \geq 0$, since $xy \geq 0$ and $k^2 \geq 0$. So $ku \in S$.

(f) $K = \{(x,y,z) \in \mathbb{R}^3 \mid x+2y+z = 1\}$; standard rule for multiplication of vectors in $\mathbb{R}^3$ by scalars.

Solution: This is not closed under multiplication by scalars. For example, $u = (1,0,0) \in K$ but $2u = (2,0,0) \notin K$

(h) $M = \{(x,x+2) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$; *Non-standard multiplication of vectors by scalars* $k \in \mathbb{R}$:

$$k \circledast (x,y) = (kx,ky-2k+2).$$

Solution: This *is* closed under this weird rule for multiplication of vectors by scalars (but not under the standard rule: see part (a)): Let $u = (x,x+2) \in M$, and $k \in \mathbb{R}$. Then$k \circledast u = k \circledast (x,x+2) = (kx,k(x+2)-2k+2) = (kx,kx+2) \in M$!

4.3 Determine whether the following subsets of $\mathbf{F}(\mathbb{R}) = \{f \mid f : \mathbb{R} \to \mathbb{R}\}$ are closed under the standard addition of functions in $\mathbf{F}(\mathbb{R})$. (Recall that $\mathbf{F}(\mathbb{R})$ consists of all real-valued functions of a real variable; i.e., all functions with domain $\mathbb{R}$, taking values in $\mathbb{R}$).

(b) $T = \{f \in \mathbf{F}(\mathbb{R}) \mid f(2) = 1\}$.

Solution: This is not closed under addition: for example, the constant function $f(x) = 1, \forall x \in \mathbb{R}$ belongs to T but $f+f$, which is the constant function 2, does not.

(d) $N = \{f \in \mathbf{F}(\mathbb{R}) \mid \text{ for all } x \in \mathbb{R}, f(x) \leq 0\}$.

Solution: This is closed under addition: Let $f,g \in N$ (so $\forall x \in \mathbb{R}, f(x) \leq 0$ and $g(x) \leq 0$. Then, since the sum of two non-positve numbers is is still non-positive, $\forall x \in \mathbb{R},, (f+g)(x) = f(x)+g(x) \leq 0$. Hence $f+g \in N$.

(f) $O = \{f \in \mathbf{F}(\mathbb{R}) \mid \text{ For all } x \in \mathbb{R}, f(-x) = -f(x)\}$.

Solution: This is the set of all so-called ‘odd’ functions, and it is closed under addition: Let $f, g \in O$. Then, $\forall x \in \mathbb{R}, f(-x) = -f(x)$ and $g(-x) = -g(x)$, so, $\forall x \in \mathbb{R}, (f+g)(-x) = f(-x) + g(-x) = -f(x) - g(x) = -(f(x) + g(x)) = -(f+g)(x)$. Hence $f + g \in O$.

4.4 Determine whether the following sets are closed under the standard rule for multiplication of functions by scalars in $\mathbf{F}(\mathbb{R})$.

(b) $T = \{f \in \mathbf{F}(\mathbb{R}) \mid f(2) = 1\}$.

Solution: This is *not* closed under multiplication by all scalars. For example, the constant function 1 belongs to T, but if we multiply this function by the scalar 2, we obtain the constant function 2, which does not belong to T.

(d) $\{f \in \mathbf{F}(\mathbb{R}) \mid \text{ For all } x \in \mathbb{R}, f(x) \leq 0\}$.

Solution: This is *not* closed under multiplication by all scalars (despite being closed under addition - see part (d) in the previous question): for example, the constant function $g(x) = -1, \forall x \in \mathbb{R}$ belongs to N but $(-1)g$, which is the constant function 1, does not.

(f) $O = \{f \in \mathbf{F}(\mathbb{R}) \mid \text{ For all } x \in \mathbb{R}, f(-x) = -f(x)\}$.

Solution: This is closed under addition: Let $f \in O$. Then, $\forall x \in \mathbb{R}, f(-x) = -f(x)$. Now let $k \in \mathbb{R}$ be any scalar. Then, $\forall x \in \mathbb{R}, (kf)(-x) = k(f(-x)) = k(-f(x)) = -kf(x) = -(kf)(x)$. Hence $kf \in O$.

4.5 Determine whether the following sets are closed under the standard operation of addition of matrices in $\mathbf{M}_{22}(\mathbb{R})$.

(b) $S = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; a + d = 0 \right\}$.

Solution: This is closed under addition: suppose $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $b = \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}$ belong to S. So $a + d = 0$ and $a' + d' = 0$. Then $A + B = \begin{bmatrix} a + a' & b + b' \\ c + c' & d + d' \end{bmatrix}$ satisfies $(a + a') + (d + d') = (a + d) + (a' + d') = 0 + 0 = 0$, and so $A + B \in S$.

(d) $U = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; ad = 0 \right\}$.

Solution: This is not closed under addition: for example, $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ both belong to U, but $A + B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ does not.

4.6 Determine whether the following sets are closed under the standard rule for multiplication of matrices by scalars in $\mathbf{M}_{22}(\mathbb{R})$.

(b) $\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; a + d = 0 \right\}$.

Solution: This is closed under multiplication by scalars: suppose $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ belongs to S. So $a+d=0$. If $k \in \mathbb{R}$ is any scalar, Then $kA = \begin{bmatrix} ka & kb \\ kc & kd \end{bmatrix}$ satisfies $ka+kd = k(a+d) = k0 = 0$, and so $kA \in S$.

(d) $U = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; ad = 0 \right\}$.

Solution: This *is* closed under multiplication by scalars (despite not being closed under addition - see part (d) of the previous question): suppose $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ belongs to U. So $ad = 0$. If $k \in \mathbb{R}$ is any scalar, Then $kA = \begin{bmatrix} ka & kb \\ kc & kd \end{bmatrix}$ satisfies $(ka)(kd) = k^2(ad) = k^2 0 = 0$, and so $kA \in U$.

4.7 The following sets have been given the indicated rules for addition of vectors, and multiplication of objects by real scalars (the so-called *'vector operations'*). If possible, check if there is a zero vector in the subset in each case. If it is possible, show your choice works in *all* cases, and if it is not possible, give an example to illustrate your answer.

(Note: in the last two parts, since the vector operations are not the standard ones, the zero vector will probably not be the one you're accustomed to.)

(b) $L = \{(x,y) \in \mathbb{R}^2 \mid x - 3y = 0\}$; standard vectors operations in $\mathbb{R}^2$.

Solution: Since the operations are standard, the zero vector is the standard one, namely $(0,0)$. Since $0 - 3(0) = 0$, $(0,0) \in L$.

(d) $S = \{(x,y) \in \mathbb{R}^2 \mid xy \geq 0\}$; standard vectors operations in $\mathbb{R}^2$.

Solution: Since the operations are standard, the zero vector is the standard one, namely $(0,0)$. Since $00 = 0 \geq 0$, $(0,0) \in S$.

(f) $K = \{(x,y,z) \in \mathbb{R}^3 \mid x + 2y + z = 1\}$; standard vectors operations in $\mathbb{R}^3$.

Solution: Since the operations are standard, the zero vector is the standard one, namely $(0,0,0)$. However, $0 + 2(0) + 0 = 0 \neq 1$, so $(0,0,0) \notin K$.

(h) $M = \{(x, x+2) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$; *Non-standard operations:–* Addition: $(x,y)\tilde{+}(x',y') = (x + x', y + y' - 2)$. Multiplication of vectors by scalars $k \in \mathbb{R}$: $k \circledast (x,y) = (kx, ky - 2k + 2)$.

Solution: Since the operations are *not* standard, the zero vector is unlikely to be the standard one. Let's find out what it might be: we need $\tilde{O} = (a,b)$, such that $(x,y)\tilde{+}(a,b) = (x,y)$, for all $(x,y) \in M$. So we need $(x,x+2)\tilde{+}(a,b) = (x,x+2)$ for all $x \in \mathbb{R}$. Well, $(x,x+2)\tilde{+}(a,b) = (x+a,(x+2)+b-2) = (x+a, x+b)$. So $(x,x+2)\tilde{+}(a,b) = (x,x+2)$ for all x iff $x = x+a$ and $x+b = x+2$ for all $x \in \mathbb{R}$. So we need $a = 0$ and $b = 2$. So the vector $\tilde{O} = (0,2)$ *works* as the zero vector in this case. Moreover, as you can see, $(0,2) \in M$, so this set with the weird operations *does indeed have a zero!*

4.8 Explain your answers to the following:

(a) Determine whether the zero function (let's denote it $\mathbf{0}$) of $\mathbf{F}(\mathbb{R})$ belongs to each of the subsets in question 3.

(3b) $T = \{f \in \mathbf{F}(\mathbb{R}) \mid f(2) = 1\}$.

Solution: Since $\mathbf{0}(2) = 0 \neq 1$, this set does not contain the zero function.

(3d) $\{f \in \mathbf{F}(\mathbb{R}) \mid \text{ For all } x \in \mathbb{R},\, f(x) \leq 0\}$. Since $\mathbf{0}(x) = 0 \leq 0$ for all $x \in \mathbb{R}$, this set does contain the zero function.

(3f) $O = \{f \in \mathbf{F}(\mathbb{R}) \mid \text{ For all } x \in \mathbb{R},\, f(-x) = -f(x)\}$. Since $\mathbf{0}(-x) = 0 = -0 = -\mathbf{0}(x)$ for all $x \in \mathbb{R}$, this set does contain the zero function.

4.9 The following sets have been given the indicated rules for addition of vectors, and multiplication of objects by real scalars. In each case, If possible, check if vector in the subset has a 'negative' in the subset.

Again, since the vector operations are not the standard ones, the negative of a vector will probably not be the one you're accustomed to seeing.

(a) $M = \{(x, x+2) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$; <u>*Non-standard Operations:–*</u> Addition:

$$(x,y)\tilde{+}(x',y') = (x+x', y+y-2).$$

Multiplication of vectors by scalars $k \in \mathbb{R}$: $k \circledast (x,y) = (kx, ky-2k+2)$.

Solution: To find the negative of a vector (if it exists), we need to know the zero. But we found this in Q. 7(h): $\tilde{0} = (0,2)$ is the zero for this weird addition. To find the negative of a vector $u = (x, x+2) \in M$, we need to solve the equation $(x,x+2)\tilde{+}(c,d) = \tilde{0} = (0,2)$ for c and d. But $(x,x+2)\tilde{+}(c,d) = (x+c, (x+2)+d-2) = (x+c, x+d)$, so we need $x+c = 0$ and $x+d = 2$. Thus, $c = -x$ and $d = 2-x$, so that the negative of $(x,x+2)$ is actually $(-x, 2-x)$.

Now, since $2-x = (-x)+2$, this puts $(-x, 2-x)$ in M, i.e.$(-x, 2-x)$ really is of the form $(x', x'+2)$ – take $x' = -x$! So this set *does* contain the negative of every element!

4.10 Explain your answers to the following:

(b) Determine whether the subsets of $\mathbf{F}(\mathbb{R})$ in question 3, equipped with the standard vector operations of $\mathbf{F}(\mathbb{R})$ are vector spaces.

(3b) $T = \{f \in \mathbf{F}(\mathbb{R}) \mid f(2) = 1\}$.

Solution: Since this set does not contain the zero function (see Q.8(a)), it is not a vector space.

(3d) $\{f \in \mathbf{F}(\mathbb{R}) \mid \text{ For all } x \in \mathbb{R},\, f(x) \leq 0\}$.

Solution: We saw in Q.4(d) that this set is not closed under multiplication by scalars, so it is not a vector space.

(3f) $O = \{f \in \mathbf{F}(\mathbb{R}) \mid \text{ For all } x \in \mathbb{R},\, f(-x) = -f(x)\}$.

Solution: We saw in previous questions that O is closed under addition, under multiplication by scalars, and has a zero. There remains the existence of negatives, and the 6 arithmetic axioms.

To see that 0 has negatives, let $f \in 0$. Then $\forall x \in \mathbb{R}, f(-x) = -f(x)$. Define a function $g : \mathbb{R} \to \mathbb{R}$ by $g(x) = -f(x), \forall x \in \mathbb{R}$. It is clear that $f + g = \mathbf{0}$, so it remains to see that $g \in O$. But, $\forall x \in \mathbb{R}, g(-x) = -f(-x) = -(-f(x)) = f(x) = -g(x)$, so indeed $g \in O$.

The arithmetic axioms are identities that hold for *all* functions in $\mathbf{F}(\mathbb{R})$, so in particular these identities are satisfied for any subset. In particular, all the arithmetic axioms hold for O.

Thus O, with the standard operations inherited from $\mathbf{F}(\mathbb{R})$, is indeed itself a vector space.

4.11 Justify your answers to the following:

(a) Equip the set $V = \mathbb{R}^2$ with the <u>*non-standard operations:–*</u> Addition:

$$(x,y)\tilde{+}(x',y') = (x+x', y+y-2).$$

Multiplication of vectors by scalars $k \in \mathbb{R}$:

$$k \circledast (x,y) = (kx, ky - 2k + 2).$$

Check that $\mathbb{R}^2$, with these new operations, is indeed a vector space.

Solution: It is clear that $\mathbb{R}^2$ with these operations is closed under these weird operations —look at the right hand side of the definitions: they live in $\mathbb{R}^2$. We saw in Q.7(h) that the vector $\tilde{0} = (0,2)$ works as a zero in the subset we called M. It's easy to check it works on all of $\mathbb{R}^2$. We also saw in Q.9(a) that the negative of (x,y) was $(-x, 2-y)$, and you can check that this works for all of $\mathbb{R}^2$.

That leaves us with 6 aritmetic axioms to check. Here, I will check only 3, and leave the rest to you. (They all hold.)

Let's check the distributive axiom: $k \circledast (u \tilde{+} v) = k \circledast u \tilde{+} k \circledast v$.

Well,

$$\begin{aligned} k \circledast ((x,y) \tilde{+} (x',y')) &= k \circledast (x+x', y+y'-2) \\ &= (k(x+x'), k(y+y'-2) - 2k + 2) \\ &= (kx + kx', ky + ky' - 2k - 2k + 2) \\ &= (kx + ky, (ky - 2k + 2) + (ky' - 2k + 2) - 2) \\ &= (kx, ky - 2k + 2) \tilde{+} (kx', ky' - 2k + 2) \\ &= k \circledast (x,y) \tilde{+} \circledast (x',y') \end{aligned}$$

Hence the distributive law holds!

Let's check the axiom: $1 \circledast u = u$: Well, $1 \circledast (x,y) = (x, y - 2(1) + 2) = (x,y)$, so that's OK too.

The last one I'll check is that for all $k,l \in \mathbb{R}$ and for all $u \in \mathbb{R}^2$,

$$(k+l) \circledast u = (k \circledast u) \tilde{+} (l \circledast u).$$

Well,

$$\begin{aligned} (k+l) \circledast (x,y) &= ((k+l)x, (k+l)y - 2(k+l) + 2) \\ &= (kx + lx, ky + ly - 2k - 2l + 2) \\ &= (kx + lx, (ky - 2k + 2) + (ly - 2l + 2) - 2) \\ &= (kx, ky - 2k + 2) \tilde{+} (lx, ly - 2l + 2) \\ &= (k \circledast (x,y) \tilde{+} (l \circledast (x,y)), \end{aligned}$$

as required.

Problems of Chapter 5

Note that in the following, I have used Theorem 6.4.1 extensively: I suggest you read ahead and acquaint yourself with this very useful result!

5.1 Determine whether each of the following is a subspace of the indicated vector space. Assume the vector space has the standard operations unless otherwise indicated.

(b) $\{(x, x-3) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$; $\mathbb{R}^2$.

Solution: Since this set does not contain $(0,0)$, it is *not* a subspace of $\mathbb{R}^2$.

(e) $\{(x,y) \in \mathbb{R}^2 \mid x-3y=0\}$; $\mathbb{R}^2$.

Solution: This is a line through the origin in $\mathbb{R}^2$ and hence *is* a subspace of $\mathbb{R}^2$. Alternatively, $\{(x,y) \in \mathbb{R}^2 \mid x-3y=0\} = \{(3y,y) \mid y \in \mathbb{R}\} = \{y(3,1) \mid y \in \mathbb{R}\} = \text{span}\{(3,1)\}$ and hence *is* a subspace of $\mathbb{R}^2$.

(g) $\{(x,y) \in \mathbb{R}^2 \mid xy \geq 0\}$; $\mathbb{R}^2$.

Solution: We saw in previous exercises that this set is not closed under addition, and hence is *not* a subspace of $\mathbb{R}^2$.

(i) $\{(x,y,z) \in \mathbb{R}^3 \mid x+2y+z=1\}$; $\mathbb{R}^3$.

Solution: This set does not contain $(0,0,0)$, and hence is *not* a subspace of $\mathbb{R}^2$.

(k) $W = \{(x,y,z,w) \in \mathbb{R}^4 \mid x-y+z-w=0\}$; $\mathbb{R}^4$.

Solution:

$$\begin{aligned} W &= \{(x,y,z,w) \in \mathbb{R}^4 \mid x = y-z+w\} \\ &= \{(y-z+w,y,z,w,) \mid y,z,w \in \mathbb{R}\} \\ &= \{y(1,1,0,0)+z(-1,0,1,0)+w(1,0,0,1) \mid y,z,w \in \mathbb{R}\} \\ &= \text{span}\{(1,1,0,0),(-1,0,1,0),(1,0,0,1)\} \end{aligned}$$

and hence *is* a subspace of $\mathbb{R}^4$.

5.2 Determine whether each of the following is a subspace of

$$\mathbf{F}(\mathbb{R}) = \{f \mid f : \mathbb{R} \to \ \mathbb{R}\},$$

with its standard operations. (Here, you'll need to use the subspace test, except perhaps in the last part.)

(b) $\{f \in \mathbf{F}(\mathbb{R}) \mid f(2)=1\}$.

Solution: This does not contain the zero function and hence is not a subspace of $\mathbf{F}(\mathbb{R})$.

(d) $\{f \in \mathbf{F}(\mathbb{R}) \mid \text{ for all } x \in \mathbb{R},\, f(x) \leq 0\}$.

Solution: We saw in previous exercises that this set is not closed under multiplication by scalars, and so is not a subspace of $\mathbf{F}(\mathbb{R})$.

(f) $O = \{f \in \mathbf{F}(\mathbb{R}) \mid \text{ For all } x \in \mathbb{R},\, f(-x) = -f(x)\}$.

Solution: Refer to solutions for exercises in the previous chapter: Q. 3(f), Q.4(f) and Q.8(a): put them together and you'll see the subspace test is carried out successfully. Hence O is indeed a subspace of $\mathbf{F}(\mathbb{R})$.

(h) $\mathbb{P} = \{p \in \mathbf{F}(\mathbb{R}) \mid p \text{ is a polynomial function in the variable } x\}$

Solution: Since the zero function is also a polynomial function, $\mathbf{0} \in \mathbb{P}$. Noting that the sum of any two polynomial functions is again a polynomial function shows $\mathbb{P}$ is closed under addition. Finally, it's also clear that a scalar multiple of a polynomial function is again a polynomial function, so $\mathbb{P}$ is closed under multiplication by scalars. Hence, by the subspace test, $\mathbb{P}$ is a subspace of $\mathbf{F}(\mathbb{R})$.

(Alternatively, note that $\mathbb{P} = \text{span}\{x^n \mid n = 0,1,2,\dots\}$ and so is a subspace of $\mathbf{F}(\mathbb{R})$. We won't often talk about the span of an infinite set of vectors (say) K, but the definition of $\text{span}\, K$ is the same: collect all (finite) linear combinations of vectors from K.)

5.3 Determine whether the following are subspaces of

$$\mathbb{P} = \{p \in \mathbf{F}(\mathbb{R}) \mid p \text{ is a polynomial function in the variable } x\},$$

with its standard operations. (In some parts, you'll be able to use the fact that everything of the form $\operatorname{span}\{v_1, \ldots, v_n\}$ is a subspace.)

(b) $\{p \in \mathbb{P} \mid \deg(p) \leq 2\}$.

Solution: $\{p \in \mathbb{P} \mid \deg(p) \leq 2\} = \{a + bx + cx^2 \mid a, b, c \in \mathbb{R}\} = \operatorname{span}\{1, x, x^2\}$ and hence this is a subspace of $\mathbb{P}$.

(d) $\{p \in \mathbb{P}_2 \mid p(1) = 0\}$.

Solution: By the Factor theorem, $\{p \in \mathbb{P}_2 \mid p(1) = 0\} = \{(x-1)q(x) \mid \deg q \leq 1\} = \{(x-1)(a + bx) \mid a, b \in \mathbb{R}\} = \{a(x-1) + bx(x-1) \mid a, b \in \mathbb{R}\} = \operatorname{span}\{x-1, x(x-1)\}$ and hence is a subspace of $\mathbb{P}$.

(f) $G = \{p \in \mathbb{P}_3 \mid p(2)\,p(3) = 0\}$.

Solution: This is not closed under addition: for example $(x-2)$ and $x-3$ both belong to G but their sum, $r(x) = 2x - 5$, does not, as $r(2)r(3) = (-1)(1) = -1 \neq 0$. Hence G is *not* a subspace of $\mathbb{P}$.

(h) $\{p \in \mathbb{P}_2 \mid p(1) + p(-1) = 0\}$.

Solution: $\{p \in \mathbb{P}_2 \mid p(1) + p(-1) = 0\} = \{a + bx + cx^2 \mid a, b, c \in \mathbb{R} \text{ and } a + b + c + a - b + c = 0\} = \{a + bx + cx^2 \mid a, b, c \in \mathbb{R} \text{ and } a + c = 0\} = \{a + bx - ax^2 \mid a, c \in \mathbb{R}\} = \operatorname{span}\{1 - x^2, x\}$ and hence this is a subspace of $\mathbb{P}$.

5.4 Determine whether the following are subspaces of $\mathbf{M}_{22}(\mathbb{R})$, with its standard operations. (In some parts, you'll be able to us the fact that everything of the form $\operatorname{span}\{v_1, \ldots, v_n\}$ is a subspace.)

(b) $X = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; a = d = 0 \quad \& \quad b = -c \right\}$.

Solution: $X = \left\{ \begin{bmatrix} 0 & -c \\ c & 0 \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; c \in \mathbb{R} \right\} = \operatorname{span}\left\{ \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\}$, and hence X is a subspace of $\mathbf{M}_{22}(\mathbb{R})$.

(d) $\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; bc = 1 \right\}$.

Solution: This set does not contain the zero matrix $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ and hence is *not* a subspace of $\mathbf{M}_{22}(\mathbb{R})$.

(f) $Z = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; ad - bc = 0 \right\}$.

Solution: This set is not closed under addition: for example both $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ belong to Z, but $A + B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ does not. Hence Z is *not* a subspace of $\mathbf{M}_{22}(\mathbb{R})$.

Problems of Chapter 6

6.1 Justify your answers to the following:

(b) Is the vector $(1,2)$ a linear combination of $(1,1)$ and $(2,2)$?

Solution: No, it isn't: suppose $(1,2) = a(1,1) + b(2,2)$. Then $a+2b = 1$ and $a+2b = 2$, which is impossible.

(d) Is the vector $(1,2,2,3)$ a linear combination of $(1,0,1,2)$ and $(0,0,1,1)$?

Solution: No, since any linear combination of $(1,0,1,2)$ and $(0,0,1,1)$ will have 0 as its second component, and the second component of $(1,2,2,3)$ is $2 \neq 0$.

(f) Is the matrix $C = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$ a linear combination of $A = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$?

Solution: No: If $C = kA + lB$ for $k, l \in \mathbb{R}$, then comparing the $(1,1)$ and the $(2,1)$–entries of both sides, we obtain $k = 1$ and $k = 2$, which is impossible.

(h) Is the polynomial $1 + x^2$ a linear combination of $1 + x - x^2$ and x?

Solution: No: if $1 + x^2 = a(1 + x - x^2) + bx$, then $(a+1)x^2 - (a+b)x + 1 - a = 0$ for all $x \in \mathbb{R}$. But a non-zero quadratic equation has at most two roots, so $a+1 = 0$. Then, $-(a+b)x + 1 - a = 0$ for all $x \in \mathbb{R}$. But a non-zero linear equation has at most one roots, so $a + b = 0$. Then we're left with $1 - a = 0$, which is impossible if $a + 1 = 0$.

(j) Is the function $\sin x$ a linear combination of the constant function 1 and $\cos x$?

Solution: No: Suppose there were scalars $a, b \in \mathbb{R}$ such that $\sin x = a1 + b\cos x$, for all $x \in \mathbb{R}$. For $x = 0$ we obtain the equation $a + b = 0$, and for $x = \pi$, we obtain the equation $a - b = 0$, implying that $a = b = 0$. But then $\sin x = 0$, for all $x \in \mathbb{R}$, which is nonsense, as $\sin(\frac{\pi}{2}) = 1 \neq 0$.

(l) If u, v and w are any vectors in a vector space V, is $u - v$ a linear combination of u, v and w?

Solution: Yes, indeed: $u - v = (1)u - (1)v + (0)w$

6.2 Justify your answers to the following:

(b) Is '$(3,4) \in \operatorname{span}\{(1,2)\}$' true? (Note that this is the same as the previous part, written using mathematical notation.)

Solution: No: if $(3,4) \in \operatorname{span}\{(1,2)\}$, then $(3,4) = a(1,2)$ for some $a \in \mathbb{R}$, which implies $a = 3$ and $a = 2$, which is nonsense.

(d) How many vectors belong to $\operatorname{span}\{(1,2)\}$?

Solution: There are infinitely many vectors in $\operatorname{span}\{(1,2)\} = \{a(1,2) \mid a \in \mathbb{R}\}$, since if $a \neq a'$ are distinct real numbers, then $a(1,2) \neq a'(1,2)$.

(f) Is $\{(1,2)\}$ a subset of $\operatorname{span}\{(1,2)\}$?

Solution: Indeed it is: since $(1,2)$ is clearly a multiple of $(1,2)$, $(1,2) \in \operatorname{span}\{(1,2)\} = \{a(1,2) \mid a \in \mathbb{R}\}$.

(h) Suppose S is a subset of a vector space V. If $S = \operatorname{span} S$, explain why S must be a subspace of V.

Solution: Since the span of any set of vectors is a subspace, and $S = \operatorname{span} S$, S is a subspace.

6.3 Give <u>two</u> distinct *finite* spanning sets for each of the following subspaces. (Note that there will be infinitely many correct answers; I've just given one example.)

(b) $\{(x,y)\in\mathbb{R}^2 \mid 3x-y=0\}$.

Solution: Since

$$\{(x,y)\in\mathbb{R}^2 \mid 3x-y=0\} = \{(x,3x) \mid x\in\mathbb{R}\} = \{x(1,3) \mid x\in\mathbb{R}\} = \{a(2,6) \mid a\in\mathbb{R}\},$$

Both $\{(1,3)\}$ and $\{(2,6)\}$ are spanning sets for $\{(x,y)\in\mathbb{R}^2 \mid 3x-y=0\}$.

(d) $U=\{(x,y,z,w)\in\mathbb{R}^4 \mid x-y+z-w=0\}$.

Solution: We saw in Q.1(k) that

$$\{(x,y,z,w)\in\mathbb{R}^4 \mid x-y+z-w=0\} = \operatorname{span}\{(1,1,0,0),(-1,0,1,0),(1,0,0,1)\}$$

So $\{(1,1,0,0),(-1,0,1,0),(1,0,0,1)\}$ is one spanning set for U. If we simply multiply these spanning vectors by non-zero scalars, we still obtain a spanning set. So, for example, $\{(2,2,0,0),(-3,0,3,0),(4,0,0,4)\}$ is another spanning set for U.

(f) $X=\left\{\begin{bmatrix} a & b\\ c & d\end{bmatrix}\in\mathbf{M}_{22}(\mathbb{R}) \;\middle|\; a=d=0 \quad \& \quad b=-c\right\}$.

Solution: We saw in Q.4(b) that $X=\operatorname{span}\left\{\begin{bmatrix} 0 & -1\\ 1 & 0\end{bmatrix}\right\}$, so $\left\{\begin{bmatrix} 0 & -1\\ 1 & 0\end{bmatrix}\right\}$ is one spanning set for X. Thus $\left\{\begin{bmatrix} 0 & -2\\ 2 & 0\end{bmatrix}\right\}$ is another.

(h) $V=\left\{\begin{bmatrix} a & 0\\ 0 & b\end{bmatrix}\in\mathbf{M}_{22}(\mathbb{R}) \;\middle|\; a,b\in\mathbb{R}\right\}$.

Solution: $V=\left\{a\begin{bmatrix} 1 & 0\\ 0 & 0\end{bmatrix}+b\begin{bmatrix} 0 & 0\\ 0 & 1\end{bmatrix} \;\middle|\; a,b\in\mathbb{R}\right\}$, so $\left\{\begin{bmatrix} 1 & 0\\ 0 & 0\end{bmatrix},\begin{bmatrix} 0 & 0\\ 0 & 1\end{bmatrix}\right\}$ is one spanning set for V. Then, $\left\{\begin{bmatrix} 2 & 0\\ 0 & 0\end{bmatrix},\begin{bmatrix} 0 & 0\\ 0 & 3\end{bmatrix}\right\}$ is another spanning set for V.

(j) $U=\left\{\begin{bmatrix} a & b\\ c & d\end{bmatrix}\in\mathbf{M}_{22}(\mathbb{R}) \;\middle|\; a+b+c+d=0\right\}$.

Solution:
Since

$$U=\left\{\begin{bmatrix} a & b\\ c & d\end{bmatrix}\in\mathbf{M}_{22}(\mathbb{R}) \;\middle|\; a=-b-c-d\right\}=\left\{\begin{bmatrix} -b-c-d & b\\ c & d\end{bmatrix}\in\mathbf{M}_{22}(\mathbb{R}) \;\middle|\; b,c,d\in\mathbb{R}\right\},$$

then $U=\left\{b\begin{bmatrix} -1 & 1\\ 0 & 0\end{bmatrix}+c\begin{bmatrix} -1 & 0\\ 1 & 0\end{bmatrix}+d\begin{bmatrix} -1 & 0\\ 0 & 1\end{bmatrix} \;\middle|\; b,c,d\in\mathbb{R}\right\}=\operatorname{span}\left\{\begin{bmatrix} -1 & 1\\ 0 & 0\end{bmatrix},\begin{bmatrix} -1 & 0\\ 1 & 0\end{bmatrix},\begin{bmatrix} -1 & 0\\ 0 & 1\end{bmatrix}\right\}$,

and so $\left\{\begin{bmatrix} -1 & 1\\ 0 & 0\end{bmatrix},\begin{bmatrix} -1 & 0\\ 1 & 0\end{bmatrix},\begin{bmatrix} -1 & 0\\ 0 & 1\end{bmatrix}\right\}$ is one spanning set for U. Clearly, $\left\{\begin{bmatrix} -2 & 2\\ 0 & 0\end{bmatrix},\begin{bmatrix} -1 & 0\\ 1 & 0\end{bmatrix},\begin{bmatrix} -2 & 0\\ 0 & 2\end{bmatrix}\right\}$

is another, distinct one — even though they do have one matrix in common, as sets they are different. Remember: two sets are the same iff they contain exactly the same elements.

(l) $\mathscr{P}_n = \{p \mid p \text{ is a polynomial function with} \deg(p) \leq n\}$.

Solution: $\mathscr{P}_n = \{a_0 + a_1x + \cdots a_nx^n \mid a_0, a_1, \ldots, a_n \in \mathbb{R}\} = \operatorname{span}\{1, x, \ldots, x^n\}$, so $\{1, x, \ldots, x^n\}$ is one spanning set for $\mathscr{P}_n$. Clearly, $\{2, x, \ldots, x^n\}$ is a different spanning set.

(n) $Y = \{p \in \mathscr{P}_3 \mid p(2) = p(3) = 0\}$.

Solution: By the Factor theorem,

$$Y = \{(x-2)(x-3)q(x) \mid \deg q \leq 1\} = \{(x-2)(x-3)(a+bx) \mid a, b \in \mathbb{R}\},$$

so $Y = \{a(x-2)(x-3) + bx(x-2)(x-3) \mid a, b \in \mathbb{R}\} = \operatorname{span}\{(x-2)(x-3), x(x-2)(x-3)\}$. Hence, $\{(x-2)(x-3), x(x-2)(x-3)\}$ is one spanning set for Y. Then, $\{2(x-2)(x-3), x(x-2)(x-3)\}$ is a different spanning set for Y.

(p) $W = \operatorname{span}\{\sin x, \cos x\}$.

Solution: This one's easy! We are explicitly given one spanning set in the definition of W, namely $\{\sin x, \cos x\}$. So $\{\sin x, 2\cos x\}$ is a different spanning set for W.

(r) $Z = \operatorname{span}\{1, \sin^2 x, \cos^2 x\}$.

Solution: Again, we are given one: $\{1, \sin^2 x, \cos^2 x\}$. Another, smaller spanning set for Z is $\{1, \sin^2 x\}$, since $\cos^2 x = 1 - \sin^2 x$. (Anything in Z is of the form $a + b\sin^2 x + c\cos^2 x$ for some $a, b, c \in \mathbb{R}$. But $a + b\sin^2 x + c\cos^2 x = a + b\sin^2 x + c(1 - \sin^2 x) = (a+c) + (b-c)\sin^2 x$, so everything in Z is in fact a linear combination of 1 and $\sin^2 x$.)

6.4 Justify your answers to the following:

(b) Suppose u and v belong to a vector space V. Show carefully that $\operatorname{span}\{u, v\} = \operatorname{span}\{u - v, u + v\}$. That is , you must show two things:

(i) If $w \in \operatorname{span}\{u, v\}$, then $w \in \operatorname{span}\{u - v, u + v\}$, and
(ii) If $w \in \operatorname{span}\{u - v, u + v\}$, then $w \in \operatorname{span}\{u, v\}$.

Solution: (i) If $w \in \operatorname{span}\{u, v\}$, then $w = au + bv$ for some scalars a, b. But

$$au + bv = \frac{(a-b)}{2}(u - v) + \frac{(a+b)}{2}(u + v),$$

so $w \in \operatorname{span}\{u - v, u + v\}$.
(ii) If $w \in \operatorname{span}\{u - v, u + v\}$, then $w = a(u - v) + b(u + v)$ for some scalars a, b. But

$$a(u - v) + b(u + v) = (a + b)u + (b - a)v \in \operatorname{span}\{u, v\},$$

so $w \in \operatorname{span}\{u, v\}$.

(d) Suppose $\operatorname{span}\{v, w\} = \operatorname{span}\{u, v, w\}$. Show carefully that $u \in \operatorname{span}\{v, w\}$.

Solution: Since $\operatorname{span}\{u, v, w\} = \operatorname{span}\{v, w\}$, and $u = 1u + 0v + 0w \in \operatorname{span}\{u, v, w\}$, then $u \in \operatorname{span}\{v, w\}$.

(f) Show that $x^{n+1} \notin \mathscr{P}_n$. [1]

[1] *Hint: Generalize the idea in the previous hint, recalling that every non-zero polynomial of degree $n+1$ has at most $n+1$ distinct roots.*

Solution: Suppose $x^{n+1} \in \mathscr{P}_n$. So, $x^{n+1} = a_0 + a_1x + \cdots + a_nx^n$ for some scalars $a_0, \ldots, a_n$. Rewriting this, we see that

$$a_0 + a_1x + \cdots + a_nx^n - x^{n+1} = 0$$

for *every real number x!* But a non-zero polynomial of degree $n+1$ (which is what we have here) has at most $n+1$ distinct roots, that is, $a_0 + a_1x + \cdots + a_nx^n - x^{n+1} = 0$ for at most $n+1$ different real numbers x. But there are more than $n+1$ real numbers, so matter how big n may be. Hence we have a contradiction, and so our original assumption $x^{n+1} \in \mathscr{P}_n$ must be false. So indeed, $x^{n+1} \notin \mathscr{P}_n$.

(h) Assume for the moment (we'll prove it later) that if W is a subspace of V, and V has a finite spanning set, then so does W.

Use this fact and the previous part to prove that $\mathbf{F}(\mathbb{R})$ does not have a finite spanning set.

Solution: If $\mathbf{F}(\mathbb{R})$ has a finite spanning set, and $\mathscr{P}$ is a subspace of $\mathbf{F}(\mathbb{R})$ by the assumption, $\mathscr{P}$ would have a finite spanning set. But the previous part states that this is impossible, so indeed $\mathbf{F}(\mathbb{R})$ cannot have a finite spanning set.

Problems of Chapter 7

7.1 Which of the following sets are linearly independent in the indicated vector space? (If you say they are, you must prove it using the definition; if you say the set is dependent, you must give a non-trivial linear dependence relation that supports your answer. For example, if you say $\{v_1, v_2, v_3\}$ is dependent, you must write something like $v_1 - 2v_2 + v_3 = 0$, or $v_1 = 2v_2 - v_3$.)

(b) $\{(1,1),(2,2)\}$; $\mathbb{R}^2$.

Solution: This set is dependent, since $2(1,1) - (2,2) = 0$.

(d) $\{(1,1),(1,2),(1,0)\}$; $\mathbb{R}^2$.

Solution: This set is dependent, since $2(1,1) - (1,2), -(1,0) = (0,0)$

(f) $\{(1,1,1),(1,0,3),(0,0,0)\}$; $\mathbb{R}^3$.

Solution: This set is dependent, as it contains the zero vector. Explicitly:

$$0(1,1,1) + 0(1,0,3) + 1\,(0,0,0) = (0,0,0).$$

(h) $\{(1,1,1),(1,0,3),(0,3,4)\}$; $\mathbb{R}^3$.

Solution: This set is independent: Suppose $a(1,1,1) + b(1,0,3) + c(0,3,4) = (0,0,0)$. Equating components of each side, we obtain the equations $a+b=0$, $a+3c=0$ and $a+3b+4c=0$. This system of equations has only one solution, namely $a=b=c=0$.

(j) $\{(0,-3),(3,0)\}$; $\mathbb{R}^2$.

Solution: This set is independent, as $a(0,-3) + b(3,0) = (0,0)$ implies $3b = 0$ and $-3a = 0$, and so $a = b = 0$.

(l) $\{(1,0,0),(2,0,-2)\}$; $\mathbb{R}^3$.

Solution: This set is independent, as $a(1,0,0)+b(2,0,-2)=(0,0,0)$ implies $a+2b=0$, and $-2b=0$, and hence $a=b=0$.

7.2 Which of the following sets are linearly independent in $\mathbf{M}_{22}$? (If you say they are, you must prove it using the definition; if you say set is dependent, you must give a non-trivial linear dependence relation that supports your answer. For example, if you say $\{v_1,v_2,v_3\}$ is dependent, you must write something like $v_1-2v_2+v_3=0$, or $v_1=2v_2-v_3$.)

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

Solution: This set is independent: Suppose $a\begin{bmatrix}1 & 0\\ 1 & 2\end{bmatrix}+b\begin{bmatrix}0 & 1\\ 0 & 1\end{bmatrix}+c\begin{bmatrix}1 & -2\\ -1 & 0\end{bmatrix}=\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}$. This yields the equations $a+c=0$, $b-2c=0$, $a-c=0$ and $2a+b=0$. The first and third of these equations imply $a=c=0$, and so using either the second or the fourth yields $b=0$. Hence $a=b=c=0$.

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

Solution: We saw in class that this set is independent. Consult your notes.

7.3 Which of the following sets are linearly independent in the indicated vector space? ? (If you say they are, you must prove it using the definition; if you say set is dependent, you must give a non-trivial linear dependence relation that supports your answer. For example, if you say $\{v_1,v_2,v_3\}$ is dependent, you must write something like $v_1-2v_2+v_3=0$, or $v_1=2v_2-v_3$.)
(b) $\{1,1+x,x^2\}$; $\mathscr{P}_2$.

Solution: This set of polynomials is independent: Suppose $a1+b(1+x)+cx^2=0$, for every $x\in\mathbb{R}$. Then, $(a+b)+bx+cx^2=0$, for every $x\in\mathbb{R}$. But a non-zero polynomial of degree 2 has at most 2 different roots, and $(a+b)+bx+cx^2=0$ has infinitely many, so this polynomial must be the zero polynomial. This means that $a+b=0$, $b=0$ and $c=0$. But this easliy implies $a=b=c=0$.

(d) $\{1,\sin x,2\cos x\}$; $\mathbf{F}(\mathbb{R})$.

Solution: This set is independent: Suppose $a1+b\sin x+c\,2\cos x=0$ for every $x\in\mathbb{R}$. In particular, for $x=0$, we obtain the equation $a+2c=0$; for $x=\pi$ we obtain $a-2c=0$, and for $x=\frac{\pi}{2}$, we obtain $a+b=0$. The first two equations here imply $a=c=0$ and then with this, the last implies $b=0$. Hence $a1+b\sin x+c\,2\cos x=0$ for every $x\in\mathbb{R}$ implies $a=b=c=0$.

(f) $\{\cos 2x,\sin^2 x,\cos^2 x\}$; $\mathbf{F}(\mathbb{R})$.

Solution: This set is dependent. Remember the double angle formula: $\cos 2x=\cos^2 x-\sin^2 x$ holds *for every* $x\in\mathbb{R}$. Hence we have the identity $\sin 2x+\sin^2 x-\cos^2 x=0$, *for every* $x\in\mathbb{R}$. This shows that three functions $\cos 2x,\sin^2 x$ and $\cos^2 x$ are linearly dependent.

(h) $\{\sin 2x,\sin x\cos x\}$; $\mathbf{F}(\mathbb{R})$.

Solution: Remember the double angle formula: $\sin 2x=2\sin x\cos x$ holds *for every* $x\in\mathbb{R}$. Hence we have the identity $\sin 2x-2\sin x\cos x=0$, *for every* $x\in\mathbb{R}$. This shows that two functions $\sin 2x$ and $\sin x\cos x$, are linearly dependent.
7.4 Justify your answers to the following:

(b) Suppose V is a vector space, and a subset $\{v_1,\dots,v_k\}\subset V$ is known to be linearly independent. Show carefully that $\{v_2,\dots,v_k\}$ is also linearly independent.

Solution: Suppose $c_2v_2+c_3v_3+\cdots+c_kv_k=0$ for some scalars $c_2,\ldots,c_k$. Then it is also true that $0\,v_1+c_2v_2+c_3v_3+\cdots+c_kv_k=0$. But as $\{v_1,\ldots,v_k\}$ is linearly independent, this implies all the scalars you see must be zero: that is, $0=c_2=c_3=\cdots=c_k=0$. Thus $c_2=c_3=\cdots=c_k=0$. Hence $c_2v_2+c_3v_3+\cdots+c_kv_k=0$ implies $c_2=c_3=\cdots=c_k=0$, and so $\{v_2,\ldots,v_k\}$ is linearly independent.

(d) Give an example of a linearly independent subset $\{v_1,v_2\}$ in $\mathbb{R}^3$, and a vector $v\in\mathbb{R}^3$ such that $\{v,v_1,v_2\}$ is linearly *dependent*.

Solution: Set $v_1=(1,0,0), v_2=(0,1,0)$ and $v=(1,1,0)$. Then, $\{(1,0,0),(0,1,0)\}$ is independent, but $\{(1,1,0),(1,0,0),(0,1,0)\}$ isn't, since $(1,1,0)-(1,0,0)-(0,1,0)=(0,0,0)$, or $(1,1,0)=(1,0,0)+(0,1,0)$.

(f) Give an example of a linearly independent subset $\{p,q\}$ in $\mathscr{P}_2$, and a polynomial $r\in\mathscr{P}_2$ such that $\{p,q,r\}$ is linearly *dependent*.

Solution: Let $p(x)=1, q(x)=x$ and $r(x)=1+x$. Then $\{1,x\}$ is independent, but $\{1,x,1+x\}$ is not: $1+x-(1+x)=0$, for all $x\in R$.

Problems of Chapter 8

8.1 Justify your answers to the following: (the setting is a general vector space V).

(b) Suppose $u\in\operatorname{span}\{v,w\}$. Show carefully that $\operatorname{span}\{v,w\}=\operatorname{span}\{u,v,w\}$.

Solution: We show, under the assumption $u\in\operatorname{span}\{v,w\}$, that $\operatorname{span}\{v,w\}\subseteq\operatorname{span}\{u,v,w\}$ and $\operatorname{span}\{u,v,w\}\subseteq\operatorname{span}\{v,w\}$.

Now it is *always* true that $\operatorname{span}\{v,w\}\subseteq\operatorname{span}\{u,v,w\}$, since every linear combination of v and w (say) $av+bw$ ($a,b\in\mathbb{R}$) is also a linear combination of u, v and w, since $av+bw=0\,u+av+bw$. (We did not need the assumption $u\in\operatorname{span}\{v,w\}$ to get this far.)

To show that $\operatorname{span}\{u,v,w\}\subseteq\operatorname{span}\{v,w\}$, we need to show that every linear combination of u, v and w is also a linear combination of v and w. Here, we will need to know that $u\in\operatorname{span}\{v,w\}$: Suppose we take any linear combination of u, v and w, say $au+bv+cw$ (with $a,b,c\in R$. Now, since $u\in\operatorname{span}\{v,w\}$, we know $u=dv+ew$ for some scalars e,d. Now we can write

$$au+bv+cw=a(dv+ew)+bv+cw=(b+ad)v+(c+ae)w,$$

which shows that every linear combination of u, v and w is also a linear combination of v and w. Hence $\operatorname{span}\{u,v,w\}\subseteq\operatorname{span}\{v,w\}$, and we're done.

(Note that this latter fact is not true unless $u\in\operatorname{span}\{v,w\}$ – for example, let $u=(1,0,0)$, $v=(0,1,0)$ and $w=(0,0,1)$. Then $\mathbb{R}^3=\operatorname{span}\{u,v,w\}\not\subseteq\operatorname{span}\{v,w\}$, the latter being only the y–z plane.)

(d) Suppose $\operatorname{span}\{v,w\}=\operatorname{span}\{u,v,w\}$. Show carefully that $\{u,v,w\}$ is linearly dependent.

Solution: Use the previous part to conclude that $u\in\operatorname{span}\{v,w\}$. Then, $u=av+bw$ for some scalar a,b, and so $u-av-bw=0$, showing that $\{u,v,w\}$ is linearly dependent: no matter what a and b are, the coefficient of u is $1\neq 0$.

(f) Suppose $\{v,w\}$ is linearly independent, and that $u\notin\operatorname{span}\{v,w\}$. Show carefully that $\{u,v,w\}$ is linearly independent.

Solution: Suppose $au+bv+cw=0$ for some scalars a,b,c. If $a\neq 0$, then we could write $u=\frac{b}{a}v+\frac{b}{a}w$, which would mean that $u\in\operatorname{span}\{v,w\}$, contrary to assumption. So a must be zero.

Hence, $au+bv+cw=0$ becomes the equation $bv+cw=0$. But $\{v,w\}$ is linearly independent, so this means that $b=c=0$.

So $au+bv+cw=0$ has now implied (under the assumptions) that $a=b=c=0$. So, $\{u,v,w\}$ is linearly independent.

(h) Suppose $\text{span}\{v,w\} \neq \text{span}\{u,v,w\}$. Show carefully that $u \notin \text{span}\{v,w\}$.

Solution: Here, we will use part (b) above: Suppose to the contrary that $u \in \text{span}\{v,w\}$. By part (b) above, that would mean $\text{span}\{v,w\} = \text{span}\{u,v,w\}$, which contradicts the assumption $\text{span}\{v,w\} \neq \text{span}\{u,v,w\}$. So $u \in \text{span}\{v,w\}$ *cannot* be true: i.e., $u \notin \text{span}\{v,w\}$.

8.2 Justify your answers to the following:

(b) Suppose $p \neq 0$ and two polynomials $\{p,q\}$ satisfy $\deg(p) < \deg(q)$ Show carefully that $\{p,q\}$ is linearly independent. Note that

Solution: Suppose $n = \deg(p) < \deg(q) = m$, and write $p(x) = a_0 + a_1x + \cdots + a_nx^n$ and $q(x) = b_0 + b_1x + \cdots + b_mx^m$. Since $n = \deg(p)$ and $\deg(q) = m$, we know $a_n \neq 0$ and $b_m \neq 0$.

Now suppose $ap+bq=0$ is the zero polynomial. This implies that

$$a(a_0 + a_1x + \cdots + a_nx^n) + b(b_0 + b_1x + \cdots + b_mx^m) = 0 \quad \text{for every } x \in \mathbb{R}$$

Since a non-zero polynomial can have only finitely many roots, the equation above implies that we're looking at the zero polynomial. In particular the coefficient of x^m must be zero. But as $n < m$, this coefficient is exactly $b\,b_m$, so $b\,b_m = 0$. Since $b_m \neq 0$, we conclude that $b = 0$.

So now we have the equation

$$a(a_0 + a_1x + \cdots + a_nx^n) = 0 \quad \text{for every } x \in \mathbb{R}$$

As before, a non-zero polynomial can have only finitely many roots, so the equation just above implies that we're looking at the zero polynomial. So the coefficient of x^n must be zero. That is, $aa_n = 0$. But $a_n \neq 0$, so $a = 0$.

So we've shown that $ap+bq=0$ implies $a=b=0$. So $\{p,q\}$ is linearly independent.

(f) * Suppose $\{u,v,w\}$ are three vectors in $\mathbb{R}^3$ such that $u \cdot v \times w \neq 0$. Prove carefully that $\{u,v,w\}$ is linearly independent.[2]

Solution: We know that $\{u,v,w\}$ is linearly independent iff none of the vectors is a linear combination of the others.

So suppose to the contrary that $u \in \text{span}\{v,w\}$. So $u = av+bw$ for some scalars $a,b \in \mathbb{R}$. Then, $u \cdot v \times w = (av+bw) \cdot v \times w \neq 0$.

Recall from high school that for any three vectors $v_1, v_2, v_3 \in \mathbb{R}^3$, we have

$$v_1 \cdot v_2 \times v_3 = v_3 \cdot v_1 \times v_2 = v_2 \cdot v_3 \times v_1.$$

So $u \cdot v \times w = (av+bw) \cdot v \times w = w \cdot (av+bw) \times v = w \cdot bw \times v = v \cdot w \times bw = 0$. But this contradicts $u \cdot v \times w \neq 0$. So, indeed, $u \notin \text{span}\{v,w\}$.

Similar arguments show that $v \notin \text{span}\{u,w\}$ and $w \notin \text{span}\{u,v\}$. So we conclude that $\{u,v,w\}$ is indeed linearly independent under the assumption that $u \cdot v \times w \neq 0$.

[2] Hint: *A geometric argument involving 'volume' is not sufficient.* Instead, first recall that $\{u,v,w\}$ is linearly independent iff none of the vectors is a linear combination of the others. Now proceed by contradiction, and by recalling from high school that for any three vectors $v_1, v_2, v_3 \in \mathbb{R}^3$ that $v_1 \cdot v_2 \times v_3 = v_3 \cdot v_1 \times v_2 = v_2 \cdot v_3 \times v_1$.

Problems of Chapter 9

9.1 Give two distinct bases for each of the following subspaces, and hence give the dimension of each subspace.

(b) $L = \{(x,y) \in \mathbb{R}^2 \mid 3x - y = 0\}$

Solution: We saw in Q.7 (b) of the exercises "VSubspacesSpan" that $L = \text{span}\{(1,3)\} = \text{span}\{(2,6)\}$. Since $\{(1,3)\}$ and $\{(2,6)\}$ are linearly independent ($\{v\}$ is independent iff $v \neq 0$), both are bases for L. Hence $\dim L = 1$.

(d) $K = \{(x,y,z,w) \in \mathbb{R}^4 \mid x - y + z - w = 0\}$

Solution: We saw in Q. 6.3(d) that $\{(1,1,0,0),(-1,0,1,0),(1,0,0,1)\}$ and $\{(2,2,0,0),(-3,0,3,0),(4,0,0,4)\}$ are spanning sets for K.

We claim both are also linearly independent: Suppose $a(1,1,0,0)+b(-1,0,1,0)+c(1,0,0,1) = (0,0,0,0)$. Equating the second component of each side yields $a = 0$; equating the third component of each side yields $b = 0$, and equating the fourth component of each side yields $c = 0$. Hence $a = b = c = 0$, and so $\{(1,1,0,0),(-1,0,1,0),(1,0,0,1)\}$ is independent.

An identical argument shows that $\{(2,2,0,0),(-3,0,3,0),(4,0,0,4)\}$ is independent. Hence both $\{(1,1,0,0),(-1,0,1,0),(1,0,0,1)\}$ and $\{(1,1,0,0),(-1,0,1,0),(1,0,0,1)\}$ are bases for K. Hence $\dim K = 3$.

(f) $S = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; a+d = 0 \right\}$.

Solution: Note first that

$$P = \left\{ \begin{bmatrix} a & b \\ c & -a \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; a,b,c \in \mathbb{R} \right\} = \text{span}\left\{ \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}.$$

So $\mathscr{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}$ spans S.

Now suppose $a \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. Then $\begin{bmatrix} a & b \\ c & -a \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, this implies $a = b = c = 0$. Hence $\mathscr{B}$ is also linearly independent, and hence is a basis of S. Another basis is given by $\left\{ \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}$. Hence $\dim S = 3$.

(h) $X = \left\{ \begin{bmatrix} 0 & -b \\ -b & 0 \end{bmatrix} \in \mathbf{M}_{22}(\mathbb{R}) \;\middle|\; b \in \mathbb{R} \right\}$.

Solution: We saw before that $X = \text{span}\left\{ \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\} = textspan\left\{ \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix} \right\}$. Since each of these spanning sets contains a single non-zero matrix, each is also linearly independent. Hence both $\left\{ \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\}$ and $\left\{ \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix} \right\}$ are bases for S, and so either shows that $\dim S = 1$.

(j) $\mathscr{P}_n$.

Solution: We saw in previous exercsies that $\mathscr{P}_n = \text{span}\{1, x, \dots, x^n\}$, so $\mathscr{B} = \{1, x, \dots, x^n\}$ spans $\mathscr{P}_n$. Suppose $a_0 + a_1 x + \cdots a_n x^n = 0$ for all $x \in \mathbb{R}$. Since a non-zero polynomial can only have finitely many roots, this shows that $a_0 = a_1 = \cdots = a_n = 0$, and so $\mathscr{B}$ is independent and hence is a basis for $\mathscr{P}_n$. It is clear that $\{2, x, \dots, x^n\}$ is another basis for $\mathscr{P}_n$. Either basis shows that $\dim \mathscr{P}_n = n+1$.

(l) $Y = \{p \in \mathscr{P}_3 \mid p(2) = p(3) = 0\}$.

Solution: We saw in previous solutions that

$$\mathscr{B} = \{(x-2)(x-3), x(x-2)(x-3)\}$$

spans Y. We can now appeal to solutions to one of the exercises on linear independence to conclude that $\mathscr{B}$ is linearly independent. Or, we could show it directly: Suppose, for some scalars a, b, that $a(x-2)(x-3) + bx(x-2)(x-3) = 0$ for all $x \in \mathbb{R}$. Setting $x = 0$ yields $a = 0$, and then setting $x = 1$ yields $b = 0$. So $a = b = 0$, and $\mathscr{B}$ is linearly independent and is hence a basis for Y. So $\dim Y = 2$. A different basis is clearly given by $\{2(x-2)(x-3), x(x-2)(x-3)\}$.

(n) $W = \text{span}\{\sin x, \cos x\}$.

Solution: We are given the spanning set $\mathscr{B} = \{\sin x, \cos x\}$ for W. We saw in class that $\mathscr{B}$ is independent, and hence is a basis of W. Thus $\dim W = 2$.

(p) $X = \text{span}\{1, \sin^2 x, \cos^2 x\}$.

Solution: Since $1 = \sin^2 x + \cos^2 x$ for all $x \in \mathbb{R}$, $1 \in sp \sin^2 x, \cos^2 x$, and so $X = \text{span}\{\sin^2 x, \cos^2 x\}$. We claim $\mathscr{B} = \{\sin^2 x, \cos^2 x\}$ is independent: Suppose there are scalars a, b such that $a \sin^2 x + b \cos^2 x = 0$ for all $x \in \mathbb{R}$. Setting $x = 0$ yields $b = 0$, and setting $x = \frac{\pi}{2}$ implies $a = 0$. Hence $a = b = 0$, and so $\mathscr{B}$ is independent, and is thus a basis of X. So $\dim X = 2$. Another basis is clearly given by $\{2\sin^2 x, \cos^2 x\}$.

9.2 Determine whether the following sets are bases of the indicated vector spaces.

(b) $\{(1,2), (-2,-4)\}$; $(\mathbb{R}^2)$. **Solution:** This is not a basis of $\mathbb{R}^2$ since $\{(1,2), (-2,-4)\}$ is dependent — the second vector is a multiple of the first.

(d) $\{(1,2), (3,4), (0,0)\}$; $(\mathbb{R}^2)$.

Solution: This set is dependent because it contains the zero vector, and so it is not a basis of $\mathbb{R}^2$.

(f) $\{(1,2,3), (4,8,7)\}$; $(\mathbb{R}^3)$.

Solution: We know that $\dim R^3 = 3$ and so every basis must contain 3 vectors. So this is not a basis of $\mathbb{R}^3$.

(h) $\{(1,0,1,0), (0,1,0,1)\}$; $(\mathbb{R}^4)$.

Solution: We know that $\dim R^4 = 4$ and so every basis must contain 4 vectors. So this is not a basis of $\mathbb{R}^4$.

(j) $\left\{ \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -2 \\ 1 & 0 \end{bmatrix} \right\}$; $(\mathbf{M}_{22})$

Solution: We know that $\dim \mathbf{M}_{22} = 4$ and so every basis must contain 4 vectors. So this is not a basis of $\mathbf{M}_{22}$.

(m) $\{1, 1+x, x^2\}$; $(\mathscr{P}_2)$.

Solution: Since we know that $\dim \mathscr{P}_2 = 3$, and we do have 3 polynomials here, it suffices to check just one of the conditions for a basis. Let's check that $\{1, 1+x, x^2\}$ spans $\mathscr{P}_2$: Simply note that for any $a, b, c, \in \mathbb{R}$, $a + bx + cx^2 = (a-b)\,1 + b(1+x) + cx^2$. This shows that $\mathscr{P}_2 = \operatorname{span}\{1, 1+x, x^2\}$, and by previous comments, that $\{1, 1+x, x^2\}$ is a basis for $\mathscr{P}_2$.

(o) $\{1, \sin x, 2\cos x\}$; $(\mathbf{F}(\mathbb{R}))$.

Solution: We saw in class (and previous solutions) that $\mathbf{F}(\mathbb{R})$ does not have a finite spanning set, so $\{1, \sin x, 2\cos x\}$ cannot be a basis for $\mathbf{F}(\mathbb{R})$.

But let's do this directly. If we can find a single function in $\mathbf{F}(\mathbb{R})$ that is not in $\operatorname{span}\{1, \sin x, 2\cos x\}$, we're done.

We claim that the function $\sin x \cos x \notin \operatorname{span}\{1, \sin x, 2\cos x\}$. Suppose the contrary, i.e., that there are scalars a, b, c such that

$$\sin x \cos x = a\,1 + b \sin x + 2c\,\cos x, \qquad \text{for all } x \in \mathbb{R}$$

Setting $x = 0$ yields the equation $0 = a + 2c$; setting $x = \frac{\pi}{2}$ gives $0 = a + b$ and setting $x = \pi$ yields $0 = a - 2c$. The first and third of these equations together imply $a = c = 0$, and then the second implies $b = 0$. But then we have the identity $\sin x \cos x = 0$ for all $x \in \mathbb{R}$, which is nonsense because the left hand side is $\frac{1}{2} \neq 0$ when $x = \frac{\pi}{4}$.

Problems of Chapter 10

10.1 Find the coordinates of the following vectors v with respect to the given bases $\mathscr{B}$ of the indicated vector space V:

(b) $v = (1,0,1)$; $\quad \mathscr{B} = \{(1,1,0), (1,-1,0), (0,0,1)\}$; $\quad V = \mathbb{R}^3$

Solution: Solving $(1,0,1) = a(1,1,0) + b(1,-1,0) + c(0,0,1)$ yields $a = b = \frac{1}{2}, c = 1$, so the coordinate vector of v with respect to $\mathscr{B}$ is $(\frac{1}{2}, \frac{1}{2}, 1)$.

(d) $v = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}$; $\quad \mathscr{B} = \left\{ \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right\}$; $\quad V = \{A \in \mathbf{M}_{22} \mid \operatorname{tr} A = 0\}$.

Solution: Clearly, $\begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} = 1 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + 0 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + 1 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, so the coordinate vector of v with respect to $\mathscr{B}$ is $(1,0,1)$.

(f) $v = \sin(x+2)$; $\quad \mathscr{B} = \{\sin x, \cos x\}$; $\quad V = \operatorname{span} \sin x, \cos x$

Solution: Since $\sin(x+2) = \sin x \cos 2 + \sin 2 \cos x$ for all $x \in \mathbb{R}$, the coordinate vector of v with respect to $\mathscr{B}$ is $(\cos 2, \sin 2)$.

Problems of Chapter 11

11.1 Find the augmented matrix of the following linear systems.

(b)

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

Solution: $\begin{bmatrix} 1 & 0 & 0 & 1 & | & 1 \\ 1 & 0 & 1 & 1 & | & 0 \\ 1 & 1 & 1 & 0 & | & -3 \\ 1 & 1 & 0 & -2 & | & 2 \end{bmatrix}$

11.2

(b) Find all (x,y) so that the matrix $\begin{bmatrix} 1 & 0 & 1 \\ x & y & 0 \end{bmatrix}$ is in reduced row-echelon form.

Solution: Clearly, x must be zero, and y could be zero or 1, so the possible pairs (x,y) are $(0,0)$ and $(0,1)$.

Problems of Chapter 12

12.1 Find the reduced row-echelon form of the following matrices:

(b) $\begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 2 \end{bmatrix}$

Solution: This matrix is already in reduced row-echelon form!

(d) $\begin{bmatrix} 1 & 2 & -1 & -1 \\ 2 & 4 & -1 & 3 \\ -3 & -6 & 1 & -7 \end{bmatrix}$

Solution:

$$\begin{bmatrix} 1 & 2 & -1 & -1 \\ 2 & 4 & -1 & 3 \\ -3 & -6 & 1 & -7 \end{bmatrix} \begin{matrix} -2R_1+R_2 \to R_2 \\ \sim \\ 3R_1+R_3 \to R_3 \end{matrix} \begin{bmatrix} 1 & 2 & -1 & -1 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & -2 & -10 \end{bmatrix} \begin{matrix} 2R_2+R_3 \to R_3 \\ \sim \\ R_2+R_1 \to R_1 \end{matrix} \begin{bmatrix} 1 & 2 & 0 & 4 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

(f) $\begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 4 & 3 \\ 1 & 0 & 2 & 1 \end{bmatrix}$

Solution:

$$\begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 4 & 3 \\ 1 & 0 & 2 & 1 \end{bmatrix} \begin{matrix} -R_1+R_3 \to R_3 \\ \sim \\ -R_1+R_4 \to R_4 \end{matrix} \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 3 & 3 & 3 \\ 0 & 1 & 1 & 1 \end{bmatrix} \begin{matrix} -3R_2+R_3 \to R_3 \\ \sim \\ -R_2+R_4 \to R_4 \end{matrix} \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} R_2+R_1 \to R_1 \\ \sim \end{matrix} \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

12.2 Find the general solutions to the linear systems whose augmented matrices are given below.

(b) $\begin{bmatrix} 1 & 0 & -1 & | & 0 \\ 0 & 1 & 2 & | & 0 \\ 0 & 0 & 0 & | & 1 \\ 0 & 0 & 0 & | & 0 \end{bmatrix}$ **Solution:** This system is inconsistent.

(d) $\begin{bmatrix} 1 & 2 & 0 & 3 & 0 & | & 7 \\ 0 & 0 & 1 & 0 & 0 & | & 1 \\ 0 & 0 & 0 & 0 & 1 & | & 2 \end{bmatrix}$

Solution: $(x,y,z,u,v) = (7-2s-3t,s,1,t,2)$; $s,t \in \mathbb{R}$, or we could write it as a set:

$$\{(7-2s-3t,s,1,t,2); \mid s,t \in \mathbb{R}\}$$

Problems of Chapters 13

13.1

(b) Find the ranks of the coefficient matrices and the ranks of the augmented matrices corresponding to the linear systems in question 12.2.

Solution:

[12.2(b)] The rank of the coefficient matrix is 2 and the rank of the augmented matrix is 3. (That's why the corresponding linear system is inconsistent.)

[12.2(d)] Both ranks are 3.

13.2 Suppose $a,c \in \mathbb{R}$ and consider the following linear system in the variables x,y and z:

$$\begin{array}{rcrcrcl} x & + & y & + & az & = & 2 \\ 2x & + & y & + & 2az & = & 3. \\ 3x & + & y & + & 3az & = & c \end{array}$$

Note that the general solution of this system may depend on the values of a and c. Let $[A\,|\,b]$ denote the augmented matrix of the system above.

(b) Find <u>all</u> values of a and c so that this system has

(i) a unique solution,

(ii) infinitely many solutions, or

(iii) no solutions.

Solution:

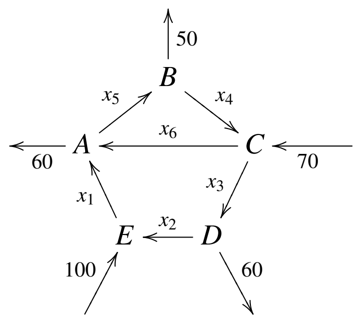
$$[A\,|\,b] = \begin{bmatrix} 1 & 1 & a & | & 2 \\ 2 & 1 & 2a & | & 3 \\ 3 & 1 & 3a & | & c \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & a & | & 2 \\ 0 & 1 & 0 & | & 1 \\ 0 & 0 & 0 & | & c-4 \end{bmatrix}$$

Since $\operatorname{rank} A = 2 < 3 = \#$ variables, for *all* values of a,c, this system *never* has a unique solution.

Since $\operatorname{rank} A = 2 = \operatorname{rank}[A\,|\,b]$ iff $c = 4$, this system will be consistent and have infinitely many solutions, since then there will be $(\#\text{variables} - \operatorname{rank} A = 3-2 =)1$ parameter in the general solution.

Since $\operatorname{rank} A = 2 < 3 = \operatorname{rank}[A\,|\,b]$ iff $c \neq 4$, this system will be inconsistent for all values of c except for $c = 4$.

13.4 Consider the network of streets with intersections A, B, C, D and E below. The arrows indicate the direction of traffic flow along the **one-way streets**, and the numbers refer to the **exact** number of cars observed to enter or leave A, B, C, D and E during one minute. Each x_i denotes the unknown number of cars which passed along the indicated streets during the same period.



(b) The reduced row-echelon form of the augmented matrix of the system in part (a) is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 1 & | & 60 \\ 0 & 1 & 0 & 0 & -1 & 1 & | & -40 \\ 0 & 0 & 1 & 0 & -1 & 1 & | & 20 \\ 0 & 0 & 0 & 1 & -1 & 0 & | & -50 \\ 0 & 0 & 0 & 0 & 0 & 0 & | & 0 \end{bmatrix}$$

Give the general solution. (Ignore the constraints from (a) at this point.)

Solution: The general solution is

$$\begin{aligned} x_1 &= 60+s-t \\ x_2 &= -40+s-t \\ x_3 &= 20+s-t \qquad\qquad s,t \in \mathbb{R} \\ x_4 &= -50+s \\ x_5 &= s \\ x_6 &= t \end{aligned}$$

Or it can be written as

$$\{(60+s-t, -40+s-t, 20+s-t, -50+s, s, t) \mid st, t \in \mathbb{R}\}.$$

13.5 State whether each of the following is (always) true, or is (possibly) false.

- If you say the statement may be false, you must give an explicit example.
- If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

(b) Every non-homogenous system of 3 equations in 2 unknowns is consistent.

Solution: This is false. For example, the system whose augmented matrix is $\begin{bmatrix} 1 & 0 & | & 1 \\ 0 & 1 & | & 0 \\ 0 & 0 & | & 1 \end{bmatrix}$ is inconsistent.

(d) Every system of 2 equations in 2 unknowns has a unique solution.

Solution: This is false. For example, the system whose augmented matrix is $\begin{bmatrix} 1 & 0 & | & 0 \\ 0 & 0 & | & 0 \end{bmatrix}$ has infinitely many solutions, namely $(0,s)$ for every $s \in \mathbb{R}$

(f) If there is a column of zeroes in the coefficient matrix of a consistent linear system, the system will have infinitely many solutions.

Solution: This is true, since once the augmented matrix has been reduced, there will still be a column of zeroes in the coefficient matrix, which means that (as it is assumed consistent), there will be at least one parameter in the general solution.

(h) If a consistent linear system has infinitely many solutions, there must be a column of zeroes in the reduced coefficient matrix (i.e. when the augmented matrix has been reduced so that the coefficient matrix is in RRE form.)

Solution: This is false. Look at Q.4 (d).

(j) If a homogenous linear system has a unique solution, it must the same number of equations as unknowns.

Solution: This is false. For example, the system whose augmented matrix is $\begin{bmatrix} 1 & 0 & | & 0 \\ 0 & 1 & | & 0 \\ 0 & 0 & | & 0 \end{bmatrix}$ has a unique solution, namely $(0,0)$.

Problems of Chapter 14

Matrix Multiplication

14.1

(b) Write the matrix product $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$ as a linear combination of the columns of $A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$.

Solution: $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a \begin{bmatrix} 1 \\ 3 \end{bmatrix} + b \begin{bmatrix} 2 \\ 6 \end{bmatrix}$.

(d) Find the matrix product $\begin{bmatrix} a \\ b \end{bmatrix} \begin{bmatrix} c & d \end{bmatrix}$.

Solution: $\begin{bmatrix} a \\ b \end{bmatrix} \begin{bmatrix} c & d \end{bmatrix} = \begin{bmatrix} ac & ad \\ bc & bd \end{bmatrix}$

(f) If $A = \begin{bmatrix} c_1 & c_2 & c_3 \end{bmatrix}$ is an $m \times 3$ matrix written in block column form, and $x = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}$ is a column vector in $\mathbb{R}^3$, express Ax as a linear combination of c_1, c_2 and c_3.

Solution: $Ax = 2c_1 + c_2 + 4c_3$.

(h) Show that if $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, then $A^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Solution: This is a straighforward computation.

(j) If C is a $m \times 4$ matrix and $D = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, express the columns of CD in terms of the columns of C.

Solution: Write $C = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 \end{bmatrix}$ in block column form. Then

$$CD = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} c_1 + c_3 & c_2 & c_1 + c_2 \end{bmatrix}$$

(l) Find all (a, b, c) so that $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} a & b \\ c & a \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Solution: Since $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} a & b \\ c & a \end{bmatrix} = \begin{bmatrix} a+2c & b+2a \\ 3a+6c & 3b+6a \end{bmatrix}$, this is the zero matrix for (a,b,c) which are solutions to

$$\begin{array}{ccccccc} a & & & + & 2c & = & 0 \\ 2a & + & b & & & = & 0 \\ 3a & & & + & 6c & = & 0 \\ 6a & + & 3b & & & = & 0 \end{array}.$$

The general solution to this system is $(a,b,c) = (-2s, 4s, s)$; $s \in \mathbb{R}$.

14.3 State whether each of the following is (always) true, or is (possibly) false. In this question, A and B are matrices for which the indicated products exist.

• If you say the statement may be false, you must give an explicit example. • If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

(b) $C(A+B) = CA + CB$

Solution: This is true: it a property of matrix multiplication we saw in class. See me if you're interested in the proof.

(d) $AB = BA$

Solution: This is not always true. For example, here's an example we saw in class:

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

(f) If $A^2 = 0$ for a square matrix A, then $A = 0$.

Solution: This is false. Look at the solution to14.1.(h).

Applications to Linear Systems

14.5 Write the matrix equation which is equivalent to each of following linear systems.

(b)

$$\begin{array}{ccccccccc} x & & & & & + & w & = & 1 \\ x & & & + & z & + & w & = & 0 \\ x & + & y & + & z & & & = & -3 \\ x & + & y & & & - & 2w & = & 2 \end{array}$$

Solution: $\begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -3 \\ 4 \end{bmatrix}$

14.6 Write the matrix equation of the linear system corresponding to each of the augmented matrices given below.

(b) $\begin{bmatrix} 1 & 0 & -1 & | & 0 \\ 0 & 1 & 2 & | & 0 \\ 0 & 0 & 0 & | & 1 \\ 0 & 0 & 0 & | & 0 \end{bmatrix}$

Solution: $\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 2 & 0 & 3 & 0 & | & 7 \\ 0 & 0 & 1 & 0 & 0 & | & 1 \\ 0 & 0 & 0 & 0 & 1 & | & 2 \end{bmatrix}$

Solution: $\begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 7 \\ 1 \\ 2 \end{bmatrix}$

14.7 State whether each of the following is (always) true, or is (possibly) false.

- If you say the statement may be false, you must give an explicit example.
- If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

(b) If $[A\,|\,b]$ is the augmented matrix of a linear system, then $\operatorname{rank} A < \operatorname{rank}[A\,|\,b]$ is possible.

Solution: This is true. For example, $\begin{bmatrix} 0 & 1 & | & 0 \\ 0 & 0 & | & 1 \end{bmatrix}$.

(d) If $[A\,|\,b]$ is the augmented matrix of a linear system, and $\operatorname{rank} A = \operatorname{rank}[A\,|\,b]$, then the system is consistent.

Solution: This is true, as we learned in class. Indeed $Ax = b$ is consistent iff $\operatorname{rank} A = \operatorname{rank}[A\,|\,b]$.

(f) If A is an $m \times n$ matrix and $Ax = 0$ has a unique solution for $x \in \mathbb{R}^n$, then the columns of A are linearly independent.

Solution: This is true, as we learned in class: write $A = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix}$ in block column form and $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$. Then $Ax = x_1c_1 + x_2c_2 + \cdots + x_nc_n$, so $Ax = 0$ has the unique solution $x = 0$ means that $x_1c_1 + x_2c_2 + \cdots + x_nc_n = 0$ implies $x_1 = x_2 = \cdots = x_n = 0$, so $\{c_1, c_2, \ldots, c_n\}$ is linearly independent.

(h) If A is an $m \times n$ matrix and $Ax = 0$ has infinitely many solutions for $x \in \mathbb{R}^n$, then the columns of A are linearly dependent.

Solution: This is true: see the solution to part (f). If $A = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix}$ in block column form and $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, then if $Ax = x_1c_1 + x_2c_2 + \cdots + x_nc_n = 0$ has more than one solution, the columns of A are indeed dependent.

(j) If A is a 6×5 matrix and $\operatorname{rank} A = 5$, then $Ax = 0$ implies $x = 0$ for $x \in \mathbb{R}^5$.

Solution: This is true, since there will be a leading one in every column of the RRE form of A, implying that there are no parameters in the general solution to $Ax = 0$, so $x = 0$ is the only solution. (Or: $\dim \ker A = \#$ columns of A $- \operatorname{rank} A = 5 - 5 = 0$, so $\ker A = \{0\}$, i.e., $Ax = 0$ implies $x = 0$.

(l) If A is a 5×6 matrix and $\operatorname{rank} A = 5$, then $Ax = b$ is consistent for every $b \in \mathbb{R}^5$.

Solution: This is true, since for every $b \in \mathbb{R}^5$, $\operatorname{rank}[A \,|\, b] = 5$: the rank of the 5×7 matrix $[A \,|\, b]$ cannot be smaller than $\operatorname{rank} A$ (5), and cannot be larger than the minimum of the number of rows (5!) and the number of columns (7). And, of course we know that $Ax = b$ is consistent iff $\operatorname{rank} A = \operatorname{rank}[A \,|\, b]$.

(n) If A is a 3×2 matrix and $\operatorname{rank} A = 1$, then $Ax = 0$ implies $x = 0$ for $x \in \mathbb{R}^2$.

Solution: This is false. For example $\operatorname{rank} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = 1$, but $Ax = 0$ has infinitely many solutions with 1 parameter.

(p) The rows of a 19×24 matrix are always linearly dependent.

Solution: This is false. For example, let $A = \begin{bmatrix} I_{19} & 0 \end{bmatrix}$, where the 0 that appears is a 19×25 zero matrix. The rows of this matrix are indeed independent, as they are the first 19 vectors in the standard basis of $\mathbb{R}^{25}$.

Problems of Chapter 15

15.1 Find a basis for the kernel (or 'nullspace') of the following matrices:

(b) $A = \begin{bmatrix} 1 & 2 & -1 & 3 \end{bmatrix}$

Solution: The general solution to $Ax = 0$ is $\{(-2s + t - 3r, s, t, r) \mid s, t, r \in \mathbb{R}\}$, so a basis for the kernel of A is $\{(-2, 1, 0, 0), (1, 0, 1, 0), (-3, 0, 0, 1)\}$.

(d) $\begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

Solution: The general solution to $Ax = 0$ is $\{(s, -s-t, -s-t, s, t \mid s, t) \in \mathbb{R}\}$, so a basis for the kernel of A is $\{(1, -1, -1, 1, 0), (0, -1, -1, 0, 1)\}$.

(f) $B = \begin{bmatrix} -4 & 2 & 2 \\ 2 & -4 & 2 \\ 2 & 2 & -4 \end{bmatrix}$

Solution: The RRE form of B is $\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$, so the general solution to $Bx = 0$ is $\{(s, s, s) \mid s, t \in \mathbb{R}\}$. Hence a basis for the kernel of B is $\{(1, 1, 1)\}$.

Problems of Chapter 16

16.1 Find the a basis for the row space and column space for each of the matrices below, and check that $\dim \operatorname{Row} A = \dim \operatorname{Col} A$ in each case.

(b) $A = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 2 \end{bmatrix}$

Solution: Row-reduction to RRE form does the trick here. Note that since A is already in RRE form, the Row Space Algorithm yields $\{(1, 0, 1, 2), (0, 1, 1, 2)\}$ as a basis for Row A and $\{(1, 0), (0, 1)\}$ as a basis for Col A.

(d) $A = \begin{bmatrix} 1 & 2 & -1 & -1 \\ 2 & 4 & -1 & 3 \\ -3 & -6 & 1 & -7 \end{bmatrix}$

Solution: Here we must first reduce A to RRE form, and find $A \sim \begin{bmatrix} 1 & 2 & 0 & 4 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Now the the Row Space Algorithm yields $\{(1, 2, 0, 4), (0, 0, 1, 5)\}$ as a basis for Row A and $\{(1, 2, -3), (-1, -1, 1)\}$ as a basis for Col A.

(f) $A = \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 4 & 3 \\ 1 & 0 & 2 & 1 \end{bmatrix}$

Solution: Here we must first reduce A to RREF, and find $A \sim \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Now the the Row Space Algorithm yields $\{(1, 0, 2, 1), (0, 1, 1, 1)\}$ as a basis for Row A and $\{(1, 0, 1, 1), (-1, 1, 2, 0)\}$ as a basis for Col A.

16.2 Find a basis of the desired type for the given subspace in each case. (Use your work from the previous question where useful.)

(b) $W = \operatorname{span}\{(1,2,-1,-1),(2,4,-1,3),(-3,-6,1,-7)\}$: any basis suffices.

Solution: We note that $W = \operatorname{Row} A$ for the matrix A of part (b) in the previous question. So we use the basis for $\operatorname{Row} A$ we found there, namely $\{(1,2,04),(0,0,1,5)\}$.

(d) $Y = \operatorname{span}\{(1,0,1,1),(-1,1,2,0),(1,1,4,2),(0,1,3,1)\}$: the basis must be a subset of the given spanning set.

Solution: We note that $Y = \operatorname{Col} A$ for the matrix A of part (f) in the previous question. So we use the basis for $\operatorname{Col} A$ we found there, namely $\{(1,0,1,1),(-1,1,2,0)\}$.

16.3

(b) We extend $\{(1,0,1,2),(0,1,1,2)\}$ to a basis of $\mathbb{R}^4$:

Set $A = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 2 \\ & & \mathbf{u}_3 & \\ & & \mathbf{u}_4 & \end{bmatrix}$. We can see two leading ones in columns 1 and 2, so if we set $\mathbf{u}_3 = (0,0,1,0)$ and $\mathbf{u}_4 = (0,0,0,1)$, then

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

It is now clear that $\operatorname{rank} A = 4$, so a desired extension is

$$\{(1,0,1,2),(0,1,1,2),(0,0,1,0),(0,0,0,1)\}$$

(d) We extend $\{(1,0,1,1),(-1,1,2,0)\}$ to a basis of $\mathbb{R}^4$:

Set $A = \begin{bmatrix} 1 & 0 & 1 & 1 \\ -1 & 1 & 2 & 0 \\ & & \mathbf{u}_3 & \\ & & \mathbf{u}_4 & \end{bmatrix}$. We row reduce as far as we can:

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ -1 & 1 & 2 & 0 \\ & & \mathbf{u}_3 & \\ & & \mathbf{u}_4 & \end{bmatrix} \sim \begin{bmatrix} \textcircled{1} & 0 & 1 & 1 \\ 0 & \textcircled{1} & 3 & 1 \\ & & \mathbf{u}_3 & \\ & & \mathbf{u}_4 & \end{bmatrix}$$

Again, we can see two leading ones in columns 1 and 2, so if we set $\mathbf{u}_3 = (0,0,1,0)$ and $\mathbf{u}_4 = (0,0,0,1)$, then

$$A = \begin{bmatrix} 1 & 0 & 1 & 1 & \\ 0 & -1 & 1 & 2 & 0 \\ 0 & 0 & 1 & 0 & \\ 0 & 0 & 0 & 1 & \end{bmatrix} \sim \begin{bmatrix} \textcircled{1} & 0 & 1 & 1 \\ 0 & \textcircled{1} & 3 & 1 \\ 0 & 0 & \textcircled{1} & 0 \\ 0 & 0 & 0 & \textcircled{1} \end{bmatrix}$$

It is now clear that $\operatorname{rank} A = 4$, so a desired extension is

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

16.4 State whether each of the following is (always) true, or is (possibly) false.

- If you say the statement may be false, you must give an explicit example.
- If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

(b) For some matrices A, $\dim \text{Row}\, A + \dim \ker A = \dim \text{Col}\, A$

Solution: Since we always have $\dim \text{Row}\, A = \dim \text{Col}\, A$, the equation above holds if and only if $\dim \ker A = 0$. There are such matrices, so the statement –which merely asserts that there are *some* matrices for which the equation holds – *is* true. For example, it is true for any invertible matrix, such as $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, where $\dim \text{Row}\, A = \dim \text{Row}\, A = 2$ and $\dim \ker A = 0$.

(d) For all matrices A, $\dim \text{Row}\, A + \dim \ker A = n$, where n is the number of columns of A.

Solution: This is always true, indeed we called it the 'conservation of dimension'. We know $\dim \ker A$ is the number of parameters in the general solution to $Ax = 0$, which we also know is the number of columns of A where there are not leading ones, which of course is the number of columns, less its rank. We also know $\text{rank} A = \dim \text{Row}\, A$. So $\dim \ker A = n - \dim \text{Row}\, A$, which is equivalent to the given equation.

(f) For all $m \times n$ matrices A, $\dim\{Ax \mid x \in \mathbb{R}^n\} + \dim\{x \in \mathbb{R}^n \mid Ax = 0\} = m$.

Solution: We actually know that $\dim\{Ax \mid x \in \mathbb{R}^n\} + \dim\{x \in \mathbb{R}^n \mid Ax = 0\} = n$ for all $m \times n$ matrices A, so the stated equation is true iff $n = m$. Thus the statement above – which asserts that this holds for all $m \times n$ matrices – is false.

For example, let $A = \begin{bmatrix} 1 & 0 \end{bmatrix}$, where $m = 1$ and $n = 2$. Then $\dim\{Ax \mid x \in \mathbb{R}^n\} = \dim \text{Col}\, A = \text{rank} A = 1$, and $\dim\{x \in \mathbb{R}^n \mid Ax = 0\} = \dim \ker A = 1$, so the equation asserted in the statement is $1 + 1 = 1$, which of course is false.

(h) For all $m \times n$ matrices A, the dot product of every vector in $\ker A$ with any of the rows of A is zero.

Solution: This is true, and is an easy consequence of block multiplication and of the definition of the kernel of A: Let $x \in \mathbb{R}^n$ belong to $\ker A$, so $Ax = 0$. Write A in block row form: $A = \begin{bmatrix} r_1 \\ \vdots \\ r_m \end{bmatrix}$, so that r_i is the i^{th} row of A. Then $Ax = \begin{bmatrix} r_1 \\ \vdots \\ r_m \end{bmatrix} x = \begin{bmatrix} r_1 \cdot x \\ \vdots \\ r_m \cdot x \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$. That is , $r_i \cdot x = 0$ for every i, $1 \leq i \leq m$.

Problems of Chapter 17

17.1 Extend the given linearly independent set of $\mathbb{R}^n$ to a basis of $\mathbb{R}^n$. (Use your work from question 1 where useful.)

(b) $\{(1,0,1),(0,1,1)\}$ $(\mathbb{R}^3)$

Solution: Since $\dim \mathbb{R}^3 = 3$ and we already have 2 linearly independent vectors, we only need to find $w \in \mathbb{R}^3$ such that rank $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ & w & \end{bmatrix} = 3$. Clearly, $w = (0,0,1)$ will do, since then $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ & w & \end{bmatrix}$ is in RRE form and its rank is clearly 3

(d) $\{(1,0,1,3)\}$ ($\mathbb{R}^4$)

Solution: Since $\dim \mathbb{R}^4 = 4$ and we only have 1 linearly independent vector, we need to find $u, v, w \in \mathbb{R}^4$ such that rank $\begin{bmatrix} 1 & 0 & 1 & 3 \\ & u & & \\ & v & & \\ & w & & \end{bmatrix} = 4$. Clearly, $u = (0,1,0,0), v = (0,0,1,0)$ and $w = (0,0,0,1)$ will suffice, since then $\begin{bmatrix} 1 & 0 & 1 & 3 \\ & u & & \\ & v & & \\ & w & & \end{bmatrix}$ is in RRE form and its rank is clearly 4.

Problems of Chapter 18

18.1 Find the inverse of each of the following matrices, or give reasons if it not invertible.

(b) $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & -2 \\ 0 & 2 & 3 \end{bmatrix}$ **Solution:** $A^{-1} = \begin{bmatrix} 1 & -3 & -2 \\ 0 & 3 & 2 \\ 0 & -2 & -1 \end{bmatrix}$

(d) $B = \begin{bmatrix} 1 & x \\ -x & 1 \end{bmatrix}$ **Solution:** $B^{-1} = \begin{bmatrix} \frac{1}{x^2+1} & -\frac{x}{x^2+1} \\ \frac{x}{x^2+1} & \frac{1}{x^2+1} \end{bmatrix}$

18.2 State whether each of the following is (always) true, or is (possibly) false. The matrices are assumed to be square.

- If you say the statement may be false, you must give an explicit example.
- If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

(b) If $A^2 = 0$ for a square matrix A, then A is not invertible.

Solution: This is true. Suppose on the contrary that A were invertible with inverse A^{-1}. Multiplying both sides of the equation $A^2 = 0$ by A^{-2}, say on the left, yields $I_n = 0$, which is nonsense. So A is not invertible..

(d) If A is invertible then the RRE form of A has a row of zeros.

Solution: This is *always* false, but here's an example: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is invertible and is in RRE form, but has no row of zeros. (We know that an $n \times n$ matrix A is invertible iff its RRE form is I_n, which has no row of zeros.)

(f) If A is a non-invertible $n \times n$ matrix then $Ax = b$ is inconsistent for every $b \in \mathbb{R}^n$.

Solution: This is false. For example, let $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Then A is not invertible, but $Ax = b$ is consistent, and indeed has infintely many solutions of the form $x = \begin{bmatrix} 1 \\ s \end{bmatrix}$, for any $s \in \mathbb{R}$.

(h) If an $n \times n$ matrix A satisfies $A^3 - 3A^2 + I_n = 0$, then A is invertible and $A^{-1} = 3A - A^2$.

Solution: This is true: rewrite $A^3 - 3A^2 + I_n = 0$ as $3A^2 - A^3 = I_n$ and then factor to obtain $A(3A - A^2) = I_n$. Now you see that $A^{-1} = 3A - A^2$.

Problems of Chapter 19

19.1 In each case, find the Fourier coefficients of the vector v with respect to the given orthogonal basis $\mathscr{B}$ of the indicated vector space W.

(b) $v = (1,2,3)$, $\mathscr{B} = \{(1,2,3),(-5,4,-1),(1,1,-1)\}$, $W = \mathbb{R}^3$.

Solution: If we write $v = c_1v_1 + c_2v_2 + c_3v_3$, then we know that $c_i = \dfrac{v \cdot v_i}{\|v_i\|^2}$ for $i = 1,2,3$. Hence $(c_1,c_2,c_3) = (1,0,0)$. (Notice that v itself is the first vector in the given orthogonal basis!)

(d) $v = (4,-5,0)$, $\mathscr{B} = \{(-1,0,5),(10,13,2)\}$ $W = \{(x,y,z) \in \mathbb{R}^3 \mid 5x - 4y + z = 0\}$

Solution: If we write $v = c_1v_1 + c_2v_2$, then we know that $c_i = \dfrac{v \cdot v_i}{\|v_i\|^2}$ for $i = 1,2$. Hence $(c_1,c_2) = (-\dfrac{2}{13}, -\dfrac{25}{273})$.

(f) $v = (1,0,1,2)$, $\mathscr{B} = \{(1,0,1,1),(0,1,0,0),(0,0,1,-1),(1,0,0,-1)\}$, $W = \mathbb{R}^4$.

Solution: If we write $v = c_1v_1 + c_2v_2 + c_3v_3 + c_4v_4$, then we know that $c_i = \dfrac{v \cdot v_i}{\|v_i\|^2}$ for $i = 1,\dots,4$. Hence $(c_1,c_2,c_3,c_4) = (\dfrac{4}{3}, 0, -\dfrac{1}{2}, -\dfrac{1}{2})$.

19.2 Find the formula for the orthogonal projection onto the subspaces in parts c), d) and e) above.

Solution:

$$\text{proj}_W(x,y,z) = \frac{(x,y,z)\cdot(-1,0,5)}{26}(-1,0,5) + \frac{(x,y,z)\cdot(10,13,2)}{273}(10,13,2).$$

After expansion and simplification, this becomes

$$\text{proj}_W(x,y,z) = \frac{1}{42}(17x + 20y - 5z, 20x + 26y + 4z, -5x + 4y + 41z)$$

(Note that computations this 'messy' will never occur on tests or exams.)

19.3 Apply the Gram-Schmidt algorithm to each of the following linearly independent sets, and check that your resulting set of vectors is orthogonal.

(b) $\{(1,0,0,1),(0,1,0,-1),(0,0,1,-1)\}$ **Solution:** With the above ordered list denoted $\{v_1,v_2,v_3\}$, we begin by setting $u_1 = v_1 = (1,0,0,1)$.

Then

$$\begin{aligned}\tilde{u}_2 &= v_2 - \frac{v_2 \cdot u_1}{\|u_1\|^2}u_1 = (0,1,0,-1) - \frac{(0,1,0,-1)\cdot(1,0,0,1)}{\|(1,0,0,1)\|^2}(1,0,0,1) \\ &= (0,1,0,-1) + \frac{1}{2}(1,0,0,1) = (\frac{1}{2},1,0,-\frac{1}{2}).\end{aligned}$$

We can rescale at this stage to eliminate fractions – in order to simplify subsequent calculations – if we wish. So we set $u_2 = (1,2,0,-1)$.

N.B. We check $u_1 \cdot u_2 = 0$ **before proceeding, and it's OK.**

Now set

$$\begin{aligned}\tilde{u}_3 &= v_3 - \frac{v_3 \cdot u_1}{\|u_1\|^2} u_1 - \frac{v_3 \cdot u_2}{\|u_2\|^2} u_2 \\ &= (0,0,1,-1) - \frac{(0,0,1,-1)\cdot(1,0,0,1)}{\|(1,0,0,1)\|^2}(1,0,0,1) \\ &\quad - \frac{(0,0,1,-1)\cdot(1,2,0,-1)}{\|(1,2,0,-1)\|^2}(1,2,0,-1) \\ &= (0,0,1,-1) + \frac{1}{2}(1,0,0,1) - \frac{1}{6}(1,2,0,-1) \\ &= (\frac{1}{3}, -\frac{1}{3}, 1, -\frac{1}{3})\end{aligned}$$

Once again we can (but we don't have to) rescale at this stage to eliminate fractions if we wish. So we set $u_3 = (1,-1,3,-1)$. **N.B. We check $u_1 \cdot u_3 = 0 = u_2 \cdot u_3$ before leaving this problem, and it's all OK.**

So after applying Gram-Schmidt (with optional rescaling), we obtain the orthogonal

$$\{(1,0,0,1),(1,2,0,-1),(1,-1,3,-1)\}.$$

(d) $\{(1,1,0),(1,0,2),(1,2,1)\}$

Solution: With the above ordered list denoted $\{v_1, v_2, v_3\}$, we begin by setting $u_1 = v_1 = (1,1,0)$. Then we set

$$\begin{aligned}\tilde{u}_2 = v_2 - \frac{v_2 \cdot u_1}{\|u_1\|^2} u_1 &= (1,0,2) - \frac{(1,0,2)\cdot(1,1,0)}{\|(1,1,0))\|^2}(1,1,0) \\ &= (1,0,2) - \frac{1}{2}(1,1,0) = (\frac{1}{2}, -\frac{1}{2}, 2).\end{aligned}$$

We can rescale at this stage to eliminate fractions – in order to simplify subsequent calculations – if we wish. So we set $u_2 = (1,-1,4)$. **N.B. We check $u_1 \cdot u_2 = 0$ before proceeding, and it's OK.**

Now set

$$\begin{aligned}\tilde{u}_3 &= v_3 - \frac{v_3 \cdot u_1}{\|u_1\|^2} u_1 - \frac{v_3 \cdot u_2}{\|u_2\|^2} u_2 \\ &= (1,2,1) - \frac{(1,2,1)\cdot(1,1,0)}{\|(1,1,0)\|^2}(1,1,0) - \frac{(1,2,1)\cdot(1,-1,4)}{\|(1,-1,4)\|^2}(1,-1,4) \\ &= (1,2,1) - \frac{3}{2}(1,1,0) - \frac{1}{6}(1,-1,4) \\ &= (-\frac{2}{3}, \frac{2}{3}, \frac{1}{3})\end{aligned}$$

Once again we can (but we don't have to) rescale at this stage to eliminate fractions if we wish. So we set $u_3 = (-2,2,1)$. **N.B. We check $u_1 \cdot u_3 = 0 = u_2 \cdot u_3$ before leaving this problem, and it's all OK.**

So after applying Gram-Schmidt (with optional rescaling), we obtain the orthogonal set (in this case an orthogonal basis of $\mathbb{R}^3$):

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

19.4 Find an orthogonal basis for each of the following subspaces, and check that your basis is orthogonal. (First, find a basis in the standard way, and then apply the Gram-Schmidt algorithm.)

(b) $U = \{(x,y,z,w) \in \mathbb{R}^4 \mid x+y-w=0\}$

Solution: Step 1: find any basis for U: A basis for U can be easily found by writing

$$U = \{(-y+w,y,z,w) \in \mathbb{R}^4 \mid y,z,w \in \mathbb{R}\} = \text{span}\{(-1,1,0,0),(0,0,1,0),(-1,0,0,1)\},$$

and noting that $y(-1,1,0,0)+z(0,0,1,0)+w(-1,0,0,1) = (-y+w,y,z,w) = (0,0,0,0)$ iff $y=z=w=0$. So the spanning set $\{(-1,1,0,0),(0,0,1,0),(-1,0,0,1)\}$ we found is indeed a basis for U.

Alternatively (and this is a better way!), since $U = \ker\begin{bmatrix}1 & 1 & 0 & -1\end{bmatrix}$, we can use the algorithm we have for finding a basis for the kernel of a matrix (which comes from Theorem 15.2.2 of "Vector Spaces First"). Then

$$\begin{aligned}\ker\begin{bmatrix}1 & 1 & 0 & -1\end{bmatrix} &= \{(-r+t,r,s,t) \mid r,s,t \in \mathbb{R}\}\\ &= \text{span}\{(-1,1,0,0),(0,0,1,0),(-1,0,0,1)\},\end{aligned}$$

and our algorithm (courtesy of Theorem 15.2.2) *guarantees without us having to check* that the spanning set $\{(-1,1,0,0),(0,0,1,0),(-1,0,0,1)\}$ we found is indeed a basis for U!

Step 2: Apply G-S, perhaps with rescaling: With the above ordered list denoted $\{v_1,v_2,v_3\}$, we begin by setting $u_1 = v_1 = (1,-1,0,0)$.

Then

$$\tilde{u}_2 = v_2 - \frac{v_2 \cdot u_1}{\|u_1\|^2}u_1 = (0,0,1,0) - \frac{(0,0,1,0)\cdot(1,-1,0,0)}{\|(1,-1,0,0)\|^2}(1,-1,0,0) = (0,0,1,0).$$

No need to rescale at this stage! So we set $u_2 = (0,0,1,0)$.

N.B. We check $u_1 \cdot u_2 = 0$ before proceeding, and it's OK.

Now set

$$\begin{aligned}\tilde{u}_3 &= v_3 - \frac{v_3 \cdot u_1}{\|u_1\|^2}u_1 - \frac{v_3 \cdot u_2}{\|u_2\|^2}u_2\\ &= (-1,0,0,1) - \frac{(-1,0,0,1)\cdot(1,-1,0,0)}{\|(1,-1,0,0)\|^2}(1,-1,0,0)\\ &\quad - \frac{(-1,0,0,1)\cdot(0,0,1,0)}{\|(0,0,1,0)\|^2}(0,0,1,0)\\ &= (-1,0,0,1) + \frac{1}{2}(1,-1,0,0) - 0\,(0,0,1,0)\\ &= (-\frac{1}{2},-\frac{1}{2},0,1)\end{aligned}$$

Once again we can (but we don't have to) rescale at this stage to eliminate fractions if we wish. So we set $u_3 = (1,1,0,-2)$. **N.B. We check $u_1 \cdot u_3 = 0 = u_2 \cdot u_3$ before leaving this problem, and it's all OK.**

So after applying Gram-Schmidt (with optional rescaling), we obtain the orthogonal basis

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

for U.

N.B. Since we have a simple description of U as $\{(x,y,z,w) \in \mathbb{R}^4 \mid x+y-w=0\}$, this time we can also check $u_1,u_2,u_3 \in U$ before leaving this problem – and it's all OK.

(d) $V = \ker \begin{bmatrix} 1 & 2 & -1 & -1 \\ 2 & 4 & -1 & 3 \\ -3 & -6 & 1 & -7 \end{bmatrix}$

Solution: Step 1: find any basis for V: $V = \ker \begin{bmatrix} 1 & 2 & -1 & -1 \\ 2 & 4 & -1 & 3 \\ -3 & -6 & 1 & -7 \end{bmatrix} = \ker \begin{bmatrix} 1 & 2 & 0 & 4 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

This kernel is $\{(-2s-4t, s, -5t, t) \mid s,t \in \mathbb{R}\} = \operatorname{span}\{(-2,1,0,0),(-4,0,-5,1)\}$, and so (courtesy of Theorem 15.2.2), a basis for V is $\{(-2,1,0,0),(-4,0,-5,1)\}$.

Step 2: Apply G-S, perhaps with rescaling: With the above ordered basis denoted $\{v_1,v_2\}$, we begin by setting $u_1 = v_1 = (-2,1,0,0)$.

Then

$$\begin{aligned}
\tilde{u}_2 &= v_2 - \frac{v_2 \cdot u_1}{\|u_1\|^2} u_1 \\
&= (-4,0,-5,1) - \frac{(-4,0,-5,1)\cdot(-2,1,0,0)}{\|(-2,1,0,0)\|^2}(-2,1,0,0) \\
&= (-4,0,-5,1) - \frac{8}{5}(-2,1,0,0) \\
&= (-\frac{4}{5}, -\frac{8}{5}, -5, -1)
\end{aligned}$$

Again we can (but we don't have to) rescale at this stage to eliminate fractions, should we wish. Set $u_2 = (4,8,25,5)$, so that $\{(-2,1,0,0),(4,8,25,5)\}$ is an orthogonal basis for V.

19.5 Find the best approximation to each of the given vectors v from the given subspace W.

(b) $v = (1,1,1)$, $W = \{(x,y,z) \in \mathbb{R}^3 \mid 5x-4y+z=0\}$

Solution: From Q. 2 (d) we know

$$\operatorname{proj}_W(x,y,z) = \frac{1}{42}(17x+20y-5z, 20x+26y+4z, -5x+4y+41z),$$

so the best approximation to $v = (1,1,1)$ from W is $\operatorname{proj}_W(1,1,1) = (\frac{16}{21}, \frac{25}{21}, \frac{20}{21})$.

N.B. <u>You must not rescale this answer!</u> This is a very, very common error at this stage. Rescaling an orthogonal basis is fine, but the projection is a fixed answer. For example, if we rescaled this to $(16,25,20)$, this is simply not the correct answer: check that $(16,25,20) \notin W$! The vector $\operatorname{proj}_W(v)$ always belongs to W.

19.6 State whether each of the following is (always) true, or is (possibly) false.

- If you say the statement may be false, you must give an explicit example.
- If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

(b) Every linearly independent set is orthogonal.

Solution: This is false: for example $\{(1,0),(1,1)\}$ is linearly independent but is not orthogonal.

(d) When finding the orthogonal projection of a vector v onto a subspace W, once the answer is obtained, it's OK to rescale the answer to eliminate fractions.

Solution: Absolutely false! See comments at the end of Q.5(b).

(f) When finding the orthogonal projection of a vector v onto a subspace W, using different orthogonal bases of W in the formula can give different answers.

Solution: This is false: it's a wonderful fact (and a consequence of Proposition 19.3.3) that even though the formula for the orthogonal projection looks different when using a different orthogonal basis, the *answer* will always be the same. That is, using different *orthogonal* bases of W in the formula will always give the same answer for the projection.

(h) To check that a vector, say u, is orthogonal to every vector in W, it suffices to check that u is orthogonal to every vector in *any* basis of W.

Solution: This is true — let's prove it.

Suppose u is orthogonal to every vector in some basis $\mathscr{B} = \{w_1, \ldots, w_k\}$ of W, i.e., $u \cdot w_i = 0$ for every i, $1 \leq i \leq k$.

Now let w be any vector in W. We show that $u \cdot w = 0$ as follows. Since $\mathscr{B}$ is a basis of W, we can write $w = a_1w_1 + a_2w_2 + \cdots + a_kw_k$ for some scalars $a_1, \ldots, a_k$. Then

$$\begin{aligned} u \cdot w &= u \cdot (a_1w_1 + a_2w_2 + \cdots + a_kw_k) \\ &= a_1(u \cdot w_1) + a_2(u \cdot w_2) + \cdots + a_k(u \cdot w_k) \\ &= a_1(0) + a_2(0) + \cdots + a_k(0) \\ &= 0. \end{aligned}$$

That is, $u \cdot w = 0$. So we've shown that if u is orthogonal to every vector in any basis of W, then u is orthogonal to every vector in W. (The converse is also clear: if u is orthogonal to every vector in W, it will surely be orthogonal to every vector in any basis of W – since these basis vectors are in W.)

Problems of Chapter 21

21.1 Find the determinants of the following matrices:

(b) $A = \begin{bmatrix} 2 & -1 & 3 \\ 3 & 0 & -5 \\ 1 & 1 & 2 \end{bmatrix}$ **Solution:** $\det A = 30$

(d) $A = \begin{bmatrix} 3 & 4 & -1 \\ 1 & 0 & 3 \\ 2 & 5 & -4 \end{bmatrix}$ **Solution:** $\det A = -10$

(f) $A = \begin{bmatrix} \lambda - 6 & 0 & 0 \\ 0 & \lambda & 3 \\ 0 & 4 & \lambda + 4 \end{bmatrix}$ **Solution:** $\det A = (\lambda - 6)(\lambda + 6)(\lambda + 2)$

(h) $A = \begin{bmatrix} -\lambda & 2 & 2 \\ 2 & -\lambda & 2 \\ 2 & 2 & -\lambda \end{bmatrix}$ **Solution:** $\det A = (\lambda - 4)(\lambda + 2)^2$

21.2 If $\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = 3$, find

(b) $A = \begin{vmatrix} b & a & c \\ e & d & f \\ h & g & i \end{vmatrix}$ **Solution:** $\det A = -3$, since $C_1 \leftrightarrow C_2$ (giving a change of sign) brings this matrix back to the original one.

(d) $\begin{vmatrix} b & 3a & c-4b \\ e & 3d & f-4e \\ h & 3g & i-4h \end{vmatrix}$ **Solution:** $\det A = -9$, since $C_1 + C_3 \rightarrow C_3$ (no change to the determinant) followed by noting the factor of 3 in C_2 (giving a factor of 3 to the determinant) yields the determinant in part (b).

21.3

(b) If B is a 4×4 matrix and $\det(2BB^T) = 64$, find $|\det(3B^2B^T)|$.

Solution: Since $64 = \det(2BB^t) = 2^4(\det B)^2$, $\det B = \pm 2$. Hence $|\det(3B^2B^t)| = 3^4|\det(B^2B^t)| = 3^4(|\det B|)^3 = 81(8) = 648$

(d) Compute the determinant of $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 9 & 10 \\ 11 & 12 \end{bmatrix} \begin{bmatrix} 13 & 14 \\ 15 & 16 \end{bmatrix}$.

Solution:

$$\begin{aligned} \det A &= \det\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 9 & 10 \\ 11 & 12 \end{bmatrix} \begin{bmatrix} 13 & 14 \\ 15 & 16 \end{bmatrix} \right) \\ &= \det \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \det \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \det \begin{bmatrix} 9 & 10 \\ 11 & 12 \end{bmatrix} \det \begin{bmatrix} 13 & 14 \\ 15 & 16 \end{bmatrix} \\ &= (-2)(-2)(-2)\det \begin{bmatrix} 13 & 14 \\ 2 & 2 \end{bmatrix} \\ &= (-2)(-2)(-2)(-2) \\ &= 16 \end{aligned}$$

21.4 State whether each of the following is (always) true, or is (possibly) false.

- If you say the statement may be false, you must give an explicit example.
- If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

In the following A and B are $n \times n$ matrices (with $n > 1$) and k is a scalar.

(b) $\det(A+B) = \det A + \det B$

Solution: This is usually false. For example if $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, then

$$\det(A+B) = \det \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 1 \neq 0 = 0 + 0 = \det A + \det B.$$

(d) $\det(kA) = k^n \det A$

Solution: This is always true, as we saw in class: multiplication of one row (or column) of A by k changes the determinant by a factor of k, so multiplication of n rows (or n columns) –which is the same as multiplying the matrix A by k – will change the determinant by n factors of k, i.e., by a factor of k^n.

(f) If A and B are the same except the first row of A is twice the first row of B, then $\det A = 2\det B$.

Solution: This is always true, as we saw in class. (Expand $\det A$ along the first row.)

21.5

(b) If u, v and w are vectors in $\mathbb{R}^3$, use properties of 3 by 3 determinants to show that

$$u \cdot v \times w = w \cdot u \times v = v \cdot w \times u$$

Solution: We know, since it's the *definition* of the determinant of a 3 by 3 matrix, that $u \cdot v \times w = \det \begin{bmatrix} u \\ v \\ w \end{bmatrix}$, where we've written the vectors as row vectors and the matrix in block row form. So $w \cdot u \times v = \det \begin{bmatrix} w \\ u \\ v \end{bmatrix} = -\det \begin{bmatrix} u \\ w \\ v \end{bmatrix} = \det \begin{bmatrix} u \\ v \\ w \end{bmatrix}$ (two row interchanges: $R_1 \leftrightarrow R_2, R_2 \leftrightarrow R_3$). Similarly, $v \cdot w \times u = \det \begin{bmatrix} v \\ w \\ u \end{bmatrix} = -\det \begin{bmatrix} u \\ w \\ v \end{bmatrix} = \det \begin{bmatrix} u \\ v \\ w \end{bmatrix}$ (two row interchanges again: $R_1 \leftrightarrow R_3, R_2 \leftrightarrow R_3$).

(h) Suppose A, B, C and D are respectively of sizes $m \times m$, $m \times n$, $n \times m$ and $n \times n$. Suppose that D is invertible. Noting that $\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I_m & 0 \\ -D^{-1}C & I_n \end{bmatrix} = \begin{bmatrix} A - BD^{-1}C & B \\ 0 & D \end{bmatrix}$, show that $\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A - BD^{-1}C)\det D$.

Solution: This follows from (g) and the multiplicative property of determinants.

Problems of Chapter 22

22.1 Find the eigenvalues of the following matrices.

(b) $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 2 \end{bmatrix}$

Solution:

$$\det\begin{bmatrix}1-\lambda & 1 & 1\\ 1 & 1-\lambda & -1\\ 1 & 1 & 2-\lambda\end{bmatrix} = \det\begin{bmatrix}1-\lambda & 1 & 1\\ 2-\lambda & 2-\lambda & 0\\ 1 & 1 & 2-\lambda\end{bmatrix} \quad (R_1+R_2\to R_2)$$

$$= (2-\lambda)\det\begin{bmatrix}1-\lambda & 1 & 1\\ 1 & 1 & 0\\ 1 & 1 & 2-\lambda\end{bmatrix} \quad (\text{Factor of } (2-\lambda) \text{ in } R_2)$$

$$= (2-\lambda)\det\begin{bmatrix}1-\lambda & 1 & 1\\ 1 & 1 & 0\\ 0 & 0 & 2-\lambda\end{bmatrix} \quad (-R_2+R_3\to R_3)$$

$$= (2-\lambda)^2\det\begin{bmatrix}1-\lambda & 1\\ 1 & 1\end{bmatrix} \quad (\text{Laplace expansion along } R_3)$$

$$= -\lambda(2-\lambda)^2$$

Hence the eigenvalues are 0 and 2. (The eigenvalue 2 has an ‘algebraic multiplicity’ of 2.)

(d) $\begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 0\\ 1 & 1 & 1\end{bmatrix}$

Solution:

$$\det\begin{bmatrix}1-\lambda & 1 & 1\\ 0 & 1-\lambda & 0\\ 1 & 1 & 1-\lambda\end{bmatrix} = (1-\lambda)\det\begin{bmatrix}1-\lambda & 1\\ 1 & 1-\lambda\end{bmatrix} \quad (\text{Laplace expansion along } R_3)$$

$$= (1-\lambda)\{(1-\lambda)^2-1\}$$

$$= (1-\lambda)\lambda(2-\lambda)$$

Hence the eigenvalues are 0, 1 and 2.

(f) $\begin{bmatrix}2 & 1 & 0\\ 0 & 2 & 1\\ 0 & 0 & 2\end{bmatrix}$

Solution: Since this matrix is an upper triangular matrix, its eigenvalues are the diagonal entries. Hence 2 is the only eigenvalue (with algebraic multiplicity 3).

(h) $\begin{bmatrix}2 & 1 & 0\\ 0 & 2 & 0\\ 0 & 0 & 2\end{bmatrix}$

Solution: Since this matrix is an upper triangular matrix, its eigenvalues are the diagonal entries. Hence 2 is the only eigenvalue (with algebraic multiplicity 3).

22.2 For each of the matrices in the previous question, find a basis for each eigenspace.

Solution:

1. (b) The eigenvalues of $A = \begin{bmatrix}1 & 1 & 1\\ 1 & 1 & -1\\ 1 & 1 & 2\end{bmatrix}$ are 0 and 2, as we saw.

$$\begin{aligned}
E_0 &= \ker(A - 0I_3) \\
&= \ker A \\
&= \ker \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 2 \end{bmatrix} = \ker \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \\
&= \{(-s, s, 0) \mid s \in \mathbb{R}\} \\
&= \operatorname{span}\{(-1, 1, 0)\}
\end{aligned}$$

Hence $\{(-1,1,0)\}$ is basis for E_0.

$$\begin{aligned}
E_2 &= \ker(A - 2I_3) \\
&= \ker \begin{bmatrix} -1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & 1 & 0 \end{bmatrix} \\
&= \ker \begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix} \\
&= \{(\frac{s}{2}, -\frac{s}{2}, s) \mid s \in \mathbb{R}\} \\
&= \operatorname{span}\{(1, -1, 2)\}
\end{aligned}$$

Hence $\{(1,-1,2)\}$ is basis for E_2.

2. (d) The eigenvalues of $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ are 0, 1 and 2, as we saw.

$$\begin{aligned}
E_0 &= \ker(A - 0I_3) \\
&= \ker A = \ker \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \\
&= \ker \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\
&= \{(-s, 0, s) \mid s \in \mathbb{R}\} \\
&= \operatorname{span}\{(-1, 0, 1)\}
\end{aligned}$$

Hence $\{(-1,0,1)\}$ is basis for E_0.

$$\begin{aligned} E_1 &= \ker(A - I_3) \\ &= \ker \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \\ &= \ker \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \\ &= \{(-s,s,0) \mid s \in \mathbb{R}\} \\ &= \operatorname{span}\{(-1,1,0)\} \end{aligned}$$

Hence $\{(-1,1,0)\}$ is basis for E_0. $E_2 = \ker(A-2I_3) = \ker \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 1 & -1 \end{bmatrix} = \ker \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} =$ $\{(s,0,s) \mid s \in \mathbb{R}\} = \operatorname{span}\{(1,0,1)\}$. Hence $\{(1,0,1)\}$ is basis for E_2.

3. (f) The single eigenvalue of $A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$ is 2, as we saw.

   $E_2 = \ker(A-2I_3) = \ker \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \{(s,0,0) \mid s \in \mathbb{R}\} = \operatorname{span}\{(1,0,0)\}$. Hence $\{(1,0,0)\}$ is basis for E_2.

4. (h) The single eigenvalue of $A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ is 2, as we saw.

   $E_2 = \ker(A-2I_3) = \ker \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \{(s,0,t) \mid s \in \mathbb{R}\} = \operatorname{span}\{(1,0,0),(0,0,1)\}$. Hence $\{1,0,0),(0,0,1)\}$ is basis for E_2.

Problems of Chapter 23

23.1 For each of the matrices A in the previous question, if possible find and invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$. If this is not possible, explain why.

Solution:

1. (b) Since $\dim E_0 + \dim E_2 = 1 + 1 = 2 < 3$, this matrix is not diagonalizable — there is no basis of $\mathbb{R}^3$ consisting of eigenvectors of A: there are at most 2 linearly independent eigenvectors of A.
2. (d) This 3×3 matrix, having 3 distinct eigenvalues, will be diag'ble. Indeed, if $v_0 = (-1,0,1)$, $v_1 = (-1,1,0)$ and $v_2 = (1,0,1)$ are written as columns, set

$$P = \begin{bmatrix} v_0 & v_1 & v_2 \end{bmatrix} = \begin{bmatrix} -1 & -1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

and

$$D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix},$$

then P will be invertible (you can check this directly, or simply note that $\dim E_0 + \dim E_1 + \dim E_2 = 1+1+1 = 3$, which guarantees P's invertibility) and $P^{-1}AP = D$.

3. (f) Since there is only one eigenvalue (2) and $\dim E_2 = 1 < 3$, this matrix is not diagonalizable — there is no basis of $\mathbb{R}^3$ consisting of eigenvectors of A: there is at most one linearly independent eigenvector of A!
4. (h) Since there is only one eigenvalue (2) and $\dim E_2 = 2 < 3$, this matrix is not diagonalizable — there is no basis of $\mathbb{R}^3$ consisting of eigenvectors of A: there are at most two linearly independent eigenvectors of A.

23.2 State whether each of the following is (always) true, or is (possibly) false.

- If you say the statement may be false, you must give an explicit example.
- If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

1. (b) The matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ has no real eigenvalues.

   Solution: This is true: $\det(A - \lambda I_2) = \lambda^2 + 1$, so $\det(A - \lambda I_2) = 0$ has no real solutions.
2. (d) If 0 is an eigenvalue of $n \times n$ matrix A, then A is not invertible.

   Solution: This is true: 0 is an eigenvalue of A iff $\det(A - 0I_n) = \det A = 0$. Hence, A is not invertible.
3. (f) Every invertible matrix is diagonalizable.

   Solution: This is false. See example (f) or (h) from the previous question, or, check for yourself that $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, which has 1 as its only eigenvalue, is not diagonalizable.
4. (h) If an $n \times n$ matrix has n distinct eigenvalues, then the matrix is diagonalizable.

   Solution: This is true, as we saw in class: for every distinct eigenvalue, we obtain an eigenvector, and we know that eigenvectors corresponding to distinct eigenvalues are linearly independent. Hence there are n linearly independent eigenvectors, which will of course be a basis for $\mathbb{R}^n$, as $\dim \mathbb{R}^n = n$.
5. (j)[3] If an $n \times n$ matrix A has eigenvalues $\lambda_1, \dots, \lambda_n$, then $\det A = \lambda_1 \dots \lambda_n$.

   Solution: Use the hint and substitute $\lambda = 0$ into the equation.

23.3 Let $A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$.

a) Compute $\det(A - \lambda I_3)$ and hence show that the eigenvalues of A are 2 and -1.

[3] Hint: Use the fact that we know $\det(A - \lambda I_n) = (-1^n)(\lambda - \lambda_1)(\lambda - \lambda_2)\dots(\lambda - \lambda_n)$.

Solution: We have

$$A-\lambda I=\begin{bmatrix}-\lambda & 1 & 1\\ 1 & -\lambda & 1\\ 1 & 1 & -\lambda\end{bmatrix}.$$

To find the determinant, we could do a cofactor expansion and then expand, or we can apply some row operations first, keeping track of any changes to the value of the determinant. (In doing so, we avoid dividing by λ since we don't know if it is zero.)

$$\begin{bmatrix}-\lambda & 1 & 1\\ 1 & -\lambda & 1\\ 1 & 1 & -\lambda\end{bmatrix}\begin{matrix}\lambda R2+R1\to R1\\ \sim\\ -R2+R3\to R3\end{matrix}\begin{bmatrix}0 & 1-\lambda^2 & 1+\lambda\\ 1 & -\lambda & 1\\ 0 & 1+\lambda & -\lambda-1\end{bmatrix}$$

So we compute (keeping an eye on common factors, to simplify our factoring work):

$$\begin{aligned}\det(A-\lambda I)&=\begin{vmatrix}0 & 1-\lambda^2 & 1+\lambda\\ 1 & -\lambda & 1\\ 0 & 1+\lambda & -\lambda-1\end{vmatrix}=-1\begin{vmatrix}1-\lambda^2 & 1+\lambda\\ 1+\lambda & -\lambda-1\end{vmatrix}\\ &=-((1-\lambda^2)(-\lambda-1)-(1+\lambda)(1+\lambda))=(1-\lambda^2)(1+\lambda)+(1+\lambda)^2\\ &=(1+\lambda)^2(1-\lambda+1)\\ &=(1+\lambda)^2(2-\lambda)\end{aligned}$$

It follows that the eigenvalues are 2 (with algebraic multiplicity 1) and -1 (with algebraic multiplicity 2).

b) Find a basis of $E_2=\{x\in\mathbb{R}^3 \mid Ax=2x\}$.

Solution:

$$\begin{aligned}E_2=\ker(A-2I)&=\ker\begin{bmatrix}-2 & 1 & 1\\ 1 & -2 & 1\\ 1 & 1 & -2\end{bmatrix}\\ &=\ker\begin{bmatrix}1 & -2 & 1\\ 0 & -3 & 3\\ 0 & 3 & -3\end{bmatrix}=\ker\begin{bmatrix}1 & 0 & -1\\ 0 & 1 & -1\\ 0 & 0 & 0\end{bmatrix}\\ &=\{(s,s,s)\mid s\in\mathbb{R}\}=\operatorname{span}\{(1,1,1)\}.\end{aligned}$$

c) Find a basis of $E_{-1}=\{x\in\mathbb{R}^3 \mid Ax=-x\}$.

Solution:

$$\begin{aligned}E_{-1}=\ker(A-(-1)I)=\ker(A+I)&=\ker\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}\\ &=\ker\begin{bmatrix}1 & 1 & 1\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\\ &=\{(-s-t,s,t)\mid s,t\in\mathbb{R}\}=\operatorname{span}\{(-1,1,0),(-1,0,1)\}.\end{aligned}$$

d) Find an invertible matrix P such that $P^{-1}AP = D$ is diagonal, and give this diagonal matrix D. Explain why your choice of P is invertible.

Solution: We can take

$$P = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Since the columns of P are eigenvectors, column multiplication shows $AP = PD$. Since eigenvectors from different eigenspaces are linearly independent, and $\dim(E_2) + \dim(E_{-1}) = 3$, the columns of P form a basis for $\mathbb{R}^3$ and so P is invertible. Thus $P^{-1}AP = D$.

e) Find an invertible matrix $Q \neq P$ such that $Q^{-1}AQ = \tilde{D}$ is also diagonal, and give this diagonal matrix $\tilde{D}$.

Solution: We may replace P with any matrix such that the columns are linearly independent eigenvectors of A (such as nonzero scalar multiples of the columns of P). For a more interesting example, you can verify that

$$Q = \begin{bmatrix} -2 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \quad \text{and} \quad \tilde{D} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

is another valid answer.

Problems of Chapter 24

24.1 State whether each of the following defines a linear transformation.

- If you say it isn't linear, you must give an explicit example to illustrate.
- If you say it is linear, you must give a clear explanation - by quoting a theorem presented in class, or by verifying the conditions in the definition *in every case*.

1. (b) $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x,y,z) = (x+2z, y)$

   Solution: This is indeed a linear transformation, since the formula comes from multiplication by the matrix $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \end{bmatrix}$: i.e., if $v = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3$ (is written as a column), then $Av = \begin{bmatrix} x+2z \\ y \end{bmatrix}$, which is the formula above (written as a column).

2. (d) $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(v) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} v$

   Solution: This is indeed a linear transformation, since we know (first example, section 24.1) that multiplication by a matrix always gives a linear transformation.

3. (f) $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(v) = \text{proj}_{(1,1,-1)}(v)$.

   Solution: This is a linear transformation. We can show this in two ways: via the definition, or by finding the standard matrix of T.
   First, we need just to write out the formula for T more explicitly:

$$(1) \quad T(v) = \text{proj}_{(1,1,-1)}(v) = \frac{v \cdot (1,1,-1)}{\|(1,1,-1)\|}(1,1,-1) = \frac{v \cdot (1,1,-1)}{3}(1,1,-1).$$

In Cartesian coordinates, if $v=(x,y,z)$, this is

$$(2)\quad T(v)=\frac{(x,y,z)\cdot(1,1,-1)}{\|(1,1,-1)\|}(1,1,-1)=\frac{(x+y-z)}{3}(1,1,-1)$$

which simplifies to $\frac{1}{3}(x+y-z,x+y-z,-x-y+z)$.

1. Via the definition:

(i) If $u,v\in\mathbb{R}^3$, then by (1) above,

$$\begin{aligned}T(u+v)&=\frac{u\cdot(1,1,-1)}{3}(1,1,-1)+\frac{v\cdot(1,1,-1)}{3}(1,1,-1)\\&=\Big(\frac{u\cdot(1,1,-1)}{3}+\frac{v\cdot(1,1,-1)}{3}\Big)(1,1,-1)\\&=\frac{(u+v)\cdot(1,1,-1)}{3}(1,1,-1)\\&=T(u)+T(v)\end{aligned}$$

(ii) If $k\in\mathbb{R}$ and $v\in\mathbb{R}^3$,

$$\begin{aligned}T(kv)&=\frac{(kv)\cdot(1,1,-1)}{3}(1,1,-1)\\&=k\Big(\frac{v\cdot(1,1,-1)}{3}(1,1,-1)\Big)\\&=kT(v)\end{aligned}$$

These show that T satisfies the two conditions of the definition, so T is linear.

2. If $v=\begin{bmatrix}x\\y\\z\end{bmatrix}$, then by (2) above, if $A=\frac{1}{3}\begin{bmatrix}1&1&-1\\1&1&-1\\-1&-1&1\end{bmatrix}$

$$T(v)=\frac{1}{3}\begin{bmatrix}x+y-z\\x+y-z\\-x-y+z\end{bmatrix}=\frac{1}{3}\begin{bmatrix}1&1&-1\\1&1&-1\\-1&-1&1\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=Av.$$

Hence, T is multiplcation by a matrix, and so by Example 24.1.1, T is linear.

4. (h) $T:\mathbb{R}^3\to\mathbb{R}^3$ defined by $T(v)=\text{proj}_v(1,1,-1)$.

Solution: This looks at first like part (f), but isn't — 'v' is in the wrong spot. Indeed this is **not** a linear transformation: look first at the formula for T:

$$T(v)=\frac{(1,1,-1)\cdot v}{\|v\|^2}(1,1,-1)$$

The dependence on v is rather complicated: it appears in the numerator as '$(1,1,-1)\cdot v$' (which is fine in itself) but the warning bells go off when we see the factor of '$\frac{1}{\|v\|^2}$'!

So let's give a counterexample to show T is not linear.

Indeed, let $v=(1,0,0)$. Then $T(v)=(1,1,-1)$, but

$$T(2v)=T(2,0,0)=\frac{1}{2}(1,1,-1)\neq 2(1,1,-1)=2T(v).$$

So T is **not** linear.

5. (j) $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(v) = 2v$. **Solution:** This is a linear transformation. It is easy to check it satisfies the conditions of the definition, but it's even easier to note that if $v \in \mathbb{R}^3$, then $T(v) = 2I_3 v$, i.e., T is multiplication by the matrix $2I_3$ and so is a linear transformation.

6. (l) $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(v) = Av$, where $A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 3 \end{bmatrix}$.

   Solution: This is a linear transformation, since it is defined by multiplication by a matrix!

24.2 In each of the following, find the standard matrix of T and use it to give a basis for $\ker T$ and $\operatorname{im} T$ and verify the conservation of dimension.

1. (b) $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x,y,z) = (2z+x, y)$

   Solution: We saw in Q.1 (b) that the standard matrix for T here is $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \end{bmatrix}$.
   Since A is already in RRE form, a basis for $\ker T = \ker A$ is easily seen to be $\{(-2,0,1)\}$. The column space algorithm yields the basis $\{(1,0),(0,1)\}$ for $\operatorname{im} T = \operatorname{Col} A$.

2. (d) $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(v) = \operatorname{proj}_{(1,1,-1)}(v)$.

   Solution: We saw in 24.1(f) that T is multiplication by $A = \frac{1}{3}\begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$.

   A short computation shows that the RRE form of A is $\begin{bmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, and so a basis for $\ker T = \ker A$ is easily seen to be $\{(-1,1,0),(1,0,1)\}$.
   The column space algorithm yields the basis $\{(1,1,-1)\}$ for $\operatorname{im} T = \operatorname{Col} A$ – as you would expect from a projection onto the line with direction $(1,1,-1)$!

3. (f) $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(v) = \operatorname{proj}_H(v)$, where H is the plane through the origin with normal $(1,1,0)$.

   Solution: We first need the formula for T. Since the subspace H is a plane in $\mathbb{R}^3$, we don't have to find an orthogonal basis of H and use the projection formula: instead, we note that if n is any normal vector for H, and v is any vector in $\mathbb{R}^3$, that

   $$v = \operatorname{proj}_n(v) + (v - \operatorname{proj}_n(v)),$$

   and that the vector $w = v - \operatorname{proj}_n(v)$ *will indeed be the orthogonal projection of v onto H*, since it satisfies both the conditions of Theorem 19.3.3! That is[4],

   $$\operatorname{proj}_H(v) = v - \operatorname{proj}_n(v).$$

   So the formula for T is $T(v) = v - \frac{v \cdot n}{\|n\|^2} n$. Let $v = (x,y,z)$. Since $n = (1,1,0)$,

   $$T(x,y,z) = (x,y,z) - \frac{x+y}{2}(1,1,0) = (\frac{y-x}{2}, \frac{x-y}{2}, z).$$

[4]First: Note that $w \cdot n = \big(v - \operatorname{proj}_n(v)\big) \cdot n = \big(v - \frac{v \cdot n}{\|n\|^2} n\big) \cdot n = v \cdot n - \frac{v \cdot n}{\|n\|^2} n \cdot n = v \cdot n - v \cdot n = 0$, so it is true that $w = v - \operatorname{proj}_n(v) \in H$. Secondly, $w - v = \big(v - \operatorname{proj}_n(v)\big) - v = -\operatorname{proj}_n(v)$ which is certainly orthogonal to H, since it is a multiple of the normal n to H. Draw yourself a picture to convince yourself geometrically.

Now it is a simple matter to check that T is multiplication by the matrix $A = \frac{1}{2}\begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$.

A short computation shows that the RRE form of A is $\begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$, and so a basis for $\ker T = \ker A$ is easily seen to be $\{(1,1,0)\}$, as you would expect for the projection onto the plane through 0 with normal $(1,1,0)$.

The column space algorithm yields the basis $\{(-1,1,0),(0,0,1)\}$ for $\text{im}\ T = \text{Col}\, A$ – and you can check yourself that $\text{span}\{(-1,1,0),(0,0,1)\} = H$, as you'd expect!

24.3 State whether each of the following is (always) true, or is (possibly) false.

- If you say the statement may be false, you must give an explicit example.
- If you say the statement is true, you must give a clear explanation - by quoting a theorem presented in class, or by giving a *proof valid for every case*.

1. (b) If $T : \mathbb{R}^4 \to \mathbb{R}^2$ is linear, then $\dim\ker T \geq 2$.

   Solution: This is true, since we know that $\dim\ker T + \dim\text{im}\ T = 4$, and that, as $\text{im}\ T$ is a subspace of $\mathbb{R}^2$, $\dim\text{im}\ T \leq \dim\mathbb{R}^2 = 2$. Thus, $\dim\ker T = 4 - \dim\text{im}\ T \geq 4-2 = 2$.

2. (d) If $T : \mathbb{R}^3 \to \mathbb{R}^2$ is linear, and $\{v_1, v_2\} \subset \mathbb{R}^3$ is linearly independent, then $\{T(v_1), T(v_2)\} \subset \mathbb{R}^2$ is linearly independent.

   Solution: This is false, indeed it is only (always) true if $\ker T = \{0\}$. For example, define $T : \mathbb{R}^3 \to \mathbb{R}^2$ by $T(x,y,z) = (0,0)$, and let $v_1 = (1,0,0)$ and $v_2 = (0,1,0)$. Then $\{T(v_1), T(v_2)\}$ is not independent as it contains the zero vector.

3. (f) If $T : \mathbb{R}^3 \to \mathbb{R}^3$ is linear, and $\ker T = \{0\}$, then $\text{im}\ T = \mathbb{R}^3$.

   Solution: This is true. Recall that $\dim\ker T + \dim\text{im}\ T = 3$ in this case, so if $\ker T = \{0\}$, then $\dim\ker T = 0$ and so $\dim\text{im}\ T = 3$. But $\text{im}\ T$ is then a 3-dimensional subspace of $\mathbb{R}^3$, which also has a dimension of 3, so by the Theorem 10.3.1, $\text{im}\ T = \mathbb{R}^3$.

24.4* State whether each of the following defines a linear transformation.

- If you say it isn't linear, you must give an explicit example to illustrate.
- If you say it is linear, you must give a clear explanation - by quoting a theorem presented in class, or by verifying the conditions in the definition *in every case*.

(b) $T : \mathscr{P} \to \mathscr{P}$ defined by $T(p)(t) = \int_0^t p(s)ds$.

Solution: You've learned in calculus (and will prove, if you take a course in analysis) that if p and q are (integrable) functions, then

$$\int_0^t (p+q)(s)ds = \int_0^t p(s)ds + \int_0^t q(s)ds.$$

Moreover you've also seen that if $k \in \mathbb{R}$ is a scalar, then $\int_0^t kp(s)ds = k\int_0^t p(s)ds$. Together, these show that T is indeed linear.

(d) $\det : \mathbf{M}_{22} \to \mathbb{R}$ defined by $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$.

Solution: The determinant is *not* linear. For example if $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, then

$$\det(A+B) = \det \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 1 \neq 0 = 0 + 0 = \det A + \det B.$$

(We saw this example in the solution to problem 21.4(b) on determinants.)

References

Amongst the many good references for this material are the following:

1. *Elementary Linear Algebra*, Howard Anton.
2. *Linear Algebra*, Tom M. Apostol.
3. *Linear Algebra with Applications*, Otto Bretscher.
4. *Linear Algebra and Its Applications*, David C. Lay.
5. *Elementary Linear Algebra*, W. Keith Nicholson (Second Edition, McGraw Hill, 2003.)
6. *Linear Algebra (Schaum's outlines)*, Seymour Lipschutz and Marc Lipson.
7. *Linear Algebra with Applications*, W. Keith Nicholson. (Sixth Edition, McGraw Hill.)
8. *Linear Algebra and Its Applications*, Gilbert Strang. (Fourth Edition, Wellesley-Cambridge Press.)
9. *Introduction to Linear Algebra*, Gilbert Strang.

For those of who would appreciate a more theoretical approach, you might try the following:

1. *Linear Algebra*, Kenneth M. Hoffman and Ray Kunze. (Second edition, Pearson)
2. *Linear Algebra*, Sterling K. Berberian.
3. *Introduction to Linear Algebra*, Serge Lang.


Photo: Ralph Nevins. Parc Provincial Driftwood Provincial Park, Ontario

Index