Lectures in Linear Algebra

David M. McClendon

Department of Mathematics
Ferris State University

last revised December 2016

Contents

Chapter 1

Vector spaces

1.1 Why do we study "linear" things?

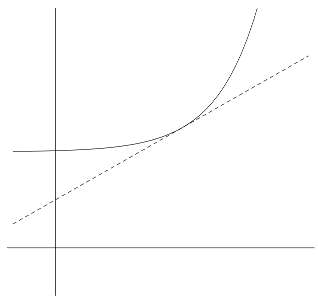
REASON # 1:
LINEAR THINGS APPROXIMATE ANYTHING THAT ISN'T LINEAR

In calculus, we learn about *derivatives*. Computing the derivative of a function is useful because something associated to a derivative solves lots of problems, including

-
-
-
-
-
-
-
- etc.

The concept behind all these applications is that if you have any function f and you want to approximate the values of f for values of x near a, you can *approximate f by a linear function* (namely, the tangent line at a):

$$\begin{aligned} f(x) &\approx L(x) \\ &= f(a) + f'(a)(x-a). \end{aligned}$$



Doing this makes sense because linear equations are easier to work with than arbitrary equations (see reason # 2 below for more on this).

Long-term goal (Math 320, perhaps): So far, you might only have heard about functions $f : \mathbb{R} \to \mathbb{R}$ (that is, functions $y = f(x)$ or $x \mapsto y$). But functions may have much more general domains and ranges; for example, a set of parametric equations describing planar motion (Math 230) can be thought of as a single function

$$f : t \mapsto (x, y)$$

which is a function $f : \mathbb{R} \to \mathbb{R}^2$.

A function $f : \mathbb{R}^2 \to \mathbb{R}^3$ might be something like

$$f(u, v) = (u^3 v, \sin(uv + 1), e^{2u} - 3).$$

A natural question is to ask what the derivative of such a function is. It should have something to do with the best linear approximation of the function (just as the derivative of a function $y = f(x)$ gives the slope of the best linear approximation to f at x). But what does linear mean in that context?

The first major goal of linear algebra is to describe what "linear" means, in a general sense.

REASON # 2 (TO STUDY LINEAR THINGS):
LINEAR THINGS ARE (RELATIVELY) EASY TO STUDY

Suppose you have an equation in one variable (say x). Every such equation can be written as $f(x) = c$, where f is some function from $\mathbb{R}$ to $\mathbb{R}$ and c is some constant. For example,

$$\sin x + x^2 - 6 = 3(x - 2) + x - 5$$

can be rewritten as

To think about such an equation conceptually/theoretically, we need to ask the following questions:

1.

2.

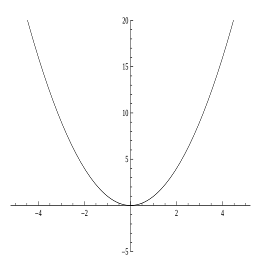
3.

Question 1 above is easy to answer theoretically:

There is at least one solution to $f(x) = c \iff$

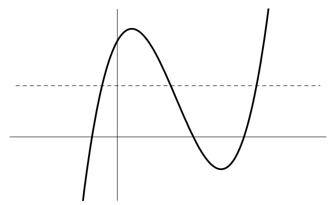
$\iff$

Example: $x^2 = c$ has at least one solution if and only if $c \geq 0$.



Unfortunately, in practice, Question 1 might be hard because we might not have access to a graph of $y = f(x)$.

Question 2 is also easy to answer theoretically:
The number of solutions to $f(x) = c$ is equal to



But in practice, Question 2 is also hard because we might not have access to a graph of $y = f(x)$, and even if we have a graph, there may be intersection points we're not aware of (that lie off the piece of the graph we see).

Question 3, in general, is impossible:
Consider the equation $x = \cos x$, i.e. $\cos x - x = 0$. We know there is a solution (since the graph of $\cos x - x$ crosses the $x-$axis), but good luck finding its exact value.

However, there is one class of equations in one variable that we can completely answer questions 1,2 and 3 for (and answer these questions **easily**). Suppose $f(x)$ is a function whose graph is a line, i.e.

In this case the equation $f(x) = c$ becomes

On the previous page we saw that every linear equation in one variable can be rewritten as $Ax = b$, where A and b are constants.

Let's look at our three questions in this situation:

Equation $\mathbf{Ax = b}$	Does the equation have a solution?	How many solutions does the equation have?	What is/are the solution(s)?
Case 1: $A \neq 0$			
Case 2: $A = 0, b = 0$			
Case 3: $A = 0, b \neq 0$			

Summarizing, we have:

Theorem 1.1 (Summary of Linear Equations in One Variable) *Every linear equation in one variable x can be rewritten as $Ax = b$ for constants A and b.*

1. *If $A \neq 0$, then an equation has exactly one solution, namely $x = \frac{b}{A} = A^{-1}b$.*
2. *If $A = 0$ and $b = 0$, then the equation has infinitely many solutions (every $x \in \mathbb{R}$ is a solution).*
3. *If $A = 0$ and $b \neq 0$, then the equation has no solution.*

Note: Once you get much beyond linear equations, these questions no longer have such nice answers.

In Math 322, we want to think about linear systems with more than one equation and/or more than one variable (in part so that we could eventually describe the derivative of a function of several variables, but also because such equations arise naturally). For example, suppose we want to solve for x and y if

$$\begin{cases} a_{11}x + a_{12}y = b_1 \\ a_{21}x + a_{22}y = b_2 \end{cases}$$

where the as and bs are constants. This is a system of two equations in two variables. There are a variety of ways one might solve this system:

Example: Solve the system

$$\begin{cases} 2x - 3y = 13 \\ x - 2y = 7 \end{cases}$$

<u>Method 1</u>: Addition / elimination

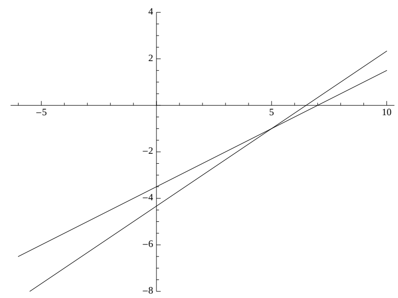
$$\begin{cases} 2x - 3y = 13 \\ x - 2y = 7 \end{cases}$$

<u>Method 2</u>: Substitution

$$\begin{cases} 2x - 3y = 13 \\ x - 2y = 7 \end{cases}$$

Method 3: Graphical (imprecise in general)

$$\begin{cases} 2x - 3y = 13 \\ x - 2y = 7 \end{cases}$$



These methods are fine if you have two equations in two variables, but what if you have m equations in n variables? An example is something like

$$\begin{cases} 3w + 2x - y + 5z = 8 \\ 2w - x + 4y - z = -3 \\ w + 5x + 2y + 6z = 17 \end{cases}$$

This is a system of three equations in four variables. How do you solve this for w, x, y, z? Does this system even have a solution? If so, how many solutions does it have?

Big picture conceptual approach:

- think of the left hand side of this system *as a single function* called T
- think of the four variables (w, x, y, z) *as a single variable* $\mathbf{x}$ (a.k.a. $\overrightarrow{x}$).
- group the right-hand side of the equation $(8, -3, 17)$ *into a single object* called $\mathbf{b}$ (a.k.a. $\overrightarrow{b}$).
- Then the system above *becomes a single equation*

$$T(\mathbf{x}) = \mathbf{b}.$$

What's hard: $\mathbf{x}$ and $\mathbf{b}$ aren't numbers. They are objects called *vectors*. Similarly, T isn't a function like any you've seen before. It turns out that T is something called a *linear transformation* (and is represented by an object called a *matrix*).

Just like what happened with our earlier example

$$Ax = b,$$

whether or not the system on the previous page (now written as a single equation)

$$T(\mathbf{x}) = \mathbf{b} \quad \text{(also written } A\mathbf{x} = \mathbf{b} \text{ where } A \text{ is a matrix)}$$

has a solution (and what that solution is) boils down to whether or not $\mathbf{b}$ is in the range of T, and this probably depends on some properties of T and/or $\mathbf{b}$. So to study these systems in general, we need to figure out what these properties are, and study the nature of linear transformations (and matrices and vectors) in general. This gets us back to the first goal of Math 322, which we already stated earlier:

The first major goal of linear algebra is to describe what "linear" means, in a general sense.

Once we do this, we will then know how to solve systems of any number of equations in any number of variables. Turns out, we will also discover lots of other useful stuff along the way.

1.2 Introducing vector spaces

Recall that our goal is to study what "linear" means, in a general sense. To get started, remember that what we know about linear equations from high school is that they all (other than vertical lines) have equation

$$y = mx + b$$

What operations are required to describe this equation?

1.

2.

Based on this, it is reasonable to expect that if we are going to define "linear" in a general sense, we probably need to assume that there is some notion of each of the two operations above.

A general setting in which we can add objects and multiply objects by real numbers is called a *vector space*, and such a setting is defined precisely on the next page.

Definition 1.2 *A* **(real) vector space** V *is a set, together with two operations:*

- ***addition:*** $+ : V \times V \to V$ *(i.e.* $(\mathbf{u}, \mathbf{v}) \mapsto \mathbf{u} + \mathbf{v}$*)*
- ***scalar multiplication:*** $\mathbb{R} \times V \to V$ *(i.e.* $(c, \mathbf{v}) \mapsto c\mathbf{v}$*)*

such that the following rules (called the "Vector Space Laws") are satisfied:

1. Addition is closed: *For all* $\mathbf{u}, \mathbf{v} \in V$, $\mathbf{u} + \mathbf{v} \in V$.
2. Scalar multiplication is closed: *For all* $c \in \mathbb{R}$ *and* $\mathbf{v} \in V$, $c\mathbf{v} \in V$.
3. Addition is commutative: *For all* $\mathbf{u}, \mathbf{v} \in V$, $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
4. Addition is associative: *For all* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.
5. Additive identity element: *There exists an element of* V *called the* **zero vector (of** V**)** *and denoted* $\mathbf{0}$, *such that* $\mathbf{v} + \mathbf{0} = \mathbf{v}$ *for all* $\mathbf{v} \in V$.
6. Additive inverses exist: *For all* $\mathbf{v} \in V$, *there exists an element* $-\mathbf{v} \in V$ *such that* $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
7. Distributivity I: *For all* $c \in \mathbb{R}$ *and all* $\mathbf{u}, \mathbf{v} \in V$, $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$.
8. Distributivity II: *For all* $c, d \in \mathbb{R}$ *and all* $\mathbf{v} \in V$, $(c + d)\mathbf{v} = c\mathbf{v} + d\mathbf{v}$.
9. Scalar multiplication is associative: *For all* $c, d \in \mathbb{R}$ *and all* $\mathbf{v} \in V$, $(cd)\mathbf{v} = c(d\mathbf{v})$.
10. Identity element for scalar multiplication: *For all* $\mathbf{v} \in V$, $1\mathbf{v} = \mathbf{v}$.

In linear algebra, real numbers are called **scalars**; *elements of the vector space are called* **vectors**.

Comment: There are other types of vector spaces (other than "real" vector spaces). These vector spaces have different types of scalars (the scalars can be complex numbers, rational numbers, or more exotic things). In Math 322, we will not deal with these.

Notation: Vectors are usually referred to by boldface letters (like **v**) when typed, and as letters with arrows over them (like $\vec{v}$) when hand-written. However, sometimes we get lazy and just refer to a vector with a letter (like v). The zero scalar is denoted 0; the zero vector is denoted **0** or $\vec{0}$.

Warning: Vectors from two different vector spaces cannot be added to one another; for example, if V and W are two unrelated vector spaces with $\mathbf{u}, \mathbf{v} \in V$ but $\mathbf{w} \in W$, then $\mathbf{u} + \mathbf{v}$ makes sense but $\mathbf{u} + \mathbf{w}$ is nonsense.

Our first example of a vector space

> **Theorem 1.3** *The set* $\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$ *is a real vector space, where the addition and scalar multiplication are defined coordinate-wise, i.e.*
>
> $$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2) \quad \textit{and} \quad c(x_1, y_1) = (cx_1, cy_1).$$

PROOF Let $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ be arbitrary elements of $\mathbb{R}^2$, and let $c, d \in \mathbb{R}$. By definition of $\mathbb{R}^2$, we have $\mathbf{u} = (u_1, u_2)$; $\mathbf{v} = (v_1, v_2)$ and $\mathbf{w} = (w_1, w_2)$. We check the ten vector space laws one by one:

1. *Addition is closed:* This is obvious by the definition of vector addition.
2. *Scalar multiplication is closed:* This is obvious by the definition of scalar multiplication.
3. *Addition is commutative:* We need to check $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$:

$$\begin{aligned} \mathbf{u} + \mathbf{v} &= (u_1, u_2) + (v_1, v_2) \\ &= (u_1 + v_1, u_2 + v_2) \quad \text{(by def'n of + in } \mathbb{R}^2\text{)} \\ &= (v_1 + u_1, v_2 + u_2) \quad \text{(since + is commutative in } \mathbb{R}\text{)} \\ &= (v_1, v_2) + (u_1, u_2) \quad \text{(by def'n of + in } \mathbb{R}^2\text{)} \\ &= \mathbf{v} + \mathbf{u}. \end{aligned}$$

4. *Addition is associative:*

$$\begin{aligned} \mathbf{u} + (\mathbf{v} + \mathbf{w}) &= (u_1, u_2) + ((v_1, v_2) + (w_1, w_2)) \\ &= (u_1, u_2) + (v_1 + w_1, v_2 + w_2) \quad \text{(by def'n of + in } \mathbb{R}^2\text{)} \\ &= (u_1 + (v_1 + w_1), u_2 + (v_2 + w_2)) \quad \text{(by def'n of + in } \mathbb{R}^2\text{)} \\ &= ((u_1 + v_1) + w_1, (u_2 + v_2) + w_2) \quad \text{(since + is associative in } \mathbb{R}\text{)} \\ &= (u_1 + v_1, u_2 + v_2) + (w_1, w_2) \quad \text{(by def'n of + in } \mathbb{R}^2\text{)} \\ &= ((u_1, u_2) + (v_1, v_2)) + (w_1, w_2) \quad \text{(by def'n of + in } \mathbb{R}^2\text{)} \\ &= (\mathbf{u} + \mathbf{v}) + \mathbf{w}. \end{aligned}$$

5. *Additive identity element:* Let $\mathbf{0} = (0, 0)$. Then

$$\begin{aligned} \mathbf{v} + \mathbf{0} &= (v_1, v_2) + (0, 0) \\ &= (v_1 + 0, v_2 + 0) \quad \text{(by def'n of + in } \mathbb{R}^2\text{)} \\ &= (v_1, v_2) \quad \text{(since 0 is additive identity in } \mathbb{R}\text{)} \\ &= \mathbf{v}. \end{aligned}$$

6. *Additive inverses exist:* Let $-\mathbf{v} = (-v_1, -v_2)$. Then

$$\begin{aligned}
\mathbf{v} + (-\mathbf{v}) &= (v_1, v_2) + (-v_1, -v_2) \\
&= (v_1 + (-v_1), v_2 + (-v_2)) \quad \text{(by def'n of } + \text{ in } \mathbb{R}^2\text{)} \\
&= (0, 0) \quad \text{(by properties of additive inverses in } \mathbb{R}\text{)} \\
&= \mathbf{0} \quad \text{(by def'n of } \mathbf{0}\text{)}.
\end{aligned}$$

7. *Distributivity I:*

$$\begin{aligned}
c(\mathbf{u} + \mathbf{v}) &= c((u_1, u_2) + (v_1, v_2)) \\
&= c(u_1 + v_1, u_2 + v_2) \quad \text{(by def'n of } + \text{ in } \mathbb{R}^2\text{)} \\
&= (c(u_1 + v_1), c(u_2 + v_2)) \quad \text{(by def'n of scalar multiplication)} \\
&= (cu_1 + cv_1, cu_2 + cv_2) \quad \text{(by distributivity of } \mathbb{R}\text{)} \\
&= (cu_1, cu_2) + (cv_1, cv_2) \quad \text{(by def'n of } + \text{ in } \mathbb{R}^2\text{)} \\
&= c(u_1, u_2) + c(v_1, v_2) \quad \text{(by def'n of scalar multiplication)} \\
&= c\mathbf{u} + c\mathbf{v}.
\end{aligned}$$

8. *Distributivity II:*

$$\begin{aligned}
(c + d)\mathbf{v} &= (c + d)(v_1, v_2) \\
&= ((c + d)v_1, (c + d)v_2) \quad \text{(by def'n of scalar multiplication)} \\
&= (cv_1 + dv_1, cv_2 + dv_2) \quad \text{(by distributivity of } \mathbb{R}\text{)} \\
&= (cv_1, cv_2) + (dv_1, dv_2) \quad \text{(by def'n of } + \text{ in } \mathbb{R}^2\text{)} \\
&= c(v_1, v_2) + d(v_1, v_2) \quad \text{(by def'n of scalar multiplication)} \\
&= c\mathbf{v} + d\mathbf{v}.
\end{aligned}$$

9. *Scalar multiplication is associative:*

$$\begin{aligned}
(cd)\mathbf{v} &= (cd)(v_1, v_2) \\
&= ((cd)v_1, (cd)v_2) \quad \text{(by def'n of scalar multiplication)} \\
&= (c(dv_1), c(dv_2)) \quad \text{(by associativity of } \cdot \text{ in } \mathbb{R}\text{)} \\
&= c(dv_1, dv_2) \quad \text{(by def'n of scalar multiplication)} \\
&= c(d(v_1, v_2)) \quad \text{(by def'n of scalar multiplication)} \\
&= c(d\mathbf{v}).
\end{aligned}$$

10. *Identity element for scalar multiplication:*

$$\begin{aligned}
1\mathbf{v} &= 1(v_1, v_2) \\
&= (1v_1, 1v_2) \quad \text{(by def'n of scalar multiplication)} \\
&= (v_1, v_2) \quad \text{(since } 1 \text{ is mult. identity in } \mathbb{R}\text{)} \\
&= \mathbf{v}.
\end{aligned}$$

Since all the laws hold, $\mathbb{R}^2$ is indeed a vector space over $\mathbb{R}$. $\square$

Note: Actually checking the vector space laws is extremely tedious.

Logical technicality: To verify the vector space laws, you need to describe a general rule for finding additive identity elements and additive inverses.

This example generalizes:

Theorem 1.4 *For any $n \geq 1$, define $\mathbb{R}^n$ to be the set of $n-$tuples of real numbers with addition and scalar multiplication defined coordinate-wise. For each n, $\mathbb{R}^n$ is a vector space (and these are the most important examples of vector spaces).*

Example: Let $\mathbf{v} = (2, -1, 3, 0)$ and $\mathbf{w} = (-3, -2, 0, 2)$.

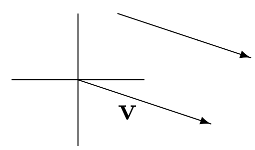
(a) To what vector space should one assume $\mathbf{v}$ and $\mathbf{w}$ belong?

(b) Compute $4\mathbf{v}$.

(c) Compute $\mathbf{v} + \mathbf{w}$.

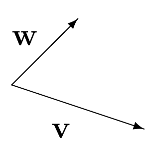
(d) Compute $2\mathbf{v} - 3\mathbf{w}$.

Pictorial representations of $\mathbb{R}^n$: To get a picture of elements of $\mathbb{R}^n$ (or any vector space, really), we think of them as "arrows":



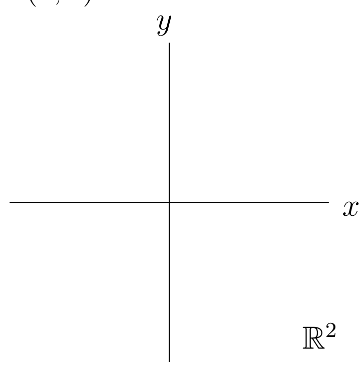
In reality, vectors in $\mathbb{R}^n$ are more like "floating" arrows, in that two vectors pointing the same direction with the same length are really the same vector (even if they are drawn in different spots).

Using the idea of vectors as "arrows", vector addition then corresponds to "head-to-tail" or "parallelogram" addition:

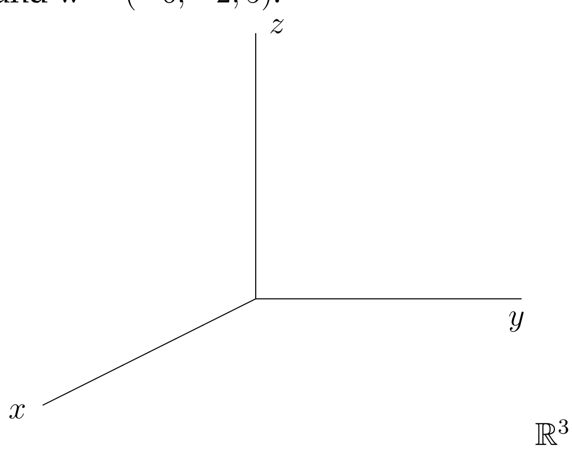


Scalar multiplication corresponds to “stretching”:

Examples:
(a) Let $\mathbf{v} = (3, -1)$ and $\mathbf{u} = (2, 1)$.



(b) Let $\mathbf{v} = (1, 2, 4)$ and $\mathbf{w} = (-6, -2, 5)$.



(c) Let $\mathbf{v} = (0, -2, 3, 1)$ and $\mathbf{w} = (2, 8, -1, 4)$.

In the definition of a vector space, we assume the 10 so-called "vector space laws". From those ten laws, it is possible to show that the following standard rules also hold in any real vector space:

Theorem 1.5 (Additional properties of real vector spaces) *Let V be a real vector space. Then:*

1. Uniqueness of additive identity: *There is only one zero vector $\mathbf{0}$ in V.*
2. Uniqueness of additive inverses: *For every $\mathbf{v} \in V$, there is exactly one vector $-\mathbf{v}$ which is the additive inverse of $\mathbf{v}$.*
3. Rule for additive inverses: *For every $\mathbf{v} \in V$, $(-1)\mathbf{v} = -\mathbf{v}$.*
4. Zero property I: *For every $\mathbf{v} \in V$, $0\mathbf{v} = \mathbf{0}$.*
5. Zero property II: *For every $c \in \mathbb{R}$, $c\mathbf{0} = \mathbf{0}$.*

1.3 Examples of vector spaces

Here is a fairly exhaustive list of vector spaces; you should familiarize yourself with this list and know the symbol(s) used to represent each space:

1. *"Traditional" vector spaces:* For any $n \in \mathbb{N}$, $\mathbb{R}^n = \{(x_1, ..., x_n) : x_j \in \mathbb{R} \, \forall \, j\}$ is a real vector space, where the addition and scalar multiplication are defined coordinate-wise. (P.S. $\forall$ means "for all"; $\exists$ means "there exists")
2. *Real numbers:* $\mathbb{R}$ is a real vector space (where the addition and scalar multiplication are the usual numerical operations). In particular, $\mathbb{R}^1 = \mathbb{R}$.
3. *Zero vector space:* $\{\mathbf{0}\}$ (the set consisting only of a zero vector) is a vector space. In particular, $\mathbb{R}^0 = \{\mathbf{0}\}$.
4. *Function spaces:* Each of the following sets is a real vector space; for all these sets of functions, the addition is described by $(f+g)(x) = f(x)+g(x)$ and the scalar multiplication is $(cf)(x) = c \cdot f(x)$. (In all cases, the additive identity element is the constant function $f(x) = 0$ and the additive inverse of f is the function $-f$.) Specific function spaces include:
   a) *Polynomials:* The set $\mathbb{R}[x]$ of polynomials whose coefficients are in $\mathbb{R}$ is a real vector space.

b) *Polynomials of bounded degree:* The set $\mathcal{P}_n$ of polynomials with real coefficients whose degree is at most n is a real vector space.
c) *Continuous functions:* The set $C(\mathbb{R}, \mathbb{R})$ of continuous functions from $\mathbb{R}$ to $\mathbb{R}$ is a real vector space.
d) *Differentiable functions:* The set of differentiable functions from $\mathbb{R}$ to $\mathbb{R}$ is a real vector space.
e) *Analytic functions:* The set C^ω of analytic functions from $\mathbb{R}$ to $\mathbb{R}$ is a real vector space (a function is *analytic* if it can be written as a power series which converges everywhere).

5. *Sequence spaces:* In these examples, the addition and scalar multiplication are defined term-by-term, i.e.

$$(x_1, x_2, x_3, \ldots) + (y_1, y_2, y_3, \ldots) = (x_1 + y_1, x_2 + y_2, x_3 + y_3, \ldots)$$
and
$$c(x_1, x_2, \ldots) = (cx_1, cx_2, \ldots).$$

a) The set $\mathbb{R}^{\mathbb{N}}$ of infinite sequences of elements of $\mathbb{R}$ is a real vector space (where the addition and scalar multiplication are done term-by-term).
b) The set $\mathbb{R}^\infty$ of infinite sequences where all but finitely many elements of the sequence are 0 also forms a real vector space. Here, some care needs to be taken to verify that addition is closed.
c) The set of convergent sequences of real numbers forms a vector space over $\mathbb{R}$.

6. *Matrix spaces:* The set of $m \times n$ matrices (this means m rows and n columns) with elements in $\mathbb{R}$, denoted $M_{mn}(\mathbb{R})$, is a real vector space where the addition and scalar multiplication are defined entry-wise. (Notation: the set of square $n \times n$ matrices with entries in $\mathbb{R}$ is denoted $M_n(\mathbb{R})$ rather than $M_{nn}(\mathbb{R})$.) Matrix spaces are discussed in more detail in Chapter 2.

Key Common Concepts

With all these examples of vector spaces,

1. You can add two things in the set, and the sum is always something in the set.
2. You can multiply things in the set by a scalar, and the product is always something in the set.

If either (1) or (2) fail, then the set is not a vector space.

Example: Suppose V and W are real vector spaces; suppose $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$ are vectors in V; suppose $\mathbf{w}_1$ and $\mathbf{w}_2$ are vectors in W; suppose all other letters represent scalars. Determine whether each of the following expressions are scalars, vectors, or if the expression is nonsense.

1. $\mathbf{v}_1 + 5c\mathbf{v}_3$

2. $c(\mathbf{v}_1 - \mathbf{v}_1)$

3. $c(d\mathbf{v}_2)$

4. $\mathbf{v}_1(\mathbf{v}_2 + c\mathbf{v}_3)$

5. $\mathbf{v}_2(c + d)$

6. $\mathbf{v}_1 + \mathbf{w}_1$

7. $\frac{1}{c}\mathbf{v}_1$

8. $0\mathbf{w}_2$

9. $c(d_1 + d_2)$

1.4 Summary of Chapter 1

1. A **(real) vector space** is a set V of objects called **vectors** which can be added to one another and multiplied by **scalars** (a scalar is a real number) in such a way that the addition and scalar multiplication obey a bunch of algebraic rules (listed earlier in this packet).

2. A vector space is the most general setting in which one can define addition and multiplication. Since linear functions are made up of addition and multiplication, this means vector spaces are the most general settings for "linear" objects.

3. Examples of vector spaces include:

   - "traditional vector spaces" $\mathbb{R}^n$, where the vectors are ordered $n-$tuples of real numbers;
   - "function spaces", where the vectors are functions;
   - "sequence spaces", where the vectors are sequences;
   - "matrix spaces", where the vectors are matrices.

   Each of these vector spaces has their own notion of addition and scalar multiplication. You cannot add vectors from two different vector spaces.

4. A good pictorial representation of a vector is a "floating arrow". In this setting, scalar multiplication corresponds to stretching and addition corresponds to "head-to-tail" addition.

5. The whole point of linear algebra is to define what **linear** means in a general sense. We are interested in doing this because linear things are relatively easy to work with, and because you can approximate any real-world problem with a linear problem.

6. Eventually we will get around to studying systems of linear equations. So far, we have looked only at systems of one linear equation in one variable. Such an equation can always be rewritten as

   $$Ax = b.$$

   - If $A \neq 0$, there is exactly one solution of this equation, namely $x = A^{-1}b$.
   - If $A = 0$ and $b = 0$, then every x is a solution of this equation (there are infinitely many solutions).
   - If $A = 0$ and $b \neq 0$, the equation has no solution.

Chapter 2

Matrices

2.1 Vocabulary associated to matrices

In the last chapter, we discussed several examples of vector spaces (sets of ordered $n-$tuples, the zero vector space, spaces of functions, etc.). Now, we introduce a last class of a vector spaces, whose importance will become apparent later.

Definition 2.1 *Given positive integers m and n, an $m \times n$* **matrix** *with entries in $\mathbb{R}$ is an array of numbers $a_{ij} \in \mathbb{R}$ where $1 \leq i \leq m$ and $1 \leq j \leq n$. We denote matrices by capital letters (usually); a matrix with entries a_{ij} is usually denoted A. We arrange the entries of the matrix in a rectangle as follows:*

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & \cdots & a_{mn} \end{pmatrix}$$

The set of matrices of size $m \times n$ with entries in $\mathbb{R}$ is denoted $M_{mn}(\mathbb{R})$. Two matrices are **equal** *if they are the same size and if all their entries coincide, i.e. $A = B$ if they are both $m \times n$ and if $a_{ij} = b_{ij}$ for all i, j.*

In particular, a_{ij} is the entry of A in the i^{th} row and j^{th} column. m is the number of rows of A; n is the number of columns of A.

Definition 2.2 *A matrix is called* **square** *if it has the same number of rows as columns. The set of square $n \times n$ matrices with entries in $\mathbb{R}$ is denoted $M_n(\mathbb{R})$ (as opposed to $M_{nn}(\mathbb{R})$).*

Example:

$$A = \begin{pmatrix} 1 & 6 \\ -4 & 5 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 5 & 7 \end{pmatrix}$$

Definition 2.3 *Given a matrix A, the* **diagonal entries** *of A are the numbers a_{11}, a_{22}, a_{33}, ...a_{nn} (this list may stop at a_{mm} depending on when you run out of entries).*

Example:

$$A = \begin{pmatrix} 2 & 7 \\ 5 & 1 \\ 4 & -2 \end{pmatrix}$$

Definition 2.4 *A matrix A is called* **diagonal** *if it is square and all of its nondiagonal entries are zero.*

Example:

$$A = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

Definition 2.5 *The* **trace** *of a matrix, denoted $tr(A)$, is the sum of the diagonal entries of that matrix.*

Example:

$$A = \begin{pmatrix} 7 & -2 & 4 \\ 1 & -2 & 5 \\ 0 & 1 & -1 \end{pmatrix}$$

Definition 2.6 *A square matrix is called* **upper triangular** *(abbreviated* **upper** Δ*) if all its entries below its diagonal are zero. A matrix is called* **lower triangular** *(abbreviated* **lower** Δ*) if all the entries above its diagonal are zero. A matrix is called* **triangular** *if it is either lower triangular or upper triangular.*

Example:

$$A = \begin{pmatrix} 2 & 3 \\ 0 & -4 \end{pmatrix} \qquad B = \begin{pmatrix} -7 & 0 & 0 \\ 3 & -2 & 0 \\ 1 & -3 & 4 \end{pmatrix}$$

Note: diagonal matrices are both upper triangular and lower triangular.

Definition 2.7 *The* **transpose** *of an* $m \times n$ *matrix* A*, denoted* A^T *or* A^t*, is the* $n \times m$ *matrix satisfying* $(a^T)_{ij} = a_{ji}$ *for all* i, j*.*

Example:

$$A = \begin{pmatrix} 1 & -4 \\ 2 & -1 \\ 3 & 0 \end{pmatrix}$$

Definition 2.8 *The* $n \times n$ **identity matrix***, denoted* I *or* I_n*, is the diagonal* $n \times n$ *matrix with all diagonal entries equal to* 1*.*

$$I = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

Definition 2.9 *Given an $m \times n$ matrix A as above, the vectors*

$$\begin{gathered}(a_{11}, a_{12}, a_{13}, ..., a_{1n}),\\(a_{21}, a_{22}, a_{23}, ..., a_{2n}),\\...,\\(a_{m1}, a_{m2}, ..., a_{mn})\end{gathered}$$

are called the **rows** *of A; note that each row of A is an element of $\mathbb{R}^n$. The vectors*

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, ..., \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

are called the **columns** *of A; note that each column of A is an element of $\mathbb{R}^m$.*

Example: For each given matrix:
- Give the size of the matrix.
- Write down the $(3, 2)-$entry of the matrix (if it exists).
- Write down the diagonal entries of the matrix.
- Write down the second row of the matrix.
- What is the trace of the matrix?
- Is the matrix square? diagonal? upper triangular? lower triangular?
- Write down the transpose of the matrix.

(a) $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$

(b) $B = \begin{pmatrix} -4 & 2 & 5 & 6 \\ 7 & 0 & -3 & 2 \\ 4 & 2 & 1 & -2 \end{pmatrix}$

(c) $C = \begin{pmatrix} -1 & 2 & 0 & -3 & 0 \\ 0 & 3 & 1 & 4 & -6 \end{pmatrix}$

- $C \in M_{25}(\mathbb{R})$, i.e. C is 2×5;
- c_{32} DNE;
- diagonal entries are -1 and 3;
- the second row of C is $(0, 3, 1, 4, -6)$ (this is an element of $\mathbb{R}^5$);
- $tr(C) = -1 + 3 = 2$;
- C is not square and hence not diagonal or triangular;
- $C^T = \begin{pmatrix} -1 & 0 \\ 2 & 3 \\ 0 & 1 \\ -3 & 4 \\ 0 & -6 \end{pmatrix}$.

(d) $D = \begin{pmatrix} 3 & 0 \\ 1 & -5 \end{pmatrix}$

2.2 Matrix operations

The first thing we want to do is show that for each m and n, the set $M_{mn}(\mathbb{R})$ of $m \times n$ matrices forms a vector space. To do this, we need to define what it means to add two matrices, and what it means to multiply a matrix by a scalar. Essentially, *addition and scalar multiplication of matrices are performed entry-by-entry*. More precisely,

Definition 2.10 *Given two matrices $A, B \in M_{mn}(\mathbb{R})$ and a scalar $r \in \mathbb{R}$, we define the matrix $A + B \in M_{mn}(\mathbb{R})$ by*

$$A + B = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} & \cdots & a_{2n}+b_{2n} \\ a_{31}+b_{31} & a_{32}+b_{32} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & \cdots & a_{mn}+b_{mn} \end{pmatrix}$$

(equivalently we set $(a + b)_{ij} = a_{ij} + b_{ij}$ for all i, j); and we define the matrix $rA \in M_{mn}(\mathbb{R})$ by

$$rA = \begin{pmatrix} ra_{11} & ra_{12} & ra_{13} & \cdots & ra_{1n} \\ ra_{21} & ra_{22} & ra_{23} & \cdots & ra_{2n} \\ ra_{31} & ra_{32} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ ra_{m1} & ra_{m2} & \cdots & \cdots & ra_{mn} \end{pmatrix}$$

(equivalently we set $(ra)_{ij} = r(a_{ij})$ for all i, j).

Note: You can only add two matrices of the same size.

Example: Let

$$A = \begin{pmatrix} 4 & 1 & -1 \\ 2 & 0 & -3 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & -1 \end{pmatrix}$$

$$C = \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}$$

Theorem 2.11 *The operations defined above make each $M_{mn}(\mathbb{R})$ into a real vector space; the additive identity element of $M_{mn}(R)$ is the $m \times n$* **zero matrix**

$$\mathbf{0} = \mathbf{0}_{m\times n} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 \end{pmatrix}.$$

Associating vectors in $\mathbb{R}^n$ to column vectors: We associate vectors in $\mathbb{R}^n$ to $n \times 1$ matrices as follows:

$$\mathbf{x} = (x_1, x_2, ..., x_n) \in \mathbb{R}^n \qquad \longleftrightarrow \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in M_{n1}(\mathbb{R}).$$

In particular, an $n \times 1$ matrix is also called a **column vector**. A column vector with n entries is the same thing as a vector in $\mathbb{R}^n$. Given a vector $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$, if we want to think of that vector as a row vector, we take the transpose of $\mathbf{x}$:

$$\mathbf{x}^T = \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \end{pmatrix} \in M_{1n}(\mathbb{R}).$$

2.3 Matrix multiplication

There is another operation one can perform on matrices, which doesn't directly have anything to do with thinking of matrices as vectors. The importance of this operation will be seen later in the course; for now we simply define it.

Definition 2.12 *Given matrices $A \in M_{mn}(F)$ and $B \in M_{pq}(F)$, if $n = p$ then we can define the* **product** *AB, which is an $m \times q$ matrix AB defined entrywise by setting*

$$(ab)_{ij} = \sum_{k=1}^{n(=p)} a_{ik} b_{kj}.$$

(If $n \neq p$, AB is undefined.)

Note:

- If A is a square matrix, we write A^2 for AA, A^3 for AAA, etc.
- If A isn't square, then A^2 is undefined.
- In general matrix multiplication is **not commutative**: $AB \neq BA$ most of the time, even if both products are defined.

Example: Let

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 1 & -3 \end{pmatrix}; \quad B = \begin{pmatrix} -2 & 1 \\ 1 & 3 \end{pmatrix}; \quad C = \begin{pmatrix} 5 & -1 \\ 1 & 2 \\ 0 & 1 \end{pmatrix}.$$

Compute each of the following quantities, or state that they are undefined (with justification).

(a) AB

(b) BA

(c) A^2

(d) B^2

(e) CB^2A

Theorem 2.13 (Properties of elementary matrix operations) *Let A, B, C be matrices with entries in $\mathbb{R}$, let I be the identity matrix of the appropriate size and let $k \in \mathbb{R}$. Then,* so long as everything is defined*, we have:*

1. $IA = A$ *and* $BI = B$
2. $A(BC) = (AB)C$
3. $k(AB) = (kA)B = A(kB)$
4. $A(B + C) = AB + AC$
5. $(A + B)C = AC + BC$
6. $(A^T)^T = A$
7. $tr(A^T) = tr(A)$
8. $(rA)^T = rA^T$
9. $(A + B)^T = A^T + B^T$
10. $(AC)^T = C^T A^T$
11. $tr(A + B) = tr(A) + tr(B)$
12. $tr(AB) = tr(BA)$

Example: Suppose $A \in M_{mn}(\mathbb{R})$, $\mathbf{b} \in \mathbb{R}^n$, $C \in M_{nn}(\mathbb{R})$, $D \in M_{nm}(R)$ (assume in this problem that neither m nor n are 1, and assume $m \neq n$). For each of the following expressions, determine if the expression is a matrix (in which you should give its size), a vector in some $\mathbb{R}^k$ (in which case you should give the value of k), a scalar, or nonsense.

(a) $A\mathbf{b}$

(b) AC^2

(c) $CA\mathbf{b}$

(d) $\mathbf{b}CI$

(e) $\mathbf{b}\mathbf{b}^T$

(f) $\mathbf{b}^T\mathbf{b}$

(g) $\mathbf{b}^T IC$

(h) $\mathbf{b}^T C\mathbf{b}$

(i) $(\mathbf{b}^T\mathbf{b})D$

(j) $DACD$

(k) $DAC^2\mathbf{b}$

(l) $\mathbf{b}^T DA$

(m) $(AC)^T$

2.4 *Mathematica* and calculator commands for matrix operations

Matrix operations for large-sized matrices can and should be done electronically using either a computer software package or calculator. In this section I give directions for matrix operations using the computer software package *Mathematica* and I give directions for matrix operations on a TI-83/84 series graphics calculator. For other calculators, consult your user manual or the internet, or ask your instructor or classmate.

Matrix operations on *Mathematica*

Defining a matrix: To store a matrix as a variable, there are two methods.

1. Type the matrix in using braces very carefully. For example, to save the matrix
$$\begin{pmatrix} 2 & 4 & 7 \\ -5 & 3 & 1 \end{pmatrix}$$
as A, execute

   `A = {{2,4,7}, {-5,3,1}}`

   Note that the entries are separated by commas, every row of the matrix needs braces around it, and the entire matrix needs braces around it.

   To type in a column vector, you need only one set of braces, so if you execute `b = {1,2,3}` this defines the column vector $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

2. Use the Basic Math Assistant Pallette. Click "Pallettes" and "Basic Math Assistant", then on the Basic Math Assistant click the fourth tab under "Basic Commands" that looks like a matrix. In the *Mathematica* notebook, type `A=`, then click the large button that looks like a matrix, then click "AddRow" or "Add Column" until the matrix is the appropriate size. Click in each box of the matrix and type in the appropriate numbers. For example, your command to define the A above would look like

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

Matrix multiplication: To multiply two matrices in *Mathematica*, you need a period between the matrices. For example, after defining matrices A and B, you can compute the matrix product by

`A.B`

The output you will get won't look like a matrix; to make it look like a matrix you can type

`A.B //MatrixForm`

For matrix powers, you will need to type `A.A` rather than `A^2`. For a larger matrix power (say A^{100}), run the following: `MatrixPower[A,100] //MatrixForm`

Other matrix operations: Once you have saved a matrix as a letter or string, you can perform standard operations on it as follows (add `//MatrixForm` to the end of the command to make the output look like a matrix):

1. For the transpose of A, execute `Transpose[A]`.
2. For the trace of A, execute `Tr[A]`.
3. To multiply A by a scalar (say 5), execute `5 A`.
4. To add two matrices (say A and B), execute `A + B`.
5. To get the i, j entry of a matrix, use double braces: execute `A[[2,3]]` (to get the $2, 3-$entry).
6. To call the $n \times n$ identity matrix, use a command like `IdentityMatrix[4]` (this generates the 4×4 identity matrix).

Matrix operations on TI-83/84-type calculators

The main button you need to find is the [MATRX] button, which may be above the [SIN] key or the second function on the [x^{-1}] button depending on your particular model.

Defining a matrix: Suppose you want to enter the matrix

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

into your calculator. There are two ways to do this:

1. On your normal screen, type `[[2,4,7][-5,3,1]]`, then hit [STO→], [MATRX] and choose a name for the matrix from the NAME menu. Note that there are no commas between the rows.

2. Alternatively, you can hit [MATRX]. You will see a menu where across the top of the screen the calculator says NAMES, MATH, EDIT. Use the right arrow to highlight EDIT, and then if necessary move the cursor up or down until you get to the letter you want to save your matrix as. Type in the size of the matrix (in this case it is 2×3), and then type the entries in one by one. Once you get done, hit [QUIT] to return to the home screen.

Matrix multiplication: To multiply two matrices (say A times B) with a TI-83/84 calculator, first save the two matrices and then return to the home screen. Now, hit [MATRX], highlight NAMES, and go down to `[A]` and hit [ENTER]. Now hit [MATRX], highlight names, and go down to `[B]` and hit [ENTER].

Your home screen should now look like `[A][B]`; now hit [ENTER] and the matrix product will be displayed.

Other matrix operations: Assuming you have saved the matrices you want to work with:

1. For the transpose of A, type [MATRX], highlight NAMES, go down to `[A]` and hit [ENTER]. Then hit [MATRX], highlight MATH and choose ^{T}. Hit [ENTER]; the home screen will look like `[A]`^{T}. Hit [ENTER] and the transpose will be displayed.
2. To multiply A by a scalar (say 5), hit 5 and then hit [MATRX], highlight NAMES, and go down to `[A]` and hit [ENTER]. Your home screen will look like `5[A]`; now hit [ENTER] and the product will be displayed.
3. To add two matrices (say A and B), use the same commands as for matrix multiplication, but put a $+$ in between the matrices.

Chapter 3

Subspaces

3.1 Introducing subspaces

Recall: We have seen several examples of real vector spaces so far:

- The zero vector space $\{\mathbf{0}\}$;
- $\mathbb{R}^n$, the space of "traditional" vectors;
- spaces of functions;
- spaces of sequences; and
- spaces of matrices.

Some of these vector spaces are subsets of one another. For example, a theorem from Calculus I says that functions which are differentiable must be continuous. Restated, this says that the set of differentiable functions from $\mathbb{R}$ to $\mathbb{R}$ is a subset of the space of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$.

When one vector space is a subset of another, we say that the first space is a "subspace" of the second. More precisely:

> **Definition 3.1** *Let V be a real vector space, and let $W \subseteq V$ be a subset of V. W is called a* **subspace** *of V if W is itself a real vector space under the operations of addition and scalar multiplication that define V.*

The only way W would not be a real vector space is if it is not closed under addition or scalar multiplication, so this definition can be restated as follows:

Definition 3.2 (equivalent to Definition 3.1) *Let V be a real vector space, and let $W \subseteq V$ be a subset of V. W is called a* **subspace** *of V if:*

1. *W* is nonempty, *i.e. there exists some $\mathbf{w} \in W$;*
2. *W* is closed under addition, *i.e. if $\mathbf{w}_1 \in W$ and $\mathbf{w}_2 \in W$, then $\mathbf{w}_1 + \mathbf{w}_2 \in W$.*
3. *W* is closed under scalar multiplication, *i.e. if $\mathbf{w}_1 \in W$ and $c \in \mathbb{R}$, then $c\mathbf{w}_1 \in W$.*

It turns out (eventually) that a deep understanding of subspaces is key to solving problems (especially solving systems of linear equations).

FIRST QUESTION: Which subsets of a vector space V are subspaces?

Theorem 3.3 *Let V be a real vector space, and let $W \subseteq V$ be a subspace. Then the zero vector $\mathbf{0}$ must be in W.*

PROOF Let W be a subspace of V. Then $W \neq \emptyset$, so there is some $\mathbf{w} \in W$. Since W is closed under scalar multiplication,

$$0\mathbf{w} = \mathbf{0} \in W$$

as desired. $\square$

Definition 3.2 and Theorem 3.3 give you a good mechanism to decide whether a subset of a vector space is a subspace. If you are given vector space V and subset $W \subseteq V$, to decide whether W is a subspace, ask the following questions (this is called the *brute-force* method for determining whether or not W is a subspace:

1. Is $\mathbf{0} \in W$? If not, W is not a subspace.

   If so, proceed to $\#$ 2.

2. Is W closed under $+$?

   If not, W is not a subspace.

   If so, proceed to $\#$ 3.

3. Is W closed under scalar multiplication?

   If not, W is not a subspace.

   If so, W is a subspace.

A "no" answer to any of these questions means W is not a subspace. A "yes" answer to all three questions means W is a subspace.

Directions: In these examples, you are given a real vector space V and a subset W of V. Determine, with justification, whether or not W is a subspace of V.

Example 1: $V =$ any vector space. $W = \{\mathbf{0}\}$.

INTUITION	RIGOROUS PROOF

Example 2: $V = M_2(\mathbb{R})$. $W =$ the set of diagonal 2×2 matrices.

INTUITION	RIGOROUS PROOF

Example 3: $V = \mathbb{R}^2$. $W =$ the set of vectors (x, y) such that $y = 3$.

INTUITION	RIGOROUS PROOF
• Is $\mathbf{0} \in W$?	

Example 4: $V = \mathbb{R}^2$. $W =$ the $x-$axis.

INTUITION	RIGOROUS PROOF
• Is $\mathbf{0} \in W$?	
• W closed under $+$?	
• W closed under $\cdot$?	

Example 5: $V = \mathbb{R}^2$. $W =$ the upper half-plane $= \{(x, y) : y \geq 0\}$

INTUITION	RIGOROUS PROOF
• Is $\mathbf{0} \in W$?	
• W closed under $+$?	
• W closed under $\cdot$?	

Example 6: $V = C(\mathbb{R}, \mathbb{R})$. $W =$ the set of differentiable functions from $\mathbb{R}$ to $\mathbb{R}$.

INTUITION	RIGOROUS PROOF
• Is $\mathbf{0} \in W$?	
• W closed under $+$?	
• W closed under $\cdot$?	

Example 7: $V = C(\mathbb{R}, \mathbb{R})$. $W =$ the set of functions passing through $(5, 0)$.

INTUITION	RIGOROUS PROOF
• Is $\mathbf{0} \in W$?	
• W closed under $+$?	
• W closed under $\cdot$?	

Example 8: $V = C(\mathbb{R}, \mathbb{R})$. $W =$ the set of functions passing through $(5, 2)$.

INTUITION	RIGOROUS PROOF
• Is $\mathbf{0} \in W$?	

Example 8: $V =$ any vector space. $W = V$.

INTUITION	RIGOROUS PROOF
• Is $\mathbf{0} \in W$?	
• W closed under $+$?	
• W closed under $\cdot$?	

How to write a proof that a subset W is a subspace of V:

You need to do **all** of these things:

1. Verify that the zero vector $\mathbf{0}$ belongs to W.
2. Take two generic elements of W (see worksheet) and add them. Verify that the sum belongs to W.
3. Take a generic element of W and multiply it by a generic constant (like r). Verify that this product belongs to W.

How to write a proof that a subset W is NOT a subspace of V:

You need to do **one** of the following three things:

1. Explain why $\mathbf{0}$ does not belong to W.
2. Alternatively, write down two **specific** elements of W (i.e. with numbers) whose sum is not in W.
3. Alternatively, write down a **specific** element of W (i.e. with numbers) and a **specific** scalar (i.e. a number) such that when you multiply that scalar by that element, the product is not in W.

Performing either (2) or (3) in the second box above is called *finding an explicit counterexample*.

Now, suppose you are given some W and asked whether or not it is a subspace. How do you know which of the two above procedures you are supposed to be carrying out?

Some intuition: In order for a subset W to be a subspace of vector space V, W should ...

1. **contain the zero vector**,
2. **be unbounded**, i.e. extend "forever" in any direction it goes (otherwise it won't be closed under scalar multiplication);
3. **be flat/straight** (otherwise it won't be closed under scalar multiplication);
4. **and be convex**, i.e. for any two points in W, the line segment connecting the points should stay entirely within W (otherwise, it won't be closed under addition).

If W is these four things, it is probably a subspace. That said, these intuitive concepts are **NOT** substitutes for a proof.

3.2 Span

Here is an important example of a class of subspaces:

Definition 3.4 *Let V be a real vector space and let $\mathbf{v} \in V$ be any vector. Define $W = \{c\mathbf{v} : c \in \mathbb{R}\}$ to be the set of scalar multiples of $\mathbf{v}$. W is called the* **span** *of $\mathbf{v}$ and denoted $Span(\mathbf{v})$ or $< \mathbf{v} >$.*

General pictorial representation of a span:

Example: $V = \mathbb{R}^2$; $\mathbf{v} = (3, 1)$.

Example: $V = \mathbb{R}^3$; $\mathbf{v} = (0, 1, 0)$.

Example: $V = C(\mathbb{R}, \mathbb{R})$, $f(x) = \sin x$.

Theorem 3.5 *Let V be a real vector space. Then the span of any vector is a subspace of V.*

PROOF We verify the three essential characteristics of being a subspace:

- $\mathbf{0} = 0\,\mathbf{v} \in Span(\mathbf{v})$
- Suppose $\mathbf{w}_1, \mathbf{w}_2 \in Span(\mathbf{v})$. That means $\mathbf{w}_1 = c_1\mathbf{v}$, $\mathbf{w}_2 = c_2\mathbf{v}$. Then

$$\mathbf{w}_1 + \mathbf{w}_2 = c_1\mathbf{v} + c_2\mathbf{v} = (c_1 + c_2)\mathbf{v} \in Span(\mathbf{v}).$$

- Suppose $\mathbf{w} \in Span(\mathbf{v})$ and $r \in \mathbb{R}$. That means $\mathbf{w} = c\mathbf{v}$ so

$$r\mathbf{w} = r(c\mathbf{v}) = (rc)\mathbf{v} \in Span(\mathbf{v}). \square$$

Definition 3.6 *Let V be a real vector space and let $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n \in V$ be any collection of vectors. Define the* **span** *of these vectors to be the set $W \subseteq V$, denoted $Span(\mathbf{v}_1, ..., \mathbf{v}_n)$ or $< \mathbf{v}_1, ..., \mathbf{v}_n >$, by*

$$W = Span(\mathbf{v}_1, ..., \mathbf{v}_n) = \{c_1\mathbf{v}_1 + ... + c_n\mathbf{v}_n : c_j \in \mathbb{R}\,\forall j\}$$

(this the set of linear combinations of the $\mathbf{v}_j$).

Theorem 3.7 (Spans are subspaces) *Let V be a real vector space. Then the span of any collection of vectors in V is a subspace of V.*

PROOF Similar to above (HW).
Hint: Two generic elements of $Span(\mathbf{v}_1, ..., \mathbf{v}_n)$ can be written down as $\sum_{j=1}^{n} c_j\mathbf{v}_j$ and $\sum_{j=1}^{n} d_j\mathbf{v}_j$, where $c_j, d_j \in \mathbb{R}$ for all j.

The word "span" can also be used as a verb:

Definition 3.8 *Let V be a real vector space (or subspace) and let $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n \in V$ be a collection of vectors. We say the set $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$* **spans** *$V$ if $V = Span(\mathbf{v}_1, ..., \mathbf{v}_n)$.*

Example: $V = \mathbb{R}^3$; $\mathbf{v}_1 = (1, 0, 0), \mathbf{v}_2 = (0, 1, 0)$

Example: $V = \mathbb{R}^3$; $\mathbf{v}_1 = (1, 2, -1), \mathbf{v}_2 = (-2, 1, 3)$

Example: $V = \mathbb{R}^3$; $\mathbf{v}_1 = (3, 1, 1)$, $\mathbf{v}_2 = (6, 2, 2)$

3.3 Distinguishing between subspaces

Questions:

1. In a general sense, what kinds of sets are subspaces?
2. How do you distinguish between vector spaces?
3. How do you distinguish between different subspaces of a vector space?

First example: What is the difference between $\mathbb{R}^2$ and $\mathbb{R}^3$?

Second example: Let $V = \mathbb{R}^2$ and let $W_1 = Span(1, 0)$; $W_2 = Span(1, 1)$; $W_3 = Span(-4, 0)$.

So what does "going in the same direction" mean, in general?

Definition 3.9 *Let V be a real vector space. Two vectors $\mathbf{v}$ and $\mathbf{w}$ in V are called* **parallel** *(denoted $\mathbf{v}||\mathbf{w}$) if one is a scalar multiple of another, i.e.*

$$\mathbf{v}||\mathbf{w} \iff (\exists\, c \in \mathbb{R} \text{ s.t. } c\mathbf{v} = \mathbf{w} \text{ or } \exists\, c \in \mathbb{R} \text{ s.t. } c\mathbf{w} = \mathbf{v}).$$

Two vectors $\mathbf{v}$ and $\mathbf{w}$ are **in the same direction** *if the c in the above definition can be taken to be greater than or equal to zero.*

Note: "s.t." means "such that".

Example: $V = \mathbb{R}^2$; $\mathbf{v} = (2, -3)$; $\mathbf{w} = (-4, 6)$.

Example: $V = \mathbb{R}^2$; $\mathbf{v} = (0, 0)$; $\mathbf{w} = (1, 2)$.

Example: $V = M_2(\mathbb{R})$; $A = \begin{pmatrix} 2 & -3 \\ 0 & 1 \end{pmatrix}$; $B = \begin{pmatrix} 8 & -12 \\ 0 & 4 \end{pmatrix}$.

The following characterization of parallelism is important because it will generalize to collections of more than two vectors.

Theorem 3.10 *Let $\mathbf{v}, \mathbf{w} \in V$, where V is a real vector space. Then $\mathbf{v}||\mathbf{w}$ if and only if there exist scalars $c_1, c_2 \in \mathbb{R}$ with c_1, c_2 not both zero, s.t. $c_1\mathbf{v} + c_2\mathbf{w} = \mathbf{0}$.*

PROOF ($\Rightarrow$) Assume $\mathbf{v}||\mathbf{w}$. Then

$$\begin{aligned} &\mathbf{v} = c\mathbf{w} \text{ for some scalar } c \\ \Rightarrow &\mathbf{v} - c\mathbf{w} = \mathbf{0} \\ \Rightarrow &1\mathbf{v} + (-c)\mathbf{w} = \mathbf{0} \\ \Rightarrow &c_1\mathbf{v} + c_2\mathbf{w} = \mathbf{0} \text{ (where } c_1 = 1,\ c_2 = -c). \end{aligned}$$

($\Leftarrow$) Suppose $c_1\mathbf{v} + c_2\mathbf{w} = \mathbf{0}$ where either c_1 or c_2 is nonzero. If $c_1 \neq 0$, divide through by c_1 to get

$$\mathbf{v} + \frac{c_2}{c_1}\mathbf{w} = \mathbf{0} \Rightarrow \mathbf{v} = \frac{-c_2}{c_1}\mathbf{w} \Rightarrow \mathbf{v}||\mathbf{w}.$$

If $c_2 \neq 0$, divide through by c_2 to get

$$\frac{c_1}{c_2}\mathbf{v} + \mathbf{w} = \mathbf{0} \Rightarrow \mathbf{w} = \frac{-c_1}{c_2}\mathbf{v} \Rightarrow \mathbf{v}||\mathbf{w}.$$

Either way, $\mathbf{v}||\mathbf{w}$ as desired. □

Example: $V = \mathbb{R}^2$; $\mathbf{v} = (1, 0)$; $\mathbf{w} = (-4, 0)$.

3.4 Affine subspaces

Definition 3.11 *Let V be a real vector space. An* **affine subspace** *A of V is a translate of a subspace, i.e. A is an affine subspace of V if there exists a vector $\mathbf{p} \in V$ and a subspace $W \subseteq V$ such that*

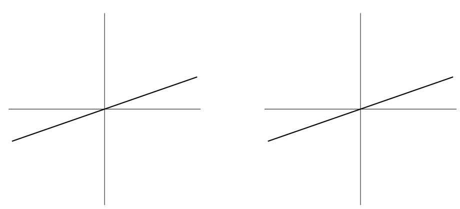
$$A = \mathbf{p} + W = \{\mathbf{p} + \mathbf{w} : \mathbf{w} \in W\}.$$

In this setting, the vector $\mathbf{p}$ is called a **translation vector** *for A, and the subspace W is called the* **associated subspace** *of A.*

First examples:

- Given real vector space V, V is a subspace of itself (hence an affine subspace of itself).
- Given real vector space V, any set $\{\mathbf{v}\}$ consisting of exactly one vector $\mathbf{v} \in V$ constitutes an affine subspace called a *point* (since any such set is the translation of the subspace $\{\mathbf{0}\}$.

Note: "Affine" is not an adjective which describes some subspaces. In fact, subspaces are special kinds of affine subspaces, not the other way around:



.

Definition 3.12 *Let V be a real vector space. A* **line** *Λ in V is an affine subspace whose associated subspace is the span of a single nonzero vector $\mathbf{v} \in V$. $\mathbf{v}$ is called a* **direction vector** *for the line.*

> **Definition 3.13** *Let Λ be a line in vector space V. A set of* **parametric equations** *for Λ is a coordinate-wise version of the equation $\mathbf{x} = \mathbf{p} + t\mathbf{v}$, where $\mathbf{p} \in \Lambda$ and $\mathbf{v}$ is a direction vector of Λ.*

Note: parametric equations for a line (or other affine subspace) are **NEVER** unique, because they depend on a choice of $\mathbf{p}$ and $\mathbf{v}$.

Ex: Find parametric equations for the line which passes through $(4, -1, 3)$ and $(2, 5, -1)$ in $\mathbb{R}^3$.

Parametric equations for a line "lay a $t-$axis" on that line, i.e. **coordinatize** the line where

$$t = 0 \leftrightarrow \text{ choice of translation vector } \mathbf{p}$$

$$\text{one unit of } t \leftrightarrow \text{ choice of direction vector } \mathbf{v}$$

Definition 3.14 *Let V be a real vector space. A* **plane** Π *in V is an affine subspace whose associated subspace is the span of two nonparallel vectors $\mathbf{v}, \mathbf{w} \in V$.*

Definition 3.15 *Let Π be a plane in vector space V. A set of* **parametric equations** *for Π is a coordinate-wise version of the equation $\mathbf{x} = \mathbf{p} + s\mathbf{v} + t\mathbf{w}$, where $\mathbf{p} \in \Lambda$ and $\mathbf{v}$ and $\mathbf{w}$ span the subspace associated to Π.*

Example: Find parametric equations of the plane in $\mathbb{R}^3$ containing the point $(3, 2, -7)$ and whose associated subspace is spanned by $(1, 5, -1)$ and $(2, 0, -3)$.

Example: Find parametric equations of the plane in $\mathbb{R}^3$ passing through $(1, 1, 4)$, $(2, 3, 1)$ and $(-2, 4, 4)$.

Parametric equations for a plane Π "lay an $s, t-$plane" on Π where

(s, t) origin $\leftrightarrow$ choice of translation vector $\mathbf{p}$
$s -$ axis $\leftrightarrow$ choice of $\mathbf{v}$
$t -$ axis $\leftrightarrow$ choice of $\mathbf{w}$

Note: The $s-$ and $t-$axes need not be $\perp$:

NEW PROBLEM: **Can we completely list all subspaces and affine subspaces of a vector space?**

In most cases, yes; for now, we do this when the vector space is $\mathbb{R}$, $\mathbb{R}^2$ and $\mathbb{R}^3$.

Theorem 3.16 *The only subspaces of $\mathbb{R}$ are the zero subspace $\{\mathbf{0}\}$ and the entire real number line $\mathbb{R}$.*

PROOF First, both $\{0\}$ and $\mathbb{R}$ are obviously subspaces of $\mathbb{R}$. So we need to show that there aren't any other subspaces of $\mathbb{R}$.

Let $W \subseteq \mathbb{R}$ be a subspace. There are two cases:

Case 1: $W = \{0\}$.

Case 2: $W \neq \{0\}$. In this situation, there is a $w \neq 0$ s.t. $w \in W$. Then $rw \in W$ for all $r \in \mathbb{R}$, so all real numbers are in W. Thus $W = \mathbb{R}$. □

Corollary 3.17 *The only affine subspaces of $\mathbb{R}$ are points and the entire real number line $\mathbb{R}$.*

PROOF Let A be an affine subspace of $\mathbb{R}$. Then $A = p + W$ where W is a subspace of $\mathbb{R}$ and $p \in \mathbb{R}$. By the previous theorem, there are two cases:

Case 1: $W = \{0\}$. Then $A = \{p\}$, a point.

Case 2: $W = \mathbb{R}$. Then $A = p + \mathbb{R} = \mathbb{R}$. □

Theorem 3.18 *The only subspaces of $\mathbb{R}^2$ are the zero subspace $\{\mathbf{0}\}$, lines passing through the origin, and all of $\mathbb{R}^2$. Consequently, the only affine subspaces of $\mathbb{R}^2$ are points, lines, and $\mathbb{R}^2$.*

PROOF It is clear that points, lines and planes are affine subspaces. What is left to prove is that there aren't any other affine subspaces; this will be done later. □

Corollary 3.19 *The only subspaces of $\mathbb{R}^3$ are the zero subspace $\{\mathbf{0}\}$, lines passing through the origin, planes passing through the origin, and all of $\mathbb{R}^3$. Consequently, the only affine subspaces of $\mathbb{R}^3$ are points, lines, planes and $\mathbb{R}^3$ itself.*

PROOF It is clear that points, lines and planes are affine subspaces. What is left to prove is that there aren't any other affine subspaces; this will be done later. □

Generic pictures of subspaces:

These pictures reflect the idea that subspaces contain the zero vector and are flat, convex, and unbounded.

Generally speaking, when you are told "W is a subspace of V", think of W as being something like a line or plane passing through the origin.

3.5 Linear independence

Motivating example: Let $V = \mathbb{R}^n$; take two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and let $W = Span(\mathbf{v}, \mathbf{w})$. What kind of object is W?

Classifying these cases depends on whether or not $\mathbf{v}$ and $\mathbf{w}$ are parallel.

Question: What is the right notion of "parallelism" for a family of $n > 2$ vectors $\mathbf{v}_1, ..., \mathbf{v}_n$?

Recall: Theorem 3.10 from earlier gives an equivalent characterization of parallelism of two vectors:

$$\mathbf{v}_1 || \mathbf{v}_2 \iff \exists c_1, c_2 \in \mathbb{R} \text{ (not both zero) s.t. } c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}.$$

Ex: $(2, 1)||(4, 2)$ since

The equation $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}$ generalizes naturally to sets of more than two vectors:

Definition 3.20 *Let V be a real vector space. A collection of vectors $\mathbf{v}_1, ..., \mathbf{v}_n \in V$ is called* **linearly dependent (lin. dep.)** *if $\exists\, c_1, ..., c_n \in \mathbb{R}$ with not all $c_j = 0$ such that $c_1\mathbf{v}_1 + ... + c_n\mathbf{v}_n = \mathbf{0}$.*

Here is an equivalent formulation of the same idea:

Definition 3.21 *Let V be a real vector space, and let $\mathbf{v}_1, ..., \mathbf{v}_n \in V$. Any expression of the form $c_1\mathbf{v}_1 + ... + c_n\mathbf{v}_n$ is called a* **linear combination** *of the $\mathbf{v}_j$. The expression $0\mathbf{v}_1 + ... + 0\mathbf{v}_n$ is called the* **trivial combination** *(this combination always equals $\mathbf{0}$) and is called a* **nontrivial combination** *of the $\mathbf{v}_j$ otherwise (i.e. if not all the c_j are zero).*

Given this definition, we can say that a set of vectors is linearly dependent if and only if they have a nontrivial combination which makes $\mathbf{0}$.

Heuristic: A lin. dep. set of vectors "repeats the same direction unnecessarily", in the same way that two parallel vectors do.

Theorem 3.22 *Let V be a real vector space. A collection of vectors $\mathbf{v}_1, ..., \mathbf{v}_n \in V$ is linearly dependent if and only if for some k, we can write $\mathbf{v}_k$ as a linear combination of the previous vectors in the list, i.e.*

$$\mathbf{v}_k = \sum_{j=1}^{k-1} d_j\mathbf{v}_j = d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + ... + d_{k-1}\mathbf{v}_{k-1}.$$

PROOF ($\Rightarrow$) Assume the $\mathbf{v}_j$ are lin. dep. Then

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_n\mathbf{v}_n = \mathbf{0}$$

where the $c_1, ..., c_n$ are scalars with not all the c_j zero. Let k be the largest subscript so that $c_k \neq 0$. Then

$$\begin{aligned} & c_1\mathbf{v}_1 + ... + c_{k-1}\mathbf{v}_{k-1} + c_k\mathbf{v}_k = \mathbf{0} \\ \Rightarrow\ & c_k\mathbf{v}_k = -c_1\mathbf{v}_1 - ... - c_{k-1}\mathbf{v}_{k-1} \\ \Rightarrow\ & \mathbf{v}_k = \frac{-c_1}{c_k}\mathbf{v}_1 - ... - \frac{c_{k-1}}{c_k}\mathbf{v}_{k-1} \\ \Rightarrow\ & \mathbf{v}_k = d_1\mathbf{v}_1 + ... + d_{k-1}\mathbf{v}_{k-1}. \end{aligned}$$

($\Leftarrow$) Suppose $\mathbf{v}_k = d_1\mathbf{v}_1 + ... + d_{k-1}\mathbf{v}_{k-1}$. Then

$$d_1\mathbf{v}_1 + ... + d_{k-1}\mathbf{v}_{k-1} - 1\mathbf{v}_k + 0\mathbf{v}_{k+1} + ... + 0\mathbf{v}_n = \mathbf{0}$$

so $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ lin. dep. by definition. □

Example: In $\mathbb{R}^3$, set $\mathbf{v}_1 = (1,0,0)$; $\mathbf{v}_2 = (0,1,0)$; $\mathbf{v}_3 = (1,2,0)$.

Why do we care about whether or not a set of vectors is linearly dependent? Because linear dependence has a lot to do with what the span of these vectors is. Consider the following theorem:

Theorem 3.23 (Removing dependent vectors doesn't change span) *Let V be a real vector space. Let $\mathbf{v}_1, ..., \mathbf{v}_n$ be a collection of linearly dependent vectors such that*

$$\mathbf{v}_k = d_1\mathbf{v}_1 + ... + d_{k-1}\mathbf{v}_{k-1}.$$

Then $Span(\mathbf{v}_1, ..., \mathbf{v}_n) = Span(\mathbf{v}_1, ..., \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, ..., \mathbf{v}_n)$, in other words $\mathbf{v}_k$ can be removed from the list without changing the span of the vectors.

PROOF Let $W_1 = Span(\mathbf{v}_1, ..., \mathbf{v}_n)$ and let $W_2 = Span(\mathbf{v}_1, ..., \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, ..., \mathbf{v}_n)$. We will show $W_1 = W_2$ by showing each set is a subset of the other:

($\supseteq$): Suppose $\mathbf{w} \in W_2$. Then by definition of span,

$$\mathbf{w} = \sum_{j=1, j\neq k}^{n} c_j\mathbf{v}_j = \sum_{j=1}^{n} c_j\mathbf{v}_j \text{ (by setting } c_k = 0)$$

Thus by definition, $\mathbf{w} \in Span(\mathbf{v}_1, ..., \mathbf{v}_n)$ so $w \in W_1$.

($\subseteq$): Suppose $\mathbf{w} \in W_1$. Then by definition of span,

$$\begin{aligned}\mathbf{w} = \sum_{j=1}^{n} c_j\mathbf{v}_j &= \sum_{j=1, j\neq k}^{n} c_j\mathbf{v}_j + c_k\mathbf{v}_k \\ &= \sum_{j=1, j\neq k}^{n} c_j\mathbf{v}_j + c_k\sum_{j=1}^{k-1} d_j\mathbf{v}_j \\ &= \sum_{j=1}^{k-1}(c_j + c_k d_j)\mathbf{v}_j + \sum_{j=k+1}^{n} c_j\mathbf{v}_j.\end{aligned}$$

Thus $\mathbf{w} \in W_2$ by definition of span.

Since $W_1 \subseteq W_2$ and $W_2 \subseteq W_1$, $W_1 = W_2$. $\square$

Definition 3.24 *Let V be a real vector space. A collection of vectors $\mathbf{v}_1, ..., \mathbf{v}_n \in V$ is called* **linearly independent (lin. ind.)** *if the collection is not linearly dependent, i.e. the only $c_1, ..., c_n \in \mathbb{R}$ such that $c_1\mathbf{v}_1 + ... + c_n\mathbf{v}_n = \mathbf{0}$ are $c_1 = c_2 = ... = c_n = 0$.*

Equivalently, a set of vectors is linearly independent if the only combination of them that makes $\mathbf{0}$ is the trivial one.

Theorem 3.25 (Elementary properties of lin. dep. and lin. ind. sets) *Let V be a real vector space. Then:*

1. *The zero vector $\mathbf{0}$ cannot be part of a linearly independent set of vectors.*
2. *A collection of one vector $\mathbf{v}$ is linearly independent if and only if $\mathbf{v} \neq \mathbf{0}$.*
3. *A collection of two vectors $\mathbf{v}, \mathbf{w}$ is linearly dependent if and only if $\mathbf{v} || \mathbf{w}$.*
4. *Any subset of a collection of linearly independent vectors is linearly independent.*
5. *Any collection of vectors containing a linearly dependent subcollection is itself linearly dependent.*

Example: Let $\mathbf{v}_1 = (1, 0, 0)$; $\mathbf{v}_2 = (0, 1, 0)$; $\mathbf{v}_3 = (1, 1, 4)$. Is the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ of vectors linearly independent, or linearly dependent? Explain.

CONCEPTUAL IDEA	ALGEBRAIC JUSTIFICATION

Example: Let $\mathbf{v}_1 = (1, -1, 2, 5)$; $\mathbf{v}_2 = (0, 2, -1, 4)$; $\mathbf{v}_3 = (-3, 0, -1, 2)$. Is the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ of vectors linearly independent, or linearly dependent? Explain.

3.6 Basis and dimension

Definition 3.26 *Let V be a real vector space (or a subspace of some other real vector space). A collection of vectors $\mathbf{v}_1, ..., \mathbf{v}_n \in V$ is called a* **basis** *of V if*

1. *$\mathbf{v}_1, ..., \mathbf{v}_n$ are linearly independent; and*
2. *$V = Span(\mathbf{v}_1, ..., \mathbf{v}_n)$.*

Note: In statement 2, it is always true that $Span(\mathbf{v}_1, ..., \mathbf{v}_n) \subseteq V$. The question is whether every $\mathbf{v} \in V$ is a linear combination of the $\mathbf{v}_j$.

Heuristic: Statement 1 says that no "directions" in V are repeated by the $\mathbf{v}_j$ (i.e. that the set of the $\mathbf{v}_j$'s isn't too big). Statement 2 says that no "directions" are "missed" by the $\mathbf{v}_j$ (i.e. that the set of the $\mathbf{v}_j$'s isn't too small).

Example: Let $V = \mathbb{R}^3$. Determine whether each of the given sets of vectors forms a basis of V.

Ex. 1: $\mathbf{v}_1 = (1, 0, 0)$; $\mathbf{v}_2 = (0, 1, 0)$.

Ex. 2: $\mathbf{v}_1 = (1, 0, 0)$; $\mathbf{v}_2 = (0, 1, 0)$; $\mathbf{v}_3 = (0, 0, 1)$; $\mathbf{v}_4 = (1, -1, 2)$.

Ex. 3: $\mathbf{v}_1 = (1,0,0)$; $\mathbf{v}_2 = (0,1,0)$; $\mathbf{v}_3 = (1,2,0)$.

Ex. 4: $\mathbf{v}_1 = (1,0,0)$; $\mathbf{v}_2 = (0,1,0)$; $\mathbf{v}_3 = (0,0,1)$.

A vector space (or subspace) is completely described by giving a basis of that space, because of the following theorem:

Theorem 3.27 (Unique Representation Theorem) *Let V be a real vector space and let $\mathcal{B} = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ be a basis of V. Then, for every $\mathbf{v} \in V$ there is exactly one choice of $c_1, ..., c_n \in \mathbb{R}$ such that*

$$\mathbf{v} = \sum_{j=1}^{n} c_j \mathbf{v}_j.$$

PROOF Since $\mathcal{B}$ is a basis, $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ span V so every $\mathbf{v} \in V$ can be written as

$$\mathbf{v} = \sum_{j=1}^{n} c_j \mathbf{v}_j = c_1\mathbf{v}_1 + ... + c_n\mathbf{v}_n$$

in **at least one** way. Now suppose this can be done in two ways, i.e.

$$\mathbf{v} = \sum_{j=1}^{n} c_j \mathbf{v}_j = \sum_{j=1}^{n} d_j \mathbf{v}_j.$$

Subtracting the right-hand side from both sides, we get

$$\sum_{j=1}^{n} c_j \mathbf{v}_j - \sum_{j=1}^{n} d_j \mathbf{v}_j = \mathbf{0}$$

so

$$\sum_{j=1}^{n} (c_j - d_j) \mathbf{v}_j = \mathbf{0}.$$

But since $\mathcal{B}$ is a basis, $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ are linearly independent, so the left-hand side of the above equation has to be the trivial combination, i.e. $c_j - d_j = 0$ for all j, i.e. $c_j = d_j$ for all j. Thus, $\mathbf{v}$ cannot be written as a linear combination of the $\mathbf{v}_j$ in more than one way. $\square$

Example: One basis of $\mathbb{R}^3$ is $\{(1,0,0),(0,1,0),(0,0,1)\}$. Take the vector $\mathbf{v} = (-2,4,3)$.

We now come to the most important theoretical principle of linear algebra, upon which the rest of the subject (including all applications) rests:

Theorem 3.28 (Exchange Lemma) *Let V be a real vector space. If $\{\mathbf{v}_1, ..., \mathbf{v}_m\}$ spans V, and if $\{\mathbf{w}_1, ..., \mathbf{w}_n\}$ is linearly independent in V, then $m \geq n$.*

PROOF We are given that $V = Span(\mathbf{v}_1, ..., \mathbf{v}_m)$.

Now $\mathbf{w}_1 \in V$, so $\mathbf{w}_1 \in Span(\mathbf{v}_1, ..., \mathbf{v}_m)$. Therefore the set

$$\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m, \mathbf{w}_1\}$$

is linearly dependent (since the last vector is in the span of the others). Writing these vectors in a different order doesn't change the fact they are linearly dependent, so

$$\{\mathbf{w}_1, \mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$$

is also lin. dep. Now discard the first $\mathbf{v}_j$ in this list which depends on the previous vectors. Discarding this $\mathbf{v}_j$ doesn't change the span of the vectors in this list, so we have

$$V = Span(\mathbf{w}_1, \mathbf{v}_1, ..., \mathbf{v}_m) = Span(\mathbf{w}_1, \mathbf{v}_1, ..., \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, ..., \mathbf{v}_m).$$

Essentially, what we have done above is "exchange" one of the $\mathbf{v}_j$s for $\mathbf{w}_1$ without changing the span of the set. This technique is why this theorem is called the Exchange Lemma.

Now we "exchange" again. $\mathbf{w}_2 \in V$, so $\mathbf{w}_2 \in Span(\mathbf{w}_1, \mathbf{v}_1, ..., \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, ..., \mathbf{v}_m)$ so $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{v}_1, ..., \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, ..., \mathbf{v}_m\}$ is lin. dep. Remove the first vector $\mathbf{v}_k$ in this list which depends on the previous vectors to obtain set

$$\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{v}_1, ..., \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, ..., \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, ...\mathbf{v}_m\}$$

which still spans V.

Keep exchanging in this fashion; if $n > m$ then we can exchange all the $\mathbf{v}$s with $\mathbf{w}$s to obtain set

$$\{\mathbf{w}_1, ..., \mathbf{w}_m\}$$

which spans V, and still have $\mathbf{w}_{m+1}$ leftover. But

$$\mathbf{w}_{m+1} \in V = Span(\mathbf{w}_1, ..., \mathbf{w}_m),$$

so $\{\mathbf{w}_1, ..., \mathbf{w}_{m+1}\}$ is linearly dependent (since the last vector is in the span of the others). But this contradicts the hypothesis that $\{\mathbf{w}_1, ..., \mathbf{w}_n\}$ is linearly independent. Thus $n \leq m$ as desired. □

Corollary 3.29 (Dimension Theorem) *Let V be a real vector space which has a basis consisting of n vectors. Then any other basis of V must also consist of n vectors.*

PROOF Suppose $\mathcal{B} = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ and $\mathcal{B}' = \{\mathbf{w}_1, ..., \mathbf{w}_m\}$ are two bases of V.

$\mathcal{B}$ is spanning set; $\mathcal{B}'$ is lin. indep $\Rightarrow n \geq m$ by Exchange Lemma.
$\mathcal{B}$ is lin. indep.; $\mathcal{B}'$ is spanning set $\Rightarrow m \geq n$ by Exchange Lemma.

Thus $m = n$. □

Definition 3.30 *Let V be a real vector space. If V is spanned by a finite set of vectors, we say V is* **finite dimensional** *and write* $\dim V < \infty$*. In this case, the* **dimension** *of V is the number of elements in any basis of V. (We define* $\dim(\{\mathbf{0}\}) = 0$ *even though* $\{\mathbf{0}\}$ *does not have a basis.) If V is not spanned by any finite set of vectors, we say V is* **infinite dimensional** *and write* $\dim V = \infty$.

Definition 3.31 *Let V be a real vector space, and $A \subseteq V$ be an affine subspace of V. The* **dimension** *of A is the dimension of the subspace $W = A - \mathbf{p}$, where $\mathbf{p}$ is any translation vector for A. In particular, a* **point** *is an affine subspace of dimension zero; a* **line** *is an affine subspace of dimension one; and a* **plane** *is an affine subspace of dimension two. A* **hyperplane** *is an affine subspace whose dimension is one less than the dimension of V.*

Example: Find a basis for, and the the dimension of, of each of the following vector spaces:

Ex. 1: $\mathbb{R}^n$

Ex. 2: $M_n(\mathbb{R})$, the set of $n \times n$ matrices with entries in $\mathbb{R}$.

Ex. 3: The set of 3×3 diagonal matrices with entries in $\mathbb{R}$.

Ex. 4: The set $\mathbb{R}[x]$ of polynomials with real coefficients.

Theorem 3.32 (Spanning Set Theorem) *If $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ spans vector space V, then some subset of $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ forms a basis of V (so $n \geq \dim V$).*

Consequence: No two vectors span $\mathbb{R}^3$; no m vectors span $\mathbb{R}^n$ if $n > m$, etc.

PROOF Start with spanning set $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$. If the vectors are lin. indep., they form a basis. Otherwise, they are lin. dep., so there is a first vector in the list which depends on the preceding vectors. Discard this vector from the list. If what is left is lin. indep., then it forms a basis; otherwise, keep discarding vectors which depend on the others until what is left is lin. indep. Since discarding dependent vectors doesn't change the span, what's left will span V and be linearly independent, hence form a basis. □

Theorem 3.33 (Linearly Independent Set Theorem) *If $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is a linearly independent set of vectors in vector space V, then $n \leq \dim V$.*

Consequence: No four vectors in $\mathbb{R}^3$ are linearly independent; no m vectors in $\mathbb{R}^n$ are linearly independent if $m > n$, etc.

PROOF Follows immediately from the Exchange Lemma, since there are $\dim V$ vectors in a basis of V (which spans V). □

Theorem 3.34 (Basis Extension Theorem) *Let V be a finite dimensional real vector space. If $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is a linearly independent set of vectors in V, then V has a basis of the form*

$$\{\mathbf{v}_1, ..., \mathbf{v}_n, \mathbf{w}_1, ..., \mathbf{w}_m\}$$

(where $\dim(V) = m + n$).

Note: m could equal 0 in this theorem (if the $\mathbf{v}_j$s already form a basis).

Proof: Let $W_0 = Span(\mathbf{v}_1, ..., \mathbf{v}_n)$. If $W_0 = V$, then we are done ($\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ forms a basis of V).

Otherwise, there is $\mathbf{w}_1 \notin W_0$. Thus $\{\mathbf{v}_1, ..., \mathbf{v}_n, \mathbf{w}_1\}$ is lin. ind. (since the first n vectors in this list are lin. ind. and the last vector is not a linear combination of the first n); let $W_1 = Span(\mathbf{v}_1, ..., \mathbf{v}_n, \mathbf{w}_1)$. If $W_1 = V$, then we are done ($\{\mathbf{v}_1, ..., \mathbf{v}_n, \mathbf{w}_1\}$ forms a basis of V).

Otherwise, there is $\mathbf{w}_2 \notin W_1$. Thus $\{\mathbf{v}_1, ..., \mathbf{v}_n, \mathbf{w}_1, \mathbf{w}_2\}$ is lin. ind. Let $W_2 = Span(\mathbf{v}_1, ..., \mathbf{v}_n, \mathbf{w}_1, \mathbf{w}_2)$. If $W_2 = V$, then we are done ($\{\mathbf{v}_1, ..., \mathbf{v}_n, \mathbf{w}_1\}$ forms a basis of V).

Otherwise, continue this process. Eventually, if $\dim V = m + n$, we get a list of lin. ind. vectors $\{\mathbf{v}_1, ..., \mathbf{v}_n, \mathbf{w}_1, ..., \mathbf{w}_m\}$. Let W_m be the span of these vectors. If

$W_m \neq V$, then there is $\mathbf{w}_{m+1} \notin W_m$ that could be added to this collection to produce a list of $n+m+1 > \dim V$ lin. ind. vectors, contradicting the Linearly Independent Set Theorem. □

Corollary 3.35 (Basis Theorem) *Let V be a real vector space and suppose $n = \dim V < \infty$. Then:*

1. *V has a basis (so long as $V \neq \mathbf{0}$).*
2. *Any set of n linearly independent vectors in V forms a basis of V.*
3. *Any set of n vectors which span V forms a basis of V.*

PROOF For (1), if $V \neq \mathbf{0}$ choose $\mathbf{v} \in V$. $\{\mathbf{v}\}$ is linearly independent; by the Basis Extension Theorem it can be extended to a basis of V.

For (2), if the set of linearly independent vectors $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ isn't a basis, then these vectors don't span V, so there is $\mathbf{w} \in V$ which is not in their span. Thus $\{\mathbf{v}_1, ..., \mathbf{v}_n, \mathbf{w}\}$ is a set of $n+1 > \dim V$ lin. ind. vectors in V, contradicting the Linearly Independent Set Theorem.

For (3), if the spanning set $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ isn't a basis, then these vectors are lin. dep. Remove all the dependent vectors from this set; this leaves a set of $< n$ vectors which still span V (since removing dependent vectors doesn't change the span of a set). This contradicts the Spanning Set Theorem. □

Corollary 3.36 *If W is a subspace of real vector space V, then $\dim W \leq \dim V$. If W is a subspace of real vector space V and $\dim W = \dim V < \infty$, then $W = V$.*

PROOF First, let's prove the first statement. If $\dim V = \infty$, the result is trivial. Otherwise, let $n = \dim W$ and let $\{\mathbf{w}_1, ..., \mathbf{w}_n\}$ be a basis of W. This basis can be extended to a basis of V, which must necessarily have $\geq n$ elements. Therefore $\dim V \geq n = \dim W$ as desired.

For the second statement, if $n = \dim V = \dim W < \infty$, let $\{\mathbf{w}_1, ..., \mathbf{w}_n\}$ be a basis of W. Thus $\{\mathbf{w}_1, ..., \mathbf{w}_n\}$ are a set of n lin. ind. vectors in V, hence form a basis of V by the Basis Theorem. Therefore $W = Span(\mathbf{w}_1, ..., \mathbf{w}_n) = V$. □

We can now revisit some ideas we encountered earlier:

Theorem 3.37 (Classification of subspaces of $\mathbb{R}$, $\mathbb{R}^2$ and $\mathbb{R}^3$)

1. *The only subspaces of $\mathbb{R}$ are $\{\mathbf{0}\}$ and $\mathbb{R}$.*
2. *The only subspaces of $\mathbb{R}^2$ are $\{\mathbf{0}\}$, lines passing through the origin, and $\mathbb{R}^2$.*
3. *The only subspaces of $\mathbb{R}^3$ are $\{\mathbf{0}\}$, lines passing through the origin, planes passing through the origin, and $\mathbb{R}^3$.*

PROOF For statement (3), notice $\dim \mathbb{R}^3 = 3$. If $W \subseteq \mathbb{R}^3$ is a subspace, then $\dim W \leq 3$. Since $\dim W$ is a nonnegative integer, we have four possibilities:

- $\dim W = 0$. Then W is a point by definition.
- $\dim W = 1$. Then W is a line by definition.
- $\dim W = 2$. Then W is a plane by definition.
- $\dim W = 3$. Then $W = \mathbb{R}^3$ by Corollary 3.36.

The proofs of the first two statements are similar. □

Corollary 3.38 (Classification of affine subspaces of $\mathbb{R}$, $\mathbb{R}^2$ and $\mathbb{R}^3$)

1. *The only affine subspaces of $\mathbb{R}$ are points, and all of $\mathbb{R}$.*
2. *The only affine subspaces of $\mathbb{R}^2$ are points, lines and all of $\mathbb{R}^2$.*
3. *The only affine subspaces of $\mathbb{R}^3$ are points, lines, planes and all of $\mathbb{R}^3$.*

PROOF This is obvious, given the preceding theorem. □

3.7 Summary of Chapter 3

Fundamental definitions:

- A **subspace** W of vector space V is a subset of V which is itself a vector space. That means that W must contain $\mathbf{0}$, be closed under addition, and closed under scalar multiplication.
- An **affine subspace** A of vector space V is a translate of a subspace of V.
- The **span** of a set of vectors is the set of linear combinations of those vectors; the span of any set of vectors is always a subspace. A set of vectors **spans** a vector space (or subspace) V if every vector in V is a linear combination of vectors in that set.
- A set of vectors $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is called **linearly independent** if the only way to write $\mathbf{0} = c_1\mathbf{v}_1 + ... + c_n\mathbf{v}_n$ is to set all the $c_j = 0$. A set of vectors is called **linearly dependent** if it is not linearly independent.
- A set of vectors is called a **basis** of vector space V if the set spans V and is linearly independent.
- Any two bases of a vector space V must have the same number of elements; the count of the number of vectors in any basis of V is called the **dimension** of V.

Fundamental theoretical concepts relating these definitions:

- A set of vectors is lin. dep. if and only if there is some vector in the list which is a linear combination of the previous vectors in the list.
- A set of one vector is lin. indep. if and only if the vector is nonzero.
- $\mathbf{0}$ is never part of a lin. indep. set of vectors.
- A set of two vectors is lin. indep. if and only if the vectors are nonparallel.
- A subset of a linearly independent set must also be lin. indep.
- If a set of vectors has a lin. dep. subset, then it is also lin. dep.
- Any set of more than $\dim V$ vectors in V must be lin. dep.
- No set of less than $\dim V$ vectors in V can span V.
- Lin. dep. vectors can be removed from a list without changing the span of that list, and every spanning set for V can be reduced to a basis of V by removing the lin. dep. vectors. Any set of $\dim V$ vectors which span V must also be a basis of V (i.e. must also be lin. indep.).

- Any lin. indep. set of vectors in V can be extended to a basis of V. Any set of $\dim V$ vectors which are lin. indep. must also be a basis of V (i.e. must also span V).
- Every vector space has a basis.
- Given a basis of V, every vector in V can be written as a linear combination of those basis vectors in exactly one way.
- If W is a subspace of V, then $\dim W \leq \dim V$.
- If W is a subspace (or affine subspace) of V and $\dim W = \dim V < \infty$, then $V = W$.

Classification of subspaces and affine subspaces: we classify these objects according to their dimension:

- The only subspaces of $\mathbb{R}$ are $\{0\}$ (dimension zero) and $\mathbb{R}$ (dimension one).

  The only affine subspaces of $\mathbb{R}$ are points (dimension zero) and $\mathbb{R}$ (dimension one).
- The only subspaces of $\mathbb{R}^2$ are $\{\mathbf{0}\}$ (dimension zero), lines passing through the origin (dimension one), and all of $\mathbb{R}^2$ (dimension two).

  The only affine subspaces of $\mathbb{R}^2$ are points (dimension zero), lines (dimension one) and all of $\mathbb{R}^2$ (dimension two).
- The only subspaces of $\mathbb{R}^3$ are $\{\mathbf{0}\}$ (dimension zero), lines passing through the origin (dimension one), planes passing through the origin (dimension two), and all of $\mathbb{R}^3$ (dimension three).

  The only affine subspaces of $\mathbb{R}^3$ are points (dimension zero), lines (dimension one), planes (dimension two), and all of $\mathbb{R}^3$ (dimension three).

In general, you should think of an affine subspace as an object akin to a point, line or plane (but perhaps larger-dimensional), and you should think of a subspace as an affine subspace which passes through $\mathbf{0}$.

Chapter 4

Dot products and orthogonality

4.1 Definitions and properties

Goals: Generalize the ideas of "length", "angle", "perpendicularity" to vectors.

Definition 4.1 *The* **dot product** *on* $\mathbb{R}^n$ *is the function* $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *(where the output associated to inputs* $\mathbf{v} = (v_1, ..., v_n)$ *and* $\mathbf{w} = (w_1, ..., w_n)$ *is written* $\mathbf{v} \cdot \mathbf{w}$ *or* $\mathbf{v}^T\mathbf{w}$*) defined by*

$$\mathbf{v} \cdot \mathbf{w} = \sum_{j=1}^{n} v_j w_j = v_1 w_1 + v_2 w_2 + ... + v_n w_n.$$

Note: $\mathbf{v} \cdot \mathbf{w}$ is a scalar, not a vector.

Note: This dot product is a special case of matrix multiplication. Writing $\mathbf{v}$ and $\mathbf{w}$ as column vectors, we see

Example: $(3, -2, 5) \cdot (1, 0, -2) =$

Spaces of functions also have a notion of dot product, which plays the same role as the above dot product does for traditional vectors:

Definition 4.2 *Let $[a,b]$ be a closed, bounded interval in $\mathbb{R}$. The* **dot product** *on $C([a,b],\mathbb{R})$ is the function $C([a,b],\mathbb{R})\times C([a,b],\mathbb{R}) \to \mathbb{R}$ (where the output associated to f and g is written $f \cdot g$) defined by*

$$f \cdot g = \int_a^b f(x)g(x)\,dx.$$

Example: Let $f(x) = x, g(x) = x^2$. Then $f, g \in C([0,1],\mathbb{R})$, and in this vector space,

$$f \cdot g =$$

Definition 4.3 *The* **dot product** *on $M_n(\mathbb{R})$ (the space of $n \times n$ matrices) is the function $M_n(\mathbb{R}) \times M_n(\mathbb{R}) \to \mathbb{R}$ (where the output associated to A and B is written $A \cdot B$) defined by $A \cdot B = tr(AB^T)$.*

Henceforth, we will call any vector space, together with its appropriate choice of dot product, a "dot product space". (There are other examples of such spaces.) This is not standard mathematics notation; usually one would call such a thing an **inner product space**, but this requires a discussion of generalizations of dot product that will not be discussed in our course.

Theorem 4.4 (Properties of dot products) *Let V be a dot product space. Then:*

1. Dot product is symmetric, *i.e.* $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$ *for all* $\mathbf{v}, \mathbf{w} \in V$.
2. Dot product is bilinear, *i.e.*

$$(\mathbf{v}_1 + \mathbf{v}_2) \cdot \mathbf{w} = \mathbf{v}_1 \cdot \mathbf{w} + \mathbf{v}_2 \cdot \mathbf{w} \text{ for all } \mathbf{v}_1, \mathbf{v}_2, \mathbf{w} \in V;$$

$$\mathbf{v}_1 \cdot (\mathbf{w}_1 + \mathbf{w}_2) = \mathbf{v}_1 \cdot \mathbf{w}_1 + \mathbf{v}_1 \cdot \mathbf{w}_2 \text{ for all } \mathbf{v}, \mathbf{w}_1, \mathbf{w}_2 \in V;$$

$$c(\mathbf{v} \cdot \mathbf{w}) = (c\mathbf{v}) \cdot \mathbf{w} = \mathbf{v} \cdot (c\mathbf{w}) \text{ for all } \mathbf{v}, \mathbf{w} \in V, \text{ all } c \in \mathbb{R}.$$

3. Dot product is positive, *i.e.* $\mathbf{v} \cdot \mathbf{v} \geq 0$ *for any* $\mathbf{v} \in V$.
4. Dot product is definite, *i.e.* $\mathbf{v} \cdot \mathbf{v} = 0$ *implies* $\mathbf{v} = \mathbf{0}$.

Note: The bilinearity property extends to the following fact: if $\mathbf{v}_i, \mathbf{w}_j \in V$ and $c_i, d_j \in \mathbb{R}$, then

$$\left(\sum_{i=1}^m c_i \mathbf{v}_i\right) \cdot \left(\sum_{j=1}^n d_j \mathbf{w}_j\right) = \sum_{i=1}^m \sum_{j=1}^n c_i d_j (\mathbf{v}_i \cdot \mathbf{w}_j).$$

PROOF Here, we'll prove this in the situation where $V = \mathbb{R}^n$ (the other examples require their own proofs, which are similar):

First, let's prove statement 1: let $\mathbf{v} = (v_1, ..., v_n)$ and $\mathbf{w} = (w_1, ..., w_n)$. Then

$$\mathbf{v} \cdot \mathbf{w} = \sum_{j=1}^{n} v_j w_j = \sum_{j=1}^{n} w_j v_j = \mathbf{w} \cdot \mathbf{v}.$$

Now, let's prove statement 3: let $\mathbf{v} = (v_1, ..., v_n)$. Then

$$\mathbf{v} \cdot \mathbf{v} = v_1 v_1 + ... + v_n = \sum_{j=1}^{n} v_j^2 \geq 0.$$

Now let's prove statement 4: let $\mathbf{v} = (v_1, ..., v_n)$. Then if

$$0 = \mathbf{v} \cdot \mathbf{v} = v_1 v_1 + ... + v_n = \sum_{j=1}^{n} v_j^2,$$

it must be that $v_j^2 = 0$ for all j (otherwise the sum would be strictly positive); thus $v_j = 0$ for all j, i.e. $\mathbf{v} = (0, 0, ..., 0) = \mathbf{0}$.

Last, let's verify the statements in 2:

First, let $\mathbf{v}_1 = (u_1, ..., u_n)$, $\mathbf{v}_2 = (v_1, ..., v_n)$ and $\mathbf{w} = (w_1, ..., w_n)$. Then

$$(\mathbf{v}_1 + \mathbf{v}_2) \cdot \mathbf{w} = \sum_{j=1}^{n} (u_j + v_j) w_j = \sum_{j=1}^{n} u_j w_j + \sum_{j=1}^{n} v_j w_j = \mathbf{v}_1 \cdot \mathbf{w} + \mathbf{v}_2 \cdot \mathbf{w}$$

and

$$\mathbf{v} \cdot (\mathbf{w}_1 + \mathbf{w}_2) = (\mathbf{w}_1 + \mathbf{w}_2) \cdot \mathbf{v} = \mathbf{w}_1 \cdot \mathbf{v} + \mathbf{w}_2 \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{w}_1 + \mathbf{v} \cdot \mathbf{w}_2.$$

Next, let $\mathbf{v} = (v_1, ..., \mathbf{v}_n)$ and $\mathbf{w} = (w_1, ..., w_n)$. Then

$$\begin{aligned} c(\mathbf{v} \cdot \mathbf{w}) = c \sum_{j=1}^{n} v_j w_j &= \sum_{j=1}^{n} (c v_j) w_j = (c\mathbf{v}) \cdot \mathbf{w} \\ &= \sum_{j=1}^{n} v_j (c w_j) = \mathbf{v} \cdot (c\mathbf{w}) \end{aligned}$$

This finishes the proof. □

This last elementary result is very important. It will be used later when analyzing matrices and studying systems of linear equations:

Theorem 4.5 (Dual relations) *Let $A \in M_{mn}(\mathbb{R})$. Then, for any $\mathbf{x} \in \mathbb{R}^n$ and any $\mathbf{y} \in \mathbb{R}^m$, we have*

$$A\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot A^T\mathbf{y}.$$

(Similarly, for any $B \in M_{nm}(\mathbb{R})$, $\mathbf{x} \cdot B\mathbf{y} = B^T\mathbf{x} \cdot \mathbf{y}$.)

PROOF This is a direct calculation:

$$\begin{aligned} A\mathbf{x} \cdot \mathbf{y} &= (A\mathbf{x})^T\mathbf{y} \\ &= \mathbf{x}^T A^T \mathbf{y} \quad \text{(Theorem 2.11 - props of matrix operations)} \\ &= \mathbf{x}^T (A^T\mathbf{y}) \\ &= \mathbf{x} \cdot (A^T\mathbf{y}). \ \square \end{aligned}$$

4.2 Norms, distances and length

Recall the definition of the absolute value of a real number:

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$

$|x|$ can also be thought of as the distance between the numbers x and 0. We want to generalize this idea to vectors:

Definition 4.6 *Let V be a dot product space. For any vector $\mathbf{v} \in V$, we define the* **norm** *(a.k.a.* **length** *a.k.a.* **magnitude** *a.k.a.* **absolute value***) of $\mathbf{v}$ to be $||\mathbf{v}|| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$.*

Note: Since dot product is positive, we'll never have to take the square root of a negative number here.

Remark: It is also useful to keep in mind that since $||\mathbf{v}|| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$,

$$||\mathbf{v}||^2 = \mathbf{v} \cdot \mathbf{v}.$$

This makes $||\mathbf{v}||^2$ easier to work with than $||\mathbf{v}||$ when writing arguments.

Example: Find $||\mathbf{v}||$ if $\mathbf{v} = (-4, 2, 5, 1, 0)$.

Example: Find $||A||$ if $A = \begin{pmatrix} 2 & -3 \\ 1 & 4 \end{pmatrix}$.

Theorem 4.7 (Properties of norms) *Let V be a dot product space and let $\mathbf{v}, \mathbf{w} \in V$. Then:*

1. Norms are positive: $||\mathbf{v}|| \geq 0$.
2. Norms are definite: $||\mathbf{v}|| = 0$ *if and only if* $\mathbf{v} = \mathbf{0}$.
3. Norms are multiplicative: $||k\mathbf{v}|| = |k| \, ||\mathbf{v}||$ *for all* $k \in \mathbb{R}$.
4. Triangle inequality: $||\mathbf{v} + \mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||$.

PROOF Statements (1) and (2) follow from statements (3) and (4) of Theorem 4.4 (properties of dot products). For property (3), observe

$$||k\mathbf{v}|| = \sqrt{k\mathbf{v} \cdot k\mathbf{v}} = \sqrt{k^2}(\mathbf{v} \cdot \mathbf{v}) = |k|\sqrt{\mathbf{v} \cdot \mathbf{v}} = |k| \, ||\mathbf{v}||.$$

We will prove property (4) in Section 4.6. □

The norm of a vector should be thought of as its distance from $\mathbf{0}$. More generally:

Definition 4.8 *Let V be a dot product space and let $\mathbf{v}, \mathbf{w} \in V$. Then the* **distance** *from $\mathbf{v}$ to $\mathbf{w}$ is $dist(\mathbf{v}, \mathbf{w}) = ||\mathbf{v} - \mathbf{w}||$.*

By the multiplicative property of norms, $||\mathbf{v} - \mathbf{w}|| = ||\mathbf{w} - \mathbf{v}||$ since

$$||\mathbf{v} - \mathbf{w}|| = ||(-1)(\mathbf{w} - \mathbf{v})|| = |-1|\,||\mathbf{w} - \mathbf{v}|| = ||\mathbf{w} - \mathbf{v}||.$$

Therefore $dist(\mathbf{v}, \mathbf{w}) = dist(\mathbf{w}, \mathbf{v})$.

Definition 4.9 *Let V be a dot product space. A vector $\mathbf{v} \in V$ is called a* **unit vector** *if $||\mathbf{v}|| = 1$.*

Example: Describe all the unit vectors in $\mathbb{R}^2$.

Theorem 4.10 (Properties of distance) *Let V be a dot product space and let $\mathbf{u}, \mathbf{v}, \mathbf{w}$ be vectors in V. Then:*

1. Distances are positive: $dist(\mathbf{v}, \mathbf{w}) \geq 0$.
2. Distances are definite: $dist(\mathbf{v}, \mathbf{w}) = 0$ *if and only if* $\mathbf{v} = \mathbf{w}$.
3. Distances are symmetric: $dist(\mathbf{v}, \mathbf{w}) = dist(\mathbf{w}, \mathbf{v})$.
4. Distances are multiplicative: $dist(k\mathbf{v}, k\mathbf{w}) = |k|\, dist(\mathbf{v}, \mathbf{w})$ *for all* $k \in \mathbb{R}$.
5. Triangle inequality: $dist(\mathbf{v}, \mathbf{w}) \leq dist(\mathbf{v}, \mathbf{u}) + dist(\mathbf{u}, \mathbf{w})$.

PROOF

1. $dist(\mathbf{v}, \mathbf{w}) = ||\mathbf{v} - \mathbf{w}|| \geq 0$.
2. Suppose $dist(\mathbf{v}, \mathbf{w}) = 0$. Then $||\mathbf{v} - \mathbf{w}|| = 0$ so $\mathbf{v} - \mathbf{w} = \mathbf{0}$ so $\mathbf{v} = \mathbf{w}$.
3. This was proven above.
4. This follows from the earlier triangle inequality (not yet proven):

$$\begin{aligned} dist(\mathbf{v}, \mathbf{w}) = ||\mathbf{v} - \mathbf{w}|| &= ||\mathbf{v} - \mathbf{u} + \mathbf{u} - \mathbf{w}|| \\ &= ||(\mathbf{v} - \mathbf{u}) + (\mathbf{u} - \mathbf{w})|| \\ &\leq ||\mathbf{v} - \mathbf{u}|| + ||\mathbf{u} - \mathbf{w}|| \\ &= dist(\mathbf{v}, \mathbf{u}) + dist(\mathbf{u}, \mathbf{w}). \ \square \end{aligned}$$

Theorem 4.11 *Let V be a dot product space. Given any nonzero vector $\mathbf{v} \in V$, there is a unit vector in the same direction as $\mathbf{v}$. This unit vector is called a* **normalized version** *of $\mathbf{v}$.*

PROOF Given $\mathbf{v} \neq \mathbf{0}$, define $\mathbf{u} = \frac{1}{||\mathbf{v}||}\mathbf{v}$. This vector is clearly in the same direction as $\mathbf{v}$;

$$||\mathbf{u}|| = \left|\left|\frac{1}{||\mathbf{v}||}\mathbf{v}\right|\right| = \frac{1}{||\mathbf{v}||}||\mathbf{v}|| = 1$$

so $\mathbf{u}$ is a unit vector. □

Example: Find a unit vector in the same direction as $(2, -3, 5, 1)$.

Theorem 4.12 (Polarization Identity) *Let V be a (real) dot product space. Given any $\mathbf{v}, \mathbf{w} \in V$,*

$$\mathbf{v} \cdot \mathbf{w} = \frac{1}{4}\left(||\mathbf{v} + \mathbf{w}||^2 - ||\mathbf{v} - \mathbf{w}||^2\right).$$

Consequence: In this class, we defined norms and distances in terms of dot products. This theorem shows that if you have a reasonable notion of distance or norm, then you can always find a dot product that created it. (This shows the real importance of dot products in that it proves that you can't talk about geometric properties of vectors without using them.)

PROOF

$$\frac{1}{4}\left[||\mathbf{v} + \mathbf{w}||^2 - ||\mathbf{v} - \mathbf{w}||^2\right]$$

$$= \frac{1}{4}\left[(\mathbf{v} + \mathbf{w}) \cdot (\mathbf{v} + \mathbf{w}) - (\mathbf{v} - \mathbf{w}) \cdot (\mathbf{v} - \mathbf{w})\right]$$

$$= \frac{1}{4}\left[\mathbf{v} \cdot \mathbf{v} + \mathbf{w} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{w} + \mathbf{w} \cdot \mathbf{w} - (\mathbf{v} \cdot \mathbf{v} - \mathbf{w} \cdot \mathbf{v} - \mathbf{v} \cdot \mathbf{w} + \mathbf{w} \cdot \mathbf{w})\right]$$
$$= \frac{1}{4}\left[4\mathbf{v} \cdot \mathbf{w}\right]$$
$$= \mathbf{v} \cdot \mathbf{w}.$$ □

Theorem 4.13 (Parallelogram Law) *Let V be a dot product space. Given any $\mathbf{v}, \mathbf{w} \in V$,*

$$||\mathbf{v}+\mathbf{w}||^2 + ||\mathbf{v}-\mathbf{w}||^2 = 2\left(||\mathbf{v}||^2 + ||\mathbf{w}||^2\right).$$

Why is this called the "parallelogram law"?

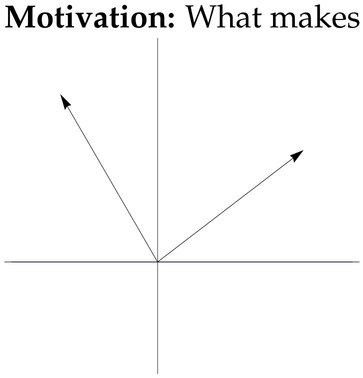
PROOF

$$\begin{aligned} ||\mathbf{v}+\mathbf{w}||^2 + ||\mathbf{v}-\mathbf{w}||^2 &= (\mathbf{v}+\mathbf{w})\cdot(\mathbf{v}+\mathbf{w}) + (\mathbf{v}-\mathbf{w})\cdot(\mathbf{v}-\mathbf{w}) \\ &= [\mathbf{v}\cdot\mathbf{v} + 2\mathbf{v}\cdot\mathbf{w} + \mathbf{w}\cdot\mathbf{w}] + [\mathbf{v}\cdot\mathbf{v} - 2\mathbf{v}\cdot\mathbf{w} + \mathbf{w}\cdot\mathbf{w}] \\ &= 2(\mathbf{v}\cdot\mathbf{v} + \mathbf{w}\cdot\mathbf{w}) \\ &= 2\left(||\mathbf{v}||^2 + ||\mathbf{w}||^2\right). \ \square \end{aligned}$$

4.3 Orthogonality

We want to extend the notion of "perpendicularity" for $\mathbb{R}^2$ to arbitrary vector spaces.

Motivation: What makes $\mathbf{v}_1 = (x_1, y_1)$ and $\mathbf{v}_2 = (x_2, y_2)$ $\perp$ in $\mathbb{R}^2$?



Definition 4.14 *Let V be a dot product space. Two vectors $\mathbf{v}, \mathbf{w}$ are called* **orthogonal** *(a.k.a.* **perpendicular***) if $\mathbf{v} \cdot \mathbf{w} = 0$, in which case we write $\mathbf{v} \perp \mathbf{w}$.*

Note: The zero vector $\mathbf{0}$ is orthogonal to every vector (since $\mathbf{0} \cdot \mathbf{v} = 0$ for any $\mathbf{v} \in V$).

Example: Determine whether or not the vectors $(2, -3, 1, -4)$ and $(0, 1, 7, 1)$ in $\mathbb{R}^4$ are orthogonal.

Example: Determine whether or not the functions $f(x) = \sin x$ and $g(x) = \cos x$ are orthogonal in $C([0, \pi], \mathbb{R})$.

Theorem 4.15 (Pythagorean Theorem) *Let V be a dot product space and let $\mathbf{v}, \mathbf{w} \in V$. Then*

$$||\mathbf{v} + \mathbf{w}||^2 = ||\mathbf{v}||^2 + ||\mathbf{w}||^2 \text{ if and only if } \mathbf{v} \perp \mathbf{w}.$$

Proof: ($\Leftarrow$) Assume $\mathbf{v} \perp \mathbf{w}$. Then

$$\begin{aligned} ||\mathbf{v} + \mathbf{w}||^2 &= (\mathbf{v} + \mathbf{w}) \cdot (\mathbf{v} + \mathbf{w}) \\ &= \mathbf{v} \cdot \mathbf{v} + \mathbf{w} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{w} + \mathbf{w} \cdot \mathbf{w} \\ &= \mathbf{v} \cdot \mathbf{v} + 0 + 0 + \mathbf{w} \cdot \mathbf{w} \\ &= ||\mathbf{v}||^2 + ||\mathbf{w}||^2. \end{aligned}$$

($\Rightarrow$) Assume $||\mathbf{v}+\mathbf{w}||^2 = ||\mathbf{v}||^2 + ||\mathbf{w}||^2$. Then

$$\begin{aligned}(\mathbf{v}+\mathbf{w})\cdot(\mathbf{v}+\mathbf{w}) &= \mathbf{v}\cdot\mathbf{v}+\mathbf{w}\cdot\mathbf{w}\\ \mathbf{v}\cdot\mathbf{v}+\mathbf{v}\cdot\mathbf{w}+\mathbf{w}\cdot\mathbf{v}+\mathbf{w}\cdot\mathbf{w} &= \mathbf{v}\cdot\mathbf{v}+\mathbf{w}\cdot\mathbf{w}\\ 2\mathbf{v}\cdot\mathbf{w} &= 0\\ \mathbf{v}\cdot\mathbf{w} &= 0\\ &\Rightarrow \mathbf{v}\perp\mathbf{w}. \square\end{aligned}$$

4.4 Projections and orthogonal complements

Definition 4.16 *Let V be a dot product space and let W be a subspace of V. Define the* **orthogonal complement of** *W, denoted $W^\perp$, to be the set of vectors orthogonal to every $\mathbf{w} \in W$; that is,*

$$W^\perp = \{\mathbf{v} \in V : \mathbf{v}\cdot\mathbf{w} = 0\,\forall \mathbf{w} \in W\}.$$

Theorem 4.17 *Let V be a dot product space and let W be a subspace of V spanned by $\mathbf{w}_1, ..., \mathbf{w}_n$. A vector $\mathbf{v} \in V$ is in $W^\perp$ if and only if $\mathbf{v}\cdot\mathbf{w}_j = 0$ for $j = 1, 2, ..., n$.*

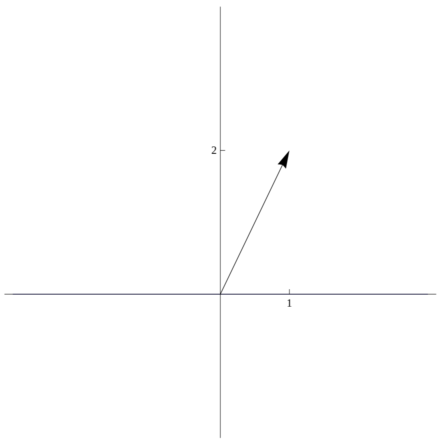
PROOF ($\Rightarrow$) follows from the definition of $W^\perp$, since $\mathbf{w}_j \in W$ for all j.

($\Leftarrow$). Let $\mathbf{w} \in W$. Since the $\mathbf{w}_j$ span W, $\mathbf{w} = \sum_j c_j\mathbf{w}_j$ so

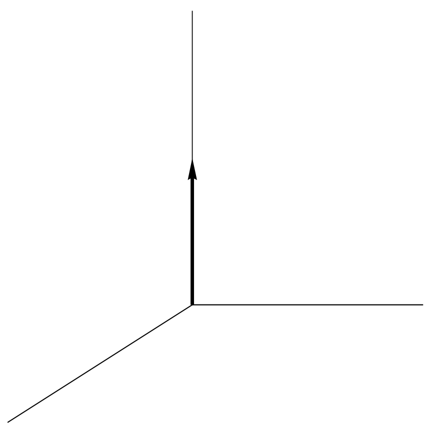
$$\mathbf{v}\cdot\mathbf{w} = \mathbf{v}\cdot\left(\sum_j c_j\mathbf{w}_j\right) = \sum_j c_j(\mathbf{v}\cdot\mathbf{w}_j) = \sum_j c_j 0 = 0$$

so $\mathbf{v}\perp\mathbf{w}$ for all $\mathbf{w} \in W$, so $\mathbf{v} \in W^\perp$ as desired. $\square$

Example 1: Let $V = \mathbb{R}^2$ and let $W = Span((1, 2))$. What is $W^\perp$?



Example 2: Let $V = \mathbb{R}^3$ and let $W = Span((0, 0, 1))$. What is $W^\perp$?



Observe: In Examples 1 and 2, we see that:

•

•

•

- every $\mathbf{v} \in V$ can be written as
$$\mathbf{v} = \pi_W(\mathbf{v}) + \pi_{W^\perp}(\mathbf{v}),$$
where $\pi_W(\mathbf{v}) \in W$ and $\pi_{W^\perp}(\mathbf{v}) \in W^\perp$:
- if $\mathcal{B}$ is a basis of W and $\mathcal{B}^\perp$ is a basis of $W^\perp$, then $\mathcal{B} \cup \mathcal{B}^\perp$ is a basis of V.

We now turn to verifying that these observations hold in general:

Theorem 4.18 (Orthogonal complements are subspaces) *Let V be a dot product space and let W be a subspace of V. Then $W^\perp$ is also a subspace of V.*

PROOF We verify the three essential characteristics of a subspace. Keep in mind that a generic element of $W^\perp$ is a vector $\mathbf{v}$ such that $\mathbf{v} \cdot \mathbf{w} = 0$ for all $\mathbf{w} \in W$.

- $\mathbf{0} \in W^\perp$ since $\mathbf{0} \cdot \mathbf{w} = 0$ for any $\mathbf{w} \in V$, hence any $\mathbf{w} \in W$.
- Let $\mathbf{v}_1, \mathbf{v}_2 \in W^\perp$. Then $\mathbf{v}_1 \cdot \mathbf{w} = 0$ and $\mathbf{v}_2 \cdot \mathbf{w} = 0$ for any $\mathbf{w} \in W$, so $(\mathbf{v}_1 + \mathbf{v}_2) \cdot \mathbf{w} = \mathbf{v}_1 \cdot w + \mathbf{v}_2 \cdot \mathbf{w} = 0 + 0 = 0$ for any $\mathbf{w} \in W$, so $\mathbf{v}_1 + \mathbf{v}_2 \in W^\perp$.
- Let $\mathbf{v} \in W^\perp$ and $r \in \mathbb{R}$. Then $\mathbf{v} \cdot \mathbf{w} = 0$ for any $\mathbf{w} \in W$, so $(r\mathbf{v} \cdot \mathbf{w}) = r(\mathbf{v} \cdot \mathbf{w}) = r(0) = 0$ for any $\mathbf{w} \in W$ so $r\mathbf{v} \in W^\perp$.

Thus $W^\perp$ is a subspace of V. □

Theorem 4.19 *Let V be a dot product space and let W be a subspace of V. Then $W \cap W^\perp = \{\mathbf{0}\}$.*

PROOF Clearly $\mathbf{0} \in W \cap W^\perp$, since $\mathbf{0}$ is in every subspace. Now suppose $\mathbf{w} \in W \cap W^\perp$. Then $\mathbf{w} \perp \mathbf{w}$, so $\mathbf{w} \cdot \mathbf{w} = 0$. By definiteness, this implies $\mathbf{w} = \mathbf{0}$. □

Theorem 4.20 (Orthogonal Decomposition Theorem (dimension 1)) *Let V be a dot product space; let $\mathbf{w} \neq 0$ be a vector in V and let $W = Span(\mathbf{w})$ (so that W is a one-dimensional subspace of V). Then for any $\mathbf{v} \in V$, we can write*

$$\mathbf{v} = \pi_{\mathbf{w}}(\mathbf{v}) + \pi_{\mathbf{w}^\perp}(\mathbf{v})$$

where $\pi_{\mathbf{w}}(\mathbf{v})||\mathbf{w}$ (i.e. $\pi_{\mathbf{w}}(\mathbf{v}) \in W$) and $\pi_{\mathbf{w}^\perp} \perp \mathbf{w}$ (i.e. $\pi_{\mathbf{w}^\perp}(\mathbf{v}) \in W^\perp$).

PROOF Given $\mathbf{v}, \mathbf{w}$, set

$$\pi_{\mathbf{w}}(\mathbf{v}) = \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}}\mathbf{w} \quad \text{and } \pi_{\mathbf{w}^\perp}(\mathbf{v}) = \mathbf{v} - \pi_{\mathbf{w}}(\mathbf{v}).$$

Clearly $\pi_{\mathbf{w}}(\mathbf{v})||\mathbf{w}$ and $\pi_{\mathbf{w}}(\mathbf{v}) + \pi_{\mathbf{w}^\perp}(\mathbf{v}) = \mathbf{v}$. It is left to show that $\pi_{\mathbf{w}^\perp}(\mathbf{v}) \perp \mathbf{w}$:

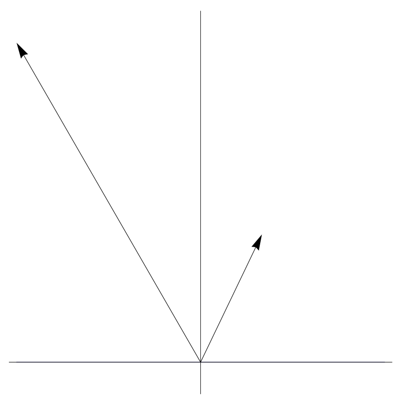
$$\begin{aligned}
\pi_{\mathbf{w}^\perp}(\mathbf{v}) \cdot \mathbf{w} = (\mathbf{v} - \pi_{\mathbf{w}}(\mathbf{v})) \cdot \mathbf{w} &= \left(\mathbf{v} - \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}}\mathbf{w}\right) \cdot \mathbf{w} \\
&= \mathbf{v} \cdot \mathbf{w} - \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}}\mathbf{w} \cdot \mathbf{w} \\
&= \mathbf{v} \cdot \mathbf{w} - \mathbf{v} \cdot \mathbf{w} \\
&= 0
\end{aligned}$$

Therefore $\pi_{\mathbf{w}^\perp}(\mathbf{v}) \perp \mathbf{w}$ as desired. □

Definition 4.21 *Let V be a dot product space, and let $\mathbf{v}, \mathbf{w} \in V$ with $\mathbf{w} \neq \mathbf{0}$. The* **projection of v onto w**, *denoted $\pi_{\mathbf{w}}\mathbf{v}$ or $proj_{\mathbf{w}}\mathbf{v}$ (and sometimes $\mathbf{v}^{\mathbf{w}}$), is the vector*

$$\pi_{\mathbf{w}}\mathbf{v} = proj_{\mathbf{w}}\mathbf{v} = \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}}\mathbf{w}.$$

Example: Find the projection of $(-3, 5)$ onto $(1, 2)$.



Example: Find the projection of $f(x) = x^2$ onto $g(x) = x$ in $C([0, 1], \mathbb{R})$.

Theorem 4.22 (Properties of projections) *Let V be a dot product space, and let $\mathbf{v}, \mathbf{w} \in V$ with $\mathbf{w} \neq \mathbf{0}$. Then:*

1. $\pi_{\mathbf{w}}\mathbf{v} = \frac{\mathbf{v}\cdot\mathbf{w}}{||\mathbf{w}||^2}\mathbf{w}$.
2. $\pi_{\mathbf{w}}\mathbf{v} = \frac{\mathbf{v}\cdot\mathbf{w}}{||\mathbf{w}||}\mathbf{u}$ *where* $\mathbf{u}$ *is a normalized version of* $\mathbf{w}$.
3. $\pi_{\mathbf{w}}\mathbf{v} = (\mathbf{v}\cdot\mathbf{u})\mathbf{u}$ *where* $\mathbf{u}$ *is a normalized version of* $\mathbf{w}$.
4. $\mathbf{v} \perp \mathbf{w}$ *if and only if* $\pi_{\mathbf{w}}\mathbf{v} = \mathbf{0}$.
5. $\mathbf{v}||\mathbf{w}$ *if and only if* $\pi_{\mathbf{w}}\mathbf{v} = \mathbf{v}$.
6. $\pi_{\mathbf{w}}\mathbf{v} \perp (\mathbf{v} - \pi_{\mathbf{w}}\mathbf{v})$.
7. $\pi_{\mathbf{w}}(\pi_{\mathbf{w}}\mathbf{v}) = \pi_{\mathbf{w}}\mathbf{v}$.

PROOF These are all calculations using the definitions. □

Theorem 4.23 (Orthogonal Decomposition Theorem (general case)) *Let V be a dot product space and let W be a finite-dimensional subspace of V. Then for any $\mathbf{v} \in V$, we can write*

$$\mathbf{v} = \pi_W(\mathbf{v}) + \pi_{W^\perp}(\mathbf{v}),$$

where $\pi_W(\mathbf{v}) \in W$ and $\pi_{W^\perp}(\mathbf{v}) \in W^\perp$. The $\pi_W(\mathbf{v})$ in this theorem is called the **projection of $\mathbf{v}$ onto** *W and the $\pi_{W^\perp}(\mathbf{v})$ is the* **component of $\mathbf{v}$ orthogonal to** *W.*

PROOF First, let $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ be a basis of W with two properties:

1. $||\mathbf{x}_j|| = 1$ for all j (i.e. $\mathbf{x}_j \cdot \mathbf{x}_j = 1$ for all j); and
2. $\mathbf{x}_i \perp \mathbf{x}_j$ for all $i \neq j$ (i.e. $\mathbf{x}_i \cdot \mathbf{x}_j = 0$ for all $i \neq j$).

Note: It is not clear that such a basis exists. We will prove in the next section that such a basis (called an *orthornormal basis* always exists.

Given such a basis, and given $\mathbf{v}$, define

$$\pi_W(\mathbf{v}) = \sum_{j=1}^{n}(\mathbf{v}\cdot\mathbf{x}_j)\mathbf{x}_j = (\mathbf{v}\cdot\mathbf{x}_1)\mathbf{x}_1 + (\mathbf{v}\cdot\mathbf{x}_2)\mathbf{x}_2 + ... + (\mathbf{v}\cdot\mathbf{x}_n)\mathbf{x}_n.$$

Since $\pi_W(\mathbf{v})$ is a combination of the $\mathbf{x}_j$, it is in $W = Span(\mathbf{x}_1, ..., \mathbf{x}_n)$ as desired.

Next, define

$$\pi_{W^\perp}(\mathbf{v}) = \mathbf{v} - \pi_W(\mathbf{v}).$$

This ensures that $\mathbf{v} = \pi_W(\mathbf{v}) + \pi_{W^\perp}(\mathbf{v})$, so all that is left to prove is that $\pi_{W^\perp}(\mathbf{v}) \in W^\perp$. To verify this, we will check that $\pi_{W^\perp}(\mathbf{v})$ is orthogonal to all the $\mathbf{x}_k$ (which is sufficient since the $\mathbf{x}_k$ span W):

$$\begin{aligned}
\pi_{W^\perp}(\mathbf{v}) \cdot \mathbf{x}_j &= (\mathbf{v} - \pi_W(\mathbf{v})) \cdot \mathbf{x}_k \\
&= \left(\mathbf{v} - \sum_{j=1}^{n} (\mathbf{v} \cdot \mathbf{x}_j)\mathbf{x}_j\right) \cdot \mathbf{x}_k \\
&= \mathbf{v} \cdot \mathbf{x}_k - \sum_{j=1}^{n} (\mathbf{v} \cdot \mathbf{x}_j)(\mathbf{x}_j \cdot \mathbf{x}_k) \\
&= \mathbf{v} \cdot \mathbf{x}_k - \mathbf{v} \cdot \mathbf{x}_k \\
&= 0.
\end{aligned}$$

Therefore $\pi_{W^\perp}(\mathbf{v})$ is orthogonal to all the $\mathbf{x}_j$, hence is in $W^\perp$ as desired. □

Theorem 4.24 (Uniqueness of orthogonal decompositions) *The $\pi_W(\mathbf{v})$ and the $\pi_{W^\perp}(\mathbf{v})$ described in Theorem 4.23 are unique (in other words, you don't get to choose between more than one possible $\pi_W(\mathbf{v})$s or more than one possible $\pi_{W^\perp}(\mathbf{v})$s).*

PROOF Suppose there are two vectors, say $\pi_W(\mathbf{v})$ and $\widehat{\pi}_W(\mathbf{v})$ which are "both" the projection of $\mathbf{v}$ onto subspace W. That means:

$$\pi_W(\mathbf{v}) \in W \qquad \widehat{\pi}_W(\mathbf{v}) \in W \qquad \mathbf{v} - \pi_W(\mathbf{v}) \in W^\perp \qquad \mathbf{v} - \widehat{\pi}_W(\mathbf{v}) \in W^\perp$$

Since W is a subspace, $\pi_W(\mathbf{v}) - \widehat{\pi}_W(\mathbf{v}) \in W$, and since $W^\perp$ is a subspace,

$$(\mathbf{v} - \pi_W(\mathbf{v})) - (\mathbf{v} - \widehat{\pi}_W(\mathbf{v})) = \pi_W(\mathbf{v}) - \widehat{\pi}_W(\mathbf{v}) \in W^\perp.$$

Since the dot product of anything in W with anything in $W^\perp$ is 0, we have

$$0 = (\pi_W(\mathbf{v}) - \widehat{\pi}_W(\mathbf{v})) \cdot (\pi_W(\mathbf{v}) - \widehat{\pi}_W(\mathbf{v})).$$

But by definiteness of the inner product, this implies $\pi_W(\mathbf{v}) = \widehat{\pi}_W(\mathbf{v})$. This means we didn't really have two different projections of $\mathbf{v}$ onto W; we only had one. □

Corollary 4.25 *Let V be a finite-dimensional inner product space and let W be a subspace of V. Given any basis $\mathcal{B}$ of W and any basis $\mathcal{B}^\perp$ of $W^\perp$, $\mathcal{B} \cup \mathcal{B}^\perp$ is a basis of V.*

PROOF We need to show that $\mathcal{B} \cup \mathcal{B}^\perp$ spans V, and that $\mathcal{B} \cup \mathcal{B}^\perp$ is linearly independent. Write $\mathcal{B} = \{\mathbf{w}_1, ..., \mathbf{w}_m\}$ and write $\mathcal{B}^\perp = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$.

To show $\mathcal{B} \cup \mathcal{B}^\perp$ spans V, let $\mathbf{v} \in V$. Then by the Orthogonal Decomposition Theorem,

$$\mathbf{v} = \pi_W(\mathbf{v}) + \pi_{W^\perp}(\mathbf{v}) = \sum_{i=1}^{m} c_j \mathbf{w}_j + \sum_{j=1}^{n} b_j \mathbf{v}_j.$$

Thus $\mathbf{v}$ can be written as a linear combination of vectors in $\mathcal{B} \cup \mathcal{B}^\perp$, so $\mathcal{B} \cup \mathcal{B}^\perp$ is a spanning set for V.

To show $\mathcal{B} \cup \mathcal{B}^\perp$ is a linearly independent set, suppose

$$\sum_{i=1}^{m} c_i \mathbf{w}_i + \sum_{j=1}^{n} b_j \mathbf{v}_j = \mathbf{0}.$$

Therefore

$$\sum_{i=1}^{m} c_i \mathbf{w}_i = -\sum_{j=1}^{n} b_j \mathbf{v}_j$$

Since the left-hand side of this equation belongs to W and the right-hand side belongs to $W^\perp$, both sides must belong to $W \cap W^\perp$, hence both sides are $\mathbf{0}$. But since $\mathcal{B}$ and $\mathcal{B}^\perp$ are bases, they are linearly independent, so all the b_j and c_i must be zero, hence $\mathcal{B} \cup \mathcal{B}^\perp$ is a linearly independent set (hence a basis). □

Corollary 4.26 (Dimension Formula) *Let V be a finite-dimensional inner product space and let W be a subspace of V. Then*

$$\dim V = \dim W + \dim(W^\perp).$$

Corollary 4.27 *Let V be a finite-dimensional inner product space and let W be a subspace of V. Then $(W^\perp)^\perp = W$.*

PROOF First, $\dim(W^\perp)^\perp + \dim W^\perp = \dim V$, so

$$\dim(W^\perp)^\perp = \dim V - \dim W^\perp = \dim W.$$

Now let $\mathbf{w} \in W$; $\mathbf{w} \perp \mathbf{v}$ for all $\mathbf{v} \in W^\perp$, so $\mathbf{w} \in (W^\perp)^\perp$. This shows $W \subseteq (W^\perp)^\perp$. But since these subspaces have the same dimension (last line on the previous page), they must be equal by an earlier theorem. □

Warning: If $\dim V = \infty$, the preceding corollary is false! It turns out that in this situation, $(W^\perp)^\perp$ can contain more vectors than W.

Definition 4.28 *Let V be a dot product space and let W be a subspace of V. Given any $\mathbf{v} \in V$, let the* **distance from $\mathbf{v}$ to** W*, denoted by $dist(\mathbf{v}, W)$, be the minimum distance from $\mathbf{v}$ to any vector in W.*

Theorem 4.29 *Let V be a dot product space and let W be a finite-dimensional subspace of V. Given any $\mathbf{v} \in V$, $dist(\mathbf{v}, W) = ||\pi_{W^\perp}(\mathbf{v})||$.*

PROOF Let $\mathbf{w} \in W$. Then $\pi_W(\mathbf{v}) - \mathbf{w}$ is in W, since it is the difference of two vectors in W and W is a subspace. Thus $(\pi_W(\mathbf{v}) - \mathbf{w}) \perp \pi_{W^\perp}(\mathbf{v})$ since $\pi_{W^\perp}(\mathbf{v}) \in W^\perp$. So by the Pythagorean Theorem,

$$||\pi_W(\mathbf{v}) - \mathbf{w}||^2 + ||\pi_{W^\perp}(\mathbf{v})||^2 = ||(\pi_W(\mathbf{v}) - \mathbf{w}) + \pi_{W^\perp}(\mathbf{v})||^2,$$

i.e.

$$||\pi_W(\mathbf{v}) - \mathbf{w}||^2 + ||\pi_{W^\perp}(\mathbf{v})||^2 = ||\mathbf{v} - w||^2,$$

so

$$||\pi_{W^\perp}(\mathbf{v})||^2 \leq ||\mathbf{v} - \mathbf{w}||^2.$$

Take the square root of both sides to get $||\pi_{w^\perp}(\mathbf{v})|| \leq dist(\mathbf{v}, \mathbf{w})$. Since this holds for any $\mathbf{w} \in W$, we have $||\pi_{W^\perp}(\mathbf{v})|| \leq dist(\mathbf{v}, W)$.

On the other hand, $||\pi_{W^\perp}(\mathbf{v})|| = ||\mathbf{v} - \pi_W(\mathbf{v})|| = dist(\mathbf{v}, \pi_W(\mathbf{v}))$ and since $\pi_W(\mathbf{v}) \in W$, this must be at least $dist(\mathbf{v}, W)$. So $||\pi_{W^\perp}(\mathbf{v})|| \geq dist(\mathbf{v}, W)$.

Putting the two inequalities from the preceding paragraphs together, we obtain the result. □

4.5 Orthonormal bases and the Gram-Schmidt procedure

Motivating example: Let $V = \mathbb{R}^4$ and let $W = Span((1, 2, 1, -1), (0, 2, 1, -2))$. Let $\mathbf{v} = (2, -1, 0, 0)$. What is $\pi_W(\mathbf{v})$?

We know (from the proof of the Orthogonal Decomposition Theorem) that if we had a basis $\{\mathbf{x}_1, \mathbf{x}_2\}$ of W with the properties that $||\mathbf{x}_j|| = 1$ for all j and $\mathbf{x}_i \perp \mathbf{x}_j$ for $i \neq j$, then

$$\pi_W(\mathbf{v}) = (\mathbf{v} \cdot \mathbf{x}_1)\mathbf{x}_1 + (\mathbf{v} \cdot \mathbf{x}_2)\mathbf{x}_2.$$

Question: How do you find $\{\mathbf{x}_1, \mathbf{x}_2\}$?

Definition 4.30 *Let V be a dot product space. A set $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ of vectors in V is called* **(pairwise) orthogonal** *if $\mathbf{x}_i \perp \mathbf{x}_j$ for all $i \neq j$. The set is called* **orthonormal** *if it is orthogonal and if $||\mathbf{x}_j|| = 1$ for all j.*

Example: $\{(2,3), (-3,2)\}$ is an orthogonal set of vectors in $\mathbb{R}^2$ which is not orthonormal.

Example: $\{\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)\}$ is an orthonormal set of vectors in $\mathbb{R}^2$.

Example: $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n\}$ is an orthonormal set of vectors in $\mathbb{R}^n$ where $\mathbf{e}_j = (0, 0, ..., 0, 1, 0, 0, .., 0)$ (the 1 is in the j^{th} position).

We are especially interested in **bases** that are orthornormal. This is because we can easily write a vector as a linear combination of vectors coming from an orthonormal basis. For example, consider $\mathbb{R}^3$ which has orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ where $\mathbf{e}_1 = (1,0,0)$; $\mathbf{e}_2 = (0,1,0)$; $\mathbf{e}_3 = (0,0,1)$.

Let $\mathbf{v} = (2, -3, 6)$. Then $\mathbf{v} =$

This generalizes:

Theorem 4.31 *Let $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ be an orthornormal basis of dot product space V. Then for every $\mathbf{v} \in V$, we have*

$$\mathbf{v} = \sum_{j=1}^{n} (\mathbf{v} \cdot \mathbf{x}_j)\mathbf{x}_j.$$

PROOF By the unique decomposition theorem, we know that we can write

$$\mathbf{v} = \sum_{k=1}^{n} c_k \mathbf{x}_k$$

uniquely. Now for each j,

$$\mathbf{v} \cdot \mathbf{x}_j = \sum_{k=1}^{n} c_k \mathbf{x}_k \cdot \mathbf{x}_j = \sum_{k=1}^{n} c_k (\mathbf{x}_k \cdot \mathbf{x}_j).$$

Since the $\mathbf{x}_j$ are orthonormal, we have $\mathbf{x}_k \cdot \mathbf{x}_j = 0$ if $\mathbf{x}_j \neq \mathbf{x}_k$ and $\mathbf{x}_j \cdot \mathbf{x}_j = ||\mathbf{x}_j||^2 = 1^2 = 1$, so the above expression becomes

$$\mathbf{v} \cdot \mathbf{x}_j = c_j.$$

Thus $\mathbf{v} = \sum_{k=1}^{n} c_k \mathbf{x}_k = \sum_{k=1}^{n} (\mathbf{v} \cdot \mathbf{x}_k)\, \mathbf{x}_k$ as desired. □

From the proof of the Orthogonal Decomposition Theorem, we get the following formula:

Corollary 4.32 (Projection formula) *Let V be an inner product space and let $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ be an orthornormal basis of a finite-dimensional subspace W. Then for every $\mathbf{v} \in V$, we have*

$$\pi_W(\mathbf{v}) = \sum_{j=1}^{n} (\mathbf{v} \cdot \mathbf{x}_j)\mathbf{x}_j.$$

This corollary gives a procedure to compute projections of vectors onto subspaces:

1. Find an orthonormal basis of the subspace; then
2. Apply the formula in the preceding corollary.

The problem is: how do we find orthonormal bases? Fortunately, there is a well-known procedure for constructing an orthonormal basis, given any basis. This procedure is called the *Gram-Schmidt procedure*.

Theorem 4.33 (Gram-Schmidt Theorem) *Given a basis $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ of a dot product space V (V could be a subspace of some other dot product space), there is a basis $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ of V such that*

1. *$\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ is an orthonormal basis, and*
2. *$Span(\mathbf{x}_1, ..., \mathbf{x}_k) = Span(\mathbf{v}_1, ..., \mathbf{v}_k)$ for all $k \leq n$.*

PROOF The orthonormal basis $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ is constructed by the Gram-Schmidt procedure described below. □

The following corollary fills in the gap in the proof of the Orthogonal Decomposition Theorem:

Corollary 4.34 *Every finite-dimensional dot product space has an orthornormal basis.*

Gram-Schmidt procedure

This series of steps constructs an orthonormal basis $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ from the given $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$:

Step 1: Define

$$\begin{aligned}
\mathbf{w}_1 &= \mathbf{v}_1 \\
\mathbf{w}_2 &= \mathbf{v}_2 - proj_{\mathbf{w}_1}\mathbf{v}_2 \\
\mathbf{w}_3 &= \mathbf{v}_3 - proj_{\mathbf{w}_1}\mathbf{v}_3 - proj_{\mathbf{w}_2}\mathbf{v}_3 \\
\vdots\ &\ \vdots \\
\mathbf{w}_n &= \mathbf{v}_n - proj_{\mathbf{w}_1}\mathbf{v}_n - proj_{\mathbf{w}_2}\mathbf{v}_n - ... - proj_{\mathbf{w}_{n-1}}\mathbf{v}_n
\end{aligned}$$

This produces an orthogonal set $\{\mathbf{w}_1, ..., \mathbf{w}_n\}$ which is a basis.

Step 2: convert the orthogonal set to an orthonormal set by normalizing: for each j, set $\mathbf{x}_j = \frac{\mathbf{w}_j}{||\mathbf{w}_j||}$.

Back to the motivating example: Let $V = \mathbb{R}^4$ and let $W = Span((1,2,1,-1), (0,2,1,-2))$. Let $\mathbf{v} = (2,-1,0,0)$. How can we write $\mathbf{v}$ as $\pi_W(\mathbf{v}) + \pi_{W^\perp}(\mathbf{v})$ where $\pi_W(\mathbf{v}) \in W$; $\pi_{W^\perp}(\mathbf{v}) \in W^\perp$?

Step 1: Use Gram-Schmidt to find orthonormal basis of W.

Step 2: Use projection formula to find $\pi_W(\mathbf{v})$.

From the previous page, we have the following orthonormal basis of W:

$$\{\left(\frac{1}{\sqrt{7}}, \frac{2}{\sqrt{7}}, \frac{1}{\sqrt{7}}, \frac{-1}{\sqrt{7}}\right), \left(\frac{-1}{\sqrt{2}}, 0, 0, \frac{-1}{\sqrt{2}}\right)\}$$

Step 3: Find $\pi_{W^\perp}(\mathbf{v})$ by setting $\pi_{W^\perp}(\mathbf{v}) = \mathbf{v} - \pi_W(\mathbf{v})$.

This gives us a (not very nice) method to compute projections of vectors onto subspaces. A less messy way to carry out the Gram-Schmidt procedure is to use the following command in *Mathematica*:

```
Orthogonalize[{{1,2,1,-1}, {0,2,1,-2}}]
```

Later in the course we will see another method of computing projections by hand that is less terrible.

4.6 Angles and the Cauchy-Schwarz inequality

Theorem 4.35 (Cauchy-Schwarz Inequality) *Let V be a dot product space. Then for any $\mathbf{v}, \mathbf{w} \in V$, $|\mathbf{v} \cdot \mathbf{w}| \leq ||\mathbf{v}||\,||\mathbf{w}||$.*

PROOF If $\mathbf{w} = \mathbf{0}$, then both sides of the C-S inequality are zero, so it is trivally true. Henceforth, assume $\mathbf{w} \neq \mathbf{0}$. Then

$$\begin{aligned}
0 \leq ||\mathbf{v} - \pi_{\mathbf{w}}\mathbf{v}||^2 &= (\mathbf{v} - \pi_{\mathbf{w}}\mathbf{v}) \cdot (\mathbf{v} - \pi_{\mathbf{w}}\mathbf{v}) \\
&= \mathbf{v} \cdot \mathbf{v} - 2\pi_{\mathbf{w}}\mathbf{v} \cdot \mathbf{v} + \pi_{\mathbf{w}}\mathbf{v} \cdot \pi_{\mathbf{w}}\mathbf{v} \\
&= ||\mathbf{v}||^2 - 2\frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}}\mathbf{w} \cdot \mathbf{v} + \left(\frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}}\mathbf{w}\right) \cdot \left(\frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}}\mathbf{w}\right) \\
&= ||\mathbf{v}||^2 - 2\frac{(\mathbf{v} \cdot \mathbf{w})^2}{\mathbf{w} \cdot \mathbf{w}} + \frac{(\mathbf{v} \cdot \mathbf{w})^2}{(\mathbf{w} \cdot \mathbf{w})^2}\mathbf{w} \cdot \mathbf{w} \\
&= ||\mathbf{v}||^2 - 2\frac{(\mathbf{v} \cdot \mathbf{w})^2}{\mathbf{w} \cdot \mathbf{w}} + \frac{(\mathbf{v} \cdot \mathbf{w})^2}{\mathbf{w} \cdot \mathbf{w}} \\
&= ||\mathbf{v}||^2 - \frac{(\mathbf{v} \cdot \mathbf{w})^2}{\mathbf{w} \cdot \mathbf{w}} \\
&= ||\mathbf{v}||^2 - \frac{(\mathbf{v} \cdot \mathbf{w})^2}{||\mathbf{w}||^2}.
\end{aligned}$$

Therefore

$$\frac{(\mathbf{v} \cdot \mathbf{w})^2}{||\mathbf{w}||^2} \leq ||\mathbf{v}||^2 \Rightarrow (\mathbf{v} \cdot \mathbf{w})^2 \leq ||\mathbf{v}||^2\,||\mathbf{w}||^2 \Rightarrow |\mathbf{v} \cdot \mathbf{w}| \leq ||\mathbf{v}||\,||\mathbf{w}||. \square$$

We are now able to prove the following result, which was mentioned earlier but not proven:

Theorem 4.36 (Triangle Inequality) *Let V be a dot product space. Then for all $\mathbf{v}, \mathbf{w} \in V$, $||\mathbf{v} + \mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||$.*

PROOF

$$\begin{aligned}
\mathbf{v} \cdot \mathbf{w} &\leq ||\mathbf{v}||\,||\mathbf{w}|| \quad \text{(C-S Inequality)} \\
\Rightarrow 2\mathbf{v} \cdot \mathbf{w} &\leq 2||\mathbf{v}||\,||\mathbf{w}|| \\
\Rightarrow \mathbf{v} \cdot \mathbf{v} + 2\mathbf{v} \cdot \mathbf{w} + \mathbf{w} \cdot \mathbf{w} &\leq ||\mathbf{v}||^2 + 2||\mathbf{v}||\,||\mathbf{w}|| + ||\mathbf{w}||^2 \\
\Rightarrow (\mathbf{v} + \mathbf{w}) \cdot (\mathbf{v} + \mathbf{w}) &\leq (||\mathbf{v}|| + ||\mathbf{w}||)^2 \\
\Rightarrow ||\mathbf{v} + \mathbf{w}||^2 &\leq (||\mathbf{v}|| + ||\mathbf{w}||)^2.
\end{aligned}$$

Take the square root of both sides to get the desired result. $\square$

Corollary 4.37 (Generalized Triangle Inequality) *Let V be a dot product space. Then for all $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n \in V$,*

$$\left|\left|\sum_{j=1}^{n} \mathbf{v}_j\right|\right| \leq \sum_{j=1}^{n} ||\mathbf{v}_j||.$$

Theorem 4.38 (Triangle Inequality) *Let V be a dot product space. Then for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, $dist(\mathbf{v}, \mathbf{w}) \leq dist(\mathbf{v}, \mathbf{u}) + dist(\mathbf{u}, \mathbf{w})$.*

PROOF Set $\mathbf{x} = \mathbf{v} - \mathbf{u}$ and $\mathbf{y} = \mathbf{u} - \mathbf{w}$ so that $\mathbf{x} + \mathbf{y} = \mathbf{v} - \mathbf{w}$. Then

$$\begin{aligned} dist(\mathbf{v}, \mathbf{w}) = ||\mathbf{v} - \mathbf{w}|| = ||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}|| &= ||\mathbf{v} - \mathbf{u}|| + ||\mathbf{u} - \mathbf{w}|| \\ &= dist(\mathbf{v}, \mathbf{u}) + dist(\mathbf{u}, \mathbf{w}). \quad \square \end{aligned}$$

Definition 4.39 *Let V be a dot product space. Given nonzero vectors $\mathbf{v}, \mathbf{w} \in V$, the* **angle between v and w** *is*

$$\theta = \arccos\left(\frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{v}||\,||\mathbf{w}||}\right).$$

Note: $\frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{v}||\,||\mathbf{w}||}$ is guaranteed to be between -1 and 1 by the C-S Inequality.

Rewritten, this says $\cos\theta = \frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{v}||\,||\mathbf{w}||}$.

Rewritten further, this says $\mathbf{v} \cdot \mathbf{w} = ||\mathbf{v}||\,||\mathbf{w}|| \cos\theta$. (This formula should look familiar if you have already taken Calculus III.)

Example: Find the angle between the vectors $(0, 1)$ and $(1, 1)$ in $\mathbb{R}^2$.

4.7 Hyperplanes

Definition 4.40 *Let V be a real vector space of dimension $n < \infty$. A* **hyperplane** *is an affine subspace of V whose dimension is $n-1$.*

Examples:

- In $\mathbb{R}$ (dimension 1), hyperplanes
- In $\mathbb{R}^2$ (dimension 2), hyperplanes
- In $\mathbb{R}^3$ (dimension 3), hyperplanes

Theorem 4.41 *Let V be a finite-dimensional dot product space and let H be a hyperplane. Then there is a vector $\mathbf{n} \in V$, called a* normal vector *to the hyperplane, and a scalar $d \in \mathbb{R}$ such that*

$$\mathbf{x} \in H \iff \mathbf{n} \cdot \mathbf{x} = d.$$

The equation $\mathbf{n} \cdot \mathbf{x} = d$ is called the **normal equation** *or* **standard equation** *of the hyperplane.*

PROOF Let W be the subspace of V associated to H. Then $\dim W^\perp = \dim V - \dim W = n - (n-1) = 1$, so $W^\perp$ is spanned by a single nonzero vector. Call this vector $\mathbf{n}$.

Now pick a fixed $\mathbf{x}_0 \in H$ (so that $H = W + \mathbf{x}_0$) and let $d = \mathbf{n} \cdot \mathbf{x}_0$. We claim that for this choice of $\mathbf{n}$ and d, $\mathbf{x} \in H \iff \mathbf{n} \cdot \mathbf{x} = d$.

($\Rightarrow$) Suppose $\mathbf{x} \in H$. Then $\mathbf{x} - \mathbf{x}_0 \in W$ so $(\mathbf{x} - \mathbf{x}_0) \perp \mathbf{n}$. Thus $(\mathbf{x} - \mathbf{x}_0) \cdot \mathbf{n} = 0$ so $\mathbf{x} \cdot \mathbf{n} = \mathbf{x}_0 \cdot \mathbf{n} = d$.

($\Leftarrow$) Suppose $\mathbf{n} \cdot \mathbf{x} = d$. Then $\mathbf{n} \cdot (\mathbf{x} - \mathbf{x}_0) = d - d = 0$ so $\mathbf{x} - \mathbf{x}_0 \in W$. Thus $\mathbf{x} \in W + \mathbf{x}_0 = H$. $\square$

Example 1: Let $V = \mathbb{R}^2$ and let H be the hyperplane $y = -2x + 4$.

Example 2: Let $V = \mathbb{R}^3$ and let H be the hyperplane passing through $(0, 1, 4)$, $(-2, -3, 8)$ and $(2, 2, -1)$.

Example 3: Find the normal equation of the plane passing through $(3, -1, 2)$ and containing the line whose parametric equations are $x = 2t + 1, y = t - 5, z = -7t + 4$.

Example 4: The two planes with equations $x + y + z = 3$ and $2x - 3y + 4z = 9$ intersect in a line. Find parametric equations of this line.

Definition 4.42 *Let* $\mathbf{a} = (a_1, a_2, a_3)$ *and* $\mathbf{b} = (b_1, b_2, b_3)$ *be two vectors in* $\mathbb{R}^3$*. The* **cross product** *of* $\mathbf{a}$ *and* $\mathbf{b}$*, denoted* $\mathbf{a} \times \mathbf{b}$*, is the vector*

$$\mathbf{a} \times \mathbf{b} = (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1).$$

Example: $(1, 4, -3) \times (2, -1, 3) =$

Theorem 4.43 (Properties of cross product) *Let* $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$*. Then:*

1. $\mathbf{a}||\mathbf{b}$ *if and only if* $\mathbf{a} \times \mathbf{b} = \mathbf{0}$*;*
2. $(\mathbf{a} \times \mathbf{b}) \perp \mathbf{a}$ *and* $(\mathbf{a} \times \mathbf{b}) \perp \mathbf{b}$*;*
3. $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$*;*
4. $(\mathbf{a} + \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \times \mathbf{c}) + (\mathbf{b} \times \mathbf{c})$ *and* $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$*;*
5. $r\mathbf{a} \times \mathbf{b} = r(\mathbf{a} \times \mathbf{b}) = \mathbf{a} \times r\mathbf{b}$ *for any* $r \in \mathbb{R}$*;*
6. *If* $\mathbf{a}$ *and* $\mathbf{b}$ *are not parallel, then* $(Span(\mathbf{a}, \mathbf{b}))^{\perp} = Span(\mathbf{a} \times \mathbf{b})$*;*
7. $||\mathbf{a} \times \mathbf{b}|| = ||\mathbf{a}||\,||\mathbf{b}|| \sin\theta$ *where* θ *is the angle between* $\mathbf{a}$ *and* $\mathbf{b}$*.*

PROOF These are all direct calculations using the definition; the important fact in linear algebra is statement (2), so we will prove the first part of that statement here:

$$\begin{aligned}(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{a} &= (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1) \cdot (a_1, a_2, a_3) \\ &= a_1(a_2b_3 - a_3b_2) + a_2(a_3b_1 - a_1b_3) + a_3(a_1b_2 - a_2b_1) \\ &= a_1a_2b_3 - a_1b_2a_3 + b_1a_2a_3 - a_1a_2b_3 + a_1b_2a_3 - b_1a_2a_3 \\ &= 0. \ \square\end{aligned}$$

Example 2 revisited: Let $V = \mathbb{R}^3$ and let H be the hyperplane passing through $(0,1,4)$, $(-2,-3,8)$ and $(2,2,-1)$.

4.8 Summary of Chapter 4

1. The dot product of two vectors is a scalar. Each vector space has its own formula for dot product; examples of dot product include:

   - if $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, $\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \mathbf{w} = \sum_{j=1}^{n} v_j w_j$.
   - if $f, g \in C([a,b], \mathbb{R})$, $f \cdot g = \int_a^b f(x)g(x)\, dx$.
   - if $A, B \in M_n(\mathbb{R})$, $A \cdot B = tr(AB^T)$.

   There are vector spaces with no dot product, but if V has a dot product formula, then we call V a **dot product space**.

2. Dot products are useful because they lead to geometric descriptions of vectors. In particular, if V is a dot product space we can define:

   a) The **norm** (length) of a vector $\mathbf{v} \in V$ is the scalar $||\mathbf{v}|| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$.
   (In other words, $\mathbf{v} \cdot \mathbf{v} = ||\mathbf{v}||^2$.)

   b) The **distance** between two vectors $\mathbf{v}, \mathbf{w} \in V$ is the scalar $dist(\mathbf{v}, \mathbf{w}) = ||\mathbf{v} - \mathbf{w}||$.

   c) A **unit vector** is a vector whose norm is 1.
   Given any nonzero $\mathbf{v}$, there is always a unit vector $\mathbf{u}$ in the same direction as $\mathbf{v}$, called a **normalized version of** $\mathbf{v}$, which is computed by setting $\mathbf{u} = \frac{1}{||\mathbf{v}||}\mathbf{v}$.

   d) Two vectors $\mathbf{v}$ and $\mathbf{w}$ are called **orthogonal** (written $\mathbf{v} \perp \mathbf{w}$) if $\mathbf{v} \cdot \mathbf{w} = 0$. Orthogonality generalizes the idea of "perpendicularity" from planar geometry.

   e) The **angle** θ between nonzero vectors $\mathbf{v}$ and $\mathbf{w}$ is $\theta = \arccos\left(\frac{\mathbf{v} \cdot \mathbf{w}}{||\mathbf{v}||\,||\mathbf{w}||}\right)$.
   (In other words, $\mathbf{v} \cdot \mathbf{w} = ||\mathbf{v}||\,||\mathbf{w}|| \cos\theta$.)
   The reason we can define angles this way is because of the Cauchy-Schwarz Inequality, which says that for any $\mathbf{v}, \mathbf{w} \in V$,

   $$|\mathbf{v} \cdot \mathbf{w}| \leq ||\mathbf{v}||\,||\mathbf{w}||.$$

3. Dot product in $\mathbb{R}^m$ and $\mathbb{R}^n$ satisfies the important **dual relation:** if $A \in M_{mn}(\mathbb{R})$, $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$, then

   $$A\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot A^T \mathbf{y}.$$

4. A basis $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ of dot product space (or subspace) V is called **orthonormal** if the vectors in the basis are pairwise orthogonal, and if every vector in the basis is a unit vector.

a) If $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ is an orthonormal basis of V, then every $\mathbf{v} \in \mathbf{v}$ satisfies

$$\mathbf{v} = \sum_{j=1}^{n} (\mathbf{v} \cdot \mathbf{x}_j)\mathbf{x}_j.$$

b) To find an orthonormal basis of V, start with any basis $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ of V and apply the **Gram-Schmidt procedure** to convert this basis into an orthonormal basis.

5. Given a dot product space V and a subspace W of V, define the **orthogonal complement** $W^\perp$ to be the set of vectors which are orthogonal to every $\mathbf{w} \in W$.

   - $W^\perp$ is a subspace of V, and $\dim V = \dim W + \dim W^\perp$.
   - $W \cap W^\perp = \{\mathbf{0}\}$.
   - $(W^\perp)^\perp = W$ if V is finite dimensional.
   - $\mathbf{v} \in W^\perp$ if and only if $\mathbf{v}$ is orthogonal to every vector in a spanning set (or basis) of W.
   - Given any $\mathbf{v} \in V$, we can write $\mathbf{v} = \pi_W(\mathbf{v}) + \pi_{W^\perp}(\mathbf{v})$ where $\pi_{W^\perp}(\mathbf{v}) \in W$ and $\pi_W(\mathbf{v}) \in W^\perp$.
     $\pi_W(\mathbf{v})$ is called the **projection of** $\mathbf{v}$ **onto** W.

     a) If $\dim W = 1$, then $W = Span(\mathbf{w})$ for some nonzero $\mathbf{w}$; in this case

     $$\pi_W(\mathbf{v}) = \pi_{\mathbf{w}}\mathbf{v} = \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}}\mathbf{w}.$$

     b) In general, $\pi_W(\mathbf{v})$ can be computed by taking a basis $\{\mathbf{w}_1, ..., \mathbf{w}_n\}$ of W, converting that basis into an orthonormal basis $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ via Gram-Schmidt, and then setting

     $$\pi_W(\mathbf{v}) = \sum_{j=1}^{n} (\mathbf{v} \cdot \mathbf{x}_j)\mathbf{x}_j.$$

     c) To compute $\pi_{W^\perp}(\mathbf{v})$, first compute $\pi_W(\mathbf{v})$ and then set

     $$\pi_{W^\perp}(\mathbf{v}) = \mathbf{v} - \pi_W(\mathbf{v}).$$

     d) $dist(\mathbf{v}, W)$, the minimum distance from vector $\mathbf{v}$ to subspace W, satisfies $dist(\mathbf{v}, W) = ||\pi_{W^\perp}(\mathbf{v})||$.

6. A **hyperplane** in vector space V is an affine subspace whose dimension is 1 less than the dimension of V.

a) If V is a dot product space, then every hyperplane has a **normal equation** $\mathbf{n} \cdot \mathbf{x} = d$.

   $\mathbf{n}$ is a called a **normal vector** to the hyperplane.

b) If $V = \mathbb{R}^2$, hyperplanes are lines, and the normal equation of a line is $ax + by = d$. (Here, the normal vector is $\mathbf{n} = (a, b)$.)

c) If $V = \mathbb{R}^3$, hyperplanes are planes, and the normal equation of a plane is

$$ax + by + cz = d$$

   (here, the normal vector is $\mathbf{n} = (a, b, c)$).

   Given two nonparallel vectors $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^3$, the normal vector $\mathbf{n}$ to the plane containing $\mathbf{a}$ and $\mathbf{b}$ is $\mathbf{n} = \mathbf{a} \times \mathbf{b}$.

Chapter 5

Linear transformations

5.1 Introduction

Recall: In Chapter 1 we discussed the problem of solving a system of several equations in several variables. Here is one such system:

$$\begin{cases} 3w + 2x - y + 5z = 8 \\ 2w - x + 4y - z = -3 \\ w + 5x + 2y + 6z = 17 \end{cases}$$

We are interested in the following questions about this system:

1.

2.

3.

Big picture conceptual approach: think of the left hand side of this system **as a single function** called T and we think of the four variables (w, x, y, z) **as a single variable** $\mathbf{x}$. Last, we group the right-hand side of the equation $(8, -3, 17)$ into a single object called $\mathbf{b}$. Then the system above **becomes a single equation**

$$T(\mathbf{x}) = \mathbf{b}.$$

At this point in the course, we know quite a bit about $\mathbf{x}$ and $\mathbf{b}$: $\mathbf{x} = (w, x, y, z)$ is a vector in $\mathbb{R}^4$ and $\mathbf{b} = (8, -3, 17)$ is a vector in $\mathbb{R}^3$. In Chapters 3 and 4 we studied vector spaces (and subspaces, and dimension, and orthogonality, etc.) in detail, so we have a good idea of how to think about the $\mathbf{x}$ and the $\mathbf{b}$. What we don't know much about yet is the T. The only thing we know is that

It turns out that this T is something called a *linear transformation*. Every linear transformation has four subspaces associated to it which determine the answers to the three questions above (thus the need to fully understand subspaces). These four subspaces come in two pairs of orthogonal complements (thus the need to understand dot products and orthogonality).

Developing an understanding of linear transformations and these associated subspaces is the content of this chapter. For now, we start with a definition:

Definition 5.1 *Let V_1 and V_2 be real vector spaces. A function $T : V_1 \to V_2$ is called a* **linear transformation** *if*

1. T preserves addition, *i.e. for all* $\mathbf{v}, \mathbf{w} \in V_1$, $T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$.
2. T preserves scalar multiplication, *i.e. for all* $\mathbf{v} \in V_1$, *and all* $r \in \mathbb{R}$, $T(r\mathbf{v}) = rT(\mathbf{v})$.

Theorem 5.2 (Linear transformations preserve zero) *If $T : V_1 \to V_2$ is a linear transformation, then $T(\mathbf{0}) = \mathbf{0}$.*

PROOF Let $\mathbf{v} \in V$ be arbitrary. Then $T(\mathbf{0}) = T(0\mathbf{v}) = 0T(\mathbf{v}) = \mathbf{0}$. □

Question: What are all the linear transformations from $\mathbb{R}$ to $\mathbb{R}$?

We have proven:

Theorem 5.3 *Every linear transformation from $\mathbb{R}$ to $\mathbb{R}$ is of the form $T(x) = cx$, where $c \in \mathbb{R}$ is a constant.*

5.2 The standard matrix of a linear transformation

Question: What are all the linear transformations from $\mathbb{R}^2$ to $\mathbb{R}^3$?

Proceeding as on the previous page, suppose $T : \mathbb{R}^2 \to \mathbb{R}^3$ is linear. Let $(a, c, e) = T(1, 0)$ and let $(b, d, f) = T(0, 1)$. Then for any $(x, y) \in \mathbb{R}^2$,

$$\begin{aligned} T(x, y) = T(x, 0) + T(0, y) \\ &= x\,T(1, 0) + y\,T(0, 1) \\ &= x(a, c, e) + y(b, d, f) \\ &= (ax, cx, ex) + (by, dy, fy) \\ &= (ax + by, cx + dy, ex + fy). \end{aligned}$$

This shows that every linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ is of the form

$$T(x, y) = (ax + by, cx + dy, ex + fy)$$

for suitable constants a, b, c, d, e and f.

There is another way to think about this form that is useful:

The work on the previous page generalizes:

Definition 5.4 *The* **standard basis of** $\mathbb{R}^n$ *is the basis* $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n\}$ *(written in that order) of* $\mathbb{R}^n$ *where*

$$\mathbf{e}_1 = (1, 0, 0, ..., 0), \qquad \mathbf{e}_2 = (0, 1, 0, ..., 0), \quad \cdots \quad \mathbf{e}_n = (0, 0, ..., 0, 1)$$

In general, $\mathbf{e}_j$ *is a vector that has a* 1 *in the* j^{th} *position and* 0*s in all other positions.*

Notice that the standard basis is an orthonormal basis of $\mathbb{R}^n$.

Theorem 5.5 *Let* $T : \mathbb{R}^n \to \mathbb{R}^m$ *be a linear transformation. Then there is a matrix* $A \in M_{mn}(\mathbb{R})$*, called the* **standard matrix of** T *such that* $T(\mathbf{x}) = A\mathbf{x}$ *for all* $x \in \mathbb{R}^n$*. Furthermore, the columns of* A *are, in order,* $T(\mathbf{e}_1)$*,* $T(\mathbf{e}_2)$*, ..., and* $T(\mathbf{e}_n)$*.*

PROOF Let $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ be the standard basis of $\mathbb{R}^n$. Let

$$A = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}.$$

Now let $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$; we claim $T(\mathbf{x}) = A\mathbf{x}$. To show this, just compute:

$$\begin{aligned} A\mathbf{x} &= x_1 T(\mathbf{e}_1) + x_2 T(\mathbf{e}_2) + ... + x_n T(\mathbf{e}_n) \quad \text{(by def'n of matrix multiplication)} \\ &= T(x_1\mathbf{e}_1) + T(x_2\mathbf{e}_2) + ... + T(x_n\mathbf{e}_n) \quad \text{(since } T \text{ preserves } \cdot) \\ &= T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + ... + x_n\mathbf{e}_n) \quad \text{(since } T \text{ preserves } +) \\ &= T(\mathbf{x}). \; \square \end{aligned}$$

Consequence: Every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is really matrix multiplication, once you write down the standard matrix.

Example: Find the standard matrix A of each of the given linear transformations (you may assume without proof that these transformations are linear). Then compute $T(2,-1)$.

1. $T : \mathbb{R}^2 \to \mathbb{R}^2$ rotates the plane by angle θ radians counterclockwiseabout the origin.

2. $T : \mathbb{R}^n \to \mathbb{R}^n$ defined by $T(\mathbf{x}) = \mathbf{x}$ (the identity transformation).

3. $T : \mathbb{R}^2 \to \mathbb{R}^2$ which projects vectors onto the subspace spanned by $(4, 1)$.

4. $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x_1, x_2, x_3) = (3x_1 + x_3, x_1 - x_2 - x_3)$.

5. $T : \mathbb{R}^2 \to \mathbb{R}^2$ where T reflects points through the line $y = 2x$.

Some other examples involving standard matrices

Example: Suppose $T : \mathbb{R}^3 \to \mathbb{R}^2$ is a linear transformation such that $T(1,0,0) = (2,-3)$, $T(0,1,0) = (-2,5)$ and $T(0,0,1) = (1,4)$. Find $T(2,-1,4)$.

Example: Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation such that $T(1,5) = (2,-7)$ and $T(2,-1) = (2,-3)$. Find $T(3,2)$.

Compositions of linear transformations

Theorem 5.6 *Let $T_1 : V_1 \to V_2$ and $T_2 : V_2 \to V_3$ be linear transformations. Then the transformation $T_2 \circ T_1 : V_1 \to V_3$ is linear as well.*

PROOF First, we show $T_2 \circ T_1$ preserves addition: let $\mathbf{v}, \mathbf{w} \in V_1$; then

$$\begin{aligned}(T_2 \circ T_1)(\mathbf{v} + \mathbf{w}) &= T_2(T_1(\mathbf{v} + \mathbf{w})) + T_2(T_1(\mathbf{v}) + T_1(\mathbf{w})) \\ &= T_2(T_1(\mathbf{v})) + T_2(T_1(\mathbf{w})) \\ &= (T_2 \circ T_1)(\mathbf{v}) + (T_2 \circ T_1)(\mathbf{w}).\end{aligned}$$

Next, we show $T_2 \circ T_1$ preserves scalar multiplication: let $\mathbf{v} \in V_1$ and $r \in \mathbb{R}$; then

$$(T_2 \circ T_1)(r\mathbf{v}) = T_2(T_1(r\mathbf{v})) = T_2(rT_1(\mathbf{v})) = rT_2(T_1(\mathbf{v})) = r\,(T_2 \circ T_1)(\mathbf{v}).$$

By definition, $T_2 \circ T_1$ is linear, since it preserves $+$ and $\cdot$. $\square$

Theorem 5.7 *Let $T_1 : \mathbb{R}^n \to \mathbb{R}^m$ and $T_2 : \mathbb{R}^m \to \mathbb{R}^p$ be linear transformations with standard matrices $A_1 \in M_{mn}(\mathbb{R})$ and $A_2 \in M_{pm}(\mathbb{R})$, respectively. Then the standard matrix of $T_2 \circ T_1 : \mathbb{R}^n \to \mathbb{R}^p$ is $A_2 A_1 \in M_{pn}(\mathbb{R})$.*

PROOF Let $\{\mathbf{e}_j\}$ represent standard basis elements of $\mathbb{R}^n, \mathbb{R}^m, \mathbb{R}^p$, etc. Let the entries of A_1 be denoted by as and the entries of A_2 be denoted by bs. Then the $(i, j)-$entry of $A_2 A_1$, by the definition of matrix multiplication, is

$$\sum_{l=1}^{m} b_{il} a_{lj}.$$

Now,

$$\begin{aligned}[(T_2 \circ T_1)(\mathbf{e}_j)]_i = [T_2(T_1(\mathbf{e}_j))]_i &= \left[T_2(j^{th} \text{ column of } A)\right]_i \\ &= [T_2(a_{1j}, a_{2j}, ..., a_{mj})]_i \\ &= \left[T_2\left(\sum_{l=1}^{m} a_{lj}\mathbf{e}_l\right)\right]_i \\ &= \left[\sum_{l=1}^{m} a_{lj} T_2(\mathbf{e}_l)\right]_i \\ &= \sum_{l=1}^{m} a_{lj} \left[l^{th} \text{ column of } A_2\right]_i \\ &= \sum_{l=1}^{m} a_{lj} \left[(i, l) - \text{entry of } A_2\right]\end{aligned}$$

$$\begin{aligned}
[(T_2 \circ T_1)(\mathbf{e}_j)]_i &= \sum_{l=1}^{m} a_{lj} \left[(i,l) - \text{entry of } A_2\right] \qquad \text{(previous page)} \\
&= \sum_{l=1}^{m} a_{lj} b_{il} \\
&= \sum_{l=1}^{m} b_{il} a_{lj}. \\
&= (i,j) - \text{entry of matrix } A_2 A_1.
\end{aligned}$$

This proves that for all i and j, $[(T_2 \circ T_1)(\mathbf{e}_j)]_i$ is the $(i,j)-$entry of A_2A_1. Therefore the j^{th} column of the standard matrix of $T_2 \circ T_1$ is the j^{th} column of A_2A_1, so the standard matrix of $T_2 \circ T_1$ is A_2A_1. $\square$

Importance: This theorem explains why matrix multiplication has the complicated definition that it does.

Example: Find the standard matrix of the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ that first multiplies vectors by $\begin{pmatrix} 2 & -1 \\ 3 & 0 \end{pmatrix}$ and then rotates the plane by $\frac{\pi}{4}$ clockwise.

5.3 Proving that a transformation is linear

Recall that by definition, $T : V_1 \to V_2$ is a linear transformation if

1. T preserves addition

2. T preserves scalar multiplication

We also proved a theorem in Section 5.1 which shows that for any linear T,

3. T preserves $\mathbf{0}$ (i.e. $T(\mathbf{0}) = \mathbf{0}$).

These three facts give you a good mechanism to decide whether a given function is a linear transformation. If you are given a function $T : V_1 \to V_2$, ask the following questions (this is called the *brute-force* method for determining whether or not T is a linear transformation:

1. Does $T(\mathbf{0}) = \mathbf{0}$?
2. Does $T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$?
3. Does $T(r\mathbf{v}) = rT(\mathbf{v})$?

A "no" answer to any of these questions means T is not linear, in which case you need to come up with a specific counterexample. A "yes" answer to all three questions means T is linear.

Note the similarities between this procedure and the procedure to prove whether or not a subset of a vector space is a subspace. More generally, we have the outline of how to write proofs of linearity / nonlinearity described below:

How to write a proof that a transformation $T : V_1 \to V_2$ is linear:

You need to do **all** of these things:

1. Verify that $T(\mathbf{0}) = \mathbf{0}$.
2. Take two generic elements of V_1, say $\mathbf{v}$ and $\mathbf{w}$. Verify that
   $T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$.
3. Take a generic element of V_1, say $\mathbf{v}$, and multiply it by a generic constant (say r). Verify that $T(r\mathbf{v}) = rT(\mathbf{v})$.

How to write a proof that a transformation $T : V_1 \to V_2$ is NOT linear:

You need to do **one** of the following three things:

1. Explain why $T(\mathbf{0}) \neq \mathbf{0}$.
2. Alternatively, write down two **specific** elements of V_1 (i.e. with numbers) (call them $\mathbf{v}$ and $\mathbf{w}$) for which $T(\mathbf{v}+\mathbf{w}) \neq T(\mathbf{v}) + T(\mathbf{w})$.
3. Alternatively, write down a **specific** element $\mathbf{v}$ of V_1 (i.e. with numbers) and a **specific** scalar r (i.e. a number) such that $T(r\mathbf{v}) \neq rT(\mathbf{v})$.

Examples: In each of the following examples, you are given a function $T : V_1 \to V_2$, where V_1 and V_2 are real vector spaces. Determine, with justification, whether or not T is a linear transformation.

Ex. 1: $V_1 = V_2 = \mathbb{R}^3$. $T(x_1, x_2, x_3) = (3x_1 + x_2 - x_3, 0, x_2)$.

Ex. 2: $V_1 = V_2 = \mathbb{R}^2$. $T(x, y) = (5x - 1, y + x)$.

Ex. 3: $V_1 = V_2 = M_3(\mathbb{R})$. $T(A) = A^T$.

Ex. 4: $V_1 = C(\mathbb{R}, \mathbb{R})$; $V_2 = \mathbb{R}^2$. $T(f) = (f(2), 3f(0))$.

Ex. 5: $V_1 = V_2 = \mathbb{R}$; $T(x) = |x|$.

Ex. 6: $V_1 = M_4(\mathbb{R})$; $V_2 = \mathbb{R}^4$. $T(A) =$ the 3rd column of A.

Ex. 7: $V_1 = V_2 = \mathbb{R}^2$. $T(\mathbf{x})$ is the reflection of point $\mathbf{x}$ through the line $x = 3$.

Ex. 8: $V_1 =$ the space of differentiable functions from $\mathbb{R}$ to $\mathbb{R}$; $V_2 = \mathbb{R}$. $T(f) = f'(0)$.

5.4 Prototypical examples of linear transformations

1. Matrix multiplication (**MOST IMPORTANT EXAMPLE**)

   Every matrix $A \in M_{mn}(\mathbb{R})$ defines a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ by $T(\mathbf{x}) = A\mathbf{x}$.

2. Projection

   Let V be any vector space and let $\mathbf{w} \in V$ be nonzero. Then the map $T : V \to V$ defined by $T(\mathbf{v}) = \pi_{\mathbf{w}}\mathbf{v}$ is linear.

   REASON:
   - $T(\mathbf{0}) = \pi_{\mathbf{w}}\mathbf{0} = \mathbf{0}$
   - $T(\mathbf{v}_1 + \mathbf{v}_2) = \pi_{\mathbf{w}}(\mathbf{v}_1 + \mathbf{v}_2) = \pi_{\mathbf{w}}(\mathbf{v}_1) + \pi_{\mathbf{w}}(\mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$
   - $T(r\mathbf{v}_1) = \pi_{\mathbf{w}}(r\mathbf{v}_1) = r\,\pi_{\mathbf{w}}(\mathbf{v}_1) = rT(\mathbf{v}_1)$.

   (Projection onto a subspace (i.e. the map $\pi_W : \mathbf{v} \to \pi_W(\mathbf{v})$ for $W \subseteq V$) is also linear.)

3. Differentiation

   Let $V = C^\infty(\mathbb{R}, \mathbb{R})$, the space of functions which are infinitely differentiable. Then $T : V \to V$ defined by $T(f) = f'$ is linear.

   REASON:
   - $T(\mathbf{0}) = 0' = 0$
   - $T(f + g) = (f + g)' = f' + g' = T(f) + T(g)$
   - $T(rf) = (rf)' = r\,f' = rT(f)$.

4. Definite Integration

   Let $V = C([a,b],\mathbb{R})$, the space of continuous functions from $[a,b]$ to $\mathbb{R}$. Then $T: V \to \mathbb{R}$, defined by $T(f) = \int_a^b f(x)\,dx$ is linear.

   REASON: • $T(\mathbf{0}) = \displaystyle\int_a^b 0\,dx = 0$

   • $T(f+g) = \displaystyle\int_a^b [f(x)+g(x)]\ dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx = T(f)+T(g)$

   • $T(rf) = \displaystyle\int_a^b rf(x)\,dx = r\int_a^b f(x)\,dx = rT(f).$

5. Evaluation

   Let V be some vector space of functions and fix $a \in \mathbb{R}$. Then $T : V \to \mathbb{R}$, defined by $T(f) = f(a)$, is linear.

   REASON: • $T(\mathbf{0}) = 0(a) = 0$

   • $T(f+g) = (f+g)(a) = f(a)+g(a) = T(f)+T(g)$

   • $T(rf) = (rf)(a) = r\,f(a) = rT(f)$.

6. Dot product by a fixed vector

   Fix a vector $\mathbf{v} \in \mathbb{R}^n$. Then the transformations $T(\mathbf{x}) = \mathbf{v}\cdot\mathbf{x}$ and $T(\mathbf{x}) = \mathbf{x}\cdot\mathbf{v}$ are both linear transformations from $\mathbb{R}^n$ to $\mathbb{R}$ (the same kind of thing works for the other dot product spaces).

7. Certain geometric transformations like rotations, reflections, etc.

8. Compositions of linear transformations

   If $T_1 : V_1 \to V_2$ and $T_2 : V_2 \to V_3$ are linear, then $T_2 \circ T_1 : V_1 \to V_3$ is linear as well (proven earlier).

9. Sums of other linear transformations

Theorem 5.8 *Let $T_1 : V_1 \to V_2$ and $T_2 : V_1 \to V_2$ be linear transformations. Then the transformation $T_1 + T_2 : V_1 \to V_2$, defined by setting*

$$(T_1 + T_2)(\mathbf{v}) = T_1(\mathbf{v}) + T_2(\mathbf{v})$$

for all $\mathbf{v} \in V_1$, is linear.

PROOF First, we show $T_1 + T_2$ preserves addition: let $\mathbf{v}, \mathbf{w} \in V_1$; then

$$\begin{aligned}(T_1 + T_2)(\mathbf{v} + \mathbf{w}) = T_1(\mathbf{v} + \mathbf{w}) + T_2(\mathbf{v} + \mathbf{w}) &= T_1(\mathbf{v}) + T_1(\mathbf{w}) + T_2(\mathbf{v}) + T_2(\mathbf{w}) \\ &= (T_1 + T_2)(\mathbf{v}) + (T_1 + T_2)(\mathbf{w}).\end{aligned}$$

Next, we show $T_1 + T_2$ preserves scalar multiplication: let $\mathbf{v} \in V_1$ and $r \in \mathbb{R}$; then

$$(T_1 + T_2)(r\mathbf{v}) = T_1(r\mathbf{v}) + T_2(r\mathbf{v}) = rT_1(\mathbf{v}) + rT_2(\mathbf{v}) = r(T_1 + T_2)(\mathbf{v}).$$

10. Scalar multiples of other linear transformations

Theorem 5.9 *Let $T : V_1 \to V_2$ be a linear transformation and let $c \in \mathbb{R}$. Then the transformation $cT_1 : V_1 \to V_2$, defined by setting*

$$(cT_1)(\mathbf{v}) = c\,T_1(\mathbf{v})$$

for all $\mathbf{v} \in V_1$, is linear.

Proof: Similar to above (need to show cT_1 preserves addition and scalar multiplication).

Remark: Examples 9 and 10 above prove that for fixed vector spaces V_1 and V_2, the set of all linear transformations from V_1 to V_2 is itself a vector space, which is denoted $L(V_1, V_2)$.

Notation: Therefore, to say "$T \in L(V_1, V_2)$" means that T is a linear transformation from V_1 to V_2.

5.5 Kernels and images

Concept: A linear transformation $T : V_1 \to V_2$ determines some subspaces of V_1 and V_2. Understanding these subspaces tells you a lot about the linear transformation, and will tell you something about the solution of equations $T(\mathbf{x}) = \mathbf{b}$.

Definition 5.10 *Let $T : V_1 \to V_2$ be a linear transformation. The* **kernel** *of T, denoted $ker(T)$, is the subset of V_1 defined by*

$$ker(T) = \{\mathbf{v} \in V_1 : T(\mathbf{v}) = \mathbf{0}\}.$$

In English, the kernel of a linear transformation is the set of vectors in the domain which get mapped to $\mathbf{0}$ under the transformation.

Theorem 5.11 (Kernels are subspaces) *Let $T : V_1 \to V_2$ be a linear transformation. Then $ker(T)$ is a subspace of V_1.*

PROOF We show that $ker(T)$ has the three essential characteristics of subspaces:

- $T(\mathbf{0}) = \mathbf{0}$, so $\mathbf{0} \in ker(T)$.
- Let $\mathbf{v}, \mathbf{w} \in ker(T)$. Thus $T(\mathbf{v}) = \mathbf{0}$ and $T(\mathbf{w}) = \mathbf{0}$. Then $T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w}) = \mathbf{0} + \mathbf{0} = \mathbf{0}$ so $\mathbf{v} + \mathbf{w} \in ker(T)$. Therefore $ker(T)$ is closed under $+$.
- Let $\mathbf{v} \in ker(T)$ and let $r \in \mathbb{R}$. Then $T(r\mathbf{v}) = rT(\mathbf{v}) = r\mathbf{0} = \mathbf{0}$ so $r\mathbf{v} \in ker(T)$ so $ker(T)$ is closed under scalar $\cdot$.

Thus $ker(T)$ is a subspace of V_1. □

Definition 5.12 *Let $T : V_1 \to V_2$ be a linear transformation. The* **image** *of T, denoted $im(T)$ or $Im(T)$, is the subset of V_2 defined by*

$$im(T) = \{\mathbf{w} \in V_2 : \exists\, \mathbf{v} \in V_1 \text{ such that } T(\mathbf{v}) = \mathbf{w}\}.$$

In English, the image of a linear transformation is its range, i.e. is the set of outputs of the linear transformation.

Theorem 5.13 (Images are subspaces) *Let $T : V_1 \to V_2$ be a linear transformation. Then $im(T)$ is a subspace of V_2.*

PROOF We show that $im(T)$ has the three essential characteristics of subspaces:

- $T(\mathbf{0}) = \mathbf{0}$, so $\mathbf{0} \in im(T)$.

- Let $\mathbf{w}_1, \mathbf{w}_2 \in im(T)$. Thus there are vectors $\mathbf{v}_1, \mathbf{v}_2 \in V_1$ such that $T(\mathbf{v}_1) = \mathbf{w}_1$ and $T(\mathbf{v}_2) = \mathbf{w}_2$. Then $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) = \mathbf{w}_1 + \mathbf{w}_2$ so $\mathbf{w}_1 + \mathbf{w}_2 \in im(T)$. Therefore $im(T)$ is closed under $+$.

- Let $\mathbf{w} \in im(T)$ and let $r \in \mathbb{R}$. Then there is a vector $\mathbf{v} \in V_1$ such that $T(\mathbf{v}) = \mathbf{w}$; we see that $T(r\mathbf{v}) = rT(\mathbf{v}) = r\mathbf{w}$ so $r\mathbf{w} \in im(T)$ so $im(T)$ is closed under scalar $\cdot$.

Thus $im(T)$ is a subspace of V_2. $\square$

Definition 5.14 *Let $T : V_1 \to V_2$ be a linear transformation. The* **rank** *of T, denoted r or $r(T)$, is the dimension of the image of T.*

The key picture associated to linear transformations

A linear transformation $T : V_1 \to V_2$ **always** suggests the following picture:

Why do we care about kernels and images?

Suppose you have an equation of the form

$$T(\mathbf{x}) = \mathbf{b}$$

where $T : V_1 \to V_2$ is linear (the goal is to find $\mathbf{x} \in V_1$, given T and $\mathbf{b} \in V_2$).

What's the relevance of the image $im(T)$?

What's the relevance of the kernel $ker(T)$?

Theorem 5.15 (Existence/uniqueness of solutions) *Suppose $T : V_1 \to V_2$ is a linear transformation. Then the equation $T(\mathbf{x}) = \mathbf{b}$...*
... has no solution if $\mathbf{b} \notin im(T)$.
... has exactly one solution if $\mathbf{b} \in im(T)$ and $\ker(T) = \{\mathbf{0}\}$.
... has infinitely many solutions if $\mathbf{b} \in im(T)$ and $\ker(T) \neq \{\mathbf{0}\}$.

Theorem 5.16 (Description of solutions) *Suppose $T : V_1 \to V_2$ is a linear transformation and suppose $\mathbf{b} \in im(T)$. Then the set of solutions of the equation $T(\mathbf{x}) = \mathbf{b}$ is the affine subspace $\mathbf{x}_p + ker(T)$, where $\mathbf{x}_p$ is any* **particular solution** *of the equation.*

PROOF (of both theorems): First of all, if $\mathbf{b} \notin im(T)$, then clearly $T(\mathbf{x}) = \mathbf{b}$ has no solution (by definition of image).

Henceforth, we assume $\mathbf{b} \in im(T)$. By definition of image, there is $\mathbf{x}_p \in V_1$ such that $T(\mathbf{x}_p) = \mathbf{b}$. Let S be the set of solutions of $T(\mathbf{x}) = \mathbf{b}$; we claim $S = \mathbf{x}_p + ker(T)$ and prove this equality by showing each set is a subset of the other:

To show $S \subseteq \mathbf{x}_p + ker(T)$: Let $\mathbf{x} \in S$. Then $T(\mathbf{x}) = \mathbf{b}$. Thus

$$T(\mathbf{x} - \mathbf{x}_p) = T(\mathbf{x}) - T(\mathbf{x}_p) = \mathbf{b} - \mathbf{b} = \mathbf{0}$$

so $\mathbf{x} - \mathbf{x}_p \in ker(T)$. Thus $\mathbf{x} = \mathbf{x}_p + (x - \mathbf{x}_p) \in \mathbf{x}_p + ker(T)$.

To show $\mathbf{x}_p + ker(T) \subseteq S$: Let $\mathbf{x} \in \mathbf{x}_p + ker(T)$. Thus $\mathbf{x} = \mathbf{x}_p + \mathbf{k}$ where $T(\mathbf{k}) = \mathbf{0}$. Therefore

$$T(\mathbf{x}) = T(\mathbf{x}_p + \mathbf{k}) = T(\mathbf{x}_p) + T(\mathbf{k}) = \mathbf{b} + \mathbf{0} = \mathbf{b}$$

so $x \in S$.

We have shown that if $\mathbf{b} \in im(T)$, the solution set S is the affine subspace $\mathbf{x}_p + ker(T)$. If $ker(T) = \{\mathbf{0}\}$, then this affine subspace is a point, hence there is one solution. Otherwise, this affine subspace has dimension ≥ 1, so it is infinite, so there are infinitely many solutions. □

Theorem 5.17 (Image-Kernel Theorem) *Let V_1 and V_2 be vector spaces where V_1 is finite-dimensional. If $T : V_1 \to V_2$ is any linear transformation, then*

$$\dim(im(T)) + \dim(ker(T)) = \dim(V_1).$$

PROOF Let $T : V_1 \to V_2$ be linear and suppose that $\dim(ker(T)) = k$ and $\dim(V_1) = n$. Since $ker(T)$ is a subspace of V_1, $k \leq n$.

Case 1: $k = n$. In this case, since $\dim(ker(T)) = \dim(V_1)$ and $ker(T) \subseteq V_1$, we can conclude that $ker(T) = V_1$, which means that $T(\mathbf{x}) = \mathbf{0}$ for all $\mathbf{x} \in V_1$. Thus $im(T) = \{\mathbf{0}\}$ and $\dim(im(T)) = 0$, so

$$\dim(im(T)) + \dim(ker(T)) = 0 + n = n = \dim(V_1) \text{ as desired.}$$

Case 2: $k < n$. In this case, $\dim((ker(T))^{\perp}) = n - k > 0$. Write down a basis of $(ker(T))^{\perp}$ and call this basis $\{\mathbf{v}_1, ...\mathbf{v}_{n-k}\}$. Let $\{x_1, ..., \mathbf{x}_k\}$ be a basis of $ker(T)$, so that $\{\mathbf{v}_1, ..., v_{n-k}, \mathbf{x}_1, ..., \mathbf{x}_k\}$ forms a basis of V_1.

We claim that $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_{n-k})\}$ is a basis of $im(T)$.

- To show $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_{n-k})\}$ spans $im(T)$:
  Let $\mathbf{w} \in im(T)$. Then there exists $\mathbf{v} \in V_1$ such that $T(\mathbf{v}) = \mathbf{w}$.

$$\begin{aligned} \mathbf{v} &= c_1\mathbf{v}_1 + ... + c_{n-k}\mathbf{v}_{n-k} + d_1\mathbf{x}_1 + ... + d_k\mathbf{x}_k \\ \Rightarrow \mathbf{w} = T(\mathbf{v}) &= c_1T(\mathbf{v}_1) + ... + c_{n-k}T(\mathbf{v}_{n-k}) + d_1T(\mathbf{x}_1) + ... + d_kT(\mathbf{x}_k) \\ &= c_1T(\mathbf{v}_1) + ... + c_{n-k}T(\mathbf{v}_{n-k}) + \mathbf{0} + ... + \mathbf{0} \\ &= c_1T(\mathbf{v}_1) + ... + c_{n-k}T(\mathbf{v}_{n-k}). \end{aligned}$$

- To show $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_{n-k})\}$ is linearly independent:
  Suppose $c_1T(\mathbf{v}_1) + ... + c_{n-k}T(\mathbf{v}_{n-k}) = \mathbf{0}$. Then

$$T(c_1\mathbf{v}_1 + ... + c_{n-k}\mathbf{v}_{n-k}) = \mathbf{0}$$

so

$$c_1\mathbf{v}_1 + ... + c_{n-k}\mathbf{v}_{n-k} \in \ker(T)\bigcap(ker(T))^{\perp}.$$

Therefore $c_1\mathbf{v}_1 + ... + c_{n-k}\mathbf{v}_{n-k} = \mathbf{0}$ so all the c_j are 0 since the $\mathbf{v}_j$ are lin. indep. Therefore $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_{n-k})\}$ is linearly independent.

Since $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_{n-k})\}$ spans $im(T)$ and is linearly independent, this list forms a basis for $im(T)$ so $\dim(im(T)) = n - k$, the number of vectors in this basis.

To conclude, we have

$$\dim(im(T)) + \dim(ker(T)) = (n-k) + k = n = \dim(V_1)$$

as desired. □

Corollary 5.18 (Rank-Nullty Theorem, version 1) *Suppose* $T : V_1 \to V_2$ *is a linear transformation with* $\dim V_1 = n < \infty$ *and* $\dim V_2 = m < \infty$*. Suppose also that* T *has rank* r*. Then* $\dim(im(T)) = r$ *and* $\dim(ker(T)) = n - r$*.*

Updated version of the key picture

A linear transformation $T : V_1 \to V_2$ **always** suggests the following picture:

5.6 Injectivity, surjectivity, bijectivity and inverses

Definition 5.19 *Let X and Y be sets and let $f : X \to Y$ be a function.*

1. *f is called* **injective** *(or $1-1$ or* **one-to-one***) if whenever $f(x) = f(y)$, it must be the case that $x = y$.*

2. *f is called* **surjective** *(or* **onto***) if for every $y \in Y$, there is an $x \in X$ such that $f(x) = y$.*

3. *f is called* **bijective** *if it both injective and surjective.*

4. *f is called* **invertible** *if there is a function $f^{-1} : Y \to X$ such that $f^{-1}(f(x)) = x$ for all $x \in X$ and $f(f^{-1}(y)) = y$ for all $y \in Y$. In this setting the function f^{-1} is called an* **inverse** *of f.*

These definitions can be restated in the language of solving equations:

1. To say f is injective means that for every $y \in Y$, the equation $f(x) = y$ has at most one solution x.

2. To say f is surjective means that for every $y \in Y$, the equation $f(x) = y$ has at least one solution x.

3. Thus f is bijective if and only if for every $y \in Y$, the equation $f(x) = y$ has exactly one solution x.

Theorem 5.20 *A function is invertible if and only if it is bijective.*

PROOF ($\Leftarrow$) Suppose f is bijective. Then for every $y \in Y$, the equation $f(x) = y$ has exactly one solution x; define $f^{-1} : Y \to X$ by setting $f^{-1}(y)$ to be the solution x of $f(x) = y$. It is easy to see that $f^{-1}(f(x)) = x$ and $f(f^{-1}(y)) = y$, so f is invertible with inverse f^{-1}.

($\Rightarrow$) Now suppose f is invertible. Then the equation $f(x) = y$ has exactly one solution, namely $x = f^{-1}(y)$, so f is bijective. □

Putting all this together, we have the following results:

Theorem 5.21 (Equivalent properties to injectivity) *Let $T : V_1 \to V_2$ be a linear transformation. Then the following are equivalent (if any one of these is true, they are all true; if any one is false, then they are all false):*

1. *T is injective.*
2. *The equation $T(\mathbf{x}) = \mathbf{b}$ always has at most one solution, no matter what $\mathbf{b} \in V_2$ is.*
3. *$ker(T) = \{\mathbf{0}\}$.*

Moreover, it is possible for T to be injective only if $\dim V_1 \leq \dim V_2$.

PROOF (1) and (2) are equivalent based on the remark on the previous page; (2) and (3) are equivalent by the Existence/Uniqueness theorem on page 19.

To prove the last statement, if T is injective, by the Image-Kernel Theorem,

$$\begin{array}{ccccc} \dim(ker(T)) & + & \dim(im(T)) & = & \dim(V_1) \\ 0 & + & \dim(im(T)) & = & \dim(V_1) \end{array}$$

Therefore $\dim(im(T)) = \dim(V_1)$. Since $im(T)$ is a subspace of V_2, we have $dim V_2 \geq \dim(V_1)$. $\square$

Theorem 5.22 (Equivalent properties to surjectivity) *Let $T : V_1 \to V_2$ be a linear transformation. Then the following are equivalent:*

1. *T is surjective.*
2. *The equation $T(\mathbf{x}) = \mathbf{b}$ always has at least one solution, no matter what $\mathbf{b} \in V_2$ is.*
3. *$im(T) = V_2$.*

Moreover, it is possible for T to be surjective only if $\dim V_1 \geq \dim V_2$.

PROOF (1) and (2) are equivalent based on the remark on the previous page; (2) and (3) are equivalent by the Existence/Uniqueness theorem on page 19.

To prove the last statement, if T is surjective, by the Image-Kernel Theorem,

$$\begin{array}{ccccc} \dim(ker(T)) & + & \dim(im(T)) & = & \dim(V_1) \\ \dim(ker(T)) & + & \dim(V_2) & = & \dim(V_1) \end{array}$$

Therefore $\dim V_2 \leq \dim V_1$. $\square$

If a linear transformation is bijective, then we can say quite a bit more:

Theorem 5.23 (Equivalent properties to bijectivity) *Let $T : V_1 \to V_2$ be a linear transformation. Then the following are equivalent:*

1. *T is bijective.*
2. *T is invertible.*
3. *The equation $T(\mathbf{x}) = \mathbf{b}$ always has exactly one solution, no matter what $\mathbf{b} \in V_2$ is.*
4. *$im(T) = V_2$ and $ker(T) = \{\mathbf{0}\}$.*

Moreover, this situation is only possible if $\dim V_1 = \dim V_2$. In this situation, the inverse $T^{-1} : V_2 \to V_1$ is always a linear transformation, and the one and only solution of equation $T(\mathbf{x}) = \mathbf{b}$ is $\mathbf{x} = T^{-1}(\mathbf{b})$.

PROOF (1) and (2) are equivalent based on Theorem 5.20; (2) and (3) are equivalent, by the remark on page 118; (3) and (4) are equivalent by the Existence/Uniqueness theorem (Theorem 5.15). Thus statements (1)-(4) are equivalent.

Now, suppose T is bijective. Then $im(T) = V_2$ so $\dim(im(T)) = \dim(V_2) = m$, and $\ker(T) = \{\mathbf{0}\}$ so $\dim(ker(T)) = 0$. Therefore by the Image-Kernel Theorem,

$$\begin{array}{ccccc} \dim(ker(T)) & + & \dim(im(T)) & = & \dim(V_1) \\ 0 & + & m & = & n \end{array}$$

so $m = n$, i.e $\dim V_1 = \dim V_2$.

Next, we show that if T is invertible, then T^{-1} is linear. We show this by brute-force:

- $T^{-1}(\mathbf{0}) = \mathbf{0}$ since $T(\mathbf{0}) = \mathbf{0}$.
- Let $\mathbf{w}_1, \mathbf{w}_2 \in V_2$ be such that $T^{-1}(\mathbf{w}_1) = \mathbf{v}_1$ and $T^{-1}(\mathbf{w}_2) = \mathbf{v}_2$. Then $T(\mathbf{v}_1) = \mathbf{w}_1$ and $T(\mathbf{v}_2) = \mathbf{w}_2$; since T is linear, $T(\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{w}_1 + \mathbf{w}_2$ so $T^{-1}(\mathbf{w}_1 + \mathbf{w}_2) = \mathbf{v}_1 + \mathbf{v}_2 = T^{-1}(\mathbf{w}_1) + T^{-1}(\mathbf{w}_2)$ so T^{-1} preserves addition.
- Let $\mathbf{w} \in V_2$ be such that $T^{-1}(\mathbf{w}) = \mathbf{v}$ and let $r \in \mathbb{R}$. Since T is linear, $T(r\mathbf{v}) = r\mathbf{w}$ so $T^{-1}(r\mathbf{w}) = r\mathbf{v} = rT^{-1}(\mathbf{v})$, so T^{-1} preserves scalar multiplication.

By the brute-force method, T^{-1} is linear.

Last, suppose T is bijective and consider the equation $T(\mathbf{x}) = \mathbf{b}$. Apply the transformation T^{-1} to both sides to get $T^{-1}(T(\mathbf{x})) = T^{-1}(\mathbf{b})$, i.e. $\mathbf{x} = T^{-1}(\mathbf{b})$. □

Some examples: For each of these linear transformations $T : V_1 \to V_2$ (you can assume they are linear without proof), your directions are:

1. Find the dimension of, and a basis for, $ker(T)$.
2. Find the dimension of, and a basis for, $im(T)$.
3. Determine whether or not T is injective, surjective, bijective or neither.
4. Sketch the key picture associated to the transformation T.
5. Determine the possible number of solutions to $T(\mathbf{x}) = \mathbf{b}$ for various $\mathbf{b} \in V_2$.

Example A: Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be the transformation defined by

$$T(x_1, x_2, x_3) = (x_1, x_3).$$

Example B: Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation defined by $T(\mathbf{x}) = A\mathbf{x}$ where

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Example C: Let $V = \mathcal{P}^3$ be the space of polynomials with degree ≤ 3 and let $T : \mathcal{P}^3 \to \mathcal{P}^3$ be defined by $T(f) = f'$.

5.7 Fundamental Theorem of Linear Algebra

Recall that given a $m \times n$ matrix A, the transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by $T(\mathbf{x}) = A\mathbf{x}$ is an important example of a linear transformation. In this setting the kernel and image of T have synonyms:

Definition 5.24 *Let $A \in M_{mn}(\mathbb{R})$ (i.e. A is an $m \times n$ matrix of real numbers).*

1. *The* **null space** *of A, denoted $N(A)$, is the set of vectors $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{0}$. (In other words, $N(A) = ker(T)$ where $T : \mathbb{R}^n \to \mathbb{R}^m$ is defined by $T(\mathbf{x}) = A\mathbf{x}$.)*
2. *The* **column space** *of A, denoted $C(A)$, is the span of the columns of A.*
3. *The* **row space** *of A, denoted $R(A)$, is the span of the rows of A.*
4. *The* **left null space** *of A is the null space of A^T, i.e. is the set of vectors $y \in \mathbb{R}^m$ such that $A^T\mathbf{y} = \mathbf{0}$. (In other words, $N(A^T) = ker(T^*)$ where $T^* : \mathbb{R}^m \to \mathbb{R}^n$ is defined by $T^*(\mathbf{y}) = A^T\mathbf{y}$.)*

Collectively, these four spaces are called the **four fundamental subspaces** *associated to the matrix A.*

Example: Let $A = \begin{pmatrix} 2 & 4 & -3 \\ 1 & 2 & 1 \end{pmatrix}$.

Observe: given $m \times n$ matrix A:

- $R(A) = C(A^T)$
- $C(A) = R(A^T)$
- If $T : \mathbb{R}^n \to \mathbb{R}^m$ is defined by $T(\mathbf{x}) = A\mathbf{x}$, then $N(A) = ker(T)$.
- If $T : \mathbb{R}^n \to \mathbb{R}^m$ is defined by $T(\mathbf{x}) = A\mathbf{x}$, then $C(A) = im(T)$.
- $N(A)$ and $R(A)$ are subspaces of $\mathbb{R}^n$.
- $N(A^T)$ and $C(A)$ are subspaces of $\mathbb{R}^m$.

The reason these sets are subspaces is because $R(A)$ and $C(A)$ are defined as spans (hence automatically subspaces), and because $N(A)$ and $N(A^T)$ are kernels of linear transformations (hence automatically subspaces).

It turns out that these subspaces are two pairs of orthogonal complements:

Theorem 5.25 (Fundamental Theorem of Linear Algebra) *Let $A \in M_{mn}(\mathbb{R})$ (i.e. A is an $m \times n$ matrix of real numbers). Then*

1. $[R(A)]^\perp = N(A)$*; and*
2. $[C(A)]^\perp = N(A^T)$.

PROOF To prove statement (1), notice

$$\begin{aligned} \mathbf{x} \in N(A) &\Leftrightarrow A\mathbf{x} = \mathbf{0} \\ &\Leftrightarrow \mathbf{x} \perp \text{(any row of } A) \\ &\iff \mathbf{x} \in [R(A)]^\perp. \end{aligned}$$

To prove statement (2), apply the same argument to matrix A^T:

$$[C(A)]^\perp = [R(A^T)]^\perp = N(A^T). \ \square$$

Corollary 5.26 (Rank Theorem) *Let $A \in M_{mn}(\mathbb{R})$. Then the following numbers are all equal to the same number, called the* **rank** *of A and denoted r or $r(A)$:*

1. $\dim(C(A))$
2. $\dim(R(A))$
3. *the number of linearly independent columns of A*
4. *the number of linearly independent rows of A*
5. *the rank of T where $T : \mathbb{R}^n \to \mathbb{R}^m$ is defined by $T(\mathbf{x}) = A\mathbf{x}$.*

PROOF It is clear that (1) and (3) are equal, and that (2) and (4) are equal. Since $C(A) = im(T)$, (1) and (5) are equal. Last, we will show (1) and (2) are equal.

Define $T : \mathbb{R}^n \to \mathbb{R}^m$ by $T(\mathbf{x}) = A\mathbf{x}$. Then by the Image-Kernel Theorem,

$$\dim(im(T)) + \dim(ker(T)) = n$$

i.e.

$$\dim(C(A)) + \dim(N(A)) = n$$

i.e.

$$\dim(C(A)) = n - \dim(N(A)).$$

But at the same time, by the FTLA $[N(A)]^\perp = R(A)$ so

$$\dim(R(A)) = n - \dim(N(A))$$

Therefore $\dim(R(A)) = \dim(C(A))$ as desired. □

Corollary 5.27 (Dimensions of fundamental subspaces) *Let $A \in M_{mn}(\mathbb{R})$ have rank r. Then:*

1. *$C(A)$ is a r-dimensional subspace of $\mathbb{R}^m$.*
2. *$R(A)$ is a r-dimensional subspace of $\mathbb{R}^n$.*
3. *$N(A)$ is an $(n - r)$-dimensional subspace of $\mathbb{R}^n$.*
4. *$N(A^T)$ is an $(m - r)$-dimensional subspace of $\mathbb{R}^m$.*

PROOF (1) and (2) follows from the Rank Theorem; (3) and (4) follow from the Rank-Nullty Theorem. □

Example: $A = \begin{pmatrix} 2 & -1 & 3 & 1 \\ 6 & -3 & 9 & 3 \end{pmatrix}$

Example: $A = \begin{pmatrix} 1 & 3 & -1 & 0 \\ 2 & -5 & 4 & 0 \\ 0 & 1 & 4 & -2 \end{pmatrix}$

5.8 More on invertibility

Recall: T is invertible $\iff$ T is bijective $\iff$ $\begin{cases} ker(T) = \{\mathbf{0}\} \text{ and} \\ im(T) = V_2 \end{cases}$

Later, we'll be able to say more about when these two conditions hold. As a preview, suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ has standard matrix A. Then, on the one hand,

$$T \text{ is invertible} \iff \begin{cases} N(A) = 0 \\ C(A) = \mathbb{R}^m \end{cases}$$

and this can only hold if $r = m = n$ (i.e. A is a square matrix). But at the same time, by definition,

$$T \text{ is invertible} \iff T^{-1} : V_2 \to V_1 \text{ exists}$$

and $T^{-1} \circ T(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x}$; $T \circ T^{-1}(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x}$.

Let B be the standard matrix of T^{-1}. Then, since $T \circ T^{-1}$ and $T^{-1} \circ T$ are both the identity transformation, the standard matrix of both $T \circ T^{-1}$ and $T^{-1} \circ T$ is I. Therefore B is some matrix such that

Such a matrix B has another name. It is called the *inverse* of A and is denoted by A^{-1} (pronounced "A inverse").

Definition 5.28 *A matrix $A \in M_{mn}(\mathbb{R})$ is called* **invertible** *if there is another matrix $A^{-1} \in M_{nm}(\mathbb{R})$ and called an* **inverse (matrix)** *of A, such that*

$$AA^{-1} = I_m \text{ and } A^{-1}A = I_n.$$

Theorem 5.29 *A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is invertible if and only if its standard matrix is invertible.*

PROOF ($\Rightarrow$) was proven above.

($\Leftarrow$) If T has standard matrix A, and A is invertible with inverse A^{-1}, define $T^{-1} : \mathbb{R}^m \to \mathbb{R}^n$ by $T^{-1}(\mathbf{x}) = A^{-1}\mathbf{x}$; it's easy to check that T and T^{-1} are inverses. □

Corollary 5.30 *If a matrix is invertible, then it is square and has full rank (i.e. $r = m = n$). Conversely, any square matrix with full rank must be invertible.*

PROOF At the top of this page.

Theorem 5.31 (Properties of inverses) *Let $A, B \in M_n(\mathbb{R})$ be invertible. Then*

1. *A has only one inverse.*
2. *$(A^{-1})^{-1} = A$.*
3. *$(AB)^{-1} = B^{-1}A^{-1}$.*
4. *A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.*

PROOF To prove (1), suppose B and C are both inverses of A. Then $AB = AC = BA = CA = I$, so $AB = AC$. Multiply both sides of this equation on the left by B to get $BAB = BAC$, i.e. $IB = IC$ i.e. $B = C$. So A cannot have two different inverses.

(2) follows from the definition of inverse.

To prove (3), notice $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$ and $(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I$.

To prove (4), start with $AA^{-1} = I$ and take the transpose of both sides to get $[AA^{-1}]^T = I^T = I$, i.e. $(A^{-1})^T A^T = I$ so by definition of inverse, $(A^{-1})^T = (A^T)^{-1}$. $\square$

Question: When is a 1×1 matrix invertible? What is the inverse of a 1×1 matrix?

Theorem 5.32 (1×1 inverses) *Let $A = \begin{pmatrix} a \end{pmatrix}$. Then A is invertible $\iff a \neq 0$, in which case*

$$A^{-1} = \begin{pmatrix} \frac{1}{a} \end{pmatrix}.$$

Question: When is a 2×2 matrix invertible? What is the inverse of a 2×2 matrix?

Answer: Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and suppose A is invertible. Write $A^{-1} = \begin{pmatrix} w & x \\ y & z \end{pmatrix}$.

Theorem 5.33 (2 × 2 inverses) *Let* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. *Then* A *is invertible* $\iff$ $ad - bc \neq 0$, *in which case*

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Example: $A = \begin{pmatrix} 3 & 2 \\ 7 & 8 \end{pmatrix}$

Question: When is a 3×3 or larger matrix invertible?

Theoretical answer: an $n \times n$ matrix is invertible $\iff$ it has full rank $\iff$ $N(A) = \{\mathbf{0}\} \iff C(A) = \mathbb{R}^n$.

Practical answer: coming in Chapter 6.

5.9 Summary of Chapter 5

1. A transformation $T : V_1 \to V_2$ is called a **linear transformation** if T preserves addition and scalar multiplication, and if T preserves the zero vector. Typical examples of linear transformations include matrix multiplication, differentiation, integration, evaluation, projection, geometric transformations like rotations and reflections, and dot product by a fixed vector. Sums, scalar multiples and compositions of linear transformations are also linear transformations.

2. The **kernel** of a linear transformation is the set of vectors in V_1 which are mapped to $\mathbf{0}$ under the transformation. $ker(T)$ is always a subspace of V_1.

3. The **image** of a linear transformation is the set of vectors in V_2 which are actual outputs of the transformation.

4. If T is a linear transformation, then the equation $T(\mathbf{x}) = \mathbf{b}$...
   ... has no solution if $\mathbf{b} \notin im(T)$
   ... has exactly one solution if $\mathbf{b} \in im(T)$ and $ker(T) = \{\mathbf{0}\}$
   ... has infinitely many solutions if $\mathbf{b} \in im(T)$ and $ker(T) \neq \{\mathbf{0}\}$
   If $\mathbf{b} \in im(T)$, then the solution set of $T(\mathbf{x}) = \mathbf{b}$ is the affine subspace $\mathbf{x}_p + ker(T)$ where $\mathbf{x}_p$ is any particular solution of $T(\mathbf{x}) = \mathbf{b}$.

5. T is called **injective** if any of these equivalent conditions hold:
   - $ker(T) = \{\mathbf{0}\}$;
   - T takes different inputs to different outputs;
   - the equation $T(\mathbf{x}) = \mathbf{b}$ has at most one solution for every $\mathbf{b} \in V_2$.

   It is possible for T to be injective only if $\dim V_1 \leq \dim V_2$.

6. $im(T)$ is always a subspace of V_2. T is called **surjective** if any of these equivalent conditions hold:
   - $im(T) = V_2$;
   - the equation $T(\mathbf{x}) = \mathbf{b}$ has at least one solution for every $\mathbf{b} \in V_2$.

   It is possible for T to be surjective only if $\dim V_1 \geq \dim V_2$.

7. A transformation T is called **bijective** if it is both injective and surjective. The following are equivalent:
   - T is bijective;
   - T is invertible (in which case T^{-1} is linear);
   - the equation $T(\mathbf{x}) = \mathbf{b}$ always has exactly one solution which is $x = T^{-1}(\mathbf{b})$.

   It is possible for T to be bijective only if $\dim V_1 = \dim V_2$.

8. The dimensions of the kernel and image of a linear transformation always add to the dimension of the domain of the transformation.

9. Every $m \times n$ matrix A has four subspaces associated to it: the **null space**, **column space**, **row space** and **left nullspace**. There is a number r called the **rank** of the matrix such that
   - $C(A)$ is a r-dimensional subspace of $\mathbb{R}^m$;
   - $R(A)$ is a r-dimensional subspace of $\mathbb{R}^n$;
   - $N(A)$ is an $(n-r)$-dimensional subspace of $\mathbb{R}^n$;
   - $N(A^T)$ is an $(m-r)$-dimensional subspace of $\mathbb{R}^m$.
   - $[R(A)]^\perp = N(A)$ and $[C(A)]^\perp = N(A^T)$ (this is the Fundamental Theorem of Linear Algebra)

10. Every linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is of the form $T(\mathbf{x}) = A\mathbf{x}$ where A is an $m \times n$ matrix called the **standard matrix** of T. The columns of A are, from left to right, $T(\mathbf{e}_1)$, $T(\mathbf{e}_2)$, ..., $T(\mathbf{e}_n)$. In this setting,
    - $im(T)$ is the same thing as $C(A)$, the span of the columns of A.
    - $ker(T)$ is the same thing as $N(A)$, the null space of A.

11. A matrix A is called **invertible** if there is another matrix A^{-1} (called the inverse of A) such that $AA^{-1} = I$ and $A^{-1}A = I$. The following are equivalent:
    - A is invertible.
    - the linear transformation $T(\mathbf{x}) = A\mathbf{x}$ is invertible.
    - A is square and has full rank (i.e. $r = m = n$).

    A 1×1 matrix $A = (a)$ is invertible if and only if $a \neq 0$, in which case $A^{-1} = \left(\frac{1}{a}\right)$.

    A 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$, in which case $A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

Chapter 6

Systems of linear equations

6.1 Definitions and theory

Motivating Examples:

1. *Determining the orthogonal complement of a subspace:* Let $V = \mathbb{R}^4$ and let

$$W = Span((2,4,0,-1),(1,-1,2,3)).$$

What is $W^{\perp}$?

2. *Determining whether a given vector lies in the span of some other vectors:* Let $V = \mathbb{R}^3$ and let

$$W = Span((1,-1,2),(2,0,-3)).$$

Is the vector $(-2,-8,17)$ in W?

3. *Word problems:*

a) Dan has some 32-cent stamps, some 29-cent stamps, and some 3-cent stamps. The number of 29-cent stamps is 10 less than the number of 32-cent stamps, while the number of 3-cent stamps is 5 less than the number of 29-cent stamps. The total value of the stamps is $\$9.45$. How many of each stamp does Dan have?

Solution: Let x = the number of 32-cent stamps; y = the number of 29-cent stamps; z = the number of 3-cent stamps. Then:

b) A cashier has 25 coins, all of which are nickels, dimes and quarters. If the total value of her coins is $\$4.90$, how many of each type of coin might she have?

Solution: Let x = the number of nickels; y = the number of dimes; z = the number of quarters. Then:

4. *Fitting data to a model:*

   Find the equation of the plane which best fits the following six points:

$$(1,2,5)$$
$$(1,3,4)$$
$$(2,3,4)$$
$$(1,4,5)$$
$$(2,2,6)$$
$$(3,3,3)$$

We know every plane has the normal equation $ax + by + cz = d$. Assuming $c \neq 0$, we can solve for z to get

$$z = \alpha + \beta x + \gamma y$$

where α, β, γ are constants.

Definition 6.1 *A* **linear equation in** n **variables** *is an equation which can be rewritten as*

$$a_1x_1 + a_2x_2 + ... + a_nx_n = b$$

where $a_1, a_2, ..., a_n, b \in \mathbb{R}$.

Definition 6.2 *A* **linear system of** m **equations in** n **variables** *(a.k.a.* $m \times n$ **linear system***) is a set of* m *linear equations in the same* n *variables:*

$$(*) \begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & & & \ddots & & \vdots & = & \vdots \\ a_{m1}x_m & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{cases}$$

A **solution** *of system* $(*)$ *is a vector* $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = (x_1, ..., x_n) \in \mathbb{R}^n$ *such that all the equations in* $(*)$ *hold.*

A system is called **consistent** *if it has a solution, and is called* **inconstent** *otherwise.*

The **solution set** *of a system is the set of all solutions; this set is a subset of* $\mathbb{R}^n$.

Two $m \times n$ *linear systems are called* **equivalent** *if they have the same solution set.*

Fundamental questions associated to linear systems:

Given an $m \times n$ linear system,

1. Is the system consistent? (In other words, is there a solution?)
2. If the system is consistent, how many solutions does it have?
3. If the system is consistent, what is the solution set?

The key idea to understanding the answers to these questions is to think of a linear system in lots of different ways.

Perspectives from which to view a system of linear equations

1. *As a system of linear equations:* (i.e. as it is written above in the definition)
2. *As a matrix equation:*
3. *As a vector equation:*
4. *As a functional equation involving a linear transformation:*

Example: Consider the system

$$\left\{\begin{array}{ccccl} 3w & -2x & +y & -4z & = 8 \\ w & +x & +7y & & = 2 \\ -w & & +y & -3z & = -3 \end{array}\right.$$

Summarizing some of what we discussed when talking about linear transformations and fundamental subspaces, we have the following theorem, which gives a complete theoretical description of the solution set of a linear system:

Theorem 6.3 (Description of the solution set of a linear system) *Let* $(*)$ *be an* $m \times n$ *system of linear equations whose matrix version is* $A\mathbf{x} = \mathbf{b}$, *and whose linear transformation version is* $T(\mathbf{x}) = \mathbf{b}$ *(where* $T : \mathbb{R}^n \to \mathbb{R}^m$ *is the linear transformation* $T(\mathbf{x}) = A\mathbf{x}$*). Then:*

1. $(*)$ *is consistent* $\iff \mathbf{b} \in C(A) \iff \mathbf{b} \in im(T)$.

   (Thus the system is inconsistent $\iff \mathbf{b} \notin C(A) \iff b \notin im(T)$*.)*

2. *If* $(*)$ *is consistent, then the solution set of the system is the affine subspace* $\mathbf{x}_p + ker(T) = \mathbf{x}_p + N(A)$ *where* $\mathbf{x}_p$ *is any one "particular" solution of the system.*

3. *If* $(*)$ *is consistent, then*

   $$\text{(*) has exactly one solution} \iff ker(T) = \{\mathbf{0}\} \iff T \text{ is injective} \iff r = n.$$

   and

   $$\text{(*) has infinitely many solutions} \iff ker(T) \neq \{\mathbf{0}\} \iff T \text{ is not injective.} \iff r \neq n.$$

The essential content of this theorem is in the following four facts:

1. A system $A\mathbf{x} = \mathbf{b}$ has no solution if and only if $\mathbf{b} \notin C(A)$.

2. A system $A\mathbf{x} = \mathbf{b}$ has exactly one solution if and only if $\mathbf{b} \in C(A)$ and $N(A) = \{\mathbf{0}\}$.

3. A system $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions if and only if $\mathbf{b} \in C(A)$ and $N(A) \neq \{\mathbf{0}\}$.

4. If the system $A\mathbf{x} = \mathbf{b}$ is consistent, then its solution set is the affine subspace $\mathbf{x}_p + N(A)$.

KEY PICTURE

Let's see how this works in a very simple setting. Suppose we have one equation in one variable, i.e. the entire linear system is just

$$Ax = b$$

where A and b are constants.

Compare this with Theorem 1.1 all the way back on page 8.

Now let's try this with two equations in two variables. Suppose we have a system

$$\begin{cases} a_{11}x + a_{12}y = b_1 \\ a_{21}x + a_{22}y = b_2 \end{cases} \quad \text{i.e.} \quad A\mathbf{x} = \mathbf{b} \text{ where } A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}; \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

6.2 Rank and dimension issues

In the last section, we saw the following: if $A \in M_{mn}(\mathbb{R})$ and $T(\mathbf{x}) = A\mathbf{x}$, then...

1. the system $A\mathbf{x} = \mathbf{b}$ has no solution if and only if $\mathbf{b} \notin C(A)$ if and only if $\mathbf{b} \notin im(T)$.
2. the system $A\mathbf{x} = \mathbf{b}$ has exactly one solution if and only if $\mathbf{b} \in C(A)$ and $N(A) = \{\mathbf{0}\}$.

   (This is equivalent to $\mathbf{b} \in im(T)$ and T being injective.)
3. A system $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions if and only if $\mathbf{b} \in C(A)$ and $N(A) \neq \{\mathbf{0}\}$.

   (This is equivalent to $b \in im(T)$ and T being not injective.)
4. If the system $A\mathbf{x} = \mathbf{b}$ is consistent, then its solution set is the affine subspace $\mathbf{x}_p + N(A) = \mathbf{x}_p + ker(T)$.

Adding to this, suppose T is surjective ($r = m$). Then $im(T) = C(A) = \mathbb{R}^m$, so every $\mathbf{b} \in \mathbb{R}^m$ lies in $im(T)$. Then

Now suppose T is injective ($r = n$). Then $ker(T) = N(A) = \{\mathbf{0}\}$, and then

Even further, suppose T is bijective ($r = m = n$). Then T is both surjective and injective, so

We can now state the following theorems involving rank, injectivity and surjectivity:

Theorem 6.4 (Equivalent properties to surjectivity) *Let $A \in M_{mn}(\mathbb{R})$ and suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is defined by $T(\mathbf{x}) = A\mathbf{x}$. Then the following are equivalent (and none of them are possible unless $m \leq n$):*

1. $r(A) = m$;
2. $im(T) = \mathbb{R}^m$;
3. *T is surjective;*
4. $C(A) = \mathbb{R}^m$;
5. $N(A^T) = \{\mathbf{0}\}$;
6. *the rows of A are linearly independent;*
7. *T "maps spanning sets to spanning sets", i.e. given any set $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ which spans $\mathbb{R}^n$, $\{T(\mathbf{v}_1), .., T(\mathbf{v}_k)\}$ spans $\mathbb{R}^m$;*
8. *$A\mathbf{x} = \mathbf{b}$ has at least one solution for every $\mathbf{b} \in \mathbb{R}^m$.*

PROOF $(1 \iff 4)$ $r(A) = dimC(A)$ so since $C(A)$ is a subspace of $\mathbb{R}^m$, it is equal to all of $\mathbb{R}^m$ if and only if its dimension is m.

$(2 \iff 4)$ because $C(A) = im(T)$.

$(2 \iff 3 \iff 8)$ by definition of surjectivity.

$(4 \iff 5)$ by the Fundamental Theorem of Linear Algebra.

$(1 \iff 6)$ by the Rank Theorem.

$(3 \Rightarrow 7)$: Let $\mathbf{w} \in \mathbb{R}^m$. Since T is surjective, $T(\mathbf{v}) = \mathbf{w}$ for some $\mathbf{v} \in \mathbb{R}^n$. Since the $\mathbf{v}_j$ span, we can write $\mathbf{v} = c_1\mathbf{v}_1 + ... + c_k\mathbf{v}_k$. Hence $\mathbf{w} = T(\mathbf{v}) = c_1T(\mathbf{v}_1) + ... + c_kT(\mathbf{v}_k)$ so $\mathbf{w} \in Span(T(\mathbf{v}_1), ..., T(\mathbf{v}_k))$ so $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_k)\}$ span $\mathbb{R}^m$.

$(7 \Rightarrow 3)$: If $T(\mathbf{v}_1), ..., T(\mathbf{v}_k)$ do not span $\mathbb{R}^m$, take $\mathbf{w}$ which does not belong to this span. Suppose $\mathbf{w} = T(\mathbf{v})$. Since $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ span $\mathbb{R}^n$, $\mathbf{v} = c_1\mathbf{v}_1 + ... + c_k\mathbf{v}_k$ so $\mathbf{w} = T(\mathbf{v}) = c_1T(\mathbf{v}_1) + ... + c_kT(\mathbf{v}_k)$, contradicting the hypothesis that $\mathbf{w}$ does not belong to the span. Therefore $\mathbf{w} \notin im(T)$ so T is not surjective. □

Note: If $m > n$, then the situation described in this theorem is impossible, because there is no surjective linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ in this case. Therefore:

Corollary 6.5 *Let $A \in M_{mn}(\mathbb{R})$ where $m > n$ (i.e. A has more rows than columns, i.e. $r \neq m$). Then there must be some $\mathbf{b} \in \mathbb{R}^m$ such that the equation $A\mathbf{x} = \mathbf{b}$ has no solution.*

Theorem 6.6 (Equivalent properties to injectivity) *Let $A \in M_{mn}(F)$ and suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is defined by $T(\mathbf{x}) = A\mathbf{x}$. Then the following are equivalent (and none of them are possible unless $m \geq n$):*

1. $r(A) = n$;
2. $R(A) = \mathbb{R}^n$;
3. $N(A) = \{\mathbf{0}\}$;
4. *T is injective;*
5. $ker(T) = \{\mathbf{0}\}$;
6. *the columns of T are linearly independent;*
7. *T "maps linearly independent sets to linearly independent sets", i.e. given any lin. ind. set $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ in $\mathbb{R}^n$, $\{T(\mathbf{v}_1), .., T(\mathbf{v}_k)\}$ is a lin. ind. set in $\mathbb{R}^m$;*
8. *$A\mathbf{x} = \mathbf{b}$ has at most one solution for every $\mathbf{b} \in \mathbb{R}^m$.*

PROOF $(1 \iff 2)$ $r(A) = dim(R(A))$ so since $R(A)$ is a subspace of $\mathbb{R}^n$, it is equal to all of $\mathbb{R}^n$ if and only if its dimension is n.

$(2 \iff 3)$ by the Fundamental Theorem of Linear Algebra.

$(3 \iff 5)$ because $N(A) = ker(T)$.

$(4 \iff 5 \iff 8)$ by definition of injectivity.

$(1 \iff 6)$ by the definition of rank.

$(4 \Rightarrow 7)$: Assume $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ are linearly independent. Suppose $c_1T(\mathbf{v}_1) + ... + c_kT(\mathbf{v}_k) = \mathbf{0}$. Then $T(c_1\mathbf{v}_1 + ... + c_k\mathbf{v}_k) = \mathbf{0}$ and This means $c_1\mathbf{v}_1 + ... + c_k\mathbf{v}_k \in ker(T)$ and since T is injective, $c_1\mathbf{v}_1 + ... + c_k\mathbf{v}_k = \mathbf{0}$. Since the $\mathbf{v}_j$ are lin. indep., all the c_j must be zero so $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_k)\}$ are linearly independent as well.

$(7 \Rightarrow 4)$: Suppose $T(\mathbf{x}) = \mathbf{0}$. If $\mathbf{x} \neq \mathbf{0}$, then $\{\mathbf{x}\}$ is a linearly independent set of one vector in $\mathbb{R}^n$. By hypothesis, $\{T(\mathbf{x})\}$ is linearly independent in $\mathbb{R}^m$ so $T(\mathbf{x}) \neq 0$, a contradiction. Therefore $\mathbf{x} = \mathbf{0}$, so $ker(T) = \{\mathbf{0}\}$ so T is injective. □

Note: If $m < n$, then the situation described in this theorem is impossible, so:

Corollary 6.7

1. *Let $A \in M_{mn}(\mathbb{R})$ where $m < n$ (i.e. A has more columns than rows.) Then the equation $A\mathbf{x} = \mathbf{b}$ never has exactly one solution.*
2. *A linear system with more variables than equations never has exactly one solution.*

Theorem 6.8 (Equivalent properties to bijectivity) *Let $A \in M_n(\mathbb{R})$ be an $n \times n$ **square** matrix and suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is defined by $T(\mathbf{x}) = A\mathbf{x}$. Then the following are equivalent:*

1. $rank(A) = n$;
2. $im(T) = C(A) = \mathbb{R}^n$;
3. $R(A) = \mathbb{R}^n$;
4. $ker(T) = N(A) = \{\mathbf{0}\}$;
5. *the rows of A form a basis of $\mathbb{R}^n$;*
6. *the rows of A are linearly independent;*
7. *the rows of A span $\mathbb{R}^n$;*
8. *the columns of A form a basis of $\mathbb{R}^n$;*
9. *the columns of A are linearly independent;*
10. *the columns of A span $\mathbb{R}^n$;*
11. *A is invertible;*
12. *There is a matrix $B \in M_n(\mathbb{R})$ such that $AB = I$;*
13. *A is row equivalent to the identity matrix I;*
14. *A has n pivots;*
15. $rref(A) = I$;
16. *$A\mathbf{x} = \mathbf{0}$ has only the solution $\mathbf{x} = \mathbf{0}$;*
17. *there is a single $\mathbf{b} \in \mathbb{R}^n$ such that the equation $A\mathbf{x} = \mathbf{b}$ has exactly one solution;*
18. *the equation $A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $\mathbf{b} \in \mathbb{R}^n$ (namely $\mathbf{x} = A^{-1}\mathbf{b}$);*
19. *T is bijective;*
20. *T is invertible (i.e. a function $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ exists);*
21. *T is injective;*
22. *T is surjective;*
23. *T maps linearly independent sets to linearly independent sets;*
24. *T maps spanning sets to spanning sets;*
25. *T maps bases to bases, i.e. for all bases $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ of $\mathbb{R}^n$, $\{T(\mathbf{v}_1), .., T(\mathbf{v}_n)\}$ is also a basis of $\mathbb{R}^n$;*
26. *given any one single basis $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ of $\mathbb{R}^n$,$\{T(\mathbf{v}_1), .., T(\mathbf{v}_n)\}$ is also a basis of $\mathbb{R}^n$;*
27. *any of properties 1-18 above hold for A^T.*

6.3 Gaussian elimination

Goal: We know what the solution to a system of linear equations looks like theoretically. What we want now is to develop a practical method that will find the solution of a system of linear equations. Ideally, this method will

1.

2.

3.

Example: Suppose $N(A) = Span((2,-1,3),(1,0,5))$ and that $A\begin{pmatrix}1\\1\\4\end{pmatrix} = \begin{pmatrix}2\\0\\3\end{pmatrix}$.

What are all the solutions $\mathbf{x}$ of the system $A\mathbf{x} = \begin{pmatrix}2\\0\\3\end{pmatrix}$?

Basic strategy: Start with some system and replace it with an equivalent system (i.e. one that has the same solution set as the original system) that is easier to solve.

Philosophy: We can

1.

2.

3.

without changing the solution set of the system.

More philosophy: the "names" of the variables don't matter, i.e.

Definition 6.9 *Given a $m \times n$ system of linear equations, the* **augmented matrix** *of the system is the $m \times (n+1)$ matrix*

$$(A \,|\, \mathbf{b}) = \left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right).$$

Example: Write the augmented matrix of the following system:

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

Example: Write the corresponding system of equations for each augmented matrix. Solve the system if it is "easy":

$$1.(A|\mathbf{b}) = \left(\begin{array}{cccc|c} 4 & -1 & 2 & 6 & 0 \\ 1 & 0 & 2 & -1 & 5 \end{array}\right)$$

$$2.(A|\mathbf{b}) = \left(\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 2 & 0 & -1 \\ 0 & 0 & 0 & 1 & 2 \end{array}\right)$$

$$3.(A|\mathbf{b}) = \left(\begin{array}{ccc|c} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 1 & -3 \end{array}\right)$$

$$4.(A|\mathbf{b}) = \left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

How to implement the basic strategy: Given a system of linear equations,

1. Write the augmented matrix $(A|\mathbf{b})$ of the system.
2. Do some "operations" on $(A|\mathbf{b})$ to produce an augmented matrix of an easy system, equivalent to the original one.
3. Solve the easy system.

What the "operations" that are allowed in Step 2? The allowable operations are

i.

ii.

iii.

Example: In each example, you are given a matrix and an indicated row operation (or operations). Perform the indicated operations:

1. Swap the first and third row, then multiply the third row by 3.

$$\begin{pmatrix} 2 & 1 \\ 0 & 4 \\ 3 & -2 \end{pmatrix}$$

2. Multiply the third row by -2, then add 4 times the first row to the third row:

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

Definition 6.10 *Two matrices are called* **row equivalent** *if one can be produced from the other by a sequence of operations of types (i), (ii) and/or (iii) above.*

Theorem 6.11 *If two systems of equations have row equivalent augmented matrices, then the systems are equivalent (they have the same solution set).*

Note: The allowable row operations are all reversible:

- To "undo" the swapping of two rows, swap the rows again.
- To "undo" the multiplying of a row by nonzero constant c, multiply the same row by $\frac{1}{c}$.
- To "undo" the adding of c times one row to another, add $-c$ times the same row to the same row.

Thus the definition of "row equivalent" is symmetric: if A is row equivalent to B, then B is row equivalent to A.

Now thinking back to the examples on pages 151 and 152, why were some of these easier to study than others?

Definition 6.12 *A matrix is in* **(row-)echelon form (ref)** *if*

1. *All rows of zeros are at the bottom of the matrix.*
2. *Defining the* **pivot** *(a.k.a.* **leading entry***) of each row to be the left-most nonzero entry in that row, the pivot of every row is to the right of the pivot of any above row. (This implies, among other things, that all entries directly below any pivot are zero.)*

Definition 6.13 *A matrix is in* **reduced (row-)echelon form (rref)** *if*

1. *It is in row echelon form.*
2. *All pivots are equal to 1.*
3. *Each pivot is the only nonzero entry in its column.*

FACT: A matrix may have several row echelon forms, but it has one and only one rref. So we can talk about **the** rref of A or $(A|\mathbf{b})$, rather than **an** rref of A or $(A|\mathbf{b})$.

Definition 6.14 *Given a matrix A, the* **pivot columns** *of A are the columns which have a pivot in the rref of A. The* **free columns** *of A are the columns which are not pivot columns.*

Note: If the matrix is augmented $(A|\mathbf{b})$, then the last column (the $\mathbf{b}$) is not a pivot nor a free column. It "doesn't count".

Now for the key idea which ties the theory to the practice:

Theorem 6.15 *If A and B are row equivalent matrices, then $R(A) = R(B)$.*

PROOF If A and B are row equivalent, then you can convert one to the other by a sequence of row operations. But none of the three types of elementary row operations changes the row space of a matrix. □

Consequence:

$$\begin{aligned}\dim(R(A)) &= \dim(R(ref(A)) \\ &= \#\text{ of nonzero rows in } ref(A) \\ &= \#\text{ of pivots of } A \\ &= \text{rank of } A \\ &= r.\end{aligned}$$

We have proven the following important theorem:

Theorem 6.16 (Rank Theorem II) *Let $A \in M_{mn}(\mathbb{R})$. Then*

$$r(A) = \#\textit{ of pivots of } A = \#\textit{ of nonzero rows in } ref(A).$$

Example: For each matrix, determine whether or not it is in row echelon form. Determine whether or not it is in reduced row echelon form. Write down all the pivots, the pivot columns and the free columns. Find the rank of A, and the dimension of the four fundamental subspaces associated to A.

1. $\left(\begin{array}{ccccc|c} 1 & 5 & -1 & 6 & 0 & 1 \\ 0 & 1 & 4 & 2 & 1 & -3 \\ 0 & 0 & 0 & 5 & 4 & 5 \\ 0 & 0 & 0 & 0 & -1 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$

2. $\begin{pmatrix} 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$

Algorithm for putting a matrix in ref or rref form by hand

Suppose you are given a matrix (it can be an augmented matrix $(A \mid \mathbf{b})$, or just a coefficient matrix A, or any other matrix) and you want to put it into echelon form. Here's a formulaic way called *Gaussian elimination* or *row reduction* which always works (this is not the only method, however).

Note: No one does this by hand if they have a computer or graphics calculator. See Section 6.7.

Part 1: Transforming a matrix into row-echelon form: "Downward reduction"

1. Place an imaginary cursor at the upper-left entry of the matrix.
2. If the cursor entry and all entries below the cursor entry are zero, move the cursor one column to the right. Repeat this step if necessary.
3. The cursor should now be in a position such that there is a nonzero entry either in the cursor position or directly below the cursor. If necessary, switch the cursor row with a row *beneath* the cursor row to put a nonzero entry in the cursor position.
4. Add multiples of the cursor row to each of the rows beneath the cursor row to create zeros in all entries beneath the cursor.
5. Move the cursor one row down and one column to the right. If the cursor is not on the bottom row, return to step 2. If you reach the last nonzero row, you are done and the matrix is in row-echelon form.

Note: These steps are called *downward reduction* because the cursor is always moving down (from the upper-left corner downward and to the right).

Part 2: Transforming a row-echelon form into a reduced row-echelon form: "Upward reduction"

1. If necessary, carry out downward reduction as in Part 1 to place the matrix in row-echelon form.
2. Multiply each pivot row by the reciprocal of its pivot, so that all the pivots become 1.
3. Place the cursor on the right-most pivot position.
4. If necessary, add multiples of the cursor row to every row above the cursor row so that all entries above the cursor become zero.
5. Once all entries above the cursor are zero, move the cursor up one row and to the left until you reach a pivot position.
6. If the cursor is on the first row, you are done. The matrix is in rref form. Otherwise, return to step 4.

Note: These steps are called *upward reduction* because the cursor is always moving up (from the bottom pivot row up and to the left).

Example: Let $A = \begin{pmatrix} 0 & 6 & 4 & -12 \\ 3 & 3 & 0 & 9 \\ 2 & 0 & -3 & 10 \end{pmatrix}$. Find $rref(A)$ (throughout these operations, the cursor position is underlined).

Part 1: Downward reduction

Matrix	Current Step	Thought Process	Subsequent Row Operation
$\begin{pmatrix} \underline{0} & 6 & 4 & -12 \\ 3 & 3 & 0 & 9 \\ 2 & 0 & -3 & 10 \end{pmatrix}$	1,2	Cursor column contains nonzero entries below cursor. Proceed to step 3.	
$\begin{pmatrix} \underline{0} & 6 & 4 & -12 \\ 3 & 3 & 0 & 9 \\ 2 & 0 & -3 & 10 \end{pmatrix}$	3	I need a nonzero entry at the cursor position.	Swap rows 1 and 2
$\begin{pmatrix} \underline{3} & 3 & 0 & 9 \\ 0 & 6 & 4 & -12 \\ 2 & 0 & -3 & 10 \end{pmatrix}$	4	I need to make all entries at below the cursor zero.	Add $-\left(\frac{2}{3}\right)$ times row 1 to row 2
$\begin{pmatrix} \underline{3} & 3 & 0 & 9 \\ 0 & 6 & 4 & -12 \\ 0 & -2 & -3 & 4 \end{pmatrix}$	5	Move the cursor down and right. Return to step 2.	
$\begin{pmatrix} 3 & 3 & 0 & 9 \\ 0 & \underline{6} & 4 & -12 \\ 0 & -2 & -3 & 4 \end{pmatrix}$	2	Cursor column contains nonzero entries below cursor; cursor entry is nonzero. Proceed to step 4.	
$\begin{pmatrix} 3 & 3 & 0 & 9 \\ 0 & \underline{6} & 4 & -12 \\ 0 & -2 & -3 & 4 \end{pmatrix}$	4	I need to make all entries at below the cursor zero.	Add $\frac{1}{3}$ times row 2 to row 3
$\begin{pmatrix} 3 & 3 & 0 & 9 \\ 0 & \underline{6} & 4 & -12 \\ 0 & 0 & -\frac{5}{3} & 0 \end{pmatrix}$	5	Move the cursor down and right. It'll be on the bottom row, so I'm done.	

Notice that this matrix is in row-echelon form. What you would typically write down on your sheet of paper is something like this:

$$\begin{pmatrix} 0 & 6 & 4 & -12 \\ 3 & 3 & 0 & 9 \\ 2 & 0 & -3 & 10 \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{pmatrix} 3 & 3 & 0 & 9 \\ 0 & 6 & 4 & -12 \\ 2 & 0 & -3 & 10 \end{pmatrix}$$
$$\xrightarrow{-\left(\frac{2}{3}\right)R_1 + R_2} \begin{pmatrix} 3 & 3 & 0 & 9 \\ 0 & 6 & 4 & -12 \\ 0 & -2 & -3 & 4 \end{pmatrix}$$
$$\xrightarrow{\frac{1}{3}R_2 + R_3} \begin{pmatrix} 3 & 3 & 0 & 9 \\ 0 & 6 & 4 & -12 \\ 0 & 0 & -\frac{5}{3} & 0 \end{pmatrix}.$$

Part 2: Upward reduction

Matrix	Current Step	Thought Process	Subsequent Row Operation
$\begin{pmatrix} 3 & 3 & 0 & 9 \\ 0 & 6 & 4 & -12 \\ 0 & 0 & -\frac{5}{3} & 0 \end{pmatrix}$	1,2	Matrix is in ref form. Need to make pivots 1.	Multiply row 1 by $\frac{1}{3}$; multiply row 2 by $\frac{1}{6}$; multiply row 3 by $-\left(\frac{3}{5}\right)$.
$\begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & 1 & \frac{2}{3} & -2 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	3	Place cursor on right-most pivot.	
$\begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & 1 & \frac{2}{3} & -2 \\ 0 & 0 & \underline{1} & 0 \end{pmatrix}$	4	I need to make all entries at above the cursor zero.	Add $-\left(\frac{2}{3}\right)$ times row 3 to row 2
$\begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & \underline{1} & 0 \end{pmatrix}$	5	Move the cursor up and left. Return to step 4.	
$\begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & \underline{1} & 0 & -2 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	4	I need to make all entries at above the cursor zero.	Add -1 times row 2 to row 1
$\begin{pmatrix} 1 & 0 & 0 & 5 \\ 0 & \underline{1} & 0 & -2 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	5	Move the cursor up and left. It'll be on the first row, so I'm done.	

Notice that this matrix is in its rref form. What you would typically write down on your sheet of paper for the upward steps is something like this:

$$\begin{pmatrix} 3 & 3 & 0 & 9 \\ 0 & 6 & 4 & -12 \\ 0 & 0 & -\frac{5}{3} & 0 \end{pmatrix} \xrightarrow{\frac{1}{3}\cdot R_1, \frac{1}{6}\cdot R_2, -\frac{3}{5}\cdot R_3} \begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & 1 & \frac{2}{3} & -2 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\xrightarrow{-\left(\frac{2}{3}\right)R_3+R_2} \begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\xrightarrow{-1\cdot R_2+R_1} \begin{pmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Putting all this together

Example:

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

Example:

$$\begin{cases} 3x_1 & +x_2 & -6x_3 & = -10 \\ 2x_1 & +x_2 & -5x_3 & = -8 \end{cases}$$

Example:

$$\begin{cases} x_1 & & +x_3 & +x_4 & = 5 \\ x_1 & +x_2 & +x_3 & +x_4 & = 6 \\ x_1 & -x_2 & +x_3 & +x_4 & = 5 \end{cases}$$

Example: Find $C(A)$ and find a basis for $C(A)$, if

$$A = \begin{pmatrix} 2 & -1 & 1 & -3 \\ 3 & 2 & 5 & -8 \\ 1 & 6 & 7 & -8 \end{pmatrix}.$$

"Old" solution:

New solution: Perform Gaussian elimination on A.

$$A = \begin{pmatrix} 2 & -1 & 1 & -3 \\ 3 & 2 & 5 & -8 \\ 1 & 6 & 7 & -8 \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_3} \begin{pmatrix} 1 & 6 & 7 & -8 \\ 3 & 2 & 5 & -8 \\ 2 & -1 & 1 & -3 \end{pmatrix} \overset{-3\times R_1+R_2}{\xrightarrow{-2\times R_1+R_3}} \begin{pmatrix} 1 & 6 & 7 & -8 \\ 0 & -16 & -16 & 16 \\ 0 & -13 & -13 & 13 \end{pmatrix}$$

$$\xrightarrow{\frac{-13}{16}\times R_2+R_3} \begin{pmatrix} 1 & 6 & 7 & -8 \\ 0 & -16 & -16 & 16 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\xrightarrow{\frac{-1}{16}\times R_2} \begin{pmatrix} 1 & 6 & 7 & -8 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Note:

Therefore in the ref form of a matrix, the linearly independent columns are the pivot columns, and the linearly dependent columns are the free columns.

Note: In general, $C(A) \neq C(ref(A))$. These subspaces only have the same dimension.

What about $R(A)$? From an earlier result, $R(A) = R(ref(A)) = R(rref(A))$ since A, $ref(A)$ and $rref(A)$ are row equivalent.

Theorem 6.17 (Bases for column space and row space) *Let $A \in M_{mn}(\mathbb{R})$. Then:*

1. *A basis for $C(A)$ consists of the pivot columns of A (not $ref(A)$ or $rref(A)$).*
2. *A basis for $R(A)$ consists of the nonzero rows in any echelon form of A.*

Procedure to find basis of null space of a matrix

The last order of business is to determine how to find a basis for the null space of a matrix:

Let $A \in M_{mn}(\mathbb{R})$. $\mathbf{x} = (x_1, ..., x_n) \in N(A) \iff A\mathbf{x} = \mathbf{0}$.

$$(A|\mathbf{0}) \overset{\text{row ops}}{\longrightarrow} rref(A|\mathbf{0}) = \left(\begin{array}{ccc|c} & & & 0 \\ & & & 0 \\ & & & 0 \\ & & & \vdots \\ & & & \vdots \\ & & & \vdots \\ & & & \vdots \\ & & & 0 \end{array}\right)$$

Example: Find a basis for $N(A)$ if $A = \begin{pmatrix} 1 & 1 & -2 & -5 \\ 2 & 0 & 2 & 2 \\ 3 & 0 & 3 & 3 \\ 4 & 0 & 4 & 4 \end{pmatrix}$.

6.4 Summary so far

Let $A \in M_{mn}(\mathbb{R})$ have rank r (r is the number of pivots in any row echelon form of A). Then:

1. The system $A\mathbf{x} = \mathbf{0}$ always has at least one solution (namely $\mathbf{x} = \mathbf{0}$). There are two possible situations, (a) or (b):
   a) $A\mathbf{x} = \mathbf{0}$ has more than one solution. This is equivalent to all of the following:
      - $N(A) \neq \{\mathbf{0}\}$
      - $\dim N(A) \geq 1$
      - $r < n$ (i.e. there is at least one free column) .

      In this case, for any $\mathbf{b} \in \mathbb{R}^n$:
      - $A\mathbf{x} = \mathbf{b}$ has no solution if and only if $\mathbf{b} \notin C(A)$;
      - $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions if and only if $\mathbf{b} \in C(A)$ (in this case the solution set of $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x}_p + N(A)$ where $\mathbf{x}_p$ is any particular solution of the system) (this case is assured if $r = m < n$);
      - $A\mathbf{x} = \mathbf{b}$ never has exactly one solution.

   b) $A\mathbf{x} = \mathbf{0}$ has exactly one solution (only $\mathbf{x} = \mathbf{0}$). This is equivalent to all of the following:
      - $N(A) = \{\mathbf{0}\}$
      - $\dim N(A) = 0$
      - $r = n$ (i.e. all the columns are pivot columns).

      In this case, for any $\mathbf{b} \in \mathbb{R}^n$:
      - $A\mathbf{x} = \mathbf{b}$ has no solution if and only if $\mathbf{b} \notin C(A)$;
      - $A\mathbf{x} = \mathbf{b}$ has exactly one solution if and only if $\mathbf{b} \in C(A)$ (this is assured if $r = m = n$; see below);
      - $A\mathbf{x} = \mathbf{b}$ never has infinitely many solutions.

      Notice that if $m < n$ (that is, there are fewer equations than variables), case (b) above is impossible because $r \leq m$ (so r cannot be equal to n).

2. In the special case where $r = m = n$ (i.e. A is square and has **full rank**), then $C(A) = R(A) = \mathbb{R}^n$ and $N(A) = N(A^T) = \{\mathbf{0}\}$. Furthermore, for every $\mathbf{b} \in \mathbb{R}^n$, the system $A\mathbf{x} = \mathbf{b}$ has exactly one solution.

   This situation is equivalent to many different things, which are described in Theorem 6.8.

3. The system $A\mathbf{x} = \mathbf{b}$ has no solution if and only if $\mathbf{b} \notin C(A)$ if and only if an echelon form of the augmented matrix $(A \,|\, \mathbf{b})$ contains a **false row** of the form $(0\ 0\ \cdots\ 0 \mid z)$ where $z \neq 0$.

6.5 Computing the inverse of a matrix

Question: Suppose $A \in M_n(\mathbb{R})$ is an invertible matrix. How do you compute its inverse?

From the long list of properties equivalent to bijectivity, we know

$$A \text{ is invertible } \iff A \text{ is row equivalent to the identity matrix } I.$$

Suppose you performed a bunch of row operations on A to produce I (i.e. to put A in rref form). If you did the same row operations to I, you will produce A^{-1} (because row operations are equivalent to left-multiplying by matrices... for more details on why this is, come to office hours).

This suggests a method, called the **Gauss-Jordan method**, to compute the inverse of a square matrix A:

1. Write an "augmented matrix" $(A|I)$.
2. Perform Gaussian elimination to put A in rref form.
   - If you get a row of zeros in the A part of the matrix at any point, A is not invertible.
   - Otherwise, you will end up with a matrix of the form $(I\,|A^{-1})$.

(In practice, one uses a computer or calculator to compute the inverse of a matrix. See Section 6.7.)

Example: Find the inverse of

$$A = \begin{pmatrix} 1 & 2 & 1 \\ -1 & 0 & 2 \\ 0 & 1 & 4 \end{pmatrix}.$$

6.6 Least-squares approximations

Motivating example: Find the equation of the plane in $\mathbb{R}^3$ passing through the following points:

$$(1,1,1) \quad (2,2,2) \quad (2,1,1) \quad (1,3,4)$$

First attempt at a solution: We know that every plane in $\mathbb{R}^3$ has normal equation

$$ax + by + cz = d$$

so by solving for z, we see that every plane (with $c \neq 0$) in $\mathbb{R}^3$ has equation

$$z =$$

To determine this equation, plug each of the given points in for x, y and z:

$$\left\{ \begin{array}{l} \\ \\ \\ \end{array} \right.$$

This translates into the matrix equation $A\mathbf{x} = \mathbf{b}$ where $\mathbf{x} = (\alpha, \beta, \gamma)$;

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}; \quad \mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 4 \end{pmatrix}$$

Notice:

$$\begin{aligned} \text{first column of } A &\leftrightarrow \text{ a column of all ones} \\ \text{last column of } A &\leftrightarrow \\ \text{second column of } A &\leftrightarrow \\ \text{the column vector } \mathbf{b} &\leftrightarrow \end{aligned}$$

Based on what we have learned, such a system is solved by performing row reductions:

$$(A \mid \mathbf{b}) = \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 3 & 4 \end{array}\right) \xrightarrow{\text{row ops}} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right) = \text{ref}(A \mid \mathbf{b})$$

Therefore

Motivating example, restated: Find the equation of the plane in $\mathbb{R}^3$ which *best fits* the following points:

$$(1,1,1) \quad (2,2,2) \quad (2,1,1) \quad (1,3,4)$$

Solution: Start by repeating the work on the previous page to obtain the matrix equation $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}; \quad \mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 4 \end{pmatrix}$$

From the last page, we know there is no solution $\mathbf{x}$ to this system, which means $\mathbf{b} \notin C(A)$. This means we can think of a picture like this:

The plane which best fits the data will be given by the $\mathbf{x}$ which makes $A\mathbf{x}$ as close to possible to $\mathbf{b}$. Thus

$A\mathbf{x}$ should be equal to

Question: How do you compute $\widehat{\mathbf{b}}$?

<u>Question</u> (from previous page): How do you compute $\widehat{\mathbf{b}}$?

<u>**Method $\#1$**</u> **(comes from Chapter 4):** $C(A)$ is spanned by the columns of A. So use Gram-Schmidt to convert the columns of A into an orthonormal basis $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$, and then use the projection formula. This gives

$$\widehat{\mathbf{b}} = (\mathbf{b} \cdot \mathbf{x}_1)\mathbf{x}_1 + (\mathbf{b} \cdot \mathbf{x}_2)\mathbf{x}_2 + (\mathbf{b} \cdot \mathbf{x}_3)\mathbf{x}_3.$$

While this method always works, there is a drawback with this method: it is computationally intensive, and doesn't tell you what $\widehat{\mathbf{x}}$ satisfies $A\widehat{\mathbf{x}} = \widehat{\mathbf{b}}$.

<u>**Method $\#2$**</u> **(new):** Since $\widehat{\mathbf{b}} \in C(A)$, the system $A\mathbf{x} = \widehat{\mathbf{b}}$ has at least one solution; call this solution $\widehat{\mathbf{x}}$. Therefore $A\widehat{\mathbf{x}} = \widehat{\mathbf{b}}$.

However, since $\widehat{\mathbf{b}}$ is the projection of $\mathbf{b}$ onto $C(A)$,

$$\mathbf{b} - \widehat{\mathbf{b}} \in [C(A)]^T = N(A^T).$$

Therefore $A^T(\mathbf{b} - \widehat{\mathbf{b}}) = \mathbf{0}$. Substituting in $A\widehat{\mathbf{x}}$ for $\widehat{\mathbf{b}}$, we see

$$\begin{aligned} & A^T(\mathbf{b} - A\widehat{\mathbf{x}}) = \mathbf{0} \\ \Rightarrow \quad & A^T\mathbf{b} - A^TA\widehat{\mathbf{x}} = \mathbf{0} \\ \Rightarrow \quad & A^T\mathbf{b} = A^TA\widehat{\mathbf{x}} \\ \Rightarrow \quad & A^T\mathbf{b} = (A^TA)\widehat{\mathbf{x}} \\ \Rightarrow \quad & (A^TA)^{-1}A^T\mathbf{b} = \widehat{\mathbf{x}} \end{aligned}$$

Therefore

$$\widehat{\mathbf{b}} = A\widehat{\mathbf{x}} = A(A^TA)^{-1}A^T\mathbf{b}.$$

$\widehat{\mathbf{x}}$ is called the **least-squares** approximation to the original equation $A\mathbf{x} = \mathbf{b}$.

Back to the motivating example:

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}; \quad \mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 4 \end{pmatrix}$$

Therefore

$$\begin{aligned}\hat{\mathbf{x}} &= (A^TA)^{-1}A^T\mathbf{b} \\ &= \left[\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 \\ 1 & 2 & 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}\right]^{-1}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 \\ 1 & 2 & 1 & 3 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ 1 \\ 4 \end{pmatrix} \\ &= \begin{pmatrix} 4 & 6 & 7 \\ 6 & 10 & 10 \\ 7 & 10 & 15 \end{pmatrix}^{-1}\begin{pmatrix} 8 \\ 11 \\ 18 \end{pmatrix} \\ &= \begin{pmatrix} 5 & -2 & -1 \\ -2 & \frac{11}{10} & \frac{1}{5} \\ -1 & \frac{1}{5} & \frac{2}{5} \end{pmatrix}\begin{pmatrix} 8 \\ 11 \\ 18 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ \frac{-3}{10} \\ \frac{7}{5} \end{pmatrix}\end{aligned}$$

Therefore the least-squares solution is $\hat{\mathbf{x}} = (\alpha, \beta, \gamma) = (0, \frac{-3}{10}, \frac{7}{5})$, so the plane best fitting the data points

$$(1,1,1) \quad (2,2,2) \quad (2,1,1) \quad (1,3,4)$$

is

Fortunately, if you have access to *Mathematica*, implementing this method is easy: type in the matrices A and $\mathbf{b}$ and then run the command

```
LeastSquares[A,b] //MatrixForm
```

This will compute and display $\hat{\mathbf{x}}$.

Drawback: The only drawback to the least-square method is that it requires the matrix A^TA to be invertible to work. Fortunately, this is assured in most situations because of the following fact:

<u>FACT</u>: Let $A \in M_{mn}(\mathbb{R})$. If the columns of A are linearly independent, then the $n \times n$ matrix A^TA is invertible.

In the context of our example, this is assured if the same x, y coordinates from the original list of points are never repeated.

To summarize:

Theorem 6.18 (Least-squares solutions) *Let $A \in M_{mn}(\mathbb{R})$ and let $\mathbf{b} \in \mathbb{R}^m$. Then, assuming the columns of A are linearly independent,*

1. *The projection of $\mathbf{b}$ onto the column space of A is*

$$\widehat{\mathbf{b}} = \pi_{C(A)}\mathbf{b} = A(A^TA)^{-1}A^T\mathbf{b}.$$

2. *The* **least-squares solution** *of the system $A\mathbf{x} = \mathbf{b}$ is the vector $\widehat{\mathbf{x}} \in \mathbb{R}^n$ given by*

$$\widehat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}.$$

$\widehat{\mathbf{x}}$ solves the equation $A\widehat{\mathbf{x}} = \widehat{\mathbf{b}}$, and is the closest approximation to any "solution" of $A\mathbf{x} = \mathbf{b}$. It isn't generally a solution of $A\mathbf{x} = \mathbf{b}$, however.

Theorem 6.19 *Let $A \in M_n(\mathbb{R})$ be an invertible square matrix and let $\mathbf{b} \in \mathbb{R}^n$. Then the least-squares solution of the system $A\mathbf{x} = \mathbf{b}$ is $\widehat{\mathbf{x}} = A^{-1}\mathbf{b}$ and is the same as the actual solution of the original system.*

PROOF If A is invertible, so is A^T and $(A^TA)^{-1} = A^{-1}(A^T)^{-1}$. Therefore

$$\widehat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} = A^{-1}(A^T)^{-1}A^T\mathbf{b} = A^{-1}I\mathbf{b} = A^{-1}b$$

and $A\widehat{\mathbf{x}} = A(A^{-1}\mathbf{b}) = \mathbf{b}$. □

We wrap this material up by restating the already-proven facts in terms of projections:

Definition 6.20 *Let $A \in M_{mn}(\mathbb{R})$ be a matrix whose columns are linearly independent. Then define the $m \times m$ matrix P to be*

$$P = A(A^TA)^{-1}A^T.$$

P is called the **projection matrix** *for A.*

Theorem 6.21 *Let $A \in M_{mn}(\mathbb{R})$ be a matrix whose columns are linearly independent, and let P be its projection matrix. Then for any $\mathbf{y} \in \mathbb{R}^m$, $P\mathbf{y} = \pi_{C(A)}\mathbf{y}$, the projection of $\mathbf{y}$ onto the column space of A.*

Example: Compute the projection of $(2, -5, 1, 7)$ onto the subspace spanned by $(1, 3, -2, 1)$ and $(4, -1, 1, 3)$.

6.7 Solving systems of equations using technology

Linear systems on *Mathematica*

Mathematica does each of the computations you need to be able to do with one command line:

1. To solve a system of linear equations $A\mathbf{x} = \mathbf{b}$ using *Mathematica*:

   a) Type in the matrix A (using the syntax from the end of Chapter 2).

   b) Type in the vector $\mathbf{b}$ (also using the syntax from the end of Chapter 2).

   c) Execute `LinearSolve[A,b]`. This will produce a particular solution $\mathbf{x}_p$ (if possible) and will tell you if there is no solution.

   d) Execute `NullSpace[A]`. This will spit out a list of vectors which form a basis of $N(A)$.

   e) The solution, as the theory tells us, is $\mathbf{x}_p + N(A)$. Make sure you write it correctly.

   As an example, suppose we were trying to solve

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

   In *Mathematica*, execute the following four commands:

   - `A = {{1,2,-3},{2,1,-5}}`
   - `b = {4,1}`
   - `LinearSolve[A,b]`
     This produces the output $\left\{\{-\frac{2}{3}\}, \{\frac{7}{3}\}, \{0\}\right\}$. This is a particular solution $\mathbf{x}_p$.
   - `NullSpace[A]`
     This produces the output $\{\{7, 1, 3\}\}$, meaning that the single vector $(7, 1, 3)$ is a basis for $N(A)$.

   The solution is therefore

   $$\mathbf{x}_p + N(A) = \left(\frac{-2}{3}, \frac{7}{3}, 0\right) + t(7, 1, 3) = \left(\frac{-2}{3} + 7t, \frac{7}{3} + t, 3t\right).$$

2. Other things *Mathematica* can do:

   a) To find the rref form of a matrix, type in the matrix (saving it as, say A) and then execute `RowReduce[A]`.

b) To find the rank of a matrix, type in the matrix and execute `MatrixRank[A]`.
c) To find a basis for the null space of a matrix, type in the matrix and execute `NullSpace[A]`.
d) To find the inverse of a square matrix, type in the matrix and execute `Inverse[A]`.
e) To find the least-squares solution of a system, type in A and $\mathbf{b}$ and execute `LeastSquares[A,b] //MatrixForm`; the output is $\hat{\mathbf{x}}$.

Linear systems on TI-83/84 type calculators

1. To solve a system of linear equations $A\mathbf{x} = \mathbf{b}$ using such a calculator:

   a) Type in the augmented matrix $(A|\mathbf{b})$ and save it (as a single matrix) using the directions of Chapter 2.
   b) Then hit `[MATRX]`, go to the right to find `MATH`, then go down to `rref(`. Hit `[ENTER]`, then pull up the matrix you have saved by hitting `[MATRX]` and scrollling down to the matrix you want. Then execute this command; your calculator will spit out the reduced row-echelon form of the augmented matrix.
   c) Solve the system of equations corresponding to the rref using the procedures described in Section 6.3.

   As an example, if we were trying to solve

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

   we would first save the matrix

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

   as `[A]`, then ask the calculator to compute `rref([A])`. You will get (after converting the answer to fractions)

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{-5}{2} & \frac{1}{2} \\ 0 & 1 & \frac{-1}{3} & \frac{7}{3} \end{pmatrix}$$

   This translates back to the system

$$\begin{cases} x + \frac{1}{2}y - \frac{5}{2}z = \frac{1}{2} \\ y - \frac{1}{3}z = \frac{7}{3} \end{cases}$$

   which you solve by hand (as in Section 6.3) to get the solution.

2. Other things the graphics calculator will do:

   a) To find the rref form of a matrix A, first type in the matrix and save it using the directions of Chapter 2. Then hit `[MATRX]`, go to the right to find `MATH`, then go down to `rref(`. Hit `[ENTER]`, then pull up the matrix you have saved by hitting `[MATRX]` and scrollling down to the matrix you want. Then execute this command; your calculator will spit out the reduced row-echelon form of A.

   b) To find the inverse of a matrix, first type in the matrix and save it using the directions of Chapter 2. Then hit `[MATRX]` and scroll down to the matrix you want to take the inverse of; hit `[ENTER]`. Then hit [x^{-1}] and execute this command; your calculator will spit out the inverse of A and will give you an error if A is not invertible.

Using an online equation solver

1. To solve a system of linear equations $A\mathbf{x} = \mathbf{b}$ using a free online tool:

   a) Open a web browser to `http://www.quickmath.com`

   b) Click "Solve" under "Equations" on the left-hand side.

   c) Click "Advanced" on the right-hand side, above the yellow box.

   d) Inside the yellow box, type in the linear equations in the left-hand box; type in the variables of the system in the middle box; click "Solve".

   e) The solution will appear below (including cases of no solution or infinitely many solutions).

2. There are probably lots of other online tools which can be used to compute bases of a null space or perform row reductions or compute inverses of matrices; I'll leave it to you to use Google to find them. (I don't suggest using WolframAlpha, because the output you get is a bit clunky.)

Chapter 7

Determinants

7.1 Definition and properties

Goal: Assign, to every square matrix, a real number which tells you something about the matrix.

Motivation: Consider a 2×2 square matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Treat the columns of A as vectors: let $\mathbf{v} = (a, c)$ and $\mathbf{w} = (b, d)$. Use those vectors to make a parallelogram:

What is the area of this parallelogram?

Definition 7.1 *Let* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{R})$. *The* **determinant** *of* A*, denoted* $\det A$*, is the number* $\det A = ad - bc$.

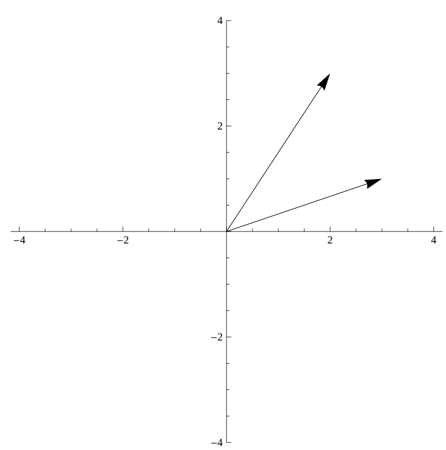
The determinant of A is also denoted $|A|$, but this a poor choice of notation that should be avoided.

Thus, if you have a parallelogram in $\mathbb{R}^2$, and you treat the sides of the parallelogram as vectors, and put those vectors in as columns of a matrix A, then

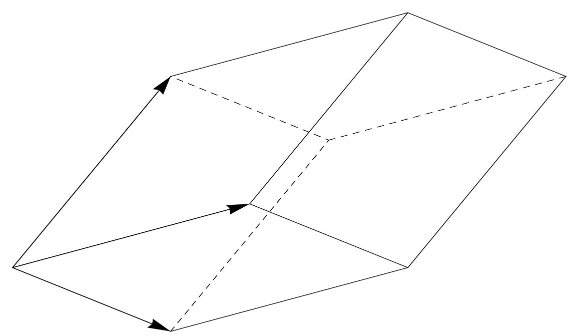
$$\text{area of the parallelogram } = |\det A|.$$

Example: Find the determinant of $\begin{pmatrix} 2 & 3 \\ 3 & 1 \end{pmatrix}$.

What, is the signficance of the fact that this determinant is negative?



In general, the columns of an $n \times n$ matrix can be thought of as n vectors in $\mathbb{R}^n$. Those n vectors can be used to make a $n-$dimensional box called a **parallelepiped**.



If you take the same vectors and put them in as columns of a matrix, this makes an $n \times n$ matrix A. The **determinant** of this matrix is either $+1$ or -1 times the volume of this parallelepiped. Whether or not it is positive or negative depends on the "orientation" of the column vectors, i.e. which order they are written in.

Remark: There is a more rigorous definition of the determinant of a matrix, but that requires mathematics beyond the scope of this course.

Theorem 7.2 (Properties of determinants) *Let $A, B \in M_n(\mathbb{R})$. Then:*

1. $\det A = \det A^T$.
2. *If B is obtained from A by swapping two columns of A, then* $\det B = -\det A$.
3. *If B is obtained from A by swapping two rows of A, then* $\det B = -\det A$.
4. *If two rows or columns of A coincide, then* $\det A = 0$.
5. $\det I = 1$, *where I is the identity matrix.*
6. *If B is obtained from A by multiplying one row (or column) of A by a constant r, then* $\det B = r \det A$.
7. $\det(rA) = r^n \det A$.
8. $\det(AB) = \det A \cdot \det B$.
9. *A is invertible if and only if* $\det A \neq 0$, *in which case* $\det(A^{-1}) = \frac{1}{\det A}$.

Fact: The only function $M_n(\mathbb{R}) \to \mathbb{R}$ which satisfies (1), (2), (5) and (6) above is the determinant.

JUSTIFICATION OF SOME OF THESE:

Note: Many of these arguments are only sketches of ideas; to formally justify many of them one needs to define determinant more rigorously.

(2) Swapping two columns of A does not change the parallelepiped whose volume is being computed, but changes the order in which the columns are written, causing the determinant to be multiplied by -1.

(4) If two columns of A coincide, then the parallelepiped is "degenerate", i.e. has no n-dimensional volume. Since the absolute value of $\det A$ is this volume, we have $|\det A| = 0$, i.e. $\det A = 0$. If two rows of A coincide, then two columns of A^T coincide so by the above, $\det A^T = 0$ so by (1) $\det A = 0$.

(5) Treating the columns of I as the sides of a parallelepiped, we get a $1 \times 1 \times 1 \cdots 1$ box with the sides orthogonal. Such a box clearly has volume 1, so $|\det I| = 1$. In fact, $\det I = 1$ because the columns of I are written in "the right order" (this is one part of the more rigorous definition of determinant).

(6) Multiplying one column of a matrix by r makes the parallelepiped "r times as big" since one side of the parallelepiped is multiplied by r. Thus the determinant is also multiplied by r, since the determinant of a matrix measures (up to the sign) this volume.

(7) This follows immediately from (6).

(9) ($\Rightarrow$) If A is invertible, then there exists matrix A^{-1} such that $AA^{-1} = I$. Then by taking determinants of both sides and applying (5) and (8), we see $\det A \cdot (\det A^{-1}) = 1$, so $\det A \neq 0$ (and $\det A^{-1} = \frac{1}{\det A}$). $\square$

7.2 Computing determinants

In this section we describe how to compute the determinant of a matrix.

1×1 matrices

The determinant of a 1×1 matrix is just the number in the matrix: if $A = (a)$, then $\det A = a$.

2×2 matrices

As we have seen, the determinant of a 2×2 matrix is given by the following formula:

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Example: $A = \begin{pmatrix} 7 & 12 \\ 3 & 5 \end{pmatrix}$

3×3 matrices

Theorem 7.3 *Let* $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \in M_3(\mathbb{R})$. *Then*

$$\det A = aei + cdh + bfg - bdi - ceg - afh.$$

The proof of this theorem is beyond the scope of this course.

The formula in Theorem 7.3 is impossible to remember, but there is a trick. Given a 3×3 matrix, copy the first two columns to the right of the matrix:

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{matrix} a & b \\ d & e \\ g & h \end{matrix}$$

Then multiply along the diagonals, and add the "upper" and "lower" products. The determinant is the bottom sum minus the top sum.

WARNING: This technique does not work for 4×4 and larger matrices.

Example: $A = \begin{pmatrix} 1 & -2 & 3 \\ 2 & 4 & -1 \\ 3 & -1 & 1 \end{pmatrix}$.

4×4 and larger matrices

The easiest and most efficient way to compute a determinant of a large matrix is to use technology (see the appendix of this packet). However, there is a way to compute determinants by hand called **evaluation by minors**.

Definition 7.4 *Let $A \in M_n(\mathbb{R})$ where $n \geq 2$ and let $i, j \in \{1, ..., n\}$. The $(i,j)-$***minor** *of A is the $(n-1) \times (n-1)$ matrix obtained by deleting the i^{th} row and j^{th} column from A; this matrix is denoted $A_{i,j}$. (Recall that $a_{i,j}$ is the $(i,j)-$entry of A.)*

Example: Let $A = \begin{pmatrix} 1 & 3 & 5 & 7 \\ 2 & -1 & 1 & 0 \\ 0 & 3 & 1 & 2 \\ 2 & 4 & -2 & -4 \end{pmatrix}$.

Here are the formulas that show you how to compute determinants of large matrices, in terms of determinants of smaller matrices (the proofs of these theorems are beyond the scope of this class):

Theorem 7.5 (Evaluation of determinants via minors along a row) *Let A be an $n \times n$ matrix. Then for any $i \in \{1, 2, ..., n\}$,*

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{i,j} \det(A_{i,j}).$$

Theorem 7.6 (Evaluation of determinants via minors along a column) *Let $A \in M_n(\mathbb{R})$. Then for any $j \in \{1, 2, ..., n\}$,*

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{i,j} \det(A_{i,j}).$$

Note: The values of $(-1)^{i+j}$ should be thought of as the following:

$$\begin{pmatrix} + & - & + & \cdots \\ - & + & - & \ddots \\ + & - & \ddots & \\ - & \ddots & & \\ \vdots & & & \end{pmatrix}$$

Example: Find the determinant of each of the following matrices:

Ex. 1: $A = \begin{pmatrix} 1 & -2 & 3 \\ 2 & 4 & -1 \\ 3 & -1 & 1 \end{pmatrix}$.

Ex. 2: $A = \begin{pmatrix} 1 & 1 & -2 & 3 \\ 2 & 0 & 4 & -1 \\ 3 & 0 & -1 & 1 \\ 3 & 2 & 0 & 1 \end{pmatrix}$.

Ex. 3: $A = \begin{pmatrix} 0 & 0 & -1 & 0 & 3 \\ 1 & 0 & 2 & 4 & 1 \\ 0 & 1 & -3 & 0 & 2 \\ 1 & 1 & 2 & 0 & -2 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}$.

Ex. 4: $A = \begin{pmatrix} 3 & 1 & 4 & -2 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & -2 & 5 \\ 0 & 0 & 0 & 10 \end{pmatrix}$.

This example generalizes:

Theorem 7.7 *If A is a triangular matrix, then $\det A$ is the product of its diagonal entries.*

7.3 Computing determinants using technology

Determinants on *Mathematica*

To find the determinant of a matrix using *Mathematica*, type in the matrix (saving it as, say A) and then execute `Det[A]`.

Determinants on TI-83/84 type calculators

To find the determinant of a matrix using a calculator, first type in the matrix and save it, following the directions of Chapter 2.

Then hit `[MATRX]`, go to the right to find `MATH`, then go down to `det(`. Hit `[ENTER]`, then pull up the matrix you have saved by hitting `[MATRX]` and scrollling down to the matrix you want. Then execute this command; your calculator will spit out the determinant of A.

7.4 Using determinants to compute cross products

Recall that if $\mathbf{v}$ and $\mathbf{w}$ are two vectors in $\mathbb{R}^3$, a vector $\mathbf{n}$ orthogonal to both $\mathbf{v}$ and $\mathbf{w}$ is given by the cross product

$$\mathbf{n} = \mathbf{v} \times \mathbf{w}.$$

We learned a formula for cross products that is hard to memorize. Here is an easier method to remember:

> **Theorem 7.8** *Let* $\mathbf{v} = (v_1, v_2, v_3)$ *and* $\mathbf{w} = (w_1, w_2, w_3)$ *be two vectors in* $\mathbb{R}^3$. *Then*
>
> $$\mathbf{v} \times \mathbf{w} = \det \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix}.$$

PROOF Evaluating the determinant by minors across the top row, we see

$$\begin{aligned} \det \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix} &= \mathbf{e}_1(v_2w_3 - w_2v_3 - \mathbf{e}_2(v_1w_3 - w_1v_3) + \mathbf{e}_3(v_1w_2 - w_1v_2) \\ &= (v_2w_3 - w_2v_3, 0, 0) - (0, v_1w_3 - w_1v_3, 0) + (0, 0, v_1w_2 - w_1v_2) \\ &= (v_2w_3 - w_2v_3, w_1v_3 - v_1w_3, v_1w_2 - w_1v_2) \end{aligned}$$

which is the same formula we already defined as $\mathbf{v} \times \mathbf{w}$. □

Warning: Strictly speaking, the expression $\begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix}$ is gibberish, because it is a matrix with some entries that are vectors and some entries that are numbers. But if you use the determinant formula on this gibberish, you get the correct cross product, so it's "okay" to use.

Example: Compute $(2, 3, -1) \times (1, 4, -3)$.

Example: Find the normal equation of the plane containing the three points $(1, 3, -2)$, $(0, 2, 1)$ and $(5, -1, 3)$.

Chapter 8

Eigentheory

8.1 Eigenvalues and eigenvectors

Problem: Given a square matrix $A \in M_n(\mathbb{R})$, compute A^r for some large whole number r.

(We'll discuss a reason why one would want to do this later, and there are many other applications of matrix powers in other areas of pure and applied mathematics.)

Example: Suppose $A = \begin{pmatrix} 2 & -3 \\ 4 & 1 \end{pmatrix}$. Then

$$A^2 = AA = \begin{pmatrix} -8 & -9 \\ 12 & -11 \end{pmatrix} \quad A^3 = A^2A = \begin{pmatrix} -52 & 15 \\ -20 & -47 \end{pmatrix} \quad A^4 = \begin{pmatrix} -44 & 171 \\ -228 & 13 \end{pmatrix} \quad \text{etc.}$$

Note that the entries of these matrices don't seem to have anything to do with A. And if you were asked to compute A^{73}, you'd be screwed.

Example: Suppose $\Lambda = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$. ($\Lambda$ is the Greek capital letter lambda.)

Then

$$\Lambda^2 =$$

$$\Lambda^3 = \Lambda^2\Lambda =$$

$$\Lambda^r =$$

This example shows that it is easy to compute the powers of a diagonal matrix. More precisely:

Theorem 8.1 *Suppose* $\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}$ *is a diagonal* $n \times n$ *matrix. Then for any* $r \in \{0, 1, 2, ...\}$, $\Lambda^r = \begin{pmatrix} \lambda_1^r & 0 & \cdots & 0 \\ 0 & \lambda_2^r & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n^r \end{pmatrix}$.

Now, suppose that given a square matrix $A \in M_n(\mathbb{R})$, you can (somehow) find a diagonal matrix $\Lambda \in M_n(\mathbb{R})$ and an invertible matrix $S \in M_n(\mathbb{R})$ such that

$$A = S\Lambda S^{-1}.$$

Then, we can compute powers of A as follows:

$$A^r = (S\Lambda S^{-1})^r =$$

This idea motivates the following definition and theorem:

Definition 8.2 *A square* $n \times n$ *matrix* A *is called* **diagonalizable** *(a.k.a.* similar to a diagonal matrix*) if there is an invertible matrix* $S \in M_n(\mathbb{R})$ *and a diagonal matrix* $\Lambda \in M_n(\mathbb{R})$ *such that* $A = S\Lambda S^{-1}$.

Writing A as $S\Lambda S^{-1}$ is called **diagonalizing** A.

From above, we have

Theorem 8.3 *If* $A \in M_n(\mathbb{R})$ *is diagonalizable, then for any* $r \in \{0, 1, 2, ...\}$, $A^r = S\Lambda^r S^{-1}$.

Therefore, to compute the powers of a diagonalizable matrix A, we could to find the S and the Λ and apply the formula in the preceding theorem.

Question: Given $A \in M_n(\mathbb{R})$, is A diagonalizable? If so, what are S and Λ?

To answer this question theoretically, we first assume A is diagonalizable and diagonalize A:

$$A = S\Lambda S^{-1}$$

Multiply both sides of this equation by S on the right to obtain

$$S\Lambda = AS. \tag{8.1}$$

Write the columns of S as $\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_n$ so that

$$S = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{pmatrix}.$$

Note that $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ must be linearly independent since S is invertible (so none of the $\mathbf{x}_j$ are the zero vector).

The left-hand side of equation (8.1) is

$$\begin{aligned} S\Lambda &= \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 s_{11} & \lambda_2 s_{12} & \cdots & \lambda_n s_{1n} \\ \lambda_1 s_{21} & \lambda_2 s_{22} & \cdots & \lambda_n s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 s_{n1} & \lambda_2 s_{n2} & \cdots & \lambda_n s_{nn} \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 \mathbf{x}_1 & \lambda_2 \mathbf{x}_2 & \cdots & \lambda_n \mathbf{x}_n \end{pmatrix}. \end{aligned}$$

The right-hand side of equation (8.1) is

$$\begin{aligned} AS &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{pmatrix} \\ &= \begin{pmatrix} a_{11}s_{11} + a_{12}s_{21} + ... + a_{1n}s_{n1} & \cdots & \cdots & a_{11}s_{1n} + a_{12}s_{2n} ... + a_{1n}s_{nn} \\ a_{21}s_{11} + a_{22}s_{21} + ... + a_{2n}s_{n1} & & & a_{21}s_{1n} + a_{22}s_{2n} + ... + a_{2n}s_{nn} \\ \vdots & & & \vdots \\ a_{n1}s_{11} + a_{n2}s_{21} + ... + a_{nn}s_{n1} & \cdots & \cdots & a_{n1}s_{1n} + a_{n2}s_{2n} + ... + a_{nn}s_{nn} \end{pmatrix} \\ &= \begin{pmatrix} A\mathbf{x}_1 & A\mathbf{x}_2 & \cdots A\mathbf{x}_n \end{pmatrix}. \end{aligned}$$

Equating the columns of the left- and right-hand sides of equation (1), we see that

$$\lambda_j \mathbf{x}_j = A\mathbf{x}_j \tag{8.2}$$

for all j.

Equation (8.2) allows us to say quite a bit:

Definition 8.4 *Let $A \in M_n(\mathbb{R})$ be a square matrix. A scalar λ is called an* **eigenvalue** *of A if there is a nonzero vector $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \lambda\mathbf{x}$. In this setting, $\mathbf{x}$ is called an* **eigenvector** *corresponding to λ. The set of eigenvalues of A is called the* **spectrum** *of A.*

The work on the previous page essentially proves the following theorem:

Theorem 8.5 *Let $A \in M_n(\mathbb{R})$ be a square matrix. A is diagonalizable if and only if it has n linearly independent eigenvectors, in which case $A = S\Lambda S^{-1}$ where Λ is a diagonal matrix whose entries are the eigenvalues of A, and S is a matrix whose columns are the corresponding eigenvectors (written in the same order as the eigenvalues are written in Λ).*

Theorem 8.6 *Let $A \in M_n(\mathbb{R})$ be a square matrix. If A has n different eigenvalues, then A is diagonalizable.*

PROOF Let $\lambda_1, ..., \lambda_n$ be the eigenvalues of A and let $\mathbf{x}_1, ..., \mathbf{x}_n$ be corresponding eigenvectors. Then for all $j \in \{1, ..., n\}$ we have

$$A\mathbf{x}_j = \lambda_j \mathbf{x}_j.$$

We claim $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ are linearly independent. To show this, suppose that they are linearly dependent and let $\mathbf{x}_k$ be the **first** dependent vector. That means $\{\mathbf{x}_1, ..., \mathbf{x}_{k-1}\}$ are linearly independent and there are scalars $c_1, ..., c_{k-1}$ such that

$$\mathbf{x}_k = c_1\mathbf{x}_1 + ... + c_{k-1}\mathbf{x}_{k-1}. \tag{8.3}$$

Now multiply both sides by A to get

$$\begin{aligned}
A\mathbf{x}_k &= A(c_1\mathbf{x}_1 + ... + c_{k-1}\mathbf{x}_{k-1}) \\
A\mathbf{x}_k &= c_1 A\mathbf{x}_1 + ... + c_{k-1}A\mathbf{x}_{k-1} \\
\lambda_k\mathbf{x}_k &= c_1\lambda_1\mathbf{x}_1 + ... + c_{k-1}\lambda_{k-1}\mathbf{x}_{k-1} \\
\lambda_k(c_1\mathbf{x}_1 + ... + c_{k-1}\mathbf{x}_{k-1}) &= c_1\lambda_1\mathbf{x}_1 + ... + c_{k-1}\lambda_{k-1}\mathbf{x}_{k-1} \\
c_1\lambda_k\mathbf{x}_1 + ... + c_{k-1}\lambda_k\mathbf{x}_{k-1}) &= c_1\lambda_1\mathbf{x}_1 + ... + c_{k-1}\lambda_{k-1}\mathbf{x}_{k-1} \\
\mathbf{0} &= c_1(\lambda_1 - \lambda_k)\mathbf{x}_1 + ... + c_{k-1}(\lambda_{k-1} - \lambda_k)\mathbf{x}_{k-1}.
\end{aligned}$$

From the previous page we have

$$\mathbf{0} = c_1(\lambda_1 - \lambda_k)\mathbf{x}_1 + ... + c_{k-1}(\lambda_{k-1} - \lambda_k)\mathbf{x}_{k-1}$$

Thus $\mathbf{x}_1, ..., \mathbf{x}_{k-1}$ are lin. dep., contradicting the hypothesis. Therefore $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ are linearly independent.

At this point we have proven:

If A has n distinct eigenvalues, it must therefore have n lin. indep. eigenvectors and therefore must be diagonalizable. This completes the proof. □

8.2 Computing eigenvalues and eigenvectors

Question: Given square matrix A, how do you compute eigenvalues and eigenvectors of A?

Answer # 1: If the size of a matrix is large (larger than 3×3, usually use a computer. In *Mathematica*, you can find the eigenvalues and eigenvectors of a matrix A by typing in the matrix and saving it as A, then executing

```
Eigensystem[A]
```

For example, suppose $A = \begin{pmatrix} 2 & -5 \\ -1 & 2 \end{pmatrix}$. After typing in this matrix, the command `Eigensystem[A]` produces the output

```
{{2 + √5, 2 - √5}, {{-√5,1},{√5,1}}}
```

This means the eigenvalues of A are $\lambda_1 = 2 + \sqrt{5}$ and $\lambda_2 = 2 - \sqrt{5}$, and the corresponding eigenvectors are $\mathbf{x}_1 = (-\sqrt{5}, 1)$ and $\mathbf{x}_2 = (\sqrt{5}, 1)$, i.e. $A = S\Lambda S^{-1}$ where

$$S = \begin{pmatrix} -\sqrt{5} & \sqrt{5} \\ 1 & 1 \end{pmatrix} \quad \Lambda = \begin{pmatrix} 2+\sqrt{5} & 0 \\ 0 & 2-\sqrt{5} \end{pmatrix}$$

Answer # 2: Suppose λ_j is an eigenvalue of A. Then $A\mathbf{x}_j = \lambda_j \mathbf{x}_j$ for some nonzero vector $\mathbf{x}_j$. Rewriting this, we get

$$A\mathbf{x}_j - \lambda_j \mathbf{x}_j = \mathbf{0}$$

Thus $\mathbf{x}_j \in N(A - \lambda_j I)$. Since $\mathbf{x}_j \neq \mathbf{0}$, that means $N(A - \lambda_j I) \neq \{\mathbf{0}\}$ so $A - \lambda_j I$ is not invertible. That means $\det(A - \lambda_j I) = 0$. We have shown:

Definition 8.7 *Let $A \in M_n(\mathbb{R})$. The* **characteristic polynomial** *of A is the expression*

$$p_A(x) = \det(A - xI).$$

(This is a polynomial of degree n.)

Theorem 8.8 *Let $A \in M_n(\mathbb{R})$. λ is an eigenvalue of A if and only if it is a root of the characteristic polynomial of A (i.e. $\det(A - \lambda I) = 0$).*

Remark: Suppose A is some matrix such that $p_A(x)$ factors as

$$p_A(x) = (x+1)(x-2)(x-3)^2(x-6)^4.$$

Then (for example), $\lambda = 3$ is an eigenvalue of A. Since 3 is a "double root" of $p_A(x)$, we say $\lambda = 3$ is an **eigenvalue of multiplicity 2**. Similarly, for this matrix A, its eigenvalues are

The right way to list the eigenvalues of a matrix is to repeat them according to their multiplicities, i.e. the eigenvalues of the A with the characteristic polynomial as above are

Theorem 8.9 *Assume the eigenvalues of a matrix A are listed according to their multiplicities. Then:*

1. *The sum of the eigenvalues is $tr(A)$.*
2. *The product of the eigenvalues is $\det(A)$.*

In other words, for the A whose characteristic polynomial

$$p_A(x) = (x+1)(x-2)(x-3)^2(x-6)^4,$$

we have

$$tr(A) =$$

$$\det(A) =$$

Example: Find eigenvalues and eigenvectors for the matrix

$$A = \begin{pmatrix} 13 & -4 \\ -4 & 7 \end{pmatrix}.$$

Putting all this together

Example: Find A^{45} if

$$A = \begin{pmatrix} 5 & 8 & 16 \\ 4 & 1 & 8 \\ -4 & -4 & -11 \end{pmatrix}.$$

Example: Find the eigenvalues of the matrix

$$A = \begin{pmatrix} 3 & 0 & 1 & 4 \\ 0 & -5 & 1 & 3 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 7 \end{pmatrix}.$$

This example generalizes:

Theorem 8.10 *The eigenvalues of a triangular matrix are its diagonal entries.*

Example: Find eigenvalues and eigenvectors for the matrix

$$A = \begin{pmatrix} -2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Application: The Fibonacci numbers $\{f_n\}$ are defined by setting $f_0 = f_1 = 1$ and setting $f_n = f_{n-2} + f_{n-1}$ for $n \geq 2$. Find a formula for the 200^{th} Fibonacci number.

8.3 Matrix exponentials

In Math 230, you learn the following formula for the exponential function e^x:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + ...$$

This formula can be used as the definition of the exponential of a square matrix A:

Definition 8.11 *Let $A \in M_n(\mathbb{R})$ be a square matrix. The* **exponential** *of A, denoted $\exp(A)$ or e^A, is the $n \times n$ matrix*

$$\exp(A) = \sum_{n=0}^{\infty} \frac{1}{n!}A^n = I + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \frac{1}{4!}A^4 + ...$$

It can be shown that this series "converges" for every matrix A (but that is beyond the scope of Math 322).

Theorem 8.12 (Properties of matrix exponentials)

1. *If Λ is diagonal with diagonal entries $\lambda_1, \lambda_2, ..., \lambda_n$, then*

$$\exp(\Lambda) = \begin{pmatrix} e^{\lambda_1} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2} & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & e^{\lambda_n} \end{pmatrix}.$$

2. *If $A = S\Lambda S^{-1}$, then $\exp(A) = S\exp(\Lambda)S^{-1}$.*

3. *If t is a scalar, then $\exp(tA) = \exp(At)$.*

PROOF (1)

(2) Recall that if $A = S\Lambda S^{-1}$, then $A^n = S\Lambda^n S^{-1}$ for all n. Therefore

$$\begin{aligned} e^A &= e^{S\Lambda S^{-1}} \\ &= \sum_{n=0}^{\infty} \frac{(S\Lambda S^{-1})^n}{n!} \end{aligned}$$

(3) Follows from the fact that $tA = At$. □

<u>**WARNING:**</u> $e^{A+B} \neq e^A e^B$ in general (but $e^{A+B} = e^A e^B$ if $AB = BA$).

This theorem suggests that to compute the exponential of a matrix, you should diagonalize it (similar to how you would compute a power of a matrix).

Example: Find $\exp(A)$ if

$$A = \begin{pmatrix} 3 & 0 & 2 \\ 0 & 3 & -2 \\ 2 & -2 & 1 \end{pmatrix}$$

Step 1: Find eigenvalues and eigenvectors:

$$\det(A - xI) = (5 - x)(3 - x)(-1 - x) \Rightarrow \lambda = 5, \lambda = 3, \lambda = -1$$

Eigenvectors:

$$\begin{aligned} \lambda = 5 &\leftrightarrow (1, -1, 1) \\ \lambda = 3 &\leftrightarrow (1, 1, 0) \\ \lambda = -1 &\leftrightarrow (1, -1, -2) \end{aligned}$$

Step 2: Diagonalize A:

$$A = S\Lambda S^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 0 & -2 \end{pmatrix} \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 0 & -2 \end{pmatrix}^{-1}$$

Step 3: Use the theorem above to write the exponential:

$$\begin{aligned}
\exp(A) &= S\exp(\Lambda)S^{-1} \\
&= \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 0 & -2 \end{pmatrix} \begin{pmatrix} e^5 & 0 & 0 \\ 0 & e^3 & 0 \\ 0 & 0 & e^{-1} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 0 & -2 \end{pmatrix}^{-1} \\
&= \begin{pmatrix} \frac{-1}{2}e^{-1} + \frac{1}{2}e^3 + e^5 & \frac{1}{2}e^{-1} + \frac{1}{2}e^3 - e^5 & e^{-1} - e^5 \\ \frac{1}{2}e^{-1} + \frac{1}{2}e^3 - e^5 & \frac{-1}{2}e^{-1} + \frac{1}{2}e^3 + e^5 & -e^{-1} + e^5 \\ -e^{-1} + e^5 & e^{-1} - e^5 & 2e^{-1} - e^5 \end{pmatrix}.
\end{aligned}$$

Note: If you were asked to compute $\exp(At)$, you would start with the same Step 1 as above. Then Step 2 would be

$$At = S(\Lambda t)S^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 0 & -2 \end{pmatrix} \begin{pmatrix} 5t & 0 & 0 \\ 0 & 3t & 0 \\ 0 & 0 & -t \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 0 & -2 \end{pmatrix}^{-1}$$

and Step 3 would be

$$\begin{aligned}
\exp(At) &= S\exp(\Lambda t)S^{-1} \\
&= \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 0 & -2 \end{pmatrix} \begin{pmatrix} e^{5t} & 0 & 0 \\ 0 & e^{3t} & 0 \\ 0 & 0 & e^{-t} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 0 & -2 \end{pmatrix}^{-1} \\
&= \begin{pmatrix} \frac{-1}{2}e^{-t} + \frac{1}{2}e^{3t} + e^{5t} & \frac{1}{2}e^{-t} + \frac{1}{2}e^{3t} - e^{5t} & e^{-t} - e^{5t} \\ \frac{1}{2}e^{-t} + \frac{1}{2}e^{3t} - e^{5t} & \frac{-1}{2}e^{-t} + \frac{1}{2}e^{3t} + e^{5t} & -e^{-t} + e^{5t} \\ -e^{-t} + e^{5t} & e^{-t} - e^{5t} & 2e^{-1} - e^{5t} \end{pmatrix}.
\end{aligned}$$

Question: Why are we interested in computing the exponential of a matrix?

Answer:

Example: Suppose y is a function of t which satisfies $y'(t) = ay(t)$ where a is a constant. Find $y(t)$.

Example: Suppose y_1 and y_2 are functions of t such that

$$\left\{ \begin{array}{l} y_1'(t) = a_{11}y_1(t) + a_{12}y_2(t) \\ y_2'(t) = a_{21}y_1(t) + a_{22}y_2(t) \end{array} \right. .$$

Describe a method for finding $y_1(t)$ and $y_2(t)$.

Example: Suppose y_1 and y_2 are functions of t such that

$$\begin{cases} y_1'(t) = 2y_1(t) - 4y_2(t) \\ y_2'(t) = -y_1(t) - y_2(t) \end{cases} .$$

Assuming that $y_1(0) = 2$ and $y_2(0) = 1$, find $y_1(t)$ and $y_2(t)$.

The preceding method can also be used to solve differential equations for a single function which involve some higher-order derivatives.

Example: Find all functions $y(t)$ which satisfy the differential equation

$$y''(t) - 4y'(t) - 21y(t) = 0.$$

Appendix A

Mathematica commands

Appendix B

Mathematica information

B.1 General *Mathematica* principles

Mathematica is an extremely useful and powerful software package / programming language invented by a mathematician named Stephen Wolfram. Early versions of *Mathematica* came out in the late 1980s and early 1990s; the most recent version (which is loaded onto machines at FSU as of 2016) is *Mathematica* 11.

Mathematica does symbolic manipulation of mathematical expressions; it solves all kinds of equations; it has a library of important functions from mathematics which it recognizes while doing computations; it does 2- and 3-dimensional graphics; it has a built-in word processor tool; it works well with Java and C++; etc. One thing it doesn't do is prove theorems, so it is less useful for a theoretical mathematician than it is for an engineer or college student.

A bit about how *Mathematica* works: When you use the *Mathematica* program, you are actually running *two* programs. The "front end" of *Mathematica* is the part that you type on and the part you see. This part actually resides on the machine at which you are seated. The "kernel" is the part of *Mathematica* that actually does the calculations. If you type in `2 + 2` and hit [SHIFT]+[ENTER], the front end "sends" that information to the kernel which actually does the computation. The kernel then "sends" the result back to the front end, which displays the output `4` on the screen. Essentially, the way one uses *Mathematica* is by typing some "stuff" in, hitting [SHIFT]+[ENTER] to execute that stuff, and getting some output back from the program.

About *Mathematica* notebooks and cells: The actual files that *Mathematica* produces that you can edit and save are called *notebooks* and carry the file designation `*.nb`; they take up little space and can easily be saved to Google docs or on a flash drive, or emailed to yourself if you want them somewhere you can retrieve them. **Suggestion:** when saving any file, include the date in the file name (so it is easier to remember which file you are supposed to be open).

A *Mathematica* notebook is broken into *cells*. A cell can contain text, input, or output. A cell is indicated by a dark blue, right bracket (a "]") on the right-hand side of the notebook. To select a cell, click that bracket. This highlights the "]" in blue. Once selected, you can cut/copy/paste/delete cells as you would highlighted blocks of text in a Word document.

To change the formatting of a cell, select the cell, then click"Format / Style" and select the style you want. You may want to play around with this to see what the various styles look like. There are three particularly important styles:

- **input:** this is the default style for new cells you type
- **output:** this is the default style for cells the kernel produces from your commands
- **text:** changing a cell to text style allows you to make comments in between the calculations

Executing mathematical commands: To execute an input cell, put the cursor anywhere in the cell and hit [ENTER]. Well, not any [ENTER]; you have to use the [ENTER] on the numeric keypad at the far-right edge of the keyboard. The [ENTER] next to the apostrophe key (a.k.a. [RETURN]) gives you only a carriage return. You can also hold down the [SHIFT] key and hit either [ENTER] or [RETURN] to execute a command.

Important general concepts re: *Mathematica* syntax

1. **Multiplication:** use a star or a space: `2 * 3` or `2 3` will multiply numbers; `a x` means a times x; `ax` means the variable ax (in *Mathematica*, variables do not have to be named after one letter; they can be named by words or other strings of characters as well).

2. **Parentheses:** used for grouping and multiplication only. Parentheses mean "times" in *Mathematica*, and always mean that you intend to **multiply** what is in front of the parenthesis by what is inside the parenthesis.

3. **Brackets:** must be used to surround the input of any function or built-in *Mathematica* command. For example, to evaluate a function $f(x)$, you would type `f[x]`, not `f(x)`. Essentially, square brackets mean "**of**" in *Mathematica*.

4. **Capitalization:** All *Mathematica* commands and built-in functions begin with capital letters. For example, to find the sine of π, typing `sin(pi)` or `sin[pi]` does you no good (the first version would be the variable "sin" times the variable "pi", for instance). The correct syntax is `Sin[Pi]`. Similarly, e is `E` and i is `I` in *Mathematica*.

5. **Spaces:** *Mathematica* commands do not have spaces in them; for example, the inverse function of sine is `ArcSin`, not `Arc Sin` or `Arcsin`.

6. **Pallettes:** Lots of useful commands are available on the Basic Math Assistant Pallette, which can be brought up by clicking "Pallettes / Basic Math Assistant" on the toolbar. If you click on a button in the pallette, what you see appears in the cell. The tab halfway down this palette marked $d \int \Sigma$ has calculus commands, and the tab to the right of the $d \int \Sigma$ has matrix commands.

7. **Logarithms:** *Mathematica* does not know what `Ln` is. For natural logarithms (base e), type "Log[ ]". For common logarithms (base 10), type "Log10[ ]".

8. $\%$ refers to the last output (like "Ans" on a TI-calculator).

9. **Help:** To get help on a command, type "?" followed by the command you don't understand. If necessary, click the $\gg$ you get at the end of the help blurb to open a help browser. You can also find out how to do lots of stuff in *Mathematica* by using Google: search for what you want help on.

10. *Mathematica* gives exact answers (i.e. not decimals) for everything if possible. If you need a decimal approximation, use the command `N[ ]`. For example, `N[Pi]` spits out 3.14159...

11. If *Mathematica* freezes up in the middle of a calculation, click "Evaluation / Abort Evaluation" on the toolbar.

B.2 *Mathematica* quick reference guides

Basic operations

	Expression	*Mathematica* syntax
SPECIAL SYMBOLS	e	`E`
	π	`Pi`
	i (i.e. $\sqrt{-1}$)	`I`
	∞	`Infinity` (or use Basic Math Assistant palette)
ARITHMETIC	$3+4x$	`3 + 4x`
	$5-7$	`5 - 7`
	$8z$	`8z` or `8 z` or `8 * z`
	xy	`x y` (don't forget the space)
	$\frac{7}{3}$	`7/3`
	$\frac{x-7+2y}{a-7b}$	To get the fraction bar, type [CONTROL]+/
		then use [TAB] to move between the top and bottom
	$\sqrt{32}$	`Sqrt[32]`
		(or type [CONTROL]+2 to get a $\sqrt{\ }$ sign)
		(or use Basic Math Assistant palette)
	$\sqrt[4]{40}$	`40^(1/4)` (or use Basic Math Assistant palette)
	$\lvert x-3\rvert$	`Abs[x-3]`
	$30!$ (factorial)	`30!`
EXPS AND LOGS	$\ln 3$	`Log[3]`
	$\log_6 63$	`Log[6,63]`
	$\log 18$	`Log10[18]` or `Log[10,18]`
	2^{7y}	`2^(7y)`
		(or type `2`, then [CONTROL]+6, then `7y`)
		(or use Basic Math Assistant palette)
	e^{x-5+x^2}	`E^(x-5+x^2)` or `Exp[x-5+x^2]`
		(or use Basic Math Assistant palette)
TRIG	$\sin\pi$	`Sin[Pi]`
	$\cos(x(y+1))$	`Cos[x(y+1)]`
	$\cot\left(\frac{2\pi}{3}+\frac{3\pi}{4}\right)$	`Cot[2 Pi/3 + 3 Pi/4]`
	$\arctan 1$	`ArcTan[1]`

Objective	*Mathematica* syntax
To call the preceding output	`%`
To get a decimal approximation to the preceding output	`N[%]` (or click **numerical value**)

B.3 Vector and matrix operations

Entering a vector into *Mathematica*

To type in a vector, you need only one set of braces, so if you execute

```
b = {1,2,3}
```

this defines the vector b to be $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$. *Mathematica* knows to treat vectors as columns.

Entering a matrix into *Mathematica*

To store a matrix as a variable, there are two methods:

1. Type the matrix in using braces very carefully. For example, to save the matrix

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

   as A, execute

   ```
   A = {{2,4,7}, {-5,3,1}}
   ```

   Note that the entries are separated by commas, every row of the matrix needs braces around it, and the entire matrix needs braces around it.

2. Use the Basic Math Assistant Pallette. Click "Pallettes" and "Basic Math Assistant", then on the Basic Math Assistant click the fourth tab under "Basic Commands" that looks like a matrix. In the *Mathematica* notebook, type `A=`, then click the large button that looks like a matrix, then click "AddRow" or "Add Column" until the matrix is the appropriate size. Click in each box of the matrix and type in the appropriate numbers. For example, your command to define the A above would look like

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

Matrix operations

To display the answer as a matrix, add `// MatrixForm` to the end of the command.

Objective	*Mathematica* syntax
Find (i, j)–entry of A	`A[[`i`,`j`]]`
Find trace of A	`Tr[A]`
Find transpose of A	`Transpose[A]`
Add two matrices of the same size	`A + B`
Multiply a matrix by a scalar	`4A` `t A`
Matrix multiply A and B	`A.B` (don't forget the period)
Find inverse of square matrix A	`Inverse[A]`
Find reduced row-echelon form of A	`RowReduce[A]`
Find particular solution of $A\mathbf{x} = \mathbf{b}$	`LinearSolve[A,b]`
Find basis of $N(A)$	`NullSpace[A]`
Find rank of A	`MatrixRank[A]`
Find least-squares solution $\widehat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$	`LeastSquares[A,b]`
Find determinant of A	`Det[A]`
Find eigenvalues and eigenvectors of A	`Eigensystem[A]`
Find eigenvalues of A	`Eigenvalues[A]`
Find eigenvectors of A	`Eigenvectors[A]`
Find $\det(A - xI)$	`CharacteristicPolynomial[A,x]`
Determine if A is diagonalizable or not	`DiagonalizableMatrixQ[A]`
Find n^{th} power of a matrix	`MatrixPower[A,`n`]` (not `A^n`)
Find exponential of a matrix	`MatrixExp[A]`

Vector operations

Objective	*Mathematica* syntax
Add two vectors **v** and **w**	`v + w`
Multiply vector by a constant	`4v` `t v`
Find j^{th} entry of **v**	`v[[j]]`
Find dot product of **v** and **w**	`v.w`
Find norm of **v**	`Norm[v]`
Find distance between **v** and **w**	`EuclideanDistance[v,w]`
Find unit vector in same direction as **v**	`Normalize[v]`
Find projection $\pi_{\mathbf{w}}(\mathbf{v})$	`Projection[v,w]`
Find angle between **v** and **w**	`VectorAngle[v,w]` (answer will be in radians)
Perform the Gram-Schmidt procedure on $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$	`Orthogonalize[`$v_1, v_2, ...v_n$`]`
Find cross product of **v** and **w**	`Cross[v,w]`

Index