Contents

Chapter 1

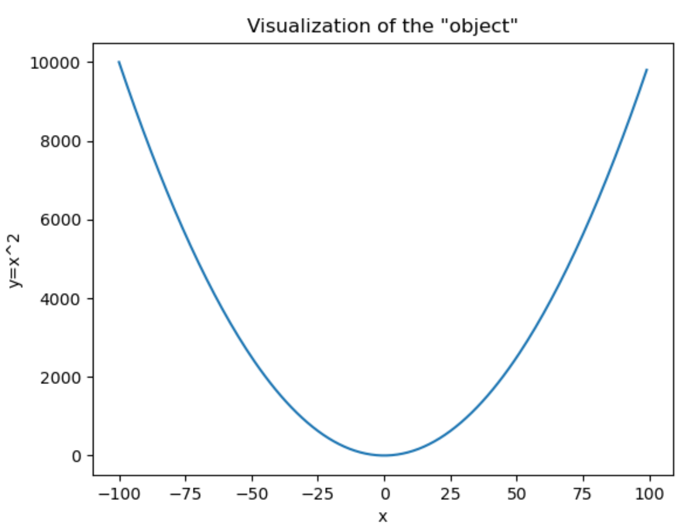
Coordinate Geometry

1.1 Introduction

Many of you have encountered some form of coordinate geometry in high school. For instance, the “standard” way to visualize a graph e.g. $f(x) = x^2$ is to visualize the points in 2-D space (x, y) where $y = x^2$. We give a demonstration in Python code.

```
import matplotlib.pyplot as plt
import numpy as np
X=np.arange(-100,100) #create list of numbers from -100 to 100
Y= X**2 #calculate the square of at each x
plt.plot(X,Y) #plot all the pairs of points in 2d plane
plt.xlabel('x')
plt.ylabel('y=x^2')
plt.title('Visualization of the "object"')
plt.show()
plt.show()
```

$x \longrightarrow x$
the “graph” object



This is also known as the Cartesian plane, named after René Descartes who invented it in the 17th century.

1.2 Visualization of geometric objects

The Cartesian plane allows us to describe shapes with equations and perform calculations with them. We first define the playing field (the Cartesian plane and higher dimensional analogues) and the players.

Definition 1.1 (Real numbers)

The set of **real numbers**, denoted as $\mathbb{R}$, is (informally) the set of all the numbers that can be written out in decimal form.

Example 1.2

The following are real numbers:

1. The integers 0, ± 1, ± 2, ...
2. Fractions in the form $\frac{a}{b}$, where a and $b \neq 0$ are integers.
3. Irrational numbers $\sqrt{2}$, π.

Remark. *The set of real numbers is known as a* ***complete field****. The definition of a complete field will be swept under the rug, but it guarantees a few things. The most important property: We will not "escape" the set by performing operations, possibly infinitely many.*

Definition 1.3 (N-dimensional space)

Let n be a positive integer. We denote the **n-dimensional real space** to be $\mathbb{R}^n$, consisting of all the n-tuples $(x_1, x_2, x_3, ..., x_n)$, where each x_j is a real number.
We call an n-tuple $(x_1, ..., x_n)$ a **point**, and two points $(x_1, ..., x_n)$ and $(y_1, ..., y_n)$ are equal if $x_j = y_j$ for all j-th entries of the tuples.

Remark. *We sometimes use* $\mathbf{x}$ *to denote* $(x_1, ..., x_n)$ *to make notation cleaner.*

1.2.1 Lines

Now that we have introduced the playing field of n-dimensional space, we can start translating the axioms of euclidean geometry to this coordinate system.

Definition 1.4 (Lines)

In euclidean geometry, a line is defined by two points. We let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. The **line** going from $\mathbf{x}$ to $\mathbf{y}$ is denoted $\overrightarrow{\mathbf{xy}}$.

What would this line look like? To get from $\mathbf{x}$ to $\mathbf{y}$, we have to traverse $y_1 - x_1$ units in the first coordinate, $y_2 - x_2$ units in the second, ..., $y_n - x_n$ in the last. We thus have a natural notation for the line $\overrightarrow{\mathbf{xy}}$.

$$\overrightarrow{\mathbf{xy}} = (y_1 - x_1, y_2 - x_2, ..., y_n - x_n).$$

This is very similar to a point as an n-tuple, but this is "spiritually" different to a point. This tuple represents the direction of line. One way to think of the correspondence between $(x_1, ..., x_n)$

point and $(x_1, ..., x_n)$ line is that $(x_1, ..., x_n)$ line is the line connecting $(0, 0, ...0)$ to $(x_1, ..., x_n)$ point. Because of this, we can identify a tuple as both the point and the line, and we call it a "vector" to abstract away from the actual geometric meaning.

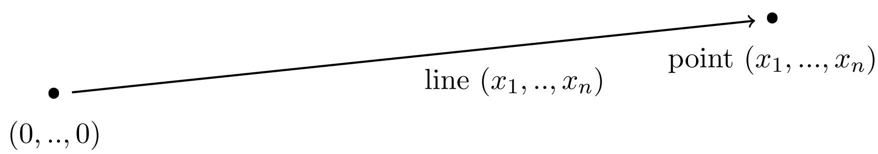


Figure 1.1: The correspondence between a line "vector" and a point "vector".

Notation. *Vectors are a very general notation of* $n-$*tuples. Depending on context, we use both of the following notations to denote the entries of* $\vec{v} \in \mathbb{R}^n$

- *"Ordered sets"* $(v_1, ..., v_n)$*, suitable dealing with points (and other geometric objects). It also looks cleaner when writing inline.*
- *"Column vectors"* $\begin{bmatrix} v_1 \\ . \\ . \\ . \\ v_n \end{bmatrix}$ *when the ordered sets are complicated to read, and when working with matrix algebra.*

1.2.2 Operation with lines

We need to translate a few more things from euclidean geometry.

Proposition 1.5

Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$, then

$$\overrightarrow{\mathbf{xy}} + \overrightarrow{\mathbf{yz}} = \overrightarrow{\mathbf{xz}},$$

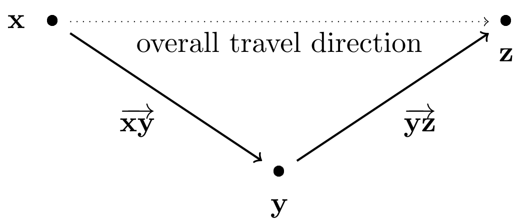
where $(a_1, ..., a_n) + (b_1, ..., b_n) \stackrel{\text{def}}{=} (a_1 + b_1, ..., a_n + b_n)$.

Proof.

$$\begin{aligned} \overrightarrow{\mathbf{xy}} + \overrightarrow{\mathbf{yz}} &= (y_1 - x_1, ..., y_1 - x_n) + (z_1 - y_1, ..., z_1 - y_n) \\ &= (y_1 - x_1 + z_1 - y_1, ..., y_n - x_n + z_n - y_n) \\ &= (z_1 - x_1, ..., z_n - x_n) \\ &= \overrightarrow{\mathbf{xz}} \end{aligned}$$

□

Geometrically, this means if you connect **x** to **y** to **z**, the overall "direction of travel" you make is **x** to **z**. This gives us a natural extension for addition of vectors by considering each entry. Similarly for scaling vectors, we just scale the entries along each dimension.



Definition 1.6 (Addition and scaling of vectors)

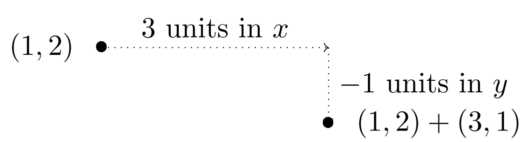
Let $\vec{a}, \vec{b}$ be two vectors in $\mathbb{R}^n$. We define the sum/difference of $\vec{a}$ and $\vec{b}$

$$\vec{a}+\vec{b} \stackrel{\text{def}}{=} (a_1+b_1, ..., a_n+b_n), \quad \vec{a}-\vec{b} \stackrel{\text{def}}{=} (a_1-b_1, ..., a_n-b_n)$$

and the scaling of $\vec{a}$ by a real number $c \in \mathbb{R}$

$$c\vec{a} \stackrel{\text{def}}{=} (ca_1, ca_2, ..., ca_n).$$

Remark. *Here we use the term "vectors", as we can in essence add points and lines together. How does one make sense of adding a line to a point? We can view this as translating the point along the path of the line, for instance, let us translate the point* $(1,2)$ 3 *units in the first coordinate and* -1 *units in the second coordinate. This will give us* $(4,1)$.



This way, we can write the line from $\vec{x}$ to $\vec{y}$ as $\vec{y}-\vec{x}$. The proof is a computational exercise.

Notation. *We now transferred for talking about points and the lines between points to addition. Therefore, we can overload the notation for points and lines as a vector* $\vec{v}$*, keeping in mind that they have the same arithmetic structure.*

In fact, most of our intuition for the real numbers translates to $\mathbb{R}$. For formality, we will list them here; in practice, we (almost always) take these properties for granted.

Proposition 1.7

Let $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$, and $a, b \in \mathbb{R}$. Then the following hold:

- *(Associativity)* $(\vec{u}+\vec{v})+\vec{w} = \vec{u}+(\vec{v}+\vec{w})$.
- *(Commutativity)* $\vec{u}+\vec{v} = \vec{v}+\vec{u}$.
- *(Identity)* The zero vector $\vec{0} \overset{\text{def}}{=} (0, 0, ..., 0) \in \mathbb{R}$ satisfies $\vec{v}+\vec{0} = \vec{v}$.
- *(Inverse)* The inverse of $\vec{v}$, $-\vec{v} \overset{\text{def}}{=} (-v_1, ..., -v_n)$ satisfies $\vec{v}+(-\vec{v}) = \vec{0}$.
- *(Scalar multiplication)* $a(b\vec{v}) = (ab)\vec{v}$.
- *(Scalar Identity)* $1\vec{v} = \vec{v}$.
- *(Distributivity 1)* $a(\vec{u}+\vec{v}) = a\vec{u}+a\vec{v}$.
- *(Distributivity 2)* $(a+b)\vec{v} = a\vec{v}+b\vec{v}$.

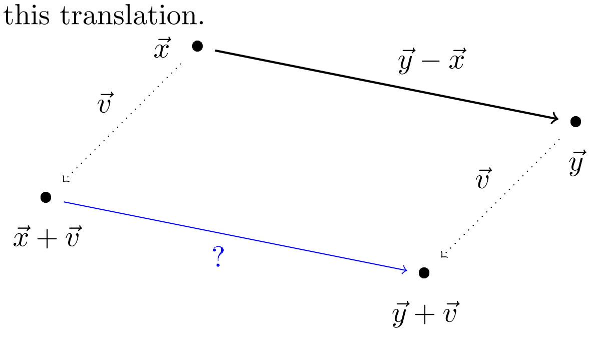
Remark. *All of these have good geometric intuition behind. For instance, the zero vector $\vec{0}$ is the "don't move" vector, corresponding to the point at the origin, or the "too short to be a line". The inverse of $\overrightarrow{\mathbf{xy}}$ is $\overrightarrow{\mathbf{yx}}$, where you go back from $\mathbf{x}$ to $\mathbf{y}$.*

Remark. *These 8 conditions are the axioms of a vector space. Later in the course, we will generalize the notion of vectors in $\mathbb{R}^n$ to other spaces (playing fields).*

Proposition 1.8

Lines are translation invariant. That is, for every $\vec{x}, \vec{y}, \vec{v} \in \mathbb{R}^n$, then the line from $\vec{x}$ to $\vec{y}$ is the same as the line from $\vec{x}+\vec{v}$ to $\vec{y}+\vec{v}$.

Let us illustrate what this statement is trying to convey. We have two points $\vec{x}$, $\vec{y}$; now we translate each of these points by $\vec{v}$, and we want the line between the points to be preserved under this translation.



The proof is one line: $(\vec{y}+\vec{v})-(\vec{y}+\vec{v}) = \vec{y}-\vec{x}+\vec{v}-\vec{v} = \vec{y}-\vec{x}$. However, an immediate consequence of this is that we can "transport" vectors in space without distorting the vector. Colloquially, *5 miles South* to you describes the same direction and length as *5 miles South* to a person a few feet away. This justifies the way we visualize the correspondence between points and vectors - we "transport" the vectors to start from the origin $(0, ..., 0)$, and the end describes the point.

Remark. *Translation (and scaling) invariance is a property of Euclidean geometry. There are some exotic geometry systems that distort distance and direction through translation and scaling. One such example is the Poincaré metric.*

Another notion we can carry from Euclidean geometry is parallel lines. Here we not only define what it means for two vectors to be parallel (never touching), we also give a definition for two vectors to be parallel but point in opposite directions.

Definition 1.9 (Parallel and Antiparallel Vectors)

Let non-zero vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$. We say that $\vec{v}$ and $\vec{w}$ are **parallel** if there is some $c > 0$ such that $\vec{v} = c\vec{w}$. We say that $\vec{v}$ and $\vec{w}$ are **antiparallel** if there is some $c < 0$ such that $\vec{v} = c\vec{w}$.

Example 1.10

Find an expression for the points on the (infinite) line passing through $P(1,1,0)$ and $Q(0,2,2)$.

Let M be a point on the line $\overrightarrow{PQ}$, then $\overrightarrow{PM}$ is parallel to $\overrightarrow{PQ}$ (or $M = P$). Either way, we would have some $c \in \mathbb{R}$ such that

$$\overrightarrow{PM} = c\overrightarrow{PQ}.$$

Indeed, we can confirm that any point represented as $P + c\overrightarrow{PQ}$ will be on the line, as $c\overrightarrow{PQ}$ is parallel to $\overrightarrow{PQ}$. One exception is to check for when $c = 0$, but we already know P is on the line. We thus have an expression for the line to be

$$(1,1,0) + c(-1,1,2)$$

for all $c \in \mathbb{R}$. Another way to put it, the line is the **image** of the function $f(c) = (1,1,0) + c(-1,1,2)$ on $\mathbb{R}$.

Remark. *This form of representing a line is known as a* ***parametric representation*** *or a* ***parametrization*** *of the line. We will give a more concrete definition of parametrization when we move beyond straight lines to "curvy" curves and surfaces in* **??**.

Remark. *Using the parametric representation for a line $l(t) = (P_1, P_2, ..., P_n) + t(v_1, v_2, ..., v_n)$, we can take slices/cross sections across each of the j-th coordinates, and get*

$$x_j = P_j + tv_j \implies t = \frac{x_j - P_j}{v_j}$$

is the equation of a line in 2d. P_j, v_j are fixed, viewing in the variables $y = x_j, x = t$, this is the equation of a line with y-intercept x_j and slope v_j. Setting the values of t for all slices to be equal,

$$t = \frac{x_1 - P_1}{v_1} = \frac{x_2 - P_2}{v_2} = ... = \frac{x_j - P_j}{v_j} = ... = \frac{x_n - P_n}{v_n}$$

This form is known as the ***symmetric equations*** *of a line.*

Example 1.11

Find an expression for the points on the line connecting **between** $P(1,1,0)$ to $Q(0,2,2)$.

This will be a segment of the line from the previous example, so the answer would be the same expression, but we limit the domain of f to be an interval on $\mathbb{R}$. Let us examine what f does to a few values of c.

c	very negative	$c = 0$	$c = 1/2$	$c = 1$	very positive
f(c)	very off the segment	P	midpoint of $\overline{PQ}$	Q	very off the segment

As we move from very negative c to very positive c, we start very away from the line segment, reach P at $c = 0$, Q at $c = 1$, then move away from the line segment. Indeed the values $(1, 1, 0) + c(-1, 1, 2)$ will be on the line segment for $c \in [0, 1]$.

Finally, the nice algebraic properties for adding and scaling vectors gives us a natural way to understand these vectors as a sum of n component vectors, one for each dimension. These *special* vectors deserve a name.

Definition 1.12 (Standard Basis Vectors and Vector Decomposition)

We denote in $\mathbb{R}^n$, the **standard basis vectors**

$$\begin{aligned}
\vec{e}_1 &\stackrel{\text{def}}{=} (1, 0, 0, ..., 0, 0, 0),\\
\vec{e}_2 &\stackrel{\text{def}}{=} (0, 1, 0, ..., 0, 0, 0),\\
\vec{e}_{n-1} &\stackrel{\text{def}}{=} (0, 0, 0, ..., 0, 1, 0),\\
\vec{e}_n &\stackrel{\text{def}}{=} (0, 0, 0, ..., 0, 0, 1),
\end{aligned}$$

so that $\vec{v} = \sum_{i=1}^{n} v_i \vec{e}_i$. In $\mathbb{R}^3$, we sometimes write

$$\vec{i} \stackrel{\text{def}}{=} \vec{e}_1, \quad \vec{j} \stackrel{\text{def}}{=} \vec{e}_2, \quad \vec{k} \stackrel{\text{def}}{=} \vec{e}_3$$

to accommodate the physicists.

Exercises

1. Compute $5\vec{v} - 2\vec{w}$ and $-3\vec{w}$ for the following pairs of vectors:

   (a) $\vec{v} = 2\vec{i} + 3\vec{j}$, $\vec{w} = 4\vec{i} - 9\vec{j}$

   (b) $\vec{v} = (1, 2, -1)$, $\vec{w} = (2, -1, 0)$

   (c) $\vec{v} = -2\vec{e}_3 + 4\vec{e}_5$, $\vec{w} = \vec{e}_1 - 4\vec{e}_5$

   (d) $\vec{v} = (\cos t, \sin t)$, $\vec{w} = (\cos t)\vec{e}_2 - (\sin t)\vec{e}_1$

2. Find parametric equations that describe the line that

   (a) passes through $(1, 2, 3)$ and $(4, -1, 2)$

   (b) passes through $(0, -2, 3)$ and $(1, 2, 3)$

1.3 Length, angles and projections

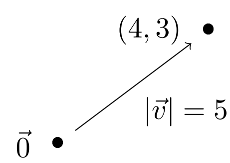
Definition 1.13 (Magnitude)

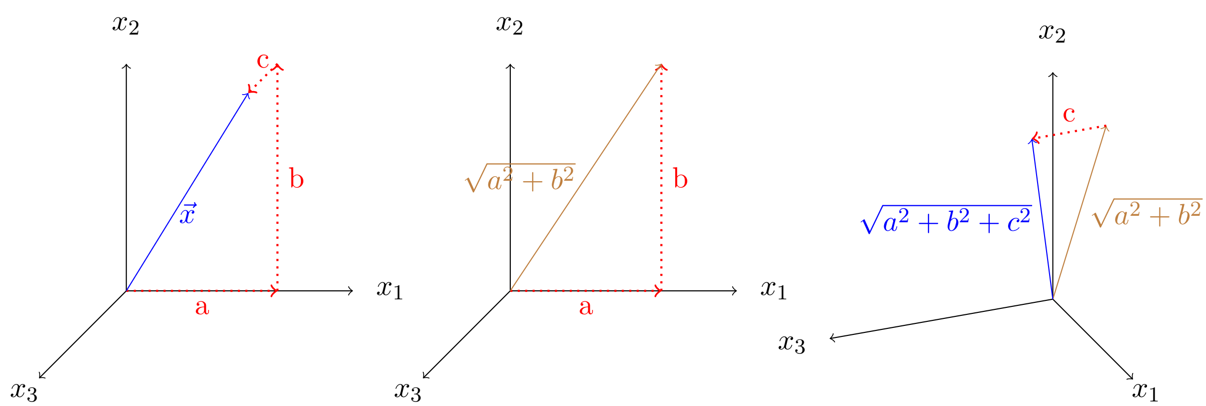
Let $\vec{v} \in \mathbb{R}^n$. The magnitude of $\vec{v}$ is denoted

$$|\vec{v}| \stackrel{\text{def}}{=} \sqrt{v_1^2 + v_2^2 + ...v_n^2}.$$

We build intuition through the lower dimensional cases. In $\mathbb{R}^2$, let us consider the point $(4, 3)$.

The magnitude of this vector is $\sqrt{3^2 + 4^2} = 5$. If this sounds very familiar, it is because this is indeed an application of Pythagorean theorem. In 3-dimensions, this still applies - take $\vec{x} = (a, b, c)$, we can traverse in each coordinate to apply Pythagorean theorem twice.





(a) $\vec{x} = (a, b, c)$. (b) Pyth. thm on a, b. (c) Reapply Pyth. thm.

Proposition 1.14

Let $\vec{v} \in \mathbb{R}^n, a \in \mathbb{R}$. Then

$$|a\vec{v}| = |a||\vec{v}|.$$

Definition 1.15 (Unit vector)

Let $\vec{v} \in \mathbb{R}^n$. We call $\vec{v}$ a **unit vector** if $|\vec{v}| = 1$.

Proposition 1.16

Let $\vec{v} \in \mathbb{R}^n$ be non-zero. Then there is a unique unit vector $|\vec{w}| = 1$ such that $\vec{v}$ is parallel to $\vec{w}$.

Proof. If $|\vec{w}| = 1, c > 0$ and $c\vec{w} = \vec{v}$, it must be that

$$c = |c||\vec{w}| = |c\vec{w}| = |\vec{v}|,$$

and

$$\vec{w} = \frac{1}{c}(c\vec{w}) = \frac{1}{|\vec{v}|}\vec{v}.$$

This shows uniqueness - there is no other possible candidate except for this vector. Now we can confirm the vector $\frac{1}{|\vec{v}|}\vec{v}$ is a unit vector and is parallel to $\vec{v}$, so there exists a unique unit vector that satisfies the properties.

□

We have now transported the notions of length and scaling into the coordinate system, and this allows us to make "measurements" such as angles and area.

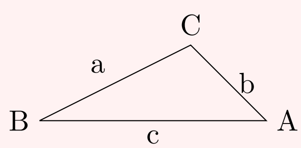
1.3.1 The Dot Product

Example 1.17

Let points A=$(5, 8)$, B=$(-2, 7)$, O=$(4, 5)$. Find the angle $\angle AOB$.

To solve this using only information about the lengths, recall the Law of Cosines:

Theorem 1.18 (Law of Cosines)

For any triangle 

,

$$c^2 = a^2 + b^2 - 2ab\cos(\angle BCA).$$

We now apply the Law of Cosines to $\triangle AOB$, so that

$$|\overrightarrow{AO}|^2 + |\overrightarrow{OB}|^2 - 2|\overrightarrow{AO}||\overrightarrow{OB}|\cos(\angle AOB) = |\overrightarrow{AB}|^2.$$

Plugging in the values, we solve

$$\begin{aligned} &((4-5)^2 + (5-8)^2) + ((-2-4)^2 + (7-5)^2) - 2\sqrt{(4-5)^2 + (5-8)^2} \\ &\times\sqrt{(-2-4)^2 + (7-5)^2}\cos(\angle AOB) = ((-2-5)^2 + (7-8)^2). \end{aligned}$$

Simplifying, we get

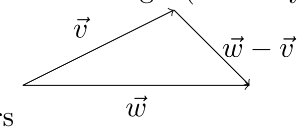
$$50 - 2\sqrt{10}\sqrt{40}\cos(\angle AOB) = 50 \implies \cos(\angle AOB) = 0.$$

So the angle is $\pi/2$.

Example 1.19

Find a closed form formula for the cosine of an angle formed by two vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$.

Let the angle formed be θ. Using the intuition from 2-D space, we can form a triangle (in a very



complex n-dimensional space). We write the Law of Cosine in terms of vectors

$$|\vec{v}|^2 + |\vec{w}|^2 - 2|\vec{v}||\vec{w}| = |\vec{w} - \vec{v}|^2.$$

This expands to

$$\sum_{i=1}^{n} v_i^2 + \sum_{j=1}^{n} w_j^2 - 2\sqrt{\sum_{i=1}^{n} v_i^2}\sqrt{\sum_{j=1}^{n} w_j^2}\ \cos(\theta) = \sum_{i=1}^{n} (w_i - v_i)^2.$$

Rearranging $(a-b)^2 = a^2 + b^2 - 2ab$,

$$2\sqrt{\sum_{i=1}^{n} v_i^2}\sqrt{\sum_{j=1}^{n} w_j^2}\ \cos(\theta) = \sum_{i=1}^{n} 2w_i v_i$$

so

$$\cos(\theta) = \frac{\sum_{i=1}^{n} w_i v_i}{\sqrt{\sum_{i=1}^{n} v_i^2}\sqrt{\sum_{j=1}^{n} w_j^2}}.$$

Definition 1.20 (Dot Product)

Let $\vec{v}, \vec{w} \in \mathbb{R}^n$, the **dot product** between $\vec{v}$ and $\vec{w}$ is

$$\vec{v} \cdot \vec{w} \overset{\text{def}}{=} \sum_{i=1}^{n} v_i w_i.$$

Proposition 1.21

Let $\vec{v} \in \mathbb{R}^n$, then $|\vec{v}|^2 = \vec{v} \cdot \vec{v}$.

Proposition 1.22

Let $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$, $c \in \mathbb{R}$. The dot product satisfies the following properties:

- *(Symmetry)* $\vec{v} \cdot \vec{w} = \vec{w} \cdot \vec{v}$.
- *(Linearity 1)* $(c\vec{v}) \cdot \vec{w} = c(\vec{v} \cdot \vec{w})(= \vec{v} \cdot (c\vec{w})$ by symmetry).
- *(Linearity 2)* $(\vec{v} + \vec{u}) \cdot \vec{w} = \vec{v} \cdot \vec{w} + \vec{u} \cdot \vec{w}$.
- *(Positive definiteness)* $\vec{v} \cdot \vec{v} \geq 0$, with equality if and only if $\vec{v} = \vec{0}$.

Proof. Here is the proof for the last property. The rest are left as an exercise We have $\vec{v} \cdot \vec{v} = |\vec{v}|^2 \geq= 0$, and if $\vec{v} \cdot \vec{v} = 0$, for all $1 \leq \jmath \leq n$, $v_j^2 \leq \sum_{i=1}^{n} v_i^2 = \vec{v} \cdot \vec{v} \leq 0$, so $v_j = 0$. This means $\vec{v} = \vec{0}$. ☐

Corollary 1.23:

1. $|\vec{v}-\vec{w}|^2 = |\vec{v}|^2 + |\vec{w}|^2 - 2\vec{v}\cdot\vec{w}$.
2. In lower dimensions $\mathbb{R}^2, \mathbb{R}^3$ If $\vec{v}, \vec{w} \neq \vec{0}$, the angle between $\vec{v}$ and $\vec{w}$ is
$$\cos^{-1}\left(\frac{\vec{v}\cdot\vec{w}}{|\vec{v}||\vec{w}|}\right).$$
In particular, the angle is $\pi/2$ when $\vec{v}\cdot\vec{w} = 0$.
3. $|\vec{v}\cdot\vec{w}| \leq |\vec{v}||\vec{w}|$.

Theorem 1.24 (Cauchy-Schwarz inequality)

Let $\vec{v}, \vec{u} \in \mathbb{R}^n$. Then
$$|\vec{v}\cdot\vec{u}| \leq |\vec{u}||\vec{v}|.$$

The proof is in one of the exercises at the end of this chapter.
The dot product allows us to make sense of “angles” in higher dimensions, so we can generalize the notion of two vectors “perpendicular” to each other.

Definition 1.25 (Orthogonality)

Let $\vec{v}, \vec{w} \in \mathbb{R}^n$. We say $\vec{v}$ and $\vec{w}$ are **orthogonal** to each other if $\vec{v}\cdot\vec{w} = 0$.

Example 1.26

Find all vectors $\vec{v} \in \mathbb{R}^3$ such that $\vec{v}\cdot(-3\vec{i} + 2\vec{j} + 6\vec{k}) = 49$. Sketch the locus of corresponding points in $\mathbb{R}^3$.

The condition expands to

$$\begin{aligned} -3v_1 + 2v_2 + 6v_3 &= 49 \\ \implies v_3 &= \frac{49 + 3v_1 - 2v_2}{6} \end{aligned}$$

so that the set of all vectors is

$$\left\{ s\vec{i} + t\vec{j} + \left(\frac{49+3s-2t}{6}\right)\vec{k} | s, t \in \mathbb{R} \right\}.$$

To plot this in 3D space, notice that this describes the equation of a plane $z = (49 + 3x - 2y)/6$. We pick any three points (of course we want those that are easy to calculate) $(0, 0, 49/6)$, $(0, 49/12, 0)$, $(-49/18, 0, 0)$.

```
def function_to_plot(X,Y):
return (49+3*X-2*Y)/6
#create figure
fig=plt.figure(figsize=(10,8))
ax = fig.add_subplot(projection='3d')
ax.view_init(elev=10, azim=120) #change view 1
```

```
#ax.view_init(elev=10, azim=0) #change view 2
X,Y=np.meshgrid(np.linspace(-10,10,10), np.linspace(-10,10,10))
Z=function_to_plot(X,Y)
surface=ax.plot_surface(X, Y, Z, alpha=0.5,
label='graph surface')
ax.set_xlabel('x')
ax.set_ylabel('y')
ax.set_zlabel('z')
#plot the curve and the cross section to integrate
ax.scatter(0,0,0,color='red',label='origin')
ax.text(0,0,0,'origin')
ax.scatter(-3,2,6,color='black')
ax.text(-3,2,6,'(-3,2,6)')
plt.legend()
plt.show()
```

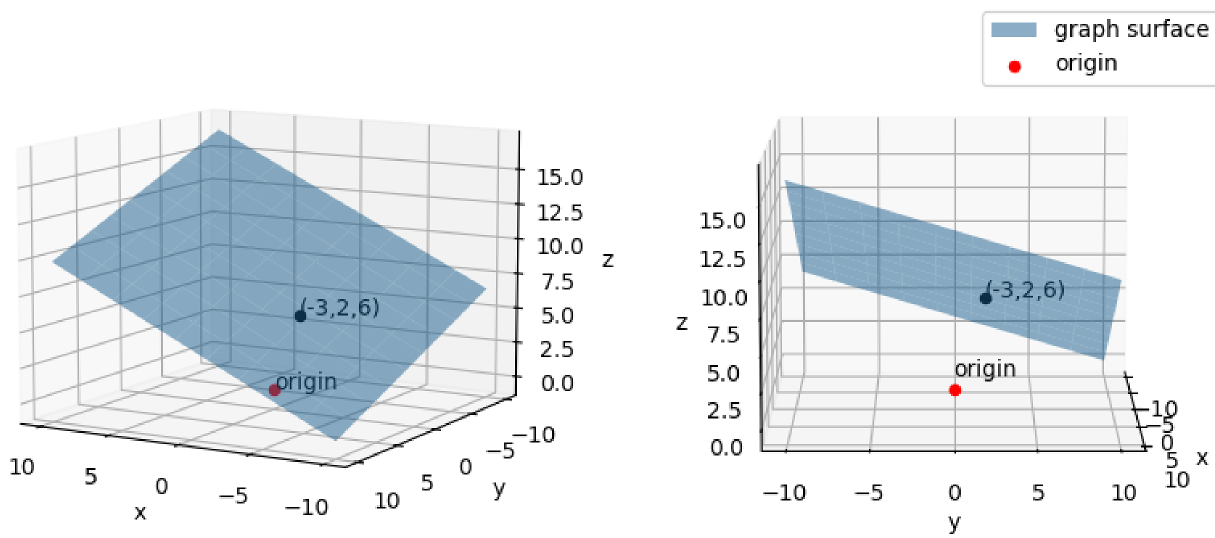


Figure 1.3: Two views of the plane

Interestingly, we see that $(-3, 2, 6)$ - the point corresponding to our vector - is a point on the plane!

Remark. *The form*

$$\left\{ s\vec{i} + t\vec{j} + \left(\frac{49 + 3s - 2t}{6} \right) \vec{k} | s, t \in \mathbb{R} \right\}$$

*is a **parametrization** of the plane described by* $-3x + 2y + 6z = 49$*, viewing this as a function* $\mathbb{R}^2 \to \mathbb{R}^3$.

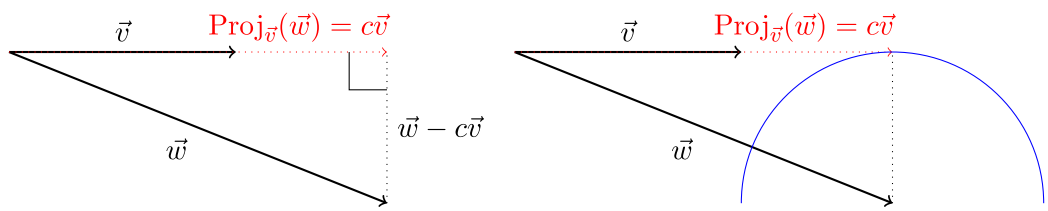
1.3.2 Projection

Definition 1.27 (Projection)

Let $\vec{v}, \vec{w} \in \mathbb{R}^n$, $\vec{v} \neq \vec{0}$, and $c \in \mathbb{R}$, such that $\vec{v} \cdot (\vec{w} - c\vec{v}) = 0$ (or equivalently, $\vec{w} - c\vec{v}$ orthogonal to $\vec{v}$).
We say that $c\vec{v} \overset{\text{def}}{=} \text{Proj}_{\vec{v}}(\vec{w})$ is the **vector projection of $\vec{w}$ on $\vec{v}$.**

To build intuition, it is always helpful to start with lower dimensions. We take a plane through $\vec{v}$ and $\vec{w}$. Looking at this slice, we can work in 2D.

We see that $c\vec{v}$ is the the point on the extension of $\vec{v}$ such that it is closest to the point $\vec{w}$. To geometrically show this idea, we draw a circle centered at $\vec{w}$ with radius $|\vec{w} - c\vec{v}|$. The line generated by $\vec{v}$ is tangent to this circle, so the contact point at $c\vec{v}$ is indeed the closest point to $\vec{w}$.



We also see from this geometric construction that 1) the projection $c\vec{v}$ is unique thus well defined, 2) $c|\vec{v}| = |\vec{w}|\cos\theta$

Proposition 1.28

The projection is unique and is given by

$$\text{Proj}_{\vec{v}}(\vec{w}) = \frac{\cos\theta|\vec{w}|}{|\vec{v}|} = \frac{\vec{v} \cdot \vec{w}}{|\vec{v}|^2} = \left(\frac{\vec{v} \cdot \vec{w}}{\vec{v} \cdot \vec{v}}\right)\vec{v}.$$

This is the first time we show that something is unique. The standard argument goes as follows: Suppose another object a' satisfies all the properties you want for a. Then you can show $a = a'$, meaning every object that satisfies the properties is a, or equivalently, a is unique.

So what happens if we pick two points P, P' on $\vec{v}$ such that both make a right angle when connected to $\vec{w}$? We would have constructed a triangle $\triangle PP'\vec{w}$ with two right angles!

Proof. Let $c, c' \in \mathbb{R}$ such that $\vec{p}_1 = c\vec{v}$ and $\vec{p}_2 = c'\vec{v}$ are both satisfy the definition of $\text{Proj}_{\vec{v}}(\vec{w})$. We want to show that $c = c'$, so we consider $\vec{u} = (c - c')\vec{v}$, the line between the two projections.
By corollary 1.23, we have two equations

$$|\vec{w} - \vec{p}_1|^2 = |\vec{w} - \vec{p}_2|^2 + |\vec{u}|^2$$
$$|\vec{w} - \vec{p}_2|^2 = |\vec{w} - \vec{p}_1|^2 + |\vec{u}|^2$$

We can solve for $|\vec{u}| = 0$, and by positive definiteness, $\vec{u} = \vec{0}$ and $c - c' = |\vec{u}|/|\vec{v}| = 0$.
To show our formula for projection works, we can just compute that

$$\vec{v} \cdot \left(\vec{w} - \left(\frac{\vec{v} \cdot \vec{w}}{\vec{v} \cdot \vec{v}}\right)\vec{v}\right) = \vec{v} \cdot \vec{w} - \left(\frac{\vec{v} \cdot \vec{w}}{\vec{v} \cdot \vec{v}}\right)(\vec{v} \cdot \vec{v}) = 0.$$

□

Exercises

3. Determine if the following pairs of vectors are orthogonal:

   (a) $3\vec{e}_1 + 4\vec{e}_2$, $-4\vec{e}_1 + 3\vec{e}_2$

   (b) $(4, -1, 2)$, $(3, 0, -6)$

   (c) $3\vec{i} - 2\vec{j}$, $-2\vec{i} - 4\vec{k}$

   (d) $\vec{e}_1 + 3\vec{e}_3 + 5\vec{e}_5 + ... + (2k-1)\vec{e}_{2k-1}$, $2\vec{e}_2 + 4\vec{e}_4 + ... + (2k)\vec{e}_{2k}$

4. Refer to the plane in Example 1.26. Show that $\text{Proj}_{-3\vec{i}+2\vec{j}+6\vec{k}}(\vec{v})$ is the same for any vector $\vec{v}$ such that $\vec{v} \cdot (-3\vec{i} + 2\vec{j} + 6\vec{k}) = 49$, and compute this projection.

5. *(Geometry of a methane molecule)* Place four points $P(0,0,0)$, $Q(1,1,0)$, $R(1,0,1)$, $S(0,1,1)$ in $\mathbb{R}^3$.

   (a) Compute the distance between any two points of $PQRS$ and show that all 6 pairs are the same. This means that $PQRS$ forms a regular tetrahedron.

   (b) Verify that the geometric center of the tetrahedron $O(1/2, 1/2, 1/2)$ is equidistant to all of the vertices of the tetrahedron.

   (c) Compute the angle between two edges of the tetrahedron, rounded to 2 decimal places. By symmetry, you only have to compute one angle.

   (d) Compute the angle $\angle POQ$, rounded to 2 decimal places. Does this angle remind you of something from Chemistry?

1.4 Cross Product

Definition 1.29 (Cross Product)

Let $\vec{v}, \vec{w} \in \mathbb{R}^3$, we define the **cross product** of $\vec{v}$ and $\vec{w}$ to be

$$\vec{v} \times \vec{w} \stackrel{\text{def}}{=} \begin{bmatrix} v_2w_3 - v_3w_2 \\ v_3w_1 - v_1w_3 \\ v_1w_2 - v_2w_1 \end{bmatrix}.$$

Remark. *Later on, we will introduce determinants of a matrix. The cross product can be understood as the determinant of the 'matrix'*

$$\begin{bmatrix} \vec{i} & v_1 & w_1 \\ \vec{j} & v_2 & w_2 \\ \vec{k} & v_3 & w_3 \end{bmatrix}.$$

This definition seems a bit unmotivating, so let us work through some examples.

Example 1.30

Compute the 9 cross products for each pair of the standard basis vectors in $\mathbb{R}^3$.

With some (heavy) computation, we find

$$\begin{array}{lll} \vec{i}\times\vec{i}=\vec{0} & \vec{i}\times\vec{j}=\vec{k} & \vec{i}\times\vec{k}=-\vec{j} \\ \vec{j}\times\vec{i}=-\vec{k} & \vec{j}\times\vec{j}=\vec{0} & \vec{j}\times\vec{k}=\vec{i} \\ \vec{k}\times\vec{i}=\vec{j} & \vec{k}\times\vec{j}=-\vec{i} & \vec{k}\times\vec{k}=\vec{0} \end{array}$$

Importantly, the cross product is **antisymmetric**, meaning $\vec{v}\times\vec{w}=-\vec{w}\times\vec{v}$. You can see special cases with the standard basis from above and confirm the general case in an exercise.

Example 1.31

Compute $\vec{v}\cdot(\vec{v}\times\vec{w})$ and $\vec{w}\cdot(\vec{v}\times\vec{w})$.

Again with some heavy computation,

$$\begin{aligned} \vec{v}\cdot(\vec{v}\times\vec{w}) &= \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}\cdot\begin{bmatrix} v_2w_3-v_3w_2 \\ v_3w_1-v_1w_3 \\ v_1w_2-v_2w_1 \end{bmatrix} \\ &= v_1v_2w_3 - v_1v_3w_2 + v_2v_3w_1 - v_2v_1w_3 + v_3v_1w_2 - v_3v_2w_1 \\ &= 0, \\ \vec{w}\cdot(\vec{v}\times\vec{w}) &= \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}\cdot\begin{bmatrix} v_2w_3-v_3w_2 \\ v_3w_1-v_1w_3 \\ v_1w_2-v_2w_1 \end{bmatrix} \\ &= w_1v_2w_3 - w_1v_3w_2 + w_2v_3w_1 - w_2v_1w_3 + w_3v_1w_2 - w_3v_2w_1 \\ &= 0. \end{aligned}$$

Which means $\vec{v}\times\vec{w}$ is orthogonal to $\vec{v}$ and $\vec{w}$!

Example 1.32

Compute $|\vec{v}\times\vec{w}|^2+(\vec{v}\cdot\vec{w})^2$.

We have

$$\begin{aligned}
|\vec{v}\times\vec{w}|^2+(\vec{v}\cdot\vec{w})^2 =&(v_2w_3-v_3w_2)^2+(v_3w_1-v_1w_3)^2+(v_1w_2-v_2w_1)^2\\
&+(v_1w_1+v_2w_2+v_3w_3)^2\\
=&(v_1w_1)^2+(v_1w_2)^2+(v_1w_3)^2\\
&+(v_2w_1)^2+(v_2w_2)^2+(v_2w_3)^2\\
&+(v_3w_1)^2+(v_3w_2)^2+(v_3w_3)^2\\
=&(v_1^2+v_2^2+v_3^2)(w_1^2+w_2^2+w_3^3)\\
=&|\vec{v}|^2|\vec{w}|^2.
\end{aligned}$$

Now we substitute $\vec{v}\cdot\vec{w}=|\vec{v}||\vec{w}|\cos\theta$, we get

$$|\vec{v}\times\vec{w}|=|\vec{v}||\vec{w}|\sqrt{1-\cos^2\theta}=|\vec{v}||\vec{w}|\sin\theta.$$

Where we know $\sin\theta\geq 0$ as θ is between 0 and π.

The value $|\vec{v}||\vec{w}|\sin\theta$ has a nice geometric meaning. It is the area spanned by the vectors $\vec{v}$ and $\vec{w}$.

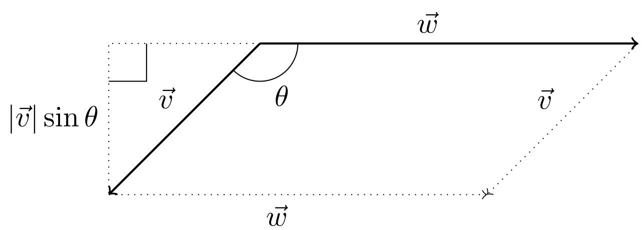


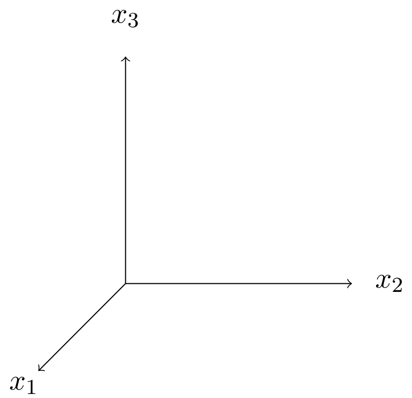
Figure 1.5: The area of a parallelogram formed by these vectors is the magnitude of the vector $\vec{v}\times|\vec{w}|$

Proposition 1.33

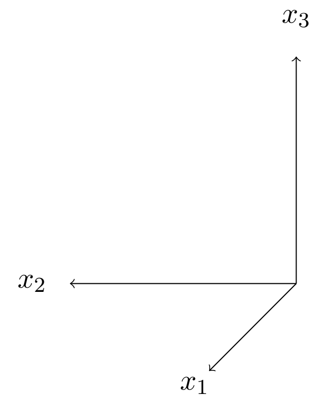
The direction of $\vec{v}\times\vec{w}$ is determined by the **right-hand rule** as follows:

> Using the right hand, align the index finger with the direction $\vec{v}$, and the middle finger with the direction of $\vec{w}$. Extend the thumb so that it is perpendicular to both the index finger and the middle finger. The thumb is pointing in the direction of $\vec{v}\times\vec{w}$.

This is a byproduct of the convention we use. In $\mathbb{R}^3$ we use what is known as a right-handed coordinate system - the vectors $\vec{i}$, $\vec{j}$, and $\vec{k}$ align with the first three fingers of the right hand respectively. If we used a left-handed coordinate system, the rule would be left-handed instead. We now have a few properties about the cross product from our computation, the first you will verify on your own:



(a) Right handed coordinate system



(b) Left handed coordinate system

Proposition 1.34

Let $\vec{v}, \vec{w}, \vec{u} \in \mathbb{R}^n$, then

- *(distributivity)* $\vec{v} \times (\vec{u} + \vec{w}) = \vec{v} \times \vec{u} + \vec{v} \times \vec{w}$ and $(\vec{v} + \vec{u}) \times \vec{w} = \vec{v} \times \vec{w} + \vec{v} \times \vec{u}$.
- *(anti-symmetry)* $\vec{v} \times \vec{w} = -\vec{w} \times \vec{v}$.
- $\vec{v} \times \vec{w}$ is orthogonal (perpendicular in 3D space) to $\vec{v}$ and $\vec{w}$.
- The magnitude $|\vec{v} \times \vec{w}|$ is given by $|\vec{v}||\vec{w}|\sin\theta$, with θ being the angle between $\vec{v}$ and $\vec{w}$, so
  - (i) The magnitude $|\vec{v} \times \vec{w}|$ also corresponds to the area of the parallelogram formed by $\vec{v}$ and $\vec{w}$.
  - (ii) If $\vec{v}$ and $\vec{w}$ are parallel or antiparallel, $\vec{v} \times \vec{w} = \vec{0}$.

Remark. *Using distributivity of the cross product, you only need to memorize the cross product of the basis vectors, and write* $\vec{v} \times \vec{w} = \sum_{i=1}^{3} \sum_{j=1}^{3} v_i w_i (\vec{e}_i \times \vec{e}_j)$.

1.4.1 Triple products

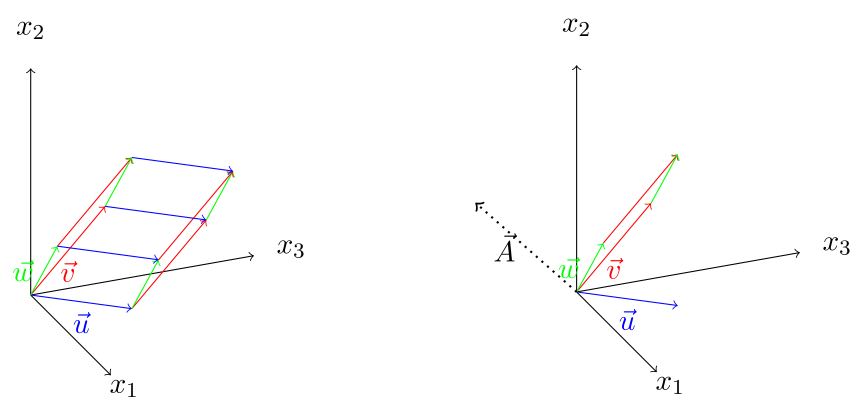
Example 1.35

Find the volume of the parallelepiped formed from $\vec{v} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \vec{w} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \vec{u} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$.

The parallelepiped is a generalization of a parallelogram to higher dimensions. Using combinations of $\vec{v}, \vec{w}, \vec{u}$, you can make the frame of a 3d solid. As in the figure on the left, edges of the same color correspond to the same vector (and thus parallel). The volume of this solid is still *base* $\times$ *height*, where the base is a 2D parallelogram formed by two vectors and the height is determined by third vector. We make an arbitrary decision and set the base to be $\vec{v}$ and $\vec{w}$. (setting any two vectors would give the same result in the end!) The area of this base is given by the cross product

$$\vec{A} = \vec{v} \times \vec{w} = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}.$$

Now we can determine the height of the paralellepiped from $\vec{u}$. We want to isolate the component of $\vec{u}$ that is orthogonal to the base. Equivalently, we want to find the component of $\vec{u}$ that is pointing in the direction of $\vec{A}$, a vector that is orthogonal to both $\vec{v}$ and $\vec{w}$!



(a) The parallelepiped, draw in an optical illusion fashion.

(b) We want to get the projection of $\vec{u}$ on $\vec{A}$.

The volume is thus

$$|\text{Proj}_{\vec{A}}(\vec{u})||\vec{v}| = \left|\frac{\vec{u} \cdot \vec{A}}{|\vec{A}|^2}\right| \times |\vec{A}| \times |\vec{A}| = |\vec{u} \cdot \vec{A}| = 2.$$

Proposition 1.36

The volume of the parallelpiped formed from $\vec{v}, \vec{w}, \vec{u}$ is

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

Remark. *This is also the expression of the (absolute value of) determinant of*

$$\begin{bmatrix} \vec{v} & \vec{w} & \vec{u} \end{bmatrix}$$

where $\vec{v}, \vec{w}, \vec{u}$ are written as column vectors. Using properties of the determinant (later chapters), you can show cycling the three vectors does not change the volume. (i.e. you can calculate using what order of the three vectors you want)

Remark. *You may notice that the expression $(\vec{v} \times \vec{w}) \cdot \vec{u}$ can take on negative volumes. In this case, the three vectors (taken in order) do not follow the right-hand rule. For instance, in the last example, $\vec{u}$ points in the 'opposite' direction as $\vec{A}$.*

Exercises

6. Find $\vec{v} \times \vec{w}$ for the following:
   (a) $\vec{v} = (4, -2, 0)$, $\vec{w} = (2, 1, -1)$
   (b) $\vec{v} = (3, 3, 3)$, $\vec{w} = (4, -3, 2)$
   (c) $\vec{v} = 2\vec{i} + 3\vec{j} + 4\vec{k}$, $\vec{w} = \vec{i} - 3\vec{j} + 4\vec{k}$
7. Find the areas for the following shapes:
   (a) The parallelogram with vertices $P(0, 0, 0), Q(1, 1, 0), R(1, 2, 1), S(0, 1, 1)$.
   (b) The triangle with vertices $A(1, 9, 3), B(-2, 3, 0), C(3, -5, 3)$.
8. Find the volume of the parallelepiped formed from the vectors $\vec{v} = (1, 1, 0), \vec{w}(0, 2, -2), \vec{u} = (1, 0, 3)$.
9. A triangular kite has vertices $P(0, 0, 10), Q(2, 1, 10), S(0, 3, 12)$ and is displaced by the wind at a velocity of $(20\vec{i} + 6\vec{j} + 4\vec{k})/s$
   (a) Find the area of the kite.
   (b) After 1/2 seconds, find the volume of the space swept by the kite. (leave the answers in $[units]^3$)

1.5 Applications - Geometry of lines and planes

Definition 1.37 (Relations between lines)

For two (infinitely extending) lines in $\mathbb{R}^n$ parametrized in s and t respectively as $l_1 = P + t\vec{v}$, $l_2 = Q + s\vec{w}$, we say the lines are

- **Parallel**, if the $\vec{v}$ and $\vec{w}$ are parallel or antiparallel.
- **Intersecting**, if l_1 and l_2 exactly one point on both l_1 and l_2.
- **Skew**, if l_1 and l_2 are not parallel/antiparallel or intersecting.

Remark. *Lines do not have direction, so there usually is no need to distinguish between parallel and antiparallel lines. One may extend the definition of antiparallel to lines from $\vec{v}$ and $\vec{w}$.*

Proposition 1.38

Determination of parallel lines are independent of parametrization. Concretely,

> Let $P_1 + t_1\vec{v}_1$ and $P_2 + t_2\vec{v}_2$ be two parametrizations of l_1, $Q_1 + s_1\vec{w}_1$, $Q_2 + s_2\vec{w}_2$ be two parametrizations of l_2. If $\vec{v}_1 = c_1\vec{w}_1$ for some $c_1 \in \mathbb{R}$, then $\vec{v}_2 = c_2\vec{w}_2$ for some (possibly different)$c_2 \in \mathbb{R}$.

The proof is not very enlightening. However, the result of this guarantees that our definition of parallel lines is precise.

Proof. The idea is to show that $\vec{v}_1$ is a scalar multiple of $\vec{v}_2$, and by the same logic $\vec{w}_1$ is a scalar multiple of $\vec{w}_2$. Since all vectors are non-zero in the parametrization, we will get the result of $\vec{v}_2$ a scalar multiple of $\vec{w}_2$.

To show $\vec{v}_1 = k\vec{v}_2$ for some k, we can pick two distinct points A, B on l_1. From the parametriation we can get $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}$ such that $P_1 + \alpha_1\vec{v}_1 = A = P_2 + \alpha_2\vec{v}_2$ and $P_1 + \beta_1\vec{v}_1 = B = P_2 + \beta_2\vec{v}_2$. Therefore we get the vector

$$\overrightarrow{AB} = (\beta_1 - \alpha_1)\vec{v}_1,$$
$$\overrightarrow{AB} = (\beta_2 - \alpha_2)\vec{v}_2.$$

As we picked distinct points A and B, we can conclude $\overrightarrow{AB} \neq \vec{0}$ and thus $\beta_1 - \alpha_1 \neq 0, \beta_2 - \alpha_2 \neq 0$, so that

$$\vec{v}_1 = \frac{\beta_2 - \alpha_2}{\beta_1 - \alpha_1}\vec{v}_2.$$

□

Example 1.39

Determine whether the lines parametrized by $l_1(t) = (1,2,1) + t(1,3,-2)$ and $l_2(t) = (3,1,0) + t(-2,-6,4)$ are parallel, intersecting, or skew. Confirm that l_1 and l_2 describe two different lines.

We notice that $-2 \times (1,3,-2) = (-2,-6,4)$, so these lines are parallel. To confirm that these two lines are not the same, we notice that $(1,2,1)$ is a point on l_1, but if we attempt to solve

$$(1,2,1) = l_2(t) = (3,1,0) + t(-2,-6,-4) \implies (-2,1,1) = t(-2,-6,-4)$$

which no t can solve! Specifically, the first coordinate forces $t = 1$ and the second coordinate forces $t = -6$.

Example 1.40

Determine whether the lines parametrized by $l_1(t) = (1,2,1) + t(1,3,-2)$ and $l_2(t) = (0,3,9) + t(0,2,3)$ are parallel, intersecting, or skew.

$(1,3,2)$ is not a multiple of $(0,2,3)$, so the lines are not parallel. We might be tempted to solve for $l_1(t) = l_2(t)$ to check for intersection, but this misses a lot of cases! We need to compare all the points of l_1 with all the points of l_2, so we need two independent variables to describe where we are on each of the lines. That is, we solve for s, t in $l_1(t) = l_2(s)$,

$$\begin{aligned} (1,2,1) + t(1,3,-2) &= (0,3,9) + s(0,2,3) \\ \implies (t+1, 3t+2, -2t+1) &= (0, 2s+3, 3s+9) \\ \implies t = -1 \text{ and } 3t+2 = 2s+3 &\text{ and } -2t+1 = 3s+9 \end{aligned}$$

$t = -1, s = -2$ solves this system of equations. We can plug in t and s in our original parametrization to find $(0,1,3)$ is indeed a point on both l_1 and l_2. We would have missed this if we set $l_1(t) = l_2(t)$!

Example 1.41

Determine whether the lines parametrized by $l_1(t) = (1,2,1) + t(1,3,-2)$ and $l_2(t) = (0,3,8) + t(0,2,3)$ are parallel, intersecting, or skew.

We repeat the same process as above to see that the lines are not parallel and solve for

$$\begin{aligned}
&(1,2,1) + t(1,3,-2) = (0,3,8) + s(0,2,3) \\
\implies &(t+1, 3t+2, -2t+1) = (0, 2s+3, 3s+8) \\
&\implies t = -1 \text{ and } 3t+2 = 2s+3 \text{ and } -2t+1 = 3s+8
\end{aligned}$$

This time, we do not have a solution - the first two equations forces $t = -1, s = -2$, and this does not solve the third. We therefore do not have a point of intersection, and the lines are skew.

Example 1.42

Determine if $P(5,6,9), Q(7,9,15), R(13,18,33)$ are **colinear** i.e. if they lie on the same line.

With the machinery we have built up, there are multiple ways to check if P, Q, R form a straight line. Here are a few ideas:

1. Check that $\overrightarrow{PQ}$ is parallel/antiparallel to $\overrightarrow{PR}$. Because the lines defined by $\overline{PQ}$ and $\overline{PR}$ are parallel and share a same point P, they are the same line.

2. Use the dot product to calculate the angle $\angle PQR = \pi$.

3. Use the cross product to calculate that $\overrightarrow{PQ} \times \overrightarrow{PR} = \vec{0}$. This means the triangle with vertices P, Q, R has no area and thus is a degenerate triangle.

Definition 1.43 (Characterization of Planes)

In euclidean geometry, planes can be characterized by any of the following ways:

- For **any three non-colinear points** P_1, P_2, P_3, there is a unique plane passing through P_1, P_2, P_3.
- For **any pair of intersecting lines**, there is a unique plane that contains both.
- For **a line** l **and a point** P, there is a unique plane that contains P and is perpendicular to l.
- For **a line** l **and a point** P **not on** l, there is a unique plane that contains both l and P.

Remark. *The third characterization is the hardest to visualize at first, but is also the easiest to describe with the analytical tools we have built towards. We can refer to Example 1.26. The plane sketched is the unique plane that contains* $(-3,2,6)$*, such that each vector in the plane is orthogonal to* $(-3,2,6)$*, so the line parametrized by* $l(t) = t(-3,2,6)$ *is perpendicular to the plane at the point of intersection* $(-3,2,6)$*.*

Example 1.44

In $\mathbb{R}^3$, determine the equation of the plane that contains $P_0(x_0, y_0, z_0)$ and is perpendicular to the line parametrized by $l(t) = Q_0 + t\vec{N}$, $\vec{N} = (a, b, c)$.

First we can exploit **translation invariance** of $\mathbb{R}^3$ and move the line to $\tilde{l} = P_0 + t\vec{N}$. Because l and $\tilde{l}$ are parallel, any line perpendicular to l will also be perpendicular to $\tilde{l}$.
Then by the characterization of a plane, any P on the plane satisfies the orthogonality relation $\overrightarrow{PP_0} \cdot \vec{N} = 0$. Expanding this, we get the equation

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0 \implies ax + by + cz = ax_0 + by_0 + cz_0.$$

Definition 1.45 (Normal form of a plane)

Let $a, b, c, d \in \mathbb{R}$, with at least one of $a, b, c \neq 0$. The equation of a plane in $\mathbb{R}^3$ written as

$$ax + by + cz = d$$

is called a **normal form** of the equation of a plane.

Remark. *As all planes can be characterized this method, all equations of planes can be put in normal form.*

Remark. *The normal form of a plane is not unique. Pick your favorite non-zero number α, the equation $\alpha ax + \alpha by + \alpha c_z = \alpha d$ describes the same plane.*

Theorem 1.46 (Normal vectors of planes)

When written in normal form, the plane is perpendicular to the vector $\vec{N} = (a, b, c)$. We call this vector $\vec{N}$ the **normal vector**.

We thus have a definition for two planes to be parallel.

Definition 1.47 (Parallel planes)

Two planes are **parallel** if the normal vectors are parallel.

A problem in the exercise will guide you through the proof of Theorem 1.46. The intuition behind the proof is the reverse direction of the equation $\overrightarrow{PP_0} \cdot \vec{N} = 0$ we derived from the last example.
We will now apply this theorem in a few examples.

Example 1.48

Find the equation of the plane that passes through the points $P(1, 0, 0), Q(0, 1, 0), R(0, 0, 1)$.

Method 1: One sees that (by coincidence) the sum of coordinates of each point are equal to 1, so immediately writes down $x + y + z = 1$. Despite being the fastest method, this is somewhat inconsistent.

Method 2: We set $ax + by + cz = d$ to be the equation, and plug in values for P, Q, R, giving the system of equations

$$\begin{aligned} a + 0b + 0c &= d \\ 0a + b + 0c &= d \\ 0a + 0b + c &= d \end{aligned}$$

This is an **underdetermined system**, meaning there are fewer equations than unknowns. The best we can say is that $a = b = c = d$. However, setting $a = b = c = d = k$ works for all $k \neq 0$, further confirming that the normal form is not unique. This method is reasonably fast when the system of equations are simple. When there are more non-zero coefficients, solving the system takes more time.

Method 3: We can compute the direction of the normal vector $\vec{N}$ of this plane. By the characterization of planes, $\vec{N}$ is orthogonal to $\overrightarrow{PQ}$ and $\overrightarrow{PR}$, two vectors in the plane. We can thus write $\vec{N}$ as a multiple of the cross product

$$\overrightarrow{PQ} \times \overrightarrow{PR} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \times \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

. Using this normal vector (or any multiple of it) and applying the theorem, we get $x + y + z = d$, and substituting P into the equation will give $x + y + z = 1$. This method is more general and consistent, as the number of operations in a cross product is constant.

> Example 1.49
>
> On the plane given by $ax + by + cz = d$, find the point on the plane that is closest to the origin.

Let r be the distance, we draw a sphere centered at the origin with radius r. This sphere is thus tangent to the surface at one point, and the vector corresponding to this point will be perpendicular to the plane. By the theorem, we can denote the point $P(\alpha a, \alpha b, \alpha c)$, with α a constant to be determined. Substituting P into the equation,

$$\alpha(a^2 + b^2 + c^2) = d \implies \alpha = \frac{d}{a^2 + b^2 + c^2}$$

The closest distance from the origin is

$$|(\alpha a, \alpha b, \alpha c)| = |\alpha|\sqrt{a^2 + b^2 + c^2} = \left|\frac{d}{\sqrt{a^2 + b^2 + c^2}}\right|$$

at the point

$$\left(\frac{da}{\sqrt(a^2 + b^2 + c^2)}, \frac{db}{\sqrt(a^2 + b^2 + c^2)}, \frac{dc}{\sqrt(a^2 + b^2 + c^2)}\right)$$

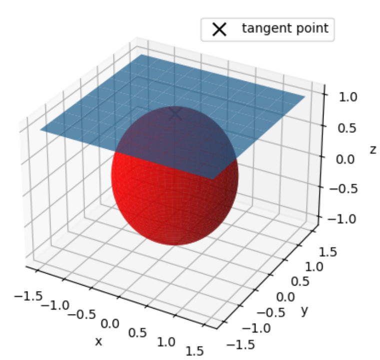


Figure 1.8: A toy example with the unit sphere of radius 1 and the plane described by $z = 1$.

1.5.1 Intersection of planes

Example 1.50

Find the intersection of the planes given by

$$6x + 2y - z = 2$$
$$\text{and } x - 2y + 3z = 5.$$

Method 1: The line lies on the first plane, so is perpendicular to its normal vector $(6, -2, -1)$. The line also lies on the second plane, so is perpendicular to its normal vector $(1, -2, 3)$. Using the cross product, the line should point in the direction of

$$\begin{bmatrix} 6 \\ 2 \\ -1 \end{bmatrix} \times \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ -19 \\ -14 \end{bmatrix}$$

Now we just need to find one point P on the line of intersection, so that the parametric equation (in t) is $P + t(4, -19, -14)$. We can impose an additional restriction that $x = 0$ (or $y = 0$ or $z = 0$) and solve the simultaneous equations

$$2y - z = 2$$
$$-2y + 3z = 5.$$

and get $(0, 11/4, 7/2)$ is a point on the intersection. The final parametric representation is $l(t) = (0, 11/4, 7/2) + t(4, -19, -14)$.

Method 2: We can solve the system of equations to find the line of intersection, and arrive at the set of symmetric equations. A similar method would be to solve for two points on the line of intersection, then get the parametric equation. This method will be introduced later

Exercises

10. Find parametric and symmetric equations for the line that
    (a) passes through $(0, 0, 0)$ and is parallel to $\vec{v} = 3\vec{i} + 4\vec{j} + 5\vec{k}$
    (b) passes through $(1, 1)$ and is perpendicular to the line described by $y = 3x + 3$
    (c) passes through $(0, 1, 2)$ and is perpendicular to plane with equation $3x + 4y = 2 + z$
    (d) passes through $(0, 1, 2)$ and is perpendicular to the yz-plane
    (e) passes through $(1, 3, 0)$ and is parallel to the line with symmetric equations $x = y - 1 = (z - 3)/2$

11. Determine if these pair of lines are parallel, intersecting, or skew:
    (a) (As described by parametric equations) $l_1(t) = (1, -1, 2) + t(2, 1, 1), l_2(t) = t(1, 0, -1)$
    (b) The line passing through $(3, 1, 3)$ and $(-2, -4, -5)$ and the line passing through $(1, 3, 5)$ and $(11, 13, 21)$.

12. Determine if $P(3,1,2), Q(-1,0,2), R(11,3,2)$ are colinear.

13. Find an equation for the plane

    (a) Containing $P(1,2,0)$ with normal vector $\vec{N} = 2\vec{i} - \vec{j} + 3\vec{k}$

    (b) containing $P(2,4,5)$ with normal vector $\vec{N} = \vec{i} + 3\vec{k}$

    (c) containing $P(1,4,3)$ and perpendicular to the line given by the parametrization $r(t) = (1+t, 2+4t, t)$

    (d) containing the origin and parallel to the plane described by $3x + 4y - 6z = -1$

    (e) containing the points $P(0,0,0), Q(1,-2,8), R(-2,-1,3)$

14. Find the intersection of the planes with equations $2x + 3y - z = 1$ and $x - y - z = 0$

15. Find the angle between the normal vectors of the following planes:

    (a) the planes with equations $x + 2y - z = 2$ and $2x - y + 3z = 1$

    (b) the plane with equation $2x + 3y - 6z = 0$ and the plane containing the points $P(1,3,-2)$, $Q(5,1,3)$, $R(1,0,1)$

1.6 End of Chapter Exercises

16. Let $PQRS$ be a parallelogram. Show that $\overline{PR}$ and $\overline{QS}$ bisect each other i.e. intersect at their midpoints.

17. Complete the proof to Proposition **??** - Show that the dot product is symmetric and linear.

18. Prove the Cauchy-Schwarz inequality. Here is a hint: Write $\vec{w} = \text{Proj}_{\vec{v}}\vec{w} + (\vec{w} - \text{Proj}_{\vec{v}}\vec{w})$, and take the squared-magnitude on each side using the dot product.

19. Prove or give a counterexample: The cross product is associative.

20. Prove the following identity: $(\vec{a} \times \vec{b}) \times \vec{c} = -(\vec{b} \cdot \vec{c})\vec{a} + (\vec{a} \cdot \vec{c})\vec{b}$.

21. Show that the cross product satisfies the Jacobi identity: For any $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^3$, $(\vec{a} \times \vec{b}) \times \vec{c} + (\vec{b} \times \vec{c}) \times \vec{a} + (\vec{c} \times \vec{a}) \times \vec{b} = \vec{0}$.

22. On the plane given by $ax + by + cz = d$, show that the closest distance from $P(x_1, y_1, z_1)$ to the plane is given by $|ax_1 + by_1 + cz_1 - d|/\sqrt{a^2 + b^2 + c^2}$. (Hint: Refer to Example 1.44)

23. Define two lines in $\mathbb{R}^3$ by the parametric equations

$$l_1(t) = \vec{r_1} + t\vec{v_1}$$
$$l_2(t) = \vec{r_2} + t\vec{v_2}$$

    We will derive a formula for the shortest distance between the two lines.

    (a) Consider the special case where the two lines are parallel. Construct two points A and B on l_1 and l_2 respectively. Find B' Such that $\overrightarrow{AB'} = \text{Proj}_{\vec{v_1}}(\overrightarrow{AB})$. Argue that B' is on the line l_1, and that the line $\overline{BB'}$ is perpendicular to l_1 and l_2. Conclude that the shortest distance between l_1 and l_2 is given by

$$\sqrt{|\vec{r_2} - \vec{r_1}|^2 - \left(\frac{(\vec{r_2} - \vec{r_1}) \cdot \vec{v_1}}{|\vec{v_1}|}\right)^2}$$

(b) Now we can assume the lines are not parallel. Find a vector $\vec{N}$ that is perpendicular to both lines.

(c) (In your mind,) look in the direction of $\vec{N}$. $\vec{N}$ will now look like a dot. The lines will seem to overlap at points A' and B' respectively. Argue that $\overline{A'B'}$ is the shortest line from l_1 to l_2, and it is given by

$$\overrightarrow{A'B'} = \mathrm{Proj}_{\vec{N}}(\overrightarrow{AB})$$

, for any A on l_1 and B on l_2.

(d) Conclude that the shortest distance in this case is given by

$$\left| \frac{(\vec{r_2} - \vec{r_1}) \cdot (\vec{v}_1 \times \vec{v}_2)}{|\vec{v}_1 \times \vec{v}_2|} \right| .$$

(e) Suppose l_1 and l_2 intersect at P. Verify that the formula evaluates to 0, thus it gives a criterion to check if two lines intersect.

Chapter 2

Linear Algebra Basics

The foundational abstraction of linear algebra is the vector space. A *vector space* is essentially a collection of objects that it makes sense to take linear combinations of. Two operations must be defined: addition and scalar multiplication. A well defined vector space meets all of the vector space axioms, which will be listed shortly. Many consequences can be drawn from these axioms, and we can build up linear algebra to solve any linear problem.

Remark. *One has to specify the field of scalars (field just means number system) related to the vector space. In most cases here, we will be talking about real vector spaces (the field of scalars is* $\mathbb{R}$*). However, vector spaces can be defined with many other fields of scalars. Examples will follow the definition.*

Definition 2.1 (Informal definition of a field)

A field $\mathbb{K}$ is a set that supports addition, subtraction, multiplication, division with the similar properties that you expect from the real numbers.

Where is this going?

In the previous chapter, we have seen vector manipulation in $\mathbb{R}^n$. This chapter tries to generalize some observations we have about $\mathbb{R}^n$ to other mathematical objects. For now, we call these sets that have '$\mathbb{R}^n$'-vector-like properties **vector spaces**. This notion is a bit abstract, and we usually default to $\mathbb{R}^n$ to gain intuition about vector spaces and motivate new definitions and techniques. It just so happens that *every finite dimensional real vector space is 'almost the same' as* $\mathbb{R}^n$, so many conclusions we get from $\mathbb{R}^n$ naturally extend to other vector spaces.

In the previous chapter, proposition 1.7 lists 8 characteristics of $\mathbb{R}^n$. Let us extract all these 8 statements, and define a vector space to be a set that satisfies these properties.

Definition 2.2 (Vector Space)

Let $\mathbb{K}$ be a field, and V be a set closed under operations addition $+ : V \times V \to V$ and multiplication $\cdot : \mathbb{K} \times V \to V$. We call V a **vector space**, or a $\mathbb{K}$**-vector space** to specify the field if the following axioms hold.

1. *(Associativity)* $x + (y + z) = (x + y) + z \ \forall x, y, z \in V$.
2. *(Commutativity)* $x + y = y + x \ \forall x, y \in V$.
3. *(Identity)* There exists some vector 0_V s.t. $x + 0_V = x \ \forall x \in V$.
4. *(Inverse)* $\forall x \in V$, $\exists y \in V$ s.t. $x + y = 0_v$.
5. *(Scalar multiplication)* $a \cdot (b \cdot x) = (a \cdot b) \cdot x \ \forall a, b \in \mathbb{K}, x \in V$.
6. *(Scalar Identity)* $1 \cdot x = x \ \forall x \in V$.
7. *(Distributivity 1)* $a \cdot (x + y) = a \cdot x + a \cdot y \ \forall a \in \mathbb{K}, x, y \in V$.
8. *(Distributivity 2)* $(a + b) \cdot x = a \cdot x + b \cdot x \ \forall a, b \in \mathbb{K}, x \in V$.

Example 2.3

The following are examples of vector spaces.

- $\mathbb{R}^n$, with addition and multiplication as we have defined so far.
- $\{0_V\}$, the set containing just the zero vector, is a vector space over any field.
- The set of polynomials with degree 3 or less. Addition and multiplication are defined as $(f + g)(x) = f(x) + g(x)$ and $cf(x) = c \times f(x)$.
- $\mathbb{R}$ is a $\mathbb{Q}$-vector space, where $\mathbb{Q}$ is the set of all rationals (fractions).

The following are non-examples of vector spaces.

- ϕ, the empty set is not a vector space over any field as it does not have a zero vector.
- The set of polynomials with degree 3 or more (using the addition and multiplication rules defined above).
- $\mathbb{Q}$ is not an $\mathbb{R}$-vector space. This is because $\pi(1) = \pi$ which is not in $\mathbb{Q}$.

More frequently, we make new vector spaces from taking subsets of existing vector spaces. For instance, there might be a subset of $\mathbb{R}^n$ that also satisfies all the axioms of being a vector space.

Definition 2.4 (Subspace)

Let V be a $\mathbb{K}$ vector space. We call $W \subseteq V$ a **subspace** of V if W forms a vector space using the inherited operations $+ : W \times W \to W$ and $\cdot : \mathbb{K} \times W \to W$.

Remark. *Because the inherited operations will automatically satisfy the axioms of a vector space,* *it suffices to show that (1) W is nonempty and (2) W is closed under vector addition and scalar multiplication to confirm that W is a subspace.*

The condition W is nonempty is required because the identity axiom requires a zero vector, which one obtains by multiplying 0 to an arbitrary vector in the subset.

Example 2.5

1. In $\mathbb{R}$-vector space $\mathbb{R}^3$, the set of all vectors in the form of $k\vec{i}$ forms a subspace.
2. Take $\mathbb{R}$ as a $\mathbb{Q}$-vector space. $\mathbb{Q} \subset \mathbb{R}$ is a subspace.
3. Take $\mathbb{R}$ as a $\mathbb{R}$-vector space. $\mathbb{Q} \subset \mathbb{R}$ is **not** a subspace because it is not closed under multiplication.

Proposition 2.6

Let $W, U \subseteq V$ be subspaces. The intersection $W \cap U$ is a subspace of V.

Proof. Both W and U contain 0_V, so the intersection is non-empty. We also have for $v_1, v_2 \in W \cap U$,

$$v_1 + v_2 \in U, \quad v_1 + v_2 \in W,$$

so addition is closed in $W \cap U$. Similarly, scalar multiplication closed under W and U, so is closed in the intersection. □

Exercises

1. Determine if the following sets are real vector spaces:
   (a) The set of $(x, y, z) \in \mathbb{R}^3$ that satisfy $2x + 3y - z = 0$.
   (b) The set of $(x, y, z) \in \mathbb{R}^3$ that satisfy $2x + 3y - z = 4$.
   (c) The set of $(x, y, z) \in \mathbb{R}^3$ that satisfy $x^2 + y^2 - z^2 = 0$.
   (d) The set of real polynomials $f(x) = ax^3 + bx^2 + cx + d$ for all real constants a, b, c, d, such that $f'(x) = 0$.
   (e) The set of continuous functions $f : \mathbb{R} \to \mathbb{R}$ such that $f(0) = 1$.
2. Give an example of a vector space that only has one subspace.
3. Give a subspace of $\mathbb{R}^4$ (that is not $\mathbb{R}^4$) that contains the vector $(-1, 3, 0, 7)$.

2.1 Span, Linear Independence

2.1.1 Span

Throughout this section, let $\mathbb{K}$ be a field. We constrain ourselves to work in V, a $\mathbb{K}$-vector space.

Definition 2.7 (Span)

Let k be some positive integer. Let $\{v_1, v_2, ..., v_k\} \subseteq V$. The **span** of $\{v_1, v_2, ..., v_k\}$ is denoted as

$$\text{span}(v_1, v_2, ..., v_k)$$

and is the *smallest* subspace of V containing $\{v_1, v_2, ..., v_k\}$. If $\text{span}(v_1, v_2, ..., v_k) = V$, we say that $\{v_1, v_2, ..., v_k\}$ is a spanning set of V, or $\{v_1, v_2, ..., v_k\}$ spans V.

The *smallest* here means that if another subspace W contains $\{v_1, ..., v_k\}$, W cannot be a subset of $\text{span}(v_1, ..., v_k)$. How do we know such a subspace exists? We can take the intersection of all the subspaces containing $\{v_1, ..., v_k\}$

$$\text{span}(v_1, ..., v_k) = \bigcap_{W \text{ subspace containing } \{v_1,...,v_k\}} W$$

which is a subspace containing $\{v_1, ..., v_k\}$ and is a subset of all other subspaces containing $\{v_1, ..., v_k\}$. We know $V \subseteq V$ is a subspace containing $\{v_1, ..., v_k\}$, so the intersection is between at least one set and is thus well-defined.

Proposition 2.8

The span of $\{v_1, v_2,\ ..., v_k\} \subseteq V$ is all linear combinations of the vectors in the set:

$$\text{span}\{v_1, v_2,\ ..., v_k\} = \{c_1v_1 + c_2v_2 + ... + c_kv_k | c_1, c_2, ...c_k \in \mathbb{K}\}$$

Proof. We first show $\text{span}\{v_1, v_2, ..., v_k\} \supseteq \{c_1v_1 + c_2v_2 + ... + c_kv_k | c_1, c_2, ...c_k \in \mathbb{K}\}$. That is, for every W subset containing $v_1, ..., v_k$, W must also contain $c_1v_1 + ... + c_kv_k$.
We now show $\text{span}\{v_1, v_2, ..., v_k\} \subseteq \{c_1v_1 + c_2v_2 + ... + c_kv_k | c_1, c_2, ...c_k \in \mathbb{K}\}$. The right side is a set that nonempty, is closed under addition and scalar multiplication, and contains $v_1 = 1v_1 + 0v_2 + ... + 0v_k$,...,$v_k = 0v_1 + ... + 0v_{k-1} + 1v_k$. Therefore, it is one of the W subspaces whose intersection is used to construct the span. □

It can be difficult to how to think about what the span of a set of vectors looks like, though it is also important to develop an intuition for it as more complex techniques are developed. It is also important to consider what vectors span a given subspace.

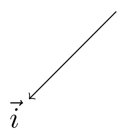
Example 2.9

Sketch, in $\mathbb{R}^3$, the following spans: $\text{span}(\{\vec{i}\})$, $\text{span}(\{\vec{i}, \vec{j}\})$, $\text{span}(\{\vec{i}, \vec{j}, \vec{k}\})$.

All linear combinations of $\vec{i}$ are vectors in the form $(a, 0, 0)$. $\vec{i}$ spans the $x - axis$. Similarly, linear combinations of $\vec{i}$ and $\vec{j}$ are all the vectors in the form $(a, b, 0)$, making up the xy-plane. Everything in $\mathbb{R}^3$ are linear combinations of $\vec{i}, \vec{j}, \vec{k}$, so $\text{span}(\{\vec{i}, \vec{j}, \vec{k}\}) = \mathbb{R}^3$.

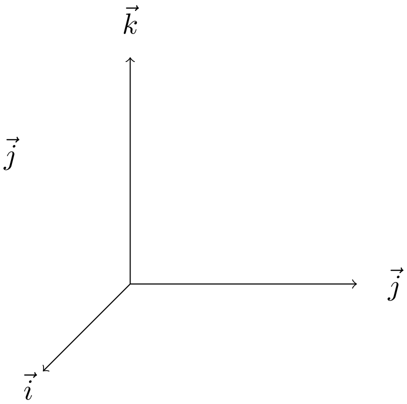
2.1.2 Linear Independence

Thinking geometrically in the previous example, the span of one vector is one-dimensional (a line), the span of two vectors is two-dimensional (a plane), and the span of three vectors is



(a) Span of $\vec{i}$

(b) Span of $\vec{i}$ and $\vec{j}$



(c) Span of $\vec{i}$, $\vec{j}$, and $\vec{k}$

three-dimensional (the whole 3D space). Specifically, we can give a correspondence between the linear combination of k vectors $\{c_1v_1, ..., c_kv_k\}$and a point in $\mathbb{R}^k$ as

$$c_1v_1 + c_2v_2 + ... + c_kv_k \sim \begin{bmatrix} c_1 \\ c_2 \\ ... \\ c_k \end{bmatrix}$$

Assuming this correspondence works, the geometry of the span should resemble $\mathbb{R}^k$. Unfortunately, this is not always true, for instance, consider the following counterexample:

Example 2.10

Sketch, in $\mathbb{R}^3$, $\text{span}(\vec{i}, \vec{k}, \vec{i} + 2\vec{k})$.

Let us consider an arbitrary linear combination of these vectors

$$\vec{v} = a\vec{i} + b\vec{k} + c(\vec{i} + 2\vec{k}) = \begin{bmatrix} a + c \\ 0 \\ b + 2c \end{bmatrix}.$$

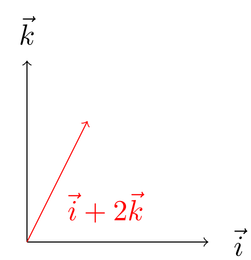


Figure 2.2: The third vector lies on the span of the first two

The second entry of $\vec{v}$ must be 0. The other two entries can be any two values (just set a to be the first value, b to be second, $c = 0$).This means the span of these three vectors is the xz-plane.

What went wrong here is that the third vector $\vec{i} + 2\vec{k}$ is already a linear combination of the first two, and so this vector is 'redundant' in the set. More precisely, $(1, 2, 0)$ and $(0, 0, 1)$ give the same linear combination of the three vectors, so the correspondence fails. We thus want to answer the following question:

> Given a set of k vectors $\{v_1, ..., v_k\}$, when do these vectors span the full k-dimensions? If they do not span the full k dimensions, how many dimensions do they span?

We already have one candidate criterion for spanning the full dimensions.

Definition 2.11 (Linear Independence)

Let $\{v_1, ..., v_k\} \subseteq V$. We say that $v_1, ..., v_k$ are **linearly independent** if every linear combination of $v_1, ..., v_k$ is unique. Precisely, if there are $c_1, ..., c_k, d_1, ..., d_k \in \mathbb{K}$ such that

$$c_1 v_1 + c_2 v_2 + ... + c_k v_k = d_1 v_1 + d_2 v_2 + ... + d_k v_k,$$

then

$$c_1 = d_1, c_2 = d_2, ..., c_k = d_k.$$

If $\{v_1, ..., v_k\}$ is not linearly independent, we say that it is **linearly dependent**.

This might be very tricky to verify, so we would like some simplier definitions for linear independence.

Proposition 2.12

The following are equivalent definitions for linear independence:

- (*) There is only one way to make 0_V. i.e. If $c_1 v_1 + ... + c_k v_k = 0_V$, then $c_1 = c_2 = ... = c_k = 0$.
- (**) If $k \geq 2$, there is no way to express one vector as a linear combination as the others. i.e. For all $1 \leq j \leq k$, $v_j \notin \text{span}(v_1, ..., v_{j-1}, v_{j+1}, ..., v_k)$.

Proof. We first show the original definition of linear independence implies (*). We already have one way to create the zero vector as a linear combination of the set, namely

$$0v_1 + 0v_2 + ... + 0v_k = 0_V.$$

By definition of linear independence, this is the only way to create the zero vector.
We now show (*) implies the original definition of linear independence. Let (*) hold. Then if

$$c_1 v_1 + ... + c_k v_k = d_1 v_1 + ... + d_k v_k,$$

then rearranging the terms we will get

$$(c_1 - d_1)v_1 + ... + (c_k - d_k)v_k = 0_V.$$

Since there is only one way to make 0_V, it must be that $c_1 - d_1 = c_2 - d_2 = ... = c_k - d_k = 0$, or $c_1 = d_1, ...c_k = d_k$. This matches our original definition of linear independence.
Now we set $k \geq 2$ and show the equivelence of (*) and (**). We now show (*) implies (**). Let (*) hold, and $v \in \text{span}(v_1, ..., v_{j-1}, v_{j+1}, ..., v_k)$. Then $v - v_j$ is a linear combination of $v_1, ..., v_k$ that is not $0v_1 + ... + 0v_k$ as the coefficient before v_j is -1. So that

$$v - v_j \neq 0_V \implies v \neq v_j.$$

What we have shown is that everything in $\text{span}(v_1, ..., v_{j-1}, v_{j+1}, ..., v_k)$ is not v_j. Which means $v_j \notin \text{span}(v_1, ..., v_{j-1}, v_{j+1}, ..., v_k)$.
One end of chapter exercise will guide you through the direction (**) implies (*). □

We now have an informal statement: *If the set is not linearly independent, then the dimension of the span must be less than k.* This is because from (**), at least one of the vectors is redundant, so that we can remove that vector and still produce the same span with $k-1$ vectors.

2.1.3 Basis

Finally, we have a special notion when the set $\{v_1, \ldots, v_k\}$ span V and are linearly independent. This means that every $v \in V$ is *a unique combination* of $\{v_1, \ldots, v_k\}$.

Definition 2.13 (Basis)

Let $\{v_1, ..., v_k\} \subseteq V$. We say that $v_1, ..., v_k$ form a **basis** for V if they span V and are linearly independent.

Example 2.14

In $\mathbb{R}^n$, the standard basis vectors $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$ are a basis. This is why we call them the standard basis.

Definition 2.15 (Dimension)

If a vector space V has a finite basis, we call V finite dimensional. Else we call V infinite dimensional.

Remark. *It is true that every basis of a vector space has the same number of elements, so we can assign a number (or infinity) to the dimension of a vector space. However, we will have to prove that later when we develop more technology.*

Theorem 2.16 (Every Vector Space has a Basis)

Let V be a $\mathbb{K}$ vector space. Then V has a basis.

Understanding this proof is optional. Our main hurdle here is that vector spaces are not necessarily 'small' to work with. For instance, $\mathbb{R}^n$ is finite dimensional, so is in some sense small. But now consider the following real vector space:

$$\tilde{V} = \{(a_1, a_2, a_3, \ldots) \mid a_j \in \mathbb{R}\}$$

Each element in this set is an infinite sequence of real numbers, and addition/scaling is entry-wise. You may argue that the infinite set $S = \{(1,0,0,\ldots), (0,1,0,\ldots), (0,0,1,\ldots), \ldots\}$ could be a basis. However, they do not span $\tilde{V}$. To show that, I claim that the subset

$$U = \{(a_1, a_2, a_3, \ldots) \in \tilde{V} \mid \text{only finitely many } a_j\text{'s are non zero}\}$$

is a subspace of $\tilde{V}$. It is a strict subset of $\tilde{V}$, contains the zero vector $(0,0,0,\ldots)$ and is closed under addition and $\mathbb{R}$-scaling. It also contains S as each element in S has 1 (thus finitely many) non-zero element(s). Thus $\operatorname{span}(S) \subseteq U \subset \tilde{V}$.

To circumvent this, this proof requires the use of the axiom of choice, which we will give a statement below.

Definition 2.17 (Axiom of Choice - optional)

The **Axiom of Choice** is one of the axioms in Zermelo-Frenkel (ZFC) Set theory. Its statement is as follows:

> Let I be a set, $(S_i)_{i\in I}$ be an indexed family of non-empty sets. Then there exists an indexed set $(s_i)_{i\in I}$ such that $s_i \in S_i$ for each $i \in I$.

Colloquially, it means there is a "choice function" that allows you to pick an element from each of the non-empty S_i's.

Remark. *What we lose from the Axiom of Choice is that the choice function is non-constructive. This means that we know that something exists without knowing exactly how to construct it. The Axiom of Choice also leads to weird consequences, for example the Banach-Tarski paradox states that you can rearrange the points of a solid 3D ball to make two solid 3D balls, each having the same size as the original. (!) While some mathematicians do not accept the Axiom of Choice and adopt ZF (ZFC without Choice), it generally makes our lives easier.*

From the Axiom of Choice, one can derive Zorn's Lemma.

Theorem 2.18 (Zorn's Lemma - optional)

Assume the Axiom of Choice. Let S be a partially ordered set, that is, there is a comparison function $\leq$ for some pairs of elements. Then if for every ascending chain in S

$$s_1, s_2, \ldots \in S \text{ such that } s_1 \leq s_2 \leq s_3 \leq \ldots,$$

There is some $t \in S$ such that $s_j \leq t$, then S has a maximal element (nothing is bigger than S).

Remark. *The example of Zorn's Lemma is as follows: Let S contain subsets of V. We have a partial ordering on S using inclusion.*

$$A \leq B \text{ if } A \subseteq B.$$

If for every chain $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ we can find $B \in S$ that contains all the A_j's, then S has a maximal element.

Proof of Every Vector Space has a Basis. Let $S = \{L \subseteq V \,|\text{L is linearly independent}\}$. We give S the partial ordering with respect to set inclusion. We first show S has a maximal element using Zorn's Lemma.

For each ascending chain $L_1 \subseteq L_2 \subseteq \ldots$, we claim $U = \bigcup_i L_i$ is a linearly independent set. We verify it directly: Let $v_1, \ldots, v_k \in U, c_1, \ldots, c_k \in \mathbb{K}$, and $\sum_i c_i v_i = 0_V$. We know that for each $1 \leq i \leq k$ there L_{a_i} such that $v_i \in L_{a_i}$. Take $a = \max_i a_i$. Then $v_i \in L_a$. Because L_a is a linearly independent set, it must be that $c_1 = \ldots = c_k = 0$.

Now we apply Zorn's Lemma to get that there is a maximal element $M \in S$. We claim Span(M). If not, we can find $v \in V$ such that $v \notin \text{Span}(M)$, but then this means $M \cup \{v\}$ is a linearly independent set. This contradicts that M is a maximal element. $\square$

Exercises

4. *(**) $\implies$ (*) in linear independence* Let $c_1 v_1 + \ldots + c_k v_k = 0_V$. If $c_1 \neq 0$, argue that $v_1 \in \text{span}(v_2, ..., v_k)$, thus this case cannot happen and $c_1 = 0$. Then show that $c_j = 0$ for all j, and thus the v_j's are linearly independent.

5. Write down a basis for the vector space of all polynomials of degree k or less.

2.2 Systems of Linear Equations

As we alluded to earlier, many types of vector spaces are in some way very similar to $\mathbb{R}^n$. For instance, the span of a set of k linearly independent real-vectors has a natural correspondence to $\mathbb{R}^k$. With the blind faith that everything here can generalize nicely back to abstract vector spaces, we limit ourselves again back to talking about $\mathbb{R}^n$.
Here, we introduce a new notation to write linear combiantions, as $c_1 v_1 + ... + c_k v_k$ is very cumbersome.

Definition 2.19 (Matrix)

Let m and n be positive integers. An $m \times n$ **matrix** is a rectangular array of m rows and n columns in the form

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

where each $c_{i,j}$ is an **entry** of the matrix. We denote the set of all $m \times n$ matrices with real entries $M_{m\times n}(\mathbb{R})$. In general, we have $M_{m\times n}(\mathbb{K})$ for entries in the field $\mathbb{K}$.

Notation. *Suggestively, we can write an $m \times n$ matrix as*

$$\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & ... & \vec{v}_n \end{bmatrix}$$

where each $\vec{v}_j \in \mathbb{R}^m$ is written as a column vector

$$\begin{bmatrix} c_{j,1} \\ c_{j,2} \\ \vdots \\ c_{j,m} \end{bmatrix}.$$

If we want to think of the matrix by its entries, we can also write the matrix as

$$\{c_{i,j}\}$$

Definition 2.20 (Matrix-vector Product)

Let

$$A = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & ... & \vec{v}_n \end{bmatrix}$$

be an $m \times n$ matrix, and

$$\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \in \mathbb{R}^n.$$

The product of A and $\vec{b}$ is evaluated as

$$A\vec{b} = b_1\vec{v}_1 + b_2\vec{v}_2 + ... + b_n\vec{v}_n.$$

Example 2.21

Let $M = \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \dots & \vec{e}_n \end{bmatrix}$ be an $n \times n$ matrix. Compute $M\vec{x}$ for any $\vec{x} \in \mathbb{R}^n$.

The computation is not too difficult. We can follow the definition to get

$$M\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + \dots + x_n\vec{e}_n = \vec{x}.$$

Basically, multiplying a vector with the matrix with columns $\vec{e}_1, \dots, \vec{e}_n$ does not change the vector. This should not be surprising, since the corresponding system of equations for $M\vec{x} = \vec{b}$ is

$$\begin{aligned} x_1 &= b_1 \\ x_2 &= b_2 \\ &\vdots \\ x_n &= b_n \end{aligned}$$

Definition 2.22 (Identity Matrix)

We denote the $n \times n$ **identity matrix** by

$$I_n \stackrel{\text{def}}{=} \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \dots & \vec{e}_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

Proposition 2.23

The matrix-vector product satisfies the following properties. Let $c \in \mathbb{R}, \vec{x}_1, \vec{x}_2 \in \mathbb{R}^n, A \in M_{m\times n}(\mathbb{R})$, then

- $A(\vec{x}_1 + \vec{x}_2) = A\vec{x}_1 + A\vec{x}_2$,
- $A(c\vec{x}_1) = c(A\vec{x}_1)$.

This means that the matrix-vector product preserves linear combinations. $A(c_1\vec{x}_1 + ... + c_k\vec{v}_k) = c_1 A\vec{x}_1 + ... + c_k A\vec{x}_k$.

Example 2.24

Determine if the vectors

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \vec{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

are linearly independent.

Recalling the definition for linear independence, we solve for

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

This is a compact way to express the system

$$\begin{aligned} c_1 + c_2 + 0 &= 0 \\ 0 + c_2 + c_3 &= 0 \\ c_1 + 0 + c_3 &= 0 \end{aligned}$$

There is a very slick way to find the solution (which involves adding all three equations together), but let us solve this methodically. This method is known as elimination, which involves isolating variables, and thus reducing the complexity of the system one by one.

First, we look at the first equation and isolate $c_1 = -c_2$. Now we can look replace all instances of c_1 in the second and third equation with $-c_2$, giving us

$$\begin{aligned} c_1 + c_2 + 0 &= 0 \\ 0 + c_2 + c_3 &= 0 \\ 0 - c_2 + c_3 &= 0 \end{aligned}$$

If we look at just the second and third equations, these only have two variables, so we have effectively decreased the complexity of the system by 1. Whatever we get from the second and third equations, we can substitute back into the first to get c_1. Now, we repeat for the second equation, $c_2 = -c_3$, and replacing all instances of c_2 we have

$$\begin{aligned} c_1 + 0 - c_3 &= 0 \\ 0 + c_2 + c_3 &= 0 \\ 0 + 0 + 2c_3 &= 0 \end{aligned}$$

The third equation now is just an equation in one variable, so we can go ahead and solve

$$\begin{aligned} c_1 + 0 - c_3 &= 0 \\ 0 + c_2 + c_3 &= 0 \\ 0 + 0 + c_3 &= 0 \end{aligned}$$

and replace all instances of c_3 with 0 to get

$$\begin{aligned} c_1 + 0 + 0 &= 0 \\ 0 + c_2 + 0 &= 0 \\ 0 + 0 + c_3 &= 0 \end{aligned}$$

or $c_1 = c_2 = c_3 = 0$. This is indeed a solution to the system, so we can now conclude that these vectors are indeed linearly independent. If we condense the systems back to matrices (concatenating the 3×3 matrix and the vector on the right) we get

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 2 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

Example 2.25

Determine the solutions to the system of equations

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

The first equation here does not let us isolate x_1, but we can simply exchange the first two equations to get

$$\left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 2 & -1 & 1 \\ 1 & 1 & 2 & 1 \end{array}\right]$$

We want to remove any dependence of x_1 for the other two equations, so we can subtract the first equation from the third to get

$$\left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 2 & -1 & 1 \\ 0 & 2 & 1 & 1 \end{array}\right]$$

We divide the second equation by 2, add this equation to the first, and subtract from the third.

$$\left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 1 & \frac{-1}{2} & \frac{1}{2} \\ 0 & 2 & 1 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & \frac{-1}{2} & \frac{1}{2} \\ 0 & 2 & 1 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & \frac{-1}{2} & \frac{1}{2} \\ 0 & 0 & 2 & 0 \end{array}\right]$$

Finally, we divide the third equation by 2, and remove all other entries of the third column.

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{array}\right]$$

So the solution is $x_1 = 1/2, x_2 = 1/2, x_3 = 0$.

2.2.1 Elementary Row Operations and Row Echelon Form

By how we manipulated the equations in the previous examples, we motivate these operations on matrices.

Definition 2.26 (Elementary Row Operations and row equivalence)

The following operations on a matrix $A \in M_{m\times n}$ are known as **elementary row operations**:

1. *(Row Swap)* Exchange any two rows.
2. *(Scaling)* Multiply a row by a **non-zero** constant.
3. *(Sum)* Add a multiple of a row to another row.

Let $B \in M_{m\times n}$. We say that A is **row equivalent** to B if A can be transformed into B By applying a sequence elementary row operations. We denote this equivalence by $A \sim B$.

Remark. *We can group the big set of $M_{m\times n}(\mathbb{R}^n)$ into classes, where each class contains matrices that are row equivalent to each other.*

Proposition 2.27

Elementary row operations are invertible. That is, if $A \sim A'$ after applying an elementary row operation, $A' \sim A$ through applying a (possibly different) row operation.

Proposition 2.28

An $m \times n$ matrix can be viewed as a system of m equations in $n-1$ variables using the representation in the previous two examples. In this representation, row equivalent matrices have the same solution sets.

The proof is not very enlightening. It boils down to checking that each of the three elementary row operations do not change the solution set of the corresponding linear system, so a sequence of them will not change the solution sets. The key takeaway from this is that we can reduce the complexity of the matrix through elementary row operations. Let us define the 'simple' forms of a matrix you can get through row operations.

Definition 2.29 (Row Echelon Form and Pivots)

A matrix is in **row echelon form** if

- Rows with all zero entries are on the bottom.
- For each row having non-zero entries, the first non-zero entry is on the right of the first non-zero entry of the row above.

In row echelon form, the first non-zero entry of a row is called a **pivot**. Colloquially, the non-zero entries form a "staircase" like shape.

Example 2.30

The following matrix is in row echelon form:

$$\begin{bmatrix} 1 & * & * & * & * \\ 0 & 0 & 2 & * & * \\ 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The pivots of this matrix, from the first row, are 1, 2 and 1.

In most cases, reducing a matrix into the row echelon form is "good enough" to find solutions. Start from the bottom row, and substitute variables go up. However, one problem with working with row echelon form is that a matrix is row equivalent to many matrices in row echelon form, so we want some 'super' row echelon form that is unique to each matrix.

Definition 2.31 (Reduced Row Echelon Form)

A matrix is in **Reduced Row Echelon Form**(rref) if

- It is in row echelon form.
- All pivots are equal to 1.
- Every pivot is the only non-zero entry in its column.

Example 2.32

The following matrix is in reduced row echelon form:

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

By how we defined the reduced row echelon form, we cannot find two distinct matrices in reduced row echelon form that are also row equivalent. This guarantees that a matrix is row equivalent to at most one reduced row echelon form matrix.

Theorem 2.33 (Existence and Uniqueness of Reduced Row Echelon Form)

Let $A \in M_{m\times n}(\mathbb{R})$. Then there exists a unique $A' \in M_{m\times n}(\mathbb{R})$ such that A' is in reduced row echelon form and $A \sim A'$. We define A' to be the **reduced row echelon form** of A, and denote

$$A' = \text{rref}(A).$$

Proof. Formalizing our steps in the previous examples, we have an algorithm for reducing a matrix into rref.

Algorithm 1: Gaussian-Jordan Reduction of $A \in M_{m\times n}(\mathbb{R}^n)$

```
i ← 1, j ← 1 while j ≤ n do
    for k in i...m do
        /* Find a row with non-zero entry in that column */
        if a_{k,j} ≠ 0 then
            swap(row i, row k)                  /* Move row k to the top row */
            row i ← 1/a_{i,j}× row i                /* Normalize pivot to 1 */
            for l in 1...m, l ≠ i do
                row l ← row l - (a_{l,j}× row i)  /* all other entries in the same column
                  = 0 */
            end
            i ← i + 1          /* Next row with pivot goes to the row below */
            Break
        end
    end
    j ← j + 1                            /* check pivot in next column */
end
```

∎

Example 2.34

Compute

$$\operatorname{rref}\left(\begin{bmatrix} 2 & -1 & 1 & 4 \\ 0 & 0 & 0 & 4 \\ 3 & 3 & 6 & 1 \\ -1 & -4 & -5 & 8 \end{bmatrix}\right)$$

We just repeat the algorithm to get

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

$$\sim \begin{bmatrix} 1 & \frac{-1}{2} & \frac{1}{2} & 2 \\ 0 & 0 & 0 & 4 \\ 3 & 3 & 6 & 1 \\ -1 & -4 & -5 & 8 \end{bmatrix} \text{ There is a non-zero entry in the first column, so we can reduce it to one}$$

$$\sim \begin{bmatrix} 1 & \frac{-1}{2} & \frac{1}{2} & 2 \\ 0 & 0 & 0 & 4 \\ 0 & \frac{9}{2} & \frac{9}{2} & -3 \\ 0 & \frac{-9}{2} & \frac{-9}{2} & 10 \end{bmatrix} \text{ add } -3 * \text{ row 1 to row 3, add row 1 to row 4}$$

$$\sim \begin{bmatrix} 1 & \frac{-1}{2} & \frac{1}{2} & 2 \\ 0 & \frac{9}{2} & \frac{9}{2} & -3 \\ 0 & 0 & 0 & 4 \\ 0 & \frac{-9}{2} & \frac{-9}{2} & 10 \end{bmatrix} \text{ Move next row with a non-zero entry}$$

$$\sim \begin{bmatrix} 1 & \frac{-1}{2} & \frac{1}{2} & 2 \\ 0 & 1 & 1 & \frac{2}{3} \\ 0 & 0 & 0 & 4 \\ 0 & \frac{-9}{2} & \frac{-9}{2} & 10 \end{bmatrix}$$

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

$$\sim \begin{bmatrix} 1 & 0 & 1 & \frac{7}{3} \\ 0 & 1 & 1 & \frac{2}{3} \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 7 \end{bmatrix} \text{ Third column has no pivot, so find pivot in fourth column}$$

$$\sim \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ Third column has no pivot, so find pivot in fourth column}$$

2.2.2 Recovering solutions from rref

Definition 2.35 (Column space)

Let $A \in M_{n\times m}(\mathbb{R})$, we define the **column space** of A $\mathrm{col}(A)$ to be the set of $\vec{b} \in \mathbb{R}^n$ for which

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

has at least one solution.

Proposition 2.36

If we write $A = [\vec{a}_1\ \vec{a}_2\ \dots\ \vec{a}_m]$ then

$$\mathrm{col}(A) = \mathrm{span}(\vec{a}_1, \dots, \vec{a}_m).$$

One of the exercises will guide you through the proof.

Producing one solution

Let us consider a general matrix in rref:

$$\left[\begin{array}{ccccc|c} 1 & * & * & * & * & a \\ 0 & 0 & 1 & * & * & b \\ 0 & 0 & 0 & 0 & 1 & c \\ 0 & 0 & 0 & 0 & 0 & d \end{array}\right]$$

If $d \neq 0$, the rref form will imply d is a pivot and thus equal to 1. So the last equation requires $0 = 1$, which means it has no solutions!

Now we set $d = 0$, so a, b, c can be anything. Regardless of the * entries, we can set for the columns with pivots, $x_1 = a, x_3 = b, x_5 = c$, and for the columns corresponding to non-pivots, $x_2 = x_4 = 0$. We thus have

Theorem 2.37 (Solutions to linear systems)

Let $A \in M_{m\times n}(\mathbb{R}), \vec{b} \in \mathbb{R}^m$, the system of equations described by $A\vec{x} = \vec{b}$ has

$$\begin{cases} \text{No solutions, if rref}\left(\left[A|\vec{b}\right]\right) \text{ has a pivot in the last column,} \\ \text{At least one solution, if the last column does not have a pivot} \end{cases}$$

Moreover, if the latter case holds, one solution (a particular solution) can be constructed from the rref by just solving the columns of rref$\left(\left[A|\vec{b}\right]\right)$ with pivots, and setting the variables corresponding to non-pivot columns to 0.

Producing all solutions

Now, suppose $A\vec{x} = \vec{b}$ has at least one solution. It can have infinitely many solutions! We want to describe this set of solutions without actually writing infinitely many vectors. One starting point is to see what properties this set satisfies. One starting point would be to guess that the set $\{\vec{x} \in \mathbb{R}^n | A\vec{x} = \vec{b}\}$ is a subspace of $\mathbb{R}^n$. This is generally not true, as every subspace contains $\vec{0}_{\mathbb{R}^n}$, and since $A\vec{0}_{\mathbb{R}^n} = \vec{0}_{\mathbb{R}^m}$, the only possible way this could be a subspace is when $\vec{b} = \vec{0}_{\mathbb{R}^m}$.

We thus start with the solution set to $A\vec{x} = \vec{0}$.

Definition 2.38 (Nullspace)

Let $A \in M_{m\times n}(\mathbb{R})$. The **nullspace** of the A, denoted $N(A)$, is the set of all solutions $\vec{x}$ for $A\vec{x} = \vec{0}$. i.e.

$$N(A) = \{\vec{x} \in \mathbb{R}^n | A\vec{x} = \vec{0}_{\mathbb{R}^m}\}.$$

Proposition 2.39

Let $A \in M_{m\times n}(\mathbb{R})$.

$$N(A) = N(\text{rref}(A)).$$

Proof. Suppose we can apply a sequence of row operations to A to get rref(A), the same row operations reduces $[A|\vec{0}]$ to $[\text{rref}(A)|\vec{0}]$. Or, $A\vec{x} = \vec{0}$ and rref$(A)\vec{x} = \vec{0}$ have the same solutions in $\vec{x}$. □

This allows us to limit the discussion of nullspaces to when A is in rref.

It turns out that the nullspace is a subspace. This is why we tend to like systems of linear equations where the right hand side is all zero. We call these systems **homogeneous**, and the solutions to homogeneous linear systems form a subspace.

Theorem 2.40 (Nullspace is a Subspace)

Let $A \in M_{m\times n}(\mathbb{R})$. $N(A)$ is a subspace of $\mathbb{R}^n$.

Proof. **Contains zero vector:** We can compute $A\vec{0}_{\mathbb{R}^n} = \vec{0}_{\mathbb{R}^m}$, so $\vec{0}_{\mathbb{R}^n} \in N(A)$.

Closed under addition: If $\vec{x}, \vec{y} \in N(A)$, we have $A(\vec{x}+\vec{y}) = A\vec{x} + A\vec{y} = \vec{0}_{\mathbb{R}^m} + \vec{0}_{\mathbb{R}^m} = \vec{0}_{\mathbb{R}^m}$. So $\vec{x}+\vec{y} \in N(A)$.

Closed under scalar multiplication: If $\vec{x} \in N(A)$, we have, for any $c \in \mathbb{R}$, $A(c\vec{x}) = c(A\vec{x}) = c\vec{0}_{\mathbb{R}^m} = \vec{0}_{\mathbb{R}^m}$. So $c\vec{x} \in N(A)$. □

Example 2.41

Find the solutions to the system represented by

$$\left[\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & -3 & 0 \\ 0 & 1 & 1 & 0 & 4 & 0 \\ 0 & 0 & 0 & 1 & -2 & 0 \end{array}\right].$$

Express the answer in the form of the span of a set of linearly independent vectors.

Let us start with the last variable x_5 and work our way towards x_1. There is no equation that fixes x_5, so let $x_5 = s$ for some $s \in \mathbb{R}$. Then the last equations fixes $x_4 = 2s$.

Again, we do not have an equation that fixes x_3, so let $x_3 = t$ for some $t \in \mathbb{R}$. This will force (from the second equation) $x_2 = -t - 4s$ and (from the first equation) $x_1 = -t + 3s$.

Our solution vector $(x_1, x_2, x_3, x_4, x_5)$ thus has the form

$$\begin{bmatrix} -t+3s \\ -t-4s \\ t \\ 2s \\ s \end{bmatrix} = s\begin{bmatrix} 3 \\ -4 \\ 0 \\ 2 \\ 1 \end{bmatrix} + t\begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

This describes exactly the span of the vectors

$$\begin{bmatrix} 3 \\ -4 \\ 0 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

To confirm these vectors are linearly independent, notice the third entry of the first vector and the fifth entry of the second vector is 0.

From this example, we have an algorithm to solve for the nullspace as a span of vectors. This is what we did in the previous example, except written in pseudocode. The idea is that each

Algorithm 2: Generating the Nullspace of Matrix A

1 $A \leftarrow \mathrm{rref}(A)$
2 $S \leftarrow$ empty set
3 **for** *each column j without a pivot* **do**
4     $\vec{v} \leftarrow \vec{e}_j$
5   **for** *each row i with a pivot* **do**
6   Locate pivot of row i in column k
7         $\vec{v} \leftarrow \vec{v} - a_{i,j}\vec{e}_k$
8   **end**
9   Add $\vec{v}$ to S
10 **end**
11 **return** $\mathrm{span}(S)$

column without a pivot denotes a free variable, while each column with a pivot is a fixed by the free variables. This is what we do in lines 5 to 8, after fixing our free variable to be 1, we subtract from the fixed variables. In particular, *there is at most one solution if every column has a pivot.* Now we return back to solving $A\vec{x} = \vec{b}$. We showed that the solutions do not form a subspace. However, these solutions are related to the nullspace.

Example 2.42

Find the solutions to the system represented by

$$\left[\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & -3 & 1 \\ 0 & 1 & 1 & 0 & 4 & 2 \\ 0 & 0 & 0 & 1 & -2 & 3 \end{array}\right].$$

Similar to computing nullspace, let us start with the last variable x_5 and work our way towards x_1. There is no equation that fixes x_5, so let $x_5 = s$ for some $s \in \mathbb{R}$. Then the last equations fixes $x_4 = 3 + 2s$.

Again, we do not have an equation that fixes x_3, so let $x_3 = t$ for some $t \in \mathbb{R}$. This will force (from the second equation) $x_2 = 2 - t - 4s$ and (from the first equation) $x_1 = 1 - t + 3s$.

Our solution vector $(x_1, x_2, x_3, x_4, x_5)$ thus has the form

$$\begin{bmatrix} 1 - t + 3s \\ 2 - t - 4s \\ t \\ 3 + 2s \\ s \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \\ 0 \end{bmatrix} + s \begin{bmatrix} 3 \\ -4 \\ 0 \\ 2 \\ 1 \end{bmatrix} + t \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

This is the nullspace, but shifted by the vector $(1, 2, 0, 3, 0)$.

Proposition 2.43

Let $\vec{x}_1, \vec{x}_2 \in \mathbb{R}^n$ be solutions to $A\vec{x} = \vec{b}$.Then

$$(\vec{x}_1 - \vec{x}_2) \in N(A).$$

Proof. We compute directly that $A(\vec{x}_1 - \vec{x}_2) = \vec{b} - \vec{b} = \vec{0}$. □

Corollary 2.44: Let $\vec{v}$ be *one* solution to $A\vec{x} = \vec{b}$. Then all the solutions of $A\vec{x} = \vec{b}$ is exactly

$$\{\vec{v} + \vec{n} | \vec{n} \in N(A)\}.$$

This means we only need one solution (particular solution), and the nullspace (homogeneous) solution to solve any matrix. This one solution can be obtained by setting all the 'free variables' in the solution to zero, and solving for all the pivots.

When producing solutions, the number of pivots of rref(A) is important. Let us define the number of pivots to be the rank of A.

Definition 2.45 (Rank)

Let $A \in M_{m\times n}(\mathbb{R})$. We define the **rank** of A

$$\text{rank}(A) = \text{the number of pivots in rref}(A).$$

We can thus formulate having zero, one or infinitely many solutions in terms of the number of pivots.

Proposition 2.46

Let $A \in M_{m\times n}(\mathbb{R}), \vec{b} \in \mathbb{R}^m$, the system of equations described by $A\vec{x} = \vec{b}$ has

$$\begin{cases} \text{At least one solution, if rank}\left(\left[A|\vec{b}\right]\right) = \text{rank}(A) \\ \text{At most one solution, if rank}(A) = n. \end{cases}$$

Going back to why we started solving linear systems, we want to find a test to see if the vectors are linearly independent, or if some vector is in the span of a set of vectors. Writing $A = [\vec{v}_1 \ \ldots \ \vec{v}_n]$, $A\vec{x} = \vec{b}$ having at least one solution means $\vec{b} \in \text{span}(\vec{v}_1, \ldots, \vec{v}_n)$, and if $A\vec{x} = \vec{0}$ has exactly one solution, the $\vec{v}_j$'s are linearly independent.

Theorem 2.47 (Rank and span/linear independence correspondence)

Let $A = [\vec{v}_1 \ \ldots \ \vec{v}_n] \in M_{m\times n}(\mathbb{R})$, $\vec{b} \in \mathbb{R}^m$. Then

- $\text{rank}\left(\left[A|\vec{b}\right]\right) = \text{rank}(A) \iff \vec{b} \in \text{span}(\vec{v}_1, \ldots, \vec{v}_n)$.
- $\text{rank}(A) = m \iff \text{span}(\vec{v}_1, \ldots, \vec{v}_n) = \mathbb{R}^m$.
- $\text{rank}(A) = n \iff \{\vec{v}_1, \ldots, \vec{v}_n\}$ is linearly independent.

Remark. *This is the first time we used the $\iff$ symbol. This means 'if and only if', or 'iff' for short. '$A \iff B$' means that if A holds, then B holds. And if B holds, A holds. For instance, 'more than 99% guarantees an A' means*

You have more than 99% in this course $\implies$ you get an A

But you can still get an A without reaching this score, for instance a 95% could get an A too. An if and only if statement (for a strict grading class) would look more like

The threshold for A in this class is a 90%.

so {You get A} $\iff$ {You get 90+%}.

Proof. The first and third statements are evident from the definitions of span and linear independence.

To show the second statement, we first prove the $\implies$ direction. Let $\vec{b} \in \mathbb{R}^m$, then reducing $[A|\vec{b}]$ would give m pivots in the first n columns, so there cannot be a pivot in the final column. (or else, we would have $m+1$ pivots in a matrix with only m rows). Therefore, the system has a solution.

For the $\impliedby$ direction, we can reverse all the elementary operations from A to rref(A) to get from $[\text{rref}(A)|\vec{e}_m]$ to $[A|\vec{w}]$ for some $\vec{w} \in \mathbb{R}^m$. This system of equations has no solution. □

Something amazing also happens when the vectors are a basis (recall it means both linearly independent and spans $\mathbb{R}^m$).

The second and third statements forces $m = \text{rank}(A) = n$, so A is a square matrix. The vectors span $\mathbb{R}^m$ iff they are linearly independent. Moreover, since rank$A = n$ for this square matrix, rref$A = I_n$, the identity matrix.

Theorem 2.48 (This One/ That One)

Let $A \in M_{n\times n}(\mathbb{R})$, and write $A = [\vec{v}_1 \dots \vec{v}_n]$. Then the following are equivalent.

1. $\{\vec{v}_n, \dots, \vec{v}_n\}$ form a basis.
2. $\{\vec{v}_n, \dots, \vec{v}_n\}$ are linearly independent.
3. $\{\vec{v}_n, \dots, \vec{v}_n\}$ span $\mathbb{R}^n$.
4. The system of equations $A\vec{x} = \vec{b}$ has exactly one solution for each $\vec{b} \in \mathbb{R}^n$.
5. The system of equations $A\vec{x} = \vec{b}$ has at most one solution for each $\vec{b} \in \mathbb{R}^n$.
6. The system of equations $A\vec{x} = \vec{b}$ has at least one solution for each $\vec{b} \in \mathbb{R}^n$.
7. $\text{rref}(A) = I_n$.
8. rref(A) has a pivot in every column.
9. rref(A) has a pivot in every row.
10. Rank$(A) = n$.
11. $N(A) = \{\vec{0}\}$.
12. $\text{col}(A) = \mathbb{R}^n$.

Remark. *This One Theorem goes by many names: Invertibility Theorem, Inverse Matrix Theorem, Inverse Function Theorem, and (my favorite) Amazingly Awesome Theorem by Prof. Cañez. Despite the lack of standardization, it is covered in every linera algebra course in some form or another. I'm putting it as 'That One Theorem', so when you quote 'this follows from That One Theorem' everyone will understand which one you are talking about.*

Remark. *This One Theorem is still incomplete. We will build towards more equivalences.*

Exercises

6. Determine if the following sets of vectors are linearly dependent or independent. If they are not linearly independent, generate the set of homogenous solutions to the corresponding linear system.

   (a) $\{(1,2,3),(0,1,1),(1,1,2)\}$

   (b) $\{\vec{e}_1,\vec{e}_2,\vec{e}_5\} \subset \mathbb{R}^7$

7. Let $m > n$. Show that every set of m vectors in $\mathbb{R}^n$ is not linearly independent. Show that every set of n vectors in $\mathbb{R}^m$ does not span m. Thus show that every basis for $\mathbb{R}^n$ has n vectors.

8. Let $\{v_1, ..., v_k\} \subset V$. Show that exactly one of the following statements hold:

   (a) For every $v \in \text{span}(\{v_1, ..., v_k\})$, there is exactly one way to write v as a linear combination of $v_1, ..., v_k$.

   (b) For every $v \in \text{span}(\{v_1, ..., v_k\})$, there is more than one way to write v as a linear combination of $v_1, ..., v_k$.

9. *Every vector space has a basis (weaker version)* This is a weaker statement of Every Vector Space has a Basis. To get around this, we use a finiteness assumption that our vector space V is a subspace of $\mathbb{R}^n$.

   (a) Suppose that our subspace is $\{\vec{0}\}$. Show that this is spanned by the empty set. The empty set is (vacuously) linearly independent*, so we have a basis. ** all elements of an empty set satisfy any condition*

   (b) Now suppose $S = \{\vec{v}_1, \ldots, \vec{v}_k\} \subset V \subseteq \mathbb{R}^n$. Show that this is a linearly independent set in V if and only if this is a linearly independent set $\mathbb{R}^n$, so we can call it linearly independent without worrying about which vector space we work in.

   (c) Let S be linearly independent. If $\text{span}(S) = V$ this will be a basis for V. Else show that there is some $\vec{v}_{k+1} \in V$ such that the set $S + \{\vec{v}_{k+1}\}$, so you can extend S to a bigger linearly independent set.

   (d) Show that the previous process terminates, i.e. you cannot add in infinitely many vectors and still make S linearly independent. Therefore, it stops at some k where $\text{span}(\{\vec{v}_1, \ldots, \vec{v}_k\}) = V$.

10. Show that every subset of a linearly independent set is linearly independent.

11. Compute the nullspace of the matrix

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

and give a basis for it. Do the same for the matrices

$$A = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 2 & -1 & 1 & 0 \\ -1 & 4 & -1 & 2 \end{bmatrix}, B = \begin{bmatrix} 1 & 3 & 4 & 0 & 2 \\ 2 & 7 & 6 & 1 & 1 \\ 4 & 13 & 14 & 1 & 3 \end{bmatrix}$$

12. Solve the folling systems of equations

(a) $\left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & -1 \\ 1 & 0 & 1 & 0 \end{array}\right]$

(b) $\left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 2 & 3 & 2 & -1 \\ 1 & 1 & -1 & 10 \end{array}\right]$

(c) $\left[\begin{array}{cccc|c} 1 & 1 & -2 & 3 & 9 \\ 2 & 1 & 0 & 1 & -18 \\ 1 & -1 & 1 & 0 & -9 \\ 3 & 1 & 2 & 1 & 9 \end{array}\right]$

13. *(HKDSE 2019 M2 Q6)* For the system of equations in x, y, z:

$$\begin{array}{rcrcrcr} x & - & 2y & - & 2z & = & \beta \\ 5x & + & \alpha y & + & \alpha z & = & 5\beta \\ 7x & + & (\alpha-3)y & + & (2\alpha+1)z & = & 8\beta \end{array}$$

Find (i) α such that the system has a unique solution, and solve for y in α, β (ii) β such that the system is inconsistent when $\alpha = -4$.

14. *(HKDSE 2021 M2 Q8)* Find the value of d such that the system of equations in x, y, z:

$$\begin{array}{rcrcrcr} x & + & (d-1)y & + & (d+3)z & = & 4-d \\ 2x & + & (d+2)y & - & z & = & 2d-5 \\ 3x & + & (d+4)y & + & 5z & = & 2 \end{array}$$

has infinitely many solutions, and solve the system of equations.

2.3 Linear Transformations

Matrix-vector multiplication preserve linear combinations. Let us name this property.

Definition 2.49 (Linear Transformation)

Let V and W be $\mathbb{K}$-vector spaces. A function $f : V \to W$ is a **linear transformation** if it preserves linear combinations. Namely, for every $c \in \mathbb{K}, \vec{v}_1, \vec{v}_2 \in V$,

- $f(\vec{v}_1 + \vec{v}_2) = f(\vec{v}_1) + f(\vec{v}_2)$,
- $f(c\vec{v}_1) = cf(\vec{v}_1)$.

Many functions in the wild are linear transformations. This includes the functions we have introduced so far.

Example 2.50

The following are examples linear transformations:

1. A matrix $A \in M_{m\times n}(\mathbb{R})$ admits a linear transformation $F : \mathbb{R}^n \to \mathbb{R}^m$ by
$$F(\vec{x}) = A\vec{x}.$$
2. Let $\vec{v} \in \mathbb{R}^n$. $f(\vec{x}) = \vec{v} \cdot \vec{x}$ is a linear transformation from $\mathbb{R}^n \to \mathbb{R}$. The projection $\mathrm{Proj}_{\vec{v}}$ is a linear transformation from $\mathbb{R}^n \to \mathbb{R}^n$.
3. Let $\vec{v} \in \mathbb{R}^3$. $f(\vec{x}) = \vec{v} \times \vec{x}$ is a linear transformation from $\mathbb{R}^3 \to \mathbb{R}^3$.
4. Recall that the set of polynomials $f(x)$ is a real vector space. The differentiation operator $\frac{d}{dx}$ is a linear transformation from the set of polyomials to itself.
5. Integration over a bounded interval $\int_a^b$ is a linear transformation from the space of integrable functions to $\mathbb{R}$.

Proposition 2.51

Let U, V, W be $\mathbb{K}$-vector spaces, and $T_1 : U \to V, T_2 : U \to V, T_3 : V \to W$ be linear transformations, then the following are also linear transformations:

- $f_1 : U \to V,\ f_1(\vec{x}) = T_1(\vec{x}) + T_2(\vec{x})$
- $f_2 : U \to V,\ f_2(\vec{x}) = cT_1(\vec{x})$ for any $c \in \mathbb{K}$
- $f_3 : U \to W,\ f_3(\vec{x}) = T_3(T_1(\vec{x}))$.

Proposition 2.52

Linear transformations are characterized by the image of a basis. Specifically, let $v_1, \ldots, v_n$ be a basis for V, and $u_1, \ldots, u_n \in U$. Then there is a unique linear transformation $T : V \to U$ satisfying

$$\begin{aligned} T(v_1) &= u_1, \\ T(v_2) &= u_2, \\ &\vdots \\ T(v_n) &= u_n. \end{aligned}$$

Proof. Let us show uniqueness of such a transformation. (That is, it might not exist for some values of u_k's, but when it does, it has to be unique.)

Since every $v \in V$ is a unique linear combination of the basis vectors for any $v = c_1v_1 + \ldots + c_nv_n$, we must have (by linearity of T),

$$T(v) = T(c_1v_1 + \ldots + c_nv_n) = c_1T(v_1) + \ldots + c_nT(v_n) = c_1u_1 + \ldots + c_nu_n.$$

This forces T to have only one form. Finally, we can check that if we define

$$T(c_1v_1 + \ldots + c_nv_n) = c_1u_1 + \ldots + c_nu_n,$$

this is well defined as each v is a unique linear combination of the basis vectors (we exactly one value of $T(v)$ assigned for each v), and it satisfies the conditions

$$\begin{aligned} T(v + w) &= T(v) + T(w) \\ T(cv) &= cT(v) \end{aligned}$$

so is a linear transformation. □

The first example of linear transformation given $A\vec{x}$ is a very general type of linear transformation. The punchline is that it is **the** example of a linear transformation.

Theorem 2.53 (Matrix Representation of Linear Transformation)

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there is a matrix $A \in M_{m\times n}$ such that for all $\vec{x} \in \mathbb{R}^n$,

$$T(\vec{x}) = A\vec{x}.$$

We denote this matrix the **matrix representation** of T.

Remark. *Importantly, this theorem is only applicable for $\mathbb{R}^n \to \mathbb{R}^m$, both are **finite dimensional**. Some linear functions work with infinite dimensional (we will give a definition later) vector spaces. Intuitively, this means the analogous 'standard basis vectors' is an infinite set $\{\vec{e}_1, \vec{e}_2, \ldots\}$.*

Before we head to the details of the proof, here is the idea behind our constuction. Let us suppose that this matrix exists; how do we compute each entry of the matrix? The trick is to think of the matrix as $\begin{bmatrix} \vec{w}_1 & \vec{w}_2 & \ldots & \vec{w}_n \end{bmatrix}$. We can get a linear combination of these to equal any of the $\vec{w}_j$'s. Namely, using the standard basis vectors, $A\vec{e}_j = \vec{w}_j$. Since $A\vec{e}_j = T(\vec{e}_j)$That means our matrix must be in the form

$$\begin{bmatrix} T(\vec{e}_1) \; T(\vec{e}_2) \; \ldots \; T(\vec{e}_n) \end{bmatrix}.$$

Now all we need to do check that this matrix works.

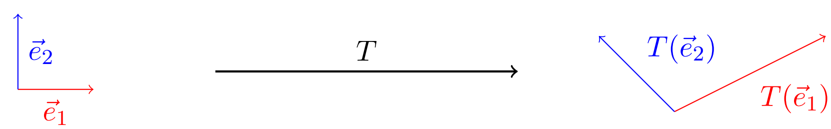
Proof. We claim that the matrix $A = \begin{bmatrix} T(\vec{e}_1) \; T(\vec{e}_2) \; \ldots \; T(\vec{e}_n) \end{bmatrix}$ satisfies $T(\vec{x}) = A\vec{x}$. Let $\vec{x} \in \mathbb{R}^n$. Then $\vec{x} = \sum_{i=1}^n x_i\vec{e}_i$. So that by linearity,

$$T(\vec{x}) = \sum_{i=1}^n x_i T(\vec{e}_i) = A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A\vec{x}.$$

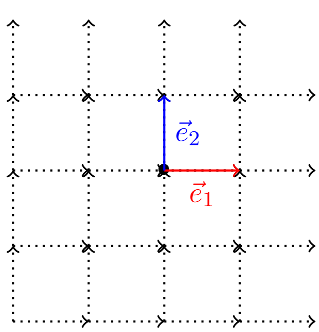
□

More intuition behind the proof

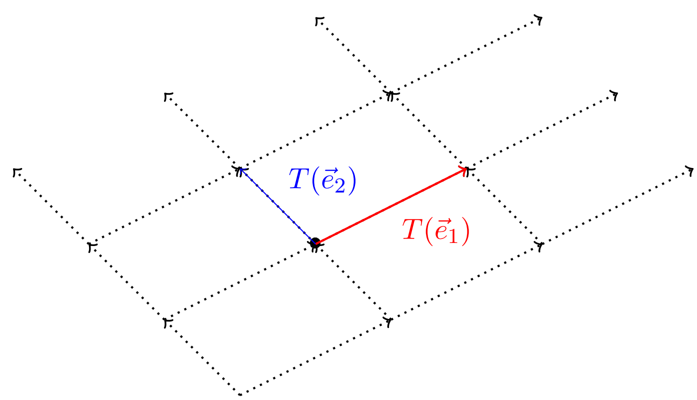
We can also think of a linear transformation geometrically. Let us track the transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ that sends $T(\vec{e}_1) = (2, 1)$ and $T(\vec{e}_2) = (-1, 1)$. Let us only consider the integer lattice



points for now. If we start at (a, b), meaning from the origin, we take a steps right and b steps up. In our transformed plane, we need to take a steps in $(2, 1)$ and b steps $(-1, 1)$ by linearity. We can geometrically track the place of these lattice points using a grid. This is just a change of our



(a) Original grid



(b) Transformed grid

coordinate system. As long as we know where the standard basis vectors go (or any basis goes), it is enough to know the whole linear transformation. The construction is you 1. decompose any vector into the linear combination of the basis vectors, then 2. add up the corresponding linear combination of the transformed basis vectors. **This intuition of linear transformations being a shift in coordinates is very useful**, and helps transform abstract problems in linear algebra back to concrete geometry.

It also helps to define analogs of column and nullspace for the transformations.

Definition 2.54 (Image and Kernel)

Let $T : V \to U$ be a linear transformation. We define the **image** of T, denoted as $\mathrm{Im}(T)$ to be the set

$$\{u \in U | T(\vec{v}) = u \text{ has a solution in v}\}.$$

We define the **kernel** of T, denoted as $\ker(T)$, to be the set

$$\{v \in V | T(v) = 0_U\}.$$

Remark. *The image and kernel of T are subspaces of U and V respectively. Indeed, if we use the matrix representation of $T : \mathbb{R}^n \to \mathbb{R}^m$, we can see the image of T is the column space of A and the kernel of T is the nullspace of A.*

Example 2.55

Express the linear transformations of the dot product, cross product, and projection as matrix-vector multiplications.

Dot Product: Let $\vec{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$. Then $\vec{v} \cdot \vec{e}_j = v_j$. So

$$\vec{v} \cdot \vec{x} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \vec{x}$$

.

Cross Product Let $\vec{v} = (v_1, v_2, v_3) \in \mathbb{R}^n$. Then

$$\vec{v} \times \vec{x} = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix} \vec{x}$$

Projection Let $\vec{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$. Then

$$\mathrm{Proj}_{\vec{v}}(\vec{x}) = \frac{1}{|\vec{v}|^2} \begin{bmatrix} v_1v_1 & v_1v_2 & \dots & v_1v_n \\ v_2v_1 & v_2v_2 & \dots & v_2v_n \\ \vdots & & \ddots & \\ v_nv_1 & v_nv-2 & \dots & v_nv_n \end{bmatrix} \vec{x}$$

where multiplication of matrix by scalar is entry wise.

Example 2.56

Let $T : \mathbb{R}^m \to \mathbb{R}^n$, $S : \mathbb{R}^n \to \mathbb{R}^p$. Let $A \in M_{m\times n}, B \in M_{n\times p}$ be the matrix representations of S and T respectively. We know that the function $F : \mathbb{R}^m \to \mathbb{R}^p$ defined by $F(\vec{x}) = S(T(\vec{x}))$ has a matrix representation. Compute this matrix.

The computation is not very interesting. However, you will do this until the day you drop ISP. If we write $A = \begin{bmatrix}\vec{x}_1 & \dots & \vec{x}_m\end{bmatrix}, \vec{b}_j \in \mathbb{R}^n$, then for each $\vec{e}_j \in \mathbb{R}^m$ we have $A\vec{e}_j = \vec{x}_j$. So that

$$S(T(\vec{e}_j)) = S(A\vec{e}_j) = S(\vec{x}_j) = B(\vec{x}_j).$$

The matrix is

$$\begin{bmatrix}B\vec{x}_1 & B\vec{x}_2 & \dots & B\vec{x}_m\end{bmatrix}.$$

Now, if we forcefully write the linear transformations in matrices, we get $S(T(\vec{x})) = BA\vec{x}$. This allows us to define the product of matrices.

Definition 2.57 (Matrix Operations)

Let $\alpha \in \mathbb{R}$, $A = \{a_{i,j}\}, B = \{b_i, j\} \in M_{m\times n}(\mathbb{R})$, $C = \{c_{i,j}\} \in M_{n\times p}(\mathbb{R})$. We define the addition of matrices

$$A + B = \{a_{i,j} + b_{i,j}\},$$

the pointwise addition of each entry. We define the scaling of a matrix

$$\alpha A = \{\alpha a_{i,j}\},$$

the pointise scaling of each entry. We also define the product of matrices

$$AC = \left\{\sum_{k=1}^{n} a_{i,k}c_{k,j}\right\}.$$

Remark. *The definition of matrix product is equivalent to what we got from getting the matrix from the composition of functions. The only difference is that we are composing a function the other way from* $\mathbb{R}^p \to \mathbb{R}^n \to \mathbb{R}^m$.

Remark. *There is a mnenomic to memorize the formula for matrix product* AB*. It involves putting* A *on the bottom left and* B *on the top right, like so:*

$$\begin{array}{cc} & \begin{bmatrix} b_{1,1} & \dots & b_{1,p} \\ \vdots & \ddots & \\ b_{n,1} & \dots & b_{n,p} \end{bmatrix} \\ \begin{bmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \ddots & \\ a_{m,1} & \dots & a_{m,n} \end{bmatrix} & \begin{bmatrix} \\ \textit{product goes here} \\ \\ \end{bmatrix} \end{array}$$

The product of an $m \times n$ *and an* $n \times p$ *matrix is an* $m \times p$ *matrix. To find the* i, j*'th entry of the product, locate the* i*-th row on matrix* A *(left) and the* j*-th column on matrix* B *(top). Treating the row and column as vectors with* n *entries, the dot product is the* i, j*-th entry of the product.*

Remark. *Viewing a column vector as a* $n \times 1$ *matrix, you can verify that the matrix-vector product is compatible with the products of two matrices.*

Proposition 2.58

The matrix operations satisfy the following properties:
Addition: Let $A, B, C \in M_{m\times n}(\mathbb{R})$, then:

- *(Associativity)* $A + (B + C) = (A + B) + C$.
- *(Commutativity)* $A + B = B + A$.
- *(Identity)* The zero matrix $0_{m\times n}$ satisfies $A + 0_{m\times n} = A$.

Scaling: Let $\alpha, \beta \in \mathbb{R}, A, B \in M_{m\times n}(\mathbb{R})$, then:

- *(Associativity)* $\alpha(\beta)A = (\alpha\beta)A$.
- *(Distributivity)* $(\alpha + \beta)(A + B) = \alpha A + \alpha B + \beta A + \beta B$.

Multiplication: Let $A, B \in M_{m\times n}(\mathbb{R}), C, D \in M_{n\times p}(\mathbb{R}), E \in M_{p\times q}(\mathbb{R})$.

$$\begin{tikzcd} \mathbb{R}^q \arrow[r, "E"] & \mathbb{R}^p \arrow[r, bend left, "C"] \arrow[r, bend right, "D"'] & \mathbb{R}^n \arrow[r, bend left, "A"] \arrow[r, bend right, "B"'] \arrow[loop below, "I_n"] & \mathbb{R}^m \end{tikzcd}$$

Then

- *(Associativity)* $A(CE) = (AC)E$.
- *Compatibility with Scaling* $A(\alpha C) = \alpha(AC) = (\alpha A)(C)$.
- *(Distributivity)* $(A + B)(C + D) = AC + AD + BC + BD$.
- *(Identity)* The identity matrix $I_n \stackrel{\text{def}}{=} \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \dots & \vec{e}_n \end{bmatrix}$ satisfies $I_n C = C$ and $BI_n = B$.

Remark. *The matrix product is generally not commutative. In fact, the product DB does not make sense, because the dimensions of the matrices are not compatible!*

Exercises

15. Compute the following products

(a) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

(c) $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

(d) $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

16. Solve

$$\begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix} X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

for the 2×2 matrix X.

17. Refer to the previous problem, express row operations as multiplication on the left. That is, if A' is A after applying a row operation, find the corresponding matrix X such that $A' = XA$.

18. Find the corresponding matrices for the following transformations:

    (a) The unique linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, T\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ -3 \end{bmatrix}$$

19. Find an example where

    (a) Two matrices $A, B \in M_{n\times n}(\mathbb{R})$ are not commutative. i.e. $AB \neq BA$.

    (b) A matrix $0_{n\times n} \neq A \in M_{n\times n}(\mathbb{R})$ such that $A^k = 0_{n\times n}$ for some $k > 1$.

20. Let $\theta, \phi \in [0, 2\pi)$. Sketch how the matrix

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

transforms the grid. Show that

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} = \begin{bmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{bmatrix}$$

and thus the set of matrix in this form are commutative. Is there a geometric argument for why these matrices must be commutative?

21. Let

$$A = \begin{bmatrix} 1 & -1 & 3 \\ 0 & -2 & 2 \end{bmatrix}, B = \begin{bmatrix} 1 & 2 \\ 1 & 0 \\ -3 & 2 \end{bmatrix}, C = \begin{bmatrix} -1 & 1 & -3 \\ 0 & 2 & -2 \end{bmatrix}, D = \begin{bmatrix} -1 & -2 & 0 \\ 1 & -2 & 1 \\ 2 & 1 & -4 \end{bmatrix}.$$

Compute the follow quantities (could be undefined): $A+3B$, $A+C$, $C+2D$, AB, BA, CD, DC.

22. Prove or provide a counterexample: for matrices A, B, C where the products are well defined, if $AB = AC$ then $B = C$.

23. Prove or provide a counterexample: if A, B are not the zero matrix, then AB is not the zero matrix (where the matrix sizes are implicitly defined).

24. Run the following code in python multiple times and pinky promise that you got the correct answer.

```python
import numpy as np
matrix1=np.random.randint(-10,11,size=(3,3))
matrix2=np.random.randint(-10,11,size=(3,3))
print("Multiply these two matrices:")
print(matrix1,"\n",matrix2)
print("Confirm that the answer is the following:")
print(matrix1.dot(matrix2))
```

2.4 Surjective and Injective Linear Transformation

Definition 2.59 (Types of functions)

Let $f : U \to W$ be a function. We say f is:

- **Surjective**, if the image of U under f, $f(U) = W$. This means for every $w \in W$, there is some $u \in U$ such that $f(u) = w$.
- **Injective**, if each $u \in U$ maps to a unique element in W. This means for every $u_1, u_2 \in U$. If $f(u_1) = f(u_2)$, $u_1 = u_2$.
- **Bijective**, if f is surjective and injective.

Example 2.60

The following functions are surjective:

- The 'extract' function $\Pi_j : \mathbb{R}^n \to \mathbb{R}$, defined by $\Pi_j(x_1, x_2, ..., x_n) = x_j$ and extracts the j-th coordinate.
- The differentiation operator $\frac{d}{dx}$ from the space of **differentiable** functions to the space of **continuous** functions. This is because all continuous functions have antiderivatives. Moreover, as all derivatives are continuous, the differentiation operator is well defined.
- Integration over the interval $[0, 1]$ sends the integrable function g to the real number $\int_0^1 g(x)\ dx$. This is surjective because the integral of $g(x) = a$ over $[0, 1]$ is a for all $a \in \mathbb{R}$.

The following functions are injective:

- The 'pad zeros' function $\iota : \mathbb{R}^n \to \mathbb{R}^{n+1}$ defined by $\iota(x_1, ..., x_n) = (x_1, ..., x_n, 0)$ and adds a 0 to the final coordinate.
- The linear function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 2x$.

The following functions are surjective:

- The identity map $Id : \mathbb{R}^n \to \mathbb{R}^n$ defined by $Id(\vec{x}) = I_n\vec{x} = \vec{x}$.

We can phrase these characterizations in of the number of solutions in u that solve $f(u) = w$.

Proposition 2.61

- f is surjective $\iff$ $f(u) = w$ has *at least one solution* in u for each $w \in W$.
- f is injective $\iff$ $f(u) = w$ has *at most one solution* in u for each $w \in W$.
- f is bijective $\iff$ $f(u) = w$ has *exactly one solution* in u for each $w \in W$.

When f is a linear transformation from $\mathbb{R}^n \to \mathbb{R}^m$, it has a matrix representation $A \in M_{m\times n}(\mathbb{R})$. To determine if f is surjective, injective, or bijective, we look at the number of solutions in $\vec{x}$ for the equation $A\vec{x} = \vec{b}$ for each $\vec{b} \in \mathbb{R}^m$.

Theorem 2.62 (Surjectivity and solutions to linear systems)

Let $T : \mathbb{R}^n \to \mathbb{R}^m$, $A \in M_{m\times n}(\mathbb{R})$, and $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$. Then the following are equivalent:

1. T is surjective.
2. $\mathrm{Im}(T) = \mathbb{R}^m$.
3. $A(\vec{x}) = \vec{b}$ is at least one solution in $\vec{x}$ for each $\vec{b} \in \mathbb{R}^m$
4. $\mathrm{rref}(A)$ has a pivot in every row.
5. The columns of A span $\mathbb{R}^m$.

Theorem 2.63 (Injectivity and solutions to linear systems)

Let $T : \mathbb{R}^n \to \mathbb{R}^m$, $A \in M_{m\times n}(\mathbb{R})$, and $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$. Then the following are equivalent:

1. T is injective.
2. $\ker(T) = \{\vec{0}\}$.
3. $A(\vec{x}) = \vec{b}$ is at most one solution in $\vec{x}$ for each $\vec{b} \in \mathbb{R}^m$.
4. $\mathrm{rref}(A)$ has a pivot in every column.
5. The columns of A are linearly independent.

Again, something interesting happens when $T : \mathbb{R}^n \to \mathbb{R}^m$ is bijective. This forces the matrix to be square, giving even more equivalences for That One Theorem.

Theorem 2.64 (This One/ That One (Linear transformation cont.))

Let $A = [\vec{v}_1 \dots \vec{v}_n] \in M_{n\times n}(\mathbb{R})$, and $T : \mathbb{R}^n \to \mathbb{R}^n$ be the unique linear transformation satisfying $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$. Then the following are equivalent.

1. $\{\vec{v}_n, \dots, \vec{v}_n\}$ form a basis.
2. $\{\vec{v}_n, \dots, \vec{v}_n\}$ are linearly independent.
3. $\{\vec{v}_n, \dots, \vec{v}_n\}$ span $\mathbb{R}^n$.

   $\vdots$

13. T is bijective.
14. T is injective.
15. T is surjective.
16. $\ker(T) = \{\vec{0}\}$.
17. $\mathrm{Im}(T) = \mathbb{R}^n$.

Exercises

25. Show that a linear transformation $T : V \to U$ is injectie $\iff$ $\ker(T) = \{0_V\}$.
26. Let $T : V \to U$, $S : U \to W$ be injective. Show that their composition $S \circ T : V \to W$ defined by $S \circ T(v) = S(T(v))$ is injective. Can $S \circ T$ still be injective if one of S or T fails to be injective?
27. Let $T : V \to U$, $S : U \to W$ be surjective. Show that their composition $S \circ T : V \to W$ is surjective. Can $S \circ T$ still be surjective if one of S or T fails to be surjective?

2.5 Inverses, Isomorphisms and determinants

2.5.1 Inverse

Definition 2.65 (Inverse function)

Let $T : U \to V$ be bijective. The **inverse of** T is the unique function T^{-1} that satisfies

$$T^{-1}(T(x)) = x,$$
$$T(T^{-1}(y)) = y$$

for all $x \in U, y \in V$.

$$\begin{tikzcd} U \arrow[rr, bend left, "T"] & & V \arrow[ll, bend left, "T^{-1}"] \end{tikzcd}$$

How do we know this inverse exists? We know that for each $y \in V$ there is exactly one $x \in U$ that satisifes $T(x) = y$, so we can set $T^{-1}(y)$ to be this particular value of x for each y. This definition of the inverse sounds tautological, so let us apply this to linear transformations.

Proposition 2.66

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a bijective linear transformation. Then T^{-1} is a linear transformation.

Proof. We want to show that for any $\vec{y}_1, \vec{y}_2 \in \mathbb{R}^n$, $c \in \mathbb{R}$, $T^{-1}(\vec{y}_1 + \vec{y}_2) = T^{-1}(\vec{y}_1) + T^{-1}(\vec{y}_2)$ and $T^{-1}(c\vec{y}_1) = cT^{-1}(\vec{y}_1)$.

The trick to solving this is to use the definition for inverses and switch back and forth with T and T^{-1}, and apply linearity of T.

$$\begin{aligned} T^{-1}(\vec{y}_1) + T^{-1}(\vec{y}_2) &= T^{-1}(T(T^{-1}(\vec{y}_1) + T^{-1}(\vec{y}_2))) \\ &= T^{-1}(T(T^{-1}(\vec{y}_1)) + T(T^{-1}(\vec{y}_2))) \text{ (as } T \text{ is linear)} \\ &= T^{-1}(\vec{y}_1 + \vec{y}_2). \end{aligned}$$

For scaling,

$$\begin{aligned} cT^{-1}(\vec{y}_1) &= T^{-1}(T(cT^{-1}(\vec{y}_1))) \\ &= T^{-1}(c(T(T^{-1}(\vec{y}_1)))) \\ &= T^{-1}(c\vec{y}_1). \end{aligned}$$

□

This means that the inverse of T has a matrix representation. If we write $T(\vec{x}) = A\vec{x}$ for some matrix A, then there is some matrix B such that $AB = BA = I_n$.

Definition 2.67 (Matrix Inverses)

Let $A \in M_{m\times n}(\mathbb{R})$. If there exists $X \in M_{n\times m}(\mathbb{R})$ such that

$$XA = I_n,$$

we call X the **left inverse** of A. Similarly, if there exists $Y \in M_{n\times m}(\mathbb{R})$ such that

$$AY = I_m,$$

we call Y the **right inverse** of A. If both the left and right inverses exist and are equal, then we say that A is **invertible**, and denote $X = Y$ the **inverse** of A, written as A^{-1}.

Remark. *If A has a left and right inverse, then the left and right inverses are equal. Indeed, we have for a left inverse X and a right inverse Y of A,*

$$X = X(I_m) = X(AY) = (XA)Y = (I_n)Y = Y.$$

As we have already seen, only bijective linear transformations have inverses, so that means we have more to add to That One Theorem. We also see that being invertible implies having a left and right inverse, so might expect that having a left or right inverse corresponds to the columns being linearly independent or span $\mathbb{R}^m$.

Example 2.68

$A \in M_{m\times n}(\mathbb{R})$ having a left/right inverse indeed corresponds to the conditions that the columns are linearly independent and the columns span $\mathbb{R}^m$. Figure out which conditions correspond to which.

There are two ways to think about this problem: one way is directly translate left and right inverses to the corresponding statements of the column vectors; the other way is to think of them in linear transformations and find out which one is surjective/injective. I will outline both methods.

Method 1: Without knowing where to go, let us suppose we want to prove A has linearly independent columns and see which extra hypothesis we need. Suppose

$$A\vec{x} = \vec{0}.$$

We want to show $\vec{x} = \vec{0}$. This means we want to cancel the A on the left, so we want to use the left inverse X.

$$\vec{x} = XA\vec{x} = X\vec{0} = \vec{0}.$$

So this means having a left inverse corresponds to linear independence. Now to prove the columns of A span $\mathbb{R}^m$, we try to solve

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

for any $\vec{b} \in \mathbb{R}^m$. To produce a solution $\vec{x}$ that can cancel A from the right, we will need the right inverse Y, and set $\vec{x} = Y\vec{b}$, then

$$A\vec{x} = AY\vec{b} = \vec{b}.$$

Method 2: We translate the problem into a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ being injective (linear independence) and surjective (span $\mathbb{R}^m$). Why does a left inverse imply injectivity? Let us suppose $T(\vec{x}) = T(\vec{y})$, then the left inverse can immediately cancel out the T from the left and give us $\vec{x} = \vec{y}$.

For surjectivity, let us break down the columns of the right inverse $Y = \begin{bmatrix} \vec{y}_1 & \dots & \vec{y}_m \end{bmatrix}$. Then by the property of matrix multipication, we can consider each column to see

$$A\vec{y}_j = \vec{e}_j.$$

Since the basis vectors can build anything in $\mathbb{R}^m$, we can just use the corresponding linear combination of the $\vec{y}_j$'s to build anything in $\mathbb{R}^m$. This means we have three extra equivalences for a matrix being invertible.

Theorem 2.69 (This/That One)

Let $A = [\vec{v}_1 \dots \vec{v}_n] \in M_{n\times n}(\mathbb{R})$, and $T : \mathbb{R}^n \to \mathbb{R}^n$ be the unique linear transformation satisfying $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$. Then the following are equivalent.

1. $\{\vec{v}_n, \dots, \vec{v}_n\}$ form a basis.
2. $\{\vec{v}_n, \dots, \vec{v}_n\}$ are linearly independent.
3. $\{\vec{v}_n, \dots, \vec{v}_n\}$ span $\mathbb{R}^n$.

$\vdots$

18. A is invertible.
19. A as a left inverse.
20. A has a right inverse.

Finally, we want to compute this inverse A^{-1}. Let us set the columns

$$A = [\vec{a}_1\ \vec{a}_2\ \dots\ \vec{a}_n],\ A^{-1} = [\vec{b}_1\ \vec{b}_2\ \dots\ \vec{b}_n],$$

then the condition $AA^{-1} = I_n$ translates to

$$\begin{aligned} A\vec{b}_1 &= \vec{e}_1, \\ A\vec{b}_2 &= \vec{e}_2, \\ &\vdots \\ A\vec{b}_n &= \vec{e}_n. \end{aligned}$$

Which means we just have to solve n different linear systems. We can do them all at once: by reducing the matrix

$$\left[\begin{array}{cccc|cccc} \vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_n & \vec{b}_1 & \vec{b}_2 & \dots & \vec{b}_n \end{array}\right]$$

to rref, we will get

$$\left[\begin{array}{c|c} I_n & * \end{array}\right]$$

if A is invertible. The $*$ will be the solutions $\vec{b}_k$'s, which means $*$ is the inverse of A.

2.5.2 Isomorphisms

Definition 2.70 (Isomorphism)

Let $T : U \to V$ be a bijective linear transformation. We call T an **isomorphism** and the vector spaces U and V are **isomorphic**. We denote this $U \simeq V$.

Recall many sections ago, we said that many types of vector spaces are very similar to $\mathbb{R}^n$. Now that we have defined isomorphic vector spaces, we have a concrete statement of what it means to be similar.

Theorem 2.71 (Finite dimensional real vector spaces $\simeq \mathbb{R}^n$)

Let V be a finite dimensional real vector space. Then $V \simeq \mathbb{R}^n$ for some $n \in \mathbb{N}$.

Proof. Since V is finite dimensional, let $\{v_1, ..., v_k\}$ be a basis for V. We just need to construct an isomorphism from $\mathbb{R}^n \to V$. What is an easy way to define this? We just need to define where each of the standard basis vectors $\vec{e}_j$ goes. This gives us a very natural choice to guess that the isomorphism $T : \mathbb{R}^k \to V$ defined by

$$T(\vec{e}_j) = v_j,$$

or equivalently

$$T(\vec{x}) = x_1 v_1 + x_2 v_2 + \ldots + x_k v_k$$

is bijective. This is indeed bijective, which you have confirmed in a previous exercise. □

Corollary 2.72: Let V be finite dimensional. Then every basis of V has the same number of vectors.

Proof. Let $\{v_1, \ldots, v_k\}$ and $w_1, \ldots, w_l$ be two bases for V. Then by the construction in the previous theorem, we have $V \simeq \mathbb{R}^k$ and $V \simeq \mathbb{R}^l$. So $\mathbb{R}^k \simeq \mathbb{R}^l$ and thus $k = l$. □

Definition 2.73 (Dimension of a vector space)

Let V be finite dimensional. The **dimension** of V is denoted $\dim(V)$, and is the number n for which $V \simeq \mathbb{R}^n$.

Understanding generalized vector spaces

We can use coordinates to understand finite dimensional vector spaces. Let V be n-dimensional, and pick a basis $\{v_1, \ldots, v_n\}$ for V. Then each element $w \in V$ can be written as a *unique linear combination* of the basis vectors, or

$$v = c_1 v_1 + c_2 v_2 + \ldots + c_n v_n = \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

Therefore, we can make sense of linear transformations between finite dimensional spaces as transformations $\mathbb{R}^n \to \mathbb{R}^m$. To make this concrete, let $\dim V = n$, $\dim\ W = m$, and pick bases $\{v_1, \ldots, v_n\}, \{w_1, \ldots, w_m\}$ for V and W respectively. Using our coordinate representation, for any transformation $T : V \to W$, we can pick a matrix representation of $\tilde{T} : \mathbb{R}^n \to \mathbb{R}^m$ such that the two paths from $\mathbb{R}^n \to \mathbb{R}^m$ lead to the same result

$$\begin{tikzcd} v & V \arrow[r, "T"] & W \arrow[d, "\text{coordinate representation of } W"] & w \\ \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} & \mathbb{R}^n \arrow[u, "\text{coordinate representation of } V"] \arrow[r, dotted, "\tilde{T}"] & \mathbb{R}^m & \begin{bmatrix} d_1 \\ \vdots \\ d_m \end{bmatrix} \end{tikzcd}$$

In particular, this correspondence of using coordinates is how we motivated the use of linear independence for computing the dimension of the span of vectors. If $v_1, \ldots, v_k$ are linearly independent, $\text{span}(v_1, \ldots, v_k)$ has a basis $\{v_1, \ldots, v_k\}$ and is isomorphic to $\mathbb{R}^k$ using coordinates.

Rank Nullity Theorem

Recall the Rank of a matrix is the number of pivots in its rref.

Proposition 2.74

Let $A \in M_{n\times m}(\mathbb{R})$. Then
$$\text{rank}(A) = \dim(\text{col}(A))$$

The intuition behind this statement is as follows: If you take all the columns that has a pivot in the rref, you will get a linearly independent set of vectors. This statement says that the span of these vectors is exactly $\text{col}(A)$, so forms a basis for the column space.

Proof. We need to show two things: First, if we isolate the pivot columns, we have a linearly independent set of vectors. Next, we have to show the span of these vectors equals the column space of A.

Let S be a sequence of row operations that reduce A to $\text{rref}(A)$.

For linear independence, we isolate the columns of A with pivot in $\text{rref}A$ to create a new matrix A' of $n \times \text{rank}(A)$. The same sequence S applied to A' will reduce it to row echelon form, with a pivot in every column. Therefore, the vectors are linearly independent.

To show they span the column space, we want to show for each $\vec{b}$ that $A\vec{x} = \vec{b}$ has a solution, $A'\vec{y} = \vec{b}$ has a solution in $\vec{y}$. To show this, we see that applying S to $\left[A \,\middle|\, \vec{b}\right]$ will give matrix in rref with no pivot in the final column. So applying S to $\left[A' \,\middle|\, \vec{b}\right]$ will also lead to a matrix in rref with no pivot in the final column. □

Similarly, we can define a term for the dimension of nullspace of a matrix.

Definition 2.75 (Nullity)

Let $A \in M_{n\times m}(\mathbb{R})$. We define the **nullity** of A
$$\text{Nullity}(A) = \dim(N(A))$$

Proposition 2.76

Nullity(A) is the number of columns in rref(A) without pivots.

We have a way to generate a basis for the nullspace, using the rows without a pivot, so this follows from the algorithm. Now we have rank corresponding to the columns with pivots and nullity corresponding to the columns without pivots. This means that for any $n \times m$ matrix, the sum of rank and nullity is always the number of columns m. This is the statement of the Rank-Nullity Theorem.

Theorem 2.77 (Rank-Nullity)

Let $A \in M_{n\times m}(\mathbb{R})$. Then

$$\mathrm{Rank}(A) + \mathrm{Nullity}(A) = m.$$

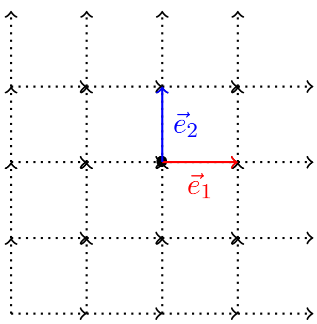
Phrased in terms of a linear transformation $T : U \to V$,

$$\dim(\mathrm{Im}(T)) + \dim(\ker(T)) = \dim(U).$$

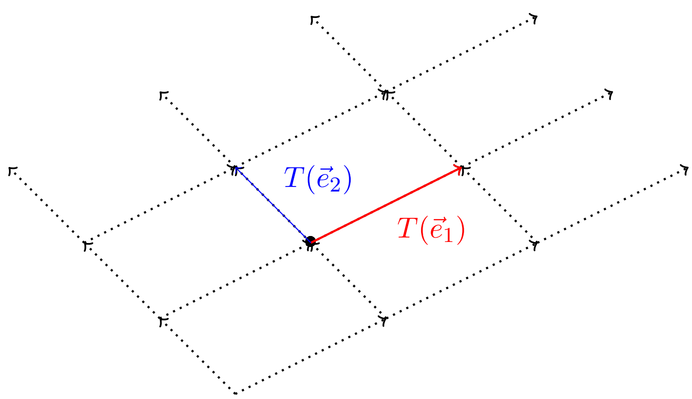
2.5.3 Determinants

We now move towards the last equivalence for That One Theorem. Recall in the intuition behind the Matrix Representation of Linear Transformations, we constructed a grid and showed that a linear transformation is a change in the coordinate system.

The original grid squares become grid 'parallelograms'. Indeed, if you extend to three dimensions,



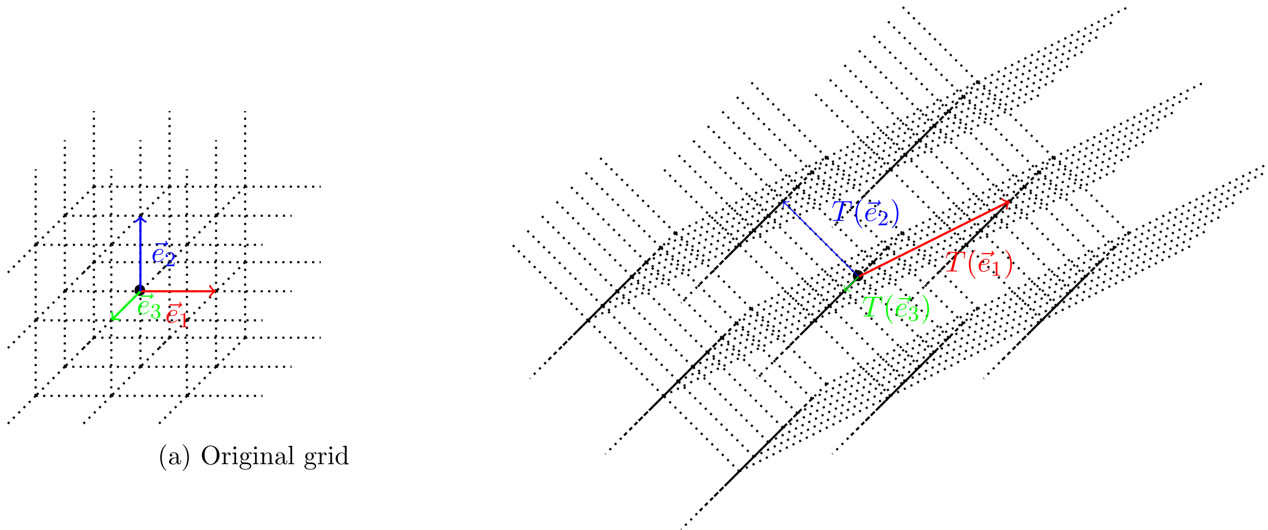
(a) Original grid



(b) Transformed grid

you can get grid 'parallelepipeds'. To say that the transformation is surjective means that the transformed unit square/cubes have non-zero area/volume and can thus tessellate.

Here we use the god-given formula for the determinant.



(a) Original grid

(b) Transformed grid

Figure 2.6: I tried my best, please use a bit of imagination

Definition 2.78 (Determinant)

We define a determinant function $\det : M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ as follows.

- $\det[a] = a$ for a 1×1 matrix $[a]$
- For higher dimensions $A = [a_{i,j}] \in M_{n\times n}(\mathbb{R})$, we recursively define

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{i,j} \det A_{i,j}$$

for any $1 \le j \le n$. Any value of j will give the same formula upon expansion.

Remark. *$A_{i,j}$ is the the* ***minor*** *of the matrix A where you delete the i-th row and j-th column. If*

$$A = \begin{bmatrix} a & b & * \\ * & * & * \\ c & d & * \end{bmatrix},$$

$$A_{2,3} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Remark. *The recursive formula also works for*

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{j,i} \det A_{j,i}$$

where instead of expanding along the j-th row, we expand along the j-th column.

Proposition 2.79

For small matrices, we have these simplified formulae.

1×1 $\det([a]) = a$.

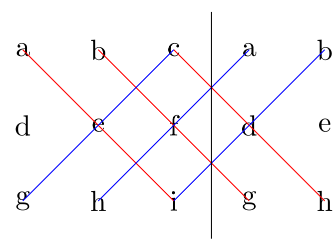
2×2 $\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$.

3×3 $\det\left(\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\right) = aei + bfg + cdh - ceg - afh - bdi$.

Remark. *The mnenomic for memorizing the 3×3 determinant is as follows: Extend the matrix as*

$$\left[\begin{array}{ccc|cc} a & b & c & a & b \\ d & e & f & d & e \\ g & h & i & g & h \end{array}\right]$$

and take the difference between the red diagonal products and the blue diagonal products.



As you have seen in a previous chapter, the formula for the determinant of a 3×3 matrix coincides with the volume of the parallelepiped spanned by the column vectors.

Example 2.80

Calculate the determinant of the rotational matrix

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

.

We just put this in the formula to get

$$\det R = \cos^2\theta - (-\sin^2\theta) = \cos^2\theta + \sin^2\theta = 1$$

regardles of the value of θ.

Example 2.81

Calculate the determinant of an upper-triangular matrix in the form

$$U = \begin{bmatrix} a_1 & * & * & \dots & * \\ 0 & a_2 & * & \dots & * \\ 0 & 0 & a_3 & \dots & * \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \dots & a_n \end{bmatrix} \in M_{n\times n}(\mathbb{R}).$$

Let's consider the special case $n = 3$. We do not know the values of $*$, but if we write out

$$\left[\begin{array}{ccc|cc} a_1 & * & * & a_1 & * \\ 0 & a_2 & * & 0 & a_2 \\ 0 & 0 & a_3 & 0 & 0 \end{array}\right],$$

the only diagonal that can possible have all non-zero entries gives the product abc. So the determinant of a triangular matrix is always the product of the diagonal terms. The determinant of an upper triangular matrix is always the product of the diagonal terms. To show this, we need to exploit the recursive formula for the determinant. If we expand along the first column, most of the terms in the sum are 0, except the first term that gives

$$(-1)^{1+1} a_1 \det\left(\begin{bmatrix} a_2 & * & \dots & * \\ 0 & a_3 & \dots & * \\ \vdots & & \ddots & \\ 0 & 0 & \dots & a_n \end{bmatrix}\right),$$

which is a_1 times the determinant of a smaller upper triangular matrix. If we know that this $(n-1)\times(n-1)$ matrix has determinant $a_2a_3\dots a_n$, this will prove the formula for a $n\times n$ matrix, then this result will prove the formula for a $(n+1)\times(n+1)$ matrix and so on! This technique is known as Mathematical Induction. If you want to prove a statement is true for all natural numbers n, you can show that it is true when $n = 1$, and show that if the statement is true for $n = k$, it is also true for $n = k + 1$.

Proof. We show that the determinant of U is $(a_1 \dots a_n)$ by induction on the value of n.

Base Case: The determinant of the matrix $[a_1]$ is a_1, which is evidently the product of all the diagonal terms.

Induction step: Suppose the determinant of a $k \times k$ upper triangular matrix is the product of the diagonal terms. We compute the determinant of a $(k+1) \times (k+1)$ matrix in the form

$$U = [u_{i,j}] = \begin{bmatrix} a_1 & * & * & \dots & * \\ 0 & a_2 & * & \dots & * \\ 0 & 0 & a_3 & \dots & * \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \dots & a_{k+1} \end{bmatrix}.$$

Applying the recursive formula and expanding along the first column, the only non-zero $u_{j,1}$ term is $u_{1,1} = a_1$, so

$$\begin{aligned} \det U &= (-1)^{1+1} a_1 \det \left(\begin{bmatrix} a_2 & * & \dots & * \\ 0 & a_3 & \dots & * \\ \vdots & & \ddots & \\ 0 & 0 & \dots & a_{k+1} \end{bmatrix} \right) \\ &= a_1 (a_2 a_3 \dots a_{k+1}) \end{aligned}$$

by applying the hypothesis on the $k \times k$ upper triangular matrix. Then the determinant of a $(k+1) \times (k+1)$ matrix is the product of the diagonal terms. By the Principle of Mathematical Induction, the determinant of an upper triangular matrix of any size is the product of its diagonal terms. □

Proposition 2.82

The elementary row operations change the determinant of a matrix in the following way:

- *(Row Swap)* The determinant is multiplied by -1.
- *(Scaling)* The determinant is multiplied by the scaling constant.
- *(Sum)* The determinant does not change.

The proof of this is not very enlightening, so we will skip it. It boils down to working with the general determinant formula and doing intense algebra manipulation.

However, we have the following corollary

Corollary 2.83: The determinant of a matrix can be computed as follows:

1. Reduce the matrix to rref (or row echelon form).
2. Take the product of the diagonal terms.
3. Multiply this product by $1/c$ for every time you scaled a row by c in computing the rref.
4. Multiply this product by -1 for every time you swaped two rows in computing the rref.
5. The final product is the determinant of the matrix.

Theorem 2.84 (This/That One (Determinants))

Let $A = [\vec{v}_1 \dots \vec{v}_n] \in M_{n\times n}(\mathbb{R})$, and $T : \mathbb{R}^n \to \mathbb{R}^n$ be the unique linear transformation satisfying $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$. Then the following are equivalent.

1. $\{\vec{v}_n, \dots, \vec{v}_n\}$ form a basis.
2. $\{\vec{v}_n, \dots, \vec{v}_n\}$ are linearly independent.
3. $\{\vec{v}_n, \dots, \vec{v}_n\}$ span $\mathbb{R}^n$.

   $\vdots$

21. $\det A \neq 0$.

Proof. Recall the motivation for introducing the determinant is the signed n-dimensional volume of the paralellotope formed by the column vectors. If (and only if) the volume is non-zero, they can tesselate the whole $\mathbb{R}^n$ space.

To show the equivalence of $\det A \neq 0$, we can use the condition from That One Theorem that $\text{rref}(A) = I_n$, and the previous corollary for computing determinants. If the determinant is 0, then the product of the diagonal terms of $\text{rref}A$ has a zero, so $\text{rref}(A) \neq I_n$. On the converse, if the determinant is non-zero, all the diagonal terms in $\text{rref}(A)$ are non-zero. The only way this can happen for a matrix in rref is when the matrix is also the identity matrix. □

Theorem 2.85 (Determinant of Product is Product of Determinant)

Let $A, B \in M_{n\times n}(\mathbb{R})$, then
$$\det(AB) = \det(A)\det(B)$$

Proof. First we consider the case where $\det(A) = 0$. By That One theorem, the column space of A is not $\mathbb{R}^n$. Let $\vec{v} \notin \text{col}(A)$. Then $\vec{v} \notin \text{col}(AB)$. So by That One Theorem again $\det(AB) = 0$.

Now we can let $\det(A) \neq 0$. By That One Theorem, we have a sequence of row opeations transforming $A \to I_n$. Denoting the row operations as matrices, we have $X_1, \dots, X_k \in M_{n\times n}(\mathbb{R})$ such that
$$X_k \dots X_1 A = I_n,$$
so
$$X_k \dots X_1 AB = I_n B = B,$$
or that the same row operations turn AB into B. Starting with $\det(B)$, we can reverse the row operations (and scaling according to the factors in corollary 2.83) to get $\det(AB) = \det(A)\det(B)$. □

Exercises

28. Compute the determinants of the following matrices: (you may find row reduction helpful)

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

(b) $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 1 & 4 & 2 & 3 \\ 4 & 3 & 2 & 1 \end{bmatrix}$

(c) $\begin{bmatrix} 0 & 0 & 0 & 0 & 3 \\ 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix}$

(d) $\begin{bmatrix} 0 & x & y \\ -x & 0 & z \\ -y & -z & 0 \end{bmatrix}$

29. For a 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, show that the matrix is invertible when $ad - bc$ is non-zero, and its inverse in this case is given by

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

30. Let $A = [a_{i,j}] \in M_{n \times n}(\mathbb{R})$. The **transpose** of A is denoted A^T and has entries $b_{i,j} = a_{j_i}$ i.e. you swap the rows and columns, as an example

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 3 \\ 0 & 0 & 3 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 2 & 0 \\ 3 & 3 & 3 \end{bmatrix}$$

Using the recursive formula of determinant, show that $\det(A)$ =$\det(A^T)$, thus A is invertible if and only if A^T is invertible.

As a wrapup, here is That One Theorem with the things we have proven so far. (There are a few more you will encounter in Spring Quarter.)

Theorem 2.86 (This/That One)

Let $A = [\vec{v}_1 \dots \vec{v}_n] \in M_{n\times n}(\mathbb{R})$, and $T : \mathbb{R}^n \to \mathbb{R}^n$ be the unique linear transformation satisfying $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$. Then the following are equivalent.

1. $\{\vec{v}_n, \dots, \vec{v}_n\}$ form a basis.
2. $\{\vec{v}_n, \dots, \vec{v}_n\}$ are linearly independent.
3. $\{\vec{v}_n, \dots, \vec{v}_n\}$ span $\mathbb{R}^n$.
4. The system of equations $A\vec{x} = \vec{b}$ has exactly one solution for each $\vec{b} \in \mathbb{R}^n$.
5. The system of equations $A\vec{x} = \vec{b}$ has at most one solution for each $\vec{b} \in \mathbb{R}^n$.
6. The system of equations $A\vec{x} = \vec{b}$ has at least one solution for each $\vec{b} \in \mathbb{R}^n$.
7. $\text{rref}(A) = I_n$.
8. $\text{rref}(A)$ has a pivot in every column.
9. $\text{rref}(A)$ has a pivot in every row.
10. $\text{Rank}(A) = n$.
11. $N(A) = \{\vec{0}\}$.
12. $\text{col}(A) = \mathbb{R}^n$.
13. T is bijective.
14. T is injective.
15. T is surjective.
16. $\ker(T) = \{\vec{0}\}$.
17. $\text{Im}(T) = \mathbb{R}^n$.
18. A is invertible.
19. A as a left inverse.
20. A has a right inverse.
21. $\det A \neq 0$.
22. $\det A^T \neq 0$.

2.6 End of Chapter Exercises

31. Let $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k \in \mathbb{R}^n$ be non-zero and pairwise orthogonal. That is $\vec{v}_i \cdot \vec{v}_j = 0$ if $i \neq j$ and > 0 if $i = j$.

    (a) Show that $\vec{v}_1, \ldots, \vec{v}_k$ are linearly independent. (Hint: what happens when you take the dot product of $c_1\vec{v}_1 + \ldots + c_k\vec{v}_k = \vec{0}$ with $\vec{v}_1$ on both sides?)

    (b) Show that $k \leq n$.

    (c) Now suppose $k = n$. Find an inverse for the matrix
    $$[\vec{v}_1 \ \ldots \ \vec{v}_n].$$

32. Recall the **transpose** of $A = [a_{i,j}] \in M_{n\times m}(\mathbb{R})$ is an $m \times n$ matrix denoted A^T and has entries $b_{i,j} = a_{j,i}$.

    (a) For vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$, we can treat $\vec{x}$ and $\vec{y}$ as $n \times 1$ column matrices. Show that $\vec{x} \cdot \vec{y} = \vec{x}^T\vec{y}$.

    (b) Show that for any $\vec{x} \in \mathbb{R}^m, \vec{y} \in \mathbb{R}^n$, $A\vec{x} \cdot \vec{y} = \vec{x} \cdot A^T\vec{y}$.

    (c) Show that for two matrices A, B, $(AB)^T = B^TA^T$, whenever the product AB is defined.

    (d) Conclude that if A is a square matrix and has an inverse, $(A^T)^{-1} = (A^{-1})^T$.

33. Let $A \in M_{n\times n}(\mathbb{R})$. Find a necessary and sufficient (if and only if) condition for $AA^T = 1$. (Hint: what is the dot product between two columns of A^T?)

34. Let $\vec{x} \in \mathbb{R}^3$, $A \in M_{3\times 3}(\mathbb{R})$. Suppose $A^2\vec{x} \neq \vec{0}$ and $A^3\vec{x} = \vec{0}$. Show that $\vec{x}, A\vec{x}, A^2\vec{x}$ span $\mathbb{R}^3$.

35. Let $A \in M_{n\times n}(\mathbb{R})$.

    (a) Let $\lambda \in \mathbb{R}$. Show that the solutions of the equation
    $$A\vec{v} = \lambda\vec{v}$$
    form a subspace of $\mathbb{R}^n$.

    (b) Show that this subspace is not the trivial subspace if and only if
    $$\det(A - \lambda I_n) = 0.$$

    (c) Conclude that if n is odd, there is some non-zero vector $\vec{v}$ such that $A\vec{v}$ is parallel to $\vec{v}$ (or $A\vec{v} = \vec{0}$). You may use the fact that all odd-degree polynomials $p(x)$ have a real root, which you can prove using the intermediate value theorem considering the values of $\lim_{x\to\infty} p(x)$ and $\lim_{x\to-\infty} p(x)$.

    (d) Suppose $\lambda_1, \ldots, \lambda_k$ are all distinct and $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^n$ non-zero vectors such that $A\vec{v}_j = \lambda_j\vec{v}_j$ for all j. Show that $\vec{v}_1, \ldots, \vec{v}_k$ are linearly independent, thus $k \leq n$. You might it helpful to consider $k = 1$ then apply induction.

    (e) Consider the sum
    $$S = \sum_{\lambda\in\mathbb{R}} \text{Nullity}(A - \lambda I_n).$$
    The previous part shows that at most n values of λ have a non-zero nullity, so we can make sense of the infinite sum as the sum over the finite number of values of λ for which the nullspace is non-trivial. Show that $S \leq n$.

36. Let a real vector space $V = \mathrm{span}(v_1, \ldots, v_n)$. Show that every set of $m > n$ vectors in V are linearly dependent. In other words, let $m > n$, and $w_1, \ldots, w_m \in V$. Show that the w_j's are linearly dependent.

37. Let V a finite dimensional vector space. Let $U, W \subseteq V$ be two subspaces of V.
    (a) Show that the intersection of U and $W \stackrel{\text{def}}{=} U \cap W$ is a subspace of V. Is the union of U and $W \stackrel{\text{def}}{=} U \cup W$ necessarily a subspace of V?
    (b) Show that the set
    $$U + W \stackrel{\text{def}}{=} \{u + v | u \in U, w \in W\}$$
    is a subspace of V.
    (c) Let $U \cap W = \{0_V\}$. Show that $\dim(U + W) = \dim(U) + \dim(W)$.
    (d) In general, show that $\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W)$. (Approach 1: Start with a basis for $U \cap W$, and extend this to a basis for U and a basis for W. Show that these vectors are linearly indpendent and span $U \cup W$) (Approach 2: Construct a surjective linear transformation from something $(\dim(U) + \dim(W))$-dimensional to $U + W$, with the kernel isomorphic to $U \cap W$. Apply a theorem that is not That One Theorem).

38. Let $P : V \to V$ be a linear transformation that satisfies $P(P(v)) = P(v)$ for every $v \in V$. Show that every $v \in V$ can be decomposed into a *unique* combination $v = v_1 + v_2$, where $v_1 \in \mathrm{Im}(P)$ and $v_2 \in \ker(P)$.

39. *(Rank Nullity abstract proof)* Let U, V be vector spaces, and U is finite dimensional. Let $T : U \to V$ be a linear transformation.
    (a) Show that $\mathrm{Im}(T)$ is finite dimensional, so we can restrict the codomain of T (through a redefinition) $T : U \to \tilde{V} = \mathrm{Im}(T)$.
    (b) Show that this $T : U \to \mathrm{Im}(T)$ is surjective.
    (c) Pick a basis for $\tilde{V}$ to be $\{\tilde{v}_1, \ldots, \tilde{v}_k\}$. For each $\tilde{v}_j$, there is some $u_j \in U$ such that $T(u_j) = \tilde{v}_j$. Show that $\{u_1, ..., u_k\}$ are linearly independent.
    (d) Show that for every $u \in U$, we can write $u = w_1 + w_2$, where $w_1 \in \mathrm{span}(u_1, \ldots, u_k)$ and $w_2 \in \ker(T)$. (Hint: what is $T(w_1 + w_2)$?)
    (e) Because $\ker(T)$ is a subspace of U, it is finite dimensional. Choose a basis $\{\tilde{u}_1, \ldots, \tilde{u}_l\}$ for $\ker(T)$. Show that $\{u_1, ..., u_k, \tilde{u}_1, \ldots, \tilde{u}_l\}$ are linearly independent, and conclude that it is a basis for U.
    (f) Finish the proof that $k + l = \dim(U)$, and why this proves the Rank-Nullity Theorem.

40. *(Adapted from HKDSE M2 2019 Q11)* Recall $M_{2\times 2}(\mathbb{R})$ is a vector space with zero vector $0_{2\times 2}$. Let $M = \begin{bmatrix} 2 & 7 \\ -1 & -6 \end{bmatrix}, I = I_{2\times 2}$.
    (a) Show that M and I are linearly independent.
    (b) Find real numbers a and b such that $M^2 = aM + bI$.
    (c) Let $V = \mathrm{span}(M, I)$. Show the map $T(A) = MA$ for all $A \in V$ defines a linear transformation $V \to V$. Find the matrix representation B for this transformation using the coordinates in the basis $\{M, I\}$.

(d) Compute $B\begin{bmatrix}1\\5\end{bmatrix}$ and $B\begin{bmatrix}1\\-1\end{bmatrix}$. Hence express $M^n = T(T(\ldots T(I)))$ in terms of M and I.

(e) Repeat (a-d) on the matrix M^{-1} to obtain a closed form expression for $M^{-n} = (M^{-1})^n$. (Hint: For (d), consider the vectors (5,1) and (1,-1))

(f) Calculate $(M^n)^{-1}$ using the expression you got from (d) and verify it evaluates to the expression in (e).

Chapter 3

Functions of Several Variables

Notation. *Here we will sometimes drop the arrow to denote vectors and coordinates in* $\mathbb{R}^n$*. Moreover, consistent with our vector notation, if the output of a function* f *is a vector in* $\mathbb{R}^n$*, then we will denote* $f_1, \ldots, f_n$ *as the corresponding entries (as functions) of the output.*

We now relax the conditions that functions have to be linear. Moreover, we can restrict the domain of the function to be some subset $\Omega \subseteq \mathbb{R}^n$, as some functions are not everywhere defined on $\mathbb{R}^m$.

Example 3.1

Find the largest domain $\Omega \subseteq \mathbb{R}^3$ such that the function

$$f(x, y, z) = \frac{1}{x^2 + y^2 + z^2 - 9}$$

is defined.

The right hand side is well defined for any $x^2 + y^2 + z^2 - 9 \neq 0$. That is, you can calculate the right hand side as long as you are not dividing by zero. This means we can take the largest domain to be $\Omega = \mathbb{R}^3 - \{(x, y, z)|x^2 + y^2 + z^3 = 9\}$. What is this mysterious region described by $x^2 + y^2 + z^2 = 9$? Recall that $\sqrt{x^2 + y^2 + z^2}$ is distance of (x, y, z) from the origin, so this traces out a sphere of radius 3.

3.1 Graphing in $\mathbb{R}^n$

Graphing has always been a useful tool for to gain intuition about the 'shape' of functions. Here we formalize the definition of the graph of a function.

Definition 3.2 (Graph)

Let $\Omega \subseteq \mathbb{R}^n$ be open, and $f : \Omega \to \mathbb{R}$. We denote the **Graph** of f to be Γ_f, and is the subset of $\mathbb{R}^{n+1}$

$$\Gamma_f \stackrel{\text{def}}{=} \{(\vec{x}, f(\vec{x})) \mid \vec{x} \in \Omega\}.$$

Intuitively, this means we take Ω and lift that set in the $(n + 1)$-st dimension according to the value of $f(\vec{x})$ evaluated at each point.

Example 3.3

Let $f(x, y) = (49 + 3x - 2y)/6$, then the graph of f is a plane in 3D space, which we have sketched in one example in the first chapter.

This is not too hard to understand for one or two variables, but gets out of hand quickly as n increases. I sometimes struggle to visualize even in the third dimension. One way to get around this is to look at the cross sections of a graph.

Definition 3.4 (Level set)

Let $\Omega \subseteq \mathbb{R}^n$ be open, and $f : \Omega \to \mathbb{R}$. Let $c \in \mathbb{R}$. The **level set** of f at c is denoted L_c, and is the set

$$L_c \stackrel{\text{def}}{=} \{\vec{x} \in \Omega \mid f(\vec{x}) = c\}.$$

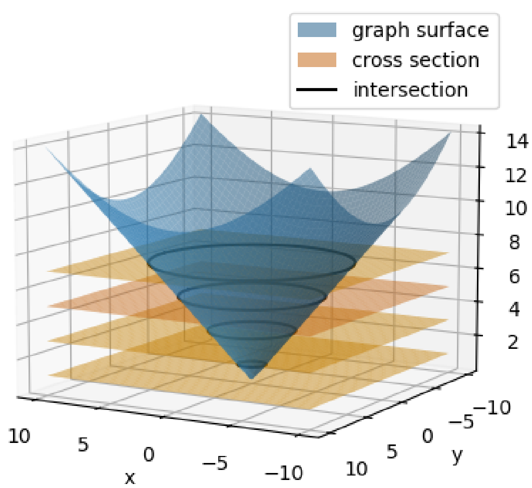
Why is this a natural definition? Consider the the intersection between Γ_f and the hyperplane described by the set of $(\vec{x}, c)$ for all $\vec{x} \in \mathbb{R}^n$. This plane will intersect all the points in L_c, except they are lifted in the $(n+1)$-st dimension by c units. That is, if you understand how all the L_c's behave for all values of c, you understand how the graph of L_c behaves.

Example 3.5

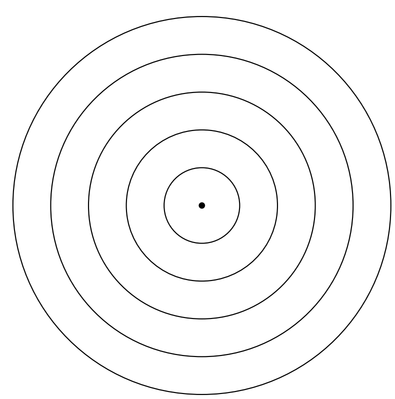
What are the level sets of the function $f(x, y) = \sqrt{x^2 + y^2}$?

The function here is not complicated at all: This is the distance (x, y) from the origin, meaning the graph of f is a cone. If we take cross section slices of positive heights, we will get circles. For negative heights, we get an empty set. For $c = 0$, L_c only contains the origin.

We can compress these level sets back to 2 dimensions.



(a) Cross sections of the cone



(b) The level sets $c = 0$ to $c = 5$, evenly spaced

.

Let us see how our flatlander friend Frankie (apparently that is what the inhabitants of $\mathbb{R}^2$ call themselves) can visualize this cone from the 2D level sets. Frankie holds a meter that constantly evaluates $f(x,y)$ at any point. He realizes that if he circles the origin at a radius of r, the meter will stay constant at a value of r. If he walks directly away from the origin, the meter will increase in value. Moreover, because the level sets $L_1, L_2, L_3, L_4, ...$ are evenly spaced, Frankie only needs to walk a the same distance away from the origin to raise the meter value the same amount. Translated back to 3D, Frankie is actually 'hiking' up the cone as he walks away from the origin. The slope (evaluated from the radial cross section) is 1, so Frankie walks k units away from the origin to raise his elevation by k, no matter where he is.

Example 3.6

What are the level sets of the function $f(x,y,z) = x^2 + y^2 + z^2$?

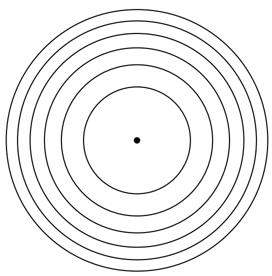


Figure 3.2: Take the 2D analog $g(x,y) = x^2 + y^2$. The level sets of $c = 0, 1, 2, \ldots$ are not evenly spaced.

Now we play the role of Frankie.

First we try to see what the level sets $f(x,y,z) = c$ look like. This is a sphere of radius $\sqrt(c)$. The level sets are also spheres, but this time they are spaced a little bit different. To go from the level set L_c to L_{c+1}, we go from the sphere of radius $\sqrt{c}$ to $\sqrt{c+1}$, and the shortest distance is

$$\begin{aligned}\sqrt{c+1} - \sqrt{c} &= \frac{(\sqrt{c+1} - \sqrt{c})(\sqrt{c+1} + \sqrt{c})}{(\sqrt{c+1} + \sqrt{c})} \\ &= \frac{1}{\sqrt{c+1} + \sqrt{c}}\end{aligned}$$

which gets smaller as c increases! This means we need to walk a shorter distance to increase our meter. Equivalently, the 'slope' in this direction increases as we get further away.

We could also hold z constant at 0 to take one more cross section $(x, y, 0, c)$ in x, y. The 'level sets' of the function $g(x,y) = f(x,y,0)$ look as follows:

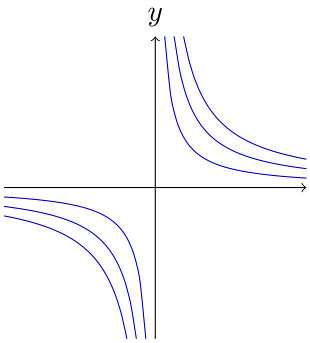
Example 3.7

Sketch the function $f(x,y) = xy$, either in 3D or through its 2D level sets.

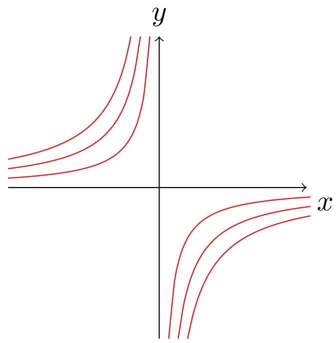
Let us begin by considering the level sets $xy = c$. This describes the graph $y = c/x$, except for the case $c = 0$, then the level set is the x and y axes.

There are many ways to visualize the actual surface from here: My goto way is to vary c and see how the level set changes. In this case, starting from positive c and decreasing, the curves get closer and closer to the x and y axis, and they 'flip' across the axis as the sign of c changes from positive to negative.

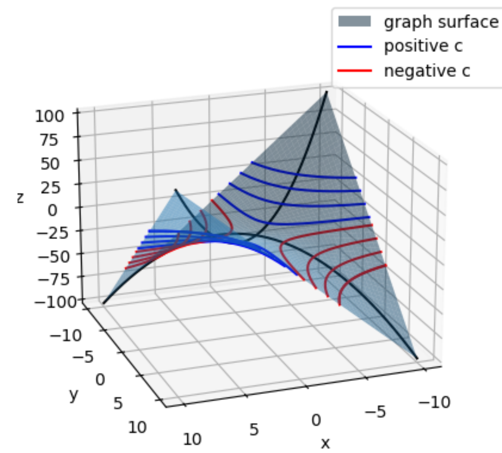
The graph looks something like a saddle, or a pringle chip. Saddle points are somewhere between a local maximum and a local minimum. The saddle point of this function is at the origin: going the direction of $x = y$, you will get $f(x,y) = x^2$ that curves upwards; going in the direction of $x = -y$, you get $f(x,y) = -x^2$ which curves downwards. This point is *in some sense* a local minimum and a local maximum depending on which way you look at it.



(a) Positive c



(b) Negative c

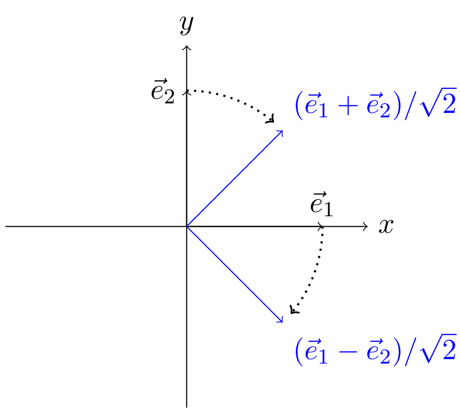


(c) A saddle

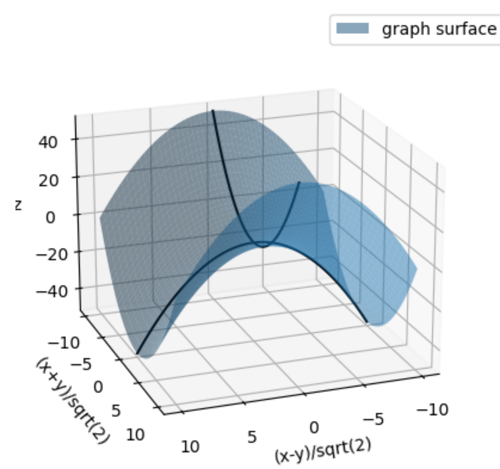
Indeed, we can rewrite

$$xy = \frac{1}{2}\left[\left(\frac{1}{\sqrt{2}}(x+y)\right)^2 - \left(\frac{1}{\sqrt{2}}(x-y)\right)^2\right],$$

so we can rotate our basis vectors $\vec{e}_1, \vec{e}_2 \in \mathbb{R}^2$ to get $(\vec{e}_1 - \vec{e}_2)/\sqrt{2}, (\vec{e}_1 + \vec{e}_2)/\sqrt{2}$. In this new basis, we can plot the rotated saddle as $(y^2 - x^2)/2$.



(a) Positive c



(b) The rotated saddle

Example 3.8

Let $a, b, c > 0$, and $k, m, l \in \mathbb{R}$. What does the surface described by

$$\frac{x^2 - 2kx}{a^2} + \frac{y^2 - 2my}{b^2} + \frac{z^2 - 2lz}{c^2} = 1$$

look like?

Let us approach a simplier version of the problem first: what is the surface described by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1?$$

We know what happens when $a = b = c = 1$. Namely, we get the unit sphere. Now if we stretch the x axis by a factor of a, we will change from a surface

$$x^2 + y^2 + z^2 = 1 \to \frac{x^2}{a^2} + y^2 + z^2 = 1!$$

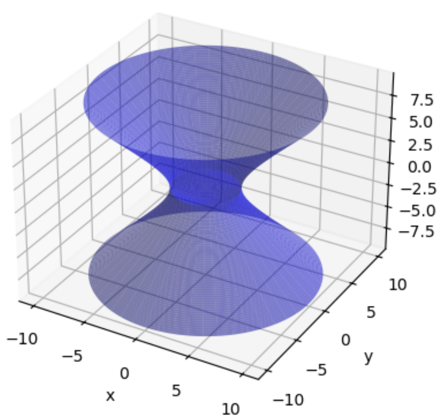
Because the x,y,z axes are orthgonal, we can also stretch the y and z axes by a factor b and c respectively to give the surface. This will give us an ellipsoid, which is just a stretched (or compressed) sphere. To deal with the original equation, we can complete the square to get

$$\begin{aligned}\frac{x^2 - kx}{a^2} + \frac{y^2 - my}{b^2} + \frac{z^2 - lz}{c^2} &= \frac{(x-k)^2 - k^2}{a^2} + \frac{(y-m)^2 - m^2}{b^2} + \frac{(z-l)^2 - l^2}{c^2} \\ &= \frac{(x-k)^2}{a^2} + \frac{(y-m)^2}{b^2} + \frac{(z-l)^2}{c^2} - \frac{k^2}{a^2} - \frac{m^2}{b^2} - \frac{l^2}{c^2},\end{aligned}$$

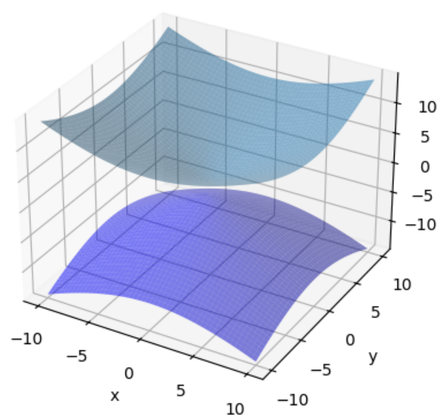
Which means we have an ellipsoid but now translated (k, m, l) units! In the later exercises, you will classify the surfaces described by

$$x^2 + y^2 - z^2 = 1, \; x^2 - y^2 - z^2 = 1,$$

which are called one and two sheeted hyperboloids respectively. In later chapters we will see a way



(a) Hyperboloid of one sheet



(b) hyperboloid of two sheets

to change our coordinate system, through some rotation, scaling and translation, to simplify any surface in the form

$$ax^2 + by^2 + cz^2 + dxy + eyz + fzx + gx + hy + iz = j$$

to look like something of the form

$$\frac{x^2}{a^2} \pm \frac{y^2}{b^2} \pm \frac{z^2}{c^2}$$

Exercises

1. Consider the level set $x^2 + y^2 - z^2 = 1$.
   (a) We take the cross section of this surface with the plane $z = k$. What does this cross section look like in x, y?
   (b) We now hold $y = 0$. What does this cross section look like?
   (c) Combine the first two parts, and try to sketch the surface in 3D. How many sheets or 'connected regions' does this hyperboloid have?

2. Consider the level set $x^2 - y^2 - z^2 = 1$.
   (a) We take the cross section of this surface with the plane $x = k$. What does this cross section look like in y, z? How does this part differ from the last problem?
   (b) We now hold $y = 0$. What does this cross section look like?
   (c) Combine the first two parts, and try to sketch the surface in 3D. How many sheets or 'connected regions' does this hyperboloid have?

3. What the largest possible domain for the function
   (a) $f(x, y) = e^{x^2 - y^2}$
   (b) $f(x, y) = \ln\big(y^2 - x^2 - 2\big)$
   (c) $f(x, y) = (x^2 - y^2)/(x - y)$
   (d) $f(x, y, z) = 1/(xyz)$
   (e) $f(x, y, z) = (z^2 - x^2 - y^2)^{-1/2}$.

4. Describe the graph of the function
   (a) $f(x, y) = 5$
   (b) $f(x, y) = 2x - y$
   (c) $f(x, y) = 1 - x^2 - y^2$
   (d) $f(x, y, z) = 4 - \sqrt{x^2 + y^2}$
   (e) $f(x, y, z) = \sqrt{24 - 4x^2 - 6y^2}$.

5. Describe the level sets of the function
   (a) $f(x, y) = x + y$
   (b) $f(x, y) = x^2 + 9y^2$
   (c) $f(x, y) = x - y^2$
   (d) $f(x, y) = x - y^3$
   (e) $f(x, y) = x^2 + 4x + y^2 + 2y + 9$
   (f) $f(x, y, z) = x^2 + y^2 - z$
   (g) $f(x, y, z) = x^2 + 2x + y^2 - 2y + z^2 + 4z$
   (h) $f(x, y, z) = z^2 - x^2 - y^2$

6. Describe the surfaces in $\mathbb{R}^3$ described by $f(x, y, z) =$

(a) $x^2 + y^2 = 16$

(b) $z^2 = 49x^2 + y^2$

(c) $z = 25 - x^2 - y^2$

(d) $x = 4y^2 - z^2$

(e) $4x^2 + y^2 + 9z^2 = 36$

(f) $4x^2 - y^2 + 9z^2 = 36$

(g) $4x^2 - y^2 - 9z^2 = 36$

7. Let $c \in \mathbb{R}$. Describe the intersection of the graph of $f(x, y) = x^2 + 9y^2$ in the given planes $x = c$, $y = c$, $z = c$.

8. Consider the (double) cone C in $\mathbb{R}^3$ described by the equation $x^2 + y^2 = z^2$:

   (a) Describe the intersection of the C with the planes $z = c$ and $x = c$. Your answer may depend on the sign of $c \in \mathbb{R}$.

   (b) In general, what is the intersection of C with the plane $ax + by + cz = d$?

3.2 Topology of $\mathbb{R}^n$

Remark (Author's notes)**.** *This part is slightly touched on in Even's textbook (but scattered across the differentiation chapter). Some past 281 instructors such as John Enns dedicate half a week on topology.*

Definition 3.9 (Open Ball)

Let $\vec{x}_0 \in \mathbb{R}^n, r > 0$. We write the **open ball of radius** r **centered at** a as $B(\vec{x}_0, r)$, and is the set of all points within r distance away from a, i.e.

$$B(\vec{x}_0, r) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid |\vec{x} - \vec{x}_o| < r\}.$$

Many statements that we will work towards are 'local', in the sense that they only dependent on the behavior of a function in a small open ball centered at a point. For instance, if a function $f : \mathbb{R} \to \mathbb{R}$ is continuous at $x = 0$, altering the value of the function at any $y \neq 0$ will not change the continuity of the function at $x = 0$. This is because the function remains unchanged in the small open ball $B(0, |y|/2)$. Similar, the differentiability (and its derivative) also only depends on the local behavior of the function.

Thus, we define a set with a special property that each point in the set is 'fully contained' inside the set.

Definition 3.10 (Open sets)

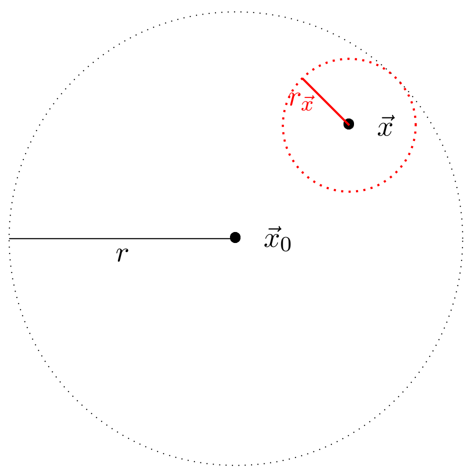
Let $\Omega \subseteq \mathbb{R}^n$. We say that Ω is **open** if for every $\vec{x} \in \Omega$, there is some radius $r_{\vec{x}} > 0$ such that $B(\vec{x}, r_{\vec{x}}) \subseteq \Omega$.

Remark. *We want to deal with open sets when we talk about the local behavior of a function. It is not useful to talk about the continuity of a function that is only defined on the integers, this set is not open because every ball centered at an integer will contain a non-integer.*

Proposition 3.11

An open ball is open.

Yeah, this needs a proof. To show this is open, let $\vec{x} \in B(\vec{x}_0, r)$. Our goal is construct a value $r_{\vec{x}} > 0$ such that $B(\vec{x}, r_{\vec{x}}) \subseteq B(\vec{x}_0, r)$.



As long as the distance between $\vec{x}$ and $\vec{x}_0$ is strictly less than r, the points within a small vicinity of $\vec{x}$ all have distance $< r$ to $\vec{x}_0$.

This can be made rigorous using the Triangle Inequality.

Proof. Let $\vec{x} \in B(\vec{x}_0, r)$. Let $r_{\vec{x}} = r - |\vec{x}_0 - \vec{x}|$. Then for any $\vec{y} \in B(\vec{x}, r_{\vec{x}})$, we have

$$\begin{aligned} |\vec{y} - \vec{x}_0| &= |\vec{y} - \vec{x} + \vec{x} - \vec{x}_0| \\ &\leq |\vec{y} - \vec{x}| + |\vec{x} - \vec{x}_0| \\ &< r - |\vec{x}_0 - \vec{x}| + |\vec{x}_0 - \vec{x}| \\ &= r, \end{aligned}$$

so $\vec{y} \in B(\vec{x}_0, r) \implies B(\vec{x}, r_{\vec{x}}) \subseteq B(\vec{x}_0, r)$. $\square$

Example 3.12

The following sets are open:

1. The top-half plane in $\mathbb{R}^2$. $\mathbb{H} \stackrel{\text{def}}{=} \{(x, y) \mid y > 0\}$. Sketch: for every $(x, y) \in \mathbb{H}$ take $r_{(x,y)} = y$.

2. The punctured ball $B^*(\vec{x}, r) \stackrel{\text{def}}{=} B(\vec{x}, r) - \{\vec{x}\}$. Sketch: for every $\vec{y} \in B^*(\vec{x}, r)$ take $r_{\vec{y}} = \min(|\vec{y} - \vec{x}|, r - |\vec{y} - \vec{x}|)$.

3. The whole space $\mathbb{R}^n$ is open. Sketch: take $r_{\vec{x}} = 1$ for any point.

4. The union of (possibly infinite) open sets $\bigcup_a G_a$, where each G_a is open. Sketch: if $\vec{x} \in \bigcup_a G_a$, then $\vec{x} \in G_a$ for some a. Because G_a is open, pick a radius such that the ball is contained in G_a.

The following sets are not open:

1. The set contianing just one point $\{\vec{x}\}$. Sketch: every ball centered at $\vec{x}$ contains points outside of $\vec{x}$.

2. The set of points $\{\vec{x} \in \mathbb{R}^n \mid |\vec{x}| \leq r\}$. Sketch: Any ball of radius $\epsilon > 0$ centered around the point $(r, 0, 0, \ldots, 0)$ contains $(r + \epsilon, 0, 0, \ldots, 0)$.

We see that a set fails to be open if there is a point in the boundary of the set. I.e. no matter how small $\epsilon > 0$ is, $B(\vec{x}, \epsilon)$ extends beyond that set.

Definition 3.13 (Boundary)

Let $\Omega \subseteq \mathbb{R}^n$. We denote the **boundary** of Ω as $\partial\Omega$. This is the set of points in $\vec{x} \in \mathbb{R}^n$ such that every open ball around $\vec{x}$ contains a point in Ω and a point not in Ω.

Colloquially, $\vec{x}$ is very (arbitrarily) close to Ω, but is also very (arbitrarily) close to Ω^c. So it is on the boundary of the two sets.

Example 3.14

In $\mathbb{R}^3$, $\partial B(\vec{0}, r)$ is the sphere of radius r. Note that a sphere refers to only the surface while a ball consists of the whole interior.

Definition 3.15 (Closure of a set and Closed sets)

Let $\Omega \subseteq \mathbb{R}^n$. We denote the **closure of** Ω as $\bar{\Omega} \stackrel{\text{def}}{=} \Omega \cup \partial\Omega$. If $\bar{\Omega} = \Omega$, we say Ω is **closed**. Equivalently, a set is closed if it contains its boundary.

Proposition 3.16

The closure of a set is closed.

What this means is that if we add in the boundary of a set, we do not create new 'boundaries'.

Proof. Let $\Omega \subseteq \mathbb{R}^n$. Let $\vec{x} \in \partial\bar{\Omega}$. Consider open ball of radius $\epsilon > 0$ centered around $\vec{x}$. We want to show that this ball contains a point in Ω and a point in $\partial\Omega$, thus $\vec{x} \in \partial\Omega$ so is already included in $\bar{\Omega}$.

This ball, by the definition of boundary of $\bar{\Omega}$, contains a point $\vec{a} \in \bar{\Omega}$ and a point in $\vec{b} \in \bar{\Omega}^c$. We first consider $\vec{b}$. Because $\vec{b} \in \bar{\Omega}^c \subseteq (\Omega \cup \partial\Omega)^c = \Omega^c \cap (\partial\Omega)^c \subseteq \Omega^c$. This point is in Ω^c too. If $\vec{a} \in \Omega$, we are done. Else, $\vec{a} \in \partial\Omega$. As $B(\vec{x}, \epsilon)$ is open, we can find some open ball $B(\vec{a}, \delta) \subseteq B(\vec{x}, \epsilon)$. Now we can reapply the definition for $\vec{a} \in \partial\Omega$ to get that there is some point $\vec{a}' \in B(\vec{a}, \delta)$ that is in Ω. So we have some $\vec{a}' \in B(\vec{x}, \epsilon)$ that is contained in Ω.

Since ϵ was arbitrary, we can conlude that every open ball centered around $\vec{x}$ contains a point in Ω and Ω^c. So $\vec{x} \in \partial\Omega \subseteq \bar{\Omega}$. $\square$

Proposition 3.17 (Sets are not doors)

Sets are not doors. A set can be both open and closed, or neither.

Proof. The empty set is both open and closed. In $\mathbb{R}^n$, $\mathbb{R}^n$ is both open and closed. The set $1 < |x| \leq 2$ is neither open nor closed in $\mathbb{R}$, as it only contains part of its boundary. $\square$

Exercises

9. **TODO:**

3.3 Limits and Continuity

Definition 3.18 (Limit Point)

Let $\Omega \subseteq \mathbb{R}^n$. A **limit point of** Ω is a point $\vec{x} \in \mathbb{R}^n$ such that every open ball centered around $\vec{x}$ contains a point $\vec{y} \in \Omega$ with $\vec{y} \neq \vec{x}$. Alternatively, using the notation of punctured balls $B^*(\vec{x}, r) \overset{\text{def}}{=} B(\vec{x}, r) - \{\vec{x}\}$, the intersection

$$B^*(\vec{x}, r) \cap \Omega$$

is non-empty for every $r > 0$.

Colloquially, you can approach (get closer and closer to) a limit point by using only points in Ω.

Remark. *When we talk about limits, we care about the behavior around some point $\vec{x}$ but not at the point $\vec{x}$ (For all we care about, $\vec{x}$ could be outside the domain of the function). This is why the condition $\vec{y} \neq \vec{x}$ is in the definition of a limit point.*

Definition 3.19 (Limit)

Let $\Omega \subseteq \mathbb{R}^n$ be open, and $f : \Omega \to \mathbb{R}^m$. Let $\vec{a}$ be a limit point of Ω, $\vec{b} \in \mathbb{R}^m$. We denote **the limit of** $f(\vec{x})$ **as** $\vec{x}$ **tends to** $\vec{a}$ as $\lim_{\vec{x} \to \vec{a}}$, and say that this limit equals $\vec{b}$ if for all $\epsilon > 0$, there is $\delta > 0$ such that

$$|f(\vec{x}) - \vec{b}| < \epsilon \text{ if } \vec{x} \in B^*(\vec{x}, \delta) \cap \Omega,$$

or equivalently,

$$|f(\vec{x}) - \vec{b}| < \epsilon \text{ if } 0 < |\vec{x} - \vec{a}| < \delta$$

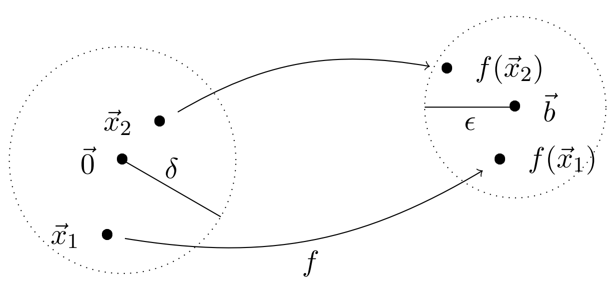
where f is defined.

Parsing the $\epsilon - \delta$ definition of limit

For an illustration. Let $\Omega = \mathbb{R}^2$ and $f : \mathbb{R}^2 \to \mathbb{R}^2$. We want to consider the limit

$$\lim_{\vec{x} \to \vec{0}} f(\vec{x}) = \vec{b}.$$

We draw the function as taking in values from the input space $\mathbb{R}^2$ and spitting them them out in the output space $\mathbb{R}^2$.



We do not know what the value $f(\vec{0})$ is - it may not exist. However, we can see if the surrounding values of f approaches some value. To make this rigorous, let's say we want to get within ϵ distance of the value $\vec{b}$. This corresponds to the open ball on the right. Now we want to pick δ such that at a very short distance δ away from $\vec{0}$, the image of $B^*(\vec{0}, \delta)$ lands within $B(\vec{b}, \epsilon)$. This corresponds to the left open ball.

That means, you can get as close as you like to $\vec{b}$ as long as you are close to $\vec{0}$. If the ϵ requirement for closeness changes from 0.1 to 0.0000001, you can just pick a smaller δ to get within this distance. Remember: the order of the limit is someone first challenges that you need to get close to the limit in the output space (ϵ), then you pick the ball around the input space (δ) that satisfies that particular value of ϵ.

Proposition 3.20

Suppose $\lim_{\vec{x} \to \vec{a}} f(\vec{x})$ exists, then it must be unique.

The idea is this: suppose you can get arbitrarily close to $\vec{b}$ and arbitrarily close to $\vec{b}'$. Then $\vec{b}$ and $\vec{b}'$ must also be arbitrarily close. But when you have two values that are arbitrarily close, then they must be the same value!

Proof. Suppose, for the sake of contradiction, we have the two limits $\vec{b} \neq \vec{b}'$ of f as $\vec{x} \to \vec{a}$. Let $\epsilon = |\vec{b} - \vec{b}'|$. Pick δ_1, δ_2 such that $|f(\vec{x}) - \vec{b}| < \epsilon$ and $|f(\vec{x} - \vec{b}'| < \epsilon|$ when $0 < |\vec{x} - \vec{a}| < \delta_1$ and $0 < |\vec{x} - \vec{a}| < \delta_2$ respectively. Then pick some $\vec{x}$ such that $0 < |\vec{x} - \vec{a}| < \min(\delta_1, \delta_2)$. For this particular value of $\vec{x}$,

$$\epsilon = |\vec{b}' - \vec{b}| = |\vec{b}' - f(\vec{x}) + f(\vec{x}) - \vec{b}| \leq |\vec{b}' - f(\vec{x})| + |\vec{b} - f(\vec{x})| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

This gives us the strict inequality $\epsilon < \epsilon$, which is impossible. □

Remark. *A consequence of this is that if a limit exists, the limit will be the same using any path of approach. Caution: It is not enough to check a few paths to confirm a limit exists. Take the function*

$$f(x, y) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

Approaching $f(0,0)$ along any straight line $y = cx$ will give a limit of 0. However, the limit does not exist, as approaching on the y-axis $x = 0$ will give a limit of 1.

Example 3.21

Compute the limit (if it exists) of

$$\lim_{(x,y)\to(0,0)} \frac{xy}{\sqrt{x^2 + y^2}}$$

Let's assume that this limit exists, what must it be? We can test the limit along the x-axis and y-axis. I.e. We set x or $y = 0$ and let the the other variable approach 0. If the limit exists, the limit should be the same no matter how which axis we approach the limit by. Then we see

$$(x = 0) \lim_{y \to 0} \frac{0y}{\sqrt{0 + y^2}} = 0$$
$$(y = 0) \lim_{x \to 0} \frac{0x}{\sqrt{x^2 + 0}} = 0$$

So the limit, if it exists, must be 0. Can we show that this exists? Let $\epsilon > 0$, then $\delta =$ [placeholder]. We want

$$\left| \frac{xy}{\sqrt{x^2+y^2}} - 0 \right| < \epsilon$$

whenever $\sqrt{x^2+y^2} < \delta$. Intuitively, this should be true as the degree of the numerator is 2 and the degree of the denominator is '1'. So the numerator shrinks to 0 much faster than the denominator. To make this rigorous, we can use the inequality $|x| = \sqrt{x^2} \leq \sqrt{x^2+y^2}$ and similarly, $|y| \leq \sqrt{x^2+y^2}$ to get

$$\left| \frac{xy}{\sqrt{x^2+y^2}} - 0 \right| \leq \left| \frac{x^2+y^2}{\sqrt{x^2+y^2}} \right| = \left| \sqrt{x^2+y^2} \right| < \delta.$$

So setting $\delta = \epsilon$ works! Now we just replace the placeholder with ϵ and pretend we knew this all along.

Proof. We claim the limit is 0. Let $\epsilon > 0$, then for all $|(x,y)| < \epsilon$, we have

$$\left| \frac{xy}{\sqrt{x^2+y^2}} - 0 \right| \leq \left| \frac{x^2+y^2}{\sqrt{x^2+y^2}} \right| = \left| \sqrt{x^2+y^2} \right| < \epsilon.$$

□

Example 3.22

Compute the limit (if it exists) of

$$\lim_{(x,y)\to(0,0)} \frac{2x^2+y^2}{x^2+y^2}.$$

Again, we test a few paths to approach this limit to see what it can be. Approaching from the y-axis $x = 0$, we get

$$\lim_{y=0} \frac{y^2}{y^2} = \lim_{y=0} 1 = 1.$$

But if we approach from the x-axis $y = 0$, we will get

$$\lim_{x=0} \frac{2x^2}{x^2} = \lim_{x=0} 2 = 2.$$

These two limits are not equal, so we can conclude that the limit does not exist.

Example 3.23

Compute the limit (if it exists) of

$$\lim_{(x,y)\to(0,0)} \frac{x^2y}{x^4+y^2}.$$

This is a tricky example. The limits along the x and y axes are both zero. In fact, the limit along every line passing the origin $y = mx$ is zero, as

$$\lim_{x\to 0} \frac{x^2(mx)}{x^4+(mx)^2} = \lim_{x\to 0} \frac{mx}{x^2+m} = 0$$

for any m. However, if we take the limit along the path $y = x^2$, we get

$$\lim_{x\to 0} \frac{x^4}{x^4 + x^4} = \frac{1}{2}.$$

The limit does not exist!

To complete our discussion of limits, the limits in higher dimensions satisfy similar properties as the 1-D limits you know and love from Calculus.

Proposition 3.24

Let $\Omega \in \mathbb{R}^n, \vec{x}_0 \in \Omega$ and $f, g : \Omega \to \mathbb{R}^m, c : \Omega \to \mathbb{R}$. Suppose that the limits $\lim_{\vec{x}\to\vec{x}_0} f(\vec{x})$, $\lim_{\vec{x}\to\vec{x}_0} g(\vec{x})$ exists, then the following are true:

1. $\lim_{\vec{x}\to\vec{x}_0} f(\vec{x}) + g(\vec{x}) = \lim_{\vec{x}\to\vec{x}_0} f(\vec{x}) + \lim_{\vec{x}\to\vec{x}_0} g(\vec{x})$
2. $\lim_{\vec{x}\to\vec{x}_0} f(\vec{x}) - g(\vec{x}) = \lim_{\vec{x}\to\vec{x}_0} f(\vec{x}) - \lim_{\vec{x}\to\vec{x}_0} g(\vec{x})$
3. $\lim_{\vec{x}\to\vec{x}_0} c(\vec{x}) f(\vec{x}) = \lim_{\vec{x}\to\vec{x}_0} c(\vec{x}) \lim_{\vec{x}\to\vec{x}_0} f(\vec{x})$
4. $\lim_{\vec{x}\to\vec{x}_0} \frac{f(\vec{x})}{c(\vec{x})} = \frac{\lim_{\vec{x}\to\vec{x}_0} f(\vec{x})}{\lim_{\vec{x}\to\vec{x}_0} c(\vec{x})}$, if the denominator is not zero.

The proof is not enlightening, and just uses a lot of $\epsilon - \delta$ definitions for limits.

Definition 3.25 (Continuity)

Let $\Omega \subseteq \mathbb{R}^n$ be open, and $f : \Omega \to \mathbb{R}^m$. We say that f is **continuous** at $\vec{x}_0 \in \Omega$ if

$$\lim_{\vec{x}\to\vec{x}_0} f(x) = f(\vec{x}_0).$$

Remark. *Another way to look at this is that you can pull the limit into the function*

$$\lim_{\vec{x}\to\vec{x}_0} f(x) = f(\lim_{\vec{x}\to\vec{x}_0} \vec{x}) = f(\vec{x}_0).$$

Example 3.26

The following functions are continuous in the largest domain that they are defined in:

1. Elementary functions, i.e. $f(x)$ is the composition of sum, products, roots, polynomial, rational trigonometric, hyperbolic, trigonometric functions.
2. Compositions of different elementary functions in multiple variables.

Exercises

10.

3.4 Differentiation in $\mathbb{R}^n$

Recall the definition of a one-dimensional derivative:

Definition 3.27 (One dimensional derivative)

Let $\Omega \subseteq \mathbb{R}$ be open, and $f : \Omega \to \mathbb{R}$. We say f is **differentiable** at $x_0 \in \Omega$ if the limit

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

exists, and we call this the **derivative** of f at x_0, and denote it $f'(x_0)$.

It is a bit hard to extend this definition to functions $\mathbb{R}^n \to \mathbb{R}^m$. Let $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$, we forcefully write

$$\vec{f}'(\vec{x}_0) = \lim_{\vec{x} \to \vec{x}_0} \frac{\vec{f}(\vec{x}) - \vec{f}(\vec{x}_0)}{\vec{x} - \vec{x}_0},$$

this equation does not really make sense, as we do not have a way to divide by a vector. There are two ways to mitigate this. The first way is to compress everything into one-dimension, and take the derivative with respect to one variable while holding all others constant.

Definition 3.28 (Partial Derivative)

Let $\Omega \subseteq \mathbb{R}^n$, and $f : \Omega \to \mathbb{R}^m$. For $1 \leq k \leq n$, we say the **partial derivative** of f with respect to x_k at $\vec{a} \in \Omega$ is the limit

$$\lim_{\tilde{a}_k \to a_k} \frac{f(a_1, \ldots, \tilde{a}_k, \ldots, a_n) - f(a_1, \ldots, a_k, \ldots, a_n)}{\tilde{a}_k - a_k}.$$

In other words, this is the one-dimensional derivate in the variable x_k.

Notation. *The partial derivative is denoted*

$$\frac{\partial f}{\partial x_k}(\vec{a}) \quad \textit{or} \quad f_{x_k}(\vec{a}).$$

Example 3.29

Let $f(x, y) = (2x + \sin(xy), -y + \cos(xy))$. Compute the partial derivatives of f in x and y.

To compute the partial derivative in x (resp. y), we hold y constant, so under constant y (resp. x),

$$\begin{aligned}\frac{\partial f}{\partial x}(x, y) =& (2 + y\cos(xy), -y\sin(xy)),\\ \frac{\partial f}{\partial y}(x, y) =& (x\cos(xy), -x\sin(xy)).\end{aligned}$$

The existence of a partial derivative implies the existence of a total derivative, so let us try to motivate the definition of that ‘real derivative’. Rethinking the one-dimensional derivative, we also use the derivative $f'(x_0)$ as a linear approximation for the function f. What this means is that for small Δx,

$$f(x_0 + \Delta x) = f(x_0) + f'(x_0)\Delta x + \text{small error}.$$

How small is this error? Moving everything else to one side and taking the limit $\Delta x \to 0$,

$$\begin{aligned}
\lim_{\Delta x \to 0} \left| \frac{\text{small error}}{\Delta x} \right| &= \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0) - f'(x_0)\Delta x}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} - f'(x_0) \\
&= f'(x_0) - f'(x_0) \\
&= 0.
\end{aligned}$$

This means our error using this linear approximation shrinks faster Δx. Thus, we say that this is the best linear approximation for functions, and visually represent it as the line tangent to the curve at the point $(x_0, f(x_0))$. Similarly, we ask in higher dimensions, if we can approximate a function to be 'locally linear', i.e. there is a linear transformation

$$\vec{f}(\vec{x_0} + \Delta\vec{x}) = f(\vec{x}_0) + T(\Delta\vec{x}) + \text{small error},$$

with this small error shrinking to $\vec{0}$ faster than $\Delta\vec{x}$, or

$$\lim_{\Delta\vec{x} \to \vec{0}} \frac{|\text{small error}|}{|\Delta\vec{x}|} = 0.$$

If there is such a transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, then T should be the higher dimensional 'derivative' of $\vec{f}$ at $\vec{x}_0$. By the Matrix Representation of Linear Transformations, let us write the derivative T as a matrix.

Definition 3.30 (Total Derivative)

Let $\Omega \subseteq \mathbb{R}^n$ be open, and $\vec{f} : \Omega \to \mathbb{R}^m$. We say f is differentiable at $\vec{x}_0 \in \Omega$ if there is $A \in M_{m\times n}$ such that

$$\lim_{\vec{x} \to \vec{x}_0} \frac{\vec{f}(\vec{x}) - \vec{f}(\vec{x}_0) - A(\vec{x} - \vec{x}_0)}{|\vec{x} - \vec{x}_0|} = \vec{0}.$$

In this case, we call A the **(total) derivative** of $\vec{f}$ at $\vec{x}_0$, and denote it $d\vec{f}|_{\vec{x}_0}$.

Theorem 3.31 (Differentiable functions are continuous)

Let $\Omega \subseteq \mathbb{R}^n$ be open, and $f : \Omega \to \mathbb{R}^m$ be differentiable everywhere on Ω. Then f is continuous.

Proof. Let $\vec{x}_0 \in \Omega$. We have

$$\lim_{\vec{x} \to \vec{x}_0} \frac{\vec{f}(\vec{x}) - \vec{f}(\vec{x}_0) - df|_{\vec{x}_0}(\vec{x} - \vec{x}_0)}{|\vec{x} - \vec{x}_0|} = \vec{0},$$

and want to show

$$\lim_{\vec{x} \to \vec{x}_0} \vec{f}(\vec{x}) - \vec{f}(\vec{x}_0) = \vec{0}.$$

Here we apply a standard trick (plus-minus), which means we massage the second equation to resemble the first:

$$\begin{aligned}\lim_{\vec{x}\to\vec{x}_0} \vec{f}(\vec{x}) - \vec{f}(\vec{x}_0) &= \lim_{\vec{x}\to\vec{x}_0} \vec{f}(\vec{x}) - \vec{f}(\vec{x}_0) - df|_{\vec{x}_0}(\vec{x}-\vec{x}_0) - df|_{\vec{x}_0}(\vec{x}-\vec{x}_0) \\ &= \lim_{\vec{x}\to\vec{x}_0} \frac{\vec{f}(\vec{x}) - \vec{f}(\vec{x}_0) - df|_{\vec{x}_0}(\vec{x}-\vec{x}_0)}{|\vec{x}-\vec{x}_0|}|\vec{x}-\vec{x}_0| - df|_{\vec{x}_0}(\vec{x}-\vec{x}_0)\end{aligned}$$

The first term is $\vec{0} * 0$ by the definition of derivative, and $|\vec{x}-\vec{x}_0|$ is a composition of elementary functions thus continuous. The second term is $df|_{\vec{x}_0}(\vec{0})$ as linear transformations are polynomials thus continuous. Therefore the whole expression equals $\vec{0}$. □

Theorem 3.32 (Uniqueness of derivative)

The derivative df, if it exists, is unique and equals the matrix

$$J = \begin{bmatrix} \frac{\partial \vec{f}}{\partial x_1} & \frac{\partial \vec{f}}{\partial x_2} & \frac{\partial \vec{f}}{\partial x_3} & \dots & \frac{\partial \vec{f}}{\partial x_n} \end{bmatrix}$$

evaluated at $\vec{x}_0$.

Definition 3.33 (Jacobian Matrix)

Suppose the partial derivatives of $\vec{f}$ exist, then we call the matrix

$$J \begin{bmatrix} \frac{\partial \vec{f}}{\partial x_1} & \frac{\partial \vec{f}}{\partial x_2} & \frac{\partial \vec{f}}{\partial x_3} & \dots & \frac{\partial \vec{f}}{\partial x_n} \end{bmatrix} =$$

the **Jacobian Matrix** of f.

Before we go on to the proof, try if you can prove this using the ideas from the Matrix Representation of a linear transformation. Here is a hint: The partial derivative can also be written as

$$\frac{\partial f}{\partial x_k}(\vec{x}_0) = \lim_{h\to 0} \frac{f(\vec{x}_0 + h\vec{e}_k) - f(\vec{x}_0)}{h}.$$

Proof of Uniqueness of Derivative. Let $\vec{x}_0 \in \Omega$. By the uniqueness of limit, we approach the limit

$$\lim_{\vec{x}\to\vec{x}_0} \frac{\vec{f}(\vec{x}) - \vec{f}(\vec{x}_0) - d\vec{f}|_{\vec{x}_0}(\vec{x}-\vec{x}_0)}{|\vec{x}-\vec{x}_0|}$$

along the line $\vec{x} = \vec{x}_0 + h\vec{e}_k$ as $h \to 0$. Therefore,

$$\begin{aligned} &\lim_{h\to 0} \frac{\vec{f}(\vec{x}_0 + h\vec{e}_k) - \vec{f}(\vec{x}_0) - d\vec{f}|_{\vec{x}_0}(h\vec{e}_k)}{|h|} = \vec{0} \\ \implies &\lim_{h\to 0} \frac{\vec{f}(\vec{x}_0 + h\vec{e}_k) - \vec{f}(\vec{x}_0) - d\vec{f}|_{\vec{x}_0}(h\vec{e}_k)}{h} = \vec{0} \\ \implies &\lim_{h\to 0} \frac{\vec{f}(\vec{x}_0 + h\vec{e}_k) - \vec{f}(\vec{x}_0)}{h} - d\vec{f}|_{\vec{x}_0}(\vec{e}_k) = \vec{0}. \end{aligned}$$

This means that the k-th column of $d\vec{f}|_{\vec{x}_0}$ is $\vec{f}_{x_k}(\vec{x}_0)$, the partial derivative of $\vec{f}$ with respect to the k-th variable. $\square$

Corollary 3.34: If a function has a total derivative, then its partial derivatives exist.

Example 3.35

Determine if the function

$$f(x,y) = x^2 + y^2.$$

is differentiable.

If this function is differentiable then the derivative must be the Jacobian $[2x\ 2y]$. Indeed, we can check for every $(x_0, y_0) \in \mathbb{R}^2$

$$\begin{aligned}
&\lim_{(x,y)\to(x_0,y_0)} \frac{x^2 + y^2 - x_0^2 - y_0^2 - \begin{bmatrix} 2x_0 & 2y_0 \end{bmatrix} \begin{bmatrix} (x-x_0) \\ (y-y_0) \end{bmatrix}}{|(x,y)-(x_0,y_0)|} \\
=& \lim_{(x,y)\to(x_0,y_0)} \frac{(x-x_0)^2 + (y-y_0)^2}{((x-x_0)^2 + (y-y_0)^2)^{1/2}} \\
=&0.
\end{aligned}$$

Therefore, this function is differentiable everywhere in $\mathbb{R}^2$.

Having partial derivatives is not a sufficient condition for differentiability.

Example 3.36

Compute the partial derivatives of the function

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2}, & \text{if } (x,y) \neq (0,0) \\ 0, & \text{if } (x,y) = 0,0 \end{cases}$$

at $(x,y) = 0, 0$.

The partial derivatives in x and y are both 0, as $f = 0$ when $x = 0$ or $y = 0$. However, this function is not differentiable. You can show this by checking that the Jacobian matrix J does not satisfy the definition of the derivative. We can also get by with less work. This function is not even continuous! In fact, if we approach the limit $(x,y) \to (0,0)$ along $y = x$, we get

$$\lim_{x\to 0} \frac{x^2}{x^2 + x^2} = \frac{1}{2} \neq f(0,0).$$

Example 3.37

Determine if the function

$$f(x, y) = \begin{cases} \frac{2x^2y+y^3}{x^2+y^2}, & \text{if } (x, y) \neq (0, 0) \\ 0, & \text{if } (x, y) = 0, 0 \end{cases}$$

is differentiable at $(x, y) = 0, 0$.

This function is continuous at $(0, 0)$. Indeed we have

$$f_x(0, 0) = 0, f_y(0, 0) = \lim_{y \to 0} \frac{1}{y} \frac{y^3}{y^2} = 1.$$

Now let us investigate the limit

$$\lim_{(x,y)\to(0,0)} \frac{\frac{2x^2y+y^3}{x^2+y^2} - \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}}{\sqrt{x^2 + y^2}} = \lim_{(x,y)\to(0,0)} \frac{2x^2y + y^3 - x^2y - y^3}{(x^2 + y^2)^{3/2}} = \lim_{(x,y)\to(0,0)} \frac{x^2y}{(x^2 + y^2)^{3/2}}.$$

This limit does not exist. Specifically, the limit as we approach along the x-axis is 0, but is not zero if we take the limit along the line $x = y$. Therefore this function is not differentiable at $(0, 0)$.

Theorem 3.38 (Differentiability)

Let $\Omega \subseteq \mathbb{R}^n$ be open, and $f : \Omega \to \mathbb{R}^m$. Let $x_0 \in \Omega$, and the partial derivatives of f are continuous in some open ball $B(x_0, \epsilon) \subseteq \Omega$. Then f is differentiable at x_0.

The proof is optional. Here is the idea: Since we compute limit of vectors as the limit in each coordinate, it suffices to check that each component of $f = (f_1, f_2, \ldots, f_m)$ is differentiable. Now let us restrict to one of these f_i's, and call it g. We already have a candidate for its derivative, namely we want to show

$$Dg(\vec{x}) = \begin{bmatrix} \frac{\partial g}{\partial x_1} & \cdots & \frac{\partial g}{\partial x_n} \end{bmatrix}.$$

Let us consider a point $\vec{x} + \vec{h}$. We want to relate $g(\vec{x})$ to $g(\vec{x} + \vec{h})$. To do this, we slowly 'walk' from $\vec{x}$ to $\vec{x} + \vec{h}$. The path γ we take is as follows:

- walk from $p_0 \stackrel{\text{def}}{=} \vec{x}$ to $p_1 \stackrel{\text{def}}{=} \vec{x} + (h_1, 0, \ldots, 0)$
- walk from $\vec{x} + (h_1, 0, \ldots, 0)$ to $p_2 \stackrel{\text{def}}{=} \vec{x} + (h_1, h_2, \ldots, 0)$
- repeat for each other coordinate
- walk from $p_{n_1} \stackrel{\text{def}}{=} \vec{x} + (h_0, \ldots, h_{n-1}, 0)$to $p_n \stackrel{\text{def}}{=} \vec{x} + (h_0, \ldots, h_{n-1}, h_n)$.

On each segment of this path, $f \circ \gamma$ is a differentiable function, and we can apply the mean value theorem to bound how much f changes.

Proof. It suffices to show the case where $m = 1$. Let $\vec{x} \in \Omega$. We want to show that

$$Df(\vec{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}.$$

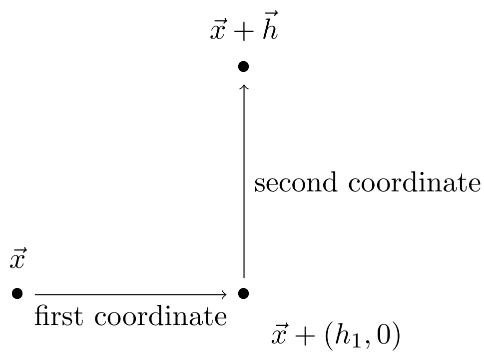


Figure 3.6: sample path for $f:\mathbb{R}^2\to\mathbb{R}$

Let $\epsilon>0$. By continuity of the partial derivatives, find $r>0$ such that for all $1\le j\le n$ and $\vec{y}<r$,

$$\left|\frac{\partial f}{\partial x_j}-\frac{\partial f}{\partial x_j}\right|\le\frac{\epsilon}{n}$$

Therefore, for all $0<\vec{h}<r$,

$$\begin{aligned}f(\vec{x}+\vec{h})-f(\vec{x})&=\sum_{j=1}^{n}f(p_j)-f(p_{j-1})\\&\le\sum_{j=1}^{n}h_j\frac{\partial f}{\partial x_j},\end{aligned}$$

where $\tilde{p}_j$ is a point on the straight line between p_{j-1} and p_j by the Mean Value Theorem. Thus,

$$\begin{aligned}\frac{1}{|\vec{h}|}|f(\vec{x}+\vec{h})-f(\vec{x})-\begin{bmatrix}\frac{\partial f}{\partial x_1}&\cdots&\frac{\partial f}{\partial x_n}\end{bmatrix}\vec{h}|\le&\frac{1}{|h|}\sum_{j=1}^{n}|h_j|\left|\frac{\partial f}{\partial x_j}-\frac{\partial f}{\partial x_j}\right|\\\le&\frac{1}{|h|}\sum_{j=1}^{n}|h|\frac{\epsilon}{n}\\\le&\epsilon.\end{aligned}$$

We thus have

$$\lim_{\vec{h}\to\vec{0}}\frac{|f(\vec{x}+\vec{h})-f(\vec{x})-\begin{bmatrix}\frac{\partial f}{\partial x_1}&\cdots&\frac{\partial f}{\partial x_n}\end{bmatrix}\vec{h}|}{|\vec{h}|}=0$$

as desired. □

Theorem 3.39 (Chain Rule)

Let $\Omega_1\subseteq\mathbb{R}^n$, $\Omega_2\subseteq\mathbb{R}^m$ be open subsets. Let $f:\Omega_1\to\Omega_2$, $g:\Omega_2\to\mathbb{R}^p$ (That is, the composition $g\circ f:\Omega_1\to\mathbb{R}^p$). If f and g are differentiable then $g\circ f$ is differentiable, and

$$d(g\circ f)(\vec{x})=dg(f(\vec{x}))df(\vec{x}).$$

The idea behind the proof of the chain rule is not too hard; the execution is just slightly frustrating. We can approximate f to its first order: for some small Δx

$$f(\vec{x}+\Delta x) = f(\vec{x}) + df(\vec{x})\Delta x + \delta_f(\Delta x),$$

with $\lim_{\Delta x\to\vec{0}} \delta_f(\Delta x)/|\Delta x| = \vec{0}$. Similarly,

$$g(f(\vec{x})+\Delta y) = g(f(\vec{x})) + dg(f(\vec{x}))\Delta y + \delta_g(\Delta y)$$

with $\lim_{\Delta y\to\vec{0}} \delta_g(\Delta y)/|\Delta y| = \vec{0}$. Therefore,

$$\begin{aligned} g(f(\vec{x})+\Delta x) &= g(f(\vec{x}) + df(\vec{x})\Delta x + \delta_f(\Delta x)) \\ &= g(f(\vec{x})) + dg(f(\vec{x}))(df(\vec{x})\Delta x + \delta_f(\Delta x)) + \delta_g(df(\vec{x})\Delta x + \delta_f(\Delta x)) \\ &= g(f(\vec{x})) + dg(f(\vec{x}))df(\vec{x})\Delta x + dg(f(\vec{x}))\delta_f(\Delta x) + \delta_g(df(\vec{x})\Delta x + \delta_f(\Delta x)). \end{aligned}$$

The first two terms is a linear approximation of $g\circ f$ with the 'derivative' equals the product of the two matrices. Therefore, we just need to show that the remaining terms go to zero (faster than Δx) if we take the limit $\Delta x \to \vec{0}$.

Remaining part of the proof (Optional). By continuity of linear transformations, we get

$$\lim_{\Delta x\to\vec{0}} \frac{dg(f(\vec{x}))\delta_f(\Delta x)}{|\Delta x|} = dg(f(\vec{x})) \lim_{\Delta x\to\vec{0}} \frac{\delta_f(\Delta x)}{|\Delta x|} = g(f(\vec{x}))\vec{0} = \vec{0}.$$

For the other term,

$$\lim_{\Delta x\to\vec{0}} \frac{\delta_g(df(\vec{x})\Delta x + \delta_f(\Delta x))}{|\Delta x|} = \lim_{\Delta x\to\vec{0}} \frac{\delta_g(df(\vec{x})\Delta x + \delta_f(\Delta x))}{|df(\vec{x})\Delta x + \delta_f(\Delta x)|}\frac{|df(\vec{x})\Delta x + \delta_f(\Delta x)|}{|\Delta x|}.$$

By continuity of $\delta_g(df(\vec{x})\Delta x + \delta_f(\Delta x))$ in Δx, the first term in the product goes to $\vec{0}$ as $\Delta x \to \vec{0}$. The second term is bounded by $|df(\vec{x})| + |\delta_f(\Delta x)/\Delta x|$. (Here we used $|df(\vec{x})|$ to denote the *operator norm* of the matrix, which will not be tested). This term is bounded, so the magnitude of the limit is bounded by

$$\lim_{\Delta x\to\vec{0}} C\frac{|\delta_g(df(\vec{x})\Delta x + \delta_f(\Delta x))|}{|df(\vec{x})\Delta x + \delta_f(\Delta x)|} = 0$$

for some constant $C > 0$. Therefore we get the definition of differentiability

$$\lim_{\Delta x\to\vec{0}} \frac{g(f(\vec{x})+\Delta x) - g(f(\vec{x})) - dg(f(\vec{x}))df(\vec{x})\Delta x}{|\Delta x|} = \vec{0}.$$

□

Corollary 3.40: Write $f(x_1,x_2,\ldots,x_n)$ and $g(y_1,y_2,\ldots,y_m)$. Then for each $1\le k\le n$, $1\le l\le m$,

$$\frac{\partial(g\circ f)_l}{\partial x_k} = \sum_{i=1}^{m} \frac{\partial g_l}{\partial y_i}\frac{\partial f_i}{\partial x_k}.$$

Example 3.41

Let $f(t) = (t,t^2,t^3), w(x,y,z) = e^{xy+z}$. Compute the derivative $(w\circ f)'(t)$ using the chain rule.

Here we have the derivative in one varible, so this coincides with the total derivative. Moreover, these functions are elementary, so the partial derivatives are continuous (so the functions are differentaible). Using the Chain rule,

$$\begin{aligned}
(w \circ f)'(t) &= \begin{bmatrix} \frac{\partial w}{\partial x}|_{(t,t^2,t^3)} & \frac{\partial w}{\partial y}|_{(t,t^2,t^3)} & \frac{\partial w}{\partial z}|_{(t,t^2,t^3)} \end{bmatrix} \begin{bmatrix} \frac{dt}{dt} \\ \frac{dt^2}{dt} \\ \frac{dt^3}{dt} \end{bmatrix} \\
&= \begin{bmatrix} e^{2t^3}t^2 & e^{2t^3}t & e^{2t^3} \end{bmatrix} \begin{bmatrix} 1 \\ 2t \\ 3t^2 \end{bmatrix} \\
&= 6t^2 e^{2t^3}.
\end{aligned}$$

As a sanity check, we can directily compute

$$\frac{d}{dt} w(f(t)) = \frac{d}{dt} e^{2t^3} = 6t^2 e^{2t^3}$$

using the one-dimensional chain rule you have learnt in high school calculus.

3.5 Applications - Geometry of Tangent Lines and Planes

This section will require material from the first chapter.

Example 3.42

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be differentiable. Find the equation of the plane tangent to the graph of f at the point $(x_0, y_0, f(x_0, y_0))$.

The plane is the best linear approximation of f,

$$f(x_0 + \Delta x, y_0 + \Delta y) \approx f(x_0, y_0) + df(x_0, y_0) \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = f(x_0, y_0) + \left.\frac{\partial f}{\partial x}\right|_{(x_0,y_0)} \Delta x + \left.\frac{\partial f}{\partial y}\right|_{(x_0,y_0)} \Delta y.$$

Therefore the tangent plane should be described by the equation

$$z = f(x, y) + \frac{\partial f}{\partial x}(x - x_0) + \frac{\partial f}{\partial y}(y - y_0) \implies \frac{\partial f}{\partial x}(x - x_0) + \frac{\partial f}{\partial y}(y - y_0) - (z - f(x, y)) = 0.$$

Example 3.43

The plane tangent to the graph of $f(x, y) = x^2 - y^2$ at $(1, 1, 0)$ is

$$2(x - 1) - 2(y - 1) - z = 0 \implies 2x - 2y - z = 0.$$

Definition 3.44 (Directional Derivatives)

Let $\Omega \subseteq \mathbb{R}^n$ be open, and $f : \Omega \to \mathbb{R}$. Let $\vec{v} \in \mathbb{R}^n$. We define the **directional derivative** of f along the direction of $\vec{v}$

$$\nabla_{\vec{v}} f(\vec{x}) \stackrel{\text{def}}{=} \lim_{h \to 0} \frac{f(\vec{x} + h\vec{v}) - f(\vec{x})}{h}.$$

Remark. *In this definition, we constrain ourselves to functions $\mathbb{R}^n \to \mathbb{R}$. i.e. the output of the function is a scalar.*

Definition 3.45 (Gradient)

Let $\Omega \subseteq \mathbb{R}^n$ be open, and $f : \Omega \to \mathbb{R}$ be differentiable. We define the **gradient** of f

$$\nabla f(\vec{x}) \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}.$$

Remark. *Using the tranpose notation, we can also write ∇f to be the column vector ($n \times 1$ matrix) df^T.*

Proposition 3.46

If f is differentiable then $\nabla_{\vec{v}} f(\vec{x}) = \nabla f(\vec{x}) \cdot \vec{v}$.

Proof. Since f is differentaible, we have

$$\lim_{\vec{x}' \to \vec{x}} \frac{f(\vec{x}') - f(\vec{x}) - df(\vec{x})(\vec{x}' - \vec{x})}{|\vec{x}' - \vec{x}|} = 0.$$

Taking this limit along the path $\vec{x}' = \vec{x} + h\vec{v}$,

$$0 = \lim_{h \to 0} \frac{f(\vec{x} + h\vec{v}) - f(\vec{x}) - hdf(\vec{x})(\vec{v})}{|h||\vec{v}|}.$$

Or,

$$\begin{aligned} 0 &= \lim_{h \to 0} \frac{f(\vec{x} + h\vec{v}) - f(\vec{x}) - hdf(\vec{x})(\vec{v})}{h} \\ &= \lim_{h \to 0} \frac{f(\vec{x} + h\vec{v}) - f(\vec{x})}{h} - df(\vec{x})(\vec{v}). \end{aligned}$$

Therefore, the directional derivative is $df(\vec{x})(\vec{v}) = \nabla f(\vec{x}) \cdot \vec{v}$. $\square$

By the Chain rule, you should recognize this as the derivative in t for the composition $f \circ \vec{r}(t)$ at $t = 0$, where $\vec{r}(0) = \vec{x}$, $\vec{r}'(0) = \vec{v}$.

In other words, suppose our n-th dimensional-lander Nadia is moving in Ω along a differentiable path $\vec{r}$. At the moment she reaches $\vec{x}$, her instantaneous velocity is $\vec{v}$, $\nabla_{\vec{v}}$ will be the rate of change of the f-meter she experiences.

Example 3.47

Let $\Omega \subseteq \mathbb{R}^n$ be open, and $f : \Omega \to \mathbb{R}$ be differentiable. Suppose that $\nabla f(\vec{x}) \neq \vec{0}$. Find the unit vector $\vec{u}$ such that $\nabla_{\vec{u}}$ is i) largest, ii) smallest.

In other words, suppose Nadia is now moving with unit speed. What direction does she need to face for the fastest ascent/descent of the f-meter?

By the Cauchy Schwarz inequality, we get

$$|\nabla_{\vec{u}} f(\vec{x})| \leq |\nabla f||\vec{u}| = |\nabla f|.$$

Therefore, the best we can hope for is a rate of $|\nabla f|$ for ascent or a rate of $-|\nabla f|$ for descent. Fortunately, this is possible with the following unit vectors:

$$\nabla f \cdot \frac{\nabla f}{|\nabla f|} = |\nabla f|,$$
$$\nabla f \cdot \frac{-\nabla f}{|\nabla f|} = -|\nabla f|.$$

This means that ∇f also points in the direction of fastest ascent, and the magnitude of ∇f describes the rate of ascent!

Remark. *You can also see ∇f and $-\nabla f$ will form an angle of 0 or π with ∇f, so you can also construct these two unit vectors using the cosine definition of dot product.*

Example 3.48 (Level sets orthogonal to gradient)

Show that levels sets are orthogonal to the gradient. Precisely, let $\Omega \subset \mathbb{R}^n$, $f : \Omega \to \mathbb{R}\gamma : [0, 1] \to \Omega$ be a curve that satisfies

$$f(t) = c$$

for a constant c. Show that

$$\gamma'(t) \cdot (\nabla f)(\gamma(t)) = 0.$$

for all $t \in [0, 1]$. Here

$$\gamma'(t) \stackrel{\text{def}}{=} D\gamma(t) = \begin{bmatrix} \gamma_1'(t) \\ \vdots \\ \gamma_n'(t) \end{bmatrix}.$$

We apply the Chain rule on the function $f(\gamma(t))$. On one hand, since $f(\gamma(t))$ is constant,

$$D(f \circ \gamma)(t) = 0.$$

Using the chain rule,

$$D(f \circ \gamma)(t) = (Df)(\gamma(t))\gamma'(t) = (\nabla f)(\gamma(t)) \cdot \gamma'(t).$$

Example 3.49

Consider the level set of the function

$$F(x, y, z) = xy - z$$

at 0. What is the equation of the tangent plane of the level set at $(0, 0, 0)$.

At the origin, $\nabla F = (0, 0, 1)$. This vector is normal to the level set, that is, it is normal to the tangent plane of the level set at $(0, 0, 0)$, therefore, the equation of the tangent plane is

$$0(x - 0) + 0(y - 0) + 1(z - 0) = 0 \implies z = 0.$$

The tangent plane is the xy-plane. You can confirm that the level set is the graph of the function $z = xy$. This is the saddle we graphed a few sections ago. At the origin, the saddle is locally flat, so has the flat tangent plane.

3.6 Implicit Differentiation

We warmup with an example of implicit differentiation in one variable.

Example 3.50

Let $x^2 + y^2 = 1$. Find dy/dx.

To solve this, we can take the derivative with respect to x on both sides to get

$$2x + 2y\frac{dy}{dx} = 0 \implies \frac{dy}{dx} = -\frac{x}{y}.$$

Let us interpret the result as a sanity check. $x^2 + y^2 = 1$ describe a circle, and dy/dx is the slope of the tangent line at any point. At $(0, 1)$ and $(0, -1)$, we get the highest and lowest points of the circle, so the slope is 0. In the first and third quadrant, the slope is negative; in the second and fourth quadrant, the slope is positive.

This dy/dx seems slightly different from the usual limit definition in the sense that y is not necessarily a function of x. However, we can locally define y in terms of x as follows.

- In the top half plane where $y > 0$, we can set $y = \sqrt{1 - x^2}$.
- In the top half plane where $y < 0$, we can set $y = -\sqrt{1 - x^2}$.

Under this construction, you can verify that $dy/dx = -x/y$ at any point on the circle where $y \neq 0$. When $y = 0$, the function breaks down as we divide by zero. This corresponds to the vertical tangent line at that point. There is no way to define y locally as a function of x.

The way we defined y as a function of x is known as an implicit function. In the sense that it is implicitly defined by the equation $x^2 + y^2 = 1$.

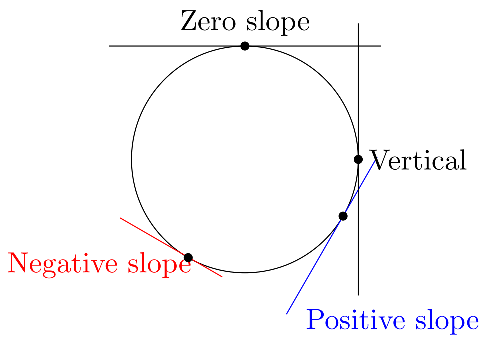


Figure 3.7: We can define y as a function of x locally on the level set $x^2 + y^2 = 1$.

Definition 3.51 (Implicit Equation)

Let $\Omega \subseteq \mathbb{R}^n$ be open, and $f : \Omega \to \mathbb{R}^m$. An **implicit equation** is a relation in the form

$$f(\vec{x}) = \vec{0}.$$

By defining $g(\vec{x}) = f(\vec{x}) - \vec{c}$ for some constant $\vec{c} \in \mathbb{R}^m$, we also get

$$f(\vec{x}) = \vec{c}$$

is an implicit equation.

Our question is as follows: given an implicit equation, can we 'solve' this equation locally? If such an equation exists, what is the partial derivatives of the variables?

We shall spoil the big result of this section.

Theorem 3.52 (Implicit function)

Let $\Omega \subseteq \mathbb{R}^{n+m}$ be open, and $f : \Omega \to \mathbb{R}^n$. We denote the coordinates of a point $(\vec{x}, \vec{y}) \in \Omega$ as $(x_1, x_2, \dots, x_n, y_1, \dots, y_m)$. Let $(\vec{a}, \vec{b}) \in \Omega$. If the matrix

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

is invertible at $(\vec{a}, \vec{b}) \in \Omega$, then there exists an open set $U \subseteq \mathbb{R}^m$ containing $\vec{b}$ and a differentiable function $g : U \to \mathbb{R}^n$ such that

$$f(g(\vec{y}), \vec{y})$$

is constant, and $g(\vec{b}) = \vec{a}$. In this case, we call the function g an **implicit function**.

Before we work through some examples, let us parse the statement of the Implicit Function

Theorem. The Jacobian matrix of f can be expressed as

$$Df \overset{\text{def}}{=} \left[Df_x \,\middle|\, Df_y \right]. = \left[\begin{array}{cccc|cccc} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} & \frac{\partial f_1}{\partial y_1} & \frac{\partial f_1}{\partial y_2} & \cdots & \frac{\partial f_1}{\partial y_m} \\ \vdots & \ddots & & \vdots & \vdots & \ddots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} & \frac{\partial f_n}{\partial y_1} & \frac{\partial f_n}{\partial y_2} & \cdots & \frac{\partial f_n}{\partial y_m} \end{array}\right].$$

This is a $n \times (n+m)$ matrix. The left matrix Df_x is exactly the matrix we want to check for the conditions of Implicit Function theorem. If it is invertible, then That One Theorem implies that there is a unique solution in $\vec{x}$ for the equation

$$Df_x\vec{x} + Df_y\vec{y} = \vec{c}$$

for any fixed $\vec{c} \in \mathbb{R}^n$ and $\vec{y} \in \mathbb{R}^m$. Therefore, if we hold $\vec{c}$ constant, we should expect that we can solve $\vec{x}$ as a function of $\vec{y}$.

Another way to look at it is to apply the Rank-Nullity theorem. Approximating the function locally as

$$f(\vec{a}, \vec{b}) + Df_x(\vec{x} - \vec{a}) + Df_y(\vec{y} - \vec{b}) + \text{error}.$$

The rank of Df is n, so the nullity is $(n+m) - n = m$. We thus have m dimensions that are in 'excess'. These m variables are the y_j's, each corresponding to one dimension in the nullspace of D_f.

Example 3.53 (Derivative of implicit function)

Work out the derivative of the implicit function $g : U \to \mathbb{R}^n$ using the Chain rule.

Recall that $f(g(\vec{y}), \vec{y})$ is constant. This is a composition of two functions. The first function sends

$$\vec{y} \mapsto (g(\vec{y}), \vec{y}),$$

which has Jacobian matrix

$$\left[\begin{array}{c} Dg \\ \hline I_m \end{array}\right]$$

and the second function is f. Since the function is constant, the derivative of the composition of these functions is $0_{m\times m}$, so we have

$$0_{m\times m} = \left[Df_x \,\middle|\, Df_y \right] \left[\begin{array}{c} Dg \\ \hline I_m \end{array}\right] = Df_x Dg + Df_y.$$

Rearranging the terms,

$$-Df_y = Df_x Dg \implies Dg = -Df_x^{-1} Df_y.$$

Example 3.54

Let $\Omega \subseteq \mathbb{R}^n$ be open, and $f : \Omega \to \mathbb{R}$. Suppose for $\vec{v} \in \Omega$ satisfies $\nabla f(\vec{v}) \neq \vec{0}$. What is the local dimension of the level set described by the implicit equation

$$f(\vec{x}) = f(\vec{v})$$

around $\vec{v}$?

We wish to apply the implicit function theorem. If the conditions of the implicit function theorem applies, then this will be a function $\mathbb{R}^{(n-1)+1} \to \mathbb{R}^1$. So the implicit function gives that the level set is locally the graph of a function in $(n-1)$ variables (locally $(n-1)$ dimensions).

To check for the conditions of implicit function theorem, we use the fact that $Df = \nabla f^T$ is non zero at $\vec{v}$, thus has rank 1. If the i-th column is the pivot column, then we can express

$$x_i = g(x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

Physically, Dg will define the tangent line/plane/hyperplane to the level set at $(\vec{v}, f(\vec{v}))$.

The tangent plane is not guaranteed to exist. For instance, take the implicit equation described by

$$x^2 - y^2 = 0.$$

Plotting this in $\mathbb{R}^2$, this gives two lines $y = x$ and $y = -x$. At almost every point (x, y) in the level set, exactly one of $y = x$ or $y = -x$ hold. Therefore, we can restrict our attention to one line, and the local dimension of the line is 1.

On the other hand, the point $(0, 0)$ lies on both lines. For a (miniature-sized) person standing at the origin, the level set is locally a cross, which is the union of two 1-dimensional objects does not look like a line or a plane.

by the Implicit Function Theorem.

Definition 3.55 (Implicit Derivative)

Let $\Omega \subseteq \mathbb{R}^n$ be open, and $f : \Omega \to \mathbb{R}$. Consider the implicit equation

$$f(\vec{x}) = c.$$

We define the **Implicit Derivative**

$$\frac{\partial x_i}{\partial x_j} \stackrel{\text{def}}{=} \frac{\partial g}{\partial x_j},$$

where g is the implicit function that solves for x_i using the variables $(x_1, \ldots, x_{i-1}, x_i, \ldots, x_n)$.

Example 3.56

Express the implicit derivative in terms of the partial derivatives of f.

We first check the conditions of the Implicit function theorem. To express x_i as an implicit function of the other variables, the square 'matrix' we need to check for invertibility is

$$\begin{bmatrix} \frac{\partial f}{\partial x_i} \end{bmatrix}.$$

This is invertible exactly when the partial derivative is non zero. Thus using Example 3.53,

$$\frac{\partial x_i}{\partial x_j} = -\frac{\frac{\partial f}{\partial x_j}}{\frac{\partial f}{\partial x_i}}.$$

You can confirm this equation for dy/dx on the circle implicitly defined by $x^2 + y^2 = 1$.

3.7 Inverse function theorem

We can also ask the question of whether a function has an inverse. That is, for a function $f : \mathbb{R}^n \to \mathbb{R}^m$, is there a function $g : \mathbb{R}^m \to \mathbb{R}^n$ that undos f. We require

$$f(g(\vec{y})) = \vec{y}$$

for all $\vec{y} \in \mathbb{R}^m$ and

$$g(f(\vec{x})) = \vec{x}$$

for all $\vec{x} \in \mathbb{R}^n$. If such a function exists, we call it the inverse of f, and denote it as f^{-1}.

Example 3.57

Suppose f^{-1} exists. What is $D(f^{-1})$?

We apply the chain rule on $f \circ f^{-1}$ and $f^{-1} \circ f$. Suppose $f(\vec{x}) = \vec{y}$ and thus $f^{-1}(\vec{y}) = \vec{x}$. On one hand, we have

$$I_m = D(f \circ f^{-1})(\vec{y}) = Df(f^{-1}(\vec{y}))D(f^{-1})(\vec{y}) = Df(\vec{x})D(f^{-1})(\vec{y}).$$

On the other hand, we have

$$I_n = D(f \circ f^{-1}) = D(f^{-1})(f(\vec{x}))Df(\vec{x}) = D(f^{-1})(\vec{y})Df(\vec{x}).$$

That means $Df(\vec{x})$ has a left and right inverse $D(f^{-1})(\vec{y})$. So $Df(\vec{x})$ is surjective and injective by Example **??**. By That One Theorem, $Df(\vec{x})$ is invertible, with $(Df(\vec{x}))^{-1} = D(f^{-1})(\vec{y})$.

The surprising fact is that the converse is true. If $Df(\vec{x})$ is invertible, then locally there is an inverse function f^{-1}.

Theorem 3.58 (Inverse Function)

Let $\Omega \subseteq \mathbb{R}^n$ be open, and $f : \Omega \to \mathbb{R}^n$. Let $\vec{a} \in \Omega$, and suppose that $Df(\vec{x})$ is invertible. Then there exists open sets U containing $\vec{a}$ and V containing $\vec{b} = f(\vec{a})$, and a function $g : V \to U$ such that

$$f(g(\vec{y})) = \vec{y}$$
$$g(f(\vec{x})) = \vec{x}$$

for all $\vec{y} \in V$ and $\vec{x} \in U$. Moreover,

$$Dg(\vec{y}) = (Df(\vec{x}))^{-1}$$

for $\vec{y} = f(\vec{x})$, and $\vec{x} \in U$.

Proof. The second statement is a result of the previous example, so we only need to show the first statement.

Consider the implicit equation in $2n$ variables

$$f(\vec{x}) - \vec{y} = \vec{0}.$$

On one hand, we can solve $\vec{y} = f(\vec{x})$ explicitly. By the inverse function theorem, this is unique at some open set around $\vec{x}$. If we can obtain $\vec{x} = g(\vec{y})$ as an implicit function, then we must have f and g are inverses of each other. Indeed, since Df is invertible, the conditions of the Implicit Function Theorem apply and we have an implicit function g such that

$$f(g(\vec{y})) - \vec{y} = \vec{0}.$$

□

Exercises

Chapter 4

Multiple Integrals

4.1 Introduction

By this point, you should have noted that the geometry of $\mathbb{R}^n$ is intimately connected to calculus. Partial derivatives and tangent planes go hand in hand. However, if there's anything to take away from your first calculus course is that derivatives also go hand in hand with integrals. So you have every right to ask "What does it mean to do integrals in higher-dimensions?" This chapter will focus on developing the core concepts that surround this question. We will make sense of integrals in 2 and 3 dimensions by generalizing our intuition for the 1D case. We will see how — in certain situations — we can convert double and triple integrals into a simple iteration of singular integrals. Finally, we will derive the multidimensional version of the change of variables formula. Along the way, various applications to other disciplines will be presented, and you are encouraged to test your understanding via the exercises in each section.

To start our discussion, let's remind ourselves of what it means to integrate a real-valued function $f(x)$ along some interval $I = \{x \in \mathbb{R} : \ a \leq x \leq b\}$,

$$\int_a^b f(x) \ \mathrm{d}x \tag{4.1}$$

Intuitively, we interpret this as the signed area determined by $f(x)$ when $x \in I$. But how do we compute this area, in general? A good strategy is to partition the interval into very small chunks that are easier to deal with, and then add up their individual contributions to the integral. To that effect, let's partition the interval I into a sequence of intervals $I_1, I_2, \cdots, I_n$ according to a sequence of points $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$. In other words, take $I_k = [x_{k-1}, x_k]$. This procedure is demonstrated in Figure 4.1. Now, take a representative point $\tilde{x}_k \in I_k$ from each interval. It doesn't matter what this point is exactly, as we are assuming the intervals are small enough such that the function doesn't change very much within each interval. The signed area of f on each of the I_k is approximately $f(\tilde{x}_k)(x_k - x_{k-1})$. Thus, since $I = I_1 \cup I_2 \cdots \cup I_n$, we can approximate the integral as follows:

$$\int_a^b f(x) \ \mathrm{d}x \approx \sum_{k=1}^{n} f(\tilde{x}_k)(x_k - x_{k-1}) \tag{4.2}$$

The right-hand side is a *Riemman sum*, as you learned in your calculus course. We can measure the accuracy of our approximation by the largest interval width $\sigma = \max_k\{x_k - x_{k-1}\}$. Then, we

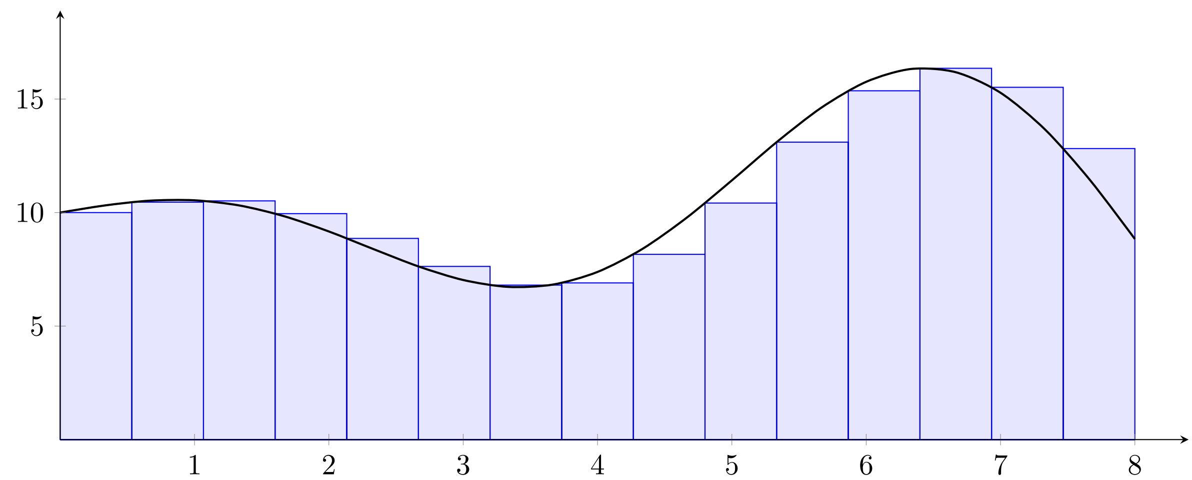


Figure 4.1: In this plot, we took $\tilde{x}_k = x_{k-1}$.

define[1] the *Riemman integral* to be:

$$\int_a^b f(x)\ \mathrm{d}x \stackrel{\text{def}}{=} \lim_{\sigma \to 0} \sum_{k=1}^{n} f(\tilde{x}_k)(x_k - x_{k-1}) \tag{4.3}$$

We can use this definition to calculate various quantities of relevance in science. For example, imagine someone gives you a rod of length L. The rod is made of an insulating material, and it is charged with a linear charge density $\rho(x)$, for $x \in [0, L]$. The total charge Q stored in the rod is given by the integral:

$$Q = \int_0^L \rho(x)\ \mathrm{d}x \tag{4.4}$$

In terms of the Riemman sum, we visualize the rod as a set of tiny point masses with charge $\rho(\tilde{x}_k)(x_k - x_{k-1})$ glued together, each contributing a small amount to the total charge of the rod. But what if someone gave us a charged sheet instead? Suppose you are also given an area charge density $\rho(x, y)$, how would you calculate the total charge of this sheet? Well, there's no reason why the divide and conquer strategy shouldn't work here as well. Let's get to work!

Suppose the sheet is represented by a region $R \in \mathbb{R}^2$. We will focus on the rectangle $A = [x_0, x_m] \times [y_0, y_n]$ that "bounds" the region. It can be subdivided into a grid of smaller rectangles obtained from subdividing the intervals $[x_0, x_m]$ and $[y_0, y_n]$ as we did with the 1D case. For a visual aid, check Figure 4.2. We want to add up the contributions just from the rectangles that lie inside the region R, so we need to evaluate ρ at a select group of rectangles while avoiding the others. This is complicated in general, so as a computational trick, we will define an auxiliary function $\tilde{\rho}(x, y)$ that agrees with $\rho(x, y)$ inside R, but is zero outside R.[2] Then, choose points $(\tilde{x}_i, \tilde{y}_j)$ inside the (i, j)-rectangle. By analogy with our intuition for the single integral, the small amount of charge contributed from that rectangle is $\tilde{\rho}(\tilde{x}_i, \tilde{y}_j)\Delta x_i \Delta y_j$ (where $\Delta x_i = x_i - x_{i-1}$ and $\Delta y_j = y_j - y_{j-1}$).

[1] Note that the Riemman sum is not a function of σ, so there are some subtleties in how we define this limit. You don't need to worry about them for now. You just need to note that the Riemman sum is definitely constrained by σ.

[2] We know that $\rho(x, y)$ is well-defined in R, but we have no idea what it does outside R. Maybe it describes the charge density of another object, or maybe it's not defined there. In the general case, we need to be careful.

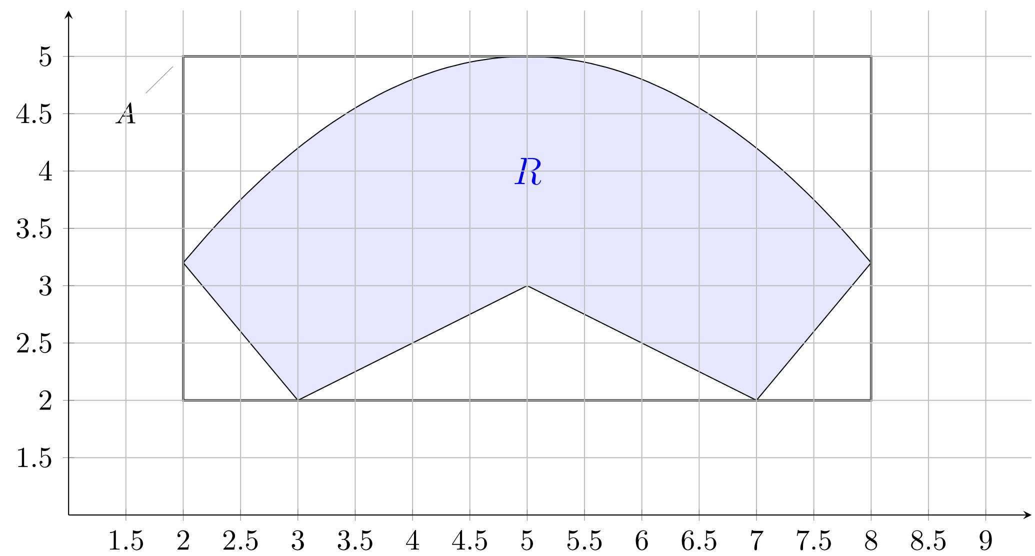


Figure 4.2: A set of small rectangles with a larger rectangle A bounding the region R.

Thus, we can approximate the total charge Q of the charged sheet by

$$Q \approx \sum_{i=1}^{m} \sum_{j=1}^{n} \tilde{\rho}(\tilde{x}_i, \tilde{y}_j) \Delta x_i \Delta y_j \tag{4.5}$$

At this point, you must be anticipating that to get the exact answer we just need to take the limit as $\sigma_x = \max_i\{\Delta x_i\}$ and $\sigma_y = \max_j\{\Delta y_j\}$ go to zero, like we did for the 1D case. However, just as we needed to be careful when taking limits in $\mathbb{R}^2$ (recall that the limit can't depend on the path taken towards the limit point), we need to be careful here. If we take the limit as $\sigma_x \to 0$ first, then $\Delta x_i \to 0$ and we end up with a bunch of rectangles of finite height and zero width. That amounts to adding zeros and our strategy fails. Similarly, it's not possible to take the limit as $\sigma_y \to 0$ first. We want both dimensions of the rectangles to become small "at the same time." A reasonable way of doing this is to declare that the largest dimension $\sigma = \max\{\sigma_x, \sigma_y\}$ of the rectangles must go to zero. Definition 4.1 is as precise as we can make this statement in this course. When the limit exists, f is said to be *integrable* over the region R. Note that being integrable depends on *both* f and R. Under different formalizations of how one actually takes this limit, the same function might be classified as integrable or not. Well-behaved functions might not be integrable if the region R is poorly-behaved, and vice versa. These pathological cases are not of interest right now, so from now on we will assume both R and f are "nice enough." In particular, if R is the union of finitely many elementary regions (as defined in the next section), then any continuous function f is integrable regardless of how you formally perform the limiting process.

4.2 Double Integrals

Based on the preceding discussion, we take Definition 4.1 to capture the basic notion of integrating over a region in $\mathbb{R}^2$. Note that the double integral is a fundamentally two-dimensional concept: the symbol $\iint$ is there just as an analogy to the one-dimensional case. As of right now, we have no idea how to compute double integrals for any reasonable function. A priori, you shouldn't look at them as two single integrals nested together. But it seems reasonable that they *should* be related to single integrals in some way. The results later in this section will indeed establish that this intuition is correct, providing a computational method to deal with multidimensional integrals. Before we engage with such powerful tools, we need first to look at these objects from a geometrical perspective.

Definition 4.1 (Double Integral)

If $R \subseteq \mathbb{R}^2$ is a finite region and $f : R \to \mathbb{R}$ is a function, then the double integral of f over the region R, is given by the following limit, when it exists:

$$\iint_R f(x, y)\ \mathrm{d}A \stackrel{\text{def}}{=} \lim_{\sigma \to 0} \sum_{i=1}^{m} \sum_{j=1}^{n} \tilde{f}(\tilde{x}_i, \tilde{y}_j)\Delta x_i \Delta y_j$$

What general feature of integral is illustrated by Figure 4.1? For each x, $f(x)$ is the height of the graph of the f at that point, so when we multiply it by the length of a thin rectangle in the Riemman sum and add all pieces together, we get the signed area determined by $f(x)$ on the interval. How does this generalize? Assume now $f(x, y)$ is a function $R \to \mathbb{R}$. For each tiny rectangle at (x, y), we are multiplying the value of a function $f(x, y)$, which is the height of the graph of f at (x, y). Thus, when we look at the graph of a two-variable function, its double integral over some region R is just the net volume covered by $f(x, y)$ in that region. Let's take this opportunity to do some programming and visualize things. Assume $f(x, y) = \sqrt{1 - x^2 - y^2}$, defined on the unit disk centered at the origin. D. We will compute an approximation for $\iint_D f(x, y)\ \mathrm{d}A$ using Python.

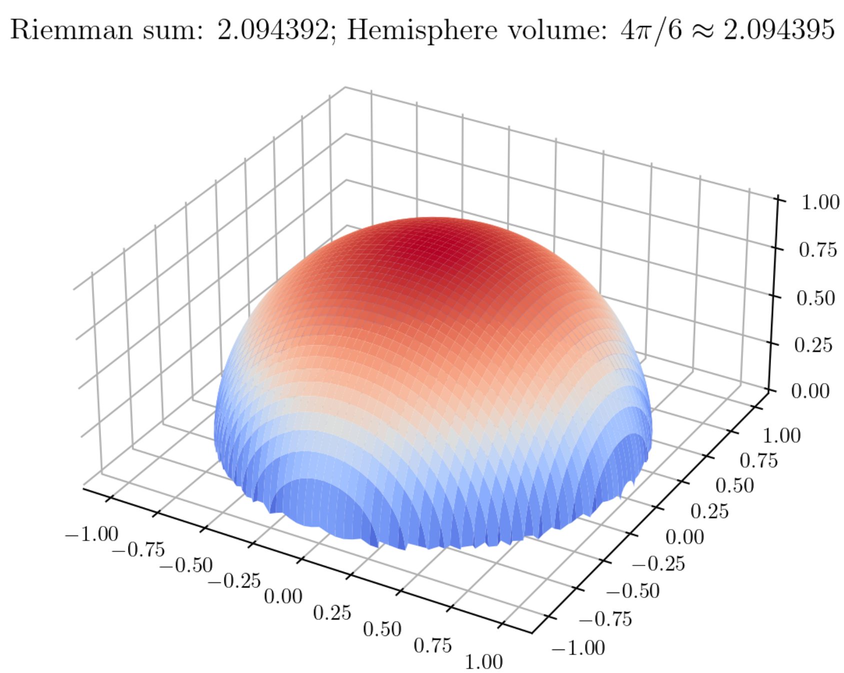


Figure 4.3: The hemisphere surface $z = \sqrt{1 - x^2 - y^2}$.

Using the code in the next page, we calculated the volume enclosed by the hemisphere surface and the xy-plane in Figure 4.3 to be 2.094392. From elementary geometry, we know the answer should be $4\pi/6 \approx 2.094395$. The error comes, of course, from the fact that the Riemman sum is just an approximation, not the exact answer. As we increase the number of sampling points (i.e. make the rectangles smaller), these errors go away and the sum gets closer to the correct value, which is the integral. Our task now will be to learn how to get the exact value using some additional facts about double integrals.

```python
import numpy as np
import matplotlib.pyplot as plt
from matplotlib import cm
# Use a 3d projection
fig, ax = plt.subplots( subplot_kw={"projection": "3d"},
                        figsize=(8,5), dpi=200)

# Increasing the number of sampling points will reduce
# the jagged edges around the equator and increasing accuracy
# of the integration.
sampling_points = np.linspace(-1, 1, 1000)

# X and Y are matrices containing the x and y coordinates of the sampling points
X, Y = np.meshgrid(sampling_points, sampling_points, indexing='ij')

# For each coordinate, we need not to sample at the last point as it
# belongs to the next interval [1, 1+delta]
Z = np.sqrt(1- X[:-1,:-1]*X[:-1,:-1] - Y[:-1,:-1]*Y[:-1,:-1])

surface = ax.plot_surface( X[:-1,:-1], Y[:-1,:-1], Z,
                           cmap=cm.coolwarm, linewidth=0,
                           antialiased=True)

ax.set_xticks([-1 + 0.25 *n for n in range(9)])
ax.set_yticks([-1 + 0.25 *n for n in range(9)])
ax.set_zticks([0.25 *n for n in range(5)])
ax.axes.set_aspect('equal')

# We took the sqrt of negative numbers many times and got NaN as a result,
# but we just want to ignore the function there, so we set it to 0
Z = np.nan_to_num(Z)

# Compute the sizes of each rectangle (these should be all the same
# except for issues with floating-point arithmetic, which is why
# we still have to compute them directly)

deltaX = np.diff(X, axis=0)
deltaY = np.diff(Y, axis=1)

# Compute Riemman sum
integral = np.sum(Z * deltaX[:,:-1] * deltaY[:-1,:])

fig.text(0.5, 0.9, s = fr'Riemman sum: ${integral:.6f}$; Hemisphere volume: $4\pi / 6 \approx {np.pi * 4/6:.6f}$', size=14, horizontalalignment='center')


fig.tight_layout()
fig.savefig('double_integral.png')
```

Let's examine the simplest case: when $R = [x_0, x_1] \times [y_0, y_1]$ is a rectangle. Imagine a function defined on this rectangle $f : R \to \mathbb{R}$. If we slice the volume V determined by the graph of f, we will obtain a collection of cross-sections whose area can be calculated using single-variable calculus: for a slicing along the x axis, at each x the area of the cross-section is $S_x = \int_{y_0}^{y_1} f(x, y)\, \mathrm{d}y$. Now, if we have a very dense slicing, then the volume can be approximated by adding up the volume of each tiny prism with base S_x and width Δx. The exact volume will be given by an integral: $\int_{x_0}^{x_1} S_x\, \mathrm{d}x$. Moreover, there was nothing special about slicing along the x axis, we could have done the same thing with the y axis. For a visual demonstration, check Figure **??**. Thus, in this case we expect a relationship between double and single integrals:

$$\iint_R f(x,y)\ \mathrm{d}A = \int_{x_0}^{x_1} \left(\int_{y_0}^{y_1} f(x,y)\ \mathrm{d}y \right)\ \mathrm{d}x = \int_{y_0}^{y_1} \left(\int_{x_0}^{x_1} f(x,y)\ \mathrm{d}x \right)\ \mathrm{d}y \tag{4.6}$$

This equivalence is nothing more than Cavalieri's principle, which was formulated in the early 17th century by Italian mathematician Bonaventura Cavalieri. Note that this principle predates the invention of integral calculus by a few decades, and had important influence in the work of Newton and Leibniz.

Theorem 4.2 (Cavalieri's Principle)

Let $\Omega_1, \Omega_2 \subseteq \mathbb{R}^3$ be finite regions in 3D. Pick a reference plane λ. If every plane parallel to λ intersects Ω_1 and Ω_2 in cross-sections of equal area, then the two regions have the same volume.

Definition 4.3 (Elementary Regions)

Theorem 4.4 (Fubini)

4.3 Iterated Integrals

4.4 Triple Integrals

4.5 Change of Variables

Chapter 5

Taylor's Theorem

5.1 Analytic Functions and Taylor Series

In this section we look to develop a method to represent functions as series. An important application of such is the use of series as solutions to differential equations.

Definition 5.2 (Power Series)

Let $x_0 \in \mathbb{R}$. A **power series** cenetered at x_0 is in the form

$$f(x) = \sum_{n=0}^{\infty} a_n (x - x_0)^n.$$

Remark. *The convention used for 0^0 when $x = x_0$ here is $0^0 = 1$. You can imagine that it is the limit of $(x - x_0)^0$ as $x \to x_0$.*

Remark. *The series might not converge. In fact, the root test for convergence gives an exact criteria for convergence.*

Definition 5.3 (Radius of Convergence)

Let $f(x)$ be a power series centered at x_0. The **radius of convergence** is the value R such that

$$\begin{cases} f(x) \text{ converges}, & \text{if } |x - x_0| < R, \\ f(x) \text{ diverges}, & \text{if } |x - x_0| > R. \end{cases}$$

Definition 5.4 (Analytic Functions)

Let $\Omega \subseteq \mathbb{R}$ be open, and $f : \Omega \to \mathbb{R}$. We say that f is **analytic** at x_0 if there exists $\epsilon > 0$, and a power series representation $p_{x_0}(x) = \sum_n a_n (x - x_0)^n$ such that $p_{x_0}(x)$ converges to $f(x)$ in $B(x_0, \epsilon)$. We say that f is analytic on (a,b) if f is analytic at every point in (a, b).

Proposition 5.5

The set of points on which f is analytic form an open set.

Being analytic is one of the strictest properties for a function. Analytic functions are infinitely differentiable (smooth).

Theorem 5.6 (Analytic Functions are Smooth)

Suppose $\sum a_n x^n$ is a power series representation for some function f with a radius of convergence $R > 0$. Then f is infinitely differentiable on $(-R, R)$.

The proof requires some analysis knowledge out of scope of the course. The hardest part is to show that you can differentiate under the summation, i.e.

$$\frac{d}{dx}\sum_n f_n(x) = \sum_n \frac{d}{dx} f_n(x).$$

Assuming this, we can get power series representations for $f'(x)$ and so on.

$$\begin{array}{llllllll}
f(x) = & a_0 & +a_1(x-x_0) & +a_2(x-x_0)^2 & +a_3(x-x_0)^3 & +a_4(x-x_0)^4 & +\dots \\
f'(x) = & & a_1 & +2a_2(x-x_0) & +3a_3(x-x_0)^2 & +4a_4(x-x_0)^3 & +\dots \\
f''(x) = & & & 2a_2 & +6a_3(x-x_0) & +12a_4(x-x_0)^2 & +\dots
\end{array}$$

Importantly, these have the same radius of convergence as the original function (a root test can confirm this), so we get a closed form for the k-th derivative of f:

$$f^{(k)}(x) = \sum_{n=k}^{\infty} \frac{n!}{(n-k)!} a_n (x-x_0)^{n-k}$$

Theorem 5.7 (Uniqueness of Power Series)

Let f be analytic at x_0. Then its power series representation $\sum_n a_n(x-x_0)^n$ is unique with coefficients

$$a_k = \frac{f^{(k)}(x_0)}{k!}.$$

Proof. We take the general form of the k-th derivative, and evaluate it at $x = x_0$. This gives us

$$f^{(k)}(x_0) = \sum_{n=k}^{\infty} \frac{n!}{(n-k)!} a_n (x_0-x_0)^{n-k}.$$

On the right hand side, all the terms with $n > k$ will evaluate to 0. The term with $n = k$ evaluates to $k!a_k$. This means

$$f^{(k)}(x_0) = k!a_k \implies a_k = \frac{f^{(k)}(x_0)}{k!}.$$

Therefore, if a power series exists, it must be in the form

$$\sum_n \frac{f^{(n)}(x_0)}{n!}(x-x_0).$$

$\square$

Definition 5.8 (Taylor Series)

The **Taylor expansion** of f centered at x_o is given by

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_o)}{n!}(x - x_o)^n.$$

A function locally equals to its Taylor series (about some point) if and only if it is analytic. We will see that if a function is analytic, it is equal to the Taylor series about a point everywhere the series converges.

Theorem 5.9 (Uniqueness of Analytic Functions)

Let $f, g : (a, b) \to \mathbb{R}$ be analytic. Suppose $f(x) = g(x)$ on some small ball $B(x_0, \epsilon)$. Then $f = g$ everywhere on (a, b).

Let's consider the function $h(x) = f(x) - g(x)$. This is analytic, as we can just take the difference of the respective coefficients in the power series for f and g. We have $h(x) = 0$ on $B(x_0, \epsilon)$. Our task now is to show that $h(x) = 0$ everywhere, so $f = g$ everywhere.

The idea now is to start with the power series about $x = x_0$

$$\sum_n 0(x - x_0)^n.$$

We can 'slide' this x_0 across a little bit to $x_1 \in B(x_0, \epsilon)$ to get

$$\sum_n 0(x + x_1 - x_0 - x_1)^n = \sum_n 0(x - x_1)^n$$

after binomial expansion of the terms $((x - x_1) + (x_1 - x_0))^n$. Since f is analytic at x_1, there is another ball centered at x_1 where $h = 0$. Therefore, we can 'slide' our center of the power series from x_1 to another x_2. If we can slide this to everywhere in (a, b) we will get that the power series representation at every point is 0.

optional material: How do we guarantee that we can slide everywhere? This requires another idea from analysis called compactness. In short, the compactness of the interval $[x_0, \tilde{x}]$ or $[\tilde{x}, x_0]$ guarantees that we can slide our center of the power series across from x_0 to any $\tilde{x} \in (a, b)$ in a finite amount of steps.

We will give another way to prove this, as we have introduced Zorn's lemma. Without loss of generality, let $x_0 < \tilde{x} \in (a, b)$ Consider S, the set of points $x \leq \tilde{x}$ that you can 'slide to' from x_0 in a finite amount of steps. I claim that every increasing sequence in S is bounded above by some element in S. Let $x_1 \leq x_2 \leq x_3 \leq \ldots$ be an increasing sequence in S. We take $y = \lim_{n\to\infty} x_n$, then the series is bounded above by y. To construct the finite sequence going from x_0 to y, we see that f is analytic at y thus it has a power series representation centered at y that converges to f for some $B(y, \delta)$. We take m large such that $x_m > y - \delta/4$. If we slide the power series centered at y to be centered at x_m, the power series converges to f at least in $B(x_m, 3\delta/4) \ni y$. That is, you can recenter the power series from x_m to y. Therefore, we take the finite sequence that recenters the power series at x_0 to x_m, then recenter that sequence at y. Therefore S contains a maximal element by Zorn's lemma. Finally, to find out what this maximal element is, we make use of the fact that the points where f is analytic is open. Therefore, the only point that can be the maximal element of

S is $\tilde{x}$, which is used as the upper limit of all elements in S. Therefore $\tilde{x}$ is the maximal element in S, thus $f = 0$ in some ball centered at $\tilde{x}$. We picked $\tilde{x}$ to be arbitrary, so $f = 0$ everywhere in (a, b).

Corollary 5.10: Let $x_0 \in (a, b)$, and f is analytic on (a, b), f equals the power series centered at x_0 where the power series converges.

Proof. The power series is analytic, and equals f on some small open ball in (a, b). □

Exercises

1. Find the Taylor Series for the given functions at the indicated points.
   (a) $f(x) = e^{-x}, x_0 = 0$.
   (b) $f(x) = e^x, x_0 = 1$.
   (c) $f(x) = 1/x, x_0 = 1$.
   (d) $f(x) = \cos(x), x_0 = \pi/2$.
   (e) $f(x) = \ln(x), x_0 = 1$.
2. Determine the radius of convergence of the given function about $x = 0$.
   (a) $f(x) = (1 + x)/(x - 2)$.
   (b) $f(x) = 2x/(1 + 2x^2)$.
   (c) $f(x) = 1/(1 - t^3)$.
   (d) $f(x) = ((t - 4)(t^2 + 3))^{-1}$.

5.2 Taylor's Theorem with Remainder

Sometimes we don't want to take the whole power series representation, but truncate the series to get an approximation for the functions.

TODO: show that the n-th order taylor series is the best n-th order approximation for a function. As Taylor Series are used to approximate functions, it is of relevance to determine the accuracy of a series in representing its desired function.

Definition 5.11 (Taylor's Formula with Remainder)

The remainder of order n of the Taylor expansion of $f(x_o)$ is represented by the function,

$$R_n(x) = f(x) - \sum_{k=0}^{n-1} \frac{f^{(k)}(x_o)}{k!}(x - x_o)^k.$$

The remainder of a Taylor expansion is the difference between the value of a function f at x and the partial sum of the n^{th} term Taylor series. The series converges if $\lim_{n\to\infty} R_n = 0$.

Theorem 5.12 ()

Let $f(x)$ be a function on the interval (a, b). f is analytic on (a, b) if there exists and $M > 0$ such that

$$|f^{(n)}(x) \leq M^n$$

for all $x \in (a, b)$ and $n \in \mathbb{N}$.

As a result of this theorem, the Taylor series expansion holds for all $x \in (a, b)$.

Proof. Let $f(x)$ be a function on the interval (a, b) and $x_o \in (a, b)$. For some $M \in \mathbb{R}$ set $C = maxM|a - x_o, M|b - x_o|$. Then the n^{th} term remainder of the Taylor expansion of $f(x)$ at x_o is given by

$$R_n = f(x) - \sum_{k=0}^{n-1} \frac{f^{(k)}(x_o)}{k!}(x - x_o)^k = \sum_{k=n}^{\infty} \frac{f^{(k)}(x_o)}{k!}(x - x_o)^k$$

Each term in this infinite series for R_n is given by

$$R_k = \frac{f^{(k)}(x_o)}{k!}(x - x_o)^k \leq \frac{M^k}{k!}$$

5.3 The Binomial Theorem

The Binomial Theorem describes the expansion of powers of a binomial expression. A binomial is an algebraic expression consisting of two terms, such as $(a + b)$. The Binomial Theorem provides a way to expand expressions of the form $(a + b)^n$, where n is a non-negative integer. The theorem can be stated as follows:

Theorem 5.13 (Binomial Theorem)

For any integer $n \geq 0$, the expansion of $(a + b)^n$ is given by:

$$(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$$

where $\binom{n}{k}$ is the binomial coefficient, defined as:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

for $k = 0, 1, 2, \ldots, n$.

In the expansion $(a + b)^n$, the sum consists of $n + 1$ terms, where each term has the form $\binom{n}{k} a^{n-k} b^k$. The key components of the expansion are:

- The *binomial coefficient* $\binom{n}{k}$, also called a combination, represents the number of ways to choose k items from n items, and is calculated using the formula:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- The powers of a and b decrease and increase respectively in each term, starting from a^n for $k = 0$ to b^n for $k = n$.
- The sum runs over all integer values of k from 0 to n.

Example 5.14

Find the binomial expansion for $(a + b)^2$ where $a, b \in \mathbb{R}$.

Using the Binomial Theorem, we can expand $(a + b)^2$ as follows:

$$(a+b)^2 = \sum_{k=0}^{2} \binom{2}{k} a^{2-k} b^k$$

This gives the following terms:

$$= \binom{2}{0} a^2 b^0 + \binom{2}{1} a^1 b^1 + \binom{2}{2} a^0 b^2$$

Using the binomial coefficients:

$$= 1 \cdot a^2 + 2 \cdot ab + 1 \cdot b^2$$

Thus, the expanded form is:

$$(a+b)^2 = a^2 + 2ab + b^2$$

Example 5.15

Find the binomial expansion of $(x + y)^3$.

For $(x + y)^3$, we apply the Binomial Theorem:

$$(x+y)^3 = \sum_{k=0}^{3} \binom{3}{k} x^{3-k} y^k$$

This expands to:

$$= \binom{3}{0} x^3 y^0 + \binom{3}{1} x^2 y^1 + \binom{3}{2} x^1 y^2 + \binom{3}{3} x^0 y^3$$

Substituting the binomial coefficients:

$$= 1 \cdot x^3 + 3 \cdot x^2 y + 3 \cdot xy^2 + 1 \cdot y^3$$

Thus, the expanded form is:

$$(x+y)^3 = x^3 + 3x^2 y + 3xy^2 + y^3$$

Example 5.16

Find the binomial expansion of $(1 + \epsilon)^n$ for small ϵ and $n \in \mathbb{N}$.

For $(1 + \epsilon)^n$, we apply the Binomial Theorem with $a = 1$ and $b = \epsilon$:

$$(1+\epsilon)^n = \sum_{k=0}^{n} \binom{n}{k} 1^{n-k} \epsilon^k$$

This simplifies to:

$$= \binom{n}{0} + \binom{n}{1}\epsilon + \binom{n}{2}\epsilon^2 + \cdots + \binom{n}{n}\epsilon^n$$

The expansion provides the terms of $(1+\epsilon)^n$ for small values of ϵ. Taking an approximation for small ϵ, we can truncate the series to obtain an approximation.

$$(1+\epsilon)^n \approx 1 + n\epsilon.$$

This result is useful for many physical applications where small perturbations are considered.

5.3.1 Properties of Binomial Coefficients

The binomial coefficients $\binom{n}{k}$ have several important properties:

- $\binom{n}{0} = \binom{n}{n} = 1$
- $\binom{n}{k} = \binom{n}{n-k}$ (Symmetry property)
- The sum of the binomial coefficients for a given n is:

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n$$

  This is known as the binomial identity.

- Pascal's identity:

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

5.3.2 Negative and Non-Integer Exponents

The Binomial Theorem can also be extended to cases where the exponent n is not a non-negative integer, although the series then becomes infinite. For example, if n is a positive integer, the expansion of $(1+x)^n$ converges to a finite sum, but if n is a negative integer or a fraction, the series may converge to an infinite sum. For $|x| < 1$, the Binomial Theorem for any real number n can be written as:

$$(1+x)^n = \sum_{k=0}^{\infty} \binom{n}{k} x^k$$

where $\binom{n}{k}$ is generalized as:

$$\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}$$

5.4 Multidimensional Taylor Series

The Taylor series is a powerful tool for approximating functions using polynomials. While the standard Taylor series applies to functions of a single variable, the *multidimensional Taylor series* extends this concept to functions of multiple variables. In this section, we will derive and explain the multidimensional Taylor series for a function of several variables. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function that is sufficiently smooth (i.e., has continuous partial derivatives) in a neighborhood of a point

$\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$. The goal is to approximate f near the point $\mathbf{a}$ using a polynomial expansion. Suppose we have a function $f(x_1, x_2, \ldots, x_n)$ of n variables, the Multidimensional Taylor series expansion around a point $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ is given by

$$f(x_1, x_2, \ldots, x_n) = f(a_1, a_2, \ldots, a_n) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a_1, a_2, \ldots, a_n)(x_i - a_i) + \cdots$$

$$+\frac{1}{k!} \sum_{i_1, i_2, \ldots, i_k = 1}^{n} \frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_k}}(a_1, a_2, \ldots, a_n)(x_{i_1} - a_{i_1})(x_{i_2} - a_{i_2}) \cdots (x_{i_k} - a_{i_k}) + \cdots$$

where the sums are taken over all possible combinations of partial derivatives of f.

Definition 5.17 (Multidimensional Taylor Series)

The general form of the Taylor series for a function $f \colon \mathbb{R}^n \to \mathbb{R}$ about a point $\mathbf{a}$ is

$$f(\mathbf{x}) = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{i_1, i_2, \ldots, i_k = 1}^{n} \frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_k}}(\mathbf{a}) \prod_{j=1}^{k} (x_{i_j} - a_{i_j})$$

We will now examine the first and second-order approximations of the function using the Taylor series. The first-order approximation, also known as the linearization of the function f around the point $\mathbf{a}$, is obtained by truncating the series after the first derivative term:

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{a})(x_i - a_i)$$

This approximation gives a linear model of the function near the point $\mathbf{a}$. The second-order approximation includes the terms up to the second derivative. It is given by:

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{a})(x_i - a_i) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a})(x_i - a_i)(x_j - a_j)$$

This provides a quadratic approximation to the function near $\mathbf{a}$, and is useful for understanding the curvature of the function around the point. Quadratic approximations are especially useful in physical applications as it indicates the stability of equilibria of dynamical systems.

Example 5.18

Compute the second-order Taylor approximation of $f(x, y) = e^{x^2+y^2}$ around the point $(0, 0)$.

First, compute the necessary partial derivatives:

$$f(x, y) = e^{x^2+y^2}$$

$$\frac{\partial f}{\partial x} = 2xe^{x^2+y^2}, \quad \frac{\partial f}{\partial y} = 2ye^{x^2+y^2}$$

$$\frac{\partial^2 f}{\partial x^2} = 2e^{x^2+y^2} + 4x^2 e^{x^2+y^2}, \quad \frac{\partial^2 f}{\partial y^2} = 2e^{x^2+y^2} + 4y^2 e^{x^2+y^2}$$

$$\frac{\partial^2 f}{\partial x \partial y} = 4xye^{x^2+y^2}$$

Evaluating these derivatives at $(0,0)$:

$$f(0,0) = e^0 = 1$$

$$\frac{\partial f}{\partial x}(0,0) = 0, \quad \frac{\partial f}{\partial y}(0,0) = 0$$

$$\frac{\partial^2 f}{\partial x^2}(0,0) = 2, \quad \frac{\partial^2 f}{\partial y^2}(0,0) = 2, \quad \frac{\partial^2 f}{\partial x \partial y}(0,0) = 0$$

Now, the second-order Taylor expansion around $(0,0)$ is:

$$f(x,y) \approx 1 + 0 \cdot (x-0) + 0 \cdot (y-0) + \frac{1}{2}\left[2(x-0)^2 + 2(y-0)^2\right]$$

$$f(x,y) \approx 1 + (x^2 + y^2)$$

Thus, the second-order approximation for $f(x,y) = e^{x^2+y^2}$ around $(0,0)$ is:

$$f(x,y) \approx 1 + x^2 + y^2$$

5.4.1 Error in the Taylor Expansion

The error in the Taylor series approximation is related to the remainder term $R_n(\mathbf{x})$, which represents the difference between the exact value of the function and its approximation up to the n-th degree. For a multivariable Taylor series, the remainder term can be written as:

$$R_n(\mathbf{x}) = \frac{1}{(n+1)!} \sum_{i_1, i_2, \ldots, i_{n+1}} \frac{\partial^{n+1} f}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_{n+1}}}(\mathbf{a}) \prod_{j=1}^{n+1} (x_{i_j} - a_{i_j})$$

This error term quantifies the difference between the approximation and the exact value of the function.

5.5 Extrema of Multivariate Functions

Just as we are interested in finding the extreme points of functions of a single variable, we likewise wish to solve for stationary points of multivariate functions. Analysis of functions of multiple variables have analogous first and second derivative tests to those learned in single variable calculus. I used the term 'stationary point' as in three dimensions in addition to maxima or minima there exist so called saddle points. A 3D representation of a saddle point is presented in figure

Definition 5.19 (Multivariable First Derivative Test)

A point $(x_o, y_o) \in \mathbb{R}^{\nvDash}$ is a *stationary point* of some function $f(x,y)$ if

$$\nabla f|_{(x,y)=(x_o,y_o)} = 0$$

Also similarly to the second derivative test in single variable calculus, we also have an analogous second derivative test in multivariable calculus to determine the classification of critical points. For this we use the discussion of multivariable Taylor series discussed in section . To second order, the Taylor expansion of some function $f(x, y)$ around (x_o, y_o) is

$$f(x, y) \approx f(x_o, y_o) + \nabla f(x_o, y_o)^T d + \frac{1}{2!} d^T H f(x_o, y_o) d + R_2(x, y)$$

From the first derivative test $\nabla f(x_o, y_o) = 0$ thereby eliminating that term. Also, we know $R_2 \to 0, (x, y) \to (x_o, y_o)$. Thus we have

$$f(x, y) \approx f(x_o, y_o) + \frac{1}{2!} d^T H f(x_o, y_o) d$$

Rearranging we have

$$f(x, y) - f(x_o, y_o) = \frac{1}{2!} d^T H f(x_o, y_o) d$$

The left side appears as the numerator of the definition of the derivative. The sign of our derivative is dependent on the Hessian matrix H which holds for all points (x, y) near (x_o, y_o).

5.6 Lagrange Multipliers

Lagrange multipliers are a method to find the extrema of a function subject to a constraint. Suppose we have a function $f(x, y)$ and a constraint $g(x, y) = c$. The method of Lagrange multipliers states that the extrema of $f(x, y)$ subject to the constraint $g(x, y) = c$ are found at points (x, y) where the gradient of f is parallel to the gradient of g. This implies

$$\nabla f(x, y) = \lambda \nabla g(x, y) \tag{5.1}$$

$$\begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \lambda \begin{bmatrix} \frac{\partial g}{\partial x} \\ \frac{\partial g}{\partial y}. \end{bmatrix} \tag{5.2}$$

The theory behind Lagrange multipliers is that we wish to find the points where $\nabla f \cdot g = 0$. This implies that the gradient of f is at some maximum or minimum (or stationary point) when subject to g. We can use the system of equations (1.2) and our constraint equation to solve for our three unknowns: λ, x, and y. This result will gives us the extrema we are looking for. This method can be generalized to any n-dimensional function $f(\mathbf{x})$ subject to a constraint $g(\mathbf{x}) = c$. In this case we will have a system with $n + 1$ equations and $n + 1$ unknowns to solve for the extrema.

Example 5.20

Find the dimensions of the box with the largest volume such that its surface area is 100 square units.

Let x, y, and z be the dimensions of the box. The volume of the box is $V(x, y, z) = xyz$ and the surface area is $S(x, y, z) = 2xy + 2yz + 2xz = 100$. We now apply our method of Lagrange Multipliers to maximize V subject to S. We have the system of equations

$$\begin{bmatrix} yz \\ xz \\ xy \end{bmatrix} = \lambda \begin{bmatrix} 2y + 2z \\ 2x + 2z \\ 2x + 2y \end{bmatrix}$$

$$2xy + 2yz + 2xz = 100.$$

Solving this system of equations we find

$$x = y = z = 5$$

$$\lambda = \frac{1}{5}.$$

Therefore, the dimensions of the box with the largest volume are 5 units by 5 units by 5 units. The corresponding maximum volume is 125 cubic units.

Chapter 6

First Order Differential Equations

6.1 Introduction

A differential equation is an equation that relates an undetermined function with one or more of its derivatives. We call equations involving only single-variable derivatives of functions *ordinary differential equations* (ODEs) and those containing partial derivatives of multivariable functions *partial differential equations*(PDEs). We will focus on the former in this course and leave study of the latter to MATH 381.

The highest order derivative occurring in an ODE defines the *order* of the differential equation. We will look at first and second order ordinary differential equations. ODEs can be either homogeneous or inhomogeneous. Homogeneous equations have all terms involving the function y or derivatives of y summed to equal 0, while inhomogeneous equations will sum to equal a nonzero term.

$$\text{Homogeneous} : f(y, y', ..., y^n, t) = 0$$

$$\text{Inhomogeneous} : f(y, y', ..., y^n) = g(t)$$

Another classification of differential equations is concerned with the linearity of the terms. We can have either linear or nonlinear equations. An ODE

$$f(y, y', ...y^n, t) = g(t)$$

is linear if f is linear with respect to terms involving the variable y or derivatives of y. The general form looks like

$$a_0(t)y^n + ... + a_n(t)y = g(t)$$

Nonlinear equations will typically have terms involving y or derivatives of y multiplied together or terms involving nonlinear functions of y such as $sin(y)$ or e^y.

6.2 Separation of Variables

A separable differential equation is any differential equation of the form,

$$N(y)\frac{dy}{dt} = M(t)$$

This allows us to multiply across by dt and integrate both sides to find a function $y(t)$.

$$\int N(y(t))\frac{dy}{dt}dt = \int M(t)dt$$

I have written $N(y) = N(y(t))$ since y is a function of t. Then we can suppose that $\frac{d}{dt}(y(t)) = N(y(t))\frac{dy}{dt}$. Which leads to the conclusion

$$y(t) = \int M(t)dt + C$$

for some constant C. You may have seen the differential treated as a fraction that can be separated and while that is sufficient for all computation purposes and will lead to the same answer, the formulation above is more mathematically rigorous.

6.3 Differential Forms

Let's take a look back at section (). The line integral of a vector field F along a curve C is defined as

$$\int_C F \cdot d\vec{r} = \int_C Mdx + Ndy$$

where $F = \langle M, N \rangle$. Given a the pair of functions $M(x, y)$ and $N(x, y)$, the expression

$$Mdx + Ndy$$

is called a *differential form*. Differential forms are generalizations of derivatives and integrals to manifolds. They are used to define integrals on curves, surfaces, volumes, and higher-dimensional geometries. A differenti 0-form is a smooth function. A differential 1-form is a linear map that takes a vector field as input and returns a function. In $\mathbf{R}^3$ this is of the form,

$$\omega = Pdx + Qdy + Rdz$$

where P, Q, R are smooth functions of x, y, z. A differential 2-form is a linear map that takes two vector fields as input and returns a function. In $\mathbf{R}^3$ this is of the form,

$$\omega = Adxdy + Bdydz + Cdzdx.$$

In general, a differential k-form is a linear map that takes k vector fields as input and returns a function. Expressed for $\mathbf{R}^n$ this is of the form,

$$\omega = \sum_{i_1<i_2<...<i_k} A_{i_1,i_2,...,i_k} dx_{i_1} dx_{i_2} ... dx_{i_k}.$$

In higher level mathematics, the wedge product $\wedge$ is used to define differential forms, however for our purposes how we have defined them is sufficient. Further study of differential forms is left to a differential geometry course.

We return to differential 2-forms in the context of first order differential equations. Suppose we have $F = \langle M, N \rangle$. Then,

$$Mdx + Ndy = F \cdot d\vec{r}.$$

We know that a vector field F is conservative if it is the gradient of a scalar field f. That is, $F = \nabla f$. Then, $F \cdot d\vec{r} = df$. Thus, if F is conservative, then $Mdx + Ndy = df$. This suggests,

$$\frac{\partial f}{\partial x} = M \quad \text{and} \quad \frac{\partial f}{\partial y} = N$$

such that,

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy.$$

Therefore, F is conservative if and only if $Mdx + Ndy$ is equal to the differential of some function f. We define such forms as exact. Recall from section (), some vector field $F = \langle M, N\rangle$ is conservative if,

$$\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}. \tag{6.1}$$

In terms of differential forms, this is equivalent to saying $Mdx + Ndy$ is closed. One must be careful here, every exact form is closed because all conservative vector fields satisfy 2.1, however not all closed forms are exact. That is, some vector fields may satisfy 2.1 but not be conservative.

6.4 Exact Equations

We now look to make use of the concept of differential forms to solve first order differential equations. Consider the differential equation,

$$\frac{dy}{dx} = f(x, y).$$

We can rewrite this as,

$$Mdx + Ndy = 0$$

Suppose $F = \langle M, N\rangle$ is exact, then

$$\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}.$$

This means there exists some function f such that $\nabla f = F = \langle M, N\rangle$. Section () tells us that exact forms correspond to conservative vector fields. Recall from our study of line integrals that the line integral of a conservative vector field F is $f(b) - f(a)$ where $\nabla f = F$ and a and b are the endpoints of the curve. Therefore, we can integrate our differential form

$$\int Mdx + Ndy$$

to find $f(x, y)$. In practice, we integrate Mdx and Ndy separately. This will gives us

$$\int Mdx + g(y)$$

and

$$\int Ndy + h(x).$$

where $g(y)$ and $h(x)$ are functions of y and x respectively. This is because the derivative of a function of x with respect to y is 0 and vice versa. We can put together the two resulting functions to find $f(x, y)$.

Example 6.2

$$\frac{dy}{dx} = -\frac{2x + y}{x + 2y}$$

We can rewrite this as,

$$(x + 2y)dy + (2x + y)dx = 0.$$

We will apply the screening test to see if this is an exact equation.

$$\frac{\partial M}{\partial y} = 1 \quad \text{and} \quad \frac{\partial N}{\partial x} = 1.$$

Therefore, this is an exact equation. We can integrate to find $f(x, y)$.

$$\int (x + 2y)dy = xy + y^2 + g(x)$$

and

$$\int (2x + y)dx = x^2 + xy + h(y).$$

We can put these together to find $f(x, y)$.

$$f(x, y) = xy + y^2 + x^2 + xy = x^2 + 2xy + y^2.$$

6.5 Integration Factors

The above formulation gives a useful technique for solving exact differential equations, however we must now consider the case where the equation is not exact. The technique we will use if called the method of integrating factors. The idea is to multiply the equation by some function $\mu(x, y)$ such that the resulting equation is exact. That is, we want to find $\mu(x, y)$ such that,

$$\mu(x, y)Mdx + \mu(x, y)Ndy = 0$$

with

$$\frac{\partial}{\partial y}(\mu M) = \frac{\partial}{\partial x}(\mu N).$$

Using chain rule we can rewrite this as,

$$\mu\frac{\partial M}{\partial y} + M\frac{\partial \mu}{\partial y} = \mu\frac{\partial N}{\partial x} + N\frac{\partial \mu}{\partial x}.$$
$$\mu[\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}] = N\frac{\partial \mu}{\partial x} - M\frac{\partial \mu}{\partial y}.$$

Unfortunately, this is a partial differential equation and is not easy to solve. However, we can make use of the fact that μ is a function of x and y. We can rewrite the above equation as,

$$\frac{\partial \mu}{\partial x} = \frac{1}{N}[\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}]\mu$$
$$\frac{\partial \mu}{\partial y} = -\frac{1}{M}[\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x}]\mu.$$

We can treat these as two separable first order differential equations.

$$\frac{\partial \mu}{\mu} = \int$$

6.6 Variation of Parameters

Following our discussion of first order homogeneous differential equations, we now move on to discussing methods of findings solutions to inhomogeneous first order differential equations.

$$\frac{dy}{dx} + a(x)y = b(x)$$

We propose a solution $y(x) = u(x)h(x)$ where $h(x)$ is the solution to the corresponding homogeneous equation.

$$\frac{dy}{dx} + a(x)y = 0$$

The solution to this equation is

$$h(x) = e^{-\int a(x)dx}.$$

Going back to our solution form $y(x) = u(x)h(x)$ and substituting into our inhomogeneous equation

$$\frac{du}{dx}h + u\frac{dh}{dx} + a(x)uh = b(x)$$

$$\frac{du}{dx}h + u\left(\frac{dh}{dx} + a(x)h\right) = b(x)$$

Since $h(x)$ is a solution to the homogeneous equation, the term in the parenthesis vanish. Therefore our differential equation becomes

$$\frac{du}{dx} = \frac{b}{h}$$

Solving for u we get

$$u = \int \frac{b(x)}{h(x)}dx$$

Lastly, multiplying by $h(x)$ to get our full solution $y(x)$

$$y(x) = h(x)\left(\int \frac{b(x)}{h(x)}dx + C\right)$$

Notice here that I have already included the constant of integration here. This is because the method of solving inhomogeneous differential equations often settles down to combining a general and particular solution. We see that the constant multiplied by $h(x)$ will give us a general solution to the homogeneous equation while the product of the term in the integral and $h(x)$ will give a particular solution.

6.7 Existence and Uniqueness of Solutions for First-Order ODEs

The study of first-order ordinary differential equations often revolves around understanding whether a solution exists for a given equation and, if so, whether that solution is unique. This section presents the fundamental results regarding existence and uniqueness, along with illustrative examples. A first-order ODE is typically written as:

$$\frac{dy}{dx} = f(x, y), \quad y(x_0) = y_0, \tag{6.2}$$

where $f(x, y)$ is a given function, and $y(x_0) = y_0$ specifies an initial condition at $x = x_0$. The goal is to determine whether there exists a function $y(x)$ satisfying Equation (6.2) and whether this solution is unique in a neighborhood of x_0. The classical result addressing this problem is the Existence and Uniqueness Theorem, often attributed to Picard-Lindelöf. The theorem is stated as follows:

Theorem 6.3 (Existence and Uniqueness)

Let $f(x, y)$ be a continuous function defined on a rectangle $R = \{(x, y) : a \leq x \leq b, c \leq y \leq d\}$ in the xy-plane. Suppose further that $f(x, y)$ satisfies a Lipschitz condition in y; that is, there exists a constant $L > 0$ such that

$$|f(x, y_1) - f(x, y_2)| \leq L|y_1 - y_2| \tag{6.3}$$

for all $(x, y_1), (x, y_2) \in R$. Then, for any point $(x_0, y_0) \in R$, there exists a unique solution $y(x)$ to the initial value problem (6.2) defined on some interval $x_0 - \delta < x < x_0 + \delta$, where $\delta > 0$.

Proof. Consider the function $y(x)$ defined and continuous on the interval $x_0 - \delta < x < x_0 + \delta$ for some $\delta > 0$ with the initial condition $y(x_0) = y_0$. Hence, the function $f(x, y(x))$ is well-defined and continuous on this interval. Suppose $f(x, y(x)) = y'(x)$. We can integrate both sides of the differential equation to obtain

$$y(x) = y(x_0) + \int_{x_0}^{x} f(t, y(x))dt. \tag{6.4}$$

Any solution $y(x)$ to 2.4 satisfies the original differential equation. We can differentiate both sides with respect to x to recover the original differential equation. Now we move on to proving uniqueness. For any two numbers a_1 and a_2 in the interval $x_0 - \delta < x < x_0 + \delta$, it follows from the Mean Value Theorem that

$$\frac{f(s, a_1) - f(s, a_2)}{a_1 - a_2} = \frac{\partial f}{\partial y}(s, z), \tag{6.5}$$

for some z between a_1 and a_2. Equation 2.5 can be rewritten as

$$|f(s, a_1) - f(s, a_2)| \leq L|a_1 - a_2|, \tag{6.6}$$

for all a_1, a_2 and x_1, x_2 in the interval $x_0 - \delta < x < x_0 + \delta$.
In other words, the conditions of the theorem states,
1. The continuity of $f(x, y)$ ensures the existence of solutions. Intuitively, if f is not continuous, the differential equation may exhibit abrupt changes that preclude the formation of a well-defined

solution.
2. The Lipschitz condition guarantees uniqueness. This condition implies that $f(x, y)$ does not change too rapidly with respect to y, preventing the trajectories of different solutions from crossing.

Example 6.4

Consider the initial value problem

$$\frac{dy}{dx} = 2x + 3y, \quad y(0) = 1.$$

Here, $f(x, y) = 2x + 3y$. Since $f(x, y)$ is a linear function of y, it satisfies the Lipschitz condition with $L = 3$. Furthermore, $f(x, y)$ is continuous everywhere. Hence, by the Existence and Uniqueness Theorem, there exists a unique solution.

Example 6.5

Consider the initial value problem

$$\frac{dy}{dx} = y^{1/3}, \quad y(0) = 0.$$

Here, $f(x, y) = y^{1/3}$ is continuous but does not satisfy the Lipschitz condition at $y = 0$. Multiple solutions exist, such as $y(x) = 0$ and $y(x) = \left(\frac{2}{3}x\right)^{3/2}$. This example highlights the necessity of the Lipschitz condition for uniqueness.

Chapter 7

Second Order Differential Equations

7.1 Introduction

Our study of differential equations continues with second order equations. Such equations are defined by their highest order derivative of the function in question being a second derivative. They are of the form

$$a(t)y'' + b(t)y' + c(t)y = f(t),$$

where $a(t)$, $b(t)$, $c(t)$, and $f(t)$ are functions of t. We will study methods to solve these equations, both homogeneous and inhomogeneous, and discuss the existence and uniqueness of solutions.

7.2 Constant Coefficients

The first technique we will study in solving second order differential equations is for cases of homogeneous equations with constant coefficients. Such equations are of the form

$$ay'' + by' + cy = 0$$

This equation suggests we are looking for solutions $y(t)$ for which the derivatives can be easily summed together to produce zero. Methods in calculus suggests the solution

$$y(t) = e^{rt}$$

Let's suppose this is the case, then

$$y'(t) = re^{rt}$$

$$y''(t) = r^2e^{rt}$$

Substituting these into our differential equation

$$a(r^2e^{rt}) + b(re^{rt}) + c(e^{rt}) = 0$$

$$(ar^2 + br + c)e^{rt} = 0$$

Since $e^{rt} \neq 0$, this implies we want to find values r which satisfy

$$ar^2 + br + c = 0$$

The fundamental theorem of algebra states that solving this equation will produce at least one complex root and 2 roots total counted for multiplicity. In this section, we will look at this case in which the equation produces two roots distinct $r_1, r_2 \in \mathbb{R}$. Thus we get two solutions,

$$y_1 = e^{r_1 t}$$

$$y_2 = e^{r_2 t}$$

We check linear independence with the Wronskian,

$$\begin{vmatrix} e^{r_1 t} & e^{r_2 t} \\ r_1 e^{r_1 t} & r_2 e^{r_2 t} \end{vmatrix} = r_2 e^{r_1} e^{r_2} - r_1 e^{r_1 t} e^{r_2 t} = (r_2 - r_1) e^{(r_1 + r_2)t} \neq 0$$

since the exponential function is never zero and r_1, r_2 are distinct. This gives us two linearly independent solutions that produce a basis for the set of solutions to this differential equation, thus our general solution is

$$y = c_1 e^{r_1 t} + c_2 e^{r_2 t}$$

where c_1, c_2 are undetermined coefficients to be determined by initial conditions.

Exercises

1. Solve the following homogeneous second-order differential equation with the given initial conditions:

$$\begin{aligned} y'' - 3y' + 2y &= 0, \\ y(0) = 1, \quad y'(0) &= 0. \end{aligned}$$

2. Solve the following homogeneous second-order differential equation with the given initial conditions:

$$\begin{aligned} y'' + 4y' + 4y &= 0, \\ y(0) = 0, \quad y'(0) &= 1. \end{aligned}$$

7.3 Complex Roots

We now look at cases of equations in the previous section for which the characteristic equation produces complex roots. However, a quick remark is needed first.

Definition 7.2 (Make this a remark somehow)

A polynomial of degree 2 with real coefficients can either have, 2 real, 2 complex, 1 repeated real or 1 repeated complex roots.

This means that any degree two polynomial cannot have one real root and one complex root. We will now look at cases for which we have two complex roots. Suppose we have a differential equation of the form

$$ay'' + by' + cy = 0$$

The previous section suggests we solve the quadratic equation

$$ar^2 + br + c = 0$$

to find r_1 and r_2 that produce solutions $y_1 = e^{r_1 t}$ and $y_2 = e^{r_2 t}$ to the differential equation. If

$$r_1 = a_1 + b_1 i$$

$$r_2 = a_2 + b_2 i$$

then our general solution becomes

$$y = c_1 e^{(a_1+b_1 i)t} + c_2 e^{(a_2+b_2 i)t}$$

However, certain cases prove it useful to find real solutions. In these cases we use Euler's Identity

$$e^{(a+bi)t} = e^{at}(cos(bt) + isin(bt))$$

. The power in this technique is that it produces two real solutions from a single complex solution. We will prove this now.

Proof. Suppose $y = u + iv$ is a complex solution to the second order homogeneous differential equation with constant coefficients

$$ay'' + by' + cy = 0$$

where u and v are real valued functions. We have

$$y' = u' + iv'$$

$$y'' = u'' + iv''$$

Substituting into our differential equation

$$a(u'' + iv'') + b(u' + iv') + c(u + iv) = 0$$

$$= (au'' + bu' + u) + i(av'' + bv' + cv) = 0$$

This suggests both the real and imaginary parts of this equation must be zero thus we have,

$$au'' + bu' + cu = 0$$

$$i(av'' + bv' + cv) = 0$$

This results suggests that both u and v are real solutions to the differential equation. Now if we let $u = cos(bt)$ and $v = sin(bt)$ we have obtained two real solutions to our differential equation from one complex solution. We still must have the $e^a t$ factor multiplied by $u + v$ and our two undetermined coefficients to be satisfied by initial conditions; therefore our general solution is

$$y = e^{at}(c_1 cos(bt) + c_2 sin(bt))$$

We check linear independence with the Wronskian

$$\left| \right|$$

From this result we see that for cases of complex roots only one root suffices to obtain a general solution.

7.4 Method of Reduction of Order

When our characteristic equations of second-order constant coefficient homogeneous equations results in repeated roots, we obtain only one solution. Therefore we look to develop a technique to find a second solution. We suggest a solution of the form

$$y(t) = v(t)y_1(t)$$

where $y_1(t)$ is the first solution found. Taking derivatives we have

$$y'(t) = v'(t)y_1(t) + v(t)y_1'(t)$$

$$y''(t) = v''(t)y_1(t) + 2v'(t)y_1'(t) + v(t)y_1''(t)$$

Substituting this into the differential equation to solve for the undetermined equation $v(t)$

$$ay'' + by' + cy = 0$$

$$a(v''y_1 + 2v'y_1' + vy_1'') + b(v'y_1 + v'y_1') + c(vy_1) = 0$$

$$av''y_1 + v'(2ay_1' + by_1) + v(ay_1'' + by_1' + cy_1) = 0$$

The third term is zero since y_1 is a solution to that differential equation which is the same as what we started out with. Therefore we have

$$av''y_1 + v'(2ay_1' + by_1) = 0$$

$$\frac{v''}{v'} = \frac{-(2ay_1' + by_1)}{ay_1}$$

We can solve this separable differential equation for v'

$$\int \frac{dv}{v} = \int (-\frac{2y_1'}{y_1} - \frac{b}{a})dt$$

This gives us the solution

$$v' = \frac{1}{y_1^2} C e^{-\int \frac{b}{a} dt}$$

We have kept the argument in the exponential in integral form as this method is generalizable to any differential equation in which we have one solution and require another for a general solution, however, in the case of constant coefficients the exponential will be $e^{bt/a}$.

Exercises

3. Given that $y_1(x) = e^x$ is a solution to the differential equation:

$$y'' - y' = 0,$$

use the method of reduction of order to find a second, linearly independent solution $y_2(x)$.

4. Suppose $y_1(x) = x$ is a solution to the differential equation:

$$x^2y'' - 3xy' + 3y = 0 \quad \text{for } x > 0.$$

Use the method of reduction of order to find a second solution $y_2(x)$ that is linearly independent of $y_1(x)$.

7.5 Variation of Parameters

Thus far we have developed techniques to solving homogeneous second order equations. We now turn our attention to finding methods to solve inhomogeneous equations.

$$ay'' + by' + cy = f(t)$$

We find solutions to inhomogeneous equations by adding a general solution to the corresponding inhomogeneous equation with a particular solution to the inhomogeneous equation.

$$y = y_g + y_p$$

Suppose we can find two linearly independent solutions to the corresponding homogeneous equation of (), y_1, y_2. The method of variation of parameters suggests we look for a particular solution of the form

$$y(t)_p = u(t)y_1(t) + u_2(t)y_2(t)$$

where $y_1(t), y_2(t)$ are solutions to the corresponding homogeneous equation and $u_1(t), u_2(t)$ are undetermined coefficients. Taking the derivative

$$y' = u_1'y_1 + u_1y_1' + u_2'y_2 + u_2y_2'$$

These calculations are made simpler if we set

$$u_1'y_1 + u_2'y_2 = 0$$

Therefore y' becomes

$$y' = u_1y_1' + u_2y_2'$$

Finding y''

$$y'' = u_1'y_1' + u_1y_1'' + u_2'y_2' + u_2y_2''$$

Substituting this into our differential equation

$$a(u_1'y_1' + u_1y_1'' + u_2'y_2' + u_2y_2'') + b(u_1y_1' + u_2y_2') + c(u_1y_1 + u_2y_2) = f(t)$$

$$u_1(ay_1'' + by_1' + cy_1) + u_2(ay_2'' + by_2' + cy_2) + au_1'y_1' + au_2'y_2' = f(t)$$

Since y_1 and y_2 are solutions to the corresponding homogeneous equation, the first two terms are zero. Thus, our equation reduces to

$$au_1'y_1' + au_2'y_2' = f(t)$$

We now have two equations and two unknowns.

$$u_1'y_1 + u_2'y_2 = 0$$

$$au_1'y_1' + au_2'y_2' = f(t)$$

From this we obtain

$$u_1' = \frac{-y_2f(t)/a}{y_1y_2' - y_1'y_2}$$

$$u_2' = \frac{y_1f(t)/a}{y_1y_2' - y_1'y_2}$$

We can integrate to find u_1 and u_2.

$$u_1 = \int \frac{-y_2 f(t)/a}{y_1 y_2' - y_1' y_2} dt$$

$$u_2 = \int \frac{y_1 f(t)/a}{y_1 y_2' - y_1' y_2} dt$$

You may notice that the argument in the denominator is the Wronksian thereby implying that if our solutions y_1, y_2 are not linearly independent, then when don't have the requisite information to form a general solution to the differential equation. It that case we must return to section 3.5 and find a second linearly independent equation via Method of Reduction of Order.

This provides us with our particular solution to the inhomogeneous differential equation

$$y_p = y_1 \int \frac{-y_2 f(t)/a}{y_1 y_2' - y_1' y_2} dt + y_2 \int \frac{y_1 f(t)/a}{y_1 y_2' - y_1' y_2} dx$$

We add this to the general solution to the corresponding homogeneous equation to obtain the full solution to the inhomogeneous equation.

$$y(x) = y_g + y_p$$

Example 7.3

Find the general solution to the differential equation

$$y'' - 3y' + 2y = 2e^x$$

The characteristic equation of the corresponding homogeneous equation is

$$r^2 - 3r + 2 = 0$$

This equation has roots $r_1 = 1$ and $r_2 = 2$. Thus the general solution to the corresponding homogeneous equation is

$$y_g = c_1 e^x + c_2 e^{2x}$$

We can use equation () compute the particular solution to the differential equation.

$$y_1 = e^x$$

$$y_2 = e^{2x}$$

$$y_1' = e^x$$

$$y_2' = 2e^{2x}$$

$$\begin{aligned}
y_p &= e^x \int \frac{-e^{2x} 2e^x}{e^x 2e^{2x} - e^{2x} e^x} dx + e^{2x} \int \frac{e^x 2e^x}{e^x 2e^{2x} - e^{2x} e^x} dx \\
&= e^x \int \frac{-2e^{3x}}{e^x 2e^{2x} - 2e^{3x}} dx + e^{2x} \int \frac{2e^{2x}}{e^x 2e^{2x} - 2e^{3x}} dx \\
&= e^x \int \frac{-2e^{3x}}{2e^{3x}} dx + e^{2x} \int \frac{2e^{2x}}{2e^{3x}} dx
\end{aligned}$$

$$
\begin{aligned}
&= e^x \int -1dx + e^{2x} \int e^{-x} dx \\
&= -e^x + e^{2x}(-e^{-x}) \\
&= -e^x - e^x \\
&= -2e^x
\end{aligned}
$$

Thus our particular solution is

$$y_p = -2e^x$$

Therefore our full solution to the differential equation is

$$y(x) = c_1 e^x + c_2 e^{2x} - 2e^x.$$

Exercises

5. Solve the non-homogeneous differential equation:

$$y'' + y = \sin(x),$$

using the method of variation of parameters. Assume that the complementary solution is given by:

$$y_c(x) = C_1 \cos(x) + C_2 \sin(x).$$

6. Use the method of variation of parameters to solve the non-homogeneous differential equation:

$$x^2 y'' - 3xy' + 3y = x^3,$$

for $x > 0$. Assume that the complementary solution is:

$$y_c(x) = C_1 x + C_2 x^3.$$

7.6 Method of Undetermined Coefficients

There are certain classes of inhomogeneous equations such that we can propose a solution form and algebraically solve for specifying parameters. Such classes usually involve equations of constant coefficients and inhomogeneous terms of familiar functions like exponentials or sinusoidals. As in the previous section, to find a full solution to an inhomogeneous equation we sum together a general solution to the corresponding homogeneous equation with a particular solution to the inhomogeneous equation. Suppose we have the second order inhomogeneous differential equation

$$a(t)y'' + b(t)y' + c(t)y = Acos(\omega t) + Bsin(\omega t)$$

We propose a particular solution of the form

$$y_p = X_1 cos(\omega t) + X_2 sin(\omega t)$$

Taking derivatives

$$
\begin{aligned}
y' &= -\omega X_1 sin(\omega t) + \omega X_2 cos(\omega t) \\
y'' &= -\omega^2 X_1 cos(\omega t) - \omega^2 X_2 sin(\omega t)
\end{aligned}
$$

Substituting into our differential equation

$$a(-\omega^2 X_1 cos(\omega t) - \omega^2 X_2 sin(\omega t)) + b(-\omega X_1 sin(\omega t) + \omega X_2 cos(\omega t) + c(X_1 cos(\omega t) + X_2 sin(\omega t))$$

$$= Acos(\omega t) + Bsin(\omega t)$$

We rearrange to get a single cosine and sine term on each side

$$(-a\omega^2 X_1 + b\omega X_2 + cX_1)cos(\omega t) + (-a\omega^2 X_2 - b\omega X_1 + cX_2)sin(\omega t)$$

$$= Acos(\omega t) + Bsin(\omega t)$$

From this it is apparent that

$$(-a\omega^2 + c)X_1 + b\omega X_2 = A$$

$$(-a\omega^2 + c)X_2 - b\omega X_1 = B$$

We can solve this system of equations to find

$$X_1 = \frac{aB - bA}{a^2\omega^2 + b^2\omega^2 - c}$$

$$X_2 = \frac{aA + bB}{a^2\omega^2 + b^2\omega^2 - c}$$

Thus our particular solution is

$$y_p = \frac{aB - bA}{a^2\omega^2 + b^2\omega^2 - c} cos(\omega t) + \frac{aA + bB}{a^2\omega^2 + b^2\omega^2 - c} sin(\omega t)$$

We add this to the general solution to the corresponding homogeneous equation to obtain the full solution to the inhomogeneous equation.

$$y(x) = y_g + y_p$$

where $y(g)$ will dependent on the form of our differential equation. Now suppose we have an inhomogeneous equation of the form

$$a(t)y'' + b(t)y' + c(t)y = e^{\alpha t}$$

We propose a particular solution of the form

$$y_p = Xe^{\alpha t}$$

Taking derivatives

$$y' = \alpha X e^{\alpha t}$$

$$y'' = \alpha^2 X e^{\alpha t}$$

Substituting into our differential equation

$$a\alpha^2 X e^{\alpha t} + b\alpha X e^{\alpha t} + cXe^{\alpha t} = e^{\alpha t}$$

$$X(a\alpha^2 + b\alpha + c) = 1$$

$$X = \frac{1}{a\alpha^2 + b\alpha + c}$$

Thus our particular solution is

$$y_p = \frac{1}{a\alpha^2 + b\alpha + c} e^{\alpha t}$$

We add this to the general solution to the corresponding homogeneous equation to obtain the full solution to the inhomogeneous equation.

$$y(x) = y_g + y_p$$

where $y(g)$ will once again be dependent on the form of our differential equation.

Example 7.4

Find the general solution to the differential equation

$$y'' + 4y = 2cos(2t)$$

Since we have a sinusoidal inhomogeneous term, we propose a particular solution of the form

$$y_p = X_1 cos(2t) + X_2 sin(2t)$$

Taking derivatives

$$y' = -2X_1 sin(2t) + 2X_2 cos(2t)$$
$$y'' = -4X_1 cos(2t) - 4X_2 sin(2t)$$

Substituting into our differential equation

$$-4X_1 cos(2t) - 4X_2 sin(2t) + 4X_1 cos(2t) + 4X_2 sin(2t) = 2cos(2t)$$
$$= 2cos(2t)$$

This implies that $X_1 = 1/2$ and $X_2 = 0$. Therefore our particular solution is

$$y_p = \frac{1}{2} cos(2t)$$

The general solution to the corresponding homogeneous equation is

$$y_g = c_1 cos(2t) + c_2 sin(2t)$$

Thus our general solution to the differential equation is

$$y(x) = c_1 cos(2t) + c_2 sin(2t) + \frac{1}{2} cos(2t)$$

Example 7.5

Find the general solution to the differential equation

$$y'' + 4y = 2e^{2t}$$

Since we have an exponential inhomogeneous term, we propose a particular solution of the form

$$y_p = Xe^{2t}$$

Taking derivatives

$$y' = 2Xe^{2t}$$
$$y'' = 4Xe^{2t}$$

Substituting into our differential equation

$$4Xe^{2t} + 4Xe^{2t} = 2e^{2t}$$
$$= 2e^{2t}$$

This implies that $X = 1/4$. Therefore our particular solution is

$$y_p = \frac{1}{4}e^{2t}$$

The general solution to the corresponding homogeneous equation is

$$y_g = c_1 cos(2t) + c_2 sin(2t)$$

Thus our general solution to the differential equation is

$$y(x) = c_1 cos(2t) + c_2 sin(2t) + \frac{1}{4}e^{2t}$$

Exercises

7. Solve the following non-homogeneous differential equation using the method of undetermined coefficients:
$$y'' - 3y' + 2y = e^{2x},$$
subject to the initial conditions $y(0) = 0$ and $y'(0) = 1$.

8. Solve the following non-homogeneous differential equation using the method of undetermined coefficients:
$$y'' + y = \sin(x),$$
subject to the initial conditions $y(0) = 1$ and $y'(0) = 0$.

7.7 Existence and Uniqueness

Chapter 8

Eigenvalues and Eigenvectors

8.1 Definition of Eigenvectors and Eigenvalues

We look to examine the behavior of linear transformations in which a vector space maps to itself. We denote $T \in \mathcal{L}(V)$ as the linear transformation $T : V \to V$ where $\mathcal{L}(V)$ is the set of all operators $\mathcal{L}(V, V)$. In order to perform operations on a subspace U of V, we look to define a special class of operators that maps U to itself.

Definition 8.1 (Invariant Subspaces)

Suppose U is a subspace of V. U is *invariant* for a given transformation $T : V \to V$, if $Tu \in U$ for any $u \in U$.

Vectors that constitute invariant subspaces and their change under T are specially defined.

Definition 8.2 (Eigenvalues and Eigenvectors)

Suppose $U \in V$ is invariant under T and u is a nonzero vector in U. Then,

$$Tu = \lambda u$$

where $\lambda \in \mathbb{F}$ is the *eigenvalue* of T and u is it's corresponding *eigenvector*.

It is important to note that for a given eigenvalue there may be multiple eigenvectors. The dimension of the subspace the eigenvectors for a given eigenvalue span (called the *eigenspace*) corresponds to the number of eigenvectors for the given eigenvalue.
Rewriting the () gives,

$$(T - \lambda I)u = 0.$$

By construction it is apparent that the set of eigenvectors of T is equal to $null(T - \lambda I)$. Since we have a nonzero vector mapping to zero, one can see that λ is an eigenvalue of T if and only if $T - \lambda I$ is not injective. And, since this gives a noninvertible square matrix by SOME THEOREM λ is an eigenvalue of T if and only if $T - \lambda I$ is not surjective as well.
By SOME THEOREM, the determinant of a noninvertible matrix is zero. This property allows us to solve for the value of λ.

Theorem 8.3 ()

Suppose $\lambda_1, \lambda_2, ..., \lambda_m$ are distinct eigenvalues of $T : V \to V$ corresponding to distinct eigenvectors $u_1, u_2, ..., u_m$. Then the eigenvectors $u_1, u_2, ..., u_k$ are linearly independent.

Proof. We proceed by contradiction. Suppose $u_1, u_2, ..., u_m$ are linearly dependent. Choose k to be the smallest integer such that,

$$u_k \in span\{u_1, u_2, ..., u_{k-1}\}.$$

Therefore u_k can be written as,

$$u_k = a_1 u_1 + a_2 u_2 + ... + a_{k-1} u_{k-1}.$$

Take the transformation T of both sides of the equation,

$$Tu_k = T(a_1 u_1 + a_2 u_2 + ... + a_{k-1} u_{k-1})$$

$$\lambda_k u_k = a_1 \lambda_1 u_1 + a_2 \lambda_2 u_2 + ... + a_{k-1} \lambda_{k-1} u_{k-1}.$$

Multiply both sides of () by λ_k and subtract () to obtain,

$$0 = a_1(\lambda_k - \lambda_1)u_1 + ... + a_{k-1}(\lambda_k - \lambda k - 1)u_{k-1}.$$

By construction, this implies that $a_i = 0$ for $i \in (1, k-1)$ since the eigenvectors are linearly independent and the eigenvalues are distinct. However, this implies $u_k = 0$, a contradiction since we don't consider $\vec{0}$ and eigenvector.

Corollary 8.4: There are at most n distinct eigenvalues for each operator on an n-dimensional vector space.

Therefore, suppose we have $T \in \mathcal{L}(V)$ with n distinct eigenvalues, then it follows that T has n distinct eigenvectors. From the previous theorem the set of eigenvectors to T must be linearly independent therefore $n \leq dim(V)$.

8.2 Computing Eigenvalues and Eigenvectors

We look to develop a method to solve for the eigenvalues and eigenvectors of some transformation $T \in \mathcal{L}(V, V)$. Suppose $T(x) = Ax$ and $n = dim(V)$, this implies that A is $n \times n$. We look for $\lambda \in \mathbf{F}$ that satisfies

$$Ax = \lambda x.$$

Right multiplying each side by the identity matrix $n \times n$ identity matrix I_n gives

$$Ax = \lambda I x.$$

Solving to isolate x produces the homogeneous equation

$$(A - \lambda I)x = 0.$$

From the previous section we know the eigenvectors of A span $null(A)$. Therefore, we look for non-trivial vectors x that solve $(A - \lambda I)$. This implies that $(A - \lambda I)$ must be non-invertible. We use the property that for non-invertible matrices the determinant is zero to solve for λ.

$$det(A - \lambda I) = 0$$

Computing the determinant of $(A - \lambda I)$ produces a polynomial $P_k(\lambda)$ where $k \leq n$.

$$P_k(\lambda) = 0$$

Solving for the roots of $P_k(\lambda)$ finds the desired eigenvalues for A. For a polynomial of degree $k \leq n$, there will be at most k eigenvalues. We substitute each computed eigenvalue into $(A - \lambda I)x = 0$ to solve for vectors x that span $null(A)$. Each x is an eigenvector of A. The space spanned by each eigenvalue λ is called the *eigenspace* of λ.

Example 8.5

Find the eigenvalues and eigenvectors of

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

We wish to find λ that satisfy,

$$\begin{bmatrix} 1 & 4 & 3 \\ 4 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} x = \lambda x$$

. Algebraically rearranging,

$$(\begin{bmatrix} 1 & 4 & 3 \\ 4 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} - \lambda I)x = 0$$

$$\begin{bmatrix} 1-\lambda & 4 & 3 \\ 4 & 1-\lambda & 0 \\ 3 & 0 & 1-\lambda \end{bmatrix} x = 0$$

Solving $det(A - \lambda I) = 0$,

$$\begin{vmatrix} 1-\lambda & 4 & 3 \\ 4 & 1-\lambda & 0 \\ 3 & 0 & 1-\lambda \end{vmatrix} = (1-\lambda)((1-\lambda)^2 - 0) - 4(4(1-\lambda) - 0) + 3(0 - 3(1-\lambda))$$

$$= (1-\lambda)^3 - 25(1-\lambda) = (1-\lambda)((1-\lambda)^2 - 25)$$

$$= (1-\lambda)(\lambda^2 - 2\lambda - 24) = (1-\lambda)(6-\lambda)(4+\lambda) = 0$$

Therefore our eigenvalues are $\lambda = 1, 6$ and -4. We substitute each eigenvalue into $(A - \lambda I)x = 0$ to find the eigenvectors of A. For $\lambda = 1$,

$$(A - 1(I))x = \begin{bmatrix} 0 & 4 & 3 \\ 4 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix} x = 0$$

By Gaussian-Jordan Reduction we get,

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

This means our eigenvector $\vec{x}$ is,

$$\vec{x} = x_3 \begin{bmatrix} 0 \\ -3/4 \\ 1 \end{bmatrix}$$

Repeating the same process for $\lambda = 6$,

$$(A - 6I) = \begin{bmatrix} -5 & 4 & 3 \\ 4 & -5 & 0 \\ 3 & 0 & -5 \end{bmatrix}.$$

By Gauss-Jordan Reduction we get,

$$\begin{bmatrix} 1 & 0 & -5/3 \\ 0 & 1 & -4/3 \\ 0 & 0 & 0 \end{bmatrix}.$$

Therefore our eigenvector is,

$$\vec{x} = x_3 \begin{bmatrix} 5/3 \\ 4/3 \\ 1 \end{bmatrix}$$

Lastly for $\lambda = 4$,

$$(A + 4I) = \begin{bmatrix} 5 & 4 & 3 \\ 4 & 5 & 0 \\ 3 & 0 & 5 \end{bmatrix}$$

. Gauss-Jordan Reduction gives us,

$$\begin{bmatrix} 1 & 0 & 5/3 \\ 0 & 1 & -4/3 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus our eigenvector is,

$$\vec{x} = x_3 \begin{bmatrix} -5/3 \\ 4/3 \\ 1 \end{bmatrix}.$$

So our set of eigenvectors for A is,

$$\left\{ \begin{bmatrix} 0 \\ -3/4 \\ 1 \end{bmatrix}, \begin{bmatrix} 5/3 \\ 4/3 \\ 1 \end{bmatrix}, \begin{bmatrix} -5/3 \\ 4/3 \\ 1 \end{bmatrix} \right\}.$$

Theorem 8.6 ()

Suppose A is an upper triangular matrix. Then the eigenvalues of A are the the entries along the diagonal. Similarly, if A were lower triangular the same result holds.

This theorem follows from the determinant of a triangular matrix being the product of the diagonal entries. Therefore, if we can subtract some λ such that one of the entries becomes zero, then the matrix determinant is zero and the value of that λ satisfies $Ax = \lambda x$.

Exercises

8.3 Diagonalization

A diagonal matrix is a matrix consisting of only diagonal entries. The diagonal matrix for an operator $T \in \mathcal{L}(V)$ is

$$\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

if,

$$\begin{aligned} Tv_1 &= \lambda_1 v_1 \\ Tv_2 &= \lambda_2 v_2 \\ &\vdots \\ Tv_n &= \lambda_n v_n. \end{aligned}$$

This suggests λ_i are the eigenvalues of T and v_i are the corresponding eigenvectors. What this tells us is that an arbitary operator has some diagonal matrix consisting of eigenvalues with respect to some basis of eigenvectors. Our goal is to find the basis of eigenvectors for some operator T such that T is diagonal.

Theorem 8.7 (The Diagonalization Theorem)

An operator $T \in \mathcal{L}(V)$ is diagonalizable if there exists a basis of V consisting of eigenvectors of T. In matrix form, suppose $T(x) = Ax$ for some $n \times n$ matrix A. Then A is diagonalizable if there exists an invertible matrix P with the eigenvectors of A as columns such that

$$D = P^{-1}AP$$

is a diagonal matrix.

This theorem tells us that a $n \times n$ matrix A is diagonalizable if the eigenvectors of A form a basis for $\mathbf{R}^n$.

Proof. Let A be a $n \times n$ matrix. Suppose P is a $n \times n$ matrix with column vectors v_i and D be a $n \times n$ diagonal matrix with diagonal entries λ_i. Then,

$$AP = A \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$$

and,

$$PD = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_n v_n \end{bmatrix}.$$

Suppose λ_i are the eigenvalues of A and v_i are the corresponding eigenvectors. Then we obtain the following result,

$$AP = PD.$$

Since P is invertible, we can multiply each side by P^{-1} to obtain,

$$A = PDP^{-1}.$$

Therefore A is diagonalizable. We have shown that a matrix A is diagonalizable if there exists an invertible matrix P such that $D = P^{-1}AP$ is a diagonal matrix. We will now develop the methodological steps to diagonalize a matrix A.

1. Find the eigenvalues of A by solving $det(A - \lambda I) = 0$.
2. Find the corresponding eigenvectors of A by solving $(A - \lambda I)v = 0$.
3. Form the matrix P with the eigenvectors of A as columns.
4. Form the matrix D with the eigenvalues of A as diagonal entries.

This will give us the diagonalization of A as $A = PDP^{-1}$.

Example 8.8

Diagonalize the matrix $A = \begin{bmatrix} 1 & 4 & 3 \\ 4 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}$.

We have already found the eigenvalues and eigenvectors of A in the previous section. The eigenvalues of A are $\lambda = 1, 6, -4$ with corresponding eigenvectors

$$\begin{bmatrix} 0 \\ -3/4 \\ 1 \end{bmatrix}, \begin{bmatrix} 5/3 \\ 4/3 \\ 1 \end{bmatrix}, \begin{bmatrix} -5/3 \\ 4/3 \\ 1 \end{bmatrix}.$$

Therefore the matrix P is,

$$P = \begin{bmatrix} 0 & 5/3 & -5/3 \\ -3/4 & 4/3 & 4/3 \\ 1 & 1 & 1 \end{bmatrix}.$$

The matrix D is,

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & -4 \end{bmatrix}.$$

Therefore the diagonalization of A is,

$$\begin{bmatrix} 1 & 4 & 3 \\ 4 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 5/3 & -5/3 \\ -3/4 & 4/3 & 4/3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} \frac{4}{125} & -\frac{3}{100} & \frac{3}{100} \\ \frac{3}{100} & \frac{2}{25} & \frac{2}{25} \\ -\frac{3}{100} & \frac{3}{100} & \frac{3}{100} \end{bmatrix}$$

Here I have skipped the calculation of P^{-1} as it is quite tedious and one can use methods from section () to compute it.

8.4 Spectral Theorem

Definition 8.9 (Self Adjoint Operators)

An operator $T \in \mathcal{L}(V)$ is *self-adjoint* if $T = T^*$.

Definition 8.10 (Hermitian Matrices)

A matrix A is *Hermitian* if $A = A^*$. If A is real, then $A = A^T$. In this case, A is symmetric.

8.5 Generalized Eigenvectors

From section () we know that an $n \times n$ matrix A with n distinct eigenvalues λ_i has n corresponding eigenvectors $\vec{v_i}$ which form a basis for $\mathbf{R}^n$. In this case each λ_i has algebraic and geometric multiplicities both equal to 1. However consider the matrix

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

It has one eigenvalue $\lambda = 1$ which has one corresponding eigenvector

$$\vec{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

We can see the eigenvectors of A do not form a basis for $\mathbf{R}^2$. The algebraic multiplicity of $\lambda = 1$ is 2, however its geometric multiplicity is only 1. Therefore we see that matrices with a set of eigenvectors which do not form a basis for the column space of A display an inequality between the geometric and algebraic multiplicities. We can generalize the above example to the following definition.

Definition 8.11 (Defective Matrices)

A $n \times n$ matrix is defective if the sum of its eigenvalues' algebraic multiplicities μ_a and the sum of geometric multiplicities μ_g has the property

$$\mu_a > \mu_g$$

The eigenvectors of defective matrices do not form a linearly independent basis for $\mathbf{R}^n$. This implies that such matrices are *non-diagonalizable*. However in cases for which we wish to diagonalize a matrix, compute a matrix exponential or find a basis consisting of eigenvectors, we seek a method to resolve this issue.

Definition 8.12 (Generalized Eigenvectors)

For a matrix A, some $\lambda \in \mathbf{F}$ is a *level j generalized eigenvector* if it satisfies

$$(A - \lambda I)^j x = 0$$

Theorem 8.13 ()

Suppose λ is an eigenvalue of A with multiplicity m. Then

$$(A - \lambda I)^j = 0$$

with $j > m$ has the same solution space as

$$(A - \lambda I)^m = 0$$

8.6 Matrix Exponentials

Suppose we have the linear system of differential equations

$$\frac{dx}{dt} = Ax.$$

Our study of differential equations suggests a solution of the form

$$x = Ce^{At}.$$

Now we are presented with the problem of computing the exponential of a matrix. Let's take the series expansion of e^x.

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + ... + \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Choosing instead to expand the matrix A we get,

$$e^A = 1 + A + \frac{A^2}{2!} + \frac{A^3}{3!} + ... + \sum_{n=0}^{\infty} \frac{A^n}{n!}.$$

We know how to compute powers of matrices however doing it in practice and taking high enough powers in the series to get a sufficient solution is quite futile (and an approximation on top of that). There are two methods building on concepts introduced earlier this chapter which allows us to exactly compute the matrix exponential, diagonalization and generalized eigenvectors. We will begin with the former which is the preferrable route.

Suppose there exists a matrix P and a diagonal matrix D such that,

$$A = PDP^{-1}$$

for some matrix A. We can substitute this into () to obtain,

$$e^A = 1 + (PDP^{-1}) + \frac{(PDP^{-1})^2}{2!} + \frac{(PDP^{-1})^3}{3!} + ... + \sum_{n=0}^{\infty} \frac{(PDP^{-1})^n}{n!}.$$

From section 4.3, we know,

$$A^n = PD^nP^{-1}$$

. Therefore, our series expansion of e^A becomes,

$$e^A = 1 + PDP^{-1} + \frac{PD^2P^{-1}}{2!} + \frac{PD^3P^{-1}}{3!} + ... + \sum_{n=0}^{\infty} \frac{PD^nP^{-1}}{n!}.$$

Factoring out P and P^{-1} we obtain,

$$e^A = P(1 + D + \frac{D^2}{2!} + \frac{D^3}{3!} + ... + \sum_{n=0}^{\infty} \frac{D^n}{n!})P^{-1}.$$

The expansion in the parenthesis is the matrix exponential of the diagonal matrix D, therefore,

$$e^A = Pe^DP^{-1}.$$

Since the exponential of a diagonal matrix is the matrix consisting of the diagonal entries exponetiated, equation () allows us to easily compute e^A.

However, in cases for which A is non-diagonalizable we must find an alternative method to solve e^A. We will use generalized eigenvectors to do this.
Generalized Eigenvectors tell us that for some λ_j, the matrix $(A - \lambda_j I)^j$ is zero. This implies that for all $m > j$, $(A - \lambda_j I)^m = 0$. Therefore, suppose λ is a jth level eigenvector. If we write $e^{(A-\lambda I)t}$ as a series expansion,

$$e^{(A-\lambda I)v} = 1 + (A - \lambda I)v + \frac{(A - \lambda I)^2 v^2}{2!} + \frac{(A - \lambda I)^3 v^3}{3!} + ... + \sum_{n=0}^{j-1} \frac{(A - \lambda I)^n v^n}{n!},$$

this series terminates at A^{j-1}. This is exactly what we are looking for in order to compute e^A.

8.7 The Fundamental Solution of a Matrix

Suppose we have a system of coupled differential equations described by:

$$
\begin{aligned}
x_1' &= a_{11}x_1 + \cdots + a_{1n}x_n \\
x_2' &= a_{21}x_1 + \cdots + a_{2n}x_n \\
&\vdots \\
x_n' &= a_{n1}x_1 + \cdots + a_{nn}x_n
\end{aligned}
$$

Where $x_1, ..., x_n$ are functions of t with derivatives $x_1', ..., x_n'$ and a_{ij} are constants.

$$
\begin{aligned}
A &= \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \\
\vec{x}(t) &= \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \\
\vec{x'}(t) &= \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix} \\
A\vec{x}(t) &= \vec{x'}(t)
\end{aligned}
$$

We solve this equation by finding an $\vec{x}(t)$ that satisfies it on some interval of t.

Definition 8.14 (Fundamental Solution of a Matrix)

For an $n \times n$ matrix A, there exists a set of n linearly independent functions $\vec{x_1}(t), ..., \vec{x_n}(t)$ which constitute an n-dimensional basis for the vector space of all solutions of A. We call this set of functions the **fundamental solution of the matrix** A.

Suppose we have the system of uncoupled differential equations

$$x_1'(t) = ax_1(t)$$

$$x_2'(t) = bx_2(t)$$

for some constants a and b. This can be written in matrix form as

$$\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} x_1'(t) \\ x_2'(t) \end{bmatrix}$$

These equations suggest the solutions to this system of differential equations are

$$x_1(t) = c_1 e^{at}$$
$$x_2(t) = c_2 e^{bt}$$

From this we suggest the solution to any linear system of differential equations $A\vec{x} = \vec{x'}$ is of the form

$$\vec{x}(t) = \vec{v}e^{\lambda t}.$$

Taking the derivative $\vec{x'}(t)$

$$\vec{x'}(t) = \lambda \vec{v} e^{\lambda t}$$

Taking equation () once more and multiplying each side by A

$$A\vec{x}(t) = A\vec{v}e^{\lambda t}$$

The left sides of equations () and () are our differential equation thus our right sides must equal.

$$A\vec{v}e^{\lambda t} = \lambda \vec{v} e^{\lambda t}$$

This suggests that vectors v and scalars λ which satisfy this system of differential equations are eigenvectors and eigenvalues of the matrix A.

Example 8.15

$$\frac{dx}{dt} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} x$$

This matrix gives the eigenvalues $\lambda = 3, -1$ corresponding to the eigenvectors

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

Therefore our solutions to the systems of differential equations are

$$\{c_1 \vec{x_1}(t) = e^{3t} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, c_2 \vec{v_2}(t) = e^{-t} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \}$$

If we have vector solutions $\vec{x_1}(t), \vec{x_2}(t) \ldots \vec{x_n}(t)$ to the system of differential equations $A\vec{x} = \vec{x'}$ then the matrix $\vec{X}(t)$ with columns $\vec{x_1}(t), \vec{x_2}(t) \ldots \vec{x_n}(t)$ is the fundamental solution of the matrix A. This matrix $\vec{X}(t)$ is the linear combination of vector solutions to the system of differential equations.

$$\vec{X}(t) = \vec{x_1}(t) + \vec{x_2}(t) + \ldots + \vec{x_n}(t)$$
$$\vec{X}(t) = \begin{bmatrix} \vec{x_1}(t) & \vec{x_2}(t) & \ldots & \vec{x_n}(t) \end{bmatrix}$$

Therefore our fundamental solution $\vec{X}(t)$ to example 4.14 is,

$$\vec{X}(t) = \begin{bmatrix} e^{3t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} & e^{-t} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \end{bmatrix}.$$